Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Genome-wide association study of face shape and appearance identifies genes of clinical and evolutionary significance

Martin P. Evison

Martin_evison@hotmail.com

Newcastle University

Marco A. Guimarães

University of São Paulo

Edna S.M. Iwamura

Federal University of São Paulo

Rebecca Darlay

Newcastle University

Ian L. Dryden

The University of Nottingham

Richard W. Vorder Bruegge

ForensicVB LLC

Heather J. Cordell

Newcastle University

Article

Keywords: GWAS, Face shape and appearance, Iris colour, Human evolution, HARES, Dysmorphias

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11394166/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Genome-wide association study of face shape and appearance identifies genes of clinical and evolutionary significance

Martin P. Evison[1*]

Marco A. Guimarães[2]

Edna S.M. Iwamura[3]

Rebecca Darlay[1]

Ian L. Dryden[4]

Richard W. Vorder Bruegge[5]

Heather J. Cordell[1]

[1] Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom

[2] Department of Pathology and Forensic Medicine, University of São Paulo Faculty of Medicine Ribeirão Preto, Ribeirão Preto, Brazil

[3] Department of Pathology, Escola Paulista de Medicina, Federal University of São Paulo, São Paulo, Brazil

[4] School of Mathematical Sciences, The University of Nottingham, Nottingham, United Kingdom

[5] ForensicVB LLC, Fredericksburg, Virginia, United States of America

*Corresponding author. Email: martin_evison@hotmail.com.

Abstract

This paper presents a genome-wide association study of face shape and appearance in healthy adults of diverse age, sex and ancestry. The data consisted of precise 3D and RGB colour measurements collected from 378 subjects genotyped at 660K single nucleotide polymorphisms. Following quality control and imputation, 36 3D distances and 18 RGB colour values were chosen for GWAS, based on utility for comparison with previous studies and questions in developmental and evolutionary biology. Analysis was applied across the autosomes using GCTA.

Visualisation performed in LocusZoom showed associations with 32 3D dimensions, many of which were associated with more than one marker. Of a total of 145 associations, nine were of genome-wide significance. Many genes detected were associated with known craniofacial traits and dysmorphias. Associations of *HERC2* and *ZNF804B* with iris pigmentation were confirmed at genome-wide significance levels. Other results suggest associations relating to the obstetrical dilemma, the length of the earlobes and changes accompanying the evolution facial morphology in humans.

The findings support the hypothesis that normal face shape variation may have evolved to signal genetic fitness and to facilitate individual identification socially, and that facial dysmorphias may result from restricted growth or overgrowth during development of genetically controlled traits.

Keywords

GWAS

Face shape and appearance

Iris colour

Human evolution

HARES

Dysmorphias

Introduction

The face and head comprise the soft and hard tissue framework housing the organs of sight, smell, taste, balance and hearing, located adjacent to the brain itself. They permit us to communicate, both vocally and via non-verbal expressions [1,2].

Morphogenesis appears conserved amongst vertebrates, whereby species-specific paths lead to assembly of the 3D face in coordination with the nervous system, sensory organs and brain [3-5]. Postnatal development involves remodelling of bone at specific sites on the growing face [6].

Recent efforts to identify genes involved in the development of face shape have relied on Genome-Wide Association Studies (GWAS) using large samples and panels of single nucleotide polymorphisms (SNPs), supported by imputation onto reference genomes. Early studies confirmed relationships with markers associated with non-syndromic cleft lip with or without cleft palate (NSCL/P) and other craniofacial syndromes, and with craniofacial development in general [7-9]. Datasets have been combined and reanalysed using methodologically uniform approaches [10-12], and extended to confirm enrichment in murine facial tissues and cranial neural crest cells [13-15]. Iris, skin and hair pigmentation have been the subject of further GWAS [16,17].

Many samples are skewed toward juveniles, however, and most have assigned individual phenotype by Procrustes superimposition rather than by direct measurement, leaving open the question of to what extent the geometry used has fidelity with real points on the living face.

In this study, measurements of 30 3D anthropometric points and 18 RGB colour values collected directly from 3D biometric images were applied in a GWAS of an ancestrally diverse sample of 378 healthy adults using a 660K SNP panel expanded to 11M SNPs on imputation. Potential associations detected using GCTA [18] were analysed in LocusZoom [19] and examined in relation to genes known to be involved in craniofacial development and dysmorphias, and in the genetics of face shape.

Results

GWAS analysis

GWAS analysis of 36 distance measurements was completed on 378 subjects (212 males and 166 females) at 552,612 SNPs and of 18 RGB colour values on 258 subjects (164 males and 94 females) at 552,580 SNPs. In both cases, the number of SNPs was increased to 10,960,275 on imputation.

All associations detected above -log10(p)=6.000 are given in Supplementary Files S1 and S2 online, and above genome-wide significance of -log10(p)=7.301 or $p=5\times10^{-8}$ in Tables 1 and 2. Descriptions of the anthropometric landmarks used are given in Table 3 and depicted in Fig. 4. 3D dimensions showing associations of above genome-wide significance are illustrated in Supplementary Figure 1 online.

Results for all 216 GWAS can be accessed online via LocusZoom (my.locuszoom.org), searchable by the study name (Evison), PMID and GWAS label ("Face GWAS" + "g" for genotyped or "i" for imputed + "as" if adjusted for age and sex). These show Manhattan and QQ plots and genomic control (λ) values for each GWAS. The chromosome number and position, rsID, and nearest genes are given. LocusZoom region plots showing linkage disequilibrium (LD) can be accessed by clicking on the desired marker in the Manhattan plot. The RGB hair colour analyses consistently yielded highly inflated probabilities (see LocusZoom plots) and were therefore discarded.

Associations with 3D distance measurements in the facial midline

A number of associations of genome-wide significance were found in the facial midline, namely *KCNQ1-AS1*, *LINC00237*, *KRI1* and *AC078980.1* with g-se, *CNTN1* with se-ls, and *AC135048.4* and *HSD3B7* with g-pg and se-pg.

KCNQ1-AS1 is an antisense RNA gene located near *CDKN1C*, which is associated with cerebral cortex development, Beckwith-Wiedemann syndrome [20], Silver-Russell syndrome [21] and IMAGe syndrome [21], each of which may incorporate a protruding forehead and triangular face [22,23]. The 11p15.5 region, incorporating *KCNQ1* and *CDKN1C*, appears to underlie overgrowth and growth-retardation syndromes involving imprinting and epigenetic regulation [23]. *LINC00237* is a lncRNA gene associated with MOMO (macrosomia, obesity, macrocephaly and ocular abnormalities) syndrome [24-26]. *KRI1* is adjacent to *ATG4D*, variants of which appear be involved in a neurodevelopmental disorder affecting speech development and incorporating facial characteristics including a depressed nasal bridge [27]. *AC078980.1* is a lncRNA gene located near microRNA gene *MIR4445*, and within 400Kb of two *DPPA* genes associated with developmental pluripotency and early embryonic development [28]. Associations of high to genome-wide

significance for these markers were also evident in the data adjusted for age and sex. Those with *AC078980.1* were supported by associations in the genotyped data and by evidence of LD in the LocusZoom region plots (Fig. 1).

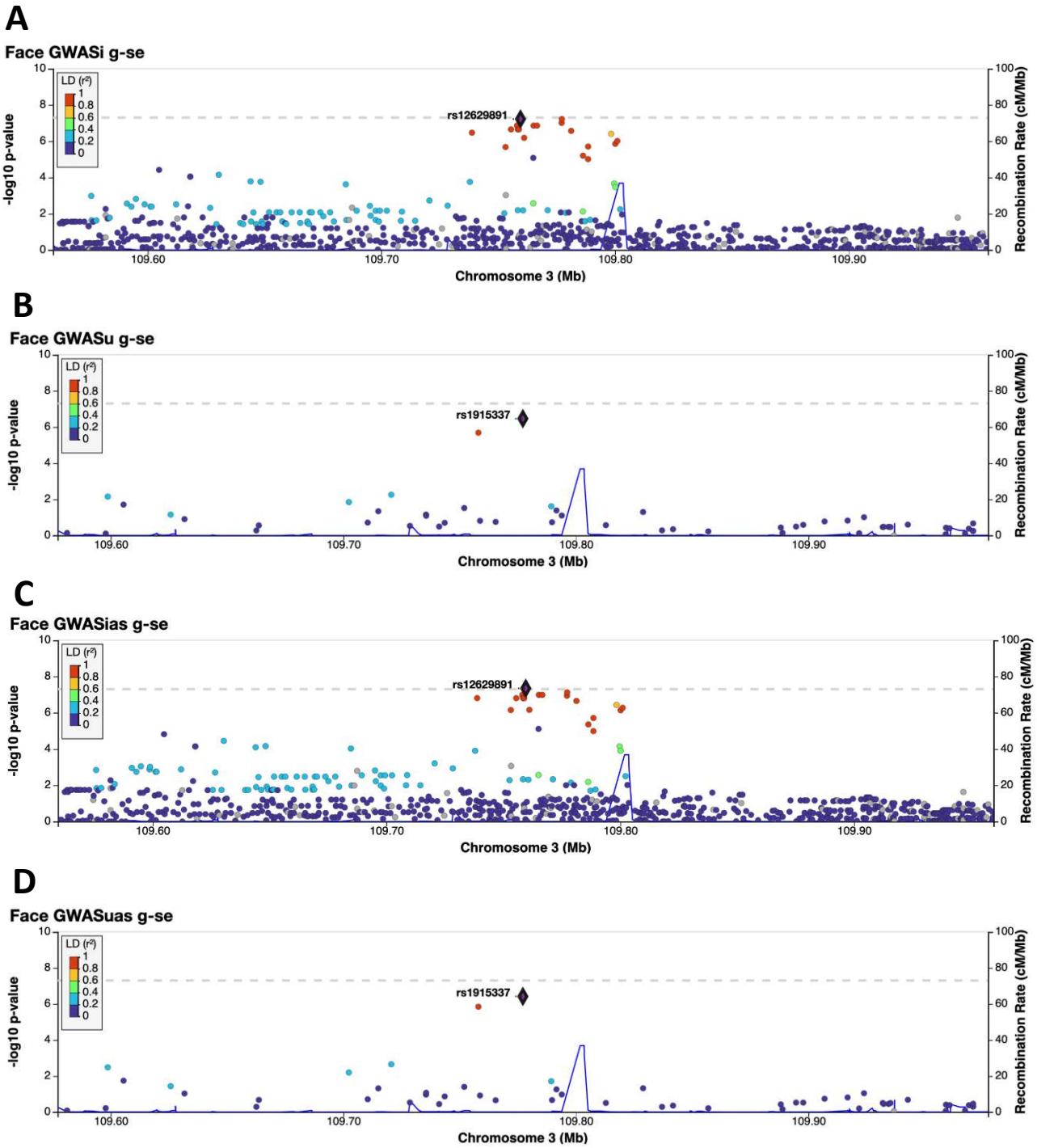
ANTXR1, *CECR2*, *MFSD14C* (*NUTMG2*) and *LINC01924* were among other genes and markers showing associations with g-se. *ANTXR1* is involved in GAPO (growth retardation, alopecia, pseudoanodontia, and optic atrophy) syndrome, symptoms of which include a high and bossed forehead [29,30]. *CECR2* is also associated with craniofacial dysmorphias [31,32], including cat eye syndrome [33]. *MFSD14C* (*NUTMG2*) may affect *PTCH1* expression leading to relevant dysmorphias [34] and *LINC01924* (*SERPIN*) appears to affect hard tissue formation [35]. These associations were reproduced in the adjusted (for age and sex) data.

Associations with other dimensions of the midface seemed to implicate a further shared set of genes or markers, *LINC01447* with g-ls, se-ls and se-pg; *LINC02671* with g-ls, g-pg and se-ls; *CNTN1* with g-ls, se-ls and se-prn, and *UNC80* with g-sl and g-pg. *LINC01447* is adjacent to *PKD1L1*, coding for a large membrane protein apparently involved in left-right symmetry and patterning [36]. *LINC02671* is near *CTNNA3*, a catenin gene involved in cell-cell adherence, which may be expressed in early brain development [37]. *CNTN1* codes for neural adhesion protein contactin-1 and is associated with both craniofacial and neurological disorders, most specifically Compton-North congenital myopathy [38,39]. *UNC80* codes for a protein important in sodium-leak channel functioning in the nervous system, mutations in which may lead to a high palate, tented upper lip and encephalopathy [40,41]. The association of *CNTN1* with se-ls was of genome-wide significance. This association and that of *LINC01447* with g-ls were reproduced in the adjusted data.

HSD3B7, found associated with g-pg and se-pg at genome-wide significance in the adjusted data only, is involved in bile acid synthesis. Rare mutations in this gene have recently been shown to result in craniosynostosis, amongst a range of other symptoms [42].

An association of *SLC41A2* with se-ls distance was evident in both imputed and genotyped data. *SLC41A2* is located near *CHST11*, which has been found essential for craniofacial development in zebrafish [43].

Fig. 1: LocusZoom region plots for the g-se 3D distance near *AC078980.1* (also online).



Images of online LocusZoom region plots for the g-se 3D distance on chromosome 3 near *AC078980.1* showing adjacent SNPs by position and -log10(p) value coded by degree of linkage disequilibrium (r^2) indexed to labelled lead SNP; *AC078980.1* is located at 109,648,107 to 109,810,861 bp: **A** imputed data. **B** genotyped data. **C** imputed data adjusted for age and sex. **D** genotyped data adjusted for age and sex

Associations with 3D distance measurements of the upper face

Of the major bilateral dimensions of the upper face, *AC008798.2*/*CCNE1* was associated with forehead to ear (g-obil, g-obir) distances, at close to genome-wide significance. *CCNE1* codes for a cyclin upregulated in various cancers [44]. Weaker associations were found with the inner eye to ear distances (enl-obil, enr-obir) and *VSX2*, which appears to affect eye organogenesis [45]. These associations were supported by evidence of LD in the LocusZoom region plots. One marker, *AC019193.4*, is associated with the width of the eye (exl-enl, exr-enr) and is located adjacent to *WWC2*, a regulator of embryonic development and angiogenesis in mice involving the retina [46]. The width of the eye and exposed white sclera is argued to be a derived feature in humans assisting in non-verbal signalling [47].

The distance between the inner canthi of the left and right eyes (enl-enr), approximating to the width of the nasal bridge, were associated with a disparate set of genes. Associations with the markers *AL138733.1*, *AL356535.1* were of genome-wide significance. These are adjacent to *MTHFD1L*, responsible for neural tube closure [48] and *AKAP12*, a controller of formation of the blood-retinal barrier [49]. Weaker associations were evident with *INPP4B*, affecting bone formation [50], *FMN1*, reported in cases of orofacial clefting [51], and *OR5V1*, responsible for odorant receptor function [52]. The association of enl-enr with *INPP4B* was also evident in the adjusted and genotyped data.

KAZN was associated with the position of the alares bilaterally, relative to the glabella (g-all, g-alr) and was also associated with the g-se distance in the midline. *KAZN* codes for kazrin, a protein essential for the formation of craniofacial tissues from the cranial neural crest and precordal plate in *Xenopus laevis* [53]. *PTPRG* was also associated with the relative position of the alares bilaterally, but in this case relative to the selion (se-all, se-alr). *PTPRG* also showed weak evidence of association in the adjusted and genotyped data. Studies in mice indicate the *PTPRG* gene product RPTP is expressed in all sensory organs, including the nose, suggesting a potential role in sensory neurons [54].

Weaker associations with nasal length (se-prn) included *AC019211.1* and *CNTN1*. *AC019211.1* showed evidence of association in the adjusted data, including the genotyped data, and of LD in the LocusZoom region plot. Gross changes in the 2q14 region incorporating *AC019211.1* are associated with craniofacial dysmorphias [55]. *CNTN1* is significantly associated with other dimensions in the facial midline (g-ls, se-ls, see above), implicating a role affecting the height of the nasal aperture.

Alar to alar distance, approximating to nasal width, was associated with a further series of genes. Associations with *SORBS1* and nearby *ALDH18A1* were evident in the unadjusted and genotyped data. *SORBS1* codes for a protein responsible for adhesion mediated cell signalling and migration [56]. *ALDH18A* codes for P5CS, a carboxylate enzyme localized to the mitochondria [57]. Mutations

in this gene have been associated with craniofacial dysmorphia, notably in de Barsy syndrome [58]. Associations with *MIR548A1HG* appear in both the unadjusted and adjusted data. This lncRNA gene is located at 6q22.3, as is *KDM1B*, which is associated with the all-alr distance at genome-wide significance in the adjusted data. Deletions in this region have been shown to lead to craniofacial dysmorphias, including nasal anomalies [59].

Also notable were many associations with cprimel-cprimer, among which those with *VPS35L* and *STK32B* occurred at high significance in the adjusted data. *VPS35L* codes for a highly conserved protein involved in maintaining the homeostasis of the plasma membrane. Mutations in *VPS35L* are associated with 3C/Ritscher-Schinzel-like syndrome [60], a cause of facial dysmorphias incorporating an upturned nose [61]. *STK32B* is a member of the serine/threonine kinase family, mutations in which cause orofacial clefting [62]. Another marker, *SNX9*, found at slightly below genome-wide significance in the unadjusted data, is included in a microdeletion known to be associated with craniofacial dysmorphia [63]. The cprimel-cprimer distance is the distance between the apex of each nostril and is approximately the same as the width of the columella at its height. It may be a proxy for the width of the philtrum, which is located at the base of the columella and is affected in clefting [64].

Associations with 3D distance measurements of the mouth and chin

A number of genes appear to be associated with the dimensions of the chin. In the midline, the sl-pg distance is weakly associated with *AC004047.1*/*TACR3*. *TACR3* codes for a receptor implicated in various neurological conditions, including orofacial allodynia [65]. *AC004047.1*/*TACR3* is also associated with the bilateral distances pg-chl and pg-chr, as are two further markers, *EIF3F* and *LINC02660*. *EIF3F* codes for a translation initiation factor, mutations in which are associated with clefting and minor facial dysmorphias [66]. This association is supported in the genotyped data. *LINC02660* is located next to *KLF6*, which codes for a transcription factor affecting odontoblast differentiation [67].

A further marker, *MIR8084*, is associated with both bilateral dimensions (pg-chl and pg-chr) in the adjusted data, and supported in the genotyped data. *MIR8084* codes for microRNA miR-8084. Although not itself implicated, a series of other microRNAs have been shown to affect frontonasal development in mice [68]. Additionally, this marker is immediately adjacent to *TRIQK*, which codes for a highly conserved small protein found in the endoplasmic reticulum [69]. Deletions in the nearby region of 8q22.1 have been shown to underlie a number of craniofacial dysmorphias, most notably Nablus syndrome [70], within which a role for noncoding RNAs has been suggested. Another gene weakly associated with pg-chr is *PCCB* (*STAG1*). This marker is more strongly associated with mouth

width (chl-chr) at just below genome-wide significance, and is supported in the genotyped data (Fig. 2).

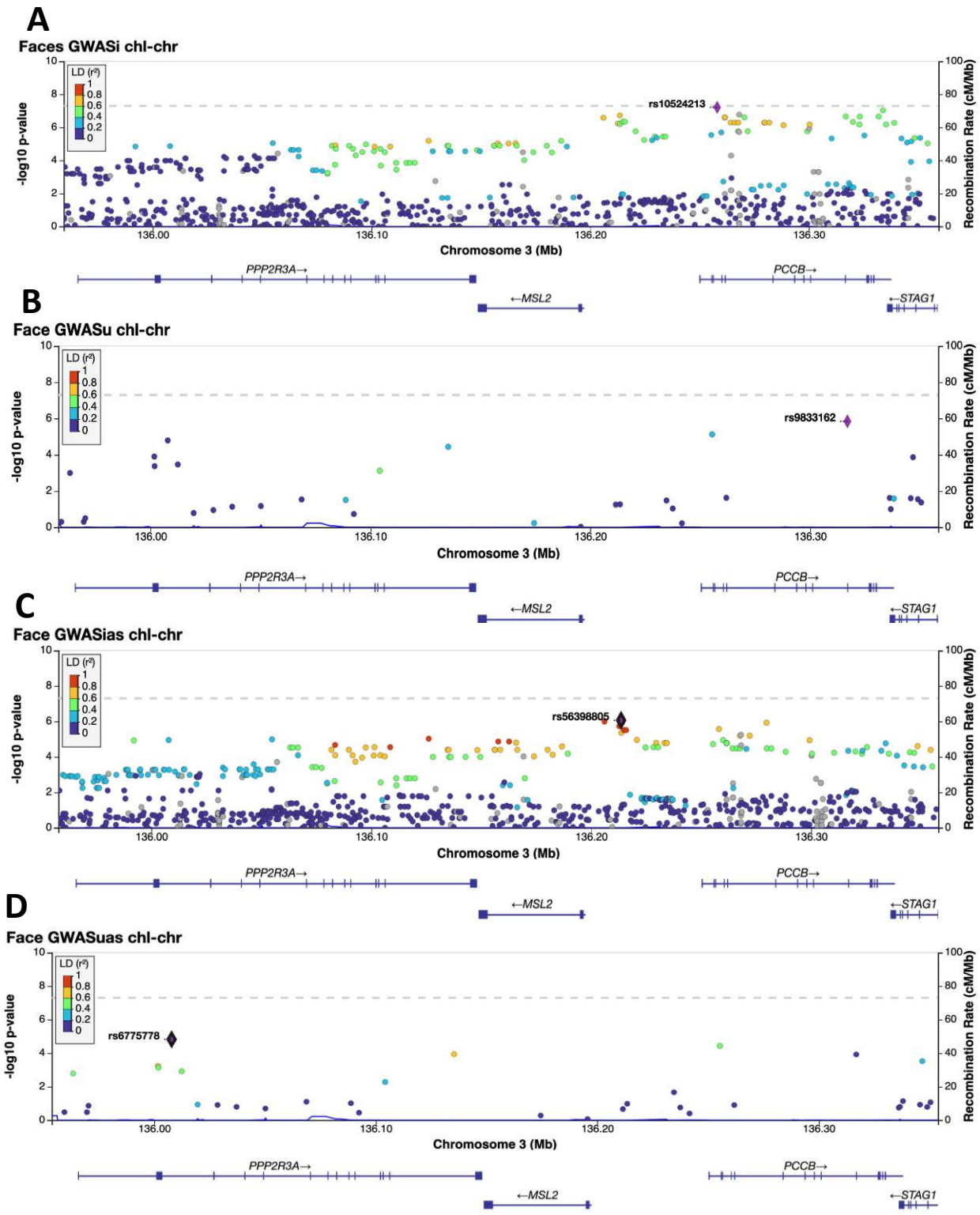
PCCB is located near *STAG1*, which codes for subunit of the cohesin complex. *STAG1* mutations are associated with a number of craniofacial dysmorphias, which can include a wide mouth [71,72] and, most notably, with STAG1 syndrome [73]. Another gene associated with mouth width is *MAST4*, an association also supported in the genotyped data. *MAST4* is involved in bone and cartilage formation, and is a critical regulator of amelogenesis [74].

Although associations with other dimensions approximating to the width of the cranial base (g-obil, g-obir; obil-obir) were detected, no convincing associations implicitly involving the two longer bilateral dimensions of the mandible (pg-obil, pg-obir) were found. Skeletal classification of the mandible reveals complex relationships between size, sex, and age, which may incorporate factors including extended growth in developing males, tooth development and, potentially, masculinisation in mature females [75]. As with the face in general [76], sex differences in the size of the mandible are considerable, and may be more diagnostic of sex than shape. Adjusting for sex in the GWAS could be anticipated to have had the unintended consequence of impeding detection of size-related associations in the major dimensions of the face, including the mandible. In this regard, it is interesting to note the p-values given in the bilateral associations of the chin (pg-chl, pg-chr) with *MIR8084* are present, but reduced in the unadjusted data, whereas those of mouth width (chl-chr) with *PCCB*(*STAG1*) and *MAST4* are present, but reduced in the adjusted data.

Associations with 3D distance measurements of the earlobe

A single gene, *CXCL13*, is weakly associated with the bilateral sbal-obil and sbar-obir dimensions, which are the distances from the lower point of attachment of the ear to the lowest point of the ear, roughly approximating to the length of the earlobe. This association is evident on both the left and right sides in both the adjusted and unadjusted data, and is supported in the genotyped data.

Fig. 2: LocusZoom region plots for the chl-chr 3D distance near *PCCB*/*STAG1* (also online).



Images from online LocusZoom region plots for the chl-chr 3D distance on chromosome 3 near *PPCB*/*STAG1* showing adjacent SNPs by position and -log10(p) value coded by degree of linkage disequilibrium (r^2) indexed to labelled lead SNP, with local gene positions shown: **A** imputed data. **B** genotyped data. **C** imputed data adjusted for age and sex. **D** genotyped data adjusted for age and sex.

Given the role of *CXCL13*/*CXCR5* in the inflammatory response [77], any causative relationship with sba-obi distance is difficult to identify. The locus encapsulates a non-coding RNA sequence, LOC1053707297 [78], which again has no obvious relationship with morphology of the pinna or earlobe. However, the adjacent *CCNG2* codes for another cyclin, which appears to be a negative regulator of cell cycle progression [79]. Also adjacent, *CNOT6L*, may promote cardiomyocyte proliferation in vitro and metastasis in certain cancers [80]. Mutations and deletions beyond 4q21 have been implicated in collagen over-accumulation in hyaline fibromatosis [81,82] and more general craniofacial dysmorphias [83], implicating a further set of genes.

The earlobe consists of areolar and adipose connective tissues beneath the epidermis, provided with a copious blood supply. The earlobe does not contain cartilage, but does contain collagen. Earlobes are especially vulnerable to the development of keloids, commonly associated with wound healing following ear piercing. Keloids are the result of excess deposition of collagen in the dermis and subcutaneous tissues. Although patterns of familial inheritance and candidate genes have been proposed, none involve chromosome 4 [84].

None of the above associations appear to repeat the findings of an earlier extensive GWAS of earlobe attachment [7,85] or of overall ear morphology [86]. However, the attachment-point-to-lowest-point (obi-sba) of ear distance is a quantitative measurement, whereas earlobe attachment is a qualitative present versus absent categorization. These results raise the question of whether earlobe length is a consequence of an overgrowth-like process associated with collagen accumulation involving *CXCL13*, *CCNG2*, *CNOT6L* or other genes in the 4q21 region.

Associations with 3D distance measurements of cranial width

Observations of covariation between head size and pelvis shape imply a strong selective effect and genetically-maintained balance in these dimensions affecting males as well as females. This may be the consequence of the obstetrical dilemma, where the selective pressure favouring large-brained babies is counteracted by the obstetric risk of childbirth [87,88].

In this study, two potential associations with the largest dimension of the skull measured (obil-obir) were detected, however both were of below genome-wide significance and poorly supported in the genotyped data. The first was with *LHX2*, which codes for a LIM homeobox transcription factor controlling 3D limb patterning and growth [89], also essential for neocortex development [90]. The second was with *CCDC141*, which has a role in left-right axis development and, potentially, in laterality defects [91]. These observations raise the possibility *LHX2* might also influence the 3D patterning of the embryonic cranial base, the breadth of which appears fundamentally related to brain

volume [92], and *CCDC141* may be complementary to *LHX2* in regulating the size and symmetry of the developing cranial base.

Among evolutionary adaptations to childbirth and bipedalism in humans, the change in shape of the ilium is thought to be the most significant [93]. One gene, *EMX2*, has been shown to affect ilium development and interact with another homeodomain transcription factor, PBX1, in pelvic formation [94]. Furthermore, *EMX2* and *LHX2* are expressed in similar gradients in the cortex [90], where both appear be under PBX control [95].

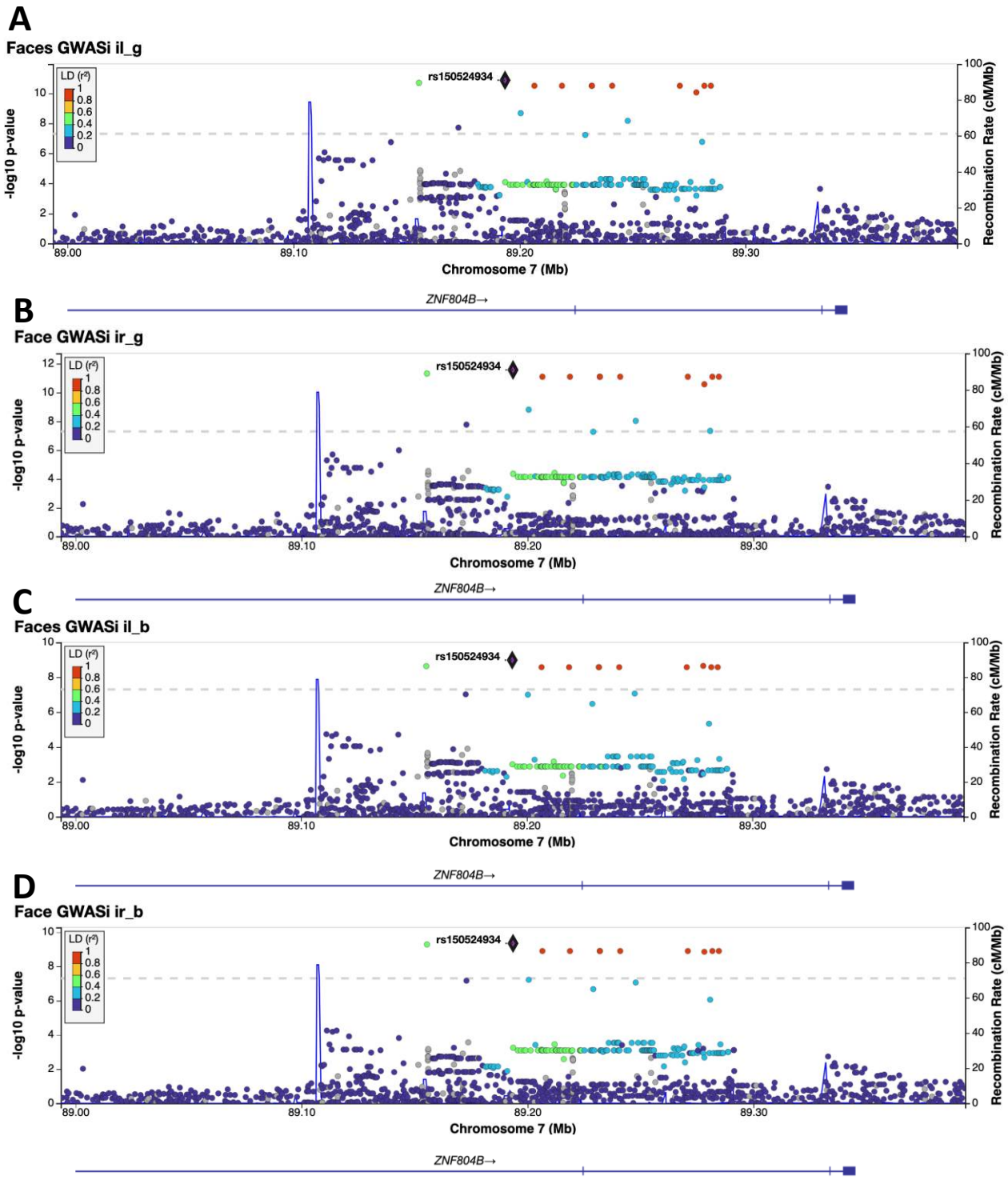
The control of both the width of the cranial base and the length of the ilium in human development invite the involvement of human accelerated regions (HARs) in the evolution of bipedalism [96,97]. A provisional search for HARs [98] near *PBX1*, *LHX2* and *EMX2*, yielded a HAR (HARsv2_0143) at position 1:164,563,928-164,564,046 within 4,862 bp of *PBX1*, which is at position 1:164,559,184-164,886,047. While no HAR has been reported in the near vicinity of *LHX2* and *EMX2*, *PBX1* is implicated in the regulation of both and regulatory changes with specific functional effects may anyway be widespread [99].

Association analysis of RGB colour values of the iris, lip vermillion and glabella.

In the imputed data, an association of RGB red colour in the iris (il-r, ir-r) with *SLIT3* was evident at genome-wide significance on the left side, repeated at just below genome-wide significance on the right. *SLIT3* is a gene implicated in vascularization and retinal axonal crossing in the facial midline in zebrafish [100]. This raises the possibility the RGB red component of iris colour is associated with vascularization. However, while these associations are repeated in the imputed data adjusted for age and sex, they are not evident in the genotyped data.

OCA2-HERC2 are well understood to control iris colour [101]. Associations with *HERC2* were found at above genome-wide significance for the RGB green (il-g, ir-g) and blue (il-b, ir-b) colour components. These associations were reproduced in the adjusted data and at above genome-wide significance in the genotyped data. Associations of above genome-wide significance were also found with *ZNF804B*, with LD consistently evident (Fig. 3).

Fig. 3: LocusZoom region plots for the il-g, ir-g, il-b and ir-b RGB colour values (also online).



Images from online LocusZoom region plots for il-g, ir-g, il-b and ir-b RGB colour values on chromosome 7 near *ZNF804B* showing adjacent SNPs by position and -log10(p) value coded by degree of linkage disequilibrium (r^2) indexed to labelled lead SNP, with the gene positions shown: **A** imputed data. **B** genotyped data. **C** imputed data adjusted for age and sex. **D** genotyped data adjusted for age and sex.

These associations were reproduced in the adjusted data, and below cutoff in the genotyped data. Very high effect sizes in the associations with colour values suggest an inflation affect attributable to small sample size. *ZNF804B* has been shown to be associated with skin pigmentation in Chinese and other East Asian populations [16]. These results extend this observation to iris pigmentation, possibly among people of East Asian ancestry contributing to the study. Differences in findings for skin versus iris pigmentation might be explained by sample biases or differences in the methods of classification: i.e. the CIELAB model [16] rather than RGB system used in this study. A polarization-sensitive anterior segment optical coherence tomography (PS-AS-OCT) study of RGB colour values underlying melanin pigmentation in the iris pigment epithelium [102] suggested high melanin concentration may be reflected in high RGB blue values and low RGB red and green values, specifically.

A further marker associated with RGB green and blue iris colour values detected was with *AL355300.1*, an association partially reproduced in the adjusted data. The relevant SNP, rs56389800, is located near to the 1,219 bp non-coding HAR enhancer HARE5 (ANC516), at position 10:36,238,421-36,239,039 [103], that is, within 3,194 bp of the SNP. HARE5 has been shown to interact most strongly with Frizzled Class Receptor FZD8, a Wnt pathway receptor implicated in brain development. Relative to other primates, HARE5 appears to cause an increased rate of neuron production and a larger cortical size, associated with early and enhanced FZD8 expression, thought to be of considerable significance to human-specific brain development [104]. FZD8 expression is early and highly correlated with that of neural stem cell markers SOX2 and PAX6 ($r > 0.90$), suggesting optic and neurocortical development is harmonious and follows a distinct path in humans [105]. PAX6 is central to eye development and is expressed in neural and pigmented layers [106]. SOX2 has a similarly complex role [107]. The association of *AL355300.1* with RGB green and blue colour values in the iris may be explained by its proximity to HARE5 and the influence of this enhancer on the formation of the neocortex, and the morphology and pigmentation of the various tissues of the eye.

Further associations of genome-wide significance were found between *FRG1DP* and *EPHA1*, and the RGB green colour value of the left and right irises (il-g, ir-g). *FRG1DP* is near *VSX1*, which appears to control red green photoreceptor pigmentation in the retina [108]. The association with *EPHA1* may be connected to melanocyte activity in the iris [109] or other roles of ephrins in the eye. These associations are reproduced in the adjusted data, but not reproduced above cutoff in the genotyped data.

An association of RGB blue iris colour values with *MYCBP2* was evident at genome-wide significance (il-b, ir-b). *MYCBP2* is involved in axonal growth and guidance. Mutations in this gene have been shown to affect the optic discs and axial length of the eyeball [110].

Among other associations also found with iris colour values are those with *EDIL3* and *MATN1* detected in both unadjusted and adjusted data. *EDIL3* is a gene implicated in retinal angiogenesis and the development of the anterior segment of the eye [111,112]. *MATN1* codes for Matrilin-1, a protein which binds to collagen fibrils and is expressed in the embryonic stages of eye development [113,114]. Collagen can be found in most tissues of the eye [115], suggesting *MATN1* may influence iris colour directly or via its interactions with collagen. A further association was found of *CLEC10A* with the RGB green colour values bilaterally (il-g, ir-g). CLEC10A is a dendritic cell marker involved in the induction in the immune response [116], including in the uvea [117].

Association analysis of the RGB colour values of the lip yielded only one bilateral result, in the unadjusted data, and at below genome-wide significance level. This was an association of the RGB blue colour value with *TSPEAR*, found on both the left and right sides (vl-b, vr-b). *TSPEAR* is associated with ectodermal dysplasia characteristically causing enlarged or everted lips [118].

A single association near *PSD3* is suggested with the glabella green (g-g) RGB value was reproduced in the adjusted and genotyped data, but at below genome-wide significance. *PSD3* is expressed in epithelial tissues and is a possible risk factor for systemic sclerosis of the skin [119].

Comparisons with other studies

Direct comparisons of SNP rsIDs with those reported in previous studies revealed no exact matches. Searches by chromosome and position within 250,000 bp, and matches by gene name are given in Supplementary File S3 online. None of these matching markers showed clear correspondence with the dimensions detected in the associations in this study. However, an association of the *EDAR* gene with RGB red component of the lip vermillion (vl-r) may relate to its role in disorders involving prominent or everted lips [120].

A search against markers implicated in human-chimpanzee divergence [121] showed a match with *NPSR1-AS1*, found near *BMPER*. *BMPER* is associated with craniofacial morphology [122] and located near one of a number of HARs. A further search against 32 human enhancer clusters showing species-biased differences in expression between humans and chimpanzees [121] yielded one enhancer cluster at 11:19,541,475 to 11:19,908,387 near to a marker at 11:19,341,729 close to *NAV2*, which codes for a crucial and conserved neuronal navigator weakly associated with g-se distance in this study, mutations in which are known to cause craniofacial dysmorphia [123].

The lack of confirmation of associations reported in existing studies may be due to the highly divergent methodological approaches taken. Nevertheless, it is noticeable other associations, such as widely reported with *PAX3* [9,15,124,125] or with overall face size [13], were not reproduced.

Discussion

These results suggest face shape is influenced by numerous genes and embryological pathways segregated according to component structures of the face, often organised early in development [15,126]. Many are known to affect specific phenotypic traits or to be involved in craniofacial dysmorphias, supporting the view the latter are frequently a consequence of overgrowth or restricted growth [127,128], possibly due to a cumulative effect of common variants [129].

HARs and other associations were implicated in evolutionary change [130], including reduction in the supra-orbital ridge, midface area, maxilla, mandible and dentition, as well as in cortex development and encephalisation. Further associations may reflect selective effects on the morphology of the eyes in ancestral and extant human groups [47,131-133].

Studies of facial attractiveness indicate symmetry may be a phenotypic indicator of genetic fitness [134-136]. A number of associations found were implicated in left-right axis organization and laterality, and may help confer symmetry and averageness, especially if under frequency-dependent selection. In addition, genetic variation in facial morphology in humans may have evolved to signal individual identity. Associations found in this study offer empirical support for this concept [137,138].

The ability to perceive and recognize faces appears to be distinctive in humans [139], highly heritable [140] and under genetic control. One specific gene, *MCTP2*, has been found to be associated with congenital prosopagnosia [141]. This gene is located at 15:94,231,366 to 15:94,483,952, which is not near any of the markers found in this study or reported HARs [98], leaving open the broad question of how face shape variation and face perception may have developed during the course of primate and human evolution.

Although the small sample size (in terms of numbers of subjects) might be considered a limitation, a strength of this study has been the use of precise measurements subject to extensive quality assurance [142] in an integrated analysis of a unique ancestrally diverse sample [5]. The results demonstrate below threshold associations are not necessarily unimportant [143,144]. They also demonstrate direct measurement of anthropometric landmarks offers a useful approach to face GWAS, generating results that are easily visualised and highly interpretable. The research may be extended to GWAS of other 3D dimensions, of PCs of overall variation or of Procrustes-adjusted large face surface datasets already collected, and by further genotyping to complement existing biometric data [145],

It may be amenable to incorporation into multi-trait analysis of GWAS summary statistics (MTAG) and combined GWAS (C-GWAS) analyses [146,147], and into studies of polygenic effects in other complex traits and diseases [148,149]. It may contribute to tools for image-based detection of syndromic pathologies or carriers [150] or for estimating appearance from crime scene DNA [151-158]. Fundamentally, they may improve understanding of craniofacial syndromes, which have a considerable effect on quality of life and on the health system [159].

Methods

Biometric data

The data consist of biometric and genetic measurements collected from healthy volunteers in São Paulo, Brazil, and Toronto, Canada, classified by self-reported age (from 17 to 74 years), sex and ancestry.

The São Paulo dataset (n=272) comprised biometric data captured using a Geometrix FaceVision FV802 Series 3D stereophotographic camera system (ALIVE Tech, Cumming, GA). This system relied on eight 1 MB cameras covering the anterior portion of the face including the ears. A proprietary Geometrix software application allowed precise 3D measurement by triangulation onto any mutually visible point in any two captured images using a manual point-and-click operation.

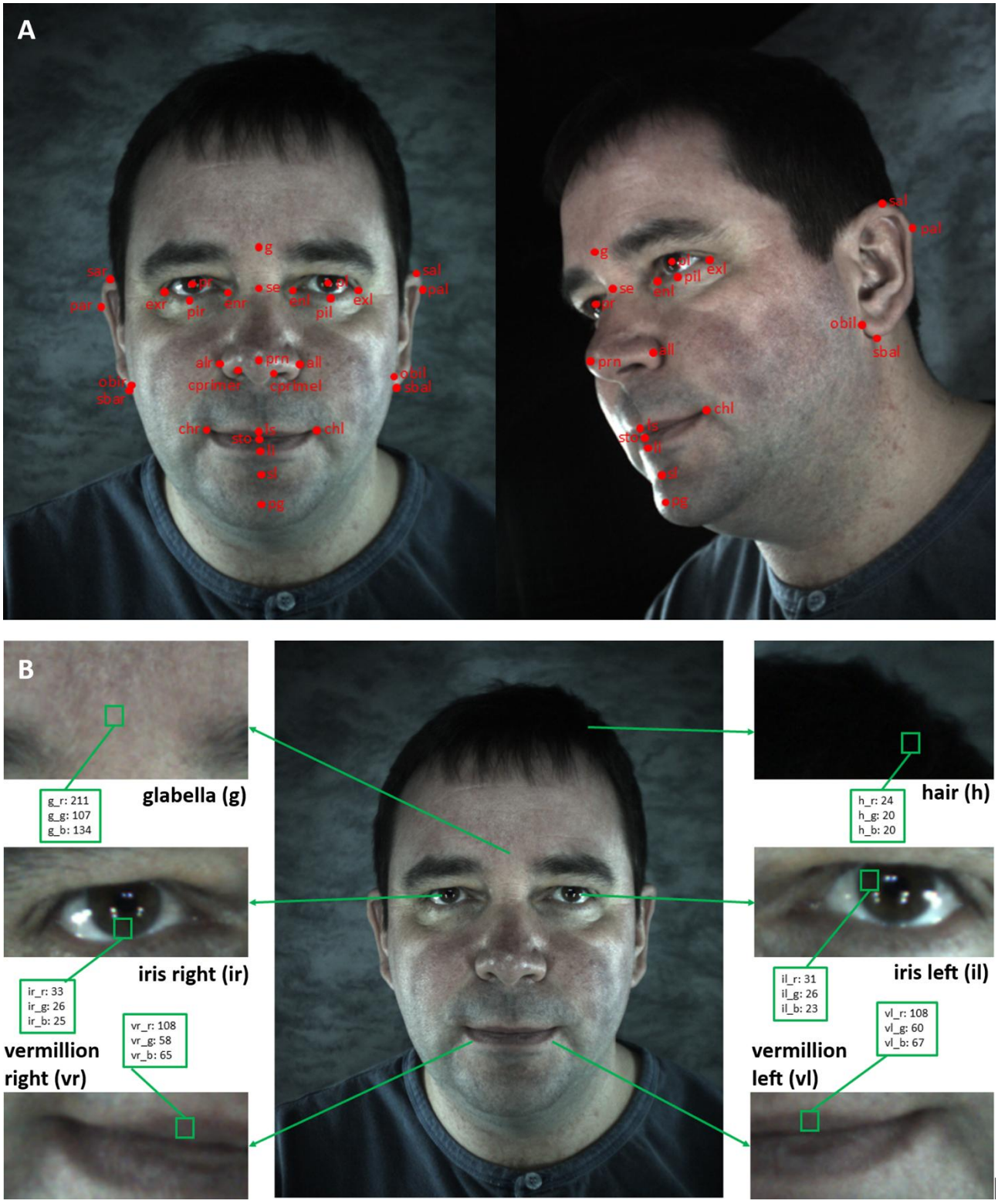
The Toronto dataset (n=123) comprised biometric data captured using a Cyberware 3030PS Head and Neck Scanner (Cyberware, Monterey, CA), a laser light detection and range finding (LiDAR) system designed for craniofacial 3D surface imaging. A proprietary Cyberware software application allowed the position of any point on a 3D surface of 55,225 vertices (256 latitudes by 256 longitudes) to be precisely measured, again using a point-and-click operation.

The sources of error inherent in the different scanning methods have been analysed in detail in previous studies, as have the relative utility in biometric analyses of the sixty or so anthropometric landmark points traditionally used. The 30 3D landmarks used (Fig. 4 and Table 3) optimally combined increased variability between individuals and reduced susceptibility to measurement error [76,160].

All biometric images were captured in a consistent head posture with the subject in a neutral facial expression. Each of 30 3D landmark positions were located by three independent observers.

RGB colour values were measured from the 1MB photographic images captured in the São Paulo dataset. Values were sampled at locations intended to survey the red (r), green (g) and blue (b) colour components of the skin at the glabella (g), hair (h), left and right iris (il, ir), and left and right lip vermillion (vl, vr). Two camera views were chosen for sampling, both in the facial midline, offering the greatest visibility of the locations chosen for sampling (see Fig. 4). RGB colour values were sampled using the Eyedropper Colour Sampler tool in Adobe Photoshop version 16.0 (Adobe, San Jose, CA). Care was taken to avoid areas of shadow, glare, artificial colour and greying with the aim of returning unmodified natural colour. Where possible, a 25-pixel array was sampled. Where this could not be achieved while avoiding unwanted pixels, a 9-pixel array was used.

Fig. 4: Capture of 3D anthropometric landmark points and RGB component colour values from stereophotographic images.



A Locations of 30 3D anthropometric landmark points (Table 3) viewed from two of the eight available stereophotography camera views. **B** Locations of six points for collection of RGB component colour values, illustrating RGB data obtained from 25-pixel arrays in Adobe Photoshop.

The individual red, green and blue values returned by the sampler tool were copied and pasted directly into an MS-EXCEL (Microsoft, Redmond, WA) spreadsheet. Each of the 18 RGB colour values were sampled by three independent observers.

Quality assurance was completed by analysis of intra-observer variation in each of the 30 3D landmark points and the separate red, green and blue measurements in each of 18 RGB colour values sampled. In all cases, individual outliers in the standard deviation of variation were remeasured until no improvement could be achieved.

After QA, mean values were calculated from the independent observations taken. For RGB colour values, mean values were further calculated from the means established from each of the two photographs analysed.

Genomic data

The genomic data was generated from saliva samples collected in São Paulo and Toronto (n=395) using Oragene DNA kits (DNA Genotek, Ottawa, ON). The São Paulo dataset could be described as admixed, but not mixed, consisting of often intermarried descendants of historic migration to Brazil and indigenous Brazilians. The Toronto dataset could be described as mixed, but not admixed, consisting of many first- and second-generation immigrants to Canada from different countries, with little or no intermarriage.

Genotyping from DNA purified from saliva was performed using the Illumina Human 660W BeadChip microarray developed in a collaboration between the Wellcome Trust Sanger Institute in Cambridge, UK, the Brigham and Women's Hospital in Boston, USA, and at the Hospital for Sick Children in Toronto, Canada, where the genotyping was carried out.

This chip includes 592,614 SNP probes offering comprehensive coverage of those identified in the CEU and CHB/JPT HapMap haplotype mapping project datasets, collected from Caucasian, Han Chinese and Japanese populations, respectively.

Quality control (QC) of the genetic data involved resolution of discordant sex, missing genotypes, heterozygosity rates, minor allele frequencies (MAF) and Hardy-Weinberg equilibrium (HWE), and detection of duplicated or related individuals carried out in PLINK [161].

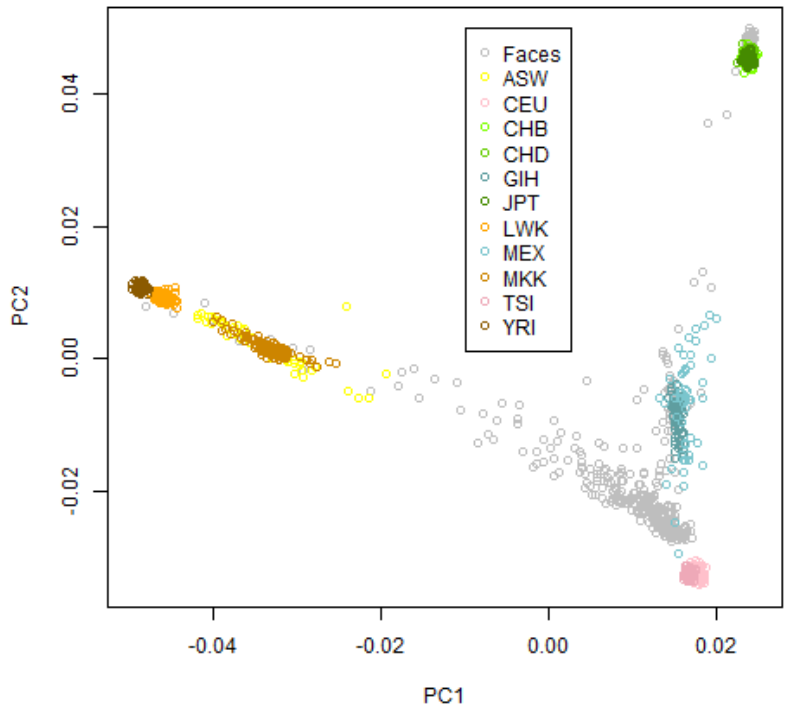
A threshold of 0.06 was chosen for removal of 11 individuals with high missingness rates. Heterozygosity rates of below 0.28 and above 0.35 led to the removal of two further individuals due

to outlying heterozygosity. Analysis of duplicated or related individuals was completed after the per-individual QC steps were completed, following which one each of four pairs of apparently related individuals were also removed. These QC steps reduced the overall sample from 395 to 378 and the RGB colour value sample from 272 to 258.

A per SNP missing genotype rate of 0.05, MAF of 0.01 and less-stringent HWE exact test p-value cut-off of 0.00000001 were chosen as thresholds for removal of SNP markers from the sample, leaving 552,612 of the original 592,614 SNP markers remaining.

Comparison by PCA of ancestry in the sample with the HapMap Phase 3 (HM3) database [162] confirmed a broad distribution across the background of the HM3 collection (Fig. 5). These distributions are not surprising given the demography of Sao Paulo, Brazil, and Toronto, Canada, and the pattern of historical and recent immigration into those cities.

Fig 5: Plot showing the first two Principal Components of the data combined with the 11 populations samples in the HM3 collection.



Key ASW: African ancestry in Southwest USA (n=90), CEU: Utah residents with Northern and Western European ancestry (n=180), CHB: Han Chinese in Beijing, China (n=90), CHD: Chinese in Metropolitan Denver, Colorado (n=100), GIH: Gujarati Indians in Houston, Texas (n=100), JPT: Japanese in Tokyo, Japan (n=91), LWK: Luhya in Webuye, Kenya (n=100), MEX: Mexican ancestry in Los Angeles, California (n=90), MKK: Maasai in Kinyawa, Kenya (n=180), TSI: Toscans in Italy (n=100), YRI: Yoruba in Ibadan, Nigeria (n=180).

Following QC, the genotyped data was imputed via the Michigan Imputation Server. The genotyped data was aligned with and imputed against the b38 (GRCh38) assembly of the Human Reference Consortium 1000G reference panel. The imputed data was initially saved in PLINK format for each of 22 autosomes, filtered for MAF > 0.1 and R^2 > 0.3, and then merged into a single dataset incorporating 10,960,275 SNPs.

Selection of phenotypic traits for GWAS analysis

Earlier research on a large 3D database [76] established the first Principal Component (PC) of face shape variation related to size, the second to dimensions of the midline, and the third to variation in bilateral dimensions. Of a possible 435 pairwise combinations of the 30 3D points measured, a set of 36 were chosen to represent these three PCs, in anticipation these dimensions may include candidates for genetic associations with major phenotypic variation. The pairwise distances chosen were, in the facial midline: g-se, g-ls, g-sl, g-pg, se-ls, se-sl, se-pg, ls-sl, ls-pg and sl-pg; in the bilateral dimensions: g-obil, g-obir, enl-obir, enr-obil, enl-obil, enr-obir, pg-obil, and pg-obir; in the midface (eyes): se-enl, se-enr, exl-enl, exr-enr and enl-enr; in the midface (nose): g-all, g-alr, se-all, se-alr, se-prn, all-alr and cprimel-cprimer; for the mouth and chin: pg-chl, pg-chr and chl-chr, and for the ears and facial width: sbal-obil, sbar-obir and obil-obir. Use of pairwise distances removes the need to adjust for differences due to rotation and transformation occurring between individual biometric scans.

The 36 dimensions chosen also facilitate comparison with previous studies of craniofacial evolution, morphology and genetics, such as the projection of the midface [130], the dimensions of the nose, brow area, cheekbones, jaw and mouth [133,163], dimensions of the nasion (selion) and philtrum and associations with NSCL/P. [8,12], and the inter-alar dimensions and spacing of endocanthions [124]. RGB colour values selected for analysis allow comparison with existing research on the genetics of iris, skin and hair pigmentation [101,151,164].

GWAS analysis

Genome-wide Complex Trait Analysis (GCTA) is a GWAS tool whose mixed linear model is suited to accommodating relatedness or population substructure [18]. GCTA was therefore for used for GWAS analysis without further specific adjustment for population stratification.

Although face shape variation evidently tends to differ by age and between males and females, explicit analysis on a large 3D database [76] showed these differences are primarily due to scale. Scale-independent differences were quite subtle and seemed to most affect the earlobes, and nose and

lips. GCTA analysis was therefore completed in two phases, first without and then with age and sex as covariates.

Quantile-Quantile (QQ) and Manhattan plots were generated from the GCTA mixed linear model (.mlma) output files produced for each of the 36 3D dimensions and 18 RGB colour values studied, and further analysed in LocusZoom [19,165], with the mlma major or "non-effect" allele (A2) set as the reference allele and the minor or "effect" allele (A1) as alternate allele.

The LocusZoom interface makes interactive reference to the 100 Genomes Project LD, GENCODE genes, HapMap-based recombination rate and the EBI GWAS catalogue [19,165]. When viewed online, LocusZoom Manhattan plots are accompanied by tables listing marker position, SNP rsID, nearest genes, and significance of association as a -log10(p) value for markers with an association of significance above -log10(p)=6.000 or $p=10^{-6}$. High significance markers can be further investigated via Locus Zoom region view plots accessed online by a point-and-click operation. These show all markers within a short and adjustable distance of the reference marker, allowing recombination rates, genome position, probability values of associations and regions of linkage disequilibrium to be visualized in colour. Locations of nearby genes in the UCSC Genome Browser are depicted. Region view plots of the results for all 216 GWAS were made publicly available via the searchable LocusZoom interface (my.locuszoom.org). Provisionally corroborative findings reported in the literature on craniofacial development and face shape genetics were discussed in relation to their developmental significance and role of any genes implicated.

Associations found in the imputed data were reviewed for consistency with genotyped data, and unadjusted results were compared with those obtained after adjustment the covariates of age and sex. In the case of bilateral measurements, associations discovered on one side were investigated on the opposite side and presence or absence of correspondence noted.

A direct comparison with the published genetics of face shape literature was made in order to find matches to reported genes and markers [7-15,64,122,124,125,131,132,154,163,166-171]. Three modes of comparison were employed: 1) by SNP rsID, 2) by chromosome and position (in bp) and 3) by gene name given as "nearest gene" by LocusZoom. For the match by chromosome and position comparison a distance of up to 250,000 bp was chosen.

Finally, a search was made for genes identified in this study adjacent to human chimpanzee-biased Human Accelerated Region (HAR) sequences, as reported in the literature [121].

References

1 Lieberman, D. *The evolution of the human head*. (Belknap Press of Harvard University Press, 2011).

2 Russell, J. A. & Fernández Dols, J. M. *The psychology of facial expression*. (Cambridge University Press, 1997).

3 Selleri, L. & Rijli, F. M. Shaping faces: genetic and epigenetic control of craniofacial morphogenesis. *Nat Rev Genet* **24**, 610-626, doi:10.1038/s41576-023-00594-w (2023).

4 LaMantia, A. S. Why Does the Face Predict the Brain? Neural Crest Induction, Craniofacial Morphogenesis, and Neural Circuit Development. *Front Physiol* **11**, 610970, doi:10.3389/fphys.2020.610970 (2020).

5 Zhou, Y., Song, H. & Ming, G. L. Genetics of human brain development. *Nat Rev Genet* **25**, 26-45, doi:10.1038/s41576-023-00626-5 (2024).

6 Enlow, D. H. *The human face; an account of the postnatal growth and development of the craniofacial skeleton*. (Hoeber Medical Division, 1968).

7 Shaffer, J. R. *et al.* Genome-Wide Association Study Reveals Multiple Loci Influencing Normal Human Facial Morphology. *PLoS Genet* **12**, e1006149, doi:10.1371/journal.pgen.1006149 (2016).

8 Liu, F. *et al.* A genome-wide association study identifies five loci influencing facial morphology in Europeans. *PLoS Genet* **8**, e1002932, doi:10.1371/journal.pgen.1002932 (2012).

9 Pickrell, J. K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet* **48**, 709-717, doi:10.1038/ng.3570 (2016).

10 White, J. D. *et al.* Insights into the genetic architecture of the human face. *Nat Genet* **53**, 45-53, doi:10.1038/s41588-020-00741-7 (2021).

11 Liu, C. *et al.* Genome scans of facial features in East Africans and cross-population comparisons reveal novel associations. *PLoS Genet* **17**, e1009695, doi:10.1371/journal.pgen.1009695 (2021).

12 Indencleef, K. *et al.* The Intersection of the Genetic Architectures of Orofacial Clefts and Normal Facial Variation. *Front Genet* **12**, 626403, doi:10.3389/fgene.2021.626403 (2021).

13 Cole, J. B. *et al.* Genomewide Association Study of African Children Identifies Association of SCHIP1 and PDE8A with Facial Size and Shape. *PLoS Genet* **12**, e1006174, doi:10.1371/journal.pgen.1006174 (2016).

14 Adhikari, K. *et al.* A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation. *Nat Commun* **7**, 11616, doi:10.1038/ncomms11616 (2016).

15 Claes, P. *et al.* Genome-wide mapping of global-to-local genetic effects on human facial shape. *Nat Genet* **50**, 414-423, doi:10.1038/s41588-018-0057-4 (2018).

16 Rawofi, L. *et al.* Genome-wide association study of pigmentary traits (skin and iris color) in individuals of East Asian ancestry. *PeerJ* **5**, e3951, doi:10.7717/peerj.3951 (2017).

17 Adhikari, K. *et al.* A genome-wide association scan in admixed Latin Americans identifies loci influencing facial and scalp hair features. *Nat Commun* **7**, 10815, doi:10.1038/ncomms10815 (2016).

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

19 Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336-2337, doi:10.1093/bioinformatics/btq419 (2010).

20 MIM130650.

21 MIM180860.

22 Laukoter, S. *et al.* Imprinted Cdkn1c genomic locus cell-autonomously promotes cell survival in cerebral cortex development. *Nat Commun* **11**, 195, doi:10.1038/s41467-019-14077-2 (2020).

23 Eggermann, T. *et al.* CDKN1C mutations: two sides of the same coin. *Trends Mol Med* **20**, 614-622, doi:10.1016/j.molmed.2014.09.001 (2014).

24 MIM614992.

25 MIM157980.

26 Vu, P. Y. *et al.* A homozygous balanced reciprocal translocation suggests LINC00237 as a candidate gene for MOMO (macrosomia, obesity, macrocephaly, and ocular abnormalities) syndrome. *Am J Med Genet A* **158A**, 2849-2856, doi:10.1002/ajmg.a.35694 (2012).

27 Morimoto, M. *et al.* Bi-allelic ATG4D variants are associated with a neurodevelopmental disorder characterized by speech and motor impairment. *NPJ Genom Med* **8**, 4, doi:10.1038/s41525-022-00343-8 (2023).

28 Eckersley-Maslin, M. A. Keeping your options open: insights from Dppa2/4 into how epigenetic priming factors promote cell plasticity. *Biochem Soc Trans* **48**, 2891-2902, doi:10.1042/BST20200873 (2020).

29 MIM230740.

30 Stranecky, V. *et al.* Mutations in ANTXR1 cause GAPO syndrome. *Am J Hum Genet* **92**, 792-799, doi:10.1016/j.ajhg.2013.03.023 (2013).

31 Fairbridge, N. A. *et al.* Cecr2 mutations causing exencephaly trigger misregulation of mesenchymal/ectodermal transcription factors. *Birth Defects Res A Clin Mol Teratol* **88**, 619-625, doi:10.1002/bdra.20695 (2010).

32 Spineli-Silva, S., Monlleo, I. L., Felix, T. M., Gil-da-Silva-Lopes, V. L. & Vieira, T. P. Overlapping Spectrum of Craniofacial Microsomia Phenotype in Cat-Eye Syndrome. *Cleft Palate Craniofac J* **61**, 1578-1585, doi:10.1177/10556656231174435 (2024).

33 MIM115470.

34 Ewing, A. D. *et al.* Microdeletion of 9q22.3: A patient with minimal deletion size associated with a severe phenotype. *Am J Med Genet A* **185**, 2070-2083, doi:10.1002/ajmg.a.62224 (2021).

35 Socorro, M. *et al.* Trps1 transcription factor represses phosphate-induced expression of SerpinB2 in osteogenic cells. *Bone* **141**, 115673, doi:10.1016/j.bone.2020.115673 (2020).

36 Field, S. *et al.* Pkd1l1 establishes left-right asymmetry and physically interacts with Pkd2. *Development* **138**, 1131-1142, doi:10.1242/dev.058149 (2011).

37 Bacchelli, E. *et al.* A CTNNA3 compound heterozygous deletion implicates a role for alphaT-catenin in susceptibility to autism spectrum disorder. *J Neurodev Disord* **6**, 17, doi:10.1186/1866-1955-6-17 (2014).

38 Compton, A. G. *et al.* Mutations in contactin-1, a neural adhesion and neuromuscular junction protein, cause a familial form of lethal congenital myopathy. *Am J Hum Genet* **83**, 714-724, doi:10.1016/j.ajhg.2008.10.022 (2008).

39 Lu, I. L. *et al.* Identification of genes associated with cortical malformation using a transposon-mediated somatic mutagenesis screen in mice. *Nat Commun* **9**, 2498, doi:10.1038/s41467-018-04880-8 (2018).

40 Stray-Pedersen, A. *et al.* Biallelic Mutations in UNC80 Cause Persistent Hypotonia, Encephalopathy, Growth Retardation, and Severe Intellectual Disability. *Am J Hum Genet* **98**, 202-209, doi:10.1016/j.ajhg.2015.11.004 (2016).

41 Valkanas, E. *et al.* Phenotypic evolution of UNC80 loss of function. *Am J Med Genet A* **170**, 3106-3114, doi:10.1002/ajmg.a.37929 (2016).

42 Yoldas Celik, M. *et al.* Expanding the genotypic spectrum of 3beta-hydroxy-delta5-C27-steroid dehydrogenase (HSD3B7) deficiency: novel mutations and clinical outcomes. *J Pediatr Endocrinol Metab* **38**, 546-550, doi:10.1515/jpem-2024-0454 (2025).

43 Habicher, J. *et al.* Chondroitin/dermatan sulfate glycosyltransferase genes are essential for craniofacial development. *PLoS Genet* **18**, e1010067, doi:10.1371/journal.pgen.1010067 (2022).

44 Zheng, X. *et al.* CCNE1 is a predictive and immunotherapeutic indicator in various cancers including UCEC: a pan-cancer analysis. *Hereditas* **160**, 13, doi:10.1186/s41065-023-00273-0 (2023).

45 Zou, C. & Levine, E. M. Vsx2 controls eye organogenesis and retinal progenitor identity via homeodomain and non-homeodomain residues required for high affinity DNA binding. *PLoS Genet* **8**, e1002924, doi:10.1371/journal.pgen.1002924 (2012).

46 Hermann, A. *et al.* The Hippo pathway component Wwc2 is a key regulator of embryonic development and angiogenesis in mice. *Cell Death Dis* **12**, 117, doi:10.1038/s41419-021-03409-0 (2021).

47 Kano, F. Evolution of the uniformly white sclera in humans: critical updates. *Trends Cogn Sci* **27**, 10-12, doi:10.1016/j.tics.2022.09.011 (2023).

48 Momb, J. *et al.* Deletion of Mthfd1l causes embryonic lethality and neural tube and craniofacial defects in mice. *Proc Natl Acad Sci U S A* **110**, 549-554, doi:10.1073/pnas.1211199110 (2013).

49 Li, X. *et al.* AKAP12 inhibits esophageal squamous carcinoma cell proliferation, migration, and cell cycle via the PI3K/AKT signaling pathway. *Mol Cell Probes* **72**, 101939, doi:10.1016/j.mcp.2023.101939 (2023).

50 Ferron, M. *et al.* Inositol polyphosphate 4-phosphatase B as a regulator of bone mass in mice and humans. *Cell Metab* **14**, 466-477, doi:10.1016/j.cmet.2011.08.013 (2011).

51 Ludwig, K. U. *et al.* Meta-analysis Reveals Genome-Wide Significance at 15q13 for Nonsyndromic Clefting of Both the Lip and the Palate, and Functional Analyses Implicate GREM1 As a Plausible Causative Gene. *PLoS Genet* **12**, e1005914, doi:10.1371/journal.pgen.1005914 (2016).

52 Volz, A. *et al.* Complex transcription and splicing of odorant receptor genes. *J Biol Chem* **278**, 19691-19701, doi:10.1074/jbc.M212424200 (2003).

53 Cho, K. *et al.* Kazrin, and its binding partners ARVCF- and delta-catenin, are required for Xenopus laevis craniofacial development. *Dev Dyn* **240**, 2601-2612, doi:10.1002/dvdy.22721 (2011).

54 Lamprianou, S. *et al.* Receptor protein tyrosine phosphatase gamma is a marker for pyramidal cells and sensory neurons in the nervous system and is not necessary for normal development. *Mol Cell Biol* **26**, 5106-5119, doi:10.1128/MCB.00101-06 (2006).

55 Goumy, C. *et al.* A novel 2q14.1q14.3 deletion involving GLI2 and RNU4ATAC genes associated with partial corpus callosum agenesis and severe intrauterine growth retardation. *Birth Defects Res A Clin Mol Teratol* **106**, 793-797, doi:10.1002/bdra.23535 (2016).

56 Zhang, M. *et al.* CAP interacts with cytoskeletal proteins and regulates adhesion-mediated ERK activation and motility. *EMBO J* **25**, 5284-5293, doi:10.1038/sj.emboj.7601406 (2006).

57 Fischer, B. *et al.* Severe congenital cutis laxa with cardiovascular manifestations due to homozygous deletions in ALDH18A1. *Mol Genet Metab* **112**, 310-316, doi:10.1016/j.ymgme.2014.05.003 (2014).

58 MIM219150.

59 Celestino-Soper, P. B. *et al.* Deletions in chromosome 6p22.3-p24.3, including ATXN1, are associated with developmental delay and autism spectrum disorders. *Mol Cytogenet* **5**, 17, doi:10.1186/1755-8166-5-17 (2012).

60 MIM619135.

61 Kato, K. *et al.* Biallelic VPS35L pathogenic variants cause 3C/Ritscher-Schinzel-like syndrome through dysfunction of retriever complex. *J Med Genet* **57**, 245-253, doi:10.1136/jmedgenet-2019-106213 (2020).

62 Ingersoll, R. G. *et al.* Association between genes on chromosome 4p16 and non-syndromic oral clefts in four populations. *Eur J Hum Genet* **18**, 726-732, doi:10.1038/ejhg.2009.228 (2010).

63 Zhong, M. L., Song, Y. M. & Zou, C. C. 6q25.1-q25.3 Microdeletion in a Chinese Girl. *J Clin Res Pediatr Endocrinol* **13**, 109-113, doi:10.4274/jcrpe.galenos.2020.2020.0008 (2021).

64 Howe, L. J. *et al.* Investigating the shared genetics of non-syndromic cleft lip/palate and facial morphology. *PLoS Genet* **14**, e1007501, doi:10.1371/journal.pgen.1007501 (2018).

65 Zhang, W. W., Wang, Y. & Chu, Y. X. Tacr3/NK3R: Beyond Their Roles in Reproduction. *ACS Chem Neurosci* **11**, 2935-2943, doi:10.1021/acschemneuro.0c00421 (2020).

66 Huffmeier, U. *et al.* EIF3F-related neurodevelopmental disorder: refining the phenotypic and expanding the molecular spectrum. *Orphanet J Rare Dis* **16**, 136, doi:10.1186/s13023-021-01744-1 (2021).

67 Chen, Z. *et al.* KLF6 facilitates differentiation of odontoblasts through modulating the expression of P21 in vitro. *Int J Oral Sci* **14**, 20, doi:10.1038/s41368-022-00172-6 (2022).

68 Iwaya, C., Yu, S. & Iwata, J. Genes Related to Frontonasal Malformations Are Regulated by miR-338-5p, miR-653-5p, and miR-374-5p in O9-1 Cells. *J Dev Biol* **12**, doi:10.3390/jdb12030019 (2024).

69 Onuma, Y., Watanabe, A., Aburatani, H., Asashima, M. & Whitman, M. TRIQK, a novel family of small proteins localized to the endoplasmic reticulum membrane, is conserved across vertebrates. *Zoolog Sci* **25**, 706-713, doi:10.2108/zsj.25.706 (2008).

70 Allanson, J. *et al.* Nablus syndrome: Easy to diagnose yet difficult to solve. *Am J Med Genet C Semin Med Genet* **178**, 447-457, doi:10.1002/ajmg.c.31660 (2018).

71 Di Muro, E. *et al.* Novel STAG1 Frameshift Mutation in a Patient Affected by a Syndromic Form of Neurodevelopmental Disorder. *Genes (Basel)* **12**, doi:10.3390/genes12081116 (2021).

72 Lehalle, D. *et al.* STAG1 mutations cause a novel cohesinopathy characterised by unspecific syndromic intellectual disability. *J Med Genet* **54**, 479-488, doi:10.1136/jmedgenet-2016-104468 (2017).

73 MIM604358.

74 Lee, D. J. *et al.* MAST4 regulates stem cell maintenance with DLX3 for epithelial development and amelogenesis. *Exp Mol Med* **56**, 1606-1619, doi:10.1038/s12276-024-01264-5 (2024).

75 Toneva, D. H., Nikolova, S. Y., Fileva, N. F. & Zlatareva, D. K. Size and shape of human mandible: Sex differences and influence of age on sex estimation accuracy. *Leg Med (Tokyo)* **65**, 102322, doi:10.1016/j.legalmed.2023.102322 (2023).

76 Evison, M. *et al.* Key parameters of face shape variation in 3D in a large sample. *J Forensic Sci* **55**, 159-162, doi:10.1111/j.1556-4029.2009.01213.x (2010).

77 Harrer, C. *et al.* The CXCL13/CXCR5 Immune Axis in Health and Disease-Implications for Intrathecal B Cell Activities in Neuroinflammation. *Cells* **11**, doi:10.3390/cells11172649 (2022).

78 NIH. (National Library of Medicine, 2025).

79 Sun, G. G., Hu, W. N., Cui, D. W. & Zhang, J. Decreased expression of CCNG2 is significantly linked to the malignant transformation of gastric carcinoma. *Tumour Biol* **35**, 2631-2639, doi:10.1007/s13277-013-1346-2 (2014).

80 Liu, Y., Ramkumar, N. & Vu, L. P. RNA deadenylation complexes in development and diseases. *Biochem Cell Biol* **101**, 131-147, doi:10.1139/bcb-2022-0325 (2023).

81 Rahman, N. *et al.* The gene for juvenile hyaline fibromatosis maps to chromosome 4q21. *Am J Hum Genet* **71**, 975-980, doi:10.1086/342776 (2002).

82 Burgi, J. *et al.* CMG2/ANTXR2 regulates extracellular collagen VI which accumulates in hyaline fibromatosis syndrome. *Nat Commun* **8**, 15861, doi:10.1038/ncomms15861 (2017).

83 Hu, X. *et al.* Further defining the critical genes for the 4q21 microdeletion disorder. *Am J Med Genet A* **173**, 120-125, doi:10.1002/ajmg.a.37965 (2017).

84 Glass, D. A., 2nd. Current Understanding of the Genetic Causes of Keloid Formation. *J Investig Dermatol Symp Proc* **18**, S50-S53, doi:10.1016/j.jisp.2016.10.024 (2017).

85 Shaffer, J. R. *et al.* Multiethnic GWAS Reveals Polygenic Architecture of Earlobe Attachment. *Am J Hum Genet* **101**, 913-924, doi:10.1016/j.ajhg.2017.10.001 (2017).

86 Adhikari, K. *et al.* A genome-wide association study identifies multiple loci for variation in human ear morphology. *Nat Commun* **6**, 7500, doi:10.1038/ncomms8500 (2015).

87 Churchill, S. E. & Vansickle, C. Pelvic Morphology in Homo erectus and Early Homo. *Anat Rec (Hoboken)* **300**, 964-977, doi:10.1002/ar.23576 (2017).

88 Fischer, B. & Mitteroecker, P. Covariation between human pelvis shape, stature, and head size alleviates the obstetric dilemma. *Proc Natl Acad Sci U S A* **112**, 5655-5660, doi:10.1073/pnas.1420325112 (2015).

89 Tzchori, I. *et al.* LIM homeobox transcription factors integrate signaling events that control three-dimensional limb patterning and growth. *Development* **136**, 1375-1385, doi:10.1242/dev.026476 (2009).

90 Shetty, A. S. *et al.* Lhx2 regulates a cortex-specific mechanism for barrel formation. *Proc Natl Acad Sci U S A* **110**, E4913-4921, doi:10.1073/pnas.1311158110 (2013).

91 Wang, P. *et al.* ccdc141 is required for left-right axis development by regulating cilia formation in the Kupffer's vesicle of zebrafish. *J Genet Genomics* **51**, 934-946, doi:10.1016/j.jgg.2024.07.014 (2024).

92 Lieberman, D. E., Pearson, O. M. & Mowbray, K. M. Basicranial influence on overall cranial shape. *J Hum Evol* **38**, 291-315, doi:10.1006/jhev.1999.0335 (2000).

93 Young, M. *et al.* The developmental impacts of natural selection on human pelvic morphology. *Sci Adv* **8**, eabq4884, doi:10.1126/sciadv.abq4884 (2022).

94 Capellini, T. D. *et al.* Control of pelvic girdle development by genes of the Pbx family and Emx2. *Dev Dyn* **240**, 1173-1189, doi:10.1002/dvdy.22617 (2011).

95 Golonzhka, O. *et al.* Pbx Regulates Patterning of the Cerebral Cortex in Progenitors and Postmitotic Neurons. *Neuron* **88**, 1192-1207, doi:10.1016/j.neuron.2015.10.045 (2015).

96 Berthelot, C. How the pelvis evolved to enable human bipedalism. *Nature* **645**, 858-859, doi:10.1038/d41586-025-02424-x (2025).

97 Senevirathne, G. *et al.* The evolution of hominin bipedalism in two steps. *Nature* **645**, 952-963, doi:10.1038/s41586-025-09399-9 (2025).

98 Girskis, K. M. *et al.* Rewiring of human neurodevelopmental gene regulatory programs by human accelerated regions. *Neuron* **109**, 3239-3251 e3237, doi:10.1016/j.neuron.2021.08.005 (2021).

99 Okamoto, A. S., Coveney, C. R., Ganapathee, D. S. & Capellini, T. D. In Vitro Massively Parallel Screening of Human Regulatory Elements Involved in Postcranial Skeletal Development for Differential Activity Compared to Chimpanzee. *Genome Biol Evol* **18**, doi:10.1093/gbe/evag121 (2026).

100 Davison, C., Bedo, G. & Zolessi, F. R. Zebrafish Slit2 and Slit3 Act Together to Regulate Retinal Axon Crossing at the Midline. *J Dev Biol* **10**, doi:10.3390/jdb10040041 (2022).

101 Suarez, P., Baumer, K. & Hall, D. Further insight into the global variability of the OCA2-HERC2 locus for human pigmentation from multiallelic markers. *Sci Rep* **11**, 22530, doi:10.1038/s41598-021-01940-w (2021).

102 Li, X. *et al.* Assessment of iris pigment epithelium depolarization and associated factors using polarization-sensitive anterior segment optical coherence tomography. *Int Ophthalmol* **45**, 122, doi:10.1007/s10792-025-03473-x (2025).

103 Franchini, L. F. & Pollard, K. S. Can a few non-coding mutations make a human brain? *Bioessays* **37**, 1054-1061, doi:10.1002/bies.201500049 (2015).

104 Liu, J. *et al.* A human-specific enhancer fine-tunes radial glia potency and corticogenesis. *Nature*, doi:10.1038/s41586-025-09002-1 (2025).

105 Boyd, J. L. *et al.* Human-chimpanzee differences in a FZD8 enhancer alter cell-cycle dynamics in the developing neocortex. *Curr Biol* **25**, 772-779, doi:10.1016/j.cub.2015.01.041 (2015).

106 Hung, S. S., Tsai, P. S., Po, C. W. & Hou, P. S. Pax6 isoforms shape eye development: Insights from developmental stages and organoid models. *Differentiation* **137**, 100781, doi:10.1016/j.diff.2024.100781 (2024).

107 Wang, X. *et al.* Developmental origin of the posterior pigmented epithelium of iris. *Cell Biochem Biophys* **71**, 1067-1076, doi:10.1007/s12013-014-0310-0 (2015).

108 Hayashi, T., Huang, J. & Deeb, S. S. RINX(VSX1), a novel homeobox gene expressed in the inner nuclear layer of the adult retina. *Genomics* **67**, 128-139, doi:10.1006/geno.2000.6248 (2000).

109 Gajdzis, M. *et al.* Ephrin Receptors (Eph): EphA1, EphA5, and EphA7 Expression in Uveal Melanoma-Associations with Clinical Parameters and Patient Survival. *Life (Basel)* **10**, doi:10.3390/life10100225 (2020).

110 Bredrup, C. *et al.* High myopia-excavated optic disc anomaly associated with a frameshift mutation in the MYC-binding protein 2 gene (MYCBP2). *Am J Ophthalmol* **159**, 973-979 e972, doi:10.1016/j.ajo.2015.01.021 (2015).

111 Micheal, S. *et al.* Whole exome sequencing identifies a heterozygous missense variant in the PRDM5 gene in a family with Axenfeld-Rieger syndrome. *Neurogenetics* **17**, 17-23, doi:10.1007/s10048-015-0462-0 (2016).

112 Shen, W. *et al.* EDIL3 knockdown inhibits retinal angiogenesis through the induction of cell cycle arrest in vitro. *Mol Med Rep* **16**, 4054-4060, doi:10.3892/mmr.2017.7122 (2017).

113 Klatt, A. R., Becker, A. K., Neacsu, C. D., Paulsson, M. & Wagener, R. The matrilins: modulators of extracellular matrix assembly. *Int J Biochem Cell Biol* **43**, 320-330, doi:10.1016/j.biocel.2010.12.010 (2011).

114 Paulsson, M. & Wagener, R. Matrilins. *Methods Cell Biol* **143**, 429-446, doi:10.1016/bs.mcb.2017.08.018 (2018).

115 Bailey, A. J. Structure, function and ageing of the collagens of the eye. *Eye (Lond)* **1 (Pt 2)**, 175-183, doi:10.1038/eye.1987.34 (1987).

116 Heger, L. *et al.* CLEC10A Is a Specific Marker for Human CD1c(+) Dendritic Cells and Enhances Their Toll-Like Receptor 7/8-Induced Cytokine Secretion. *Front Immunol* **9**, 744, doi:10.3389/fimmu.2018.00744 (2018).

117 Lin, J. B. *et al.* Aqueous Macrophages Contribute to Conserved CCL2 and CXCL10 Gradients in Uveitis. *Ophthalmol Sci* **4**, 100453, doi:10.1016/j.xops.2023.100453 (2024).

118 Bowles, B. *et al.* TSPEAR variants are primarily associated with ectodermal dysplasia and tooth agenesis but not hearing loss: A novel cohort study. *Am J Med Genet A* **185**, 2417-2433, doi:10.1002/ajmg.a.62347 (2021).

119 Martin, J. E. *et al.* Identification of CSK as a systemic sclerosis genetic risk factor through Genome Wide Association Study follow-up. *Hum Mol Genet* **21**, 2825-2835, doi:10.1093/hmg/dds099 (2012).

120 Thaweesapphithak, S. *et al.* Novel EDAR death domain variants in Thai patients with ectodermal dysplasia: clinical, molecular, and scoping review insights. *BMC Oral Health* **25**, 1737, doi:10.1186/s12903-025-07103-x (2025).

121 Prescott, S. L. *et al.* Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* **163**, 68-83, doi:10.1016/j.cell.2015.08.036 (2015).

122 Zhang, M. *et al.* Genetic variants underlying differences in facial morphology in East Asian and European populations. *Nat Genet* **54**, 403-411, doi:10.1038/s41588-022-01038-7 (2022).

123 Accogli, A. *et al.* Loss of Neuron Navigator 2 Impairs Brain and Cerebellar Development. *Cerebellum* **22**, 206-222, doi:10.1007/s12311-022-01379-3 (2023).

124 Paternoster, L. *et al.* Genome-wide association study of three-dimensional facial morphology identifies a variant in PAX3 associated with nasion position. *Am J Hum Genet* **90**, 478-485, doi:10.1016/j.ajhg.2011.12.021 (2012).

125 Li, Y. *et al.* EDAR, LYPLAL1, PRDM16, PAX3, DKK1, TNFSF12, CACNA2D3, and SUPT3H gene variants influence facial morphology in a Eurasian population. *Hum Genet* **138**, 681-689, doi:10.1007/s00439-019-02023-7 (2019).

126 Santagati, F. & Rijli, F. M. Cranial neural crest and the building of the vertebrate head. *Nat Rev Neurosci* **4**, 806-818, doi:10.1038/nrn1221 (2003).

127 Burkardt, D. D., Tatton-Brown, K., Dobyns, W. & Graham, J. M., Jr. Approach to overgrowth syndromes in the genome era. *Am J Med Genet C Semin Med Genet* **181**, 483-490, doi:10.1002/ajmg.c.31757 (2019).

128 Kamien, B. *et al.* A Clinical Review of Generalized Overgrowth Syndromes in the Era of Massively Parallel Sequencing. *Mol Syndromol* **9**, 70-82, doi:10.1159/000484532 (2018).

129 Aponte, J. D. *et al.* Shared genetic basis and structure of syndromic and normal facial variation. *Am J Hum Genet* **113**, 1987-2000, doi:10.1016/j.ajhg.2026.07.014 (2026).

130 Lacruz, R. S. *et al.* The evolutionary history of the human face. *Nat Ecol Evol* **3**, 726-736, doi:10.1038/s41559-019-0865-7 (2019).

131 Endo, C. *et al.* Genome-wide association study in Japanese females identifies fifteen novel skin-related trait associations. *Sci Rep* **8**, 8974, doi:10.1038/s41598-018-27145-2 (2018).

132 Jacobs, L. C. *et al.* Intrinsic and extrinsic risk factors for sagging eyelids. *JAMA Dermatol* **150**, 836-843, doi:10.1001/jamadermatol.2014.27 (2014).

133 Guo, J. *et al.* Variation and signatures of selection on the human face. *J Hum Evol* **75**, 143-152, doi:10.1016/j.jhevol.2014.08.001 (2014).

134 Little, A. C., Jones, B. C. & DeBruine, L. M. Facial attractiveness: evolutionary based research. *Philos Trans R Soc Lond B Biol Sci* **366**, 1638-1659, doi:10.1098/rstb.2010.0404 (2011).

135 Lee, A. J. *et al.* Facial averageness and genetic quality: Testing heritability, genetic correlation with attractiveness, and the paternal age effect. *Evol Hum Behav* **37**, 61-66, doi:10.1016/j.evolhumbehav.2015.08.003 (2016).

136 Mitchem, D. G. *et al.* Estimating the sex-specific effects of genes on facial attractiveness and sexual dimorphism. *Behav Genet* **44**, 270-281, doi:10.1007/s10519-013-9627-5 (2014).

137 Sheehan, M. J. & Nachman, M. W. Morphological and population genomic evidence that human faces have evolved to signal individual identity. *Nat Commun* **5**, 4800, doi:10.1038/ncomms5800 (2014).

138 Joshi, R. S. *et al.* Look-alike humans identified by facial recognition algorithms show genetic similarities. *Cell Rep* **40**, 111257, doi:10.1016/j.celrep.2022.111257 (2022).

139 Bruce, V. & Young, A. W. *Face perception*. (Psychology Press, 2012).

140 Wilmer, J. B. *et al.* Human face recognition ability is specific and highly heritable. *Proc Natl Acad Sci U S A* **107**, 5238-5241, doi:10.1073/pnas.0913053107 (2010).

141 Sun, Y. *et al.* Human genetics of face recognition: discovery of MCTP2 mutations in humans with face blindness (congenital prosopagnosia). *Genetics* **227**, doi:10.1093/genetics/iyae047 (2024).

142 Anderson, D. *et al.* First genome-wide association study in an Australian aboriginal population provides insights into genetic risk factors for body mass index and type 2 diabetes. *PLoS One* **10**, e0119333, doi:10.1371/journal.pone.0119333 (2015).

143 Castellucci, L. C. *et al.* A Genome-wide Association Study Identifies SERPINB10, CRLF3, STX7, LAMP3, IFNG-AS1, and KRT80 As Risk Loci Contributing to Cutaneous Leishmaniasis in Brazil. *Clin Infect Dis* **72**, e515-e525, doi:10.1093/cid/ciaa1230 (2021).

144 Jamieson, S. E. *et al.* Common and Rare Genetic Variants That Could Contribute to Severe Otitis Media in an Australian Aboriginal Population. *Clin Infect Dis* **73**, 1860-1870, doi:10.1093/cid/ciab216 (2021).

145 Evison, M. P. *A Genome-Wide Association Study of Face Shape and Appearance* Master of Philosophy thesis, University of Newcastle upon Tyne, (2026).

146 Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet* **50**, 229-237, doi:10.1038/s41588-017-0009-4 (2018).

147 Xiong, Z. *et al.* Combined genome-wide association study of facial traits in Europeans increases explained variance and improves prediction. *Nat Commun* **16**, 6562, doi:10.1038/s41467-025-61761-7 (2025).

148 Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177-1186, doi:10.1016/j.cell.2017.05.038 (2017).

149 Kullo, I. J. Clinical use of polygenic risk scores: current status, barriers and future directions. *Nat Rev Genet* **27**, 246-263, doi:10.1038/s41576-025-00900-8 (2026).

150 Bannister, J. J. *et al.* Comparing 2D and 3D representations for face-based genetic syndrome diagnosis. *Eur J Hum Genet* **31**, 1010-1016, doi:10.1038/s41431-023-01308-w (2023).

151 Walsh, S. *et al.* IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Sci Int Genet* **5**, 170-180, doi:10.1016/j.fsigen.2010.02.004 (2011).

152 Frudakis, T. N. *Molecular photofitting : predicting ancestry and phenotype using DNA*. (Elsevier/Academic Press, 2008).

153 Claes, P. *et al.* Modeling 3D facial shape from DNA. *PLoS Genet* **10**, e1004224, doi:10.1371/journal.pgen.1004224 (2014).

154 Qiao, L. *et al.* Genome-wide variants of Eurasian facial shape differentiation and a prospective model of DNA based face prediction. *J Genet Genomics* **45**, 419-432, doi:10.1016/j.jgg.2018.07.009 (2018).

155 Tozzo, P., Politi, C., Delicati, A., Gabbin, A. & Caenazzo, L. External visible characteristics prediction through SNPs analysis in the forensic setting: a review. *Front Biosci (Landmark Ed)* **26**, 828-850, doi:10.52586/4991 (2021).

156 Jiao, M. *et al.* De Novo Reconstruction of 3D Human Facial Images from DNA Sequence. *Adv Sci (Weinh)* **12**, e2414507, doi:10.1002/advs.202414507 (2025).

157 Wagner, J. K., Claessens, N., Maloney, C. M. & Claes, P. Comment on "De Novo Reconstruction of 3D Human Facial Images from DNA Sequence". *Adv Sci (Weinh)* **13**, e13207, doi:10.1002/advs.202513207 (2026).

158 Jiao, M., Li, J., Zuo, C., Wang, S. & Chen, L. Reply to "Comment on 'De Novo Reconstruction of 3D Human Facial Images From DNA Sequence'". *Adv Sci (Weinh)* **13**, e73979, doi:10.1002/advs.73979 (2026).

159 Buchanan, E. P., Xue, A. S. & Hollier, L. H., Jr. Craniofacial syndromes. *Plast Reconstr Surg* **134**, 128e-153e, doi:10.1097/PRS.0000000000000308 (2014).

160 Evison, M. P. & Vorder Bruegge, R. W. *Computer-aided forensic facial comparison*. 1st edn, (Taylor & Francis Group, 2010).

161 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575, doi:10.1086/519795 (2007).

162 HapMap3. (2026).

163 Boehringer, S. *et al.* Genetic determination of human facial morphology: links between cleft-lips and normal variation. *Eur J Hum Genet* **19**, 1192-1197, doi:10.1038/ejhg.2011.110 (2011).

164 Lucock, M. D. The evolution of human skin pigmentation: A changing medley of vitamins, genetic variability, and UV radiation during human expansion. *Am J Biol Anthropol* **180**, 252-271, doi:10.1002/ajpa.24564 (2023).

165 Boughton, A. P. *et al.* LocusZoom.js: interactive and embeddable visualization of genetic association study results. *Bioinformatics* **37**, 3017-3018, doi:10.1093/bioinformatics/btab186 (2021).

166 Xiong, Z. *et al.* Novel genetic loci affecting facial shape variation in humans. *Elife* **8**, doi:10.7554/eLife.49898 (2019).

167 Wu, W. *et al.* Whole-exome sequencing identified four loci influencing craniofacial morphology in northern Han Chinese. *Hum Genet* **138**, 601-611, doi:10.1007/s00439-019-02008-6 (2019).

168 Cha, S. *et al.* Identification of five novel genetic loci related to facial morphology by genome-wide association studies. *BMC Genomics* **19**, 481, doi:10.1186/s12864-018-4865-9 (2018).

169 Crouch, D. J. M. *et al.* Genetics of the human face: Identification of large-effect single gene variants. *Proc Natl Acad Sci U S A* **115**, E676-E685, doi:10.1073/pnas.1708207114 (2018).

170 Lee, M. K. *et al.* Genome-wide association study of facial morphology reveals novel associations with FREM1 and PARK2. *PLoS One* **12**, e0176566, doi:10.1371/journal.pone.0176566 (2017).

171 Qian, W. *et al.* Genetic evidence for facial variation being a composite phenotype of cranial variation and facial soft tissue thickness. *J Genet Genomics* **49**, 934-942, doi:10.1016/j.jgg.2022.02.020 (2022).

Acknowledgements

The authors thank Angela Martins, Raffaela Francisco, Teresa Cabral, Maria da Mota and Daniel Fu Ren for their assistance with data collection and administration, and Stephen W. Scherer and The Centre for Applied Genomics/CGEn at SickKids Hospital in Toronto, where the genotyping was performed.

Funding

Support for genotyping was provided by a University of Toronto Connaught Award and biometric equipment by TSWG award T216E to M.P.E. This research was funded in whole, or in part, by the Wellcome Trust [Grant number 219424/Z/19/Z to HJC]. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Author information

Authors and affiliations

Population Health Sciences Institute (PHSI), Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom

Martin Evison, Rebecca Darlay, Heather Cordell

Department of Pathology and Forensic Medicine, University of São Paulo Faculty of Medicine Ribeirão Preto, Ribeirão Preto, Brazil

Marco Guimaraes

Department of Pathology, Escola Paulista de Medicina, Federal University of São Paulo, São Paulo, Brazil

Edna Iwamura

School of Mathematical Sciences, The University of Nottingham, Nottingham, United Kingdom

Ian Dryden

ForensicVB LLC, Fredericksburg, Virginia, United States of America

Richard W. Vorder Bruegge

Contributions

Overall study design: M.P.E. 3D shape and biometric analysis of the face: I.L.D., M.P.E., R.V.B. Data acquisition and processing: M.P.E, M.A.G., E.S.M.I., GWAS analysis and interpretation: M.P.E, H.J.C., R.D. Overall interpretation of findings and writing of first draft: M.P.E. Review, editing and approval of final draft: all authors.

Data availability

Summary statistics used in GCTA analysis for 36 distance measurements completed on 378 subjects and for 18 RGB colour values completed on 258 subjects can accessed at https://my.locuszoom.org using the generic search term “Face GWAS” or, for individual studies, the study name (Evison),

PMID and GWAS label (“Face GWAS” + “g” for genotyped or “i” for imputed + “as” if adjusted for age and sex). Informed consent does not allow sharing of individual-level genetic, geometric or biometric data.

Ethics declarations

Ethical review

This study was reviewed by the University of Toronto (Office of Research Ethics Protocol 22698) and the University of São Paulo (Comité de Ética em Pesquisa Hospital das Clinicas da Faculdade de Medicine de Ribeirão Preto da Universidade de São Paulo Processo 984/2007), who approved all experimental protocols. The research was conducted with informed consent of the donors in accordance with relevant guidelines and regulations of these institutions.

The facial images used in Figure 4 and Supplementary Figure 1 of the manuscript are included with the informed consent of the donor, coauthor MPE.

Competing interests

The authors declare no competing interests

Table 1 Pairwise distances generating associations of genome-wide significance.

Phenotype	Chr:Pos[b]	Allele[c]	rsID	Nearest gene	-log10(p)	b[d]
g-se	11:2,864,406	C/T	rs77802206	*KCNQ1-AS1*	9.123	5.305
g-se[a]	11:2,864,406	C/T	rs77802206	*KCNQ1-AS1*	9.122	5.295
g-se[a]	19:10,563,883	C/T	rs142240681	*KRI1*	7.509	5.704
g-se	20:21,080,435	C/G	rs144020852	*LINC00237*	7.382	4.543
g-se[a]	3:109,759,563	A/G	rs12629891	*AC078980.1*	7.346	1.648
g-pg[a]	16:30,986,831	T/C	rs200458768	*AC135048.4, HSD3B7*	7.898	-5.770
se-ls	12:41,082,188	C/A	rs192897690	*CNTN1*	8.452	-5.325
se-ls[a]	12:41,082,188	C/A	rs192897690	*CNTN1*	7.641	-4.490
se-pg[a]	16:30,986,831	T/C	rs200458768	*AC135048.4, HSD3B7*	7.744	-4.728
se-enr	1:8,252,255	C/T	rs116227640	*AL358876.2*	8.021	3.299
se-enr[a]	1:8,252,255	C/T	rs116227640	*AL358876.2*	7.739	2.987
exl-enl	4:183,007,311	A/T	rs78191818	*AC019193.4*	7.717	1.770
enl-enr	6:151,225,805	T/C	rs6908043	*AL138733.1, AL356535.1*	7.412	5.735
enr-obir[a]	2:154,685,188	A/G	rs144283929	*AC061961.1*	8.156	7.884
enr-obir[a]	7:76,071,718	G/A	rs140879076	*AC005077.4*	7.485	8.922
enr-obir[a]	16:49,176,277	A/G	rs75174906	*AC044798.3*	7.411	6.678
all-alr[a]	6:18,203,788	G/A	rs112479069	*KDM1B*	7.777	-4.559
cprimel-cprimer	16:19,691,944	G/A	rs150699629	*VPS35L*	8.706	4.838
cprimel-cprimer	4:5,464,924	T/A	rs114045090	*STK32B*	7.980	5.309
cprimel-cprimer[a]	16:19,691,944	G/A	rs150699629	*VPS35L*	7.696	4.232
cprimel-cprimer	3:59,952,694	C/T	rs111975747	*FHIT*	7.549	-5.172
cprimel-cprimer[a]	5:136,567,071	C/T	rs182193009	*AC112178.1*	7.407	3.892

[a]GWAS adjusted for age and sex, [b]Chr:Pos: Chromosome and position (bp), [c]the effect allele is given first, [d]b: effect size.

Table 2 RGB colour values generating associations of genome-wide significance.

Phenotype	Chr:Pos[b]	Allele[c]	rsID	Nearest gene	-log10(p)	b[d]
il-r	5:169,377,956	A/T	rs72839821	*SLIT3*	7.781	10.966
il-g	7:89,193,584	T/C	rs150524934	*ZNF804B*	10.866	14.326
il-g	10:36,235,227	G/T	rs56389800	*AL355300.1*	9.593	16.553
il-g	20:28,871,721	A/G	rs1482118314	*FRG1DP*	8.209	13.513
il-g	15:28,165,345	C/T	rs12916300	*HERC2*	7.995	6.157
il-g[a]	7:89,193,584	T/C	rs150524934	*ZNF804B*	7.987	11.388
il-g[a]	7:143,395,091	C/T	rs117651735	*EPHA1*	7.866	23.293
il-g	1:43,142,099	G/T	rs55725655	*FAM183A*	7.748	18.089
il-g	7:143,395,091	C/T	rs117651735	*EPHA1*	7.309	23.834
il-b	7:89,193,584	T/C	rs150524934	*ZNF804B*	8.980	10.584
il-b	15:28,165,345	C/T	rs12916300	*HERC2*	8.384	5.244
il-b[a]	15:28,165,345	C/T	rs12916300	*HERC2*	7.717	4.695
il-b	13:77,199,751	C/A	rs190618997	*MYCBP2*	7.515	21.156
il-b	10:36,235,227	G/T	rs56389800	*AL355300.1*	7.306	11.695
ir-g	7:89,193,584	T/C	rs150524934	*ZNF804B*	11.577	14.744
ir-g	10:36,235,227	G/T	rs56389800	*AL355300.1*	10.954	17.686
ir-g[a]	7:89,193,584	T/C	rs150524934	*ZNF804B*	8.579	11.648
ir-g	6:142,893,161	C/A	rs150629833	*HIVEP2*	8.636	32.655
ir-g	20:28,871,721	A/G	rs1482118314	*FRG1DP*	8.123	13.367
ir-g	7:143,357,936	G/A	rs79172074	*FAM131B*	7.894	21.427
ir-g	1:119,648,688	T/A	rs183872671	*PHGDH*	7.793	34.457
ir-g[a]	7:143,395,091	C/T	rs117651735	*EPHA1*	7.761	22.750
ir-g	17:7,070,585	A/G	rs8070374	*CLEC10A*	7.593	9.345
ir-g[a]	1:53,756,994	T/C	rs142689889	*NDC1*	7.533	21.143
ir-g[a]	5:84,061,259	A/G	rs150664316	*EDIL3*	7.521	25.719
ir-g	5:77,552,446	T/C	rs115556475	*WDR41*	7.517	9.345
ir-g	12:65,973,239	G/A	rs148869671	*HMGA2*	7.394	30.004
ir-g	18:39,817,601	A/G	rs11660497	*MIR924HG*	7.378	30.004
ir-g[a]	10:36,235,227	G/T	rs56389800	*AL355300.1*	7.333	13.211
ir-b	7:89,193,584	T/C	rs150524934	*ZNF804B*	9.343	10.818
ir-b	1:30,658,245	G/A	rs114640642	*MATN1*	9.071	23.356
ir-b	10:36,235,227	G/T	rs56389800	*AL355300.1*	8.087	12.368
ir-b[a]	1:30,658,245	G/A	rs114640642	*MATN1*	8.047	20.388
ir-b[a]	1:53,756,994	T/C	rs142689889	*NDC1*	7.803	17.746
ir-b	18:39,817,601	A/G	rs11660497	*MIR924HG*	7.521	31.870
vl-r	2:109,483,310	A/G	rs73953132	*SH3RF3*	8.246	42.542
vl-r	4:143,582,915	C/T		*FREM3*	7.473	61.098

[a]GWAS adjusted for age and sex, [b]Chr:Pos: Chromosome and position (bp), [c]the effect allele is given first, [d]b: effect size.

Table 3: Description of 3D anthropometric landmarks.

Label	Name	Description
g	Glabella	The most prominent midline point between the eyebrows
sl	Sublabiale	Determines the lower border of the lower lip and upper border of the chin
pg	Pogonion	The most anterior point of the chin
en	Endocanthion (l, r)	The point at the inner commissure of the eye fissure
ex	Exocanthion (l, r)	The point at the outer commissure of the eye fissure
P	Pupil (l, r)	Determined when the head is in the rest position and the eye is looking straight forward
pi	Palpebrale inferius (l, r)	The lowest point in the mid portion of the free margin of each lower eyelid
se	Selion	The deepest landmark located at the bottom of the nasofrontal angle
prn	Pronasale	The most protruded point of the apex nasi
al	Alar (l, r)	The most lateral point on each alar contour
cprime	Highest point of columella (l, r)	The point at each columella crest, level with the tip of the corresponding nostril
ls	Labiale superius	The midpoint of the upper vermillion line
li	Labiale inferius	The midpoint of the lower vermillion line
sto	Stomion	The intersection of the vertical facial midline and the horizontal labial fissure
ch	Cheilion (l, r)	The point located at each labial commissure
sa	Superaurale (l, r)	The highest point on the free margin of the auricle
sba	Subaurale (l, r)	The lowest point on the free margin of the ear lobe
pa	Postaurale (l, r)	The most posterior point on the free margin of the ear
obi	Otobasion inferius (l, r)	The point of attachment of the ear lobe on the cheek

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigure1.pptx
- SupplementaryFileS1.xlsx
- SupplementaryFileS2.docx
- SupplementaryFileS3.docx