Heliyon 10 (2024) e39173

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

High heritability of human facial traits reveals associations with *CNTLN, BRCA1*, and *TMPRSS6* loci in Korean families

Donghyun Lee [a,1], Hyo-Jeong Ban [b,1], Kyung-Won Hong [c], Jong Young Lee [a,**], Seongwon Cha [b,*]

[a] *Oneomics Co., Ltd., Bucheon-si, Gyeonggi-do, 14585, South Korea*
[b] *KM Data Division, Korea Institute of Oriental Medicine, Daejeon, 34054, South Korea*
[c] *Theragen Bio Co., Ltd., Seongnam-si, Gyeonggi-do, 13493, South Korea*

ARTICLE INFO

Keywords:
Facial trait
CNTLN
BRCA1
TMPRSS6
Single nucleotide polymorphism

ABSTRACT

Facial features are determined by interactions between genetic and environmental factors. However, genes underlying facial similarities in individuals from the same family remain less explored. To identify genetic variants associated with heritable facial features, we investigated familial (parent–offspring) associations and estimated familial correlation and heritability using 39 facial measurements in 408 individuals from 117 Korean families. Facial trait heritability ranged from 0.124 to 0.669. Longitudinal facial growth-related traits were highly heritable, including distances from the nasion to right alare ($h^2 = 0.668898$), pogonion to midendocanthion ($h^2 = 0.661557$), subnasale to midendocanthion ($h^2 = 0.656882$), and morphological facial height ($h^2 = 0.654376$). We identified the top three significant genome-wide associated variants in the eye, nose, and lip–jaw regions. *CNTLN* (rs10511632: beta $= -0.02696$, $p = 1.146 \times 10^{-9}$) and *BRCA1* (rs397509305: beta $= 0.02741$, $p = 7.17 \times 10^{-9}$) loci were associated with distance from the nasion to the right alare. The *TMPRSS6* (rs228913: beta $= 0.05101$, $p = 3.68 \times 10^{-9}$) locus was associated with the distance from the labiale superius to the pogonion and lower facial height. These associations were maintained in an independent unrelated population. In conclusion, we identified new gene variants associated with longitudinal facial morphology that may affect individual facial differences, which has important implications for clinical and forensic applications.

1. Introduction

Readily visible differences in craniofacial characteristics among various facial features (the head, nose, mouth, chin, and lips),

Abbreviations: enL-exL, left palpebral fissure length; ls-pg, labiale superius and pogonion; men-pg, pogonion and midendocanthion; enR-enL, intercanthal width; sn-men, midendocanthion; n-alL, nasion to left alare; n-alR, nasion to right alare; alR-alL, nasal width; prn-sn, nasal tip height; sn-pg, lower facial height; ls-cphL, left length of Cupid's bow; prn-alL, pronasale to right alare; CNCC, cranial neural crest cell; REML, restricted maximum likelihood; mAF, minor allele frequency.

* Corresponding author.
** Corresponding author.
E-mail addresses: jylee@oneomics.co.kr (J.Y. Lee), scha@kiom.re.kr (S. Cha).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e39173
Received 18 April 2024; Received in revised form 8 October 2024; Accepted 8 October 2024
Available online 10 October 2024
2405-8440/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

determined by cartilaginous and skeletal structures, make each individual human face unique and identifiable [1]. Understanding the genetic basis of the facial morphology of an individual has several important implications for clinical and forensic applications, in which genetic information provides important insights [1,2].

Similar facial features observed in families, such as between parents and offspring [3], suggest the existence of gene variants with considerable effects on the facial morphology of an individual. Importantly, special regions of the brain are dedicated to facial recognition in humans and other primates [4]. Several studies, including human heritability studies using twin and parent–offspring designs, have suggested a strong genetic basis for the observed variations in facial morphology [5,6].

Genetic variants and normal-range variations in facial morphology have been studied in various ethnic groups; for example, genome-wide association studies (GWASs) have reported over 70 genetic loci associated with human facial features [7–15]. The limited overlap of findings from independent GWASs is likely attributable to methodological differences in facial phenotyping and biological factors, including a different genetic architecture of functional facial variations across human populations. For example, studies have previously reported the morphological divergence between European and Asian populations with respect to the nose, brow ridges, and cheek–jaw regions [16]. However, studies identifying face-associated genetic variants in familial populations are lacking.

This study aimed at (1) identifying single nucleotide polymorphisms (SNPs) highly associated with 39 facial measurements and (2) estimating the SNP-based heritability for each facial trait and genetic correlation in parent–offspring in Korean families. We suggest that a combination of family-based and population-based association analyses can provide an accurate and powerful approach for identifying and characterizing the full range of craniofacial shape variants.

2. Materials and methods

2.1. Ethics statement

This study was performed in accordance with the principles of the Declaration of Helsinki. All participants provided written informed consent (for minors aged under 18, the consent was obtained from minors in addition to parental/guardian consent.), and this study was approved by the Institutional Review Board of the Korea Institute of Oriental Medicine (IRB No.: I-2007/006-002, I-1702/001–002).

2.2. Study participants

Family participants were recruited from South Korea (16 sites) from 2007 to 2012 by the Korean Medicine Data Center of the Korean Institute of Oriental Medicine. Among the 422 participants selected, 14 with unclear family relationships or plastic surgery were excluded, leaving 408 participants. Of these, 200 were men (ages 10.1–78.0), and 208 were women (ages 7.3–75.4). All participants belonged to 117 families (Supplementary Table S1), in which there were no half-sibs and inbreeding. All pedigrees were two generation parent-offspring, and four large families including three generations were divided into two-generation parent-offspring pedigrees (Supplementary Fig. S1 and Fig. S2).

Replication participants were selected from the recruited population in South Korea (Ansan and Ansung) from 2009 to 2012 for the Korean Genome and Epidemiology Study (KoGES) [17]. Facial analyses have been described in a previous study [8]. All 5643 KoGES participants (2648 men and 2995 women) were included in the present study.

2.3. Facial traits

Researchers ensured that participants did not wear any cosmetic makeup and were photographed in frontal view using a digital camera (DSLR Nikon D90 with a Nikon AF 50-mm F1.8D lens, 3216 × 2136 pixels; Nikon, Tokyo, Japan). Photographs for facial analysis were taken by trained investigators according to a standard protocol, as reported in a previous study [8]. The participants were photographed with a neutral expression in both frontal and profile views under the following standard conditions: the hair was pulled back with a hair band; the center points of the two pupils and the two points connected between the facial contour and upper auricular perimeters were on the same horizontal line; and a ruler was placed approximately 10 mm below the chin to convert pixels into millimeters [8,18]. To extract feature points from the two-dimensional (2D) face images of all participants, we used a facial landmark detector included in the Dlib library, which detects faces and identifies facial landmarks [19]. In total, 25 points (Supplementary Table S2, Supplementary Fig. S3) were marked on the face, and the distances between pairs of points were used as variables. In total, 39 facial traits were measured by connecting two points using a linear distance (Supplementary Table S3), after which the distances between the centers of both eyes were normalized to 1 for all images; that is, the 39 traits were given as relative proportional distance values to the eye center distance. The eyes, nose, lip–jaw, and facial size, as four facial regions, were associated with 8, 12, 16, and 3 traits, respectively.

2.4. SNP genotyping

Genomic DNA was extracted from the peripheral blood of family participants using the QiAmp DNA blood kit (QIAGEN, Germantown, MD, USA). All family samples included in the analysis were genotyped using the Axiom APMRA kit (Thermo Fisher Scientific, Waltham, MA, USA), which contains over 750,000 SNPs, including 50,000 novel markers covering East and South Asian

populations based on the human genome version 19 (build 37). Data cleaning was performed using PLINK 1.9 (https://www.cog-genomics.org/plink/1.9/) [20]. Before performing the association analysis, we performed GWAS quality-control procedures. In total, 336,712 from 668,758 variants, including small insertions/deletions (indels), were considered for analysis. These variants passed the quality-control test with a genotype call rate ≥90 % and a minor allele frequency (mAF) ≥0.01, while an exact test of the Hardy–Weinberg equilibrium produced a *p*-value $\geq 1 \times 10^{-4}$. Thereafter, we checked the occurrence of Mendelian errors to exclude SNPs that failed to show transmission according to Mendelian inheritance laws in family-based studies. SNP genotyping for replication samples has been described in a previous study [8] and was performed using an Affymetrix Genome-Wide Human SNP array 5.0 (Affymetrix, Santa Clara, CA, USA).

2.5. Narrow-sense heritability in facial shape

GCTA (Genome-wide Complex Trait Analysis) was used to estimate the proportion of variance in each facial trait that was explained by all GWAS SNPs. We estimated the narrow-sense heritability (h^2) that can be explained by common genetic variations using the restricted maximum likelihood (REML) method in the GCTA software. In this study, narrow-sense heritability stands for the ratio of the variance that can be explained by genetic factors and phenotypic variance ($h^2 = V(G)/Vp$). The REML method provides more unbiased estimates of variance components compared with the maximum likelihood method [21]. We estimated the amount of phenotypic variance captured by SNPs via a linear mixed model using age and sex of the genetic relationship matrix (GRM) as fixed effect covariates, as well as a random effect with a covariance matrix corresponding to the GRM. A likelihood ratio test was performed by default, examining the significance of the random effect of the GRM on the fit of the model [22].

2.6. Family-based genetic associations in facial shape

Analyses of family-based genetic associations of facial traits were performed using PLINK (version 1.9) [20]. To ensure accurate family relationships, we used identity by descent (IBD) calculations using the –genome option in PLINK to verify participant relatedness. This process included calculating identity by state (IBS) based on the average of shared alleles between pairs of individuals. For the secondary subset, family relationships were confirmed in the entire sample set using IBD analysis. The IBD results confirmed that all participants were from different families. Large families were not excluded during analysis because family structure was considered when performing the association test. The –qfam-total option of the PLINK program used for this analysis allowed the utilization of overall association information while considering family structure, potentially detecting stronger association signals.

PLINK provides a QFAM analysis (QFAM-total) that tests the association of quantitative traits in family-based data. The standard regression of a phenotype on a genotype ignores family structure, whereas QFAM uses a permutation procedure (permuting genotype rather than phenotype) to control for the nonindependence of individuals within the same family [20]. Linear models were used to test the presence of genetic associations between each of the extracted factors and each SNP under an additive genetic model. QFAM-total is a total association test that splits the between- and within-family components and performs a linear regression of the phenotype on the genotype, as follows:

$$\widehat{y}_{ij} = \mu + \beta_b b_{ij} + \beta_w w_{ij},$$

where $\widehat{y}_{ij}$ is the vector of the phenotype of individual j in family i, and the between-family b_{ij} and within-family w_{ij} components are calculated using parental genotypes, if they are available for both parents; otherwise, the genotypes of siblings are used. To correct the data structure, 100,000 permutations were used [23]. The cut-off *p*-value was determined using the criterion of a Bonferroni-adjusted threshold for significance of 1.48×10^{-7} (0.05/336,712). The traditional genome-wide significance level was considered below 5.0×10^{-8}.

2.7. Replication analysis, annotation, and allele frequency comparison

The genome-wide significant variants for facial traits, which were identified using family-based association analyses, were tested for replication in an unrelated population via multiple linear regression analysis adjusted for age and sex using PLINK [20]. In cases where the Affymetrix Genome-Wide Human SNP array 5.0 genotype platform did not include the face-associated variants from family-based analyses, we extracted proxies of these variants in the same linkage disequilibrium block using the PLINK program under criterion $r^2 > 0.5$. Statistical significance was set at $p < 0.05$.

We have provided a comprehensive annotation of the identified SNPs. We included information on the genomic functional regions, regulatory elements, and *in silico* scores obtained from predictive algorithms for functional changes owing to variants (eQTL [24], ENCODE [25], CADD [23], and Funseq2 [26]). We also presented differences in the allele frequency between East Asian populations and those from other regions, along with signs of positive selection (Fixation Index and Integrated Haplotype Score).

3. Results

3.1. Heritability of 39 facial phenotypes

Supplementary Table S4 provides the heritability results for four facial areas both in trio familial individuals and in an unrelated

population. For familial-based heritability, we found that in the eye area of the face, heritability estimates (h^2) ranged between 0.260 and 0.473, with the left palpebral fissure length (enL-exL) showing the highest heritability value (0.473). The heritability estimates in the nose area ranged between 0.124 and 0.669, with the distance from the nasion to the right alare (n-alR) showing the highest heritability value (0.669). Similarly, we observed that our heritability estimates for the lip–jaw area ranged between 0.126 and 0.629, with the distance between the labiale superius and pogonion (ls-pg) showing the highest heritability value (0.629). Regarding the facial size, we estimated heritability values that ranged between 0.558 and 0.662, with the distance between the pogonion and midendocanthion (men-pg) showing the highest heritability value (0.662). Notably, we determined that the n-alR and men-pg exhibited the highest heritability values among all four facial areas, suggesting that nose and facial size are the most heritable features.

For population-based heritability, overall heritability values were much lower than those for familial individuals. The values ranged from 0.00211 of the left length of Cupid's bow (ls-cphL) to 2.4586 of philtrum length (sn-ls). Interestingly, the highest heritability value in familial subjects was observed with n-alR, which showed a below-average heritability value (0.072578) in the unrelated population. This may be an indication that heritable traits in the nose area would be more clearly identifiable in a familial study.

3.2. Family-based genome-wide analysis

We performed a family-based genome-wide analysis of 39 facial traits determined from 25 landmarks across four areas of the face in 117 Korean families (including children, youth, and adults). The average age of the father, mother, and offspring (male and female) was 54.1 ± 9.9, 51.3 ± 10.0, 23.7 ± 10.6, and 23.5 ± 10.2 years, respectively (Supplementary Table S1). We performed family-based genetic association analyses using 336,712 variants from a Precision Medicine Research Array using an Affymetrix Axiom array kit. We identified several SNPs at 37 loci as being significantly associated with the investigated facial traits (Supplementary Table S5).

Our analysis identified a total of 18 genome-wide significant loci (Table 1). In particular, we found that in the eye region, intercanthal width (enR-enL) and enL-exL were associated with locus 6q25.3, with the associated SNP (rs6900678; $p = 4.44 \times 10^{-8}$) being located within intron 8 of 20 on AT-rich interaction domain 1B (*ARID1B*). In the nose region, the subnasale to midendocanthion (sn-men), nasion to left alare (n-alL), nasion to right alare (n-alR), nasal width (alR-alL), and nasal tip height (prn-sn) were associated with SNPs in 14 loci. We detected that among these, the top three SNPs, rs10511632, rs78535580, and rs397509305, were located in the intergenic region upstream of the centlein gene (*CNTLN*) and the deletion region of BRCA1 DNA repair associated (*BRCA1*). In the lip–jaw area, the labiale superius to the pogonion (ls-pg), lower facial height (sn-pg), and left length of Cupid's bow (ls-cphL) were associated with loci 22q12.3 (both ls-pg and sn-pg), 20q13.33, and 5p13.1, respectively. We found that the most significant SNP rs228913 ($p = 3.57 \times 10^{-8}$) was located within intron 1 of 16 on transmembrane serine protease 6 (*TMPRSS6*). However, we did not identify any genome-wide significant loci for facial size.

We detected two SNPs with multiple associations with 39 constituent linear distances. More specifically, SNP rs228913 was associated with both sn-pg and ls-pg in the lip–jaw area. However, SNP rs10511632 was associated with the nasion to the right and the left alares (n-alL and n-alR) in the nose area. The most significant GWAS SNPs ($p < 10^{-9}$) detected, including rs228913 (*TMPRSS6*), rs10511632 (*CNTLN*), and rs397509305 (*BRCA1*), are shown using a composite Manhattan plot with $-\log_{10}(p\text{-value})$ against the chromosome position (Fig. 1).

Table 1
Associations ($p < 5 \times 10^{-8}$) in the QFAM analysis among family participants.

Locus	SNP	Position	Ref	Alt	Beta	*p*-value	Facial traits	Gene
Eyes (1 locus)								
6q25.3	rs6900678	157,116,146	G	A	0.0171	4.44E-08	enR-enL	*ARID1B*
6q25.3	rs6900678	157,116,146	G	A	−0.0161	4.54E-08	enL-exL	*ARID1B*
Nose (14 loci)								
9p22.2	rs10511632	16,999,400	A	G	−0.0270	1.14E-09	n-alR	*CNTLN*
9p22.2	rs10511632	16,999,400	A	G	−0.0254	2.19E-09	n-alL	*CNTLN*
17q21.31	rs397509305	43,095,848	TT	–	0.0274	7.17E-09	n-alR	*BRCA1*
15q26.1	rs1057518236	92,949,036	C	T	0.0289	1.60E-08	n-alR	*CHD2*
10q21.1	rs12572437	53,619,108	G	A	−0.0119	1.98E-08	prn-sn	*MBL2*
22q13.31	rs62228060	45,980,504	A	G	0.0258	2.00E-08	n-alR	*WNT7B*
9p13.1	rs62569967	39,088,520	C	A	0.0188	2.11E-08	alR-alL	*CNTNAP3*
1p36.22	rs6680343	12,624,498	C	T	0.0113	2.20E-08	alR-alL	*AADACL4*
16p13.2	rs80338701	8,811,088	C	A	0.0398	2.59E-08	n-alR	*PMM2*
7p21.1	rs1351470759	18,585,524	T	A	0.0316	2.88E-08	n-alR	*HDAC9*
2p22.3	rs117763249	33,314,870	G	A	0.0309	4.30E-08	n-alR	*LTBP1*
14q32.13	rs886037671	95,117,723	C	A	0.0247	4.49E-08	n-alR	*DICER1*
16q24.1	rs79842977	84,650,919	G	A	0.0244	4.55E-08	n-alR	*KLHL36*
2q31.2	rs727505323	178,531,084	CACTTGGT	–	0.0260	4.62E-08	n-alR	*TTN-AS1;TTN*
13q12.12	rs1057516543	23,354,931	C	–	0.0256	4.75E-08	n-alR	*SACS*
Lip–jaw (3 loci)								
22q12.3	rs228913	37,108,117	A	C	0.0510	3.68E-09	sn-pg	*TMPRSS6*
22q12.3	rs228913	37,108,117	A	C	0.0395	3.57E-08	ls-pg	*TMPRSS6*
20q13.33	rs6090046	64,115,023	T	G	0.0327	3.74E-08	sn-pg	*MYT1*
5q13.1	rs36019094	40,237,272	C	A	−0.00634	4.69E-08	ls-cphL	intergenic

3.3. Replication analysis, annotation, and allele frequency comparison

We performed linear regression analyses of 18 loci–SNPs (LD proxy SNPs) in an unrelated population ($n = 5643$) to replicate our previous analysis used to detect the familial association of facial phenotypes. Our results showed that the associations were reproduced in the nose and lip–jaw regions but not in the eye region and facial size in the unrelated population (Table 2), in which none of the 18 SNPs showed substantial changes in associations after additionally adjusting for the body mass index (data not shown). In particular, we observed that the associations for the three most relevant loci (*CNTLN*, *BRCA1*, and *TMPRSS6*) were also maintained in the unrelated population. We also found that the *CNTLN* and *TMPRSS6* variants were associated with the same phenotypes in family participants: n-alR (rs374785 proxy for rs10511632: beta = −0.00255, $p = 0.00337$) and sn-pg (rs16997734 proxy for rs228913: beta = −0.0207, $p = 0.0350$), respectively. In addition, we detected that the *BRCA1* variant associated with n-alR in the familial analysis was associated with a similar phenotype in the nose region, that is, the pronasale to the right alare (prn-alare) (rs35931760 proxy for rs397509305: beta = 0.00272, $p = 0.00956$).

Annotation of the identified SNPs has been given in Supplementary Table S6. The coding variant rs397509305 indicates a potential significant functional impact on the *BRCA1* gene product owing to a frameshift mutation resulting in a premature stop codon. Furthermore, we evaluated whether the SNPs were associated with regulatory elements such as histone modifications and

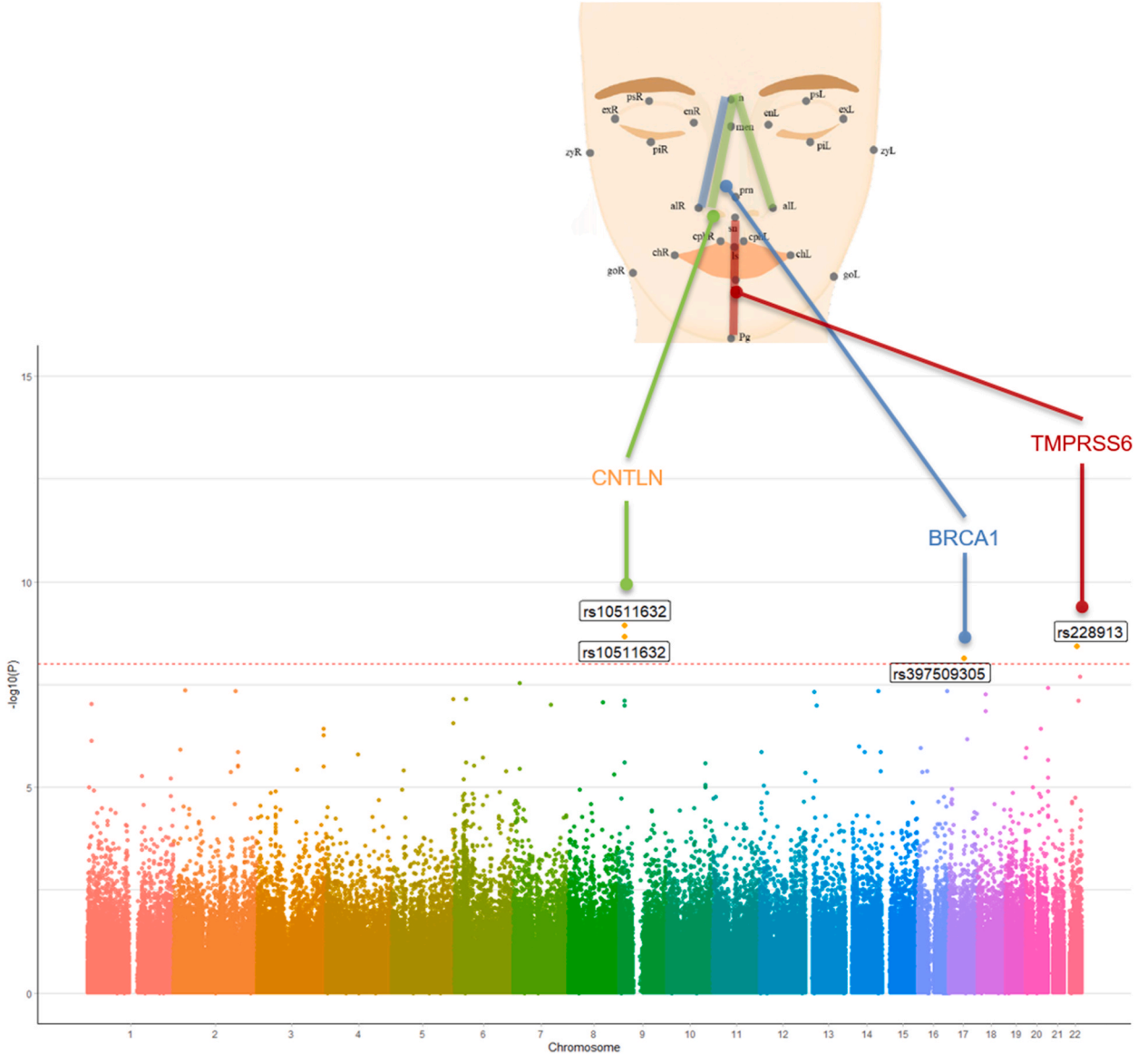

Fig. 1. Composite Manhattan plot showing four peaks of face-associated single nucleotide polymorphisms (SNPs) with $p < 10^{-8}$ (red dotted line). Each peak represents SNP, near gene, and associated facial traits.

Table 2
Replication analysis of 18 loci–SNPs in an unrelated population.

Family participants		Unrelated participants (replication)							
SNP	Traits	Lead SNP	r^2 in LD	Allele	mAF	Traits	Beta	SE	*p*-value
Nose									
rs10511632	n-alR	rs374785	0.869	C > T	0.181425	n-alR	−0.00255	0.00120	0.0337
rs10511632	n-alL	rs374785	0.869	C > T	0.181425	n-alL	−0.00242	0.00107	0.0232
rs397509305	n-alR	rs34981932	0.929	T > G	0.140202	prn-alL	−0.00195	0.000819	0.0171
rs1057518236	n-alR	rs11634204	0.686	G > A	0.125514	prn-sn	0.00468	0.00203	0.0217
rs12572437	prn-sn	rs11003713	0.987	G > A	0.161814	sn-men	−0.00229	0.000887	0.00968
rs6680343	alR-alL	rs3000896	0.989	C > T	0.0836576	prn-alL	0.00524	0.00225	0.0199
rs80338701	n-alR	SNP_A-1891017	0.894	C > T	0.195295	n-alR	0.00241	0.00114	0.0352
rs80338701	n-alR	rs6498096	0.999	T > C	0.173545	n-alL	0.00243	0.00109	0.0258
rs886037671	n-alR	rs11625274	0.782	G > T	0.439854	alR-alL	0.00115	0.000581	0.0476
rs727505323	n-alR	SNP_A-1915444	0.767	A > G	0.427954	n-alR	−0.00189	0.000919	0.0403
rs727505323	n-alR	SNP_A-1966036	0.999	G > A	0.178689	n-alL	0.00231	0.00106	0.0298
Lip and Jaw									
rs228913	sn-pg, ls-pg	rs5756524	0.612	G > A	0.261062	sn-ls	0.00795	0.00225	0.000424
rs36019094	ls-cphL	rs13181878	0.711	C > T	0.295301	ls-li	−0.00232	0.000926	0.0124

Abbreviations: LD, linkage disequilibrium; mAF, minor allele frequency; SE, standard error; n-alR, nasion to right alare; n-alL, nasion to left alare; prn-sn, nasal tip height; sn-pg, lower facial height; ls-pg, labiale superius and pogonion; ls-cphL, left length of Cupid's bow; prn-alL, pronasale to right alare; sn-men, midendocanthion; alR-alL, nasal width; prn-sn, nasal tip height.

transcription factor binding sites. Specifically, rs10511632 demonstrated a significant eQTL association with gene expression levels in non-sun-exposed suprapubic skin tissue, suggesting a regulatory role. ENCODE data also showed that these SNPs are related to histone modifications and transcription factor binding in relevant cell lines.

Our analysis revealed allele frequency differences between East Asian populations and those of other regions for certain SNPs (Supplementary Table S7). This difference was particularly notable between East Asian and African populations for rs10511632 and rs228913. However, generally low Fixation Index values suggest that the overall level of genetic differentiation is not high. While strong evidence of positive selection was not observed, there is a possibility of weak selective pressure for rs397509305.

4. Discussion

The results of the present study suggest that the identified genetic variants in the 117 Korean families (parents and offspring) examined were significantly related to traits with strong heritability for facial phenotypes. In addition, we identified the top three genetic loci (*CNTLN*, *BRCA1*, and *TMPRSS6*) strongly associated with nose and lip–jaw traits, even in an unrelated population. We identified novel SNPs associated with facial phenotypes (especially vertical length) in familial participants.

Heritability highlights the importance of genetic factors in explaining differences among individuals and allows an immediate comparison of the same trait across populations and different traits within a population [27]. In the present study, 39 traits from four facial regions were used to identify vertical length traits with high heritability (>0.60), such as sn-men, n-sn, n-alL, n-alR, ls-pg, sn-pg, n-pg, and men-pg, in Korean families. By contrast, horizontal traits were found to be less heritable. The widely variable facial traits, such as facial height, may be genetically heritable, such that the facial longitudinal growth in the lower face is greater than that in other facial regions from the fetal stage to adulthood [28,29].

We found a strong association between the *CNTLN* variant and the distance from nasion to alares (n-alR, n-alL). *CNTLN* is located on chromosome 9 (9p22.2), and its functions include protein kinase binding and protein binding and bridging, according to Gene Ontology database annotations. *CNTLN* was first identified in the rat brain at an aphidicolin-inducible common fragile site prone to mutation and epigenetic changes; therefore, *CNTLN* is a hotspot for genomic instability [30]. Common fragile sites are a part of normal chromosome structure and are composed of unstable DNA stretches that form gaps and breaks in metaphase chromosomes after partial inhibition of DNA synthesis. Fragile sites are more frequently observed in patients with schizophrenia (SCZ) and have been reported to colocalize with SCZ-linked genes [31]. Moreover, patients with SCZ reportedly have substantial facial dysmorphology, including overall facial widening, increased width of the nose, narrowing of the mouth, and upward displacement of the chin; among these, nose length, lip thickness, and tragion height differ between men and women [31]. *CNTLN* is also reportedly associated with the height of the upper lip vermilion [32]. In conclusion, previous studies suggest that *CNTLN* may be correlated with general facial morphology, which is in agreement with the findings of the present study.

The *BRCA1* variant is strongly associated with the distance between the nasion and right alare (n-alR). *BRCA1*, which is associated with DNA repair, is located on chromosome 17 (17q21) and encodes a 190 kDa nuclear phosphoprotein involved in maintaining genomic stability and regulating DNA damage sensors and signal transducers. It also acts as a tumor suppressor protein. Mutations in *BRCA1* are responsible for approximately 40 % of inherited breast cancers and more than 80 % of inherited breast and ovarian cancers [33]. Various disease-related studies have shown that *BRCA1* affects craniofacial development, such as embryonic orofacial primordia, and may act as a molecular hub during lip and palate morphogenesis; non-syndromic cleft lip and palate, one of the most common human craniofacial deformities, is associated with the dysregulation of the gene regulatory network for DNA damage and cell cycle control via BRCA1 [34]. Furthermore, the expression of *BRCA1* is important for craniofacial bone development in animal models, as

cranial neural crest cells (CNCCs) drive the generation of mesenchymal cells, but not epithelial cells, resulting in craniofacial skeletal defects [35]. Hence, cellular defense against DNA damage may be involved in determining the susceptibility to both facial malformation and cancer [34]. As mentioned in the results, rs397509305 has shown a weak positive selective pressure, which could be interesting given the functional importance of the *BRCA1* gene. These results demonstrate subtle differences in population-specific adaptation and evolution. Further studies encompassing broader genomic regions could reveal more definitive patterns.

Among the various SNPs identified in Korean families, SNP rs228913 ($p = 3.68 \times 10^{-9}$), which is related to the lip–jaw area ($p = 3.57 \times 10^{-8}$), is located in the intron of *TMPRSS6* (22q12.3), which is a type II transmembrane serine proteinase attached to the cell surface. *TMPRSS6* (matriptase-2) is a part of the signaling pathway that controls the levels of hepcidin, a protein that regulates iron balance in the body [36]. However, the relationship between the *TMPRSS6* gene and facial morphology remains unknown.

The identification of genetic variants associated with facial traits (nose and lip-jaw) in Korean families has clinical significance. The *CNTLN*, *BRCA1*, and *TMPRSS6* variants have application potential in early diagnosis and treatment planning for congenital facial anomalies linked with brain dysmorphogenesis such as SCZ (*CNTLN* variant), cleft palate and oncogenetics (*BRCA1* variant), or iron deficiency anemia (*TMPRSS6* variant) [31,34,36]. By deepening our understanding of gene-phenotype relationships, our discoveries might also provide new avenues for clinical applications based on facial features, including the development of personalized medical approaches, improvement of the accuracy of genetic counseling, and creation of non-invasive diagnostic tools.

Though our results can be instrumental in deciphering the relationship between genetic variants and phenotypic variations in humans, the present study has a limitation as the facial-associated variants were not experimentally validated. Although we replicated the familial associations in another unrelated population, it is necessary to validate these associations with facial malformation in vivo and in vitro, for example, in a zebrafish model and via differential gene expression in CNCCs, respectively. We also acknowledge the limitation of using 2D face pictures instead of 3D images. Although we understand that 2D images cannot fully capture all aspects of 3D facial traits, we made a concerted effort to maximize the facial information extracted from our available data. The measures taken to achieve the best-possible results were: (1) we added numerous distance variables for each facial feature, including the eyes, nose, lip–jaw area, and overall facial size, and this approach allowed us to enhance the depth and breadth of facial information analyzed, even within the constraints of 2D imagery; and (2) we incorporated multiple measurements for each facial region with the aim of providing a more nuanced and detailed representation of facial characteristics.

5. Conclusions

In summary, we performed facial measurement and quantitative genetic analyses to estimate narrow-sense heritability attributed to common genetic variations and the pairwise genetic correlations of 39 traits with facial morphology in Korean families. We identified the top three associated loci for the eyes, nose, and lip–jaw facial regions. These highly associated SNPs were related to nine facial traits (two in the eyes, four in the nose, and three in the lip–jaw). Based on our results, we propose that common variants are largely associated with longitudinal variations in the nose and jaw in the general Korean population. Collectively, this familial study on facial morphology enhances our understanding of the relationships between genetic variants and variations in human facial features.

CRediT authorship contribution statement

Donghyun Lee: Writing – original draft, Formal analysis. **Hyo-Jeong Ban:** Writing – review & editing, Formal analysis. **Kyung-Won Hong:** Resources, Investigation, Data curation. **Jong Young Lee:** Writing – original draft, Conceptualization. **Seongwon Cha:** Writing – review & editing, Supervision, Conceptualization.

Ethics declaration

All participants provided written informed consent. For participants aged under 18, the consent was obtained from the participants as well as from their parent/guardian. This study protocol was approved by the Institutional Review Board of the Korea Institute of Oriental Medicine (IRB No.: I-2007/006-002, I-1702/001–002).

Data availability statement

The data will be made available by the corresponding author upon reasonable request.

Funding

This research was supported by the Bio&Medical Technology Development Program of the National Research Foundation funded by the Ministry of Science, ICT & Future Planning (NRF-2018M3E3A1057298) and the research program of the Korea Institute of Oriental Medicine (KSN1732120), Republic of Korea.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing

interests: Hyo-Jeong Ban reports financial support was provided by Korea Ministry of Science and ICT and by Korea Institute of Oriental Medicine. Seongwon Cha reports financial support was provided by Korea Ministry of Science and ICT and by Korea Institute of Oriental Medicine. Jong Young Lee reports a relationship with Oneomics Co., Ltd. that includes: board membership. Donghyun Lee reports a relationship with Oneomics Co., Ltd. that includes: employment. Kyung-Won Hong reports a relationship with Theragen Bio Co., Ltd. that includes: employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The KoGES study was conducted using bioresources from the National Biob3ank of Korea and the Korea Disease Control and Prevention Agency, Republic of Korea (KBN-2020-076). Familial and facial phenotypic data were provided by the Korea Medicine Data Center of the Korea Institute of Oriental Medicine. We would like to thank Editage (www.editage.co.kr) for editing and reviewing this manuscript for English language.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e39173.

References

[1] S. Richmond, L.J. Howe, S. Lewis, E. Stergiakouli, A. Zhurov, Facial genetics: a brief overview, Front. Genet. 9 (2018) 462, https://doi.org/10.3389/fgene.2018.00462.
[2] D.J.J. Farnell, S. Richmond, J. Galloway, A.I. Zhurov, P. Pirttiniemi, T. Heikkinen, V. Harila, H. Matthews, P. Claes, Multilevel principal components analysis of three-dimensional facial growth in adolescents, Comput. Methods Progr. Biomed. 188 (2020) 105272, https://doi.org/10.1016/j.cmpb.2019.105272.
[3] H.J. Kim, S.W. Im, G. Jargal, S. Lee, J.H. Yi, J.Y. Park, J. Sung, S.I. Cho, J.Y. Kim, J.I. Kim, J.S. Seo, Heritabilities of facial measurements and their latent factors in Korean families, Genomics Inform 11 (2013) 83–92, https://doi.org/10.5808/GI.2013.11.2.83.
[4] J.M. Leppanen, C.A. Nelson, The development and neural bases of facial emotion recognition, Adv. Child Dev. Behav. 34 (2006) 207–246, https://doi.org/10.1016/s0065-2407(06)80008-x.
[5] N. Martinez-Abadias, M. Esparza, T. Sjovold, R. Gonzalez-Jose, M. Santos, M. Hernandez, Heritability of human cranial dimensions: comparing the evolvability of different cranial regions, J. Anat. 214 (2009) 19–35, https://doi.org/10.1111/j.1469-7580.2008.01015.x.
[6] M. Seselj, D.L. Duren, R.J. Sherwood, Heritability of the human craniofacial complex, Anat. Rec. 298 (2015) 1535–1547, https://doi.org/10.1002/ar.23186.
[7] K. Adhikari, M. Fuentes-Guajardo, M. Quinto-Sanchez, J. Mendoza-Revilla, J. Camilo Chacon-Duque, V. Acuna-Alonzo, C. Jaramillo, W. Arias, R.B. Lozano, G. M. Perez, J. Gomez-Valdes, H. Villamil-Ramirez, T. Hunemeier, V. Ramallo, C.C. Silva de Cerqueira, M. Hurtado, V. Villegas, V. Granja, C. Gallo, G. Poletti, L. Schuler-Faccini, F.M. Salzano, M.C. Bortolini, S. Canizales-Quinteros, M. Cheeseman, J. Rosique, G. Bedoya, F. Rothhammer, D. Headon, R. Gonzalez-Jose, D. Balding, A. Ruiz-Linares, A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation, Nat. Commun. 7 (2016) 11616, https://doi.org/10.1038/ncomms11616.
[8] S. Cha, J.E. Lim, A.Y. Park, J.H. Do, S.W. Lee, C. Shin, N.H. Cho, J.O. Kang, J.M. Nam, J.S. Kim, K.M. Woo, S.H. Lee, J.Y. Kim, B. Oh, Identification of five novel genetic loci related to facial morphology by genome-wide association studies, BMC Genom. 19 (2018) 481, https://doi.org/10.1186/s12864-018-4865-9.
[9] J.B. Cole, M. Manyama, E. Kimwaga, J. Mathayo, J.R. Larson, D.K. Liberton, K. Lukowiak, T.M. Ferrara, S.L. Riccardi, M. Li, W. Mio, M. Prochazkova, T. Williams, H. Li, K.L. Jones, O.D. Klein, S.A. Santorico, B. Hallgrimsson, R.A. Spritz, Genomewide association study of african children identifies association of SCHI P1 and PDE8A with facial size and shape, PLoS Genet. 12 (2016) e1006174, https://doi.org/10.1371/journal.pgen.1006174.
[10] D.J.M. Crouch, B. Winney, W.P. Koppen, W.J. Christmas, K. Hutnik, T. Day, D. Meena, A. Boumertit, P. Hysi, A. Nessa, T.D. Spector, J. Kittler, W.F. Bodmer, Genetics of the human face: identification of large-effect single gene variants, Proc Natl Acad Sci U S A 115 (2018) E676–E685, https://doi.org/10.1073/pnas.1708207114.
[11] M.K. Lee, J.R. Shaffer, E.J. Leslie, E. Orlova, J.C. Carlson, E. Feingold, M.L. Marazita, S.M. Weinberg, Genome-wide association study of facial morphology reveals novel associations with FREM1 and PARK2, PLoS One 12 (2017) e0176566, https://doi.org/10.1371/journal.pone.0176566.
[12] F. Liu, F. van der Lijn, C. Schurmann, G. Zhu, M.M. Chakravarty, P.G. Hysi, A. Wollstein, O. Lao, M. de Bruijne, M.A. Ikram, A. van der Lugt, F. Rivadeneira, A. G. Uitterlinden, A. Hofman, W.J. Niessen, G. Homuth, G. de Zubicaray, K.L. McMahon, P.M. Thompson, A. Daboul, R. Puls, K. Hegenscheid, L. Bevan, Z. Pausova, S.E. Medland, G.W. Montgomery, M.J. Wright, C. Wicking, S. Boehringer, T.D. Spector, T. Paus, N.G. Martin, R. Biffar, M. Kayser, A genome-wide association study identifies five loci influencing facial morphology in Europeans, PLoS Genet. 8 (2012) e1002932, https://doi.org/10.1371/journal.pgen.1002932.
[13] L. Paternoster, A.I. Zhurov, A.M. Toma, J.P. Kemp, B. St Pourcain, N.J. Timpson, G. McMahon, W. McArdle, S.M. Ring, G.D. Smith, S. Richmond, D.M. Evans, Genome-wide association study of three-dimensional facial morphology identifies a variant in PAX3 associated with nasion position, Am. J. Hum. Genet. 90 (2012) 478–485, https://doi.org/10.1016/j.ajhg.2011.12.021.
[14] J.K. Pickrell, T. Berisa, J.Z. Liu, L. Segurel, J.Y. Tung, D.A. Hinds, Detection and interpretation of shared genetic influences on 42 human traits, Nat. Genet. 48 (2016) 709–717, https://doi.org/10.1038/ng.3570.
[15] J.R. Shaffer, E. Orlova, M.K. Lee, E.J. Leslie, Z.D. Raffensperger, C.L. Heike, M.L. Cunningham, J.T. Hecht, C.H. Kau, N.L. Nidey, L.M. Moreno, G.L. Wehby, J. C. Murray, C.A. Laurie, C.C. Laurie, J. Cole, T. Ferrara, S. Santorico, O. Klein, W. Mio, E. Feingold, B. Hallgrimsson, R.A. Spritz, M.L. Marazita, S.M. Weinberg, Genome-wide association study reveals multiple loci influencing normal human facial morphology, PLoS Genet. 12 (2016) e1006149, https://doi.org/10.1371/journal.pgen.1006149.
[16] J. Guo, J. Tan, Y. Yang, H. Zhou, S. Hu, A. Hashan, N. Bahaxar, S. Xu, T.D. Weaver, L. Jin, M. Stoneking, K. Tang, Variation and signatures of selection on the human face, J. Hum. Evol. 75 (2014) 143–152, https://doi.org/10.1016/j.jhevol.2014.08.001.
[17] Y.S. Cho, M.J. Go, Y.J. Kim, J.Y. Heo, J.H. Oh, H.J. Ban, D. Yoon, M.H. Lee, D.J. Kim, M. Park, S.H. Cha, J.W. Kim, B.G. Han, H. Min, Y. Ahn, M.S. Park, H. R. Han, H.Y. Jang, E.Y. Cho, J.E. Lee, N.H. Cho, C. Shin, T. Park, J.W. Park, J.K. Lee, L. Cardon, G. Clarke, M.I. McCarthy, J.Y. Lee, J.K. Lee, B. Oh, H.L. Kim, A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits, Nat. Genet. 41 (2009) 527–534, https://doi.org/10.1038/ng.357.
[18] J.H. Do, B. Ku, J.S. Jang, H. Kim, J.Y. Kim, Analysis of Sasang constitutional types using facial features with compensation for photographic distance, Integr Med Res 1 (2012) 26–35, https://doi.org/10.1016/j.imr.2012.09.003.
[19] D.E. King, Dlib-ml: a machine learning toolkit, J. Mach. Learn. Res. 10 (2009) 1755–1758.

[20] C.C. Chang, C.C. Chow, L.C. Tellier, S. Vattikuti, S.M. Purcell, J.J. Lee, Second-generation PLINK: rising to the challenge of larger and richer datasets, GigaScience 4 (2015) 7, https://doi.org/10.1186/s13742-015-0047-8.
[21] J. Yang, S.H. Lee, M.E. Goddard, P.M. Visscher, GCTA: a tool for genome-wide complex trait analysis, Am. J. Hum. Genet. 88 (2011) 76–82, https://doi.org/10.1016/j.ajhg.2010.11.011.
[22] J.B. Hall, W.S. Bush, Analysis of heritability using genome-wide data, Curr Protoc Hum Genet 91 (1) (2016) 30 31–31 30 10, https://doi.org/10.1002/cphg.25.
[23] G.R. Abecasis, L.R. Cardon, W.O. Cookson, A general test of association for quantitative traits in nuclear families, Am. J. Hum. Genet. 66 (2000) 279–292, https://doi.org/10.1086/302698.
[24] G.T. Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues, Science 369 (2020) 1318–1330, https://doi.org/10.1126/science.aaz1776.
[25] Y. Luo, B.C. Hitz, I. Gabdank, J.A. Hilton, M.S. Kagda, B. Lam, Z. Myers, P. Sud, J. Jou, K. Lin, U.K. Baymuradov, K. Graham, C. Litton, S.R. Miyasato, J. S. Strattan, O. Jolanki, J.W. Lee, F.Y. Tanaka, P. Adenekan, E. O'Neill, J.M. Cherry, New developments on the Encyclopedia of DNA Elements (ENCODE) data portal, Nucleic Acids Res. 48 (2020) D882–D889, https://doi.org/10.1093/nar/gkz1062.
[26] Y. Fu, Z. Liu, S. Lou, J. Bedford, X.J. Mu, K.Y. Yip, E. Khurana, M. Gerstein, FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer, Genome Biol. 15 (2014) 480, https://doi.org/10.1186/s13059-014-0480-5.
[27] P.M. Visscher, W.G. Hill, N.R. Wray, Heritability in the genomics era–concepts and misconceptions, Nat. Rev. Genet. 9 (2008) 255–266, https://doi.org/10.1038/nrg2322.
[28] P.M. Som, T.P. Naidich, Illustrated review of the embryology and development of the facial region, part 2: late development of the fetal face and changes in the face from the newborn to adulthood, AJNR Am J Neuroradiol 35 (2014) 10–18, https://doi.org/10.3174/ajnr.A3414.
[29] J. Yang, B. Benyamin, B.P. McEvoy, S. Gordon, A.K. Henders, D.R. Nyholt, P.A. Madden, A.C. Heath, N.G. Martin, G.W. Montgomery, M.E. Goddard, P. M. Visscher, Common SNPs explain a large proportion of the heritability for human height, Nat. Genet. 42 (2010) 565–569, https://doi.org/10.1038/ng.608.
[30] K. Maccaroni, E. Balzano, F. Mirimao, S. Giunta, F. Pelliccia, Impaired replication timing promotes tissue-specific expression of common fragile sites, Genes 11 (2020), https://doi.org/10.3390/genes11030326.
[31] R.J. Hennessy, P.A. Baldwin, D.J. Browne, A. Kinsella, J.L. Waddington, Frontonasal dysmorphology in bipolar disorder by 3D laser surface imaging and geometric morphometrics: comparisons with schizophrenia, Schizophr. Res. 122 (2010) 63–71, https://doi.org/10.1016/j.schres.2010.05.001.
[32] B. Bonfante, P. Faux, N. Navarro, J. Mendoza-Revilla, M. Dubied, C. Montillot, E. Wentworth, L. Poloni, C. Varon-Gonzalez, P. Jones, Z. Xiong, M. Fuentes-Guajardo, S. Palmal, J.C. Chacon-Duque, M. Hurtado, V. Villegas, V. Granja, C. Jaramillo, W. Arias, R. Barquera, P. Everardo-Martinez, M. Sanchez-Quinto, J. Gomez-Valdes, H. Villamil-Ramirez, C.C. Silva de Cerqueira, T. Hunemeier, V. Ramallo, F. Liu, S.M. Weinberg, J.R. Shaffer, E. Stergiakouli, L.J. Howe, P. G. Hysi, T.D. Spector, R. Gonzalez-Jose, L. Schuler-Faccini, M.C. Bortolini, V. Acuna-Alonzo, S. Canizales-Quinteros, C. Gallo, G. Poletti, G. Bedoya, F. Rothhammer, C. Thauvin-Robinet, L. Faivre, C. Costedoat, D. Balding, T. Cox, M. Kayser, L. Duplomb, B. Yalcin, J. Cotney, K. Adhikari, A. Ruiz-Linares, A GWAS in Latin Americans identifies novel face shape loci, implicating VPS13B and a Denisovan introgressed region in facial variation, Sci. Adv. 7 (2021), https://doi.org/10.1126/sciadv.abc6160.
[33] R. Prakash, Y. Zhang, W. Feng, M. Jasin, Homologous recombination and human health: the roles of BRCA1, BRCA2, and associated proteins, Cold Spring Harb Perspect Biol 7 (2015) a016600, https://doi.org/10.1101/cshperspect.a016600.
[34] G.S. Kobayashi, L. Alvizi, D.Y. Sunaga, P. Francis-West, A. Kuta, B.V. Almada, S.G. Ferreira, L.C. de Andrade-Lima, D.F. Bueno, C.E. Raposo-Amaral, C.F. Menck, M.R. Passos-Bueno, Susceptibility to DNA damage as a molecular mechanism for non-syndromic cleft lip and palate, PLoS One 8 (2013) e65677, https://doi.org/10.1371/journal.pone.0065677.
[35] K. Kitami, M. Kitami, M. Kaku, B. Wang, Y. Komatsu, BRCA1 and BRCA2 tumor suppressors in neural crest cells are essential for craniofacial bone development, PLoS Genet. 14 (2018) e1007340, https://doi.org/10.1371/journal.pgen.1007340.
[36] L. Silvestri, A. Pagani, A. Nai, I. De Domenico, J. Kaplan, C. Camaschella, The serine protease matriptase-2 (TMPRSS6) inhibits hepcidin activation by cleaving membrane hemojuvelin, Cell Metab 8 (2008) 502–511, https://doi.org/10.1016/j.cmet.2008.09.012.