Classification of early age facial growth pattern and identification of the genetic basis in two Korean populations

Mi‑Yeon Cha1, Yu‑Jin Hong2, Ja‑Eun Choi1, Tae‑Song Kwon3, Ig‑Jae Kim2 & Kyung‑Won Hong1*

Childhood to adolescence is an accelerated growth period, and genetic features can influence differences of individual growth patterns. In this study, we examined the genetic basis of early age facial growth (EAFG) patterns. Facial shape phenotypes were defined using facial landmark distances, identifying five growth patterns: continued-decrease, decrease-to-increase, constant, increase-to-decrease, and continued-increase. We conducted genome-wide association studies (GWAS) for 10 horizontal and 11 vertical phenotypes. The most significant association for horizontal phenotypes was rs610831 (TRIM29; β = 0.92, p‑value = 1.9 × 10−9) and for vertical phenotypes was rs6898746 (ZSWIM6; β = 0.1103, p‑value = 2.5 × 10−8). It is highly correlated with genes already reported for facial growth. This study is the first to classify and characterize facial growth patterns and related genetic polymorphisms.

Differences in the relative size, shape, and spatial arrangement (vertical, horizontal, and depth) of various facial features (e.g., eyes, nose, and lips) make each individual human face unique. Therefore, skull growth and facial morphology are of interest to many scientific disciplines, especially anthropology, genetics, and forensic science. Our face shapes change continuously from infancy to adulthood. During the early ages, from 1 to 20 years, our face shapes grow rapidly, and genetic features may be responsible for individual differences in facial phenotypes. The period from childhood to adolescence is characterized by accelerated growth, and developmental modeling of facial morphology is useful for forensic and biomedical practices. Because the number of missing persons is increasing every year and technology is required to estimate the present face from the past.

To understand early age facial growth, an important point is considering of the growth direction. The past 10 years of facial morphology research has benefited from the development of image recognition technology, which can quickly and accurately capture the details of the face. Similarly, the development of genotyping technology facilitates the exploration of genetic impacts on human facial morphology. Although various studies have been conducted to examine facial growth, most studies have focused on identifying the causes of craniofacial abnormalities. In the study of the healthy individuals of Europeans ancestry, some genes such as MAFB, PAX9 associated with craniofacial development or syndromes. The genetic loci associated with facial phenotypes were reported genes such as PRDM16, PAX3, and TP63. Research to investigate the biological basis underlying the normal range of facial variability has only recently been conducted. Over the past few years, as facial recognition technology has improved, substantial progress has been made in the identification of loci related to facial traits in published genome-wide association studies (GWAS). The starting point of GWAS analyses for facial morphology begins with craniofacial development or the identification of genetic loci associated with genetic facial deformities and syndromes. According to recently reported GWAS results, studies on human phenotypes have identified and reported multiple loci associated with normal facial surface morphology.

While facial variation is influenced such as age and nutritional status, striking facial similarities within families reveal a strong genetic component. However, genetic and GWAS studies are mainly studies of facial morphology in adults. The association between facial phenotype and SNP has been reported in European adolescents, and facial changes (face height, eye width, and nose width) in 15-year-old British children have
been studied, but mostly adults. Therefore, understanding the relationship between facial shape and genetic loci during rapid facial changes period helps innate understanding of the face. It is expected to affect the overall industry necessary for children’s awareness. This study was conducted on Korean subjects, the purpose is to increase the probability of identity inference by predicting the faces of long-term missing children. Therefore, it is important to understand the research in the period when the face shape changes rapidly.

A reported study of facial morphology in Koreans identified five GWAS loci associated with facial traits not for facial growth. In this study, we aimed to analyze early age facial growth (EAFG) patterns using facial landmark distances measured from normal facial surface phenotypes to examine the possible genetic basis of individual differences among two Korean populations.

**Results**

**The overall population characteristics and measurements.** We collected current and past photos from participants. For each participant, one current photo was obtained through studio photography, and 5 to 7 past frontal photos were collected. For the age of the past photos, 1 to 2 photos were collected and used for the study, each of which was less than 5 years old, 5 to 10 years old, 10 to 15 years old, and 15 to 20 years old. Supplementary Table S1 lists the number of photos collected by each participant and their age at the time of taking the photos in detail. The measurement data and characteristics information of each participant according to the facial area are shown in Fig. 1. Supplementary Table S2 shown the characteristics of the participant.

We quantified the facial features of two independent populations who were recruited during separate periods: 172 individuals in Population 1 (POP1) were recruited from January 2019 to July 2020, and 100 individuals in Population 2 (POP2) were recruited from July 2020 to September 2020. We collected a total of 172 current photos and 884 past photos for POP1 and 100 current photos and 600 past photos for POP2. The 21 facial phenotypes of each current and past profile photograph were determined by measuring the distances between 19 facial landmarks. Facial landmarks and measurement areas are depicted in Fig. 1, and the measurement results for each area are summarized in Table 1. All direct measurements were normalized against the distance between the center of left and right irises. The facial phenotypes were categorized into two facial groups: Category 1 was described as the horizontal index (H1–H10), and Category 2 was described as the vertical index (V1–V11).

The measurements from each past photograph were compared against those of the current photograph using non-linear regression methods to determine the facial changes over time. The regression patterns determined by the visual inspection of individual’s changes were used for the genetic association study (Supplementary Fig. S1).

**Defining each measurement and analyzing the time series of facial measurements according to age.** The measurements from each past photograph were compared against those of the current photograph using non-linear regression methods to determine the facial changes over time. The facial growth patterns of each individual are determined by the visual inspection of individual’s changes were used for the genetic association study.

To understand the changes in facial measurements with age, a graph of the time series for each individual facial measurement was plotted for 21 phenotypes according to the age relative to the current age using a non-linear model. Supplementary Fig. S1A–C graphically represent individual growth patterns for each representative eye, nose, and mouth phenotype. Table 1 summarizes the distribution of facial growth patterns by face region in Pop1 and Pop2. We identified five EAFG patterns: Pattern 1 (DD), continued decrease; Pattern 2 (DI), decrease to increase; Pattern 3 (CC), constant; Pattern 4 (ID), increase to decrease; and Pattern 5 (II), continued increase (Fig. 2). The X-axis represents age, and the Y-axis represents facial distances between two selected points.

---

**Figure 1.** Measuring 19 facial landmarks using 21 facial measurements between facial landmark pairs.

| Area | Phenotype | Distance between two facial landmarks | Description |
|------|-----------|--------------------------------------|-------------|
| H1   | 1-2       | Intercommissural width               |
| H2   | 3-4       | Outercommissural width               |
| H3   | 5-6       | Interpalpebral superciliary length   |
| H4   | 7-8       | Interpalpebral infraorbital length   |
| H5   | 1-3       | Left palpebral fissure length        |
| H6   | 2-4       | Right palpebral fissure length       |
| H7   | 18-19     | Mouth width                          |
| H8   | 13-14     | Nasal width                          |
| H9   | 12-13     | Subnasale to left alare              |
| H10  | 12-14     | Subnasale to right alare             |
| V1   | 5-7       | Left palpebral fissure height        |
| V2   | 6-8       | Right palpebral fissure height       |
| V3   | 12-13     | Philtrum height                      |
| V4   | 16-17     | Lobule inferius to pogionation       |
| V5   | 18-15     | Midlacrimal to palpebral supercilium |
| V6   | 10-12     | Midlacrimal to subnasale             |
| V7   | 9-11      | Nasal bridge height                  |
| V8   | 9-12      | Nasal height                         |
| V9   | 10-17     | Midlacrimal to pogionation           |
| V10  | 9-17      | Nasion to pogionation                |
| V11  | 12-17     | Subnasale to pogionation             |
Therefore, gender was not used as a covariate in the GW AS analysis in this study.

no significant difference in the distribution of patterns between males and females in the facial growth pattern.

measurements, we identified significant and suggestive SNPs. The phenotype used in this study was analyzed in Tables 2 and 3. The criteria for dividing DD, DI, CC, ID, and II patterns were established and validated using regional signal plots (Supplementary Fig. S4). The top significant SNPs for each significant cluster are described other significant or suggestive SNPs). The significant or suggestive and clustered SNP loci were illustrated using

in 104 loci showed a clustered pattern (i.e., significant or suggestive SNPs co-segregated with more than three

a single significant SNP without co-segregation with other SNPs nearby the significant SNP), and 729 SNPs

Table 1. EAFG pattern distributions for each phenotype in two populations: populations 1 and 2.

based on the 19 features measured. Among the EAFG patterns, the higher the frequency, the darker the

Genotype analysis. Genome-wide single-nucleotide polymorphism (SNP) genotypes were obtained from an 800 K SNP microarray experiment using an Axiom array followed by imputation using 1000 Genomes Phase 3 data, resulting in a total of 7,375,270 polymorphic SNPs included in the GWAS. We conducted a GWAS for the combined analysis of POP1 and POP2, in addition to separate analyses for POP1 and POP2. The significant or suggestive SNPs from the combined analysis were determined based on the criteria of a p-value < 5 × 10^{-8} for significant SNPs and 5 × 10^{-8} ≤ p-value < 1 × 10^{-5} for suggestive SNPs. In addition, SNPs in the individual POP1 and POP2 analysis with p-value < 0.05 were considered significant. The combined GWAS results are illustrated using quantile–quantile (QQ) plots (Supplementary Fig. S2) and Manhattan plots (Supplementary Fig. S3) for each phenotype. A total of 97 SNPs satisfied the genome-wide significance criteria (p-values < 5 × 10^{-8}), and 759 SNPs were identified with suggestive association p-values (5 × 10^{-8} ≤ p-value < 1 × 10^{-5}) for 21 facial phenotypes (Supplementary Table S3). The SNPs were analyzed by clustering patterns: 77 SNPs were singletons (i.e., a single significant SNP without co-segregation with other SNPs nearby the significant SNP), and 729 SNPs in 104 loci showed a clustered pattern (i.e., significant or suggestive SNPs co-segregated with more than three other significant or suggestive SNPs). The significant or suggestive and clustered SNP loci were illustrated using regional signal plots (Supplementary Fig. S4). The top significant SNPs for each significant cluster are described in Tables 2 and 3. The criteria for dividing DD, DI, CC, ID, and II patterns were established and validated using quantitative trait association analysis (Wald test). Among the 5 EAFG patterns in the horizontal and the vertical measurements, we identified significant and suggestive SNPs. The phenotype used in this study was analyzed by coding facial growth patterns from the past to the present from 1 to 5 (Supplementary Table S4). There was no significant difference in the distribution of patterns between males and females in the facial growth pattern. Therefore, gender was not used as a covariate in the GWAS analysis in this study.

The SNPs of the existing Facial Measurement GWAS results were checked once more in the results of this study, and the results were included in a separate Supplementary Table S5. We could test 20 SNPs that was
| Chromosome | Position | SNP ID | Allele 1 | Allele 2 | Genotype 1 | Genotype 2 | Genotype 3 | Genotype 4 | Genotype 5 |
|------------|----------|--------|----------|----------|------------|------------|------------|------------|------------|
| 1 | 14613193 | rs3732968 | AA | GG | 0.19 0.81 | 0.16 0.84 | 0.15 0.85 | 0.14 0.86 | 0.13 0.87 |
| 2 | 22051232 | rs2920727 | CC | TT | 0.18 0.82 | 0.17 0.83 | 0.16 0.84 | 0.15 0.85 | 0.14 0.86 |
| 3 | 33431192 | rs10916560 | CC | TT | 0.2 0.8 | 0.19 0.81 | 0.18 0.82 | 0.17 0.83 | 0.16 0.84 |
| 4 | 44654083 | rs12950897 | CC | TT | 0.2 0.8 | 0.19 0.81 | 0.18 0.82 | 0.17 0.83 | 0.16 0.84 |
| 5 | 55809023 | rs5761382 | CC | TT | 0.2 0.8 | 0.19 0.81 | 0.18 0.82 | 0.17 0.83 | 0.16 0.84 |
| 6 | 66936009 | rs12081848 | CC | TT | 0.2 0.8 | 0.19 0.81 | 0.18 0.82 | 0.17 0.83 | 0.16 0.84 |
| 7 | 78041582 | rs11357059 | CC | TT | 0.2 0.8 | 0.19 0.81 | 0.18 0.82 | 0.17 0.83 | 0.16 0.84 |
| 8 | 89207883 | rs13239420 | CC | TT | 0.2 0.8 | 0.19 0.81 | 0.18 0.82 | 0.17 0.83 | 0.16 0.84 |
| 9 | 99068510 | rs13031707 | CC | TT | 0.2 0.8 | 0.19 0.81 | 0.18 0.82 | 0.17 0.83 | 0.16 0.84 |
| 10 | 10049721 | rs1196881 | CC | TT | 0.2 0.8 | 0.19 0.81 | 0.18 0.82 | 0.17 0.83 | 0.16 0.84 |
| 11 | 11000038 | rs11019602 | CC | TT | 0.2 0.8 | 0.19 0.81 | 0.18 0.82 | 0.17 0.83 | 0.16 0.84 |
| 12 | 12050803 | rs11909631 | CC | TT | 0.2 0.8 | 0.19 0.81 | 0.18 0.82 | 0.17 0.83 | 0.16 0.84 |
| 13 | 13031620 | rs12990793 | CC | TT | 0.2 0.8 | 0.19 0.81 | 0.18 0.82 | 0.17 0.83 | 0.16 0.84 |
| 14 | 14089781 | rs13239403 | CC | TT | 0.2 0.8 | 0.19 0.81 | 0.18 0.82 | 0.17 0.83 | 0.16 0.84 |
| 15 | 15091868 | rs1196881 | CC | TT | 0.2 0.8 | 0.19 0.81 | 0.18 0.82 | 0.17 0.83 | 0.16 0.84 |
| 16 | 16058593 | rs1196881 | CC | TT | 0.2 0.8 | 0.19 0.81 | 0.18 0.82 | 0.17 0.83 | 0.16 0.84 |
| 17 | 17090981 | rs1196881 | CC | TT | 0.2 0.8 | 0.19 0.81 | 0.18 0.82 | 0.17 0.83 | 0.16 0.84 |
| 18 | 18039872 | rs1196881 | CC | TT | 0.2 0.8 | 0.19 0.81 | 0.18 0.82 | 0.17 0.83 | 0.16 0.84 |
| 19 | 19055943 | rs1196881 | CC | TT | 0.2 0.8 | 0.19 0.81 | 0.18 0.82 | 0.17 0.83 | 0.16 0.84 |
| 20 | 20070096 | rs1196881 | CC | TT | 0.2 0.8 | 0.19 0.81 | 0.18 0.82 | 0.17 0.83 | 0.16 0.84 |

*Continued*
Table 2. Features of suggestive and significant SNPs in horizontal regions associated with the 5 EAFG patterns identified in the combined sample and each individual population sample. Significant values are in bold.

*Physical positions are based on NCBI build 37 of the human genome.

Discussion

Human development is characterized by distinct developmental processes, especially during adolescence, and the speed and direction of craniofacial development differ for each person. Therefore, under the assumption that the speed and direction of development would differ across individuals, we calculated the changes in facial measurements according to age over time for each individual and for each indicator and divided these changes into five major categories. An index was calculated based on the positioning of landmarks in the facial profile picture, according to a widely used method for current facial analyses. GWAS analysis of facial development patterns was performed by recoding each of the five growth patterns as individual values.

Despite the significant difference from the existing approach such as GWAS analysis of face measurements or measurements related to differences in facial shape between individuals, but an analysis of childhood facial growth patterns. A limitation of this study was that there were not enough participants and some factors that could affect facial growth were not measured. The facial measurements were appeared 2D images rather than 3D images. To understand facial growth patterns, we tried to measure for facial traits, such as take a photograph of early age participants in the studio and recruited past pictures. Through like this study, further study will be more understanding for facial morphology and facial growth.

Overall, while some difference in facial growth patterns were found, these patterns were not consistent across all individuals. The results suggest that there is a need for further research to understand the factors that influence facial growth in different populations and to develop more effective interventions for facial development.
| Gene       | Protein Annotation | Chromosome | Start | End | RefSeq Accession | Description |
|------------|--------------------|-------------|-------|-----|------------------|-------------|
| ACVR2A     | activin A receptor type II A | 5 | 44,237,200 | 44,275,564 | ACVR2A         | growth factor, receptor |
Table 3. Features of suggestive and significant SNPs in vertical regions associated with the 5 EAGF patterns identified in the combined sample and in each individual population. Significant values are in bold. *Physical positions are based on NCBI build 37 of the human genome.

diagonal length on the left side of the nose, which tends to increase with age. H9 is the width of the nose, which tends to increase with age because the lower lateral cartilage and the skin surrounding the ends of the septum weaken, losing elasticity.

Among the vertical axis lengths, many indicators showed similar patterns of increase as the horizontal area, and the vertical axis indicators of the face appear to affect each other during growth. The vertical lengths of the eyes, V1 and V2, showed the greatest tendency toward an increasing pattern. The length of the nose, measured by V5–V6, commonly increases until the age of 20 years. The characteristics of the nose are well known and include major changes, such as drooping tips10. The bone base that supports the nose in youth, a pair of
nal nasal bones, and the ascending process of the maxilla are responsible for many of the soft tissue changes that are observed in the nose during aging.

The pattern frequencies measured for EAG showed that although we used independent populations, our results were replicated in each population. As shown in Table 1, approximately 70% of the facial development patterns were replicated in each group. The index with the highest frequency was replicated in each group, indicating a common pattern across populations. Although a few indices showed a different pattern, these unique indices clustered into large categories (increase or decrease).

However, no analysis model exists for facial growth, and the classification of facial growth as a visual expression clustering model is limiting. In this study, we analyzed by applying the -assoc option provided by PLINK software, and the results of this analysis are based on statistical models called likelihood ratio test and Wald test. The reason for applying this analysis is that the phenotype we are targeting is not a general quantitative phenotype, but multinomial variables called facial growth pattern. The currently available method for genome-wide analysis of these variables and multiple SNPS was the statistical model provided by PLINK software. Therefore, the significance between the SNP and the phenotype discovered in this study can be understood as an analysis result of whether the SNP has the explanatory power to explain the phenotype. Some genetic studies based on multinomial variables, and among them, we can check an example of applying the same likelihood ratio test as ours.

In addition, the number of samples cannot be considered representative of all Koreans. However, this study represents the first attempt to classify the pattern of facial growth, and when data from two independent groups collected at the same time are analyzed and compared, the common result (the frequency of the pattern is more than 70% coincident) overcomes these limitations.

Most facial changes occur before age 18, but growth and facial remodeling have been shown to continue throughout life. The facial skeleton is generally believed to expand continuously throughout life, which is reflected in the gradual increase in certain facial anthropometric measurements with age, such as anterior nasal cavity and facial width. Certain measurements increase significantly with aging, but some measurements are reduced. The chin length becomes shorter as the mandible of the face grows backward due to aging, resulting in a shorter overall face length.

Some extrinsic variables like gender, body mass are known to affect facial morphology. The main influence of gender on facial phenotypes was reported as nasal area and upper facial area, and body mass index (BMI) was reported as a face width characteristic. Obesity-related sites such as cheeks and neck were excluded from the measurement. So, it is thought that the effect of the degree of obesity in this study is relatively small.

GWAS results provide a hypothesis-free approach to identifying unique genetic variations that underlie craniofacial shape differences within populations. A total of 97 significant or suggestive SNPs in 19 gene loci were reported in this study. For 19 loci showing significant and suggestive phenotypic associations, substantial literature was identified associating these loci with facial development, as shown in Fig. 3. In the current work, we found 10 suggestive SNPs in the horizontal region: FOXK1, IGF1R, FAM161A, POUSF2, DYNC111, SFSWAP, TRIM29, RAPGEF1, PCDH7, and CXCR4. We also found 9 suggestive SNPs in the vertical region: ZSWIM6, CSN3, ATXN7, COL18A1, CHST9, CTNNA3, ASTN2, TUSC3, and MTCL1. The gene annotations from the UCSC database (https://genome.ucsc.edu) were used to predict the functional effects of the variants. The genes reported to affect embryonic development from UCSC database included CXCR4 in the horizontal region and CSN3 and TUSC3 in the vertical region. The genes reported to affect cranial growth and brain development included ZSWIM6, ATXN7, CSTN2, and MTCL1 in the vertical region. In addition, genes related to molecular mechanisms in the regulation of skeletal muscle and cartilage included FOXK1, RAPGEF1, and IGF1R in the horizontal region, and COL18A1 and CHST9 in the vertical region. The genes associated with retinal circuit components and the growth of sensory organs were FAM161A, POUSF2, and SFSWAP in the horizontal region. Finally, the genes related to frontonasal and dysmorphic facial features were PCDH7, TRIM29, and DYNC111 in the horizontal region and CTNNA3 in the vertical region.

The authors have generated an interesting report using 2 dimensional images to perform a facial shape GWAS study focused on the fixed time points, but on ontogenetic growth trajectories. To my knowledge, this is a novel analysis. It is also performed in an interesting way, using a mixture of photo types. The argument for this paper is framed around using this for the purpose of being able to improve the accuracy of age progression for missing children, but there is very little discussion of that and this is a very intriguing basic science question.

The facial growth patterns that occur in childhood remain poorly understood and estimating facial growth simply through photographic analysis or the use of existing facial indicators can be difficult. In addition, an individual's unique facial morphologies can be difficult to quantify using simple photographic indicator analysis. The characteristics of facial growth should consider differences in each individual's innate genetic makeup. This study is meaningful because we have classified and characterized facial growth patterns that occur in childhood and will contribute to research on face growth, face recognition, and potentially contribute to finding missing children in the future. This study can serve as a basis for understanding facial morphology and can be expanded to various research fields exploring facial growth, including forensic sciences for both adults and children.

Materials and methods

Study participants. The facial images were obtained from the Human ICT, a company specializing in collecting face data. Related individuals among the participants were not included in the analysis. The facial measurement data were obtained from the National Project of the Missing Child conducted by the Korea Institute of Science and Technology (KIST). Two independent populations (POP1 and POP2) between the ages of 18 and 20 were recruited in different periods: 172 individuals in POP1 were recruited from January 2019 to July 2020,
and 100 individuals in POP2 were recruited from July 2020 to September 2020. Relatives were not included, only individuals were included. Two types of facial photos were collected: a current picture and older pictures of the same individual. Each individual's current photo was taken in a studio using a Canon EOS 1300D camera with 2592 × 1728 resolution at 400 lx illumination. The participants were asked to close their mouths, hold their faces in a neutral expression, and prevent hair from covering their foreheads. The older photos comprised various ages that can constitute individual chronologies, and a minimum of 4 to a maximum of 21 photos were collected from each participant. Among older facial photos, high-quality photos, such as passport photos and school graduation photos, were prioritized. We also asked the participants to supply photos in which the face was in a neutral expression, not family photos, and torn photos or those that were creased over the face were excluded. Age information was collected for each of the past photos. All facial photographs were digitized with a scanner. We collected a total of 172 current photos and 884 past photos for POP1 and 100 current photos and 600 past photos for POP2.

Ethics statement. The research was conducted in accordance with the principles described in the Declaration of Helsinki. The Institutional Review Board of Theragen Bio Institute approved this study (internal review board No.: 700062-20181130-GP-006-01), and all participants provided written informed consent.

Craniofacial measurements. To extract features for craniofacial measurements in current and past facial photographs, we first detected the facial region in each photograph using Dlib. Dlib detects the face and automatically identifies facial landmarks after the facial region is detected. Among the numerous facial feature points detected, we selected 19 feature points and utilized two facial landmark detectors to automatically extract those feature points, as shown in Fig. 1. One of the detectors was used to extract the facial landmarks using an hourglass network-based feature adaptation network (FAN) approach, whereas the other detector was an in-house program that combined Dlib and Stasm. The FAN method stacked three hourglass networks, including residual architecture, which is parallel, hierarchical, and multi-scale blocks, to enhance the performance of the feature localization. Stasm is an active shape model (ASM) method with feature descriptors that we fused to the Dlib program. All detectors were programmed in C++ and python in a Qt environment. Facial detection and feature point extraction were performed as automatic processes but could also be manually modified to obtain more accurate feature locations by a well-trained operator (accuracy: 98.8% on average).

Euclidean distances between two selected points based on the 19 selected features were calculated, as shown in Supplementary Table 2. A total of 21 facial metric values were calculated. Before this process, the distances

Figure 3. Drawings indicate the facial phenotypes and genes associated with the 5 EAFG patterns identified using genome-wide significant associations.
between the centers of both eyes were normalized to 1 for all images to avoid issues associated with scale differences between components and differences in the Z-axis between the subjects and the camera. Distance calculation programs were implemented in Visual Studio C++.

**Face measurement quality controls and sample filtering.** We clustered measurements into groups of horizontal and vertical measurements and selected 10 phenotypes to represent the horizontal index and 11 phenotypes to represent the vertical index. The 21 facial phenotypes were measured in each current and past profile photograph, using the 19 facial landmarks in Supplementary Table 2. For all facial phenotypes, we performed data quality controls on volunteers, regardless of the trait being studied. We drew boxplot diagrams for each of the 21 measurements to exclude outlier data from the top 2% and the bottom 2% from all past and current measurements using the R programming language used in R packages\(^{52}\). Obesity-related sites such as cheeks and neck were excluded from the measurement. An ordinal multinomial model was applied to show that the results were not due to bias. In this study, we analyzed by applying the --assoc option provided by PLINK software, and the results of this analysis are based on statistical models called likelihood ratio test and Wald test\(^{21,22}\).

**Early age facial growth pattern (EAFG) analysis.** Facial growth shows large variations among individuals; therefore, we graphed the time series of individual facial measurements based on age relative to the current age (Supplementary Fig. S1) using a non-linear model in QQ plot in R package\(^{53}\). We defined a total of 5 EAFG growth patterns that were clustered as follows (Fig. 2): Pattern 1 (DD), continued decrease; Pattern 2 (DI), decrease to increase; Pattern 3 (CC), constant; Pattern 4 (ID), increase to decrease, and Pattern 5 (II), continue increased. Among 21 phenotypes, we coded the 5 EAFG patterns and summarized each individual with similar aging trends in Table 1.

**Genotype data.** Oral swab samples (KIST) were obtained, and DNA was extracted using ExgeneTM Tissue SV (GeneAll, Seoul, Korea). All DNA samples were amplified and randomly portioned into 25–125-bp fragments, which were purified, resuspended, and hybridized to an Axiom array (TPMRA chip, Thermo Fisher, Seoul, Korea), a customized array based on the Asian Precision Medicine Research Array (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Following hybridization, the bound targets were washed under stringent conditions to remove non-specific background to minimize noise resulting from random ligation events. The SNP set was filtered based on genotype call rates (≥ 0.98) and MAF (≥ 0.10). Hardy–Weinberg equilibrium (HWE) was calculated for individual SNPs using an exact test. All SNPs reported in this manuscript demonstrated HWE p-values > 0.0001. After filtering, 560,795 polymorphic SNPs were analyzed on chromosomes 1–22.

**Imputation of SNPs.** We conducted an imputation analysis to increase the genome coverage. Imputation of genotypes was performed using minimap\(^4\) at the Michigan Imputation Server (MIS) using the 1000G Phase 3 v4 reference panel\(^5\). We uploaded phased GWAS genotypes and received imputed genomes in return. After imputation, 7,375,270 polymorphic SNPs were analyzed on chromosomes 1–22. INFO score is over than 0.8.

**Genome-wide association studies of EAFG patterns.** To identify not only individual indicators but also indicators that commonly affect facial growth, we performed an analysis that combined the phenotypes of POP1 and POP2. We also conducted a genome-wide association scan of the coded 1 to 5 EAFG growth patterns using asymptotic analyses (likelihood ratio test and Wald test) using the combined population of POP1 and POP2. Population-specific and combined population analyses were performed using PLINK version 1.9 (https://www.cog-genomics.org/plink/)\(^55\), SPSS (IBM SPSS Statistics Inc., New York, U.S.)\(^96\), and R Statistical Software\(^52\). We calculated the beta coefficient and the standard error (SE) values for the association study. To compare the GWAS results for each population, we conducted a replication study using 172 samples from POP1 and 100 samples from POP2. We selected the genetic markers associated with the 5 EAFG patterns in each GWAS, determined by association p-values < 1 × 10^{-5} in the combined dataset and p-values < 0.05 for the individual population datasets and the replication study. In this study, GWAS analysis was performed based on about 800,000 SNPs. The Bonferroni correction p-value threshold is applied. The results are shown in Tables 2 and 3 and Manhattan plots are depicted in Supplementary Fig. S3. The QQ plot, generated using R Statistical Software\(^52\), of the observed p-values showed minimal inflation of the GWAS results from the combined population sample (Supplementary Fig. S2).

**Annotation of SNP-associated genes.** To identify and annotate genes that are functionally related to suggestive and significant SNPs identified in the GWAS, SNP locus data were obtained from the UCSC Genome Browser (Genome Bioinformatics Group, University of Santa Cruz, Santa Cruz, CA, USA). The gene annotations from the UCSC database and Genotype–Tissue Expression (GTEx) database (GTEx Analysis Release v.8, http://www.gtexportal.org/) were used to predict the functional effects of the variants. We constructed regional plots of association for regions of interest using the program LocusZoom (Supplementary Fig. S4)\(^7\).

**Data availability**

Raw genotype or phenotypic data cannot be used due to limitations imposed by ethics. The Summary statistics obtained here are based on the GWAS analysis and can be accessed with the supplementary materials.

Received: 22 December 2021; Accepted: 5 August 2022
Published online: 15 August 2022
References

1. Marcuccio, R., Hallgrensson, R. & Young, N. M. Facial morphogenesis: Physical and molecular interactions between the brain and the face. Curr. Top. Dev. Biol. https://doi.org/10.1016/bs.ctdb.2015.09.001 (2015).

2. Richmond, S., Howe, L. J., Lewis, S., Stergiakouli, E. & Zhurow, A. Facial genetics: A brief overview. Front. Genet. https://doi.org/10.3389/fgen.2018.00462 (2018).

3. Som, P. M. & Naidich, T. P. 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. Am. J. Neuroradiol. 35, 10–18. https://doi.org/10.3174/ajnr.A4314 (2014).

4. Kader, F. & Ghai, M. DNA methylation and application in forensic sciences. Forensic Sci. Int. https://doi.org/10.1016/j.forsciint.2015.01.037 (2015).

5. De Jong, M. A. et al. Automated human skull landmarking with 2D Gabor wavelets. Phys. Med. Biol. https://doi.org/10.1088/1361-6560/aa8f0a (2018).

6. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. Nat. Rev. Genet. https://doi.org/10.1038/nrg.2016.59 (2016).

7. Hammond, P. et al. Discriminating power of localized three-dimensional facial morphology. Am. J. Hum. Genet. https://doi.org/10.1016/j.ajhg.2011.12.021 (2012).

8. Shaffer, J. R. et al. Genome-wide association study reveals multiple loci influencing normal human facial morphology. PLoS Genet. https://doi.org/10.1371/journal.pgen.1006149 (2016).

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

10. Paternoster, L. et al. Genome-wide association study of three-dimensional facial morphology identifies a variant in PAX3 associated with nasion position. AJHG. 90, 478–485. https://doi.org/10.1016/j.ajhg.2011.12.021 (2012).

11. Boerhringer, S. et al. Genetic determination of human facial morphology: Links between cleft-lips and normal variation. Eur. J. Hum. Genet. 19, 1192–1197. https://doi.org/10.1038/ejhg.2011.110 (2011).

12. Amini, F. & Borzabadi-Farahani, A. Heritability of dental and skeletal cephalometric variables in monozygous and dizygozous iranian twins. Orthod. Waves. 68(2), 72–79 (2009).

13. Carson, E. A. Maximum likelihood estimation of human craniofacial heritabilities. Am. J. Phys. Anthropol. 131(2), 169–180 (2006).

14. Johannsdottir, B., Thorarinsson, F., Thorardason, A. & Magnusson, T. E. Heritability of craniofacial characteristics between parents and offspring estimated from lateral cephalograms. Am. J. Orthod. Dentofac. Orthop. 127(2), 200–207 (2005).

15. 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(3), 478–485. https://doi.org/10.1016/j.ajhg.2011.12.021 (2012).

16. Liu, F. et al. A genome-wide association study identifies five loci influencing facial morphology in Europeans. PLoS Genet. 8(9), e1002932. https://doi.org/10.1371/journal.pgen.1002932 (2012).

17. Toma, A. M. et al. The assessment of facial variation in 4747 British school children. Eur. J. Orthod. 34(6), 655–664. https://doi.org/10.1093/ajo/ref122 (2012).

18. Das, S. et al. Next-generation genotype imputation service and methods. Nat. Genet. https://doi.org/10.1038/ng.3656 (2016).

19. Edelstein, D. R. Aging of the normal nose in adults. Laryngoscope. 106, 1–23. https://doi.org/10.1097/00005537-199609001-00001 (1996).

20. Mendelson, B. & Wong, S. C. H. Changes in the facial skeleton with aging: Implications and clinical applications in facial rejuvenation. Aesthetic Plast. Surg. 36, 753–760. https://doi.org/10.1007/s00266-012-9904-3 (2012).

21. Qian, M. & Shao, Y. A likelihood ratio test for genome-wide association under genetic heterogeneity. Ann. Hum. Genet. 77(2), 174–182 (2013).

22. German, C. A., Sinsheimer, J. S., Jakobsen, J. R., Hession, T. J. & Treder, J. E. Soft tissue profile changes from 5 to 45 years of age. J. Biomed. Biotechnol. https://doi.org/10.1155/2012/834578 (2012).

23. Bishara, S. E., Bonfante, B. et al. Identification of five novel genetic loci related to facial morphology by genome-wide association studies. BMC Genom. https://doi.org/10.1186/s12864-018-4865-9 (2018).

24. Lee, B. J., Do, J. H. & Kim, J. Y. A classification method of normal and overweight females based on facial features for automated medical applications. J. Biomed. Biotechnol. https://doi.org/10.1155/2012/934578 (2012).

25. Cha, S. et al. Identification of five novel genetic loci related to facial morphology by genome-wide association studies. BMC Genom. https://doi.org/10.1186/s12864-018-4865-9 (2018).

26. Yazdani, S. et al. Identification of five novel genetic loci related to facial morphology by genome-wide association studies. BMC Genom. https://doi.org/10.1186/s12864-018-4865-9 (2018).

27. Muller, F. et al. Identification of five novel genetic loci related to facial morphology by genome-wide association studies. BMC Genom. https://doi.org/10.1186/s12864-018-4865-9 (2018).

28. Farlie, P. G., Baker, N. L., Yip, P. & Tan, T. Y. Frontonasal dysplasia: Towards an understanding of molecular and developmental aetiology. Mol. Syndromol. https://doi.org/10.1159/000450533 (2016).

29. Yan, J. et al. COP9 signalosome subunit 3 is essential for maintenance of cell proliferation in the mouse embryonic epiblast. Mol. Cell. Biol. https://doi.org/10.1128/mcb.23.19.6798-6808.2003 (2003).

30. Lu, H. C. et al. Disruption of the ATXN1-CIC complex causes a spectrum of neurobehavioral phenotypes in mice and humans. Nat. Genet. https://doi.org/10.1038/ng.3808 (2017).
formed genetic analyses. I.-J.K. designed the project. K.-W.H. provided guidance on aspects of study design and institutional affiliations.

Publisher’s note
Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to K.-W.H.

10.1038/s41598-022-18127-6.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-18127-6.

Acknowledgements
We would like to thank the volunteers for their enthusiastic support for this study. We are very grateful to the institutions that allowed the use of their facilities for the assessment of volunteers, including Theragen Bio Co., and Center for Imaging Media Research, Korea Institute of Science and technology. This study was supported by the National research foundation of Korea (NRF-2018M3E3A1057354).

Author contributions
T.-S.K. contributed to volunteer recruitment or data collection. Y.-J.H. performed image analyses. J.-E.C. performed genetic analyses. I.-J.K. designed the project. K.-W.H. provided guidance on aspects of study design and corresponding author. M.-Y.C. wrote the paper with input from coauthors.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-18127-6.

Correspondence and requests for materials should be addressed to K.-W.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022