scientific reports

Check for updates

OPEN

Exploring the association between SNPs and facial morphology in a Spanish population

Belén Navarro-López[1,2]✉, Franziska Wilke[3], Victoria Suárez-Ulloa[4]✉, Miriam Baeta[1,2], Rubén Martos-Fernández[5], Olatz Moreno-López[1,6], Iñigo Olalde[1,7], Begoña Martínez-Jarreta[8,9], Susana Jiménez[10], Susan Walsh[3] & Marian M. de Pancorbo[1,8]

Understanding and predicting human external phenotypes, particularly facial shape, is of great value for individual identification. However, facial morphology is a highly complex trait. Despite its complexity, recent genome wide association studies (GWAS) have shed light on potential SNPs associated with facial features, offering a first glimpse into the likely genetic background of individual appearance. In this paper we have selected a set of 116 candidate SNPs and studied their association with facial phenotypes in a Spanish population of 412 individuals, highlighting a wide spectrum of facial morphologies worthy of investigation. We performed canonical correlation analysis (CCA) between each SNP and the observed spacial variation in facial shape, from its representation by a dense mesh of 7160 quasi-landmarks, revealing significant associations within different facial segments. In particular, ten SNPs are highlighted for their strong association within this Spanish population, some of them uncovering correlations with novel facial regions. These findings underline the importance and usefulness of conducting candidate SNP studies, not only to validate existing associations but also to unveil novel correlations within subpopulations.

Keywords Forensic DNA phenotyping, Candidate SNPs, Facial morphology, Correlation study, Spanish population

DNA analysis is an essential tool in forensic biology, enabling individual identification through the comparison of genetic profiles based on short tandem repeat markers (STRs) and single nucleotide polymorphisms (SNPs). However, cases may remain unsolved due to the absence of a match to a genetic profile database or an individual that is part of the investigation. Recently, advances in the field of forensic genetics have led to the development of a subfield called Forensic DNA Phenotyping (FDP), aiming to determine externally visible characteristics (EVCs) from DNA that may aid in identification as a "biological witness". By obtaining physical appearance data, FDP assists in guiding the search for suspects in criminal cases or the identification of missing persons[1].

EVCs are complex traits, often influenced by numerous genetic and environmental factors[2]. Consequently, identifying the genes involved and developing DNA-based prediction models is challenging. However, over the past decade, different assays have been designed and validated to analyze appearance-related SNPs simultaneously, such as the VISAGE Enhanced Tool for Appearance and Ancestry[3].

Facial shape is one of the most recognizable features in humans[2,4], as there are numerous subtle differences in size and form, influenced by a significant genetic component, that together contribute to each individual's facial distinctiveness[5,6]. This suggests that many of the facial traits can be predicted from DNA, with increased accuracy as more genetic determinants are uncovered[2]. The initial approach to understand the genetic basis of human facial variation focused on candidate gene studies, targeting genetic loci involved in craniofacial development or syndromes[2,7]. More recently, genome-wide association studies (GWAS) have proven to be a

[1]BIOMICs Research Group, Department of Zoology and Animal Cellular Biology, Lascaray Research Center, Faculty of Pharmacy, University of the Basque Country UPV/EHU, 01006 Vitoria-Gasteiz, Spain. [2]Bioaraba Health Research Institute, 01009 Vitoria-Gasteiz, Spain. [3]Indiana University Indianapolis (IUI), Indianapolis, IN 46202, USA. [4]Panacea Cooperative Research S. Coop, 24001 Castilla y León, Spain. [5]Department of Legal Medicine, Toxicology, and Physical Anthropology, University of Granada, 18071 Granada, Spain. [6]Department of Physical Anthropology, Society of Sciences Aranzadi, 20014 Donostia, Spain. [7]Ikerbasque-Basque Foundation of Science, 48009 Bilbao, Spain. [8]Faculty of Medicine, University of Zaragoza, 50009 Zaragoza, Spain. [9]Aragon Health Research Institute (IIS-Aragón), 50009 Zaragoza, Spain. [10]Department of Pathology and Surgery, University of Miguel Hernández, 03550 Alicante, Spain. ✉email: belennavarrolopez1997@gmail.com; victoria.ulloa@panacea-coop.com

valuable tool for exploring the correlation between normal facial variation and phenotype informative SNPs[1]. To date, these studies have identified numerous genetic loci implicated in facial morphology[4,5,8–33]. Interestingly, these findings suggest that many of these loci may play a role in the regulation of cranial neural crest cells (CNCC), which constitute most of the cranial mesenchyme and subsequently differentiate into bone, cartilage, and connective tissue of the face and head[12,34]. However, it should be noted that SNPs identified by GWAS are not always replicated in subsequent studies, whether involving individuals from diverse populations or even among individuals with shared European ancestry[9,14]. Additionally, evidence indicates that facial appearance varies across different regions of Europe[35], implying that findings from studies on European populations may not be generalizable to all subgroups within a particular ancestry. Therefore, it is of interest to analyze these candidate markers using specific subpopulations to help establish the significance of their association within various groups.

It is also worth noting that advancements in 3D imaging systems have facilitated comprehensive evaluations of global and local variation, enabling detailed analyses of associations between genetic markers and even minor facial traits[12,36]. 3D scans allow the capture of morphological information that was not readily possible using 2D approaches. These advancements have opened up new possibilities for conducting easily comparable candidate SNPs studies across diverse populations. Given the observed variation in the significance of findings among previous investigations, candidate association studies play an important role in pinpointing which SNPs can be robust, exhibiting the strongest correlations and contributing towards global predictions, thus serving as reliable tools for predicting facial appearance. It is important to consider that exploring correlations between genetic markers and facial morphology needs the investigation of panmictic populations, characterized by high morphological diversity.

Understanding facial genetics offers significant potential not only in forensic science but also across several other disciplines. The insights gained from this field could have clinical applications in medicine, orthodontics, and aesthetics, as well as contribute to advancements in paleoanthropology[37–42].

Given all the aforementioned considerations, we focused our research on the Spanish population, a previously unexplored yet promising option because of its well-defined biological environment and the extensive genetic contribution of various subpopulations[43]. Its diverse genetic background, coupled with significant internal migration[44], creates a heterogeneous population in terms of facial traits, providing a wide range of diversity for most facial features within a single population. Therefore, our research focuses on studying previously identified genetic markers within the Spanish population to assess their correlation with facial morphology registered using 3D imaging techniques.

Materials and methods
Sample collection
The samples used for this analysis included 412 individuals collected from a Spanish population. Ethical approval for this study (M10_2021_143) was provided by the Ethics Committee for Research on Human Subjects of the University of the Basque Country, CEISH-UPV/EHU, BOPV 32, 17/2/2014. All participants gave their written informed consent prior to participation and completed a form to collect data on variables which may influence facial shape such as sex, age, weight, height, history of facial surgery, and presence of pathologies or traumas (Table 1). Each participant gave a saliva sample taken by sterile swab in triplicate. In addition, a facial 3D scan was performed using a white light led 3D scanner Academia 3D/20 (Creaform, Canada) following the published protocols by the 3D Facial Norms Project[45] and Heike et al.[46]. Participants were asked to remove accessories, tie back their hair, and sit with a neutral facial expression with eyes closed. Each scan was scrutinized in real-time for quality.

Exclusion criteria were applied to remove those participants with (a) a medical history encompassing facial surgery, pathologies, or traumas that may have induced alterations in facial morphology, (b) missing information on sex, age, weight, height, or (c) insufficient image quality.

Genotyping
DNA extraction and quantification
DNA from saliva swab samples was isolated via a salting out method[48] using the DNA Purification System PuregenTM (Gentra System, Inc., Minneapolis, USA). Quantity as well as quality of the DNA obtained after

Variable	Category	N	%
Sex	Male	208	50.49
	Female	204	49.51
Age[47]	Young (18–35)	315	76.46
	Middle Age (36–60)	87	21.11
	Older adulthood	10	2.43
BMI (as classified by the Centers for Disease Control and Prevention)	Underweight	18	4.37
	Healthy weight	264	64.08
	Overweight	90	21.84
	Obese	40	9.71

Table 1. Categories into which all data included in this study have been divided (*N* = 412).

extraction was evaluated by spectrophotometry with a NanoDrop™ One (Thermofisher Scientific, Waltham, USA) and fluorimetry with a Qubit® 2.0 Fluorometer and the Qubit® dsDNA HS Assay Kit, 0,1-120 ng (Thermofisher Scientific). Once quantified, DNA was diluted in Milli-Q water and stored at -20 °C.

SNP selection and analysis
A set of 116 SNPs related to facial morphology previously published in literature were studied (Supplementary Table S1). The main criterion for selection was a minor allele frequency (MAF) for the Iberian population ≥ 0.05, so that these markers can provide enough genetic variation to perform association analyses.

All 116 SNPs underwent evaluation for Fluidigm technology analysis (Fluidigm Corp., California, USA), yet only 94 were suitable for this methodology. Thus, these 94 SNPs (Supplementary Table S1) were genotyped employing one Juno 96×96 and two 48.48 Fluidigm SNPtypes genotyping systems. For the remaining 22 SNPs (Supplementary Table S1), there were observed incompatibilities with this technique, therefore three panels were designed for the simultaneous analysis of variants at the selected loci using SNaPshot minisequencing technology (Applied Biosystems, Foster City, USA) (Supplementary Fig. S1).

SNaPshot primer design
Primers design as well as melting temperature, GC% content and the possible formation of hairpin and dimers on each pair of primers checking was carried out using PerlPrimer v1.1.21 software[49]. Web-based AutoDimer software[50] was used to study compatibility of primers in multiplex reactions. Finally, specificity was evaluated with primer-BLAST[51]. Primers used in amplification range in size from 106 to 279 bp. Attempts were made in all cases to select those primers that produced the shortest possible amplicons, so that they could be used in samples with high limitations in their DNA content and/or moderate DNA degradation levels. Minisequencing primers were designed manually to hybridize the adjacent region of the SNP. Nevertheless, melting temperatures as well as potential unfavourable reactions between them were also studied with the web-based AutoDimer software[50]. These primers were augmented with a 10- to 65-bp non-binding tail at the 5′ end in order to assure various sizes of the minisequencing products, with a minimum difference of 5 bp.

SNaPshot multiplex PCR amplification
PCR multiplex amplification was carried out as follows: 5 µL of Multiplex PCR Master Mix (QIAGEN, Hilden, Germany), 1 µL of primermix (final concentration of each primer shown in Supplementary Table S2), 2 ng of genomic DNA and Milli-Q water in order to reach a final reaction volume of 10 µL. The amplification was performed in a C1000™ Thermal Cycler (BioRad, Hercules, USA) under the following conditions: initial denaturalization at 95 ºC for 15 min; 35 cycles at 95 °C for 30 s, 60 °C for 50 s, and 65 °C for 40 s; and a final extension of 6 min at 65 °C. Amplified DNA was treated with ExoSAP (Takara Bio Inc., Japan) to eliminate the remaining primers and nucleotides: 1 µL of ExoSAP per 2.5 µL of PCR product, incubated at 37 °C for 5 min followed by enzymatic inactivation at 80 °C for 1 min.

SNaPshot minisequencing and capillary electrophoresis
Minisequencing was carried out in a final volume of 7 µL containing 2 µL of SNaPshot™ Multiplex Kit reaction mix (Applied Biosystems, Foster City, USA), 1 µL of primermix (final concentration of each primer shown in Supplementary Table S2), 3 µL of Milli-Q water and 1 µL of purified multiplex PCR product. Thermocycling conditions in a C1000™ Thermal Cycler (BioRad) were 25 cycles at 96 °C for 10 s, 55 °C for 5 s, and 60 °C for 30 s. Minisequencing products were further purified by enzymatic digestion adding 1 µL of SAP (Takara Bio Inc.) to 2 µL of product and incubated at 37 °C for 60 min to remove any remaining nucleotides, followed by enzyme denaturation at 80 °C for 15 min. Finally, minisequencing products were analysed by mixing 1 µL of purified product, 12 µL of Hi-DI formamide (Applied Biosystems), and 0.38 µL of Gene-Scan 120LIZ (Applied Biosystems). After denaturation, capillary electrophoresis was conducted in an ABI PRISM® 3130 Genetic Analyzer (Applied Biosystems). Data were analysed using GeneMapper® Software v4.0 (Applied Biosystems).

Population genetic parameters
Allele and genotype frequencies were calculated using Arlequin v3.5.2.2 software[52]. Differences in population parameters were assessed by means of a Chi-square test conducted with IBM SPSS Statistics v25 software[53], using previously published European and Iberian data as reference populations[54], as these are the most closely related populations to this Spanish cohort. Hardy-Weinberg Equilibrium (HWE) was also analyzed using Arlequin v3.5.2.2 software. Linkage analysis was performed to determine combined segregation. Data derived from linkage disequilibrium (LD) analysis were used for haplotype inference, employing the expectation maximization algorithm to establish combined segregation patterns. All computations were performed using Arlequin v3.5.2.2 software. *P*-value threshold for considering HWE and LD was set post the Bonferroni correction ($\alpha = 0.05/116 = 4.310\text{E}{-}04$).

Processing 3D facial imaging and landmarking
Post-acquisition, the 3D images were processed using the ACADEMIA Software Bundle (VXmodel, VXinspect and VXElements) for both quality and background noise elimination. Each cleaned 3D image was exported as a Wavefront Object file and imported into the Skeleton-ID software[55] for manual placement of five landmarks (endocanthion L/R, chelion L/R, and pronasale).

Following this step, alignment of each 3D surface image was achieved across 7160 spatially dense quasi-landmarks. This process involved the automatic nonrigid mapping of a specially constructed template face or anthropometric mask[36] onto each image using the MeshMonk image-processing pipeline[56].

Subsequently, all facial scans were symmetrized, wherein each image was averaged with its reflection obtained by inverting the X coordinate and aligned in space using a Least Squares Procrustes Superimposition. From this point onward, all statistical analyses were based on functions of the statistical toolbox in Matlab 2023b. All code used for analyses was adapted from previous publications by White et al. and Wilke et al.[14,35].

Correction and segmentation of facial shape

The symmetric 3D images were corrected for all covariates (sex, age, age squared, height, and weight) using Partial Least Squares Regression (PLSR). Following facial adjustments, each face was segmented into 63 facial modules, distributed across four main quadrants (QI: upper lip + philtrum, QII: nose, QIII: lower face, and QIV: upper face), according to the labels of Sero et al.[57]. It facilitated the examination of both global and local influences on facial variation. Subsequently, each segmented module underwent independent Generalized Procrustes Analysis, followed by Principal Component Analysis (PCA) and Horn's parallel analysis to capture the predominant variance the facial segments exhibited while reducing the dimensionality of the dataset, as previously described in White et al.[14].

Association analysis

For each of the 63 facial segments, the association between each SNP and the phenotypic variation was assessed through canonical correlation analysis (CCA). On the one hand, the independent variable (each SNP) was coded according to the additive genetic model based on the presence of the major allele (AA = 0, Aa = 1, aa = 2). On the other hand, the dependent variable (facial shape) was represented as PCs, that represent the direction of phenotypic effects in shape space.

CCA extracts the linear combination of PCs from a facial segment that has the highest correlation with a given SNP and assigns each individual a CCA score[14]. The correlation value between these scores and the associated SNPs was subjected to significance testing using Rao's F-test approximation (right-tail, one-sided). Finally, considering the multiple comparisons conducted, a False Discovery Rate (FDR) adjusted significance threshold was calculated at a level of 0.05. A SNP was considered significant if at least one of the 63 facial segments reached the FDR threshold. In addition, to illustrate the associations, normal displacement maps (magnitude of inward/outward movement) were plotted for each significant SNP and segment.

Results

In this study, we investigated the association between genetics and facial phenotyping variation within a Spanish population. A total of 116 candidate SNPs previously linked to facial morphology were selected based on published literature (Supplementary Table S1). The selection process involved an extensive review of existing studies, encompassing SNPs identified in both single and multiple studies, regardless of the populations in which the associations were initially reported. To ensure adequate genetic variability for robust association analyses, as mentioned in Material and Methods, SNPs with a MAF ≥ 0.05 in the Iberian population were included. After this initial selection, and considering our limitations, priority was given to SNPs replicated in multiple studies and those identified in populations of European ancestry. This strategy was employed to enhance the relevance and representativeness of the findings within the Spanish cohort. CCA was subsequently performed to explore the associations between the selected SNPs and different facial regions. Detailed results for each facial segment are presented below.

Population genetic parameters

Allele and genotype frequencies are presented in Supplementary Table S3. Regarding their allele frequency distributions, there were no significant deviations from previously published Iberian data, except for two SNPs (rs11713363 and rs6695444). In these cases, the distribution of major and minor alleles seemed to be inverted when compared to the European and Iberian data[54]. No significant departure from HWE was observed for the majority of SNPs after applying the Bonferroni correction (α = 0.05/116 = 4.310E−04) apart from rs145365164 (*P*-value = 2.700E−04), rs1863716 (*P*-value = 1.000E−04), rs6129266 (*P*-value = 0), and rs6695444 (*P*-value = 0) (Supplementary Table S3). LD was calculated between all the SNPs. Statistically significant values were observed between SNPs located within the same candidate gene, suggesting co-segregation (Supplementary Table S4). This relationship is attributed to the close proximity of loci in terms of base pairs.

Association analysis

Association analyses conducted in this study proved to be significant for several SNPs (Supplementary Table S5). Findings suggested that, out of the 116 SNPs examined, 101 may potentially influence facial morphology in our population (*P*-value CCA < 0.05). All of these SNPs have been previously linked to various facial traits by other authors[4,5,8–33], underscoring the role of these markers in shaping facial morphology. However, within this subset of 101 SNPs, only ten remained statistically significant after FDR adjustment (*P*-value FDR < 0.05) (Fig. 1). Detailed results categorizing the face into four main quadrants as illustrated in Fig. 2 (QI: upper lip + philtrum, QII: nose, QIII: lower face, and QIV: upper face), are presented below.

Upper lip + philtrum (Quadrant I)
Potential significant correlations (*P*-value CCA < 0.05) were identified between 51 SNPs and the upper lip + philtrum area. However, among them, only two SNPs (rs702489 and rs12786942) reached the FDR threshold (*P*-value FDR < 0.05) and were considered significant (Supplementary Table S6). Both markers revealed a previously unreported association with this region, having been linked to the nose and upper face in prior studies[8,24].
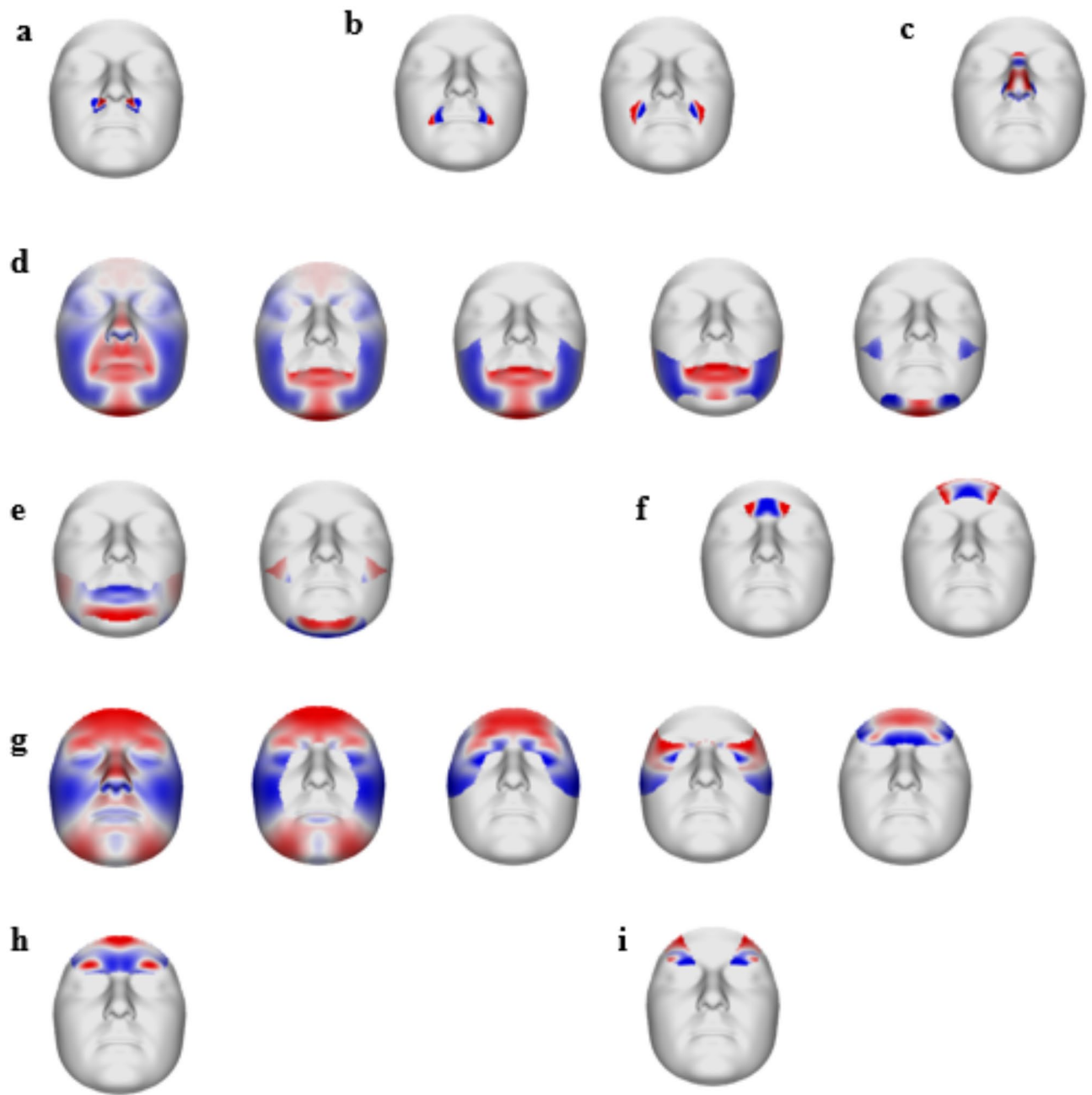
Fig. 1. Facial representations highlighting the regions with statistically significant associations with SNP rs702489 (**a**), rs12786942 (**b**), rs227833 (**c**), rs11768577 (**d**), rs7167736 (**e**), rs3936018 and rs72691108 (**f**), rs6535972 (**g**), rs17275866 (**h**), and rs7738892 (**i**). Images obtained using Matlab™ version 2023b.

Significant correlations were observed between SNP rs702489 and the first quadrant, including segments 30 and 61 (Supplementary Fig. S2a). The effect of rs702489: G > A (each copy of G increases the effect following an additive model) is shown in an outward projection of the region below the nasal alas and the laterals of the philtrum, as well as an inward orientation of the area surrounding the aforementioned region (Fig. 1a, Supplementary Fig. S2b).

Furthermore, statistically significant associations were found between the upper lip and philtrum region, specifically segments 31, 57, and 62 and SNP rs12786942 (Supplementary Fig. S3a). The impact of rs12786942: A > T (each copy of A increases the effect following an additive model) is related to an inward projection of the inner laterals of the upper lip and philtrum. In contrast, the outer lateral regions of these areas display an outward orientation (Fig. 1b, Supplementary Fig. S3b–c).
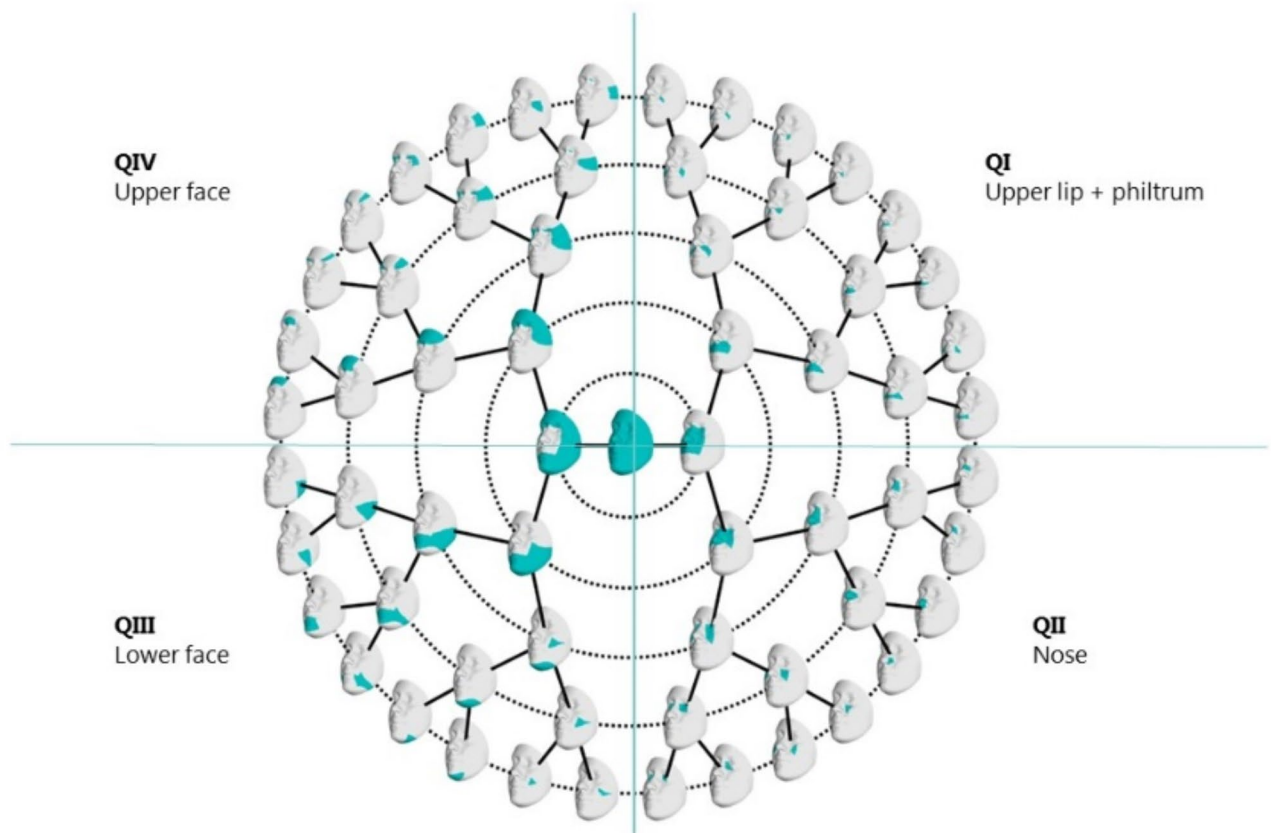
Fig. 2. Facial segmentation scheme divided into four main quadrants, adapted from the Facial Segmentation figure available on Figshare (identifier https://doi.org/10.6084/m9.figshare.7649024), as referenced in Sero et al.[57].

Nose (Quadrant II)
Findings revealed associations (*P*-value CCA < 0.05) between the nose and 50 SNPs. However, only one SNP (rs227833) reached the FDR threshold (*P*-value FDR < 0.05) (Supplementary Table S6).

As mentioned earlier, significant associations were identified between specific segments of the nose area, segments 13 and 27 and SNP rs227833 (Supplementary Fig. S4a). rs227833: G > C (each copy of G increases the effect following an additive model) is linked to an outward protrusion of the nasal root and dorsum nasi. Simultaneously, the nasal bridge, lateral ales and subnasal region exhibit an inward projection (Fig. 1c, Supplementary Fig. S4b).

Lower face (Quadrant III)
Although potential significant correlations (*P*-value CCA < 0.05) were found between the lower face and 55 SNPs, only rs11768577 and rs7167736 were considered significant as they met the FDR threshold (*P*-value FDR < 0.05) (Supplementary Table S6).

Significant associations were identified between SNP rs11768577 and the lower face, including segments 1, 2, 5, 10, 11, 20, 21, 22, 23, 41, 43, 44, 45, and 46 (Supplementary Fig. S5a). The impact of rs11768577: T > G (each copy of T increases the effect following an additive model) is shown with an outward orientation of the nasal, labial and central chin regions, contrasted with an inward depression of the remaining facial features. More specifically, there is an outward orientation of the central part of the chin and the region under the lips and, simultaneously, with an inward depression in the lateral areas of the mandible and cheeks (Fig. 1d, Supplementary Fig. S5b–f).

In addition, significant correlations were also found between the lower face, specifically segments 10, 11, 21, 22, 42, 44, and SNP rs7167736 (Supplementary Fig. S6a). The effect of rs7167736: T > G (each copy of T increases the effect following an additive model) is associated with an inward projection of the lower lip and the region below, as well as the lateral areas of the menton. The chin and the central part of the menton shows an outward orientation (Fig. 1e, Supplementary Fig. S6b–c).

SNP	Association with facial morphology through previous studies	Association with facial morphology through the present study
rs6535972	Upper face[10]	Global and Upper face
rs11768577	Chin[8]	Global and Lower face
rs227833	Nose[4,12,15]	Nose
rs7167736	Chin[8]	Lower face
*rs17275866**	*Chin[8]*	*Upper face*
*rs702489**	*Nose[8]*	*Upper lip + philtrum*
*rs12786942**	*Upper face[24]*	*Upper lip + philtrum*
rs3936018	Upper face[9,14,30]	Upper face
rs72691108	Upper face[12]	Upper face
rs7738892	Upper face[10]	Upper face

Table 2. Statistically significant SNPs (*P*-value FDR < 0.05) identified in our study and their associations with facial morphology, from both previous studies and the current investigation. SNPs with novel associations are highlighted in italics and marked with an asterisk.

Upper face (Quadrant IV)

Associations (*P*-value CCA < 0.05) were observed between 50 SNPs and the upper face. Among them, five SNPs (rs3936018, rs72691108, rs6535972, rs17275866, and rs7738892) reached the FDR threshold (*P*-value FDR < 0.05) and were considered significant (Supplementary Table S6).

For both SNPs rs3936018 and rs72691108, significant associations were found including segments 38 and 39, located in the forehead area (Supplementary Fig. S7a). The impact associated with rs3936018: T > C and rs72691108: G > A (each copy of T or G respectively increases the effect following an additive model) is shown as a depression of the forehead centrally and a protrusion laterally (Fig. 1f, Supplementary Fig. S7b–c).

Statistically significant associations were also observed between SNP rs6535972 and overall facial structure (segments 1 and 2). Additionally, specific segments of the upper face, such as segments 4, 8, and 9 displayed significant correlations too (Supplementary Fig. S8a). As illustrated in Supplementary Fig. S7, rs6535972: G > C (each copy of G increases the effect following an additive model) is predominantly linked to an outward protrusion of the central forehead and the upper periorbital region, and to an inward projection of the laterals of the forehead and the lower periorbital region (Fig. 1g, Supplementary Fig. S8b–f).

Moreover, SNP rs17275866 was significantly associated with the upper face, a novel and previously undescribed region, specifically segments 9, 18, and 37 (Supplementary Fig. S9a). The effect of rs17275866: G > C (each copy of G increases the effect following an additive model) is found to be related to an inward depression of the forehead region, with the exception of specific areas in the eyebrow and the upper part of the forehead, which exhibit an outward protrusion (Fig. 1h, Supplementary Fig. S9b).

Finally, significant associations were observed between rs7738892 and two segments of the upper face, segments 18 and 37 (Supplementary Fig. S10a). SNP rs7738892: C > T (each copy of G increases the effect following an additive model) impacts the frontal forehead region, leading to an inward orientation of the eyebrow area and an outward projection of the area above the temples (Fig. 1i, Supplementary Fig. S10b).

Discussion

The face is a distinctive and highly variable component of an individual's appearance[58]. Investigating facial diversity through DNA enhances our understanding of genetic bases underlying human facial variation and has significant forensic implications. Identifying genetic markers to predict facial appearance from DNA is a key goal in FDP. This study provides support for genetic markers associated with facial morphology in a Spanish population. We started with 116 candidate SNPs previously associated with facial morphology in other populations. Of these, 101 SNPs exhibited a *P*-value < 0.05 following CCA, suggesting that the influence on facial shape observed in prior studies is also present in our Spanish population. However, after FDR adjustment, only ten SNPs demonstrated statistically significant associations within normal-range facial shape phenotypes in our cohort (Supplementary Table S6). Hence, we will focus our discussion on these ten SNPs, as they have shown a robust and strong signal.

Notably, seven of the identified SNPs corroborate associations with facial regions previously established in different studies[4,8,9,11,12,14,15,30], emphasizing their significance across various populations. In addition, within this set of ten SNPs, three novel associations with distinct facial region have been uncovered (Table 2). Specifically, our findings highlight the association of rs17275866 with the upper face, diverging from its previous associations with the chin[8]. Furthermore, our study establishes an association between rs702489 and the philtrum region, notwithstanding prior investigations linking this SNP with nose morphology[8]. Lastly, rs12786942 shows correlation with the lateral areas of the upper lip and philtrum, which contrasts with its prior associations with the upper face[24]. These findings underscore the intricate genetic mechanisms influencing facial morphology and suggest that associations observed in populations with European ancestry, when considered as a single group, may require further examination within smaller, specific subpopulations to determine if similar effects are present. Thus, they highlight the value of comprehensive investigations in different regions to elucidate the multifaceted nature of genetic contributions towards facial traits.

When examining the SNPs that have shown a strong correlation with facial morphology within our cohort, their distribution aligns with that observed in the European population[54]. Biologically, these SNPs are associated

with or located near genes involved in morphofacial development. More detailed information on these associations can be found in Supplementary Note 1.

Considering that the findings hereby presented could help in the development of predictive models for facial morphological traits, it is important to highlight regions where multiple SNPs have demonstrated influence in our population. For instance, both rs6565972 and rs11768577 exhibit a global effect, affecting the whole face. On the one hand, the presence of the G allele in rs6535972 correlates with an outward protrusion of the frontal, nasal, labial, and chin regions, along with an inward projection of the cheeks (Supplementary Fig. S7). On the other hand, the T allele in rs11768577 is associated with an outward orientation of the nasal, labial, and central chin regions, juxtaposed with an inward depression of the remaining facial features (Supplementary Fig. S4). Although both SNPs affect the entire face, more localized analyses suggest that rs6565972 may have a more pronounced effect on the upper face, while rs11768577 might have a greater impact on the lower face. To better understand how these SNPs interact and the effects of their combined genotypes, future studies assigning varying weights to each of these markers depending on the facial segment will be necessary.

This situation is repeated in other facial quadrants, particularly in the upper face, where rs17275866, rs3936018, rs72691108, rs7738892, and rs11768577 (previously mentioned with both global and local effects) have shown a relationship. In the lower face different SNPs, including rs11768577 and rs7167736, are also involved. The remaining SNPs have demonstrated more specific correlations, rs702489 and rs12786942 with different upper lip segments, and rs227833 with the nasal area. However, even in these cases, to understand the phenotype fully, the influence of markers rs6565972 and rs11768577, which exhibit global effects, must also be considered. This underscores the complexity of predicting facial shape. Although associations between individual SNPs and various traits are observed, these markers do not act in isolation and it is important to understand how multiple SNPs work in concert to produce a single phenotype. An individual has different possible combinations for each marker, which could result in one SNP being associated with a wider face and another with a narrower face.

Following the implication of multiple SNPs in the same region, our LD results suggest co-segregation of rs3936018 and rs72691108, as well as other markers (*TBX15-WARS2* rs1229119/rs1766789/rs3936018/ rs72691108). Additionally, from our set of ten SNPs, rs17275866, rs11768577, and rs77388952 have also shown co-segregation with different previously associated face markers (*LINC02268* rs17275866/rs60975547, *SEM1-LOC105375412* rs11768577/ rs62470051, and *KIF6* rs2092524/rs7738892 respectively) (Supplementary Table S4). This opens a new avenue for studying the influence of SNPs not only individually but as haplotypes, which can provide significant new information regarding facial appearance.

Another challenge in developing DNA-based prediction models is the small proportion of facial variance explained by face-associated SNPs. Sexual dimorphism, along with other covariates such as age or ancestry, play an essential role in the human face. In fact, up to one-third of the variance in several facial features can be attributed to sex and ancestry[58]. Therefore, we corrected our 3D images for different covariates (sex, age, height, and weight) using PLSR before conducting any correlation analysis between each SNP and the phenotypic variation. Ancestry was not considered as all individuals were from the same population and no facial variation differences were observed between individuals from different regions (Supplementary Fig. S11). This can be attributed to the substantial interregional genetic admixture within Spain due to internal migration, making birthplace an insufficient indicator of geographic ancestry. To minimize potential misclassification, we treated our population as uniform within itself and geographic region was not included as a variable in our analysis. Once the set of ten SNPs was established, we assessed the percentage of variability explained by each marker (Supplementary Table S7 and Fig. S12). The low value obtained, with all ten SNPs combined explaining only 0.705% of the variance, is consistent with previous findings, where each SNP explained less than 1% of the overall phenotypic facial variation[10,11].

This highlights that, although FDP has made significant progress in predicting certain traits like pigmentation, hair morphology, age, and ancestry, accurately estimating facial appearance remains challenging. The polygenic nature of facial features, the low variability explained by each SNP using the methods highlighted here, and the large, sometimes irreproducible, number of markers found in GWAS across different populations, are some of the barriers in this field. However, studies like this one elucidate which genetic loci may be most informative about facial morphological traits and thus, providing valuable insights for future predictive models.

Conclusions

In summary, we describe a study of candidate SNPs focused on facial phenotyping within a Spanish population. This research not only reinforces numerous previously identified loci, albeit with generally weak associations likely due to sample size limitations, but also reveals strong associations for several genetic loci in the Spanish population. Furthermore, we have found novel and previously undescribed links between ten candidate SNPs and specific facial regions. Among them, two are directly correlated with unspecific facial morphology, providing valuable insights into general facial appearance. Additionally, three SNPs show associations with wide facial regions, while the remaining five SNPs have more localized effects, shedding light on the genetics underlying facial morphology development.

Our findings highlight the importance of conducting candidate SNPs studies across diverse reference populations to identify markers that are robust while providing accurate information for FDP applications. By employing consistent methodologies, including the same facial mesh, segmentation techniques, and SNPs established in prior research, we obtained significant results that, in some cases, differ slightly from prior observations. While much research has traditionally focused on European populations as a homogeneous group, our study shows that a Spanish population can exhibit differences. Even on a smaller level, these results suggest that it might be valuable to not only conduct large-scale European studies, but also to consider specific, smaller

subpopulation studies. This approach could enhance the reproducibility of results and provide more solid and precise data for future predictive modeling.

For future research, a continued investment in developing a consortium that includes experts in forensic genetics, anthropology, and bioinformatics will be invaluable. Such collaboration will allow the pooling of sufficient knowledge and resources to scale and optimize our findings. Continued study in this FDP subfield is essential to achieve the precision and accuracy necessary for developing potential DNA-based prediction models with practical use in real-world forensic scenarios.

Data availability
The datasets used and/or analyzed during the current study available from the MMP on reasonable request.

References
1. Kayser, M. Forensic DNA phenotyping predicting human appearance from crime scene material for investigative purposes. *Forens. Sci. Int. Genet.* **18**, 33–48 (2015).
2. Richmond, S., Howe, L. J., Lewis, S., Stergiakouli, E. & Zhurov, A. Facial genetics: A brief overview. *Front. Genet.* **9**, 462 (2018).
3. Xavier, C. et al. Development and inter-laboratory evaluation of the VISAGE enhanced tool for appearance and ancestry inference from DNA. *Forens. Sci. Int. Genet.* **61**, 102779 (2022).
4. Li, Y. et al. LYPLAL1, PRDM16, PAX3, DKK1, TNFSF12, CACNA2D3, and SUPT3H gene variants influence facial morphology in a Eurasian population. *Hum. Genet.* **138** (6), 681–689 (2019).
5. Liu, F. et al. A genome-wide association study identifies five loci influencing facial morphology in Europeans. *PLoS Genet.* **8**(9), e1002932 (2012).
6. Tsagkrasoulis, D., Hysi, P., Spector, T. & Montana, G. Heritability maps of human face morphology through large-scale automated three-dimensional phenotyping. *Sci. Rep.* **7**, 45885 (2017).
7. Bhat, M. The human face: Genes, embryological development and dysmorphology. *Int. J. Dev. Biol.* **64** (4-5-6), 383–391 (2020).
8. Pickrell, J. K. et al. Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48** (7), 709–717 (2016).
9. Hoskens, H. et al. 3D facial phenotyping by biometric sibling matching used in contemporary genomic methodologies. *PLoS Genet.* **17** (5), e1009528 (2021).
10. Xiong, Z. et al. Combining genome-wide association studies highlight novel loci involved in human facial variation. *Nat. Commun.* **13** (1), 7832 (2022).
11. Xiong, Z. et al. Novel genetic loci affecting facial shape variation in humans. *eLife* **8**, e49898 (2019).
12. Claes, P. et al. Genome-wide mapping of global-to-local genetic effects on human facial shape. *Nat. Genet.* **50** (3), 414–423 (2018).
13. Bonfante, B. et al. A GWAS in Latin Americans identifies novel face shape loci, implicating VPS13B and a Denisovan introgressed region in facial variation. *Sci. Adv.* **7** (6), eabc6160 (2021).
14. White, J. D. et al. Insights into the genetic architecture of the human face. *Nat. Genet.* **53** (1), 45–53 (2021).
15. Huang, Y. et al. A genome-wide association study of facial morphology identifies novel genetic loci in Han Chinese. *J. Genet. Genom.* **48** (3), 198–207 (2021).
16. Cho, H. W., Ban, H. J., Jin, H. S., Cha, S. & Eom, Y. B. Effect of genetic variants in UBE2O and TPK1 on facial morphology of Koreans. *Forens. Sci. Res.* **8** (1), 62–69 (2023).
17. Adhikari, K. et al. A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation. *Nat. Commun.* **7**, 11616 (2016).
18. Qiao, L. et al. Genome-wide variants of Eurasian facial shape differentiation and a prospective model of DNA based face prediction. *J. Genet. Genom.* **45** (8), 419–432 (2018).
19. Cha, S. et al. Identification of five novel genetic loci related to facial morphology by genome-wide association studies. *BMC Genom.* **19** (1), 481 (2018).
20. Liu, C. et al. Genome scans of facial features in East Africans and cross-population comparisons reveal novel associations. *PLoS Genet.* **17** (8), e1009695 (2021).
21. 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**(8), e1006174 (2016).
22. Lee, M. K. et al. Genome-wide association study of facial morphology reveals novel associations with FREM1 and PARK2. *PLoS One* **12**(8), e1006174 (2017).
23. 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 (2012).
24. Shaffer, J. R. et al. Genome-Wide association study reveals multiple loci influencing normal human facial morphology. *PLoS Genet.* **12**(8), e1006149 (2016).
25. Knol, M. J. et al. Genetic architecture of orbital telorism. *Hum. Mol. Genet.* **31** (9), 1531–1543 (2022).
26. Ludwig, K. U. et al. Imputation of orofacial clefting data identifies novel risk loci and sheds light on the genetic background of cleft lip ± cleft palate and cleft palate only. *Hum. Mol. Genet.* **26** (4), 829–842 (2017).
27. Naqvi, S. et al. Precise modulation of transcription factor levels identifies features underlying dosage sensitivity. *Nat. Genet.* **55**, 841–851 (2023).
28. Crouch, D. J. M. et al. Genetics of the human face: Identification of large-effect single gene variants. *Proc. Natl. Acad. Sci. USA* **115** (4), E676–E685 (2018).
29. Wu, W. et al. Whole-exome sequencing identified four loci influencing craniofacial morphology in Northern Han Chinese. *Hum. Genet.* **138** (6), 601–611 (2019).
30. Goovaerts, S. et al. Joint multi-ancestry and admixed GWAS reveals the complex genetics behind human cranial vault shape. *Nat. Commun.* **14** (1), 7436 (2023).
31. Shaffer, J. R. et al. Multiethnic GWAS reveals polygenic architecture of earlobe attachment. *Am. J. Hum. Genet.* **101** (6), 913–924 (2017).
32. Jacobs, L. C. et al. Intrinsic and extrinsic risk factors for sagging eyelids. *JAMA Dermatol.* **150** (8), 836–843 (2014).
33. Indencleef, K. et al. The intersection of the genetic architectures of orofacial clefts and normal facial variation. *Front. Genet.* **12**, 626403 (2021).
34. Snider, T. N. & Mishina, Y. Cranial neural crest cell contribution to craniofacial formation, pathology, and future directions in tissue engineering. *Birth Defect. Res. C Embryo Today* **102** (3), 324–332 (2014).
35. Wilke, F. et al. Exploring regional aspects of 3D facial variation within European individuals. *Sci. Rep.* **13** (1), 3708 (2023).

36. Claes, P., Walters, M. & Clement, J. Improved facial outcome assessment using a 3D anthropometric mask. *Int. J. Oral Maxillofac. Surg.* **41**, 324–330 (2012).
37. Gokhman, D. et al. Reconstructing Denisovan anatomy using DNA methylation maps. *Cell* **179** (1), 180–192e10 (2019).
38. Baab, K. L. The taxonomic implications of cranial shape variation in Homo erectus. *J. Hum. Evol.* **54** (6), 827–847 (2008).
39. Dempsey, R. F. et al. Nonsyndromic craniosynostosis. *Clin. Plast. Surg.* **46** (2), 123–139 (2019).
40. De Bock, F., Braun, V. & Renz-Polster, H. Deformational plagiocephaly in normal infants: a systematic review of causes and hypotheses. *Arch. Dis. Child.* **102** (6), 535–542 (2017).
41. Jheon, A. H. et al. Moving towards precision orthodontics: An evolving paradigm shift in the planning and delivery of customized orthodontic therapy. *Orthod. Craniofac. Res.* **1**, 106–113 (2017).
42. Matthews, H. S. et al. Large-scale open-source three-dimensional growth curves for clinical facial assessment and objective description of facial dysmorphism. *Sci. Rep.* **11** (1), 12175 (2021).
43. Olalde, I. et al. The genomic history of the Iberian Peninsula over the past 8000 years. *Science* **363**, 1230–1234 (2019).
44. Núñez, C. E. Within the European periphery: Education and labor mobility in twentieth-century Spain. *Paedagog Hist.* **39** (5), 621–649 (2003).
45. Samuels, B. D. et al. FaceBase 3: Analytical tools and FAIR resources for craniofacial and dental research. *Development* **147** (18), dev191213 (2020).
46. Heike, C. L., Upson, K., Stuhaug, E. & Weinberg, S. M. 3D digital stereophotogrammetry: A practical guide to facial image acquisition. *Head Face Med.* **6**, 18 (2010).
47. Serrulla, F. Recomendaciones en antropología forense asociación española de antropología y odontología forense. AEAOF (2013).
48. Miller, S. A., Dykes, D. D. & Polesky, H. F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16** (3), 1215 (1988).
49. Marshall, O. J. PerlPrimer: Cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics* **20** (15), 2471–2472 (2004).
50. Vallone, P. M. & Butler, J. M. AutoDimer: A screening tool for primer-dimer and hairpin structures. *Biotechniques* **37** (2), 226–231 (2004).
51. Ye, J. et al. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinform.* **13**, 134 (2012).
52. Excoffier, L. & Lischer, H. E. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and windows. *Mol. Ecol. Resour.* **10** (3), 564–567 (2010).
53. IBM Corp. IBM SPSS Statistics for Windows, Version 28.0. (2021).
54. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526** (7571), 68–74 (2015).
55. Skeleton-id. https://skeleton-id.com/
56. White, J. D. et al. MeshMonk: Open-source large-scale intensive 3D phenotyping. *Sci. Rep.* **9** (1), 6085 (2019).
57. Sero, D. et al. Facial recognition from DNA using face-to-DNA classifiers. *Nat. Commun.* **10** (1), 2557 (2019).
58. Claes, P. et al. Modeling 3D facial shape from DNA. *PLoS Genet.* **10** (3), e1004224 (2014).

Acknowledgements

Author contributions

Conceptualization: MMP; methodology, BNL, MB, RMF, OML, IO, BMJ, and SJ; software: BNL and FW; formal analysis: BNL, FW, VSU, SW; supervision: MMP; writing—original draft preparation: BNL, VSU, and MMP; writing—review & editing: FW, MB, RMF, OML, IO, BMJ, SJ, and SW; project administration: MMP; funding acquisition: MMP. All authors have read and agreed to the published version of the manuscript.

Funding

Declarations

Competing interests

The VSU is partner and researcher at Panacea Cooperative Research. Her work in this project was funded by a grant Torres Quevedo PTQ2022-012683. Identities of volunteer data donors were not accessible to Panacea or any researcher involved following approved protocols by the ethical committee of University of the Basque Country UPV/EHU M10/2021/143. The remaining authors, including BNL, FW, MB, RMF, OML, IO, BMJ, SJ, SW, and MMP, declare no conflicts of interest.

Ethics approval

The study was conducted in accordance with the Declaration of Helsinki, and approved by Ethics Committee for Research on Human Subjects of the University of the Basque Country, CEISH-UPV/EHU, BOPV 32, 17/2/2014 (M10_2021_143, published on April the 22nd, 2021).

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-98748-9.

Correspondence and requests for materials should be addressed to B.N.-L. or V.S.-U.