



Common *cis*-regulatory variation modifies the penetrance of pathogenic *SHROOM3* variants in craniofacial microsomia

Hao Zhu, Jiao Zhang, Soumya Rao, et al.

Genome Res. 2025 35: 1065-1079 originally published online April 15, 2025

Access the most recent version at doi:[10.1101/gr.280047.124](https://doi.org/10.1101/gr.280047.124)

References This article cites 91 articles, 14 of which can be accessed free at:
<http://genome.cshlp.org/content/35/5/1065.full.html#ref-list-1>

Creative Commons License This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <https://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Common *cis*-regulatory variation modifies the penetrance of pathogenic *SHROOM3* variants in craniofacial microsomia

Hao Zhu,¹ Jiao Zhang,² Soumya Rao,³ Matthew D. Durbin,⁴ Ying Li,⁵ Ruirui Lang,¹ Jiqiang Liu,¹ Baichuan Xiao,¹ Hailin Shan,¹ Ziqiu Meng,¹ Jinmo Wang,¹ Xiaokai Tang,¹ Zhenni Shi,¹ Liza L. Cox,³ Shouqin Zhao,⁵ Stephanie M. Ware,^{4,6} Tiong Y. Tan,⁷ Michelle de Silva,⁷ Lyndon Gallacher,⁷ Ting Liu,⁸ Jie Mi,⁹ Changqing Zeng,¹⁰ Hou-Feng Zheng,^{11,12} Qingguo Zhang,² Stylianos E. Antonarakis,^{13,14,15} Timothy C. Cox,^{3,16} and Yong-Biao Zhang^{1,17,18}

¹School of Engineering Medicine, Beihang University, Beijing 100191, China; ²Plastic Surgery Hospital, Chinese Academy of Medical Sciences, Beijing 100144, China; ³Department of Oral & Craniofacial Sciences, University of Missouri-Kansas City, Kansas City, Missouri 64108, USA; ⁴Department of Pediatrics, Herman B. Wells Center for Pediatric Research, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA; ⁵Department of Otolaryngology-Head and Neck Surgery, Beijing Tongren Hospital, Capital Medical University, Beijing 100051, China; ⁶Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA; ⁷Victorian Clinical Genetics Service, Royal Children's Hospital and Department of Pediatrics, University of Melbourne, Victoria 3052, Australia; ⁸Department of Ophthalmology, Daping Hospital, Army Medical University, Chongqing 400000, China; ⁹Center for Non-Communicable Disease Management, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing 100045, China; ¹⁰Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China; ¹¹Center for Health and Data Science (CHDS), the Second Affiliated Hospital of Soochow University, Suzhou, Jiangsu 215004, China; ¹²Diseases & Population (DaP) Geninfo Laboratory, School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310024, China; ¹³Department of Genetic Medicine and Development, University of Geneva Medical Faculty, Geneva 1211, Switzerland; ¹⁴Medigenome, Swiss Institute of Genomic Medicine, 1207 Geneva, Switzerland; ¹⁵iGE3 Institute of Genetics and Genomes in Geneva, Geneva 1211, Switzerland; ¹⁶Department of Pediatrics, University of Missouri-Kansas City, Kansas City, Missouri 64108, USA; ¹⁷Key Laboratory of Big Data-Based Precision Medicine and Key Laboratory of Innovation and Transformation of Advanced Medical Devices, Ministry of Industry and Information Technology, Key Laboratory of Biomechanics and Mechanobiology, Ministry of Education, Beijing 100191, China; ¹⁸National Medical Innovation Platform for Industry-Education Integration in Advanced Medical Devices (Interdiscipline of Medicine and Engineering), Beihang University, Beijing 100083, China

Pathogenic coding variants have been identified in thousands of genes, yet the mechanisms underlying the incomplete penetrance in individuals carrying these variants are poorly understood. In this study, in a cohort of 2009 craniofacial microsomia (CFM) patients of Chinese ancestry and 2625 Han Chinese controls, we identified multiple predicted pathogenic coding variants in *SHROOM3* in both CFM patients and healthy individuals. We found that the penetrance of CFM correlates with specific haplotype combinations containing likely pathogenic-coding *SHROOM3* variants and CFM-associated expression quantitative trait loci (eQTLs) of *SHROOM3* expression. Further investigations implicate specific eQTL combinations, such as rs1001322 or rs344131, in combination with other significant CFM-associated eQTLs, which we term combined eQTL phenotype modifiers (CePMods). We additionally show that rs344131, located within a regulatory enhancer region of *SHROOM3*, demonstrates allele-specific effects on enhancer activity and thus impacts expression levels of the associated *SHROOM3* allele harboring any rare coding variant. Our findings also suggest that CePMods may serve as pathogenic determinants, even in the absence of rare deleterious coding variants in *SHROOM3*. This highlights the critical role of allelic expression in determining the penetrance and severity of craniofacial abnormalities, including microtia and facial asymmetry. Additionally, using quantitative phenotyping, we demonstrate that both microtia and facial asymmetry are present in two separate *Shroom3* mouse models, the severity of which is dependent on gene dosage. Our study establishes *SHROOM3* as a likely pathogenic gene for CFM and demonstrates eQTLs as determinants of modified penetrance in the manifestation of the disease in individuals carrying likely pathogenic rare coding variants.

[Supplemental material is available for this article.]

Corresponding authors: zhangyongbiao@buaa.edu.cn, coxtc@umkc.edu, stylianos.antonarakis@unige.ch

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.280047.124>.

© 2025 Zhu et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <https://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Incomplete penetrance represents a complex phenomenon whereby the presence of a pathogenic variant in the genome results in a disease phenotype in some but not all individuals that carry the variant (Vogt 1926; Cooper et al. 2013). This variability is attributed to a multitude of factors, including but not limited to genetic modifiers, environmental conditions, and epigenetic modifications. Variants in regulatory elements that impact gene expression, otherwise known as expression quantitative trait loci (eQTLs), serve as one of the genetic modifiers that influence the penetrance of pathogenic variants (Castel et al. 2018). These common variants, which modify gene expression in specific haplotype combinations with rare pathogenic alleles, “interact” with the latter, thereby modulating their penetrance (Emison et al. 2005; Lim et al. 2024).

The etiology of many diseases often adheres to a genetic model, encompassing diverse combinations of regulatory and pathogenic variants (Emison et al. 2005; Rahit and Tarailo-Graovac 2020). In such diseases, the penetrance of pathogenic variants is influenced not only by the rare genetic alteration itself but also by the presence of a concurrent regulatory variant that dictates allele-specific expression (ASE) levels (Castel et al. 2018). However, to date, the combinatorial patterns of parental alleles, comprising both deleterious and regulatory variants, that give rise to birth defects remain inadequately investigated.

Craniofacial microsomia (CFM) is a condition that encompasses considerable phenotypic variability and low penetrance and is often recognized by notable facial asymmetry involving the external ear, mandible, and maxilla. A previous large genome-wide association study (GWAS) of CFM patients identified numerous risk loci, including *SHROOM3*, which carries the highest burden of deleterious coding variants among CFM-associated genes (Supplemental Fig. 1; Zhang et al. 2016; Xu et al. 2021). *Shroom3* shows high expression in the pharyngeal arches, which develop into various craniofacial structures, including the external ear (Hildebrand and Soriano 1999). In mice, homozygous loss of function of *Shroom3* has been amply described as having variably penetrant neural tube defects (exencephaly, spina bifida), facial clefts, and cardiac and kidney anomalies, among other anomalies (Hildebrand and Soriano 1999; McGreevy et al. 2015; Durbin et al. 2020; Lawlor et al. 2023). Although small ears were noted in the original report (Hildebrand and Soriano 1999), this phenotype was not emphasized, and heterozygotes appear grossly normal. Further studies are needed to elucidate how *SHROOM3* impacts craniofacial phenotypes, including asymmetry and microtia, and to establish its role in modifying susceptibility to CFM.

In this study, we aim to elucidate the role of variants in *SHROOM3* in modifying the penetrance and expressivity of CFM by focusing on combinatorial eQTLs that interact with deleterious coding variants. We also investigate the contribution of *SHROOM3* to phenotypic variability in CFM by utilizing quantitative assessments in mouse models.

Results

Putative deleterious rare coding variants in *SHROOM3*

We conducted whole-genome sequencing on 2009 Chinese CFM patients and enrolled, as controls, 2625 healthy Chinese individuals from the SG10K data set (Wu et al. 2019). Previous association studies had identified several CFM-related risk genes (Zhang et al. 2016; Xu et al. 2021). Initially, we focused our efforts on interrogating the predicted deleterious variants within risk genes directly

encompassed by GWAS signals in both cohorts (Supplemental Fig. 1). *SHROOM3* was found to harbor the highest number of predicted deleterious coding variants in our cohorts and therefore became the focus of this study.

Within the CFM cohort, 15,481 variants were detected across the *SHROOM3* locus (including 100 kb on either side of the gene; Chr 4: 76,335,229–76,883,253, hg38), with 43 rare coding variants identified as putatively deleterious. These variants were carried by 65 patients (Fig. 1; Table 1; Supplemental Table 1). Additionally, we included 206 CFM patients from the Gabriella Miller Kids First Pediatric Research (GMKF) program, 591 individuals (264 affected) available through the FaceBase program, and 12 CFM families who were mostly of Australasian/European descent and found an additional nine putative deleterious rare coding variants (Table 1; Supplemental Table 1). In 2625 healthy Chinese individuals, *SHROOM3* contained 32 predicted deleterious coding variants, carried by 85 individuals (Supplemental Table 2). These predicted deleterious coding variants within CFM patients and healthy Chinese provided the opportunity to investigate the incomplete penetrance potentially modified by eQTLs.

A gene burden analysis indicated a significant enrichment of deleterious variants in CFM patients compared with the 2625 healthy Han Chinese ($P=0.008$, SKAT burden test). Notably, compared with the 2009 patients, the subset of 65 patients showed a marked preference for the conchal type auricular presentation (51 out of 65, 78%) as per the Nagata classification, in contrast to 28% in the entire cohort ($P=2.88 \times 10^{-8}$, Pearson’s chi-squared test) (Supplemental Table 3). The OMENS classification further identified O1 (34%; abnormal orbital size) and S0 (48%; no obvious soft tissue or muscle deficiency) as the primary categories (Supplemental Table 3). This provides evidence that *SHROOM3* is primarily associated with the development of the upper auricle.

Identifying potential regulatory variants through an association analysis

To investigate regulatory variants that have the potential to modify the effect of deleterious rare coding variants, we undertook an association analysis on variants in the *SHROOM3* region with the 2009 Chinese CFM patients and 2625 healthy Han Chinese from the SG10K data set. Association tests revealed a large number of common variants significantly associated with CFM, surpassing a stringent P -value threshold of $<5 \times 10^{-8}$, using the first 10 eigenvalues as covariates for adjusting population stratification (Fig. 1B; Supplemental Fig. 2; Supplemental Table 4). These results replicated previous observations ascertained from chip-based GWASs (Zhang et al. 2016). A total of 84 of the associated variants are *SHROOM3* eQTLs (denoted in GTEx database) (Supplemental Table 4; Supplemental Fig. 3; The GTEx Consortium 2020), and 30 of the associated variants are mapped in regions with promoter/enhancer histone marks (Supplemental Table 5). The most significant variant is rs10017322 with a P -value of 2.34×10^{-44} (after Bonferroni correction), which is an eQTL for *SHROOM3*.

Fine-mapping eQTL modifiers of predicted pathogenic *SHROOM3* variants in CFM patients

An evaluation of linkage disequilibrium (LD) patterns divided the common variants into five distinct haplotype blocks (H1–H5) situated within the *SHROOM3* locus (Supplemental Fig. 4). To fine-map candidate pathogenic eQTLs influencing the penetrance of CFM through altered ASE, we utilized conditional association

were eQTLs (Supplemental Table 6). Combining the results from these fine-mapping methods, we identified two unlinked eQTLs, rs10017322 and rs344131 ($r^2=0.05$), consistently flagged as putatively pathogenic eQTLs across all methods (Fig. 2A).

To elucidate the haplotype combinations of the two putatively modifying eQTLs and deleterious variants between the case (65 CFM Chinese patients) and control (85 healthy Han Chinese) samples with monoallelic, likely deleterious, *SHROOM3* variants, we performed haplotype phasing. For inferring the haplotypes of variants identified in the *SHROOM3* region, we employed SHAPEIT5, a phasing method optimized for accurate processing of rare variants in extensive sequencing data sets (Hofmeister et al. 2023). Acknowledging that not all coding predicted deleterious variants are necessarily pathogenic, we applied stringent criteria for defining pathogenic variants. The stringent criteria are as follows: (1) For loss-of-function (LoF) variants, they must be predicted as high-confidence (HC) by loss-of-function transcript effect estimator (LOFTEE), and (2) for nonsynonymous variants, they must be located within functional domains and either predicted as pathogenic by AlphaMissense or predicted as ambiguous by AlphaMissense with an ESM1b score below -10 .

This stringent selection resulted in five predicted pathogenic variants in CFM Chinese patients and two in healthy Chinese (Fig. 1A; Table 1; Supplemental Tables 1, 2). Upon analyzing the haplotype combinations, we observed that in CFM Chinese patients, all five predicted pathogenic mutant alleles were in the same haplotype with the high-expression allele (G allele) of rs10017322, whereas in controls, the mutant alleles of the two predicted pathogenic variants were combined with the low-expression allele (A allele) of rs10017322 (Fig. 2B; Supplemental Fig. 3). To mitigate potential population selection bias owing to the absence of LoF variants in the control samples, we obtained data from various healthy population databases, including the 1000 Genomes Project (Sudmant et al. 2015), the Westlake BioBank for Chinese (WBBC) (Cong et al. 2022), and the Chinese Academy of

Sciences Precision Medicine Initiative (CASPMI) (Du et al. 2019). We identified the p.(Thr753ProfsTer88) variant in the Finnish population from the 1000 Genomes Project, the p.(Gln1117GluTer73) variant from WBBC, and the p.(Ser1535ProfsTer4) and p.(Glu1872Ter) variants from CASPMI. All these LoF variants are on the same haplotype as the low-expression allele (A allele) of rs10017322. This combination implies a modified penetrance in controls, influenced by the reduced expression level of the pathogenic allele of rare variants (Fig. 2B).

We utilized the missense score (see Methods) to evaluate the functional effects of all identified variants in *SHROOM3* and confirmed that all five predicted pathogenic variants exhibited significant effects (Fig. 2C). To capture the relationship between eQTLs and predicted pathogenic variants, we introduced the eQTL-P model. This model defines an eQTL-P as a combination in which the mutant allele of a predicted pathogenic variant is in *cis* with a high-expression allele of an eQTL (e.g., rs10017322), potentially increasing the penetrance of the pathogenic allele. Using this model, we identified six additional variants following the same pattern, in which the mutant allele is combined in *cis* with the high-expression allele of rs10017322 (Fig. 1, variants marked in bold; Table 1, marked as eQTL-P; Supplemental Table 1).

However, when conditioning on the carrier status of all 11 presumptively pathogenic variants (the five predicted pathogenic variants plus the six deleterious variants that follow the combination pattern), only minimal alterations were observed in the significant CFM-associated variants (Supplemental Fig. 5B). Furthermore, on deeper examination, conditioning on all rare variants within a 1 kb region of all *cis*-regulatory (*cis*-reg) elements and the 11 presumptively pathogenic variants, a considerable number of variants remained significantly associated with CFM (Supplemental Fig. 5C). These results indicate that other variants such as eQTLs identified from fine-mapping analysis, along with rare coding variants, may also play a pivotal role in the complex molecular mechanism of phenotypic penetrance.

Table 1. Likely pathogenic variants in *SHROOM3* for CFM patients from Chinese, European, and Hispanic ancestries

Chromosome position (hg38)	<i>SHROOM3</i> cHGVS NM_020859.4	<i>SHROOM3</i> pHGVS	AlphaMissense prediction	Domain	Model	Populations
Chr 4: 76,555,619	c. 179G>T	p.(Gly60Val)	Pathogenic	PDZ	eQTL-P	GMKF
Chr 4: 76,555,718	c. 278C>T	p.(Ser93Phe)	Ambiguous	PDZ	eQTL-P	CCFM
Chr 4: 76,555,759	c. 319C>T	p.(Arg107Cys)	Ambiguous	PDZ		CCFM
Chr 4: 76,555,760	c. 320G>A	p.(Arg107His)	Ambiguous	PDZ		CCFM
Chr 4: 76,730,823	c. 475C>T	p.(Arg159Ter)				FaceBase
Chr 4: 76,739,581	c. 1408G>C	p.(Val470Leu)	Ambiguous			CCFM
Chr 4: 76,740,538	c. 2365G>A	p.(Glu789Lys)	Pathogenic		eQTL-P	CCFM
Chr 4: 76,741,361	c. 3188G>C	p.(Arg1063Pro)	Ambiguous	ASD1	eQTL-P	GMKF
Chr 4: 76,741,475	c. 3302_3347del	p.(Ala1102ArgfsTer23)			eQTL-P	CCFM
Chr 4: 76,741,847	c. 3674T>C	p.(Leu1225Pro)	Pathogenic			CCFM
Chr 4: 76,741,904	c. 3731C>T	p.(Pro1244Leu)	Ambiguous			CCFM
Chr 4: 76,755,001	c. 4518C>G	p.(Asp1506Glu)	Ambiguous			GMKF
Chr 4: 76,770,802	c. 5526C>G	p.(Asp1842Glu)	Ambiguous	ASD2		CCFM
Chr 4: 76,770,821	c. 5545C>T	p.(Leu1849Phe)	Pathogenic	ASD2	eQTL-P	CCFM
Chr 4: 76,778,885	c. 5699G>T	p.(Arg1900Leu)	Pathogenic	ASD2	eQTL-P	CCFM

(Domain) Protein domain; (Populations) the variant identified from Chinese CFM (CCFM) or Gabriella Miller Kids First (GMKF). eQTL-P marks the pathogenic variant in combination with rs10017322, with red mark predicted pathogenic variants.

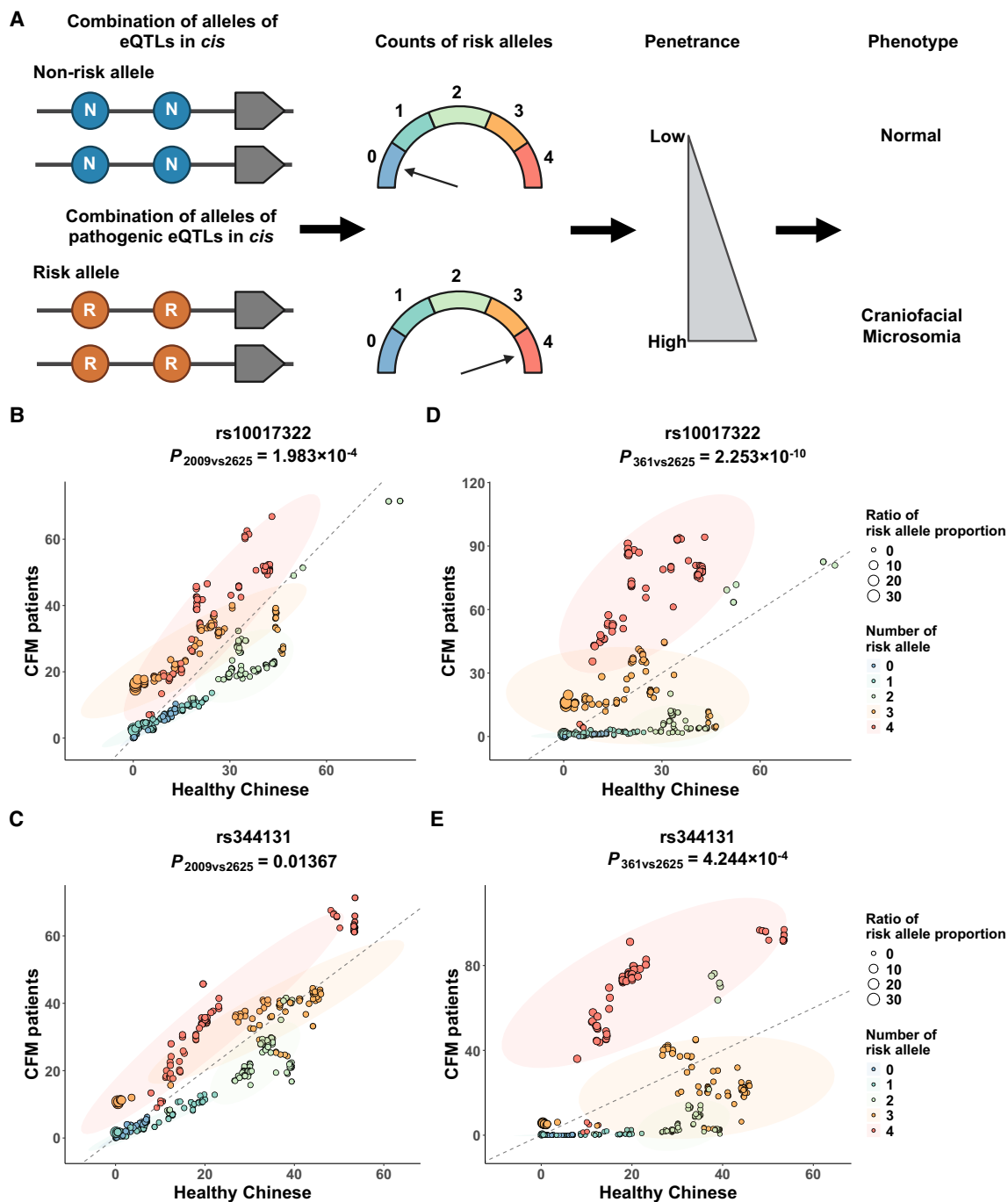


Figure 3. Patterns of haplotype combinations involving putatively pathogenic eQTLs. (A) Conceptual diagram hypothesizing that the combination of risk alleles from *SHROOM3* eQTLs correlates with the CFM phenotype. The greater the number of risk alleles carried, the higher the likelihood of being a CFM patient. (B,C) Risk allele combinations involving the putatively pathogenic eQTL rs10017322/rs344131 and other *SHROOM3* eQTLs in 2009 Chinese CFM patients and 2625 healthy Chinese individuals. Carriers of four risk alleles show a significant enrichment ($P=0.0001983$) in CFM patients. (D,E) Risk allele combinations as in B,C but analyzed in 361 CFM patients who strongly contribute to the *SHROOM3* signal compared with 2625 healthy Chinese. Individuals carrying four risk alleles demonstrate a more significant enrichment ($P=2.253 \times 10^{-10}$) in CFM patients. Ratio of risk allele proportion indicates the proportion of the risk allele compared with CFM patients and healthy Chinese.

These results suggest that rs344131 acts as a potential pathogenic eQTL in CFM.

The significant enrichment of specific eQTL combinations in CFM patients suggests a role for common regulatory variants in CFM etiology. To validate this hypothesis, we focused on an en-

riched eQTL pair, rs344131 and rs61090632 (Fig. 4C). We cloned enhancer sequences of EH38E2305115 and EH38E2304970, containing different alleles of rs344131 (A/G, G being the risk allele) and rs61090632 (T/A, A being the risk allele), into firefly luciferase reporter plasmids. Luciferase assays revealed that combinations

containing risk alleles significantly reduced enhancer activity compared with the wild-type (AT) combination (Fig. 4D). These findings suggest that combinations of deleterious regulatory alleles may impair enhancer function, potentially contributing to CFM pathogenesis.

Craniofacial phenotypes in *Shroom3* mouse models

To better appreciate the role of *SHROOM3* in craniofacial development and the potential impact of reduced *SHROOM3* expression levels as a contributor to CFM, we analyzed facial morphology initially in the *Shroom3* gene trap (*Shroom3st*) mutant line. We qualitatively and quantitatively assessed *Shroom3* heterozygous null (*Shroom3^{st/st}*), homozygous null (*Shroom3^{st/st}*), and background-matched wild-type littermate control (*Shroom3^{+/+}*) embryos following three-dimensional (3D) optical projection tomography (OPT) imaging. As previously reported, all *Shroom3^{st/st}* homozygotes presented with exencephaly and variable penetrant additional features, including spina bifida (40%) and orofacial/oronasal clefts (30%) (Martin et al. 2019). Many homozygotes appeared to exhibit midface and mandibular hypoplasia, evident by a slightly protruding tongue, and small and sometimes

protruding external ears (Fig. 5A). Heterozygotes appeared grossly normal.

Landmark-based quantitative analysis of facial features was then performed, including maxillary and mandibular lengths, auricular size, and various measures of facial asymmetry. Head width measurements, as a proxy for overall size, showed that the head size of homozygotes was notably smaller than that of controls ($P = 0.0258$) (Supplemental Fig. 6). Overall, heterozygotes tended to be slightly smaller than littermate controls, although this did not reach statistical significance. Maxillary and mandibular length measurements revealed homozygotes to have markedly shorter mandibles and maxillary tissue (mandibles: $P = 0.0124$; maxillary tissue: $P = 0.0018$) (Fig. 5B–D), confirming the gross observations. These measurements remained significant even after normalization to head width. The mandibles and maxillary tissue of younger heterozygous embryos tended to be shorter than that of controls, but only the mandibles reached significance ($P = 0.0272$) (Fig. 5B–D, top). These findings suggest a subtle overall delay in craniofacial growth in heterozygotes but a larger and consistent impact on maxillary and mandibular growth in homozygotes.

In line with initial gross observations, measurements of auricular size confirmed that all homozygous embryos had marked

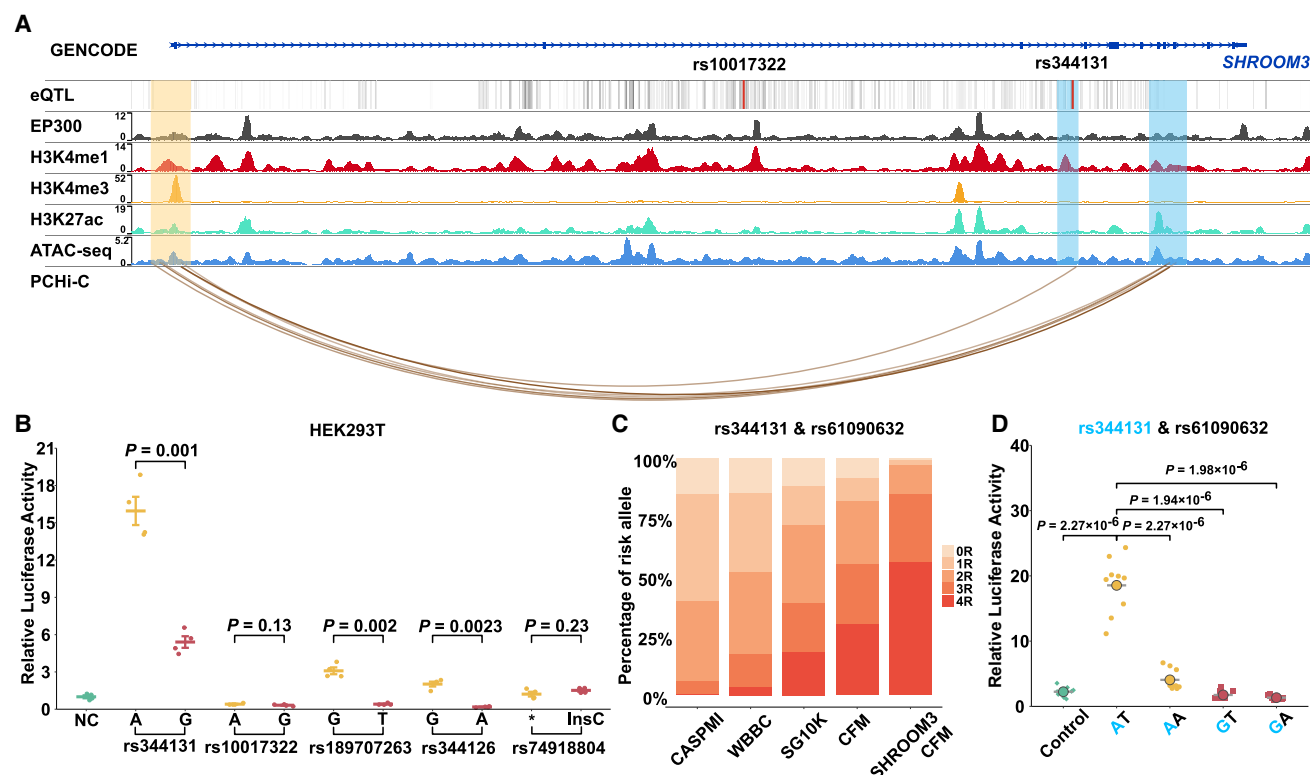


Figure 4. In vitro analysis of eQTL effects on enhancer activity. (A) Epigenetic landscape at *SHROOM3* in human neural crest cells (hCNCCs). This panel presents an integrated epigenomic analysis, combining in vitro PCHI-C, ATAC-seq, and ChIP-seq data from hCNCCs (Rada-Iglesias et al. 2012; Xu et al. 2024). It displays the epigenetic markers H3K4me1, H3K4me3, and H3K27ac in hCNCCs. The promoter is highlighted with light orange shading. Enhancers are highlighted with blue shading, with eQTLs within these enhancers marked by black lines (darkness represents the significance level). In the PCHI-C section, differential chromatin interactions connect enhancers to the *SHROOM3* promoter, with line colors indicating the strength of interaction. (B) Luciferase assays to assess the effects of eQTLs on *SHROOM3* enhancers, including an empty vector as a control. Data from three independent experiments, each comprising three technical replicates, are depicted. (C) Enrichment analysis of risk allele combinations for eQTLs rs344131 and rs61090632 in CFM patients compared with control populations. Significant enrichment is observed in both the entire CFM cohort ($P < 0.008884$) and the *SHROOM3* CFM subset ($P < 1.71 \times 10^{-24}$) using chi-square tests. Control populations: (CASPMI) Chinese Academy of Sciences Precision Medicine Initiative, (WBBC) Westlake BioBank for Chinese, and (SG10K) Singapore 10K Genome Project. Patient cohorts: CFM ($n = 2009$) and *SHROOM3* CFM ($n = 361$). (D) Luciferase assay results demonstrating the combined effect of alleles rs344131 and rs61090632 in HEK-293T cells. P -values are calculated using the Student's t -test.

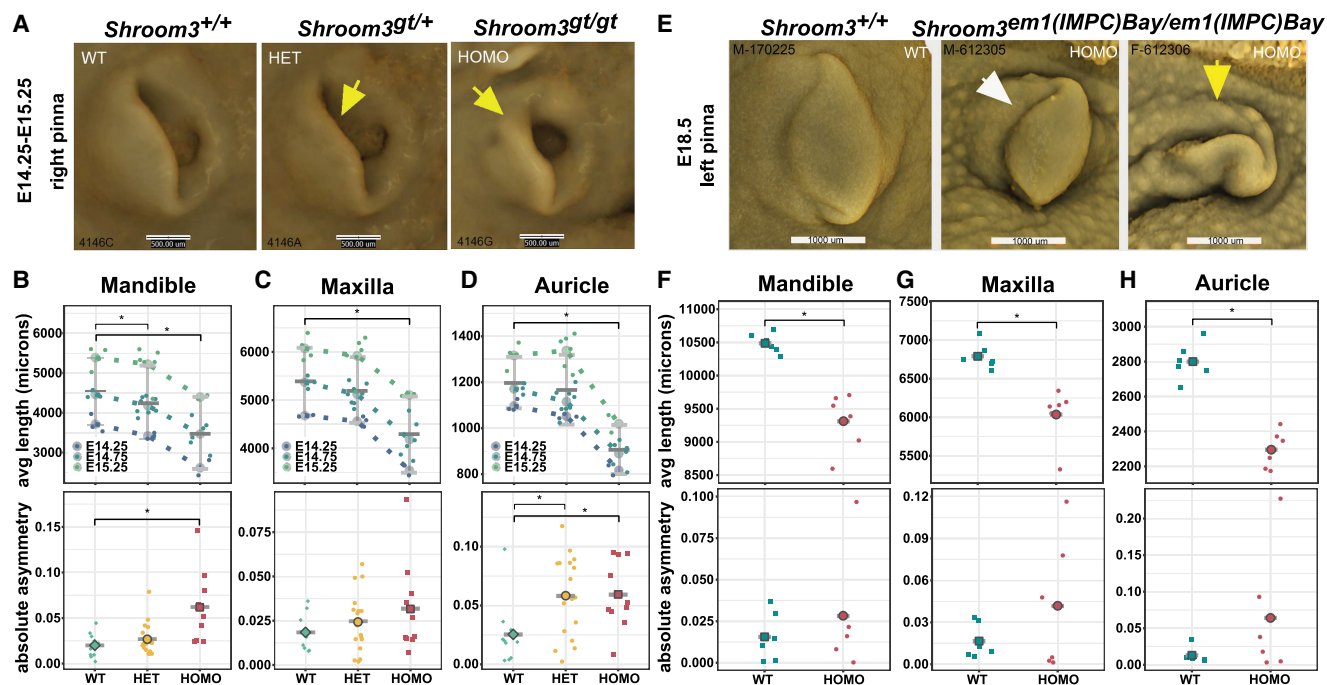


Figure 5. In vivo analysis of *Shroom3* zygosity in mice is associated with microtia and facial asymmetry. (A) Lateral view of 3D rendered developing auricles (right side) from E14.25 wild-type (*Shroom3*^{+/+}), heterozygote (*Shroom3*^{gt/+}), and homozygote (*Shroom3*^{gt/gt}) littermates following optical projection tomography. Note the marked microtia in the homozygote and the subtle dorsal auricular deficiency in the heterozygote (both marked by yellow arrows in the respective images) compared with the wild-type littermate. (B–D) Landmark-based quantitative assessment of auricular, mandibular and maxillary length (top) and absolute asymmetry (bottom) in wild-type, *Shroom3*^{gt/+}, and *Shroom3*^{gt/gt} embryos between E14.25 and E15.25. In the graphs of the average lengths of the mandible, maxilla, and auricle, dotted lines “connect” the means of the different genotypes from the same litter (stage). Mean \pm SD indicated for the average length measurements; the mean (large shape with horizontal line dissecting it) is shown for the absolute asymmetry measurements. Asterisks denote statistical significance ($P < 0.05$). (E) Lateral view of 3D rendered auricles (left side) from an E18.5 wild type (*Shroom3*^{+/+}) and two age-matched homozygote (*Shroom3*^{em1(IMPC)Bay/em1(IMPC)Bay}) embryos derived from microCT scans generated by the International Mouse Phenotyping Center (IMPC; <https://www.mousephenotype.org>). The first homozygote (central) shows the more typical presentation of a small pinna that is inwardly curved and posteriorly rotated (white arrow). The second homozygote had significantly asymmetric pinna phenotypes - the auricular presentation of the left ear (yellow arrow) was severe, whereas the right pinna was similar in presentation to the first homozygote. (F–H) Landmark-based quantitative assessment of mandibular, maxillary, and auricular length (top) and absolute asymmetry (bottom) in E18.5 wild-type (*Shroom3*^{+/+}) and *Shroom3*^{em1(IMPC)Bay/em1(IMPC)Bay} null embryos. The mean (large shape with horizontal line dissecting it) is shown for each. Asterisks denote statistical significance ($P < 0.05$).

microtia that often showed significant asymmetry in presentation (length: $P = 0.0004$; absolute asymmetry: $P = 0.0169$) (Fig. 5B–D). The differences remained significant even after normalization of head size. A side-by-side comparison of auricular morphology with controls (Fig. 5A) confirmed that homozygotes have cupped and protruding auricles, with reduction of the size of the helix. This presentation somewhat resembles the concha-type microtia, the type predominant in the cohort harboring deleterious variants in *SHROOM3*. The microtia was sometimes more striking on one side. The average auricular length in heterozygotes tended to be slightly smaller in the younger stage embryos, although the differences were not statistically significant. Again, in a side-by-side comparison with wild-type littermates, the upper aspect of the auricular helix in heterozygotes appeared subtly hypoplastic in some embryos (Fig. 5A, cf. left and middle images), a feature not captured by the length measurements. Even so, heterozygotes showed as much asymmetry in auricular size as homozygotes (absolute asymmetry: $P = 0.0146$) (Fig. 5B–D), suggesting that auricular size is most sensitive to *Shroom3* gene dosage. To determine if this asymmetry extended across the entire face, the left and right side mandibular and maxillary lengths were also measured, and the ratio of the measurements was used to calculate the absolute asymmetry, the degree of asymmetry regardless of the direction of

asymmetry, present in each structure. As with ear size, both heterozygotes and homozygotes exhibited considerably more variation in asymmetry of the mandible and the maxilla compared with the littermate controls. However, as a group, the differences only reached significance for the mandibular measurements for homozygotes ($P = 0.0078$) (Fig. 5B–D). Nevertheless, the trend in asymmetry corresponded with *Shroom3* zygosity.

To validate these findings, we also downloaded and assessed the available microcomputed tomography (microCT) data for a *Shroom3* knockout (KO) mutant (*Shroom3*^{em1(IMPC)Bay/em1(IMPC)Bay}) generated by the International Mouse Phenotyping Consortium (IMPC). In total, scan data for six homozygous E18.5 embryos (four male and two female) were available. Data for four male and two female age-matched control embryos were also downloaded. Initial gross examination of the 3D renderings revealed that all mutants presented with exencephaly, and five of the six had striking midline facial clefts, more severe than that seen in the *Shroom3*^{gt/gt} mutant. Notably, despite the severity of the midline clefts, some homozygotes still had at least one normally formed nostril, and in one embryo, both were normally formed. Each of these embryos, as well as the one homozygote without a midfacial cleft (Supplemental Fig. 7), exhibited apparent facial hypoplasia involving both the snout and mandible. The pinnae, however,

were all slightly dysmorphic, posteriorly rotated, and small, with one homozygote showing severe concha-type microtia on one side (Fig 5E, yellow arrow). Most homozygotes also presented with variable, and often asymmetric, overt ocular phenotypes, including exophthalmia. Similar to the *Shroom3^{st/st}* embryos, all E18.5 *Shroom3^{em1(IMPC)Bay/em1(IMPC)Bay}* homozygotes were slightly smaller than controls.

We then performed a quantitative assessment using the same landmarks employed on the younger *Shroom3^{st/st}* embryos. To adjust for the size difference between mutants and controls at this age, we normalized all facial measurements to hindlimb foot length. Following normalization, *Shroom3^{em1(IMPC)Bay/em1(IMPC)Bay}* embryos were confirmed to have hypoplastic maxillae and mandibles (each ~12% shorter than wild-type controls), with the pinnae ~18% shorter in length (see Fig 5F–H, top). The pinna helix was also slightly dysmorphic in all mutants, showing a “curled inward” appearance (Fig 5E, white arrow) unlike controls, which had a flattened appearance. Measures of asymmetry (absolute asymmetry) for the auricle, mandible, and maxillae of controls and homozygotes showed similar trends to the findings in the *Shroom3^{st/st}* mutant, with mutants showing notably increased asymmetry for all three measures (Fig 5F–H, bottom).

Our data therefore show that reduced levels of *Shroom3* (via allele loss) affects the development of maxillary, mandibular, and auricular structures with 100% penetrance in homozygotes and is associated with increased asymmetry of facial structures in both heterozygotes and homozygotes. The findings in the mouse models support the notion that changes in the total level of expression of *Shroom3* influence CFM susceptibility and therefore likely also penetrance when combined with the impact of a pathogenic coding variant.

Discussion

To date, millions of disease-associated pathogenic/likely pathogenic (P/LP) variants have been identified or predicted, such as the more than 0.31 million P/LP variants deposited in the ClinVar database (Landrum et al. 2020). However, the penetrance of P/LP variants is highly variable, particularly for autosomal dominant phenotypes (Schmidt et al. 2024). Population-based studies of 157 diseases have shown that P/LP variants were associated with a mean penetrance of only 6.9% (Forrest et al. 2022), posing challenges for genetic counseling and diagnosis (Szeri et al. 2022; Tingaud-Sequeira et al. 2022). Multiple factors affect the penetrance of P/LP variants, including genetic variants such as eQTLs regulating the expression of the *trans*-alleles, *cis* regulatory variation (*cis* eQTLs), genetic modifiers unlinked with the causative gene, polygenic risk, imprinting, epigenetic regulation, and environmental factors (Gaudet et al. 2010; Castel et al. 2018; Hsu et al. 2021; Kingdom and Wright 2022; Mao et al. 2023). The mechanisms underlying modified penetrance have been partially elucidated and include the following: (1) variation in allelic expression resulting from allelic imbalance, duplication compensation, random monoallelic expression, or tissue-specific or biased mRNA isoform usage (Borel et al. 2015; Lek et al. 2016; Dick et al. 2016; Cummings et al. 2020; Servetti et al. 2021; Einson et al. 2023; Lobanova and Zhenilo 2024) and (2) altered gene expression mediated by genetic or epigenetic modifiers, such as *cis*-acting regulatory elements, *trans*-acting transcription factors, or epigenetic modifications (Castel et al. 2018; Delaneau et al. 2019; Lee et al. 2020; Tolmacheva et al. 2020). Despite the progress in understanding the factors that influence the penetrance of P/LP

variants, the specific reasons why these variants cause disease in some individuals but not others remain largely elusive. This issue highlights the need to (1) develop more refined P/LP variant classification systems that include allelic gene expression driving penetrance (Forrest et al. 2022; Shekari et al. 2023; Schmidt et al. 2024), and (2) elucidate the mechanisms underlying the incomplete penetrance of P/LP variants (Castel et al. 2018).

In CFM patients and healthy Chinese individuals, we identified 13 predicted pathogenic variants whose penetrance was modified by eQTLs. Castel et al. (2018) proposed a “modified penetrance” model to explain how *cis*-regulatory variants (eQTLs) modified the penetrance of coding variants through regulating the allelic expression level. Compatible with this model, we found that the haplotype configuration of predicted pathogenic variants and eQTLs of *SHROOM3* was different between CFM patients and healthy Chinese. We also found an aggregation of risk alleles of *SHROOM3* eQTLs in phenotype-positive individuals, suggesting that the penetrance of the eQTLs may be further modified by additional eQTLs in CFM patients. We termed these combinatorial effects of eQTLs as combined eQTL phenotype modifiers (CePMods), which reflects their capacity to alter phenotype penetrance through complex haplotype interactions. Such findings may have significant impact in genetic counseling. Besides *SHROOM3*, pathogenic variants in other recently identified CFM risk genes, including *FOXI3*, *SF3B2*, and *ROBO1/2*, also show incomplete or variable penetrance (Supplemental Table 7), and thus, further work is needed to determine if P/LP variants in these loci are similarly modified by single eQTLs or CePMods (Timberlake et al. 2021; Quiat et al. 2022; Mao et al. 2023).

Variants in enhancers, similar to the enhancer that contains rs344131, are promising candidates as genetic modifiers. One such example is a common noncoding variant within a conserved enhancer-like sequence of the *RET* gene that modifies Hirschsprung disease risk (Emison et al. 2005). Variants have also been described that optimize the binding affinity of transcription factors to enhancers, resulting in greater penetrance and more severe polydactyly phenotypes (Lim et al. 2024). However, Barton et al. (2022) found incomplete penetrance of cystic fibrosis carrier phenotypes that cannot be explained by the “modified penetrance” model and thus requires more research to identify the functional determinants of these mechanisms. Our study also found that the “modified penetrance” model does not fully explain the *SHROOM3* association signal (Fig. 2A; Supplemental Fig. 5B) and proposed a haplotype eQTL combination model that needs further investigation in human diseases and model organisms.

SHROOM3, in addition to having previously been associated in GWAS with CFM, has also been implicated as a genetic risk factor in neural tube defects, cleft lip and palate, and chronic kidney disease (Lemay et al. 2015; Zhang et al. 2016; Chen et al. 2018; Prokop et al. 2018; Deshwar et al. 2020; Perez et al. 2023). It is tempting to speculate that different *SHROOM3* eQTLs influence, at least in part, the genetic risk of these different conditions, that is, based on the embryonic tissue(s) in which the enhancers that harbor the eQTLs are active. In support of this notion that expression levels contribute to risk, we have investigated two separate *Shroom3* mutant mouse models and report here significant facial and auricular presentations in addition to the previously described anomalies in homozygotes. Our quantitative analyses show that *Shroom3* zygosity in mice is correlated with the severity in craniofacial dysmorphism that includes reduced auricular size and decreased mandibular and maxillary length. Notably, increased facial and auricular asymmetry was found in heterozygous and

homozygous of both *Shroom3* models. Although only homozygotes showed overt microtia, presenting as concha-type microtia (i.e., small cupped ears with reduced helices), these observations are consistent with the notion that the level of *Shroom3* activity/expression is a critical factor in the presentation of microtia and, more broadly, the facial asymmetry characteristic of CFM. *In toto*, our study identifies *SHROOM3* as a likely pathogenic gene for CFM and highlights the model of modified penetrance in individuals carrying likely pathogenic rare coding variants. However, large family-based studies are required to validate the proposed inheritance pattern, and mouse models incorporating both eQTL and P/LP variants are needed to further substantiate our hypothesis.

In summary, we demonstrate that the combination of likely pathogenic variants with eQTLs can modify the penetrance of pathogenic variants. Notably, CePMods may exert damaging effects even in the absence of rare pathogenic variants when likely pathogenic eQTLs are combined. This underscores their role as key contributors to phenotypic variability and a potential mechanism for the persistence of deleterious effects.

Methods

Study cohorts

Patients with CFM were recruited from various sources. All procedures adhered to the Helsinki Declaration principles and followed approved protocols. Informed consent for biological investigations was obtained from all participants or their legal guardians.

The cohort of 2009 Chinese CFM patients was assembled from several institutions, including Peking Union Medical College Hospital, Chinese Academy of Medical Sciences Plastic Surgery Hospital, and Beijing Tongren Hospital at Capital Medical University. CFM patients were classified employing both the Nagata and OMENS systems by seasoned clinicians. Selection of participants did not involve gender or age screening; however, individuals with non-CFM-related anomalies were excluded. The project received ethical endorsements from multiple committees, including the ethics committees of Beihang University (BM20210057), Beijing Tongren Hospital, Capital Medical University (KY058), and the Plastic Surgery Hospital of the Peking Union Medical College (201707). The control group, comprising 2625 healthy Han Chinese individuals, was sourced from the SG10K pilot data set of Singapore (application no. SG10KP00003).

Analysis of whole-genome sequencing from the NIH GMKF CFM and microtia patient cohorts (phs002130.v1.p1.c1 and phs002172.v1.p1.c1, respectively), was conducted by coauthors S.R. and T.C.C. under approved access (project ID 31609). The CFM cohort contains mostly trios of varied ethnicity, with 69 affected individuals. The microtia cohort consists of 180 affected individuals from a combination of trios, duos, and singletons, all of Hispanic ancestry. The patients and unaffected relatives were collected from Ecuador, Colombia, and the United States, as previously described (Quiat et al. 2022).

High-throughput sequencing and variant calling

For the 2009 Chinese CFM patients, paired-end (2×150 bp) whole-genome sequencing was carried out on them by using the platform of T7 sequencer of BGI. Across all sequenced samples, the average depth was around $10\times$. The sequencing reads underwent analysis utilizing tools including Burrows-Wheeler aligner (v0.7.17; Li and Durbin 2009), SAMtools (v1.12; Danecek et al. 2021), Picard (v2.27.0; <https://broadinstitute.github.io/picard/>), the Genome

Analysis Toolkit (v4.1.4.1; Auwera et al. 2013), and CNVkit (0.9.11; Talevich et al. 2016). Sequenced reads were aligned to the hg38 reference human genome. For the Australasian cohort, whole-genome sequencing and initial variant calling were performed by the Children's Mercy Genomic Medicine Center as part of the Genomic Answers for Kids (GA4K) program, with subsequent analyses and filtering performed by the Cox laboratory. For the GMKF CFM and microtia cohorts, VCF files of whole-genome sequencing were directly obtained and all *SHROOM3* locus variants extracted for analysis by the Cox laboratory under prior project approval.

Association study

The case group comprised 2009 Chinese individuals, whereas the control group consisted of 2625 healthy Han Chinese from the pilot data of SG10K project, with Han Chinese ancestry confirmed through PCA analysis. Quality-control parameters were set as follows: sample call rate $>98\%$, SNP call rate $>95\%$, a Hardy-Weinberg equilibrium threshold of 0.001 (determined using Fisher's exact test) in the control cohort, and a minor allele frequency (MAF) >0.001 in either the case or control cohort; 4187 variants were finally utilized for subsequent analysis of association study using PLINK (Purcell et al. 2007). Single-variant association tests were conducted. The threshold for genome-wide significance was established at a P -value $< 6 \times 10^{-8}$. Variants with $MAF > 0.1$ were retained for LD analysis.

SKAT burden test

We performed a SNP-set kernel association test using the SKAT package v2.2.5 (Wu et al. 2011) to analyze deleterious variants in the *SHROOM3* gene in 2009 Chinese CFM patients and 2625 healthy Han Chinese samples from the SG10K data set.

Fine mapping

For fine-mapping causal variants, we employed FINEMAP v1.4.2, CAVIAR v2.2, PAINTOR v3.0, and PLINK's conditional analysis function, utilizing default parameters for the first three methods. The conditional analysis in PLINK followed a modified approach from our previous work (Zhang et al. 2016). Specifically, after establishing a final list of associated variants, we iteratively used the lead SNP in each block as a covariate for conditional analysis on the remaining variants within that block. This iterative process continued until no additional independent associated variants were detected in each block. Notably, we did not consider the presence of independent associated variants in other blocks during each block's conditional analysis. This approach accounts for the low LD among variants across different blocks and their potential interactions, which could influence the outcomes of the association tests.

LD block analysis

To delineate LD blocks, we incorporated variants with a MAF greater than 0.1 that showed significant association. We computed the pairwise r^2 values among these variants and visualized the LD blocks using an R package (R Core Team 2024).

Annotating common variants

We annotated all significantly associated variants using HaploReg V4.2 (Ward and Kellis 2012), employing default parameters with the exception of LD calculation, which was adjusted according to the respective populations. Additionally, we utilized HaploReg to annotate deleterious-associated variants to identify those

potentially affecting *SHROOM3* expression. Variants situated in enhancer or promoter regions, as indicated by histone or DNase marks in stem cells, and those annotated as eQTLs were classified as expression-affecting variants.

For *SHROOM3*, we compiled all eQTLs across various cell types and tissues as cataloged in the GTEx database. In instances in which an eQTL exhibited both positive and negative normalized effect sizes (NESs) across different cell types and tissues, we selected the extreme value from the predominant trend to represent its NES. In all other cases, we simply chose the largest NES value.

Identification of likely pathogenic rare variants

A total of 4187 variants underwent annotation using ANNOVAR (Wang et al. 2010) in the 2009 Chinese CFM. We subjected these variants, along with those found in the 2625 healthy Chinese samples, to the following filtering process: (1) to retain LoF variants, we exclusively focused on missense, stop-gain, stop-loss, splicing, and frameshift variants; (2) to predict deleterious variants, employing three predictive criteria—PolyPhen-2, SIFT, CADD scores (>15)—we retained variants predicted to be deleterious; and (3) to filter variants based on MAF in the public databases, for high-frequency variants with a $MAF > 0.005$ in any subpopulations in ExAC, gnomAD, or the 1000 Genomes Project, we retained those with a minimum of two deleterious predictions, alongside adherence to stringent conservation benchmarks (GERP score > 2 , phyloP100way_vertibrate > 1.6 , and phastCons100way_vertibrate score > 0.5). Regarding rare variants with $MAF < 0.005$ in all subpopulations of ExAC, gnomAD, SG10K, and the 1000 Genomes Project, retention was dependent on being predicted as deleterious by at least one prediction algorithm. Those rare variants that passed all the above three criteria were considered as deleterious variants.

To ascertain potentially pathogenic variants, we utilized two advanced artificial intelligence methods: ESM1b (Brandes et al. 2023) and AlphaMissense (Cheng et al. 2023). ESM1b, a sophisticated protein language model, assesses the potential harm of variants by evaluating the implications of missense variants on a genome-wide scale. AlphaMissense further refines this prediction by integrating structural analysis and evolutionary conservation to determine functional impact and disease association. For this investigation, variants located within recognized functional domains of *SHROOM3*, identified as pathogenic by AlphaMissense, or deemed ambiguous by AlphaMissense and with an ESM1b score below -10 were classified as likely pathogenic variants.

We utilized LOFTEE to identify HC LoF variants. We consider LoFs to result in a truncated nonfunctional protein unless they trigger nonsense-mediated mRNA decay (NMD). To determine whether LoFs escape NMD, we employed NMDetective with a cutoff of 0.52 and applied the rules proposed by Lindeboom et al. (2019) and Karczewski et al. (2020).

The Missense score was calculated using a modified approach based on the methodology outlined by Jurgens et al. (2022). Missense variants were annotated using the variant effect predictor (VEP; v112.0). These tools were classified into two categories. First, the qualitative prediction algorithm category included AlphaMissense (Cheng et al. 2023), ESM1b (Brandes et al. 2023), SIFT (Vaser et al. 2016), PolyPhen-2 HDIV (Adzhubei et al. 2010), PolyPhen-2 HVAR (Adzhubei et al. 2010), LRT (Chun and Fay 2009), MutationTaster (Schwarz et al. 2014), FATHMM (Rogers et al. 2018), PROVEAN (Choi and Chan 2015), MetaSVM (Kim et al. 2017), MetaLR (Chen et al. 2023), MCAP (Jagadeesh et al. 2016), fathmm-MKL (Shihab et al. 2015), MutationAssessor (Reva et al. 2011), and ALoFT (Balasubramanian et al. 2017). Second, the quantitative prediction algorithm group consisted of

tools such as CADD (Rentzsch et al. 2019), VEST3 (Carter et al. 2013), DANN (Quang et al. 2015), GERP (Huber et al. 2020), phyloP100way (Pollard et al. 2010), phyloP20way (Pollard et al. 2010), phastCons100way (Siepel et al. 2005), phastCons20way (Siepel et al. 2005), and SiPhy (Garber et al. 2009). For qualitative algorithms, variants classified as deleterious (“D”) by a given tool were assigned one point. For quantitative algorithms, variants with scores exceeding the 90th percentile of all predicted variants in the data set were also awarded one point per algorithm. The proportion of the deleterious score was then calculated as the ratio of the total accumulated scores to the maximum possible scores across all algorithms, further adjusted by incorporating functional domain information and ensuring that variants located in critical functional domains were given appropriate weight during the scoring process. Additionally, HC LoF variants annotated by LOFTEE were directly assigned a deleterious score of one, bypassing the scoring criteria used for missense variants.

Haplotype prediction and combination analysis

We first used SHAPEIT5 (v5.1.0; Hofmeister et al. 2023) to construct the haplotype of 4187 variants identified in *SHROOM3*. Common variants with a MAF greater than 0.005 were phased using the phase_common tool, whereas for rare variants (those with MAF less than 0.005), we utilized the phase_rare tool to phase these variants onto a scaffold of common variants (Hofmeister et al. 2023). Subsequently, we extracted predicted pathogenic coding variants and eQTL sites from the phased haplotype data and combined them to analyze the expression levels of each eQTL allele on the *SHROOM3* gene in conjunction with the alternative alleles of predicted pathogenic variants, where $a(g)$ represents the genotype combination of four types of pathogenic variants with different eQTL loci:

$$a(g) = \begin{cases} i, & g \in (\text{High_cis}/\text{High_trans}) \\ ii, & g \in (\text{High_cis}/\text{Low_trans}) \\ iii, & g \in (\text{Low_cis}/\text{High_trans}) \\ iv, & g \in (\text{Low_cis}/\text{Low_trans}) \end{cases}$$

High-contribution CFM samples to *SHROOM3* signal

To elucidate the contributors to the *SHROOM3* association signal, we implemented a three-step methodology. First, we generated a reference list of CFM-associated variants that included those variants with an initial P -value less than 5×10^{-8} (chi-square test in PLINK) in the association test. Subsequently, we analyzed the directional contribution of each case to the variant significance in the reference list. This step involved assessing the change in significance level for all variants in the reference list when each patient was individually excluded. We tracked the change in the significance level of all variants for each patient removal. To quantify each patient’s contribution, we employed an indicator function $\beta(i,j)$, where β equals one for a positive contribution and -1 for no contribution. In this function, “ i ” represents CFM patient samples, “ j ” denotes the variants in the reference list, and $p(i,j)$ indicates the P -value of variant “ j ” after excluding patient “ i ”, with $\beta(i,j)$ defined as

$$\beta(i,j) = \begin{cases} 1, & p_{(i,j)} > p_{\text{initial}} \\ -1, & p_{(i,j)} < p_{\text{initial}} \end{cases}$$

The final step involved identifying contributors to the *SHROOM3* signal. We calculated the $\beta(i)$ for each patient and selected the top 5% of samples as potential contributors. The contribution score for each sample at the *SHROOM3* loci, denoted as ϵ , was calculated as the sum of $\beta(i)$ across all patients (N), where N is the total

number of CFM patients:

$$\varepsilon = \sum_{i=1}^N \beta(i).$$

We repeated these three steps for three cycles until the *SHROOM3* association signal was no longer detectable. Samples consistently identified in this process were classified as major contributors to the association signal.

Luciferase assay

We identified 84 associated variants as *SHROOM3* eQTLs, as annotated in the GTEx database. Additionally, 30 eQTLs were mapped to regions marked by promoter or enhancer histone modifications. Based on promoter capture Hi-C analyses, 10 of these variants are associated with *SHROOM3* promoter activity in hCNCCs. Among these 10 eQTLs, five exhibit strong LD with rs10017322 or rs344131 and were excluded from further luciferase experiments. Finally, we selected rs10017322, rs344131, rs74918804, rs189707263, and rs344126 for luciferase reporter assays.

The HEK-293T cell line (1101HUM-PUMC000091) was obtained from the Cell Resource Center, Peking Union Medical College, and maintained in DMEM (Gibco) supplemented with 10% FBS (Gibco) in a humidified incubator at 37°C with 5% CO₂. The sequence of the *SHROOM3* enhancer was cloned into a luciferase vector (Genechem GV238), and alternative alleles of rs10017322, rs344131, and rs61090632 were generated using the Q5 site-directed mutagenesis kit (New England BioLabs E0554). For luciferase assays, HEK-293T cells were transfected with *SHROOM3* enhancer luciferase vectors and pRL-TK vectors using the Lipofectamine 3000 transfection reagent (Thermo Fisher L3000008). After 24 h, cells were harvested and lysed, and fluorescence intensity was measured using the dual-luciferase assay kit (TransGen FR201) following the manufacturer's instructions on CLARIOstar microplate reader (BMG Labtech).

Epigenome data

The hCNCC PChIP data utilized in this study are from Xu et al. (2024). hCNCC ChIP-seq data sets are available at the NCBI Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) under the following accession numbers: EP300 (GSM714804), H3K27ac (GSM714807), H3K4me1 (GSM714808), and H3K4me3 (GSM714809) (Rada-Iglesias et al. 2012). All Epigenome data were visualized using the WashU Epigenome Browser (<https://epigenomegateway.wustl.edu/browser/>).

Mouse model

Two *Shroom3* mouse models were assessed. *Shroom3* gene trap (*Shroom3*^{+/*gt*} and *Shroom3*^{gt/*gt*}) mice were obtained from the Jackson Laboratories, and *Shroom3* KO (*Shroom3*^{em1(IMPC)Bay/em1(IMPC)Bay}) data sets were available through the IMPC website (<https://www.mousephenotype.org>). The *Shroom3*^{gt/*gt*} mice were initially identified and described by Hildebrand and Soriano (1999) with homozygotes fully penetrant for exencephaly, with secondary palatal clefting, spina bifida, cardiac anomalies, kidney and thyroid defects, as well as with ocular abnormalities all showing variable degrees of penetrance (Chung et al. 2010; Plageman et al. 2010, 2011; Lang et al. 2014; Durbin et al. 2020). Heterozygotes have not been reported as exhibiting an overt phenotype. IMPC summary data for the *Shroom3*^{em1(IMPC)Bay/em1(IMPC)Bay} mice indicate highly penetrant exencephaly, midfacial clefting, and variable spina bifida and tail kinking, among other features (<https://www.mousephenotype.org/data>).

OPT and quantitative assessment of embryonic craniofacial tissue

Shroom3^{+/*gt*}, *Shroom3*^{gt/*gt*}, and *Shroom3*^{+/*+*} littermate controls were collected at approximately embryonic day 14.5 (E14.5) and fixed overnight in 4% paraformaldehyde. Embryos were then rinsed briefly in phosphate buffered saline and embedded in 1.1% low-melting-point agarose in Milli-Q water. Embedded embryos were dehydrated in 100% methanol for 3 days, changing the methanol each day, and then cleared using 1:2, benzyl alcohol:benzyl benzoate (BABB) for 3 days with daily changes of BABB. Samples were then imaged in fresh BABB using a Bioptonic 3001M OPT (Bruker/Skyscan). Imaging was performed at standard 512 × 512 resolution under UV light (425/40 nm excitation; 475 nm emission) with zooming to an image pixel size of ~25 microns, 0.9° rotation step, and 360° rotation. Raw imaging data were then reconstructed into multiplanar slices (.bmp) with Nrecon V1.7.4.2, using Gaussian smoothing (smoothing set at one), and then imported into Drishti V3.0 volume exploration software (Limaye 2012; <https://github.com/nci/Drishti>) for qualitative and quantitative assessment of virtual 3D renderings of each scanned sample. Consistent scan, reconstruction, and rendering settings were used for each specimen.

Nine 3D landmark coordinates were collected from each embryo (see Supplemental Fig. 7). These landmarks enabled proxy measurements of the following: left and right-side maxillary length (distal nasal tip to postorbital genal whisker follicle), pinna height (rostral to caudal aspect of auricle), head width (left to right side caudal aspect of auricle), and mandibular length (distal tip of mandible to caudal aspect of auricle). Visual assessment of the different harvested litters suggested subtle differences in embryonic age between litters: E14.25 to E15.25. Therefore, for all measurements except assessments of asymmetry, the litter stage was considered. To assess asymmetry across the embryonic face, the following ratios were calculated: mandibular asymmetry (left vs. right mandibular length), maxillary asymmetry (left vs. right maxillary length) and, head width to auricular length ratio (left and right sides). There was no statistical bias in the sidedness of any asymmetry, and so, data were reported as absolute asymmetry: Abs(1 – left/right). All measurements were graphed using SuperPlotsOfData and were statistically analyzed using a Welch's *t*-test within this program (Goedhart 2021). Differences were considered significant for comparisons yielding *P* < 0.05.

Quantitative assessment of microCT data

High-resolution microCT scan data (NRRD file format) of iodine contrast-stained E18.5 *Shroom3* KO and control embryos were downloaded from the IMPC website. The data were then imported into Drishti V3.0 as described above for qualitative and quantitative assessment. The same landmark-based analysis was then performed as described for the *Shroom3*^{gt/*gt*} analyses, with the exception that size normalization between the control and mutant embryos was performed using left hindlimb foot length measurements, which were also made on the rendered scans.

Data access

All raw and processed sequencing data generated in this study have been submitted to the National Genomics Data Center (GSA; <https://ngdc.cnbc.ac.cn/gsa-human/>) under accession numbers HRA005132, HRA004333, HRA003925, and HRA003924. Raw, reconstructed, and rendered OPT scan data for all mouse embryo analyses are available in figshare (<https://doi.org/10.6084/m9.figshare.28440413.v1>), and microCT data sets are available through the IMPC website with IDs 157760 (wild type) and

612303 (homozygous) (<https://www.mousephenotype.org/embryoviewer/?mgi=MGI%3A1351655>). The source data for this paper are available as Supplemental Source Data.

Competing interest statement

S.E.A. is a cofounder and CEO of Medigenome and serves on the scientific advisory board of the “Imaging” institute in Paris. The remaining authors declare no competing interests.

Acknowledgments

We thank the patients and their families for their participation in the study. The study was partially supported by the National Natural Science Foundation of China (32470644 and 82171844 to Y.-B.Z.), the Beijing Natural Science Foundation Program and Scientific Research Key Program of Beijing Municipal Commission of Education (KZ202010025039), the Child Care foundation and a European Research Council grant 249968 to S.E.A., and both a Stowers Family Endowment and BioNexusKC Patton Trust Research Grant to T.C.C. The RDNOW program acknowledges financial support from the Royal Children’s Hospital Foundation, the Murdoch Children’s Research Institute, and the Harbig Foundation. The research conducted at the Murdoch Children’s Research Institute was supported by the Victorian Government’s Operational Infrastructure Support Program.

Author contributions: Y.-B.Z., T.C.C., and S.E.A. designed and supervised the study. J.Z., Q.Z., S.Z., T.Y.T., J.M., C.Z., H.Z., and P.C. recruited case samples. H.Z., R.L., Z.S., J.L., S.R., and T.C.C. planned and conducted laboratory experiments. M.D.D. and S.M.W. generated the *Shroom3* and control embryos. L.L.C. and T.C.C. performed the qualitative and quantitative assessment of the *Shroom3* and control embryos. H.Z., B.X., H.S., Z.M., T.L., X.T., and J.W. analyzed data. H.Z., S.E.A., Y.-B.Z., and T.C.C. drafted and revised the manuscript. All authors have reviewed and contributed to the manuscript.

References

- Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. *Nat Methods* **7**: 248–249. doi:10.1038/nmeth0410-248
- Auweru GAVD, Carneiro MO, Hartl C, Poplin R, Angel Gd, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. 2013. From FastQ data to high-confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinform* **43**: 11.10.11–11.10.33. doi:10.1002/0471250953.bi1110843
- Balasubramanian S, Fu Y, Pawashe M, McGillivray P, Jin M, Liu J, Karczewski KJ, MacArthur DG, Gerstein M. 2017. Using ALoFT to determine the impact of putative loss-of-function variants in protein-coding genes. *Nat Commun* **8**: 382. doi:10.1038/s41467-017-00443-5
- Barton AR, Hujoel MLA, Mukamel RE, Sherman MA, Loh P-R. 2022. A spectrum of recessiveness among Mendelian disease variants in UK Biobank. *Am J Hum Genet* **109**: 1298–1307. doi:10.1016/j.ajhg.2022.05.008
- Benner C, Spencer CCA, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. 2016. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**: 1493–1501. doi:10.1093/bioinformatics/btw018
- Borel C, Ferreira Pedro G, Santoni F, Delaneau O, Fort A, Popadin Konstantin Y, Garieri M, Falconnet E, Ribaux P, Guipponi M, et al. 2015. Biased allelic expression in human primary fibroblast single cells. *Am J Hum Genet* **96**: 70–80. doi:10.1016/j.ajhg.2014.12.001
- Brandes N, Goldman G, Wang CH, Ye CJ, Ntranos V. 2023. Genome-wide prediction of disease variant effects with a deep protein language model. *Nat Genet* **55**: 1512–1522. doi:10.1038/s41588-023-01465-0
- Carter H, Douville C, Stenson PD, Cooper DN, Karchin R. 2013. Identifying Mendelian disease genes with the variant effect scoring tool. *BMC Genom* **14**: S3. doi:10.1186/1471-2164-14-S3-S3
- Castel SE, Cervera A, Mohammadi P, Aguet F, Reverter F, Wolman A, Guigo R, Iossifov I, Vasileva A, Lappalainen T. 2018. Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk. *Nat Genet* **50**: 1327–1334. doi:10.1038/s41588-018-0192-y
- Chen Z, Kuang L, Finnell RH, Wang H. 2018. Genetic and functional analysis of *SHROOM1-4* in a Chinese neural tube defect cohort. *Hum Genet* **137**: 195–202. doi:10.1007/s00439-017-1864-x
- Chen Y, Liu L, Li J, Jiang H, Ding C, Zhou Z. 2023. MetaLR: Meta-tuning of Learning Rates for transfer learning in medical imaging. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, pp. 706–716. doi:10.1007/978-3-031-43907-0_67
- Cheng J, Novati G, Pan J, Bycroft C, Žemgulytė A, Applebaum T, Pritzel A, Wong LH, Zielinski M, Sargeant T, et al. 2023. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* **381**: eadg7492. doi:10.1126/science.adg7492
- Choi Y, Chan AP. 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**: 2745–2747. doi:10.1093/bioinformatics/btv195
- Chun S, Fay JC. 2009. Identification of deleterious mutations within three human genomes. *Genome Res* **19**: 1553–1561. doi:10.1101/gr.092619.109
- Chung M-I, Nascone-Yoder NM, Grover SA, Drysdale TA, Wallingford JB. 2010. Direct activation of *Shroom3* transcription by Pitx proteins drives epithelial morphogenesis in the developing gut. *Development* **137**: 1339–1349. doi:10.1242/dev.044610
- Cong P-K, Bai W-Y, Li J-C, Yang M-Y, Khederzadeh S, Gai S-R, Li N, Liu Y-H, Yu S-H, Zhao W-W, et al. 2022. Genomic analyses of 10,376 individuals in the Westlake BioBank for Chinese (WBBC) pilot project. *Nat Commun* **13**: 2939. doi:10.1038/s41467-022-30526-x
- Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. 2013. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum Genet* **132**: 1077–1130. doi:10.1007/s00439-013-1331-2
- Cummings BB, Karczewski KJ, Kosmicki JA, Seaby EG, Watts NA, Singer-Berk M, Mudge JM, Karjalainen J, Satterstrom FK, O’Donnell-Luria AH, et al. 2020. Transcript expression-aware annotation improves rare variant interpretation. *Nature* **581**: 452–458. doi:10.1038/s41586-020-2329-2
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* **10**: giab008. doi:10.1093/gigascience/giab008
- Delaneau O, Zayhtyska M, Borel C, Giannuzzi G, Rey G, Howald C, Kumar S, Ongen H, Popadin K, Marbach D, et al. 2019. Chromatin three-dimensional interactions mediate genetic effects on gene expression. *Science* **364**: eaat8266. doi:10.1126/science.aat8266
- Deshwar AR, Martin N, Shannon P, Chitayat D. 2020. A homozygous pathogenic variant in *SHROOM3* associated with anencephaly and cleft lip and palate. *Clin Genet* **98**: 299–302. doi:10.1111/cge.13804
- Dick IE, Joshi-Mukherjee R, Yang W, Yue DT. 2016. Arrhythmogenesis in Timothy Syndrome is associated with defects in Ca²⁺-dependent inactivation. *Nat Commun* **7**: 10370. doi:10.1038/ncomms10370
- Du Z, Ma L, Qu H, Chen W, Zhang B, Lu X, Zhai W, Sheng X, Sun Y, Li W, et al. 2019. Whole genome analyses of Chinese population and de novo assembly of a northern Han genome. *Genomics Proteomics Bioinformatics* **17**: 229–247. doi:10.1016/j.gpb.2019.07.002
- Durbin MD, O’Kane J, Lorentz S, Firulli AB, Ware SM. 2020. SHROOM3 is downstream of the planar cell polarity pathway and loss-of-function results in congenital heart defects. *Dev Biol* **464**: 124–136. doi:10.1016/j.ydbio.2020.05.013
- Einson J, Glinos D, Boerwinkle E, Castaldi P, Darbar D, Andrade MD, Ellinor P, Fornage M, Gabriel S, Germer S, et al. 2023. Genetic control of mRNA splicing as a potential mechanism for incomplete penetrance of rare coding variants. *Genetics* **224**: iyad115. doi:10.1093/genetics/iyad115
- Emison ES, McCallion AS, Kashuk CS, Bush RT, Grice E, Lin S, Portnoy ME, Cutler DJ, Green ED, Chakravarti A. 2005. A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. *Nature* **434**: 857–863. doi:10.1038/nature03467
- Forrest IS, Chaudhary K, Vy HMT, Petrazzini BO, Bafna S, Jordan DM, Rocheleau G, Loos RJE, Nadkarni GN, Cho JH, et al. 2022. Population-based penetrance of deleterious clinical variants. *JAMA* **327**: 350–359. doi:10.1001/jama.2021.23686
- Garber M, Guttman M, Clamp M, Zody MC, Friedman N, Xie X. 2009. Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics* **25**: i54–i62. doi:10.1093/bioinformatics/btp190
- Gaudet MM, Kirchoff T, Green T, Vijai J, Korn JM, Guiducci C, Segrè AV, McGee K, McGuffog L, Kartsonaki C, et al. 2010. Common genetic variants and modification of penetrance of BRCA2-associated breast cancer. *PLoS Genet* **6**: e1001183. doi:10.1371/journal.pgen.1001183

- Goedhart J. 2021. SuperPlotsOfData—a web app for the transparent display and quantitative comparison of continuous data from different conditions. *Mol Biol Cell* **32**: 470–474. doi:10.1091/mbc.E20-09-0583
- The GTEx Consortium. 2020. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**: 1318–1330. doi:10.1126/science.aaz1776
- Hildebrand JD, Soriano P. 1999. Shroom, a PDZ domain-containing actin-binding protein, is required for neural tube morphogenesis in mice. *Cell* **99**: 485–497. doi:10.1016/S0092-8674(00)81537-8
- Hofmeister RJ, Ribeiro DM, Rubinacci S, Delaneau O. 2023. Accurate rare variant phasing of whole-genome and whole-exome sequencing data in the UK Biobank. *Nat Genet* **55**: 1243–1249. doi:10.1038/s41588-023-01415-w
- Hormozdiari F, Zhu A, Kichaev G, Ju CJT, Segrè AV, Joo JWJ, Won H, Sankararaman S, Pasaniuc B, Shifman S, et al. 2017. Widespread allelic heterogeneity in complex traits. *Am J Hum Genet* **100**: 789–802. doi:10.1016/j.ajhg.2017.04.005
- Hsu F-C, Roberts NJ, Childs E, Porter N, Rabe KG, Borgida A, Ukaegbu C, Goggins MG, Hruban RH, Zogopoulos G, et al. 2021. Risk of pancreatic cancer among individuals with pathogenic variants in the *ATM* gene. *JAMA Oncol* **7**: 1664–1668. doi:10.1001/jamaoncol.2021.3701
- Huber CD, Kim BY, Lohmueller KE. 2020. Population genetic models of GERP scores suggest pervasive turnover of constrained sites across mammalian evolution. *PLoS Genet* **16**: e1008827. doi:10.1371/journal.pgen.1008827
- Jagadeesh KA, Wenger AM, Berger MJ, Guturu H, Stenson PD, Cooper DN, Bernstein JA, Bejerano G. 2016. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat Genet* **48**: 1581–1586. doi:10.1038/ng.3703
- Jurgens SJ, Choi SH, Morrill VN, Chaffin M, Pirruccello JP, Halford JL, Weng LC, Nauffal V, Roselli C, Hall AW, et al. 2022. Analysis of rare genetic variation underlying cardiometabolic diseases and traits among 200,000 individuals in the UK Biobank. *Nat Genet* **54**: 240–250. doi:10.1038/s41588-021-01011-w
- Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, et al. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**: 434–443. doi:10.1038/s41586-020-2308-7
- Kichaev G, Yang W-Y, Lindstrom S, Hormozdiari F, Eskin E, Price AL, Kraft P, Pasaniuc B. 2014. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet* **10**: e1004722. doi:10.1371/journal.pgen.1004722
- Kim S, Jhong J-H, Lee J, Koo J-Y. 2017. Meta-analytic support vector machine for integrating multiple omics data. *BioData Min* **10**: 2. doi:10.1186/s13040-017-0126-8
- Kingdom R, Wright CF. 2022. Incomplete penetrance and variable expressivity: from clinical studies to population cohorts. *Front Genet* **13**: 920390. doi:10.3389/fgene.2022.920390
- Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, Hart J, Hoffman D, Jang W, Kaur K, Liu C, et al. 2020. ClinVar: improvements to accessing data. *Nucleic Acids Res* **48**: D835–D844. doi:10.1093/nar/gkz972
- Lang RA, Herman K, Reynolds AB, Hildebrand JD, Plageman TF. 2014. p120-catenin-dependent junctional recruitment of Shroom3 is required for apical constriction during lens pit morphogenesis. *Development* **141**: 3177–3187. doi:10.1242/dev.107433
- Lawlor A, Cunanan K, Cunanan J, Paul A, Khalili H, Ko D, Khan A, Gros R, Drysdale T, Bridgewater D. 2023. Minimal kidney disease phenotype in *Shroom3* heterozygous null mice. *Can J Kidney Health Dis* **10**: 20543581231165716. doi:10.1177/20543581231165716
- Lee RVD, Correard S, Wasserman WW. 2020. Deregulated regulators: disease-causing *cis* variants in transcription factor genes. *Trends Genet* **36**: 523–539. doi:10.1016/j.tig.2020.04.006
- Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**: 285–291. doi:10.1038/nature19057
- Lemay P, Guyot M-C, Tremblay É, Dionne-Laporte A, Spiegelman D, Henrion É, Diallo O, Marco PD, Merello E, Massicotte C, et al. 2015. Loss-of-function de novo mutations play an important role in severe human neural tube defects. *J Méd Genet* **52**: 493–497. doi:10.1136/jmedgenet-2015-103027
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**: 1754–1760. doi:10.1093/bioinformatics/btp324
- Lim F, Solvason JJ, Ryan GE, Le SH, Jindal GA, Steffen P, Jandu SK, Farley EK. 2024. Affinity-optimizing enhancer variants disrupt development. *Nature* **626**: 151–159. doi:10.1038/s41586-023-06922-8
- Limaye A. 2012. Drishti: a volume exploration and presentation tool. In *Dev X-Ray Tomogr VIII*, Vol. 8506, p. 85060X. Society of Photo-Optical Instrumentation Engineers, San Diego. doi:10.1117/12.935640
- Lindeboom RGH, Vermeulen M, Lehner B, Supek F. 2019. The impact of nonsense-mediated mRNA decay on genetic disease, gene editing and cancer immunotherapy. *Nat Genet* **51**: 1645–1651. doi:10.1038/s41588-019-0517-5
- Lobanova YV, Zhenilo SV. 2024. Genomic imprinting and random mono-allelic expression. *Biochemistry (Mosc)* **89**: 84–96. doi:10.1134/S000629792401005X
- Mao K, Borel C, Ansar M, Jolly A, Makrythanasis P, Froehlich C, Iwaszkiewicz J, Wang B, Xu X, Li Q, et al. 2023. *FOXP3* pathogenic variants cause one form of craniofacial microsomia. *Nat Commun* **14**: 2026. doi:10.1038/s41467-023-37703-6
- Martin JB, Muccioli M, Herman K, Finnell RH, Plageman TF. 2019. Folic acid modifies the shape of epithelial cells during morphogenesis via a Fcrl1 and MLCK dependent mechanism. *Biol Open* **8**: bio041160. doi:10.1242/bio.041160
- McGreevy EM, Vijayraghavan D, Davidson LA, Hildebrand JD. 2015. Shroom3 functions downstream of planar cell polarity to regulate myosin II distribution and cellular organization during neural tube closure. *Biol Open* **4**: 186–196. doi:10.1242/bio.20149589
- Perez KKD, Chung S, Head ST, Epstein MP, Hecht JT, Wehby GL, Weinberg SM, Murray JC, Marazita ML, Leslie EJ. 2023. Rare variants found in multiplex families with orofacial clefts: Does expanding the phenotype make a difference? *Am J Med Genet A* **191**: 2558–2570. doi:10.1002/ajmg.a.63336
- Plageman TF, Chung M-I, Lou M, Smith AN, Hildebrand JD, Wallingford JB, Lang RA. 2010. Pax6-dependent *Shroom3* expression regulates apical constriction during lens placode invagination. *Development* **137**: 405–415. doi:10.1242/dev.045369
- Plageman TF, Zacharias AL, Gage PJ, Lang RA. 2011. Shroom3 and a Pitx2-N-cadherin pathway function cooperatively to generate asymmetric cell shape changes during gut morphogenesis. *Dev Biol* **357**: 227–234. doi:10.1016/j.ydbio.2011.06.027
- Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. 2010. Detection of non-neutral substitution rates on mammalian phylogenies. *Genome Res* **20**: 110–121. doi:10.1101/gr.097857.109
- Prokop JW, Yeo NC, Ottmann C, Chhetri SB, Florus KL, Ross EJ, Sosonkina N, Link BA, Freedman BI, Coppola CJ, et al. 2018. Characterization of coding/noncoding variants for *SHROOM3* in patients with CKD. *J Am Soc Nephrol* **29**: 1525–1535. doi:10.1681/ASN.2017080856
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**: 559–575. doi:10.1086/519795
- Quang D, Chen Y, Xie X. 2015. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics* **31**: 761–763. doi:10.1093/bioinformatics/btu703
- Quiat D, Kim SW, Zhang Q, Morton SU, Pereira AC, DePalma SR, Willcox JAL, McDonough B, DeLaughter DM, Gorham JM, et al. 2022. An ancient founder mutation located between *ROBO1* and *ROBO2* is responsible for increased microtia risk in Amerindigenous populations. *Proc Natl Acad Sci* **119**: e2203928119. doi:10.1073/pnas.2203928119
- Rada-Iglesias A, Bajpai R, Prescott S, Bruggmann SA, Swigut T, Wysocka J. 2012. Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell* **11**: 633–648. doi:10.1016/j.stem.2012.07.006
- Rahit KMT, Tarailo-Graovac M. 2020. Genetic modifiers and rare Mendelian disease. *Genes (Basel)* **11**: 239. doi:10.3390/genes11030239
- R Core Team. 2024. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>.
- Rentsch P, Witten D, Cooper GM, Shendure J, Kircher M. 2019. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* **47**: D886–D894. doi:10.1093/nar/gky1016
- Reva B, Antipin Y, Sander C. 2011. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res* **39**: e118. doi:10.1093/nar/gkr407
- Rogers MF, Shihab HA, Mort M, Cooper DN, Gaunt TR, Campbell C. 2018. FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics* **34**: 511–513. doi:10.1093/bioinformatics/btx536
- Schmidt RJ, Steeves M, Bayrak-Toydemir P, Benson KA, Coe BP, Conlin LK, Ganapathi M, Garcia J, Gollob MH, Jobanputra V, et al. 2024. Recommendations for risk allele evidence curation, classification, and reporting from the ClinGen low penetrance/risk allele working group. *Genet Med* **26**: 101036. doi:10.1016/j.gim.2023.101036
- Schwarz JM, Cooper DN, Schuelke M, Seelow D. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* **11**: 361–362. doi:10.1038/nmeth.2890
- Servetti M, Pisciotto L, Tassano E, Cerminara M, Nobili L, Boeri S, Rosti G, Lerone M, Divizia MT, Ronchetto P, et al. 2021. Neurodevelopmental disorders in patients with complex phenotypes and potential complex

- genetic basis involving non-coding genes, and double CNVs. *Front Genet* **12**: 732002. doi:10.3389/fgene.2021.732002
- Shekari S, Stankovic S, Gardner EJ, Hawkes G, Kentistou KA, Beaumont RN, Mörseburg A, Wood AR, Prague JK, Mishra GD, et al. 2023. Penetrance of pathogenic genetic variants associated with premature ovarian insufficiency. *Nat Med* **29**: 1692–1699. doi:10.1038/s41591-023-02405-5
- Shihab HA, Rogers MF, Gough J, Mort M, Cooper DN, Day INM, Gaunt TR, Campbell C. 2015. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics* **31**: 1536–1543. doi:10.1093/bioinformatics/btv009
- Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**: 1034–1050. doi:10.1101/gr.3715005
- Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G, Fritz MH-Y, et al. 2015. An integrated map of structural variation in 2,504 human genomes. *Nature* **526**: 75–81. doi:10.1038/nature15394
- Szeri F, Miko A, Navasiolava N, Kaposi A, Verschuere S, Molnar B, Li Q, Terry SF, Boraldi F, Uitto J, et al. 2022. The pathogenic c.1171A>G (p.Arg391Gly) and c.2359G>A (p.Val787Ile) ABCC6 variants display incomplete penetrance causing pseudoxanthoma elasticum in a subset of individuals. *Hum Mutat* **43**: 1872–1881. doi:10.1002/humu.24498
- Talevich E, Shain AH, Botton T, Bastian BC. 2016. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput Biol* **12**: e1004873. doi:10.1371/journal.pcbi.1004873
- Timberlake AT, Griffin C, Heike CL, Hing AV, Cunningham ML, Chitayat D, Davis MR, Doust SJ, Drake AF, Duenas-Roque MM, et al. 2021. Haploinsufficiency of *SF3B2* causes craniofacial microsomia. *Nat Commun* **12**: 4680. doi:10.1038/s41467-021-24852-9
- Tingaud-Sequeira A, Trimouille A, Sagardoy T, Lacombe D, Rooryck C. 2022. Oculo-auriculo-vertebral spectrum: new genes and literature review on a complex disease. *J Med Genet* **59**: 417–427. doi:10.1136/jmedgenet-2021-108219
- Tolmacheva EN, Kashevarova AA, Nazarenko LP, Minaycheva LI, Skryabin NA, Lopatkina ME, Nikitina TV, Sazhenova EA, Belyaeva EO, Fonova EA, et al. 2020. Delineation of clinical manifestations of the inherited Xq24 microdeletion segregating with sXCI in mothers: two novel cases with distinct phenotypes ranging from UBE2A deficiency syndrome to recurrent pregnancy loss. *Cytogenet Genome Res* **160**: 245–254. doi:10.1159/000508050
- Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, Martin HC, Lappalainen T, Posthuma D. 2021. Genome-wide association studies. *Nat Rev Methods Prim* **1**: 59. doi:10.1038/s43586-021-00056-9
- Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. 2016. SIFT missense predictions for genomes. *Nat Protoc* **11**: 1–9. doi:10.1038/nprot.2015.123
- Vogt O. 1926. Psychiatrisch wichtige Tatsachen der zoologisch-botanischen Systematik. *Z Gesamte Neurol Psychiatr* **101**: 805–832. doi:10.1007/BF02878364
- Wang K, Li M, Hakonarson H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**: e164. doi:10.1093/nar/gkq603
- Ward LD, Kellis M. 2012. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* **40**: D930–D934. doi:10.1093/nar/gkr917
- Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. 2011. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genetics* **89**: 82–93. doi:10.1016/j.ajhg.2011.05.029
- Wu D, Dou J, Chai X, Bellis C, Wilm A, Shih CC, Soon WWJ, Bertin N, Lin CB, Khor CC, et al. 2019. Large-scale whole-genome sequencing of three diverse Asian populations in Singapore. *Cell* **179**: 736–749.e15. doi:10.1016/j.cell.2019.09.019
- Xu X, Wang B, Jiang Z, Chen Q, Mao K, Shi X, Yan C, Hu J, Zha Y, Ma C, et al. 2021. Novel risk factors for craniofacial microsomia and assessment of their utility in clinic diagnosis. *Hum Mol Genet* **30**: 1045–1056. doi:10.1093/hmg/ddab055
- Xu X, Chen Q, Huang Q, Cox TC, Zhu H, Hu J, Han X, Meng Z, Wang B, Liao Z, et al. 2024. Dysregulation in spatiotemporal expression of *HMX1* coordinated by multifaceted enhancers drives an auricular disorder. bioRxiv doi:10.1101/2024.04.01.587639
- Zhang Y-B, Hu J, Zhang J, Zhou X, Li X, Gu C, Liu T, Xie Y, Liu J, Gu M, et al. 2016. Genome-wide association study identifies multiple susceptibility loci for craniofacial microsomia. *Nat Commun* **7**: 10605. doi:10.1038/ncomms10605

Received October 7, 2024; accepted in revised form March 10, 2025.