Genome-wide analysis of copy-number variation in humans with cleft lip and/or cleft palate identifies *COBLL1*, *RIC1*, and *ARHGEF38* as clefting genes

Authors

Lisa A. Lansdon, Amanda Dickinson, Sydney Arlis, ..., Douglas W. Houston, Jeffrey C. Murray, J. Robert Manak

Correspondence

john-manak@uiowa.edu

The contribution of copy-number variants to cleft lip with or without cleft palate (CL/P) has been relatively understudied. Using a strategy to identify likely higher effect size microdeletions, we identify *COBLL1*, *RIC1*, and *ARHGEF38* as genes associated with CL/P that play important roles in vertebrate craniofacial development.

Lansdon et al., 2023, The American Journal of Human Genetics *110*, 71–91
January 5, 2023
https://doi.org/10.1016/j.ajhg.2022.11.012

CellPress

Genome-wide analysis of copy-number variation in humans with cleft lip and/or cleft palate identifies *COBLL1*, *RIC1*, and *ARHGEF38* as clefting genes

Lisa A. Lansdon,[1,2,3,4,5] Amanda Dickinson,[6] Sydney Arlis,[2] Huan Liu,[7] Arman Hlas,[2] Alyssa Hahn,[3] Greg Bonde,[7] Abby Long,[2] Jennifer Standley,[1] Anastasia Tyryshkina,[8] George Wehby,[9] Nanette R. Lee,[10] Sandra Daack-Hirsch,[11] Karen Mohlke,[12] Santhosh Girirajan,[8] Benjamin W. Darbro,[1,3] Robert A. Cornell,[3,7] Douglas W. Houston,[2,3] Jeffrey C. Murray,[1,3] and J. Robert Manak[1,2,3,*]

Summary

Cleft lip with or without cleft palate (CL/P) is a common birth defect with a complex, heterogeneous etiology. It is well established that common and rare sequence variants contribute to the formation of CL/P, but the contribution of copy-number variants (CNVs) to cleft formation remains relatively understudied. To fill this knowledge gap, we conducted a large-scale comparative analysis of genome-wide CNV profiles of 869 individuals from the Philippines and 233 individuals of European ancestry with CL/P with three primary goals: first, to evaluate whether differences in CNV number, amount of genomic content, or amount of coding genomic content existed within clefting subtypes; second, to assess whether CNVs in our cohort overlapped with known Mendelian clefting loci; and third, to identify unestablished Mendelian clefting genes. Significant differences in CNVs across cleft types or in individuals with non-syndromic versus syndromic clefts were not observed; however, several CNVs in our cohort overlapped with known syndromic and non-syndromic Mendelian clefting loci. Moreover, employing a filtering strategy relying on population genetics data that rare variants are on the whole more deleterious than common variants, we identify several CNV-associated gene losses likely driving non-syndromic clefting phenotypes. By prioritizing genes deleted at a rare frequency across multiple individuals with clefts yet enriched in our cohort of individuals with clefts compared to control subjects, we identify *COBLL1*, *RIC1*, and *ARHGEF38* as clefting genes. CRISPR-Cas9 mutagenesis of these genes in *Xenopus laevis* and *Danio rerio* yielded craniofacial dysmorphologies, including clefts analogous to those seen in human clefting disorders.

Introduction

Cleft lip and/or cleft palate (CL/P) is a common birth defect occurring on average in one in every 1,000 live births.[2] Approximately 70% of all clefts are isolated occurrences (non-syndromic [NSCL/P]), with the remaining individuals presenting with additional clinical phenotypes (syndromic [SCL/P]).[3] Studies sub-stratifying individuals with clefts by cleft type (cleft lip and palate [CLP], cleft lip only [CL], cleft palate only [CPO]), cleft laterality [unilateral, bilateral], or sidedness [left, right]) have demonstrated their overlapping and unique epidemiology.[4] CPO occurs more frequently in females than males while CL/P has increased prevalence in males,[5] and left-sided, unilateral clefts are the most common cleft type while bilateral clefts occur least frequently.[6] Although NSCL/P and NSCL have historically been grouped as etiologically similar entities,[2,7] recent studies have identified genetic factors which contribute uniquely to each cleft subtype.[4,8–11]

Copy-number variants (CNVs), defined as abnormal gains or losses of portions of chromosomal DNA greater than 1 kilobase (kb) in size, are common causes of disease,[12–18] and CNVs carried by individuals with CL/P have begun to explain a portion of the missing heritability of clefting.[19–27] Published estimates for the detection of pathogenic CNVs in individuals with SCL/P range from 21.4% (31/145)[28] to 60% (3/5),[27] while a detection rate of 7.2% (9/125)[28] has been cited for individuals with NSCL/P. However, due to the inherent heterogeneity of syndromes associated with CL/P, studying the CNV landscape of individuals with NSCL/P compared with that of SCL/P has been challenging, and to our knowledge CNV data from large NSCL/P cohorts have not been utilized to conduct comparative analyses within NSCL/P subtypes. One systematic review published in 2012 reported that clinically significant chromosomal defects detected using routine karyotyping or microarray were observed nearly exclusively in individuals with SCL/P in postnatal cohorts.[29] Several important limitations of the study noted by the authors include the inconsistent use of

[1]Department of Pediatrics, University of Iowa, Iowa City, IA 52242, USA; [2]Department of Biology, University of Iowa, Iowa City, IA 52242, USA; [3]Interdisciplinary Genetics Program, University of Iowa, Iowa City, IA 52242, USA; [4]Department of Pathology and Laboratory Medicine, Children's Mercy Kansas City, Kansas City, MO 64108, USA; [5]Department of Pathology, University of Missouri – Kansas City School of Medicine, Kansas City, MO 64108, USA; [6]Virginia Commonwealth University, Richmond, VA 23284, USA; [7]Department of Anatomy and Cell Biology, University of Iowa, Iowa City, IA 52242, USA; [8]Pennsylvania State University, University Park, PA 16802, USA; [9]College of Public Health, University of Iowa, Iowa City, IA 52242, USA; [10]Office of Population Studies Foundation, Inc., University of San Carlos, Cebu City, Philippines; [11]College of Nursing, University of Iowa, Iowa City, IA 52242, USA; [12]University of North Carolina, Chapel Hill, NC 27514, USA
*Correspondence: john-manak@uiowa.edu
https://doi.org/10.1016/j.ajhg.2022.11.012.

karyotyping versus microarray across clefting subgroups and studies, as well as considerable variability in how syndromic versus non-syndromic individuals were defined.[29] Furthermore, a landmark paper which defined CNV profiles of individuals with developmental delay found that individuals with neurodevelopmental phenotypes and craniofacial findings, including CL/P, harbored larger CNVs than individuals with autism spectrum disorder or epilepsy alone.[17] Conversely, a research study employing early microarray technology identified sub-karyotypic deletions overlapping known cleft loci in individuals with SCL/P and NSCL/P at nearly equivalent frequencies and sizes.[26] These included a classic ~2.7 megabase (Mb) deletion of 22q11.21 overlapping the DiGeorge syndrome (MIM: 188400) locus, deletions of *IRF6* (MIM: 607199) ranging from 100 kilobases (kb) to 1 Mb in size (in individuals with Van der Woude syndrome [MIM: 119300] within their SCL/P cohort),[26] and large deletions in two individuals with NSCL/P (one with a 3.2 Mb deletion at chromosome 6q25.1q25.2 and one with a 2.2 Mb deletion at 10q26.11q26.13 overlapping *FGFR2* [MIM: 176943]).[26] Other studies of individuals affected with clefting phenotypes have used microarrays to identify rare, likely etiologic deletions overlapping known Mendelian clefting loci,[25] as well as putative clefting regions,[20,23,25,27,28] and have begun to explore the role of higher-frequency variants (known as copy-number polymorphisms, or CNPs) in clefting.[19,24] However, these studies were conducted on relatively small collections of individuals with CL/P, precluding the discovery of additional robust genetic associations. Here we present the largest study to date of CNVs in individuals with SCL/P and NSCL/P, allowing us to comprehensively assess the respective CNV profiles of individuals between and within these subgroups, compare these CNVs to known causes of Mendelian clefting, and also identify clefting loci within individuals with NSCL/P.

In this study, we generated high-resolution comparative genomic hybridization array data (aCGH) using a cohort of 869 individuals of Filipino ancestry (792 of whom had NSCL/P and 77 of whom had SCL/P). We then assessed the total number of CNVs, CNV load, and CNV burden between cleft type and within NSCL/P subgroups and observed no significant differences between any groups. In the second part of our study, we identified 28 individuals with CNVs overlapping known syndromic loci in addition to 160 clinically relevant clefting genes overlapped by at least one CNV in our cohort. Finally, using a proprietary gene discovery pipeline, we identified genes deleted at a rare frequency in individuals with clefts whose loss was enriched in our affected cohort compared to control subjects, leading to prioritization of three putative NSCL/P clefting loci: *RIC1 homolog, RAB6A GEF complex partner* (*RIC1* [MIM: 610354]), *Rho guanine nucleotide exchange factor 38* (*ARHGEF38* [MIM: 619919]), and *Cordon-bleu WH2 repeat protein like 1* (*COBLL1* [MIM 610318]). F0 CRISPR-Cas9 deletion of these candidate genes in *Xenopus laevis* (African clawed frog) and *Danio rerio* (zebrafish) resulted in craniofacial dysmorphologies, including medial clefts in frogs that are analogous to human clefts.

Material and methods

Sample collection

Samples were gathered from individuals seen during surgical screening as part of Operation Smile medical missions in the Philippines (910)[30] or evaluated at the University of Iowa (308 individuals of European ancestry and 101 individuals of unknown ancestry) as part of contact during clinical care or epidemiologic surveys. All individuals were recruited following signed informed consent obtained in compliance with Institutional Review Board (IRB No. 199804081 [Philippines] and IRB No. 199804080 [Iowa]). After bioinformatic quality controls of the array-based comparative genomic hybridization (aCGH) data,[20] 1,108 probands were analyzed (1,025 NSCL/P; 83 SCL/P).

Array-based comparative genomic hybridization

Comparative genomic hybridization was performed as recommended by the manufacturers (Roche NimbleGen cgh_cnv_userguide_v7p0; Agilent G4410-90020v3_1_CGH_ULS_Protocol). Briefly, 1 μg (Agilent: 0.5 μg) of DNA from an individual with a cleft was labeled with Cy3-coupled nonamers and 1 μg (0.5 μg) of control DNA (from an unaffected male from the Philippines for male and female samples gathered during Operation Smile medical missions in the Philippines, and an unaffected male of European ancestry for samples gathered at the University of Iowa) was labeled with Cy5-coupled nonamers. Each labeled DNA sample was co-hybridized to a Roche NimbleGen (Human CGH 2.1M Whole-Genome Tiling v2.0D Array) or Agilent (SurePrint G3 Human CGH Microarray Kit 1 × 1M) human whole-genome tiling array, and the array was processed, scanned, and analyzed as previously described.[15,31]

Copy-number-variant calling and quality control

BioDiscovery's Nexus Copy Number FASST2 Segmentation Algorithm, a Hidden Markov Model (HMM)-based approach, was used to make initial copy number calls as previously described,[20] and quality control (QC) steps were performed as described below. Microarray data were QC'd based on several data metrics generated by NimbleGen's DEVA, Agilent's Feature Extraction and Nexus software.[31] Arrays were retained if less than eight or seven metrics (for NimbleGen or Agilent, respectively) fell outside of two standard deviations of the mean.

For CNV call QC, duplicate arrays per individual and sex discordant arrays were removed from further analysis. X-shift values of females were adjusted to control for hybridization against a male control using the median X-shift across all females by array platform. Due to the increased fragmentation of calls made by Nexus software, all calls of the same type (gains versus losses) were merged using BEDTOOLS when breakpoints were within one base pair of each other. Nexus CNV calls were verified using DEVA's segMNT algorithm (NimbleGen) or CytoGenomic's ADM-2 algorithm (Agilent). Parameters were adjusted from default settings to more similarly call CNVs when compared to Nexus by setting the minimum segment difference to 0.3 and requiring a minimum of 3 probes per segment in DEVA. Calls between Nexus and either calling algorithm were compared using BEDTOOLS intersect, requiring 70% overlap in either direction to be retained for further analysis. To reduce false positive calls, we required a minimum

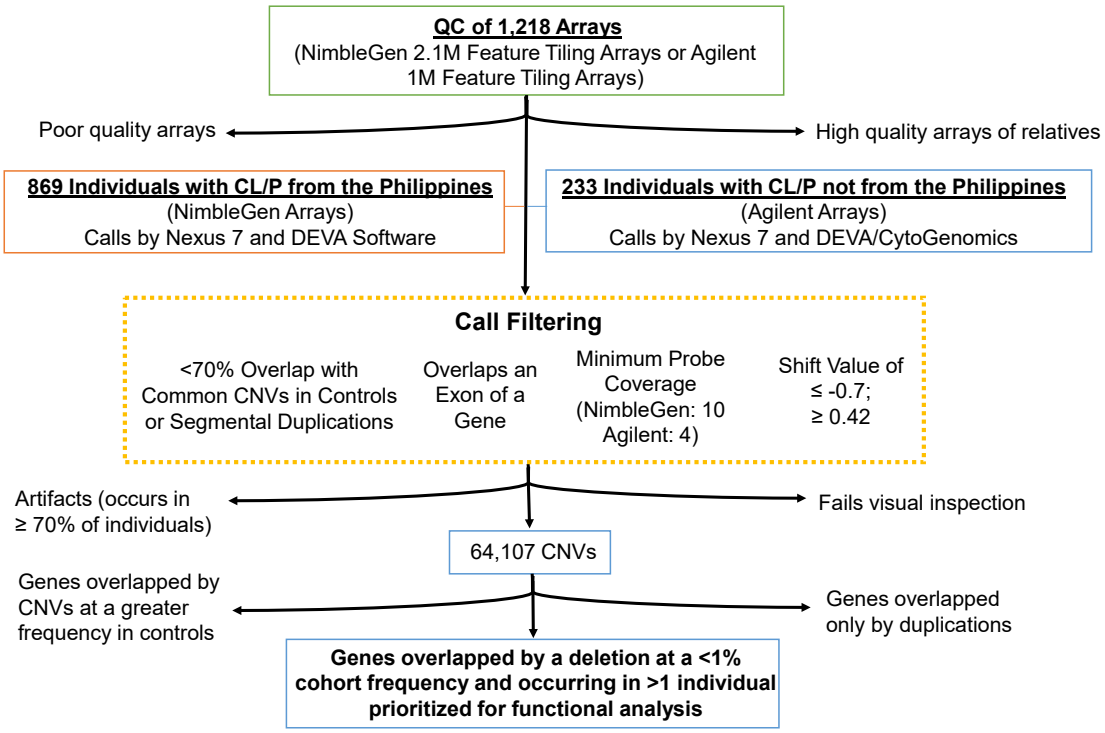

Figure 1. Bioinformatics copy-number-variant prioritization pipeline
Of the original 1,218 arrays, 1,102 passed quality controls and were used for downstream analyses. Copy-number variants (CNVs) which were overlapping an exon of a gene which passed minimum quality-control metrics (probe coverage and shift value) but occurring in less than 70% of the cohort or sharing less than 70% overlap with common CNVs or segmental duplications, were visually inspected. Genes that were recurrently deleted but at a frequency of less than 1% of the cohort with CL/P and at a higher frequency in individuals with clefts versus control subjects were prioritized for functional analysis.

number of probes encompassed by the CNV (NimbleGen: 10, Agilent: 4), a shift value of ≥ 0.42 or ≤ -0.7 for gains or losses, respectively, and that less than 70% of the CNV was overlapped by a segmental duplication, centromere, telomere, or pseudoautosomal region. All calls occurring in $\geq 70\%$ of individuals run on each platform or $\geq 70\%$ of the population were also removed, as these were likely the result of array-specific or control-specific artifacts, respectively. Finally, any arrays with the total number of calls falling two standard deviations from the mean or containing less than 20 CNV calls were considered outliers and removed from further analysis. This resulted in 869 high-quality NimbleGen arrays and 239 high-quality Agilent arrays for analysis, of which all CNV calls were visually inspected (Figure 1).

Due to the fact that no large-scale studies of CNVs in control populations from the Philippines have been published, CNV calling was performed on 1,783 control samples obtained from females from the Philippines enrolled in the Cebu Longitudinal Health and Nutrition survey.[32,33] Samples were genotyped on Affymetrix 5.0 arrays, analyzed using PennCNV,[34] and compared to calls within the individuals with CL/P for functional validation prioritization. Copy-number variants (CNVs) from select trios (when available) were validated in Illumina Omni 5 Exome BeadChip microarrays with PennCNV v.1.0.5.[34] In brief, PennCNV uses Log R Ratio (LRR), a measure of signal intensity, and B Allelic Frequency (BAF), a measure of allelic ratio, to infer the presence of duplications and deletions in microarrays. The CNV calls generated by PennCNV were annotated with RefSeq[35] hg19 (GRCh37) gene coordinates using a custom script to confirm the presence of deletions of interest, and the presence of deletions in *COBLL1* were confirmed by visual inspection of LRR and BAF plots.

Summary statistic generation

From our list of high-confidence CNV calls (see "copy-number-variant calling and quality control"), we employed the BEDTOOLS intersect function to calculate frequency of each CNV within the cohort by population requiring a 70% reciprocal overlap. Calls were separated into three lists for analysis: (1) all calls (Table S1), (2) semi-rare calls (occurring at a 1%–5% frequency), and (3) rare calls (occurring at a less than 1% frequency). Due to the fact that CNV number was largely influenced by hybridization platform and population, only the cohort for which we had sufficient power (samples hybridized using the NimbleGen platform) was used for the generation of the summary statistics. All statistical comparisons were conducted using VassarStats: http://vassarstats.net/.

Comparison with genomic disorder loci

All identified variants in samples gathered from individuals from the Philippines or at the University of Iowa with CL/P were compared to an in-house curated list of known genomic disorders (utilized by The University of Iowa's clinical cytogenetics laboratory; Table S2) using the BEDTOOLS intersect function and requiring a minimum of 70% reciprocal overlap (see Tables S2 and S3 for a summary and complete list of CNVs identified in our affected cohort, respectively).

Comparison with known clefting loci

All identified variants in our cohort of individuals with CL/P were compared to two lists of genes which contribute to Mendelian CL/P disorders: a list of 358 genes which are putatively involved in CL/P formation[36] and 336 clinically relevant genes involved

in Mendelian clefting[37] (see Table S4 for a list of Mendelian clefting genes overlapped by CNVs in our samples).

Rare variant analysis

High-confidence CNVs identified in samples obtained from individuals in the Philippines hybridized on the NimbleGen platform were filtered to identify calls overlapping coding regions of the genome. We compared the frequency of these calls within our cohort to a control dataset of individuals from the Philippines (see "copy-number-variant calling and quality control"), as well as an in-house curated list of CNVs from the Database of Genomic Variants[20] using the BEDTOOLS 70% intersect function and custom Python scripts. For our functional analysis, we focused on replicated deletions passing visual inspection which overlapped protein coding genes and occurred at a frequency of less than 1% but were deleted at a higher frequency in the individuals with NSCL/P from the Philippines versus control subjects. Three hundred twenty genes fit these criteria (Table S5) which were further prioritized for functional validation based on of several annotations, including haploinsufficiency score, presence of a deletion in our European cleft cohort, expression patterns in mouse (MGI: www.informatics.jax.org),[38] fish (ZFIN: zfin.org),[39] and frog (Xenbase: www.xenbase.org),[40] known human disease association (OMIM: omim.org), constraint (gnomAD: gnomad.broadinstitute.org),[41] gene function (NCBI: www.ncbi.nlm.nih.gov), and presence of craniofacial anomalies in individuals harboring deletions of these genes (DECIPHER: https://decipher.sanger.ac.uk/).[42]

Pathway analysis

Chromosomal position over-representative analysis was performed using GeneTrail 3.2[43] for genes deleted in one or more individuals with NSCL/P from the Philippines that were deleted at a higher frequency in affected individuals versus control subjects, yet at a frequency of less than 1% of the cohort overall (calls that occur at a less than 1% frequency are listed in Table S5), in addition to genes deleted in only one individual within the NSCL/P cohort from the Philippines yet at a higher frequency in affected individuals versus control subjects (singleton deletions fulfilling these criteria can be found in Table S6). Default values were used including Benjamini-Yekutieli to control for false discovery rate in multiple testing.[44] Pathway analysis was performed using Ingenuity Pathway Analysis (IPA) Core Analysis with default settings including consideration of direct and indirect relationships, experimentally observed and high (predicted) confidence levels, and all available species, as well as an "Enriched diseases and functions" output (Qiagen). Also included in the analysis pipeline was a Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. Canonical pathways that achieved statistical significance after Benjamini-Hochberg correction were included in the results (Tables S7A–S7H).

Xenopus laevis

Staging

Xenopus laevis embryos were obtained using standard procedures[45] approved by the VCU Institutional Animal Care and Use Committee (IACUC protocol number AD20261) and University of Iowa IACUC (protocol number 2021664). Embryos were staged according to Nieuwkoop and Faber.[46] Stages are also reported as hours post-fertilization at 23°C for better comparisons across vertebrates.

In situ *hybridization*

Since *X. laevis* is allotetraploid and thus has two homeologs of a large number of genes (denoted S and L), we designed our *in situ* probes to maximize the likelihood that both homeologs would be detected. To design *in situ* probes for *arhgef38, cobll1,* and *ric1* that are most likely to detect both homeologs of each gene, we aligned the L and S sequences using MultAlin and identified the most similar 800 base pair (bp) region from highly conserved exons (exons 10–13 for *arhgef38,* exon 13 for *cobll1*, and exons 19–22 for *ric1*). The sequences were synthesized by IDT and inserted into the pUCIDT vector at the EcoRV site. Antisense and sense probes were synthesized using T3 and T7 polymerases, respectively. Probes were diluted to 1 μg/mL in hybridization buffer prior to use. Whole-mount *in situ* hybridization was performed as previously described.[47] A minimum of six embryos were used for each probe at each stage.

Morpholino knockdown

Splice-blocking antisense morpholinos (MOs) were designed and purchased from GeneTools (sequences in Table S8). A standard control MO was obtained from GeneTools. All MOs were labeled with fluorescein which allowed separation of un-injected individuals from injected, fluorescent animals by 24 h of development. Microinjections were performed using an Eppendorf Femtojet micro-injector and a Zeiss Discovery V8 stereoscope. Embryos were placed in a dish lined with nylon Spectra mesh (1,000 μm opening and 1,350 μm thickness) at the bottom to hold embryos in place and filled with 3% Ficoll 400 (Fisher, cat # BP52 5) dissolved in 0.1X MBS. MOs were diluted to 34 ng/μL and 17 ng/μL, and 1–2 nL were injected into each embryo (effective concentrations reported in Table S8). To assess whether the MOs resulted in changes in mRNA structure, PCR was performed with Apex Hotstart Taq master mix (Bioline, cat # 42-144) on a BioRad MJ Mini Personal Thermocycler. The PCR products were analyzed on a 2% agarose gel prepared with molecular grade agarose (Bioline, cat # BIO-41025) in TAE buffer.

CRISPR-Cas9 mutagenesis

sgRNA was designed using the ChopChop software.[48] A sequence was chosen that best targeted the desired gene with no off-targets and high efficiency (see sequences in Table S9). The sgRNAs were purchased from Synthego and diluted as recommended in low EDTA TE buffer. Then, 200 pg of sgRNA was incubated with 2 pg of Cas9 protein (PNA Bio Inc., cat # CP01) for 10 min and 1–2 μL was injected into each embryo. A negative control consisted of the same concentrations of Cas9 protein. Since there was no way to visually determine whether an embryo was mutant, all embryos injected were counted but only those with some phenotype were used to calculate percentages with craniofacial defects. The DNA was extracted from 10 randomly selected embryos with a craniofacial malformation using the HotShot protocol.[49] Each embryo was immersed in 40 μL of an alkaline lysis buffer (25 mM NaOH, 0.2 mM Na-EDTA) and heated for 40 min at 95°C. The solution was then cooled, and an equal volume of neutralization buffer (40 mM Tris-HCL) was added. One mL of this solution was used in a standard PCR reaction with Hotstart Taq master mix and primers that flanked the predicted mutation site (primer sequences can be found in Table S9). The product was then sent for purification and sequencing at Genewiz (Azenta Life Sciences) using the same forward primers.

Imaging facial features of X. laevis

At stage 42–43, tadpoles were anesthetized in 1% tricaine for 10 min and then fixed in 4% paraformaldehyde overnight at 4°C. A No. 15 scalpel (VWR, cat # 82029-856) and Dumont No. 5 forceps (Fisher, cat # NC9404145) were used to make two cuts to isolate the head: first at the posterior end of the gut and then second caudal to the gills. Isolated heads were mounted in small

holes or depressions in either agarose or clay-lined dishes containing PBS with 0.1% Tween (PBT). The faces were imaged using a Discovery V8 stereoscope fitted with an Axiovision digital camera (Zeiss).

Alcian blue staining

At stage 45, tadpoles were anesthetized in 1% tricaine for 10 min and then fixed in Bouin's solution overnight at room temperature. After washing out the fixative in 70% EtOH, tadpoles were soaked for 3–4 days in an Alcian blue solution (20% acetic acid, 80% EtOH, 0.1 mg/mL Alcian blue). Tadpoles were washed with acid alcohol (AA; 1% HCl, 70% EtOH) and rehydrated in PBT, and pigment was removed by soaking in 3% H_2O_2 for 45–60 min. Tadpoles were then washed in 1% KOH and mounted in 75% glycerol for imaging.

Danio rerio

D. rerio *maintenance*

Danio rerio embryos, larvae, and adults were reared as described previously[50] in the University of Iowa Zebrafish Facility, and approved by the University of Iowa IACUC (ACURF protocol number 1003051). Animals were staged by hours or days postfertilization at 28.5°C (hpf or dpf, respectively).[51]

CRISPR-Cas9 mutagenesis

Fertilized *Danio rerio* embryos were co-injected at the one- to two-cell stage with sgRNA (200–400 pg per embryo) and/or Cas9 protein (IDT) at 2 ng per embryo. The efficacy of each sgRNA within *Danio rerio* embryos was vetted by high-resolution melt analysis on eight individual injected embryos (sgRNA sequences are listed in Table S10).

Alcian blue staining

Danio rerio larvae were euthanized then fixed overnight in 4% paraformaldehyde (PFA). After a wash in phosphate-buffered solution (0.8% NaCl, 0.02% KCl, 0.02 M PO4 [pH 7]) with 0.25% Tween 20 (PBST), pigment was removed by soaking in a 3% H_2O_2 and 0.5% KOH medium for 20–30 min. Larvae were then washed in PBST and soaked overnight in an acid alcohol (AA) solution containing Alcian blue (0.37% HCl, 70% EtOH, and 0.1% Alcian blue). The larvae were then washed extensively in AA, rehydrated in PBST, and mounted in 4% methyl cellulose for imaging.

Results

Comparative analysis of CNV profiles between clefting subtypes

We analyzed aCGH data from 869 individuals from the Philippines, 792 with NSCL/P (206 CL, 531 CLP, 54 CPO, 1 unknown) and 77 with SCL/P (8 CL, 56 CLP, 9 CPO, 4 unknown), hybridized on NimbleGen 2.1M feature whole-genome tiling arrays that passed quality controls (see material and methods). Data analysis was performed using Nexus Copy Number (version 7.5; BioDiscovery), DEVA (version 1.2; Roche NimbleGen), BEDTOOLS, and several in-house custom Python scripts (see material and methods). We conducted a comparative analysis of individuals with NSCL/P and SCL/P considering total number of CNVs detected (Figures S1A–S1D), the amount of genomic content overlapped by CNV events ("CNV load;" Figures S1E–S1H), and the number of protein-coding genes overlapped by CNVs ("CNV burden;" Figures S1I–S1L). Overall, we observed no significant differences between groups (2-sided Mann-Whitney; Bonferroni corrected p value required for significance: <0.00185), even when stratifying by CNV type (loss or gain) or CNV frequency (all CNVs, Figure S1; or CNVs occurring at a 1%–5% frequency or <1% frequency, data not shown). We also assessed the 792 individuals from the Philippines with NSCL/P for differences between CNV number, load, and burden after stratifying by genotypic sex (500 male, 292 female; Figures S2A, S2E, and S2I, respectively), cleft type (206 CL, 531 CLP, 54 CPO; Figures S2B, S2F, and S2J, respectively), unilateral cleft sidedness (346 left, 170 right; Figures S2C, S2G, and S2K, respectively), and cleft laterality (206 bilateral, 516 unilateral; Figures S2D, S2H, and S2L, respectively), and observed no significant differences in CNV profiles between these NSCL/P subgroups (2-sided Mann-Whitney and Kruskal-Wallis; Bonferroni corrected p value required for significance: <0.00139). Finally, a qualitative investigation into the largest CNV occurring within each individual by cohort (total, NSCL/P, or SCL/P) showed a moderate increase in gains sized 300–400 kb within the SCL/P cohort, but otherwise no readily apparent differences in the largest gains and losses were observed (Figure S3).

Detection of CNVs overlapping known genomic disorders

To identify known pathogenic CNVs within our cohort of individuals with CL/P, we compared all CNVs passing quality controls to a list of loci previously implicated in genomic disorders (Table S2). Twenty-eight individuals (25 from the Philippines; 3 of European ancestry) were found to have CNVs which shared 70% reciprocal overlap with a known syndromic disorder locus (Tables 2, S2, and S3). The most common finding was an ~80 kb duplication of the *HOXD* cluster at 2q31.1 in 11 individuals (4 SCLP; 4 NSCL; 3 NSCLP), followed by four ~1.6 Mb type II deletions (1 SCPO, 3 NSCLP) and two ~1.2 Mb type II duplications (2 NSCPO) of 22q11.2. Two individuals had CNVs overlapping the thrombocytopenia-absent radius (TAR) susceptibility region on 1q21.1 (one gain, one loss; both with NSCLP [MIM: 274000]), two had ~1.3 Mb deletions of 3q29 (2 NSCLP), and two individuals had ~540 kb deletions of 16p11.2 (1 NSCLP; 1 NSCPO). In addition, one individual each was found to have an ~874 kb duplication of the Williams-Beuren syndrome locus on 7q11.23 (NSCLP [MIM: 194050]), an ~744 kb duplication of 15q11.2 (NSCL), an ~786 duplication of 16p13.11 (NSCLP), an ~280 kb duplication of Xq28 (NSCL), and confirmation of trisomy 21 in an individual with Down syndrome (MIM: 190685) and SCLP. Finally, we note that two individuals diagnosed with Van der Woude syndrome (VWS) who had been previously reported as negative for sequence variants within *IRF6* and *GRHL3* ([MIM: 608317]; known Mendelian causes of VWS[52,53]) were assessed for CNVs overlapping these genes or nearby non-coding regions, and no variants were detected.

Detection of CNVs overlapping known clefting loci

In order to determine whether any of the CNVs passing our quality control filters overlapped with genes associated with Mendelian clefting disorders, we compared any genes within the deleted or duplicated interval with a list of candidate genes from 2017,[36] as well as a list of clinically relevant genes from 2020,[37] associated with clefting phenotypes. This yielded a list of 123 candidate genes and 160 clinically relevant genes implicated in clefting that were overlapped by at least one CNV in our cohort (56 of which overlapped between both lists; Table S4). Further restriction of this list by deletions occurring in $\leq$1% of our cohort and with a population frequency cutoff of 0.1% in our Filipino and DGV curated controls (see material and methods) resulted in a list of 51 genes (16 candidate: *ANK1* [MIM: 612641], *AUTS2* [MIM: 607270], *COMT* [MIM: 116790], *CRLF1* [MIM: 604237], *GMPPB* [MIM: 615320], *LMNA* [MIM: 150330], *OTX2* [MIM: 600037], *RAI1* [MIM: 607642], *RBM8A* [MIM: 605313], *SCLT1* [MIM; 611399], *SMOC1* [MIM; 608488], *SPEG* [MIM: 615950], *TAC3* [MIM: 162330], *TACR3* [MIM: 162332], *TBX4* [MIM: 601719], and *TPM2* [MIM: 190990]; 24 clinically relevant: *ARID3B* [MIM: 612457], *BMPR1A* [MIM: 601299], *CHD7* [MIM: 608892], *DHCR7* [MIM: 602858], *FZD2* [MIM: 600667], *GDF11* [MIM: 603936], *GREM1* [MIM: 603054], *IQGAP2* [MIM: 605401], *ISM1* [MIM: 615793], *KDM6A* [MIM: 300128], *KMT2A* [MIM: 159555], *MMP3* [MIM: 185250], *MSX1* [MIM: 142983], *NBAS* [MIM: 608025], *NECTIN1* [MIM: 600644], *NEDD4L* [MIM: 606384], *PORCN* [MIM: 300651], *RYR1* [MIM: 180901], *SCAMP1* [MIM: 606911], *SIX1* [MIM: 601205], *SNAP29* [MIM: 604202], *STK11* [MIM: 602216], *TBX1* [MIM: 602054], and *UFD1* [MIM: 601754]; and 11 appearing on both lists: *ASXL1* [MIM: 612990], *B3GLCT* [MIM: 610308], *BCOR* [MIM: 300485], *DYNC2H1* [MIM: 603297], *FGFR2* [MIM: 176943], *HDAC8* [MIM: 300269], *IFT140* [MIM: 614620], *MSX2* [MIM: 123101], *SKI* [MIM: 164780], *WDR11* [MIM: 606417], and *YAP1* [MIM: 606608]). The majority of the aforementioned genes were overlapped by one or two deletions in the cohort (43/51; 84%), and *TACR3* (candidate gene) was the gene most frequently overlapped by rare losses, with eight individuals with clefts harboring a deletion. *TBX1* (clinically relevant) was deleted in five individuals, and *COMT* (candidate), *NEDD4L* (clinically relevant), *SNAP29* (clinically relevant), *STK11* (clinically relevant), and *UFD1* (clinically relevant) were each overlapped by four rare deletions. Of note, *COMT*, *SNAP29*, *UFD1L*, and *TBX1* were overlapped by the same deletion in four or five different individuals, respectively, within the 22q11.2 locus, and likely indicate the presence of DiGeorge syndrome in these individuals.

Pathway analysis

Over-representation analysis using chromosomal position was performed for genes deleted at a rare frequency within the NSCL/P cohort from the Philippines (arbitrarily defined as deleted in less than 1% of the affected cohort) that were also deleted at a higher frequency in individuals with NSCL/P versus control subjects. The most statistically significant results for genes deleted in more than one individual with NSCL/P were chromosomal loci at which known genomic disorders have been identified including 3q29, 4p16.1, and 22q11.21 (Table S7A). These disorders all contain CL/P within their associated spectrum of disease.[54–58] This analysis was repeated for genes deleted in only one individual in our NSCL/P cohort from the Philippines, and several emerging and potential disease loci were identified. These include 7q35, 11q22.1q11.2, 14q32.32q32.33, and 19p13.3 (Table S7B). In addition, we observed that these singleton deletions overlapped well-known genomic disorder loci, including 1q21.1 and 16p11.2.

To further explore the genes which were recurrently deleted or only deleted in one individual in our NSCL/P cohort from the Philippines, pathway analysis was performed using the QIAGEN Ingenuity Pathway Analysis (IPA) platform (QIAGEN IPA: https://digitalinsights.qiagen.com/IPA).[59] Although none of the pathways identified for recurrently deleted genes reached significance, singleton deleted genes showed statistically significant enrichment for 15 canonical pathways including inhibition of matrix metalloproteases, synaptogenesis signaling pathway, role of osteoblasts, osteoclasts, and chondrocytes in rheumatoid arthritis, and axonal guidance signaling (Table S7C). Next, we preselected genes based on DECIPHER (DECIPHER: https://decipher.sanger.ac.uk/) haploinsufficiency scores of $<$10 for our analysis pipeline, since scores in this range denote likely haploinsufficiency loci (genes for which loss of one copy is deleterious and likely to cause a disease phenotype). For singleton deleted genes with haploinsufficiency scores $<$10, numerous other pathways emerged, including epithelial adherens junction signaling, RANK signaling in osteoclasts, regulation of the epithelial-mesenchymal transition pathway, gap junction signaling, apoptosis signaling, WNT/beta-catenin signaling, and integrin signaling, among others (Table S7D). Additionally, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of these genes revealed 18 enriched cellular pathways including adherens junction, apoptosis, axon guidance, and focal adhesion (Table S7E).[60–62]

In order to further mine our dataset, we used IPA/KEGG to look for canonical pathways, KEGG, and functions pathways using a combined list of all genes deleted in one or more individuals at a less than 1% cohort frequency with HI scores of 10 or less. We chose to focus on this list given the strong haploinsufficiency scores which suggest that they would be the most likely genes to show disease phenotypes upon monoallelic loss. Among the statistically significant categories known to be important for cranial neural crest cell (CNCC) function were the function annotations "migration of cells" (adjusted $p = 6.92 \times 10^{-4}$) which includes 27 genes, "formation of cellular protrusions" (adjusted $p = 3.86 \times 10^{-5}$) which includes 19 genes, and "organization of cytoskeleton" (adjusted $p = 1.09 \times 10^{-3}$)

Table 1. Top seven candidate genes with strong likelihood of involvement in craniofacial development

Gene	Number probands with gene deleted	Gene function
Rho guanine nucleotide exchange factor 38 (*ARHGEF38*)	4	unknown; Rho signaling
Cordon-bleu WH2 repeat protein-like 1 (*COBLL1*)	3	promotes actin filament formation and dendritic branching via WH2 domain
Exocyst complex component 4 (*EXOC4*)	3	part of the exocyst complex which targets exocytic vesicles for docking on the plasma membrane; essential for epithelial polarity and interacts with actin cytoskeleton
Lipoprotein lipase (*LPL*)	4	triglyceride hydrolase and ligand factor for receptor-mediated lipoprotein uptake
Plakophilin 2 (*PKP2*)	4	Armadillo repeats allow localization to cell desmosomes and nuclei, linking cadherins to intermediate filaments in the cytoskeleton; may regulate β-catenin
RAB6A GEF complex partner 1 (*RIC1*)	3	necessary for nucleotide exchange on Rab6A; Rab6A functions in the exocytic pathway and interacts with ARHGEF10
von Willebrand factor D and EGF domains (*VWDE*)	7	enables cell adhesion in the blood stream

Top candidate genes deleted in our cohort are indicated in the first column, with the number of probands carrying deletions of each gene indicated in the second column. The third column describes any known functions of each gene.

which includes 20 genes (Tables S7F–S7H). These analyses also led to several additional enriched pathways, including ephrin receptor signaling.

Identification of putative clefting loci

Due to our previous identification of a clefting locus, *Isthmin 1* (*ISM1* [MIM: 615793]), in a cohort of 140 individuals from the Philippines with NSCL/P,[20] we employed a similar filtering strategy to identify putative clefting candidates within the 792 probands from the Philippines with NSCL/P. Since our strategy relies upon population genetics data in which high effect size variants are generally rare in frequency in the population (due to selection against such variants), we thus focused on protein coding genes that were overlapped by a deletion (which are generally more deleterious than gains) at a rare frequency in individuals with NSCL/P. We stringently defined this threshold as deletions occurring at a less than 1% frequency yet occurring in two or more affected individuals that were also found at a higher frequency in individuals with NSCL/P than control subjects. These filtering criteria resulted in 320 genes for further prioritization (a list of these genes can be found in Table S5). We then assessed each call for replication within the cohort of European ancestry with NSCL/P (229) and used haploinsufficiency score (<10 predicted as likely haploinsufficient in DECIPHER; DECIPHER: https://decipher.sanger.ac.uk/); gnomAD constraint score (gnomAD: https://gnomad.broadinstitute.org); overlap by CNVs in DECIPHER within individuals with craniofacial phenotypes[42]; craniofacial expression pattern (if known) in *Danio rerio* (ZFIN: https://zfin.org), *Xenopus* (Xenbase: www.xenbase.org), and mouse (MGI: www.informatics.jax.org/); ortholog presence in *Danio rerio* and *Xenopus laevis*; known biological function (National Center for Biotechnology Information [NCBI]; NCBI: https://www.ncbi.nlm.nih.gov/gene); and previous implication in clefting[36,37] to select the most promising Mendelian clefting candidate genes (all annotations can be found in Table S5). Deletions overlapping a maximum of three genes were preferentially selected for ease of functional interpretation. Given the available data for each gene and their fulfillment of the above criteria, we identified seven genes that had a strong likelihood of being associated with craniofacial development (Table 1): *R. guanine nucleotide exchange factor 38* (*ARHGEF38*); *Cordon-bleu WH2 repeat protein-like 1* (*COBLL1*); *Exocyst complex component 4* (*EXOC4* [MIM: 608185]); *Lipoprotein lipase* (*LPL* [MIM: 246650]); *Plakophilin 2* (*PKP2* [MIM: 602861]); *RAB6A GEF complex* (*RIC1*); and *von Willebrand factor D and EGF domains* (*VWDE*) (Log2 plots of these deletions generated by Nexus software for *EXOC4*, *LPL*, *PKP2*, and *VWDE* can be found in Figure S4; for *ARHGEF38*, *COBLL1* and *RIC1*, see Figure 2).

For the individuals with CL/P harboring deletions overlapping these candidate Mendelian NSCL/P genes, we assessed whether the CNV event was *de novo* or inherited by hybridizing parent samples on Agilent 1M feature whole-genome tiling arrays or Illumina HumanOmni5Exome SNP arrays, when available. Of note, due to the low quality of some of the parental samples, parentage could not be confirmed in all individuals. In one tested trio per gene, we detected a *de novo* deletion overlapping *EXOC4, LPL,* and *RIC1*, whereas a second deletion of *RIC1* was inherited from an affected mother, and one deletion overlapping *COBLL1* was inherited from an unaffected father. Although not selected for functional analysis, we also identified three

Table 2. Summary of copy-number variants overlapping genomic disorder loci

Genomic disorder locus	CNV type	Chr	Start coordinate (hg18)	Stop coordinate (hg18)	CNV size range (kb)	Number of individuals	Ancestral country of origin	Cleft type	Cleft lip laterality	Genotypic sex	Diagnosis (syndromic)
1q21.1 TAR susceptibility	del	1	144,112,609	144,525,602	413	1	P	NSCLP	B	F	N/A
1q21.1 TAR susceptibility	dup	1	144,112,609	144,494,342	382	1	E	NSCLP	Un	M	N/A
HOXD cluster	dup	2	176,665,222	176,750,540	73–85	11	P	NSCLP:3; NSCL:4; SCLP:4	L:2; R:3; B: 6	M:6; F:5	C&L:1; Ank:1; Un:2
3q29	del	3	197,513,580	198,827,680	1314	2	P	NSCLP	L:1; R:1	M:1; F:1	N/A
WBS distal	dup	7	74,907,157	75,754,545	847	1	P	NSCLP	L	F	N/A
15q11.2 BP1-2	dup	15	20,428,073	21,172,461	744	1	P	NSCL	L	F	N/A
16p13.11	dup	16	15,413,831	16,199,769	786	1	P	NSCLP	L	M	N/A
16 p11.2 (593kb)	del	16	29,557,497	30,107,356	540–549	2	P:1; E:1	NSCLP:1; NSCPO:1	L	M	N/A
Trisomy 21	dup	21	26,384,126	46,944,323	20,560	1	P	SCLP	L	M	T21
DGS/VCFS Type II	dup	22	17,271,718	18,691,318	1,226–1,290	2	P:1; E:1	NSCPO	N/A	M	N/A
DGS/VCFS Type II	del	22	17,402,877	18,691,904	1,289–1,650	4	P	NSCLP:3; SCPO:1	L:2; B:1	M:2; F:2	PR
Xq28	dup	X	153,260,249	153,540,650	280	1	P	NSCL	L	F	N/A

Twenty-eight individuals were found to have copy-number variants overlapping known genomic disorder loci (see Table S2 for a complete list of loci and Table S3 for the coordinates of the call identified in each proband). hg18, human genome build coordinates GRCh18; kb, kilobase; TAR, thrombocytopenia absent radius; WBS, Williams-Buren syndrome; BP1-2, breakpoints 1 and 2; DGS/VCFS, DiGeorge/velocardiofacial syndrome; del, deletion; dup, duplication; chr, chromosome; P, Philippines; E, Europe; NSCLP, non-syndromic cleft lip and cleft palate; NSCL, non-syndromic cleft lip; SCLP, syndromic cleft lip and cleft palate; NSCPO, non-syndromic cleft palate only; SCPO, syndromic cleft palate only; B, bilateral; Un, unknown; L, left; R, right; N/A, not applicable; F, female; M, male; C&L, cleft and limb defects; Ank, ankyloblepharon; T21, Trisomy 21; PR, Pierre Robin.

heterozygous, putative *de novo* deletions overlapping *HNRNPL* (MIM: 603083), *EBI3* (MIM: 605816), and *CHP2*, heterozygous deletions of *FAM149A*, *PCDH9* (MIM: 603581), and *INPP5F* (MIM: 609389) (each of which were inherited from an affected mother), and two putative *de novo*, heterozygous 3q29 deletions (MIM: 609425).

Functional validation of putative clefting genes

Rho and Rab GTPase signaling components are known to play a role in a variety of processes associated with cellular dynamics including the formation of adhesion junctions, actin organization, cell division, cell migration, and membrane trafficking),[63–67] all of which are linked to craniofacial development. Additionally, genes involved in both Rho and Rab signaling are associated with craniofacial anomalies.[68–73] Thus, we chose the three genes in our list of top candidates that play roles in Rho/Rab GTPase signaling for functional validation in *Xenopus laevis* (*X. laevis*): *ARHGEF38*, *COBLL1*, and *RIC1*. *In situ* hybridization using probes designed to detect both homeologs of each gene (see material and methods) revealed expression of all three genes in the head region of *X. laevis* embryos during early events in craniofacial development such as CNCC migration and branchial arch specification (Figure S5). At stage 25, *arhgef38* is expressed ubiquitously in the head, including the presumptive branchial arches, with concentrated expression in the brain and spinal cord as well as the cement gland. Stages 27 and 29 show similar expression patterns to stage 25, although strong expression in the cement gland is no longer seen at stage 27 and is instead more pronounced in the otic vesicle and developing eye at stage 29. At stage 25, *cobll1* is strongly expressed in the otic vesicle, hatching gland, and cement gland as well as scattered epidermal cells. At stages 27 and 29, *cobll1* is ubiquitously expressed throughout the embryo including the presumptive branchial arches, with the strongest expression concentrated in the otic vesicle, cement gland, brain, and spinal cord. At stage 29, expression is further pronounced in the region of (and anterior to) the first branchial arch that gives rise to the midface. Finally, at stage 25, *ric1* shows very low levels of expression, while at stage 27 its expression is observed throughout the head, including the presumptive branchial arches and eye. Stage 29 expression is similar to that of

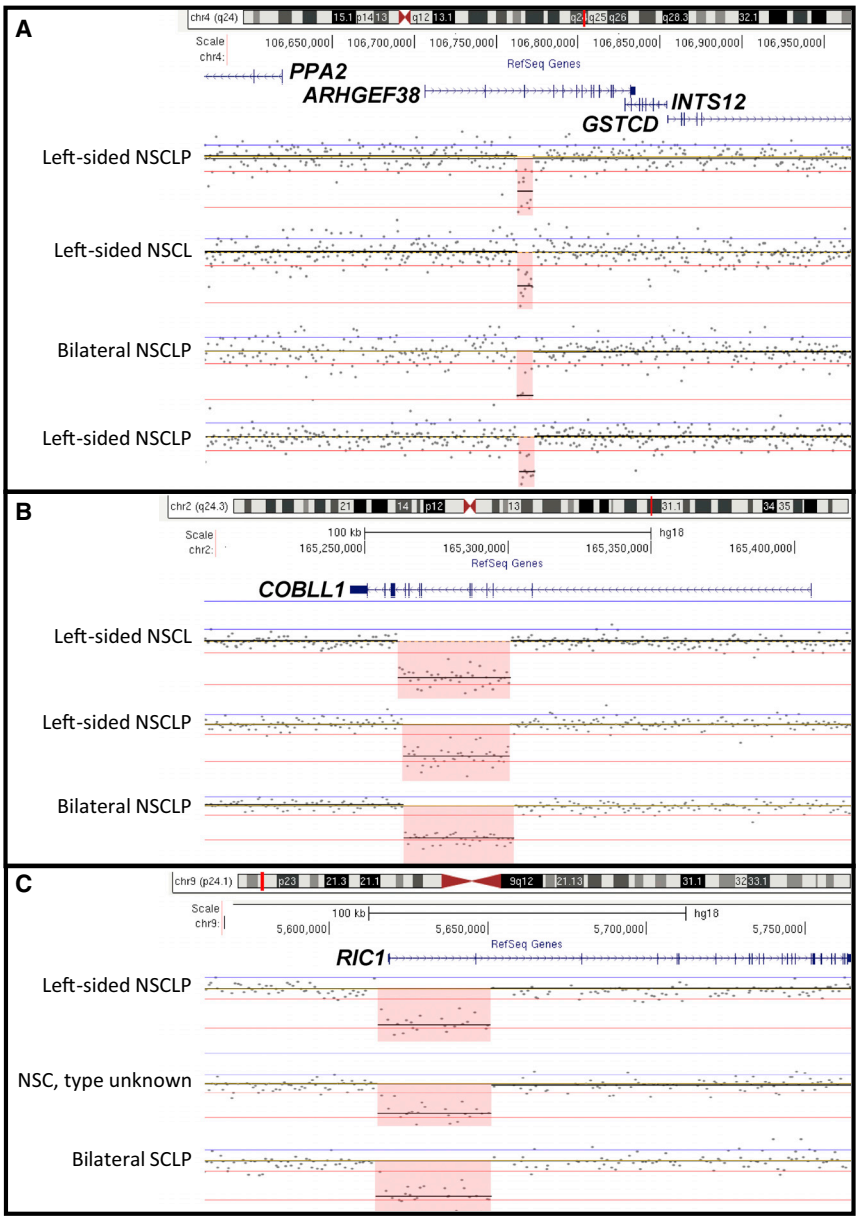

Figure 2. Log2 plots of top candidate clefting genes

Log2 plots of (A) *ARHGEF38*, (B) *COBLL1*, and (C) *RIC1* were generated using BioDiscovery's Nexus 7 software and the UCSC genome browser. NSCLP, non-syndromic cleft lip and palate; NSCL, non-syndromic cleft lip only; NSC, non-syndromic cleft; SCLP, syndromic cleft lip and palate.

stage 27, with refined expression in the posterior-most branchial arches, otic vesicle, brain, and eye. In summary, all three genes show expression in the region of the branchial arches (the equivalent of the mammalian pharyngeal arches) which give rise to many of the structures forming the jaws and palate.

To determine whether *cobll1*, *ric1*, and *arhgef38* are required for craniofacial development, we used both antisense oligos (morpholinos; MOs) and CRISPR-Cas9 to target individual homeologs of each gene. MOs or short guide RNAs (sgRNAs)/Cas9 targeting the S or L homeolog of each gene (Tables S8 and S9) were injected into the 1-cell stage of *X. laevis* embryos and their effectiveness was determined. To confirm that the MOs caused splicing defects, we used RT-PCR and primers flanking the exon targeted for deletion. Results indicated that indeed alternative gene products were produced in morphant tadpoles with craniofacial abnormalities (Figures S6–S8). The size of the alternative gene products was consistent with the predicted exon deletions. The effectiveness of CRISPR techniques were assessed by sequencing randomly selected mutants (with a craniofacial malformation) using primers that flanked the sgRNA target. Results indicated that all 10 CRISPR mutants for each gene tested had alternative sequences near the sgRNA target sites as predicted (Figures S6–S8C and S8D).

We next assessed the shape of the mouth for evidence of dysmorphologies by imaging and observing the faces of morphants and mutants at stages 42–43 (80–87 hours postfertilization; hpf), by which time the tissues that form the roof of the oral cavity have migrated to the region and have begun to expand and differentiate. A small percentage (ranging from 2.6% to 8.2%) of the control tadpoles injected with either control MOs or sgRNAs/Cas9 were smaller but despite this did not display obvious changes in craniofacial morphology including the shape of the mouth (Figure 3).

The *ric1.L* and *ric1.S* morphants (86.5% and 69.5%, respectively) and mutants (54.4% and 63.3%, respectively) had craniofacial defects which included narrower faces and eyes that were closer set, so much so that sometimes they appeared fused (Figures 3E–3I). Of the craniofacial malformations observed in these tadpoles, a portion also had triangular-shaped mouths that appeared cleft-like, indicative of primary palate malformation (*ric1.L* MO = 23.1%, *ric1.S* MO = 20.9%, *ric1.L* CR = 21.4%, *ric1.S* CR = 16.7%, Figures 3E–3I). In addition, for many of the *ric1* morphants, the buccopharyngeal membrane (a layer of cells that covers the mouth opening) failed to break down (Figure S8H).

The *cobll1.L* and *cobll1.S* morphants (86.0% and 92.0%, respectively) and mutants (61.5% and 54.8%, respectively) also had craniofacial defects which included narrower

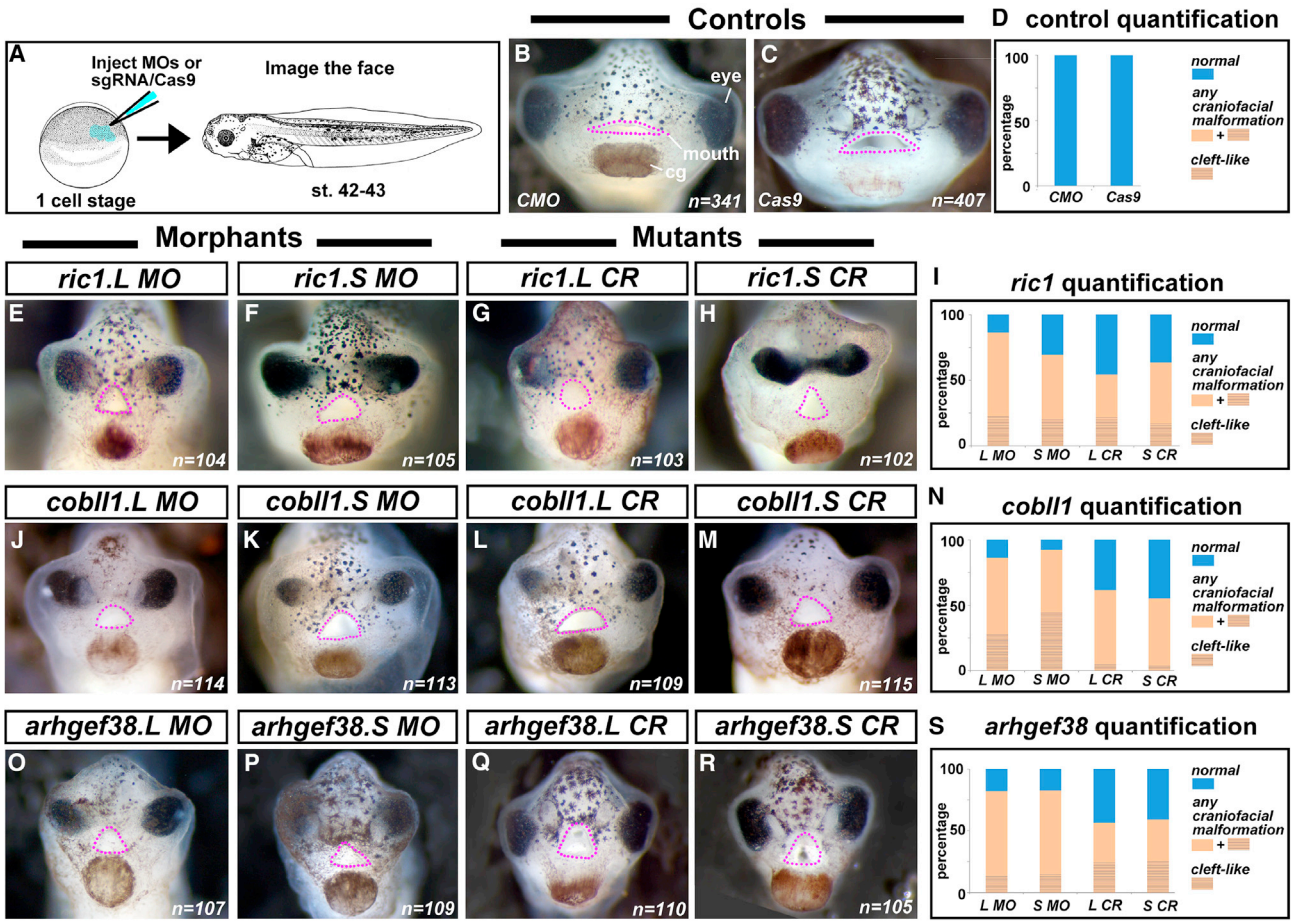

Figure 3. Knockdown of *Ric1*, *Cobll1*, and *Arhgef38* cause craniofacial malformations in *Xenopus laevis*

(A) Schematic showing injection of reagents at the one-cell stage followed by imaging at stage 42–43 (80–87 hpf). Xenopus illustrations © Natalya Zahn (2022).[74]

(B and C) Frontal view of the face of representative tadpoles injected with control morpholinos (MOs) or Cas9.

(D) Stacked bar graphs showing that 100% of the tadpoles were normal with respect to their craniofacial morphology.

(E–H) Frontal views of the faces of representative tadpoles injected with splice-blocking MOs or short guide RNA (sgRNA)/Cas9 targeting *ric1.L* and *ric1.S*, respectively (three biological replicates for each).

(I) Stacked bar graphs showing the percentage of the tadpoles that had normal faces, had craniofacial malformations, or triangular mouths that appeared cleft-like.

(J–M) Frontal views of the faces of representative tadpoles injected with MOs or sgRNA/Cas9 targeting *cobll1.L* and *cobll1.S* (two biological replicates for each).

(N) Stacked bar graphs showing the percentage of the tadpoles that had normal faces, had craniofacial malformations, or triangular mouths that appeared cleft-like.

(O–R) Frontal views of the faces of representative tadpoles injected with MOs or sgRNA/Cas9 targeting *arhgef38.L* and *arhgef38.S* (two biological replicates for each).

(S) Stacked bar graphs showing the percentage of the tadpoles that had normal faces, had craniofacial malformations, or triangular mouths that appeared cleft-like. The tadpole mouth opening is outlined in pink dots. Numbers of tadpoles are reported in the bottom right corner. CMO, control morpholino; CR, CRISPR-Cas9; st, stage; cg, cement gland; L, *Xenopus laevis* L homeolog; S, *Xenopus laevis* S homeolog.

faces and smaller eyes (Figures 3J–3N). Of the craniofacial malformations observed in these tadpoles, a portion had cleft-like triangular-like shaped mouths (*cobll1.L* MO = 27.2%, *cobll1.S* MO = 45.1%, *cobll1.L* CR = 5.5%, *cobll1.S* CR = 3.5%, Figures 3J–3N). Notable was the lower penetrance of mouth shape defects in the *cobll1* mutants compared with their morphant counterparts. Such lower penetrance has been observed in mosaic F0 *X. laevis* mutants when a large field of mutant cells is required to observe a phenotype.[75] Further mechanistic studies would be necessary to uncover the reason for lower penetrance of the craniofacial phenotype in *cobll1* mutants.

The *arhgef38.L* and *arhgef38.S* morphants (82.24% and 82.57%, respectively) and mutants (56.36% and 59.05%, respectively) had craniofacial defects which again included narrower faces and smaller eyes (Figures 3O–3S). Of the craniofacial malformations observed in these tadpoles, a subset also had cleft-like triangular shaped mouths (*arhgef38.L* MO = 13.08%, *arhgef38.S* MO = 14.68%, *arhgef38.L* CR = 24.55%, *arhgef38.S* CR = 25.71%, Figures 3O–3S). It is

important to note that in these knockdown experiments, the range of defects was similar across both homeologs of each gene suggesting that they did not have distinctive roles in the embryo, and that both homeologs contributed to craniofacial development (Figures S6–S8). Further, we noted striking similarities as well as a similar spectrum of craniofacial morphology between morphants and mutants, suggesting that the malformations were not likely caused by off-target effects (Figures S6–S8). Intriguingly, this spectrum of moderate to severe craniofacial abnormalities recapitulates the phenotypes observed in our previous modeling of a clefting gene, *ism1*.[20]

Since the shape and structure of the face is determined in part by the cranial cartilages, these structures were examined in *arhgef38, ric1*, and *cobll1* morphants. Collagen labeling revealed a reduction in cartilages of the face and this reduction was more profound in tadpoles with severe craniofacial defects (Figure S9). In particular, we observed defects in the ceratobranchial, ceratohyal, Meckel's, and trabecular cartilages in each of our morphant groups, in addition to other less common cartilage defects. Importantly, a reduction in these elements in the morphants is consistent with the expression of *arhgef38, ric1,* and *cobll1* in the branchial arches, precursors of the cranial skeleton.

As an additional test of whether the three candidate genes contribute to craniofacial development, we examined phenotypes in *Danio rerio* (*D. rerio*) larvae injected with sgRNAs targeting each gene (along with Cas9) at the single-cell stage (Table S10). Since *cobll1* has two paralogs (*cobll1a*, *cobll1b*) in *D. rerio*, we chose *cobll1b* for functional analysis given its reported expression in head regions during development,[76] in addition to its higher conservation to human *COBLL1*. We confirmed the efficacy of all sgRNAs using high-resolution melt analysis on eight individually injected embryos (see material and methods). The resulting embryos are expected to be genetically mosaic with a subset of cells exhibiting bi-allelic mutation of the targeted gene.[77] We examined the sgRNA plus Cas9 protein injected (mutant) embryos at 48 h postfertilization (hpf) for gross morphological defects of the head, including edema, ectopic blood in the fourth ventricle, and abnormal head shape (Figure S10A). More than 10% of mutant embryos had gross head deformities at 48 hpf. Surviving larvae were fixed at 4 days postfertilization (dpf) and processed to reveal the cartilage, and then examined for phenotypic abnormalities (Figure S10A–S10F). At 4 dpf, larvae sorted earlier based on abnormal head shape revealed the characteristic phenotypes described in Figure S10A (*arhgef38*, disorganized arches, collapsed ceratohyal; *cobll1b*, small bent Meckel's and ceratohyal arches). F0 larvae injected with sgRNAs targeting *ric1* recapitulated the hypoplastic Meckel's cartilage and abnormal ceratohyal cartilages phenotypes reported in *ric1* mutants,[73] and larvae injected with sgRNAs targeting *radil1* (which was deleted in only one individual in our cohort and used as a positive control) demonstrated the expected head deformities, edema, and cartilage defects.[78] Collectively, our knockdown studies in *X. laevis* and *D. rerio* support *ric1, cobll1*, and *arhgef38* as having important roles in midface development and being required for proper craniofacial development.

Discussion

Copy-number variants (CNVs) have been shown to contribute to orofacial clefting,[19–21,23–27] and prior studies have largely focused on either the detection of CNVs overlapping known Mendelian clefting loci, the utilization of segregation in large pedigrees, or common CNVs that contribute to clefting phenotypes. To our knowledge, no comparative analyses of CNV profiles between individuals with varying cleft subtypes have been performed to date, and a limited number of studies have pursued functional validations of CNVs in individuals with CL/P for confirmation of a gene's role in clefting. Our study furthers the investigation of the contribution of CNVs to CL/P by assessing CNVs occurring in individuals with NSCL/P versus SCL/P, as well as within NSCL/P subgroups. In addition, we report CNVs within these cohorts which overlap with known genomic disorder or Mendelian clefting loci. Finally, using an in-house analysis strategy to increase the likelihood of identifying clefting driver genes, we report the identification and validation in two vertebrate model organisms of NSCL/P candidates, *COBLL1*, *RIC1*, and *ARHGEF38,* which are overlapped by rare, recurrent deletions within our cohort of affected individuals.

By utilizing whole-genome tiling arrays and employing a series of stringent filters to identify high-confidence CNV calls in a cohort of 869 individuals with clefts from the Philippines (792 NSCL/P, 77 SCL/P), we assessed the total number of genomic CNVs, CNV load, and CNV burden within individuals with NSCL/P and SCL/P. No significant differences were observed between the two cohorts, suggesting that the overall number of CNVs and CNV content may not vary by cleft type. However, it is important to note that the inherent heterogeneity within the SCL/P cohort may be a confounding factor in drawing any strong conclusions about similarities or differences in global CNV profiles between individuals with NSCL/P versus those with SCL/P. In accordance with previously published work,[79–82] we observed that a higher proportion of the detected CNVs regardless of cleft type were gains rather than losses. This can likely be attributed to the fact that gains are generally considered to be better tolerated in the population than deletions.[83] We also note that the largest CNVs occurring within individuals are more often gains, and that the largest deletions occurring within individuals with CL/P (regardless of cleft type) are more frequently <500 kb (Figure S3). Of note, we observed a slight, qualitative increase in gains of 2–3 Mb in individuals with NSCL/P (18%) versus individuals with SCL/P (14%), and an elevated percent of individuals with SCL/P (22%) harboring gains 300–400 kb compared to

those with NSCL/P (10.9%). Additional investigations which consider the genetic content of these regions and frequency of each CNV event are needed to determine whether there is clinical relevance to these findings.

Individuals with NSCL/P may be subcategorized by cleft type (cleft lip only [CL], cleft palate only [CPO], and cleft lip and palate [CLP]; listed from least to most prevalent), sex (occurring twice as frequently in males as females), cleft laterality (unilateral clefts versus the less common bilateral clefting), and cleft sidedness (with left-sided clefts occurring at a higher frequency than right-sided clefts).[9] Recently, genetic modifiers significantly associated with these subgroups have been discovered and suggest a genetic basis for this phenotypic heterogeneity.[8,84] Stratification of our NSCL/P cohort by these subgroups demonstrated no difference in number of CNVs, CNV load, or CNV burden, suggesting that there is no direct correlation between an individual's genomic CNV profile and clefting epidemiology.

The assessment of regions previously implicated in genomic disorders resulted in the identification of known pathogenic CNVs in 28 of the individuals with CL/P in our study (Tables 2, S2, and S3). Six individuals had alterations of the 22q11.2 locus which has been associated with 22q11.2 deletion syndrome (also known as velocardiofacial or DiGeorge syndrome) and 22q11.2 duplication syndrome (MIM: 608363).[54] Recurrent CNV gains and losses of this region are mediated by flanking low-copy repeats (LCRs) resulting in either a common 2.54 Mb CNV or smaller atypical or nested copy-number changes. All four individuals from our cohort carrying deletions of this region had smaller ~1.3 to ~1.6 Mb deletions, and the two duplications were ~1.2 and ~1.3 Mb in size. The reported phenotypic features of individuals with these syndromes most commonly include congenital heart defects, palatal anomalies (including CP), immune deficiency, hearing loss, characteristic facial features, and learning difficulties.[54] The majority of our individuals with CN alterations of this region (5/6) reportedly had NSCPO, three of whom also had CL (two unilateral; one bilateral). However, it is important to note that the phenotyping of the individuals from the Philippines was time limited, and investigators were unable to assess cardiac or other internal abnormalities. In addition, the parents of these individuals were not evaluated, so it is possible that features consistent with 22q11.2 deletion syndrome were missed. These data could also suggest that non-syndromic clefting may also be associated with smaller CNVs of this region or that individuals who present with clefting as the sole observable phenotype may, in fact, have later-onset syndromic findings, or subtle features requiring additional clinical testing (such as echocardiogram or behavioral assessment) for detection.

Two individuals each had ~1.3 Mb deletions of 3q29, ~545 kb deletions of 16p11.2 (MIM: 611913) and alterations of the 1q21.1 thrombocytopenia-absent radius (TAR) susceptibility locus (one individual with an ~413 kb deletion and one individual with an ~382 kb duplication). Recurrent, ~1.6 Mb deletions of 3q29 are associated with neurodevelopmental findings including intellectual disability, autism spectrum disorder (ASD), and attention-deficit/hyperactivity disorder, as well as failure to thrive, patent ductus arteriosus, gastrointestinal concerns, and others.[85] Rarely, CL/P[55,56] or submucous CPO[86] have been reported in individuals with this syndrome. Notably, both of our individuals with 3q29 deletions had non-syndromic, unilateral clefts, suggesting that smaller deletions of this region may be a rare cause of NSCLP, or that additional syndromic features of individuals with deletions within the 3q29 microdeletion syndrome region may be subtle and easily overlooked during routine clinical phenotyping.

The recurrent ~593 kb microdeletions overlapping 16p11.2 are associated with variable delayed language development, learning difficulties, and/or ASD.[87] Although CL/P is not considered part of the primary clinical phenotype, CPO has been reported in individuals with microdeletions or microduplications of 16p11.2, and CLP has also been observed in individuals with recurrent 16p11.2 microdeletions.[88] Microdeletions of this locus were observed in our cohort in one individual with NSCPO and one with NSCLP. Due to the variable penetrance of this microdeletion syndrome[87] as well as the rarity of CL/P reported in affected individuals, it is unclear whether this microdeletion is contributing to the observed clefting phenotypes or if they are of a separate etiology.

Finally, ~200 kb recurrent deletions of the TAR locus at 1q21.1 have been reported in individuals with bilateral absent radii and thrombocytopenia in addition to other variable clinical features.[89] Proximal microdeletions and microduplications, as well as larger microdeletions involving this locus (including a 1.7 Mb interstitial deletion), have been reported in individuals with SCLP,[90–92] but to our knowledge NSCLP has not been associated with a microduplication or microdeletion of this region. Of note, we also identified several singleton gains and losses overlapping known genomic disorder loci, as well as a 14.4 Mb gain of 13q25 in one individual diagnosed with NSCLP (Tables S2 and S3).

Intriguingly, 11 individuals (4 SCLP with bilateral clefts; 4 NSCL; 3 NSCLP) had ~80 kb duplications of the *HOXD* cluster. Although variants within this region have been previously implicated in limb defects, to our knowledge one person has been reported with a cleft, resulting from a translocation disrupting this cluster.[93] Although additional work within a larger, high-powered study is needed to better understand the clinical relevance of CNVs within this region, connections between *Hox* loci and craniofacial development have been well documented during mouse embryogenesis. For example, in both the hindbrain and in CNCCs, *Hox* genes are expressed in a nested fashion to combinatorially specify regional properties of the head,[94–96] and perturbation studies have shown that *Hox* genes are required for neural crest cell specification,

migration, and differentiation.[95,97,98] However, the first branchial arch (which patterns the maxilla) appears to be devoid of *Hox* expression,[99] complicating interpretation of the results. One potential explanation is that gains of *HOXD* result in ectopic expression of the genes within the first branchial arch, altering their developmental trajectory. More work is clearly needed to mechanistically connect *HOXD* cluster duplications to orofacial clefting.

We next conducted an analysis of CNVs in our cohort overlapping putative or known Mendelian clefting loci and identified deletions overlapping 227 clefting genes (123 candidate[36] and 160 clinically relevant,[37] with 56 overlapping between both lists; Table S4). After restricting by cohort and population frequency, we observed *TACR3* was overlapped with the highest number of deletions (eight individuals), followed by *TBX1* (five individuals), and *COMT, NEDD4L*, *SNAP29*, *STK11*, and *UFD1* (four individuals each). *TACR3* has been associated with autosomal-recessive congenital hypogonadotrophic hypogonadism with or without anosmia. To our knowledge, deletions of *TACR3* have not been observed in individuals with clefts outside of the present study. However, individuals with variants in other genes associated with hypogonadotrophic hypogonadism (such as *FGFR1* and *CHD7*) have been reported to have CL/P,[100–103] suggesting that alterations of genes involved in this disease pathway may also contribute to clefting phenotypes. *TBX1* falls within the 22q11.2 microdeletion syndrome locus.[54] Murine studies assessing the impact of *Tbx1* deletion in the palate demonstrated that knockout mice have a small mandible and tongue compared to wild-type controls, and that this loss resulted in the dysregulation of several genes previously implicated in cleft palate in humans, including *MYH3* and *NEB*.[104] Additional work has further established *TBX1* as part of the gene regulatory network required for palatal formation,[105–107] strongly supporting a role for dysregulation of this gene in cleft formation.

Of the genes that were deleted in four individuals in our cohort (*COMT, NEDD4L*, *SNAP29*, *STK11*, and *UFD1*), four of them (*NEDD4L*, *SNAP29, STK11*, and *UFD1*) have been observed in clinical syndromes. Variants in *NEDD4L* are associated with autosomal-dominant periventricular nodular heterotopia syndrome (MIM: 617201), which includes cleft palate, syndactyly, and neurodevelopmental delay.[108,109] Variants in *SNAP29* cause autosomal-recessive cerebral dysgenesis, neuropathy, ichthyosis, and keratoderma (CEDNIK [MIM: 609528]) syndrome,[110,111] have been more recently reported in association with Pelizaeus-Merzbacher-like disorder,[110] and result in variable yet complex phenotypic features (including CL/P) when observed in combination with 22q11.2 deletion.[112] Loss-of-function variants in *STK11* are associated with Peutz-Jeghers syndrome (PJS [MIM: 175200]), an autosomal-dominant cancer predisposition syndrome. To our knowledge, focal deletions encompassing *STK11* have not been reported in individuals with CL/P; however, a deletion of 19p13.3 encompassing *STK11* and neighboring genes was reported in an individual with submucosal cleft palate, mild developmental delay, and seizures.[113] As a result, upper and lower gastrointestinal screening was performed which identified polyps consistent with PJS. These data suggest that deletions within 19p13.3 may result in clefting phenotypes, and further support the recommendation that individuals with deletions overlapping *STK11* identified as an incidental finding would benefit from additional clinical screening. Finally, *UFD1* (also known as *UFD1L*) falls within the 22q11.2 recurrent deletion locus. The identification of four deletions in our cohort overlapping *UFD1* suggests that this gene may be an additional contributor to clefting phenotypes within this genomic disorder region.

For the next phase of our study, we turned to enrichment strategies to explore whether particular genomic regions, or gene pathways, showed statistically significant enrichments connected to clefting or craniofacial development. In order to identify genomic locations that might be enriched in our disease cohort, we performed an over-representation analysis of genes deleted in one or more individuals and at a greater frequency in individuals with clefts versus controls. This analysis revealed several chromosomal regions associated with syndromic CNV disorders that include clefting or other craniofacial anomalies (Tables S7A and S7B), including 1q21.1 and 22q11.21, and provide additional evidence that microdeletions of these regions are likely associated with clefting. We next sought to determine whether this list of deleted genes was enriched for pathways associated with craniofacial development using the IPA platform, including KEGG analysis. Although we were unable to find enriched pathways for recurrently deleted genes, numerous pathways were significantly enriched for singleton deleted genes or singleton deleted genes with haploinsufficiency scores of 10 or less (Tables S7C–S7H; corrected p values <0.05). Several of these pathways play key roles in CNCC function (WNT/beta-catenin signaling; EMT transition; adherens junction pathway; focal adhesion pathway; migration of cells; formation of cellular protrusions; organization of cytoskeleton; Ephrin B/Ephrin receptor signaling; integrin signaling; HIF1alpha signaling; apoptosis signaling; inhibition of matrix metalloproteases; role of osteoblasts, osteoclasts, and chondrocytes in rheumatoid arthritis), whose alteration is associated with neurocristopathies which includes clefting.[114–118] Collectively, these pathways are represented by a total of 54 genes identified as being deleted in our cohort. It is important to point out that several members in these pathways are shared between the other enriched pathways, suggesting that they likely play key roles in different aspects of craniofacial development and that their loss would be particularly deleterious. Finally, while not found in any significantly enriched IPA pathways, the Rho and Rab GTPase components that we have functionally validated as bona fide craniofacial patterning genes (*RIC1*, *ARHGEF38*, *COBLL1;* see below) all play roles in cellular processes that occur during CNCC

development, as does the clefting-associated *ISM1* gene that we previously identified in an earlier subset of the larger clefting cohort reported here.[20]

While statistical methodologies are important tools for associating genes and genomic regions with a particular disease process, functional validation studies in a suitable model organism are critical for proving a gene's involvement. The substantial amount of CNV data generated from our large cohort of individuals with CL/P allowed us to utilize an unconventional search strategy for identifying putative clefting loci. We made use of population genetics data supporting that rare variants are expected to drive disease phenotypes[119,120] and hypothesized that any genes which were rarely deleted in individuals with clefts yet were deleted with greater frequency in affected individuals versus control subjects might have a higher effect size (i.e., haploinsufficiency loci) and allow for functional validation. Indeed, we previously reported a proof-of-principle study on a subset of the current cohort (~140 individuals) which resulted in the identification of a Mendelian clefting locus, *ISM1*.[20]

Due to their connections to related biological pathways associated with CNC cellular dynamics (Rho, Rab GTPase signaling)[65–67] and because Rho and Rab signaling components have been associated with craniofacial anomalies,[68–73] we selected *RIC1, ARHGEF38,* and *COBLL1* for functional follow-up in *X. laevis* and *D. rerio*. Although these models do not form a secondary palate, conserved genetic networks regulate the formation of the face across vertebrates,[121–130] and the primary palate and the roof of the mouth in *X. laevis* are thought to be analogous to the primary and secondary palate in mammals.[123,131,132] Therefore, our experiments tested whether the genes are involved in craniofacial morphogenesis rather than specifically in palate formation.

First, we performed *in situ* hybridization of the three genes at three developmental stages in *X. laevis* and observed expression of each gene in the cranial region that will later give rise to structures forming the face (Figure S5), thus suggesting a potential role for *ric1, arhgef38*, and *cobll1* during craniofacial development. To further test this hypothesis, we targeted these genes for knockdown/mutation using morpholino (MO) and F0 CRISPR-Cas9 ablation, respectively, in *X. laevis*. Prior studies have demonstrated that malformation of the structures important for primary palate development results in a change in the shape of the mouth in *X. laevis* tadpoles. Specifically, the dorsal aspect of the mouth becomes narrower than the ventral aspect, creating a more triangular shape which can appear as a median oral cleft.[123] Observation and imaging of *X. laevis* tadpoles injected with MOs and sgRNAs targeting each candidate gene resulted in the majority of injected tadpoles developing craniofacial defects (Figures 3I, 3N, and 3S). Such defects included midface hypoplasia and a reduction in cranial cartilages. Further, a portion of tadpoles with malformations affecting the craniofacial region also had malformed mouths resembling a median cleft in the primary palate, similar to what has been observed in knockdowns in *X. laevis* of other genes and pathways associated with orofacial clefting.[20,123]

To determine whether the role of these genes in craniofacial morphogenesis was evolutionarily conserved, we performed additional loss-of-function studies using CRISPR-Cas9 ablation in *D. rerio*. Although an abnormal craniofacial skeleton is a relatively frequent finding in *D. rerio* models,[130] we did observe characteristic changes in the craniofacial skeleton for *arhgef38*, *cobll1b*, and *ric1* mutants, and we replicated the previously reported findings of reduced Meckel's cartilage with CRISPR-Cas9 mutagenesis of *radil1* (which was deleted in one individual with CLP in our cohort) as an internal control (Figure S10).[78] Collectively, these results indicate that *ric1, arhgef38*, and *cobll1* are required for craniofacial morphogenesis, and our data in developmental models suggest that decreased function of these three genes in humans could result in craniofacial birth defects such as cleft lip and palate.

Intriguingly, our top three candidate genes are involved in either the Rho or Rab GTPase signaling pathways, which have been implicated in a wide array of cellular dynamics including the formation of adhesion junctions, actin organization, cell division, cell migration, and membrane trafficking.[65–67] Notably, a recent paper demonstrated the importance of actomyosin dynamics in secondary palate formation whereby midline epithelial seam cells are removed through actomyosin-dependent streaming migration of epithelial trails and islands in order to allow confluence of mesenchymal secondary palate cells.[133] In addition, several pathway components of Rho or Rab signaling have been associated with orofacial clefting and other craniofacial anomalies.[68–73] *COBLL1* is a WH2 domain containing protein which is involved in F-actin binding and filament formation,[134] and genes involved in actin cytoskeletal formation and organization have been previously associated with NSCL/P[135,136] and some forms of SCL/P.[137] Carroll and colleagues identified *Cobll1* in mice due to its sequence similarity to *Cordon-bleu* (*Cobl*), yet described nonoverlapping expression patterns between the two genes.[138] While mouse *Cobl* shows strong expression in the neural tube, *Cobll1* is distinctly expressed in the first branchial arch (its first embryonic region of expression, which gives rise to the maxilla and mandible), branchial clefts, and nasal placodes.[139] These structures are well conserved among vertebrates as regions critical for craniofacial development. *Cobll1* encodes a Rho GTPase signaling effector that is expressed in *Xenopus* CNCCs[140] and in the region of the branchial arches (this report), with Rho GTPase signaling being required for both cell division and migration.[64] Collectively, these data implicate *Cobll1* in developmental events associated with craniofacial development.[141]

The *Drosophila* ortholog of *RIC1* was originally identified as a gene important for expression of *N-Cadherin* (*CDH2* [MIM: 114020]) within photoreceptor cell synapses in

Drosophila[142] and is a binding partner of *Rab6* whose function is tightly regulated by guanine nucleotide exchange factors (GEFs) and GTPase activating proteins (GAPs) within the Rho signaling pathway.[142–147] While our work implicates *RIC1* in NSCL/P, *RIC1* variants were recently shown to be responsible for a Mendelian syndrome that includes cataract, tooth abnormality, intellectual disability, facial dysmorphism, attention-deficit hyperactivity disorder, and clefting (CATIFA [MIM: 618761]),[73] in addition to a related syndrome including brain atrophy, microcephaly, and CL/P.[148] Work in *D. rerio* has demonstrated a role for *ric1* in normal skeletogenesis, where it is required for procollagen secretion from craniofacial chondrocytes, with mutant animals presenting with craniofacial anomalies including flattened heads and reduced jaw size.[73] Similar to zebrafish, *Xenopus* with decreased *ric1* also had craniofacial defects and reduced jaw cartilages. In *Xenopus*, the *RIC1* paralog Ric-8A is required for CNCC migration, failure of which leads to craniofacial anomalies including clefting,[70,141] and *Ric1* itself is expressed in CNCCs[140] and in the region of the branchial arches (this report). Collectively, these data provide evidence for a role of *RIC1* in craniofacial patterning.

Although to our knowledge the function of *ARHGEF38* has never been investigated, as a member of the GEF family of GTPase regulatory proteins it is likely involved in the regulation of cellular dynamics in an interplay with GAPs, and both GEFs and GAPs are known to control cell migration.[63,64] Consistent with this hypothesis, a recent report correlated high levels of *ARHGEF38* with aggressive prostate cancer, where the authors proposed a role for *ARHGEF38* in promoting prostate cancer cell migration.[149] Moreover, similar to *RIC1* and *COBLL1*, *ARHGEF38* is expressed in CNCCs in *Xenopus*[140] as well as other head structures, while pathogenic variants in GEF-associated GAPs have been previously identified in humans with NSCL/P,[71] thus suggesting *ARHGEF38* may be an additional player in this signaling pathway.

We conclude that although this study supports a role for deletions overlapping *COBLL1, RIC1*, and *ARHGEF38* as causal variants for CL/P formation, additional studies using larger cohorts are needed to more fully define the contribution of rare and common CNVs to clefting. Future investigations within our cohort and others could include utilizing sequencing data to determine whether sequence-level variants are present within the non-deleted allele, and additional functional studies should consider whether any CNVs detected within our cohort disrupt regulatory regions of true causative loci which themselves are not encompassed by the CNVs. These investigations, in combination with familial studies to assess segregation and penetrance, will help to further our understanding of the contribution of CNVs in CL/P formation.

Data and code availability

Plasmids are available upon request. The data from the individuals with CL/P discussed in this publication have been deposited in NCBI's Gene Expression Omnibus.[1] The accession number for the data reported in this paper is GEO: GSE212296 (GEO: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE212296). The control data discussed in this publication are based on data available within the dbGaP website under phs000523.v1.p1 (dbGaP: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000523.v1.p1).

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2022.11.012.

Acknowledgments

We are ever grateful to the families who participated in this research and the many nurses, doctors, dentists, speech pathologists, and others who provided care both in the U.S. and through Operation Smile in the Philippines. Edith Villanueva and William and Kathy Magee deserve particular thanks for their years of dedication to this project at Iowa and through their non-profits Project Hope and Operation Smile. Special thanks to the team of skilled undergraduates who made this project possible including Claire Olson, Maddie Lorentzen, Mason LaMarche, Regan Benbow, Nick Stange, and Alana Jones. In addition, we thank Manak Lab research assistant Josh Wankum for processing arrays. We also thank Dr. Elizabeth Leslie for her invaluable input throughout this project. This work was supported by National Institutes of Health grants to J.R.M. (R01 DE-021071), D.W.H. (R01 GM-083999), J.C.M. (R37 DE-08559), R.A.C. (R01 DE-027983), and L.A.L. (T32 GM-008629).

Declaration of interests

The authors declare no competing interests.

Received: June 2, 2022
Accepted: November 18, 2022
Published: December 8, 2022

Web resources

dbGaP, https://www.ncbi.nlm.nih.gov/projects/gap/
DECIPHER, https://decipher.sanger.ac.uk/
Gene Expression Omnibus, https://www.ncbi.nlm.nih.gov/geo/
gnomAD, gnomad.broadinstitute.org
Mouse Genome Database (MGI), www.informatics.jax.org
National Center for Biotechnology Information, https://www.ncbi.nlm.nih.gov/
OMIM, https://www.omim.org/
QIAGEN IPA, https://digitalinsights.qiagen.com/IPA
VassarStats, http://vassarstats.net/
Xenbase, www.xenbase.org
The Zebrafish Information Network (ZFIN), zfin.org

References

1. Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. *30*, 207–210. https://doi.org/10.1093/nar/30.1.207.

2. Leslie, E.J., and Marazita, M.L. (2013). Genetics of cleft lip and cleft palate. Am. J. Med. Genet. C Semin. Med. Genet. *163C*, 246–258. https://doi.org/10.1002/ajmg.c.31381.
3. Beaty, T.H., Marazita, M.L., and Leslie, E.J. (2016). Genetic factors influencing risk to orofacial clefts: today's challenges and tomorrow's opportunities. F1000Res. *5*, 2800. https://doi.org/10.12688/f1000research.9503.1.
4. Carlson, J.C., Anand, D., Butali, A., Buxo, C.J., Christensen, K., Deleyiannis, F., Hecht, J.T., Moreno, L.M., Orioli, I.M., Padilla, C., et al. (2019). A systematic genetic analysis and visualization of phenotypic heterogeneity among orofacial cleft GWAS signals. Genet. Epidemiol. *43*, 704–716. https://doi.org/10.1002/gepi.22214.
5. Mossey, P.A., Little, J., Munger, R.G., Dixon, M.J., and Shaw, W.C. (2009). Cleft lip and palate. Lancet *374*, 1773–1785. https://doi.org/10.1016/S0140-6736(09)60695-4.
6. Gundlach, K.K.H., and Maus, C. (2006). Epidemiological studies on the frequency of clefts in Europe and worldwide. J. Cranio-Maxillo-Fac. Surg. *34*, 1–2. https://doi.org/10.1016/S1010-5182(06)60001-2.
7. Dixon, M.J., Marazita, M.L., Beaty, T.H., and Murray, J.C. (2011). Cleft lip and palate: understanding genetic and environmental influences. Nat. Rev. Genet. *12*, 167–178. https://doi.org/10.1038/nrg2933.
8. Carlson, J.C., Taub, M.A., Feingold, E., Beaty, T.H., Murray, J.C., Marazita, M.L., and Leslie, E.J. (2017). Identifying genetic sources of phenotypic heterogeneity in orofacial clefts by targeted sequencing. Birth Defects Res. *109*, 1030–1038. https://doi.org/10.1002/bdr2.23605.
9. Genisca, A.E., Frías, J.L., Broussard, C.S., Honein, M.A., Lammer, E.J., Moore, C.A., Shaw, G.M., Murray, J.C., Yang, W., Rasmussen, S.A.; and National Birth Defects Prevention Study (2009). Orofacial clefts in the national birth defects prevention study, 1997-2004. Am. J. Med. Genet. *149A*, 1149–1158. https://doi.org/10.1002/ajmg.a.32854.
10. Leslie, E.J., Liu, H., Carlson, J.C., Shaffer, J.R., Feingold, E., Wehby, G., Laurie, C.A., Jain, D., Laurie, C.C., Doheny, K.F., et al. (2016). A Genome-wide association study of nonsyndromic cleft palate identifies an etiologic missense variant in GRHL3. Am. J. Hum. Genet. *98*, 744–754. https://doi.org/10.1016/j.ajhg.2016.02.014.
11. Mangold, E., Böhmer, A.C., Ishorst, N., Hoebel, A.K., Gültepe, P., Schuenke, H., Klamt, J., Hofmann, A., Gölz, L., Raff, R., et al. (2016). Sequencing the GRHL3 coding region reveals rare truncating mutations and a common susceptibility variant for nonsyndromic cleft palate. Am. J. Hum. Genet. *98*, 755–762. https://doi.org/10.1016/j.ajhg.2016.02.013.
12. Greenway, S.C., Pereira, A.C., Lin, J.C., DePalma, S.R., Israel, S.J., Mesquita, S.M., Ergul, E., Conta, J.H., Korn, J.M., McCarroll, S.A., et al. (2009). De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot. Nat. Genet. *41*, 931–935. https://doi.org/10.1038/ng.415.
13. Itsara, A., Cooper, G.M., Baker, C., Girirajan, S., Li, J., Absher, D., Krauss, R.M., Myers, R.M., Ridker, P.M., Chasman, D.I., et al. (2009). Population analysis of large copy number variants and hotspots of human genetic disease. Am. J. Hum. Genet. *84*, 148–161. https://doi.org/10.1016/j.ajhg.2008.12.014.
14. Rosenfeld, J.A., Ballif, B.C., Torchia, B.S., Sahoo, T., Ravnan, J.B., Schultz, R., Lamb, A., Bejjani, B.A., and Shaffer, L.G. (2010). Copy number variations associated with autism spectrum disorders contribute to a spectrum of neurodevelopmental disorders. Genet. Med. *12*, 694–702. https://doi.org/10.1097/GIM.0b013e3181f0c5f3.
15. Bassuk, A.G., Muthuswamy, L.B., Boland, R., Smith, T.L., Hulstrand, A.M., Northrup, H., Hakeman, M., Dierdorff, J.M., Yung, C.K., Long, A., et al. (2013). Copy number variation analysis implicates the cell polarity gene glypican 5 as a human spina bifida candidate gene. Hum. Mol. Genet. *22*, 1097–1111. https://doi.org/10.1093/hmg/dds515.
16. Hilger, A.C., Dworschak, G.C., and Reutter, H.M. (2020). Lessons learned from CNV analysis of major birth defects. Int. J. Mol. Sci. *21*, E8247. https://doi.org/10.3390/ijms21218247.
17. Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V., et al. (2011). A copy number variation morbidity map of developmental delay. Nat. Genet. *43*, 838–846. https://doi.org/10.1038/ng.909.
18. Girirajan, S., Campbell, C.D., and Eichler, E.E. (2011). Human copy number variation and complex genetic disease. Annu. Rev. Genet. *45*, 203–226. https://doi.org/10.1146/annurev-genet-102209-163544.
19. Younkin, S.G., Scharpf, R.B., Schwender, H., Parker, M.M., Scott, A.F., Marazita, M.L., Beaty, T.H., and Ruczinski, I. (2014). A genome-wide study of de novo deletions identifies a candidate locus for non-syndromic isolated cleft lip/palate risk. BMC Genet. *15*, 24. https://doi.org/10.1186/1471-2156-15-24.
20. Lansdon, L.A., Darbro, B.W., Petrin, A.L., Hulstrand, A.M., Standley, J.M., Brouillette, R.B., Long, A., Mansilla, M.A., Cornell, R.A., Murray, J.C., et al. (2018). Identification of Isthmin 1 as a novel clefting and craniofacial patterning gene in humans. Genetics *208*, 283–296. https://doi.org/10.1534/genetics.117.300535.
21. Cai, Y., Patterson, K.E., Reinier, F., Keesecker, S.E., Blue, E., Bamshad, M., and Haddad, J., Jr. (2017). Copy number changes identified using whole exome sequencing in nonsyndromic cleft lip and palate in a honduran population. Birth Defects Res. *109*, 1257–1267. https://doi.org/10.1002/bdr2.1063.
22. Szczałuba, K., Nowakowska, B.A., Sobecka, K., Smyk, M., Castaneda, J., Dudkiewicz, Z., Kutkowska-Kaźmierczak, A., Sąsiadek, M.M., Śmigiel, R., and Bocian, E. (2015). High-resolution array comparative genomic hybridization utility in polish newborns with isolated cleft lip and palate. Neonatology *107*, 173–178. https://doi.org/10.1159/000368878.
23. Conte, F., Oti, M., Dixon, J., Carels, C.E.L., Rubini, M., and Zhou, H. (2016). Systematic analysis of copy number variants of a large cohort of orofacial cleft patients identifies candidate genes for orofacial clefts. Hum. Genet. *135*, 41–59. https://doi.org/10.1007/s00439-015-1606-x.
24. Younkin, S.G., Scharpf, R.B., Schwender, H., Parker, M.M., Scott, A.F., Marazita, M.L., Beaty, T.H., and Ruczinski, I. (2015). A genome-wide study of inherited deletions identified two regions associated with nonsyndromic isolated oral clefts. Birth Defects Res. A Clin. Mol. Teratol. *103*, 276–283. https://doi.org/10.1002/bdra.23362.
25. Shi, M., Mostowska, A., Jugessur, A., Johnson, M.K., Mansilla, M.A., Christensen, K., Lie, R.T., Wilcox, A.J., and Murray, J.C. (2009). Identification of microdeletions in candidate genes for cleft lip and/or palate. Birth defects research. Birth Defects Res. A Clin. Mol. Teratol. *85*, 42–51. https://doi.org/10.1002/bdra.20571.

26. Osoegawa, K., Vessere, G.M., Utami, K.H., Mansilla, M.A., Johnson, M.K., Riley, B.M., L'Heureux, J., Pfundt, R., Staaf, J., van der Vliet, W.A., et al. (2008). Identification of novel candidate genes associated with cleft lip and palate using array comparative genomic hybridisation. J. Med. Genet. *45*, 81–86. https://doi.org/10.1136/jmg.2007.052191.
27. Lei, T.Y., Wang, H.T., Li, F., Cui, Y.Q., Fu, F., Li, R., and Liao, C. (2016). Application of high resolution SNP arrays in patients with congenital oral clefts in south China. J. Genet. *95*, 801–809.
28. Cao, Y., Li, Z., Rosenfeld, J.A., Pursley, A.N., Patel, A., Huang, J., Wang, H., Chen, M., Sun, X., Leung, T.Y., et al. (2016). Contribution of genomic copy-number variations in prenatal oral clefts: a multicenter cohort study. Genet. Med. *18*, 1052–1055. https://doi.org/10.1038/gim.2015.216.
29. Maarse, W., Rozendaal, A.M., Pajkrt, E., Vermeij-Keers, C., Mink van der Molen, A.B., and van den Boogaard, M.J.H. (2012). A systematic review of associated structural and chromosomal defects in oral clefts: when is prenatal genetic analysis indicated? J. Med. Genet. *49*, 490–498. https://doi.org/10.1136/jmedgenet-2012-101013.
30. Murray, J.C., Daack-Hirsch, S., Buetow, K.H., Munger, R., Espina, L., Paglinawan, N., Villanueva, E., Rary, J., Magee, K., and Magee, W. (1997). Clinical and epidemiologic studies of cleft lip and palate in the Philippines. Cleft Palate-Craniofacial J. *34*, 7–10. https://doi.org/10.1597/1545-1569(1997)034<0007:CAESOC>2.3.CO;2.
31. Brophy, P.D., Alasti, F., Darbro, B.W., Clarke, J., Nishimura, C., Cobb, B., Smith, R.J., and Manak, J.R. (2013). Genome-wide copy number variation analysis of a Branchio-oto-renal syndrome cohort identifies a recombination hotspot and implicates new candidate genes. Hum. Genet. *132*, 1339–1350. https://doi.org/10.1007/s00439-013-1338-8.
32. Lange, L.A., Croteau-Chonka, D.C., Marvelle, A.F., Qin, L., Gaulton, K.J., Kuzawa, C.W., McDade, T.W., Wang, Y., Li, Y., Levy, S., et al. (2010). Genome-wide association study of homocysteine levels in Filipinos provides evidence for CPS1 in women and a stronger MTHFR effect in young adults. Hum. Mol. Genet. *19*, 2050–2058. https://doi.org/10.1093/hmg/ddq062.
33. Wu, Y., Li, Y., Lange, E.M., Croteau-Chonka, D.C., Kuzawa, C.W., McDade, T.W., Qin, L., Curocichin, G., Borja, J.B., Lange, L.A., et al. (2010). Genome-wide association study for adiponectin levels in Filipino women identifies CDH13 and a novel uncommon haplotype at KNG1-ADIPOQ. Hum. Mol. Genet. *19*, 4955–4964. https://doi.org/10.1093/hmg/ddq423.
34. Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F.A., Hakonarson, H., and Bucan, M. (2007). PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. *17*, 1665–1674. https://doi.org/10.1101/gr.6861907.
35. Pruitt, K.D., Brown, G.R., Hiatt, S.M., Thibaud-Nissen, F., Astashyn, A., Ermolaeva, O., Farrell, C.M., Hart, J., Landrum, M.J., McGarvey, K.M., et al. (2014). RefSeq: an update on mammalian reference sequences. Nucleic Acids Res. *42*, D756–D763. https://doi.org/10.1093/nar/gkt1114.
36. Funato, N., and Nakamura, M. (2017). Identification of shared and unique gene families associated with oral clefts. Int. J. Oral Sci. *9*, 104–109. https://doi.org/10.1038/ijos.2016.56.
37. Bishop, M.R., Diaz Perez, K.K., Sun, M., Ho, S., Chopra, P., Mukhopadhyay, N., Hetmanski, J.B., Taub, M.A., Moreno-Uribe, L.M., Valencia-Ramirez, L.C., et al. (2020). Genome-wide enrichment of De Novo coding mutations in orofacial cleft trios. Am. J. Hum. Genet. *107*, 124–136. https://doi.org/10.1016/j.ajhg.2020.05.018.
38. Bult, C.J., Blake, J.A., Smith, C.L., Kadin, J.A., Richardson, J.E.; and Mouse Genome Database Group (2019). Mouse Genome Database (MGD) 2019. Nucleic Acids Res. *47*, D801–D806. https://doi.org/10.1093/nar/gky1056.
39. Ruzicka, L., Howe, D.G., Ramachandran, S., Toro, S., Van Slyke, C.E., Bradford, Y.M., Eagle, A., Fashena, D., Frazer, K., Kalita, P., et al. (2019). The zebrafish information network: new support for non-coding genes, richer Gene Ontology annotations and the alliance of genome resources. Nucleic Acids Res. *47*, D867–D873. https://doi.org/10.1093/nar/gky1090.
40. Karimi, K., Fortriede, J.D., Lotay, V.S., Burns, K.A., Wang, D.Z., Fisher, M.E., Pells, T.J., James-Zorn, C., Wang, Y., Ponferrada, V.G., et al. (2018). Xenbase: a genomic, epigenomic and transcriptomic model organism database. Nucleic Acids Res. *46*, D861–D868. https://doi.org/10.1093/nar/gkx936.
41. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141, 456 humans. Nature *581*, 434–443. https://doi.org/10.1038/s41586-020-2308-7.
42. Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R.M., and Carter, N.P. (2009). DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources. Am. J. Hum. Genet. *84*, 524–533. https://doi.org/10.1016/j.ajhg.2009.03.010.
43. Gerstner, N., Kehl, T., Lenhof, K., Müller, A., Mayer, C., Eckhart, L., Grammes, N.L., Diener, C., Hart, M., Hahn, O., et al. (2020). GeneTrail 3: advanced high-throughput enrichment analysis. Nucleic Acids Res. *48*, W515–W520. https://doi.org/10.1093/nar/gkaa306.
44. Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. Ann. Statist. *29*, 1165–1188. https://doi.org/10.1214/aos/1013699998.
45. Sive, H., Grainger, R., and Harland, R. (2000). Early Development of Xenopus laevis: A Laboratory Manual (Cold Spring Harbor Press).
46. Nieuwkoop, P.D.a.F., J. (1994). Normal Table of Xenopus Laevis (daudin): A Systematical and Chronological Survey of the Development from the Fertilized Egg till the End of Metamorphosis (Garland Publishers).
47. Hulstrand, A.M., and Houston, D.W. (2013). Regulation of neurogenesis by Fgf8a requires Cdc42 signaling and a novel Cdc42 effector protein. Dev. Biol. *382*, 385–399.
48. Labun, K., Montague, T.G., Krause, M., Torres Cleuren, Y.N., Tjeldnes, H., and Valen, E. (2019). CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. Nucleic Acids Res. *47*, W171–W174. https://doi.org/10.1093/nar/gkz365.
49. Truett, G.E., Heeger, P., Mynatt, R.L., Truett, A.A., Walker, J.A., and Warman, M.L. (2000). Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). Biotechniques *29*, 52–54. https://doi.org/10.2144/00291bm09.

50. Westerfield, M. (2000). The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish (Danio rerio), 4 Edition (University of Oregon Press).
51. Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., and Schilling, T.F. (1995). Stages of embryonic development of the zebrafish. Dev. Dyn. *203*, 253–310.
52. Kondo, S., Schutte, B.C., Richardson, R.J., Bjork, B.C., Knight, A.S., Watanabe, Y., Howard, E., de Lima, R.L.L.F., Daack-Hirsch, S., Sander, A., et al. (2002). Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. Nat. Genet. *32*, 285–289. https://doi.org/10.1038/ng985.
53. Peyrard-Janvid, M., Leslie, E.J., Kousa, Y.A., Smith, T.L., Dunnwald, M., Magnusson, M., Lentz, B.A., Unneberg, P., Fransson, I., Koillinen, H.K., et al. (2014). Dominant mutations in GRHL3 cause Van der Woude syndrome and disrupt oral periderm development. Am. J. Hum. Genet. *94*, 23–32. https://doi.org/10.1016/j.ajhg.2013.11.009.
54. McDonald-McGinn, D.M., Hain, H.S., Emanuel, B.S., and Zackai, E.H. (1999). 22q11.2 Deletion Syndrome. In GeneReviews, M.P. Adam, H.H. Ardinger, and R.A. Pagon, eds. (University of Washington).
55. Willatt, L., Cox, J., Barber, J., Cabanas, E.D., Collins, A., Donnai, D., FitzPatrick, D.R., Maher, E., Martin, H., Parnau, J., et al. (2005). 3q29 microdeletion syndrome: clinical and molecular characterization of a new syndrome. Am. J. Hum. Genet. *77*, 154–160. https://doi.org/10.1086/431653.
56. Petrin, A.L., Daack-Hirsch, S., L'Heureux, J., and Murray, J.C. (2011). A case of 3q29 microdeletion syndrome involving oral cleft inherited from a nonaffected mosaic parent: molecular analysis and ethical implications. Cleft Palate-Craniofacial J. *48*, 222–230. https://doi.org/10.1597/09-149.
57. Ingersoll, R.G., Hetmanski, J., Park, J.W., Fallin, M.D., McIntosh, I., Wu-Chou, Y.H., Chen, P.K., Yeow, V., Chong, S.S., Cheah, F., et al. (2010). Association between genes on chromosome 4p16 and non-syndromic oral clefts in four populations. Eur. J. Hum. Genet. *18*, 726–732. https://doi.org/10.1038/ejhg.2009.228.
58. Paradowska-Stolarz, A.M. (2014). Wolf-Hirschhorn syndrome (WHS) - literature review on the features of the syndrome. Adv. Clin. Exp. Med. *23*, 485–489. https://doi.org/10.17219/acem/24111.
59. Krämer, A., Green, J., Pollard, J., Jr., and Tugendreich, S. (2014). Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics *30*, 523–530. https://doi.org/10.1093/bioinformatics/btt703.
60. Kanehisa, M. (2019). Toward understanding the origin and evolution of cellular organisms. Protein Sci. *28*, 1947–1951. https://doi.org/10.1002/pro.3715.
61. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M. (2021). KEGG: integrating viruses and cellular organisms. Nucleic Acids Res. *49*, D545–D551. https://doi.org/10.1093/nar/gkaa970.
62. Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *28*, 27–30. https://doi.org/10.1093/nar/28.1.27.
63. Bos, J.L., Rehmann, H., and Wittinghofer, A. (2007). GEFs and GAPs: critical elements in the control of small G proteins. Cell *129*, 865–877. https://doi.org/10.1016/j.cell.2007.05.018.
64. Ridley, A.J. (2015). Rho GTPase signalling in cell migration. Curr. Opin. Cell Biol. *36*, 103–112. https://doi.org/10.1016/j.ceb.2015.08.005.
65. Villarroel-Campos, D., Bronfman, F.C., and Gonzalez-Billault, C. (2016). Rab GTPase signaling in neurite outgrowth and axon specification. Cytoskeleton *73*, 498–507. https://doi.org/10.1002/cm.21303.
66. Lawson, C.D., and Ridley, A.J. (2018). Rho GTPase signaling complexes in cell migration and invasion. J. Cell Biol. *217*, 447–457. https://doi.org/10.1083/jcb.201612069.
67. Nassari, S., Del Olmo, T., and Jean, S. (2020). Rabs in signaling and embryonic development. Int. J. Mol. Sci. *21*, E1064. https://doi.org/10.3390/ijms21031064.
68. Oshima-Nakayama, M., Yamada, A., Kurosawa, T., Aizawa, R., Suzuki, D., Saito, Y., Kassai, H., Sato, Y., Yamamoto, M., Shirota, T., et al. (2016). Cdc42 is crucial for facial and palatal formation during craniofacial development. Bone Rep. *5*, 1–6. https://doi.org/10.1016/j.bonr.2016.01.001.
69. Fortugno, P., Josselin, E., Tsiakas, K., Agolini, E., Cestra, G., Teson, M., Santer, R., Castiglia, D., Novelli, G., Dallapiccola, B., et al. (2014). Nectin-4 mutations causing ectodermal dysplasia with syndactyly perturb the rac1 pathway and the kinetics of adherens junction formation. J. Invest. Dermatol. *134*, 2146–2153. https://doi.org/10.1038/jid.2014.119.
70. Leal, J.I., Villaseca, S., Beyer, A., Toro-Tapia, G., and Torrejón, M. (2018). Ric-8A, a GEF for heterotrimeric G-proteins, controls cranial neural crest cell polarity during migration. Mech. Dev. *154*, 170–178. https://doi.org/10.1016/j.mod.2018.07.004.
71. Leslie, E.J., Mansilla, M.A., Biggs, L.C., Schuette, K., Bullard, S., Cooper, M., Dunnwald, M., Lidral, A.C., Marazita, M.L., Beaty, T.H., and Murray, J.C. (2012). Expression and mutation analyses implicate ARHGAP29 as the etiologic gene for the cleft lip with or without cleft palate locus identified by genome-wide association on chromosome 1p22. Birth defects research. Birth Defects Res. A Clin. Mol. Teratol. *94*, 934–942. https://doi.org/10.1002/bdra.23076.
72. Parada-Sanchez, M.T., Chu, E.Y., Cox, L.L., Undurty, S.S., Standley, J.M., Murray, J.C., and Cox, T.C. (2017). Disrupted IRF6-NME1/2 complexes as a cause of cleft lip/palate. J. Dent. Res. *96*, 1330–1338. https://doi.org/10.1177/0022034517723615.
73. Unlu, G., Qi, X., Gamazon, E.R., Melville, D.B., Patel, N., Rushing, A.R., Hashem, M., Al-Faifi, A., Chen, R., Li, B., et al. (2020). Phenome-based approach identifies RIC1-linked Mendelian syndrome through zebrafish models, biobank associations and clinical studies. Nat. Med. *26*, 98–109. https://doi.org/10.1038/s41591-019-0705-y.
74. Zahn, N., James-Zorn, C., Ponferrada, V.G., Adams, D.S., Grzymkowski, J., Buchholz, D.R., Nascone-Yoder, N.M., Horb, M., Moody, S.A., Vize, P.D., and Zorn, A.M. (2022). Normal table of xenopus development: a new graphical resource. Development *149*, dev200356. https://doi.org/10.1242/dev.200356.
75. Bharathan, N.K., and Dickinson, A.J.G. (2019). Desmoplakin is required for epidermal integrity and morphogenesis in the Xenopus laevis embryo. Dev. Biol. *450*, 115–131. https://doi.org/10.1016/j.ydbio.2019.03.010.
76. Thisse, B., Pflumio, S., Fürthauer, M., Loppin, B., Heyer, V., Degrave, A., Woehl, R., Lux, A., Steffan, T., Charbonnier, X.Q., and Thisse, C. (2001). Expression of the Zebrafish Genome during Embryogenesis. https://zfin.org/ZDB-PUB-010810-1#summary.
77. Li, M., Zhao, L., Page-McCaw, P.S., and Chen, W. (2016). Zebrafish genome engineering using the CRISPR-Cas9 system.

Trends Genet. *32*, 815–827. https://doi.org/10.1016/j.tig.2016.10.005.

78. Smolen, G.A., Schott, B.J., Stewart, R.A., Diederichs, S., Muir, B., Provencher, H.L., Look, A.T., Sgroi, D.C., Peterson, R.T., and Haber, D.A. (2007). A Rap GTPase interactor, RADIL, mediates migration of neural crest precursors. Genes Dev. *21*, 2131–2136. https://doi.org/10.1101/gad.1561507.
79. Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., and Lee, C. (2004). Detection of large-scale variation in the human genome. Nat. Genet. *36*, 949–951. https://doi.org/10.1038/ng1416.
80. Kidd, J.M., Cooper, G.M., Donahue, W.F., Hayden, H.S., Sampas, N., Graves, T., Hansen, N., Teague, B., Alkan, C., Antonacci, F., et al. (2008). Mapping and sequencing of structural variation from eight human genomes. Nature *453*, 56–64. https://doi.org/10.1038/nature06862.
81. Girirajan, S., Dennis, M.Y., Baker, C., Malig, M., Coe, B.P., Campbell, C.D., Mark, K., Vu, T.H., Alkan, C., Cheng, Z., et al. (2013). Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. Am. J. Hum. Genet. *92*, 221–237. https://doi.org/10.1016/j.ajhg.2012.12.016.
82. Girirajan, S., Johnson, R.L., Tassone, F., Balciuniene, J., Katiyar, N., Fox, K., Baker, C., Srikanth, A., Yeoh, K.H., Khoo, S.J., et al. (2013). Global increases in both common and rare copy number load associated with autism. Hum. Mol. Genet. *22*, 2870–2880. https://doi.org/10.1093/hmg/ddt136.
83. Lee, C., Iafrate, A.J., and Brothman, A.R. (2007). Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. Nat. Genet. *39*, S48–S54. https://doi.org/10.1038/ng2092.
84. Carlson, J.C., Standley, J., Petrin, A., Shaffer, J.R., Butali, A., Buxó, C.J., Castilla, E., Christensen, K., Deleyiannis, F.W.D., Hecht, J.T., et al. (2017). Identification of 16q21 as a modifier of nonsyndromic orofacial cleft phenotypes. Genet. Epidemiol. *41*, 887–897. https://doi.org/10.1002/gepi.22090.
85. Gladys Mulle, J., Gambello, M.J., Sanchez Russo, R., Murphy, M.M., Burrell, L., Klaiman, C., White, S., Saulnier, C.A., Walker, E.F., Cubells, J.F., et al. (2016). 3q29 Recurrent Deletion.. GeneReviews. https://pubmed.ncbi.nlm.nih.gov/27656750/.
86. Dasouki, M.J., Lushington, G.H., Hovanes, K., Casey, J., and Gorre, M. (2011). The 3q29 microdeletion syndrome: report of three new unrelated patients and in silico "RNA binding" analysis of the 3q29 region. Am. J. Med. Genet. *155A*, 1654–1660. https://doi.org/10.1002/ajmg.a.34080.
87. Miller, D.T., Chung, W., Nasir, R., Shen, Y., Steinman, K.J., Wu, B., and Hanson, E. (2009). 16p11.2 Recurrent Microdeletion. GeneReviews. https://www.ncbi.nlm.nih.gov/books/NBK11167/.
88. Shinawi, M., Liu, P., Kang, S.H.L., Shen, J., Belmont, J.W., Scott, D.A., Probst, F.J., Craigen, W.J., Graham, B.H., Pursley, A., et al. (2010). Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. J. Med. Genet. *47*, 332–341. https://doi.org/10.1136/jmg.2009.073015.
89. Toriello, H.V. (1993). Thrombocytopenia Absent Radius Syndrome. In GeneReviews. https://www.ncbi.nlm.nih.gov/books/NBK23758/.
90. Gamba, B.F., Zechi-Ceide, R.M., Kokitsu-Nakata, N.M., Vendramini-Pittoli, S., Rosenberg, C., Krepischi Santos, A.C.V., Ribeiro-Bicudo, L., and Richieri-Costa, A. (2016). Interstitial 1q21.1 microdeletion is associated with severe skeletal anomalies, dysmorphic face and moderate intellectual disability. Mol. Syndromol. *7*, 344–348. https://doi.org/10.1159/000450971.
91. Midro, A., Hubert, E., Preferansow, J., and Iwaszkiewicz-Pawłowska, A. (1993). TAR syndrome with orofacial clefting. Genet. Couns. *4*, 187–192.
92. Rosenfeld, J.A., Traylor, R.N., Schaefer, G.B., McPherson, E.W., Ballif, B.C., Klopocki, E., Mundlos, S., Shaffer, L.G., Aylsworth, A.S.; and 1q211 Study Group (2012). Proximal microdeletions and microduplications of 1q21.1 contribute to variable abnormal phenotypes. Eur. J. Hum. Genet. *20*, 754–761. https://doi.org/10.1038/ejhg.2012.6.
93. Stevenson, D.A., Bleyl, S.B., Maxwell, T., Brothman, A.R., and South, S.T. (2007). Mandibulofacial dysostosis in a patient with a de novo 2;17 translocation that disrupts the HOXD gene cluster. Am. J. Med. Genet. *143A*, 1053–1059. https://doi.org/10.1002/ajmg.a.31715.
94. Hunt, P., Gulisano, M., Cook, M., Sham, M.H., Faiella, A., Wilkinson, D., Boncinelli, E., and Krumlauf, R. (1991). A distinct Hox code for the branchial region of the vertebrate head. Nature *353*, 861–864. https://doi.org/10.1038/353861a0.
95. Trainor, P.A., and Krumlauf, R. (2000). Patterning the cranial neural crest: hindbrain segmentation and Hox gene plasticity. Nat. Rev. Neurosci. *1*, 116–124. https://doi.org/10.1038/35039056.
96. Trainor, P.A., and Krumlauf, R. (2001). Hox genes, neural crest cells and branchial arch patterning. Curr. Opin. Cell Biol. *13*, 698–705. https://doi.org/10.1016/s0955-0674(00)00273-8.
97. Gavalas, A., Trainor, P., Ariza-McNaughton, L., and Krumlauf, R. (2001). Synergy between Hoxa1 and Hoxb1: the relationship between arch patterning and the generation of cranial neural crest. Development *128*, 3017–3027.
98. Minoux, M., and Rijli, F.M. (2010). Molecular mechanisms of cranial neural crest cell migration and patterning in craniofacial development. Development *137*, 2605–2621. https://doi.org/10.1242/dev.040048.
99. Parker, H.J., Pushel, I., and Krumlauf, R. (2018). Coupling the roles of Hox genes to regulatory networks patterning cranial neural crest. Dev. Biol. *444*, S67–S78. https://doi.org/10.1016/j.ydbio.2018.03.016.
100. Xu, H., Niu, Y., Wang, T., Liu, S., Xu, H., Wang, S., Liu, J., and Ye, Z. (2015). Novel FGFR1 and KISS1R mutations in chinese kallmann syndrome males with cleft lip/palate. BioMed Res. Int. *2015*, 649698. https://doi.org/10.1155/2015/649698.
101. Aoyama, K., Mizuno, H., Tanaka, T., Togawa, T., Negishi, Y., Ohashi, K., Hori, I., Izawa, M., Hamajima, T., and Saitoh, S. (2017). Molecular genetic and clinical delineation of 22 patients with congenital hypogonadotropic hypogonadism. J. Pediatr. Endocrinol. Metab. *30*, 1111–1118. https://doi.org/10.1515/jpem-2017-0035.
102. Tommiska, J., Känsäkoski, J., Christiansen, P., Jørgensen, N., Lawaetz, J.G., Juul, A., and Raivio, T. (2014). Genetics of congenital hypogonadotropic hypogonadism in Denmark. Eur. J. Med. Genet. *57*, 345–348. https://doi.org/10.1016/j.ejmg.2014.04.002.

103. Riley, B.M., Mansilla, M.A., Ma, J., Daack-Hirsch, S., Maher, B.S., Raffensperger, L.M., Russo, E.T., Vieira, A.R., Dodé, C., Mohammadi, M., et al. (2007). Impaired FGF signaling contributes to cleft lip and palate. Proc. Natl. Acad. Sci. USA *104*, 4512–4517. https://doi.org/10.1073/pnas.0607956104.
104. Funato, N., and Yanagisawa, H. (2018). Deletion of the T-box transcription factor gene, Tbx1, in mice induces differential expression of genes associated with cleft palate in humans. Arch. Oral Biol. *95*, 149–155. https://doi.org/10.1016/j.archoralbio.2018.08.001.
105. Li, J., Rodriguez, G., Han, X., Janečková, E., Kahng, S., Song, B., and Chai, Y. (2019). Regulatory mechanisms of soft palate development and malformations. J. Dent. Res. *98*, 959–967. https://doi.org/10.1177/0022034519851786.
106. Li, R., Chen, Z., Yu, Q., Weng, M., and Chen, Z. (2019). The function and regulatory network of Pax9 gene in palate development. J. Dent. Res. *98*, 277–287. https://doi.org/10.1177/0022034518811861.
107. Gao, S., Moreno, M., Eliason, S., Cao, H., Li, X., Yu, W., Bidlack, F.B., Margolis, H.C., Baldini, A., and Amendt, B.A. (2015). TBX1 protein interactions and microRNA-96-5p regulation controls cell proliferation during craniofacial and dental development: implications for 22q11.2 deletion syndrome. Hum. Mol. Genet. *24*, 2330–2348. https://doi.org/10.1093/hmg/ddu750.
108. Broix, L., Jagline, H., Ivanova, E., Schmucker, S., Drouot, N., Clayton-Smith, J., Pagnamenta, A.T., Metcalfe, K.A., Isidor, B., Louvier, U.W., et al. (2016). Mutations in the HECT domain of NEDD4L lead to AKT-mTOR pathway deregulation and cause periventricular nodular heterotopia. Nat. Genet. *48*, 1349–1358. https://doi.org/10.1038/ng.3676.
109. Kato, K., Miya, F., Hori, I., Ieda, D., Ohashi, K., Negishi, Y., Hattori, A., Okamoto, N., Kato, M., Tsunoda, T., et al. (2017). A novel missense mutation in the HECT domain of NEDD4L identified in a girl with periventricular nodular heterotopia, polymicrogyria and cleft palate. J. Hum. Genet. *62*, 861–863. https://doi.org/10.1038/jhg.2017.53.
110. Mah-Som, A.Y., Skrypnyk, C., Guerin, A., Seroor Jadah, R.H., Vardhan, V.N., McKinstry, R.C., and Shinawi, M.S. (2021). New cohort of patients With CEDNIK syndrome expands the phenotypic and genotypic spectra. Neurol. Genet. *7*, e553. https://doi.org/10.1212/NXG.0000000000000553.
111. Sprecher, E., Ishida-Yamamoto, A., Mizrahi-Koren, M., Rapaport, D., Goldsher, D., Indelman, M., Topaz, O., Chefetz, I., Keren, H., O'brien, T.J., et al. (2005). A mutation in SNAP29, coding for a SNARE protein involved in intracellular trafficking, causes a novel neurocutaneous syndrome characterized by cerebral dysgenesis, neuropathy, ichthyosis, and palmoplantar keratoderma. Am. J. Hum. Genet. *77*, 242–251. https://doi.org/10.1086/432556.
112. McDonald-McGinn, D.M., Fahiminiya, S., Revil, T., Nowakowska, B.A., Suhl, J., Bailey, A., Mlynarski, E., Lynch, D.R., Yan, A.C., Bilaniuk, L.T., et al. (2013). Hemizygous mutations in SNAP29 unmask autosomal recessive conditions and contribute to atypical findings in patients with 22q11. J. Med. Genet. *50*, 80–90. https://doi.org/10.1136/jmedgenet-2012-101320.
113. Scollon, S., McWalter, K., Abe, K., King, J., Kimata, K., and Slavin, T.P. (2012). Haploinsufficiency of STK11 and neighboring genes cause a contiguous gene syndrome including Peutz-Jeghers phenotype. Am. J. Med. Genet. *158A*, 2959–2962. https://doi.org/10.1002/ajmg.a.35629.
114. Pegoraro, C., and Monsoro-Burq, A.H. (2013). Signaling and transcriptional regulation in neural crest specification and migration: lessons from xenopus embryos. Wiley Interdiscip. Rev. Dev. Biol. *2*, 247–259. https://doi.org/10.1002/wdev.76.
115. Szabó, A., and Mayor, R. (2018). Mechanisms of neural crest migration. Annu. Rev. Genet. *52*, 43–63. https://doi.org/10.1146/annurev-genet-120417-031559.
116. Shellard, A., and Mayor, R. (2016). Chemotaxis during neural crest migration. Semin. Cell Dev. Biol. *55*, 111–118. https://doi.org/10.1016/j.semcdb.2016.01.031.
117. Gouignard, N., Andrieu, C., and Theveneau, E. (2018). Neural crest delamination and migration: Looking forward to the next 150 years. Genesis *56*, e23107. https://doi.org/10.1002/dvg.23107.
118. Szabó, A., and Mayor, R. (2016). Modelling collective cell migration of neural crest. Curr. Opin. Cell Biol. *42*, 22–28. https://doi.org/10.1016/j.ceb.2016.03.023.
119. Bomba, L., Walter, K., and Soranzo, N. (2017). The impact of rare and low-frequency genetic variants in common disease. Genome Biol. *18*, 77. https://doi.org/10.1186/s13059-017-1212-4.
120. Kaiser, J. (2012). Human genetics. Genetic influences on disease remain hidden. Science *338*, 1016–1017. https://doi.org/10.1126/science.338.6110.1016.
121. Eberhart, J.K., Swartz, M.E., Crump, J.G., and Kimmel, C.B. (2006). Early Hedgehog signaling from neural to oral epithelium organizes anterior craniofacial development. Development *133*, 1069–1077.
122. Swartz, M.E., Nguyen, V., McCarthy, N.Q., and Eberhart, J.K. (2012). Hh signaling regulates patterning and morphogenesis of the pharyngeal arch-derived skeleton. Dev. Biol. *369*, 65–75. https://doi.org/10.1016/j.ydbio.2012.05.032 S0012-1606(12)00291-6.
123. Kennedy, A.E., and Dickinson, A.J.G. (2012). Median facial clefts in Xenopus laevis: roles of retinoic acid signaling and homeobox genes. Dev. Biol. *365*, 229–240. https://doi.org/10.1016/j.ydbio.2012.02.033.
124. Tahir, R., Kennedy, A., Elsea, S.H., and Dickinson, A.J. (2014). Retinoic acid induced-1 (Rai1) regulates craniofacial and brain development in Xenopus. Mech. Dev. *133*, 91–104. https://doi.org/10.1016/j.mod.2014.05.004.
125. Wahl, S.E., Kennedy, A.E., Wyatt, B.H., Moore, A.D., Pridgen, D.E., Cherry, A.M., Mavila, C.B., and Dickinson, A.J.G. (2015). The role of folate metabolism in orofacial development and clefting. Dev. Biol. *405*, 108–122. https://doi.org/10.1016/j.ydbio.2015.07.001.
126. Dubey, A., and Saint-Jeannet, J.P. (2017). Modeling human craniofacial disorders in Xenopus. Curr. Pathobiol. Rep. *5*, 79–92. https://doi.org/10.1007/s40139-017-0128-8.
127. Devotta, A., Juraver-Geslin, H., Gonzalez, J.A., Hong, C.S., and Saint-Jeannet, J.P. (2016). Sf3b4-depleted Xenopus embryos: A model to study the pathogenesis of craniofacial defects in Nager syndrome. Dev. Biol. *415*, 371–382. https://doi.org/10.1016/j.ydbio.2016.02.010.
128. Van Otterloo, E., Williams, T., and Artinger, K.B. (2016). The old and new face of craniofacial research: How animal models inform human craniofacial genetic and clinical data. Dev. Biol. *415*, 171–187. https://doi.org/10.1016/j.ydbio.2016.01.017.
129. Dougherty, M., Kamel, G., Grimaldi, M., Gfrerer, L., Shubinets, V., Ethier, R., Hickey, G., Cornell, R.A., and Liao, E.C. (2013). Distinct requirements for wnt9a and irf6 in

extension and integration mechanisms during zebrafish palate morphogenesis. Development *140*, 76–81. https://doi.org/10.1242/dev.080473.

130. Duncan, K.M., Mukherjee, K., Cornell, R.A., and Liao, E.C. (2017). Zebrafish models of orofacial clefts. Dev. Dyn. *246*, 897–914. https://doi.org/10.1002/dvdy.24566.
131. Dickinson, A.J.G., and Sive, H.L. (2009). The Wnt antagonists Frzb-1 and Crescent locally regulate basement membrane dissolution in the developing primary mouth. Development *136*, 1071–1081. https://doi.org/10.1242/dev.032912.
132. Dickinson, A.J.G. (2016). Using frogs faces to dissect the mechanisms underlying human orofacial defects. Semin. Cell Dev. Biol. *51*, 54–63. https://doi.org/10.1016/j.semcdb.2016.01.016.
133. Teng, T., Teng, C.S., Kaartinen, V., and Bush, J.O. (2022). A unique form of collective epithelial migration is crucial for tissue fusion in the secondary palate and can overcome loss of epithelial apoptosis. Development *149*, dev200181. https://doi.org/10.1242/dev.200181.
134. Izadi, M., Schlobinski, D., Lahr, M., Schwintzer, L., Qualmann, B., and Kessels, M.M. (2018). Cobl-like promotes actin filament formation and dendritic branching using only a single WH2 domain. J. Cell Biol. *217*, 211–230. https://doi.org/10.1083/jcb.201704071.
135. Leslie, E.J., Carlson, J.C., Shaffer, J.R., Feingold, E., Wehby, G., Laurie, C.A., Jain, D., Laurie, C.C., Doheny, K.F., McHenry, T., et al. (2016). A multi-ethnic genome-wide association study identifies novel loci for non-syndromic cleft lip with or without cleft palate on 2p24.2, 17q23 and 19q13. Hum. Mol. Genet. *25*, 2862–2872. https://doi.org/10.1093/hmg/ddw104.
136. Qian, Y., Li, D., Ma, L., Zhang, H., Gong, M., Li, S., Yuan, H., Zhang, W., Ma, J., Jiang, H., et al. (2016). TPM1 polymorphisms and nonsyndromic orofacial clefts susceptibility in a Chinese Han population. Am. J. Med. Genet. *170A*, 1208–1215. https://doi.org/10.1002/ajmg.a.37561.
137. De Groote, P., Tran, H.T., Fransen, M., Tanghe, G., Urwyler, C., De Craene, B., Leurs, K., Gilbert, B., Van Imschoot, G., De Rycke, R., et al. (2015). A novel RIPK4-IRF6 connection is required to prevent epithelial fusions characteristic for popliteal pterygium syndromes. Cell Death Differ. *22*, 1012–1024. https://doi.org/10.1038/cdd.2014.191.
138. Carroll, E.A., Gerrelli, D., Gasca, S., Berg, E., Beier, D.R., Copp, A.J., and Klingensmith, J. (2003). Cordon-bleu is a conserved gene involved in neural tube formation. Dev. Biol. *262*, 16–31.
139. Aylward, A., Cai, Y., Lee, A., Blue, E., Rabinowitz, D., Haddad, J., Jr.; and University of Washington Center for Mendelian Genomics (2016). Using whole exome sequencing to identify candidate genes with rare variants in nonsyndromic cleft lip and palate. Genet. Epidemiol. *40*, 432–441. https://doi.org/10.1002/gepi.21972.
140. Briggs, J.A., Weinreb, C., Wagner, D.E., Megason, S., Peshkin, L., Kirschner, M.W., and Klein, A.M. (2018). The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. Science *360*, eaar5780. https://doi.org/10.1126/science.aar5780.
141. Cordero, D.R., Brugmann, S., Chu, Y., Bajpai, R., Jame, M., and Helms, J.A. (2011). Cranial neural crest cells on the move: their roles in craniofacial development. Am. J. Med. Genet. *155A*, 270–279. https://doi.org/10.1002/ajmg.a.33702.
142. Tong, C., Ohyama, T., Tien, A.C., Rajan, A., Haueter, C.M., and Bellen, H.J. (2011). Rich regulates target specificity of photoreceptor cells and N-cadherin trafficking in the Drosophila visual system via Rab6. Neuron *71*, 447–459. https://doi.org/10.1016/j.neuron.2011.06.040.
143. Ishida, M., E Oguchi, M., and Fukuda, M. (2016). Multiple types of guanine nucleotide exchange factors (GEFs) for Rab small GTPases. Cell Struct. Funct. *41*, 61–79. https://doi.org/10.1247/csf.16008.
144. Novick, P. (2016). Regulation of membrane traffic by Rab GEF and GAP cascades. Small GTPases *7*, 252–256. https://doi.org/10.1080/21541248.2016.1213781.
145. Rafighdoost, H., Hashemi, M., Narouei, A., Eskanadri-Nasab, E., Dashti-Khadivaki, G., and Taheri, M. (2013). Association between CDH1 and MSX1 gene polymorphisms and the risk of nonsyndromic cleft lip and/or cleft palate in a southeast Iranian population. Cleft Palate. Craniofac. J. *50*, e98–e104. https://doi.org/10.1597/12-144.
146. Kumari, P., Singh, S.K., and Raman, R. (2018). A novel noncoding RNA within an intron of CDH2 and association of its SNP with non-syndromic cleft lip and palate. Gene *658*, 123–128. https://doi.org/10.1016/j.gene.2018.03.017.
147. Brito, L.A., Yamamoto, G.L., Melo, S., Malcher, C., Ferreira, S.G., Figueiredo, J., Alvizi, L., Kobayashi, G.S., Naslavsky, M.S., Alonso, N., et al. (2015). Rare variants in the epithelial cadherin gene underlying the genetic etiology of nonsyndromic cleft lip with or without cleft palate. Hum. Mutat. *36*, 1029–1033. https://doi.org/10.1002/humu.22827.
148. Patel, N., Anand, D., Monies, D., Maddirevula, S., Khan, A.O., Algoufi, T., Alowain, M., Faqeih, E., Alshammari, M., Qudair, A., et al. (2017). Novel phenotypes and loci identified through clinical genomics approaches to pediatric cataract. Hum. Genet. *136*, 205–225. https://doi.org/10.1007/s00439-016-1747-6.
149. Liu, K., Wang, A., Ran, L., Zhang, W., Jing, S., Wang, Y., Zhang, X., Liu, G., Sen, W., and Song, F. (2020). ARHGEF38 as a novel biomarker to predict aggressive prostate cancer. Genes Dis. *7*, 217–224. https://doi.org/10.1016/j.gendis.2019.03.004.