Original research

New locus underlying auriculocondylar syndrome (ARCND): 430 kb duplication involving TWIST1 regulatory elements

Vanessa Luiza Romanelli Tavares,1,2,3,4,5,6 Sofia Ligia Guimarães-Ramos,1 Yan Zhou,1 Cibele Masotti,1,3,4 Suzana Ezquina,1,4 Danielle de Paula Moreira,1 Henk Buermans,5 Renato S. Freitas,6 Johan T. Den Dunnen,5 Stephen R. F. Twigg1,2, Maria Rita Passos-Bueno1

ABSTRACT

Background Auriculocondylar syndrome (ARCND) is a rare genetic disease that affects structures derived from the first and second pharyngeal arches, mainly resulting in micrognathia and auricular malformations. To date, pathogenic variants have been identified in three genes involved in the EDN1-DLX5/6 pathway (PLCB4, GNAI3 and EDN1) and some cases remain unsolved. Here we studied a large unsolved four-generation family.

Methods We performed linkage analysis, resequencing and Capture-C to investigate the causative variant of this family. To test the pathogenicity of the CNV found, we modelled the disease in patient craniofacial progenitor cells, including induced pluripotent cell (iPSC)-derived neural crest and mesenchymal cells.

Results This study highlights a fourth locus causative of ARCND, represented by a tandem duplication of 430 kb in a candidate region on chromosome 7 defined by linkage analysis. This duplication segregates with the disease in the family (LOD score=2.88) and includes HDAC9, which is located over 200 kb telomeric to the top candidate gene TWIST1. Notably, Capture-C analysis revealed multiple cis interactions between the TWIST1 promoter and possible regulatory elements within the duplicated region. Modelling of the disease revealed an increased expression of HDAC9 and its neighbouring gene, TWIST1, in neural crest cells. We also identified decreased migration of iPSC-derived neural crest cells together with dysregulation of osteogenic differentiation in iPSC-affected mesenchymal stem cells.

Conclusion Our findings support the hypothesis that the 430 kb duplication is causative of the ARCND phenotype in this family and that deregulation of TWIST1 expression during craniofacial development can contribute to the phenotype.

INTRODUCTION

Auriculocondylar syndrome (ARCND) (OMIM #602483, #614669 and #615706), also referred to as ‘question mark ear syndrome’, is a rare Mendelian disorder with a prevalence of under 1 in 1000000 (Orphanet; http://www.orpha.net/consorc/cgi-bin/index.php). ARCND is characterised by micrognathia, question mark ears, mandibular condyle hypoplasia, and other less common features such as microstomia, glossoptosis, postauricular tags and prominent cheeks.12 There is wide clinical variability, including cases with isolated ear anomalies. Treatment is mainly corrective through surgical intervention for mandibular ramus lengthening using distraction osteogenesis and accompanied by orthodontic treatments and speech therapy.13 Understanding the aetiology of this disorder and elucidating genetic causes improve counselling and may lead to the development of preventive or therapeutic strategies, besides deepening our knowledge of craniofacial development.

The main structures affected in ARCND are derived from the first and second pharyngeal arches that are colonised by neural crest cells (NCCs), originating from the neural plate border by epithelial mesenchymal transition, and mesenchymal stem cells. NCCs are multipotent cells with high migratory ability that can differentiate into several derivatives such as cartilage, bone, peripheral neurons, melanocytes and glia, and they have a central role in craniofacial development.4 Disruption in these migratory, patterning or differentiation processes may result in congenital craniofacial malformations.5 So far, pathogenic variants in patients with ARCND have been found in genes of the EDN1-endothelin-1 receptor type A (EDNRA) pathway, which are expressed by the neural crest-derived ectomesenchymal cells of pharyngeal arches and are responsible for the patterning of the mandibular domain in the first arch.6-8 The most commonly mutated gene in individuals with ARCND is PLCB4 (MIM 600810; 58% of patients), followed by GNAI3 (MIM 139370; 19% of cases) and EDN1 (MIM 131240 and MIM 139370; 15% of cases).9-12 Approximately 8% of ARCND cases remain unsolved.9,13 Most of the knowledge concerning ARCND pathogenesis comes from functional studies using animal models of craniofacial development, such as mouse and zebrafish.14,15 However, animal models may not completely reflect what happens in human development,16 and human stem cells or induced pluripotent cells (iPSCs) represent a complementary model system to study development in a human-specific context.17-20 iPSCs, which can differentiate into cell types such as NCCs and NCC-derived mesenchymal-like stem cells, can provide insight...
into human craniofacial development where facial structures derived from the first pharyngeal arch are compromised, as successfully exemplified in the case of Richieri-Costa-Pereira syndrome (RCPS), RCPS, caused by biallelic, hypomorphic alleles at the DEAD-box helicase EIF4A3, shares overlapping clinical features with ARCND particularly in mandible underdevelopment. In this work, we studied a previously reported ARCND family and performed genetic and functional investigations using patient iPSCs that had been differentiated into NCC (iPSC-derived NCs) and mesenchymal-like stem cells (mMSC) derivatives. Our findings suggest that duplication of sequences at the HDAC9 locus can lead to the development of ARCND, possibly by disruption of regulatory elements that control expression of the neighbouring TWIST1 gene.

**MATERIALS AND METHODS**

**Patients and DNA samples**

The Brazilian family (referred to as F1; online supplemental table 1) is a non-consanguineous family with 11 members showing the typical characteristics of ARCND. Of those documented, the majority presented with question mark ears (8 of 10), microstomia (8 of 10) and micrognathia (6 of 10), with considerable intrafamilial variability observed. Genomic DNA was extracted from peripheral blood lymphocytes according to Miller et al.

**Sanger sequencing**

Primers and conditions used in PCR amplification prior to Sanger sequencing were as described in Romanelli Tavares et al. The data were analysed using Sequencher V.5.1 software (http://genecodes.com/). Variants present in the 1000 Genomes Database, dbSNP150 (through University of California, Santa Cruz Genome Browser, UCSC; https://genome.ucsc.edu/), Genome Aggregation Database (https://gnomad.broadinstitute.org/), or in the Online Archive of Brazilian Mutations (http://abraom.ib.usp.br/) were considered unlikely to be pathogenic.

**Linkage analysis**

Nine affected individuals (II-4, II-6, II-8, III-5, III-10, III-13, III-14, IV-3 and IV-6) and three unaffected individuals (III-6, III-11 and IV-4) were genotyped using the GeneChip Human Mapping 50K Array Xba 240 (Affymetrix), according to the manufacturer’s protocol. The genotype data were analysed using Affymetrix Genotyping Console. The overall quality of the samples was estimated through quality control (QC) algorithm (dynamic model algorithm with QC call rate) using a threshold of 90%.

Linkage analysis was performed with the easyLINKAGE-Plus V.5.08 package. Mendelian inconsistencies were removed using PedCheck V.1.0.159 and the non-Mendelian inconsistencies with Merlin V.1.0.1 software. The logarithmic odds (LOD) score was obtained using the parametric multipoint test with GeneHunter V.2.1r5. Analysis parameters were defined as autosomal dominant, estimated penetrance K=0.9, disease allele frequency estimated at 0.0001 and marker spacing at 0.0010 cM, and map distances were acquired from AFRY 100K deCODE Human GRCh37/hg19, and SNPs with a call rate less than 90% were removed.

The candidate region was confirmed by amplification of microsatellite markers from the ABI PRISM Linkage Mapping Set V2.0 (Perkin-Elmer, Applied Biosystems, Foster City, California), read in a MegaBACE 1000 automatic sequencer (Amersham, GE Healthcare, Little Chalfont, UK) according to the manufacturer’s protocol and analysed with MegaBACE Profiler software (Amersham, GE Healthcare). Endavour gene prioritisation was applied to the candidate Chr7 region (https://endeavour.esat.kuleuven.be) to generate a candidate gene list. Training gene lists (reference genes) were compiled according to the following criteria: (a) genes with a central role in the formation of the structures affected in patients with ARCND (eg, ears, mandible and mandibular condyle); or (b) genes related to the embryonic developmental processes involved in the formation of some of the structures affected in patients with ARCND (eg, neural crest and formation of the first and second pharyngeal arches) (online supplemental table 2).

**Whole exome sequencing**

Whole exome libraries were generated using either the Agilent SureSelect Human All Exon 50 Mb Kit (patients IV-3 and IV-6; performed at the Center for Human and Clinical Genetics, Leiden University Medical Center, The Netherlands) or the Illumina TruSeq Kit (patients III-10 and III-13; performed at Luiz de Queiroz College of Agriculture, Sao Paulo, Piracicaba, Brazil). Whole exome sequencing (WES) was carried out on Illumina HiSeq 2000 (2 x 100 bp paired-end run). Sequences were aligned to the human reference GRCh37 (hg19) using the Burrows-Wheeler Aligner (BWA). Processing and variant calling were performed along with batch samples using the Unified Genotyper tool (Genome Analysis Toolkit, GATK) (http://www.broadinstitute.org/gatk/), using default parameters, with exception to the following changes: minIndelCnt 3; minIndelFrac 0.020; contamination 0.02; metrics snps.metrics; stand_call_conf 30.0; stand_emit_conf 10.0; min_base_quality_score 12; dcov 300; baq CALCULATE_AS_NECESSARY. Annotation was done with Annovar (http://annovar.openbioinformatics.org/). Variants were selected if they had been approved by the filter quality (PASS), frequency ≤0.01 in 1000 Genomes Database (https://www.internationalgenome.org/) and Exome Variant Server (ESP6500; https://evs.gs.washington.edu/ESV/), heterozygous genotype in all four affected individuals sequenced, and with allele count ≤4 in the local sequenced cohort of patients without ARCND (total allele number = 132).

**Targeted sequencing**

Targeted sequencing at the HDAC9/TWIST1 locus was performed using a resequencing capture panel designed to the TWIST1 gene and flanking regions (2.4 Mb with boundaries selected according to human to mouse synteny; chr7:17,346,143–19,695,462, GRCh38). Genomic DNA from family members (three affected and one unaffected family members: IV-3, IV-6, III-5, II-7) was fragmented by sonication, ligated to indexed Illumina sequencing adapters and amplified. Purified libraries were mixed with the biotinylated probe mixture (SeqCap EZ Choice Library System, Roche Nimblegen) and enriched DNA for the targeted regions sequenced on an Illumina HiSeq 2500. Sequencing adapter sequences and low-quality bases were removed using Trimmomatic (V.0.32; parameter SLIDINGWINDOW: 4:2032) and the trimmed read pairs were then aligned to human reference genome hg19 using BWA (V.0.7.12) with default parameters. The aligned reads were analysed using amplimap (V.0.2.9) and coverage calculated using BEDTools V.0.25.0. Variants were called separately in each sample using Platypus (V.0.8.1) and then concatenated, merged and normalised using BCFtools (V.1.5; https://github.com/samtools/bcftools) and annotated with Annovar.
For the breakpoint isolation, we examined the resequencing data at the duplication junctions and designed the following primers to amplify the breakpoint: F-5’-CCCATGCTCAT TCTTCTTTG-3’ and R-5’-TGCGAGCTTATGTCTTCTT-3’.

**Capture-C**

To identify the chromatin regions that the TWIST1 promoter interacts with, we used a Capture-C approach. For the chromatin template we prepared human mesenchymal cells from human embryonic calvaria (three different samples at 12–14 postconception weeks, provided by the Human Developmental Biology Resource, UK). We removed the skin and dura mater and then dissected a bony strip (approximately 0.5 mm wide) that included the frontal bone, coronal suture and parietal bone and placed this in a gelatinised culture dish containing the following selective medium: BHK-21 Glasgow MEM (Gibco 21700-025)—to this 500 mL we added glutamine (Gibco 25030-024), 2 mM sodium pyruvate (Gibco 11360-039), 100 U/mL penicillin/streptomycin, non-essential amino acids (Gibco 11140-035), 10% Fetal Bovine Serum (FBS) (Gibco 10270), 0.1 mM β-mercaptoethanol and Lif (inhouse prepared Lif-containing medium from modified Chinese Hamster Ovary (CHO) cells).

Cells were allowed to grow out of the bone for 3–5 days and then collected (discarding the bony strips) and cultured again using the same media. Cells were passaged at least twice more before collection for 3C library preparation. Cells (10–15 million) were fixed in formaldehyde and then lysed prior to digestion of the cross-linked DNA template with DpnII and DNA ligation. Following DNA purification, the 3C library was sonicated and then replaced and cell migration images were acquired at 0 hour and 24 hours. All samples were assessed simultaneously in two independent experiments. The percentage of the wound covered by migrating cells after 24 hours was quantified in ARCDN and control NCCs using ImageJ.

**Generation of iPSC, NCC and MSC**

Three ARCDN samples and three control samples were used for generation of iPSCs. One of the control iPSCs used in this study (F7405-1) had been generated with retroviral transduction and was described and characterised elsewhere. The other cells were established from erythroblast cultures derived from peripheral blood (Romaniello Tavares VL, et al. J Med Genet 2021;0:1–11. doi:10.1136/jmedgenet-2021-107825). Fluorescence was detected using the QuantStudio 5 System (Life Technologies) following the manufacturer’s recommendations. Total RNA was extracted from cells with the Nucleospin RNA II Extraction Kit (Macherey-Nagel) according to the manufacturer’s recommendations. Total RNA was converted into cDNA using SuperScript IV (Life Technologies) and oligo-dT primers. Real-time QPCR reactions were performed with 2X Fast SYBR Green PCR Master Mix (Applied Biosystems) and 50–400 nM of each primer. Fluorescence was detected using the QuantStudio 5 System (Life Technologies) under a standard temperature protocol. Primer pairs were either designed with Primer-BLAST or retrieved from PrimerBank and supplied by Extenx (online supplemental table 4). geNorm (https://genorm.cmgg.be) was used to determine the normalisation factor (using gene expression of TATA-box binding protein (TBP), hydroxymethylbilane synthase (HMBS) or glyceraldehyde-3-phosphate dehydrogenase (GAPDH)) and calculate normalisation factors (E−ΔC) for the normalisation factor (using gene expression of TATA-box binding protein (TBP), hydroxymethylbilane synthase (HMBS) or glyceraldehyde-3-phosphate dehydrogenase (GAPDH)) and calculate normalisation factors (E−ΔC) for in the online supplemental material along with the antibodies used (online supplemental table 3). To assess EDN1/EDNRA pathway-related gene expression, NCCs were treated with EDN1 100nM for 24 hours.
The maximum theoretical LOD score for this family was equal to 2.93.

### Statistical analysis

All experiments were performed in triplicate, unless stated otherwise. Statistical comparisons were performed using GraphPad Prism V.5 software. Unpaired student’s t-test and two-way analysis of variance (ANOVA) values were represented as mean ± SE. The level of statistical significance was set at p < 0.05.

### RESULTS

#### Evidence of a fourth locus for ARCND

Sanger sequencing of PLCB4, GNAI3 and EDN1 did not reveal any pathogenic variants in the coding regions, 5’ UTRs (untranslated regions) or splice sites of these genes, suggesting that a different locus might underlie the ARCND in the family. Next, we carried out a linkage analysis that revealed three regions with positive LOD scores on chromosomes 7, 14 and 18 (online supplemental figure 1). The highest LOD score (2.88), which is close to the threshold value of ≥3.0 for significance, and the maximum theoretical LOD score for this family (2.93), was observed in a region of about 17.6 Mb on chromosome 7 (chr7: 14395902–32017194 (hg38); table 1). Genotyping of microsatellite markers narrowed the chromosome 7 linkage region to chr7: 14395902–28158440 (hg38) (online supplemental figure 2).

We performed WES on four affected individuals. After filtering (as described in the methods), only one variant from the candidate regions on chromosomes 7, 14 and 18 remained, a synonymous change located in TWIST1 (NM_014817.3:c.345G>A; p.(=)), classified as a variant of uncertain significance (BP4, PM2 and PP4 according to the American College of Medical Genetics and Genomics guidelines). Sanger sequencing of additional family members (nine affected and five non-affected individuals) showed that the variant did not segregate with the disease and therefore it was not considered further. Variants in candidate genes within the endothelin pathway (EDN1, EDNRA, DLX5, DLX6, FURIN and ECE1) were also investigated with WES in the same manner, but no obvious pathogenic variants were detected. We then took a gene prioritisation approach, and TWIST1 was the top-ranked gene in this analysis using two training lists, with a p value equal to 0.00053 and 0.00027 (using training lists ‘a’ and ‘b’, respectively; online supplemental table 3). TWIST1, a basic transcription factor of the helix-loop-helix (bHLH) family, is expressed in cranial mesoderm and neural crest-derived mesenchyme, which are tissues involved in craniofacial development. Its role in mandibular condyle and mandible formation has also been demonstrated. These observations led us to further investigate the genomic region surrounding TWIST1.

Targeted resequencing of a 2.4 Mb region around TWIST1 revealed no potential pathogenic variants in TWIST1, but did detect a tandem duplication within HDAC9 that was only present within the three affected family members tested (figure 1A and online supplemental figure 3). We designed primers to the sequences either side of the duplication and characterised the duplication breakpoint (figure 1B, C). The duplication spanned 430,302 bp (NC_000007.14:g.18437239_18867540dup) telomeric to TWIST1, covering most of the HDAC9 gene. Multiple transcript isoforms of HDAC9 are duplicated in their entirety (including transcript isoforms 3, 8, 9, 10 and 11); however, full-length catalytically active transcript isoforms (transcript isoforms 1, 5, 6 and 7) extend beyond the breakpoint and therefore are likely to be disrupted by this duplication. Breakpoint amplification and Sanger sequencing within the family demonstrated segregation of the duplication with the phenotype (figure 1B). All unaffected individuals were negative, suggesting full penetrance in this family (online supplemental figure 4).

Analysis of CNVs in DECIPHER ([https://decipher.sanger.ac.uk](https://decipher.sanger.ac.uk)) revealed 19 patients with copy-number gains in the region. With the exception of a single small duplication (patient identification 276644), all the CNVs that overlap the one described here are much larger and encompass multiple genes. Nevertheless, 7 out of 19 have descriptions that include ear malformation among other features (patient identification: 393911, 393942, 395511, 396512, 280316, 396373 and 394346) include enhancer eTW6 (Hs2307) that regulates the expression of Twist1 and other two regulatory elements (eTW7 (Hs2306) and eTW5) Hs644; VISTA Enhancer Browser). Similarly, 7 out of 19 DECIPHER patients presented with micrognathia (patient identification: 393911, 393942, 395511, 396512, 280316, 396373 and 394346); these duplications also include eTW6, except for individual 2363. Interestingly, one of the DECIPHER duplications (276644; 179 kb) is enclosed entirely within the ARCND duplicated region; however, this patient does not have ARCND clinical features (2021, Olivier Faire, L., personal communication). We note that the non-overlapping sequence between patient 276644 and the ARCND duplication includes the aforementioned regulatory Twist1 elements (online supplemental figure 5 and table 6).

To explore how the duplication identified above, which is over 200 kb telomeric to the candidate gene TWIST1, could be pathogenic, we carried out a Capture-C analysis using the TWIST1 promoter as the viewpoint. This demonstrated that there were multiple contacts between TWIST1 and regions to either side. The highest frequency of interactions was telomeric to TWIST1, particularly within the HDAC9 gene and the region spanned by the duplication (figure 1A). This implies that this region contains regulatory elements involved in the control of TWIST1 expression, providing a possible pathogenic mechanism for the duplication.

### Table 1: Regions with positive logarithmic odds (LOD) scores obtained by linkage analysis

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Maximum LOD score</th>
<th>rsID</th>
<th>Physical position hg38 (bp)</th>
<th>To rsID</th>
<th>Physical position hg38 (bp)</th>
<th>Region size (Mb)</th>
<th>Number of genes (NCBI RefSeq curated)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chr7</td>
<td>2.88</td>
<td>rs1036140</td>
<td>14395902</td>
<td>rs28190</td>
<td>32017194</td>
<td>17.6</td>
<td>146</td>
</tr>
<tr>
<td>Chr14</td>
<td>2.41</td>
<td>rs10484206</td>
<td>49103955</td>
<td>rs10498419</td>
<td>49978197</td>
<td>0.87</td>
<td>14</td>
</tr>
<tr>
<td>Chr18</td>
<td>1.67</td>
<td>rs1398193</td>
<td>48460768</td>
<td>rs768360</td>
<td>5023703</td>
<td>1.8</td>
<td>20</td>
</tr>
</tbody>
</table>

The maximum theoretical LOD score for this family was equal to 2.93.
Analysis of the duplication in iPSC and craniofacial progenitor cells

To further investigate the pathogenicity of the duplication, we used an in vitro approach to model the disease, generating iPSC from affected and unaffected individuals in the family. The experimental design was based on recapitulating different stages of early embryonic development that are relevant to the ARCND phenotype, most particularly iNCC (iPSC-derived NCCs) and nMSC (inNCC-derived mesenchymal-cell like). All cell types were fully characterised and showed cellular specific expression of relevant markers and typical cell morphology (online supplementary figures 6–8).

Expression analysis of HDAC9, TWIST1 and ARCND-related markers

Previous analysis of HDAC9 has shown that it contains regulatory elements important for TWIST1 expression.49 51 Together with our prioritisation and Capture-C analysis, this prompted us to investigate the expression of both of these genes in iPSC and derived cell types. TWIST1 and HDAC9 mRNA in iPSC did not show any difference between patients and controls (data not shown). However, an increase in HDAC9 (3.15-fold, unpaired t-test p=0.009) and TWIST1 (2.03-fold, unpaired t-test p=0.03) mRNA was observed in the ARCND-iNCC (figure 2A,B) compared with controls (unpaired t-test).

In order to evaluate if there is deregulation of the EDN1 pathway in the ARCND cells, we investigated the expression of the key downstream target of this pathway, DLX5,15 52 as well as other genes shown to be activated, BARX1, NKKX3.2, GSC, DLX3 and HAND2.53 54 Expression of BARX1, NKKX3.2 and GSC was not significantly different between controls and ARCND iNCCs (two-way ANOVA, p>0.05; online supplemental figure 9), while HAND2, DLX3 and DLX5 mRNA levels were too low to be measured (data not shown).

Analysis of ARCND iNCC and nMSC function

To screen for cellular phenotypes, we assessed cell cycle and migration of ARCND iNCC compared with controls, as alterations in these cellular functions are considered to be underlying mechanisms in several NCC-related diseases.17 55 56 Although no significant difference in cell cycle distribution was detected between patient and control cells, a significant decrease in migratory capacity of ARCND iNCCs was observed compared with controls (4.3-fold decrease, Student’s t-test, p=0.0009) (figure 2C–E).

Marked mandibular hypoplasia is often seen in patients with ARCND, which could be caused by dysregulation of osteogenic differentiation. Therefore, we investigated this process in nMSCs. Our data showed that during osteogenic differentiation, ALP enzymatic activity was significantly diminished in ARCND-nMSCs after 9 days of osteoinduction (decrease of 20.3-fold, paired t-test p=0.029; figure 3A,B). In addition, alizarin red staining revealed a subtle decrease in matrix mineralisation in ARCND-nMSCs in comparison with controls (t-test p>0.05; figure 3C). Next, we assessed the expression of key osteogenesis genes (figure 3D–I). ALP showed a statistically significant downregulation (p=0.035), whereas MSX2 expression was higher in ARCND-nMSC compared with the controls, although not reaching statistical significance. No significant difference
was seen in the expression of RUNX2, TWIST1, BGLAP and COL1A1. Together, these results indicate a delay or impairment of osteogenic differentiation.

**DISCUSSION**

In this work, we used linkage and resequencing analysis to reveal a new ARCND locus on chromosome 7. We identified a novel 430 kb CNV that duplicated sequences beginning ~280 kb telomeric of TWIST1, the gene prioritised as the best candidate for the phenotype in the linkage region. The CNV, which segregated with the disease in a four-generation large Brazilian family, duplicated possible regulatory element sequences within HDAC9, which we demonstrated, through a chromosome conformation capture assay, make contact with the TWIST1 promoter. This aligns with previous studies of this region that show that mouse Twist1 regulatory regions can be found within both introns and exons of the Hdac9 gene. Notably, a 23 kb deletion that included three of these regulatory element sequences, which may also be bound by craniofacial transcription factors Lmx1b and Tfap2, was associated with a reduction in Twist1 expression.

**Figure 2** RT-qPCR assessment of (A) HDAC9 and (B) Twist1 showing upregulated expression in ARCND-derived NCC. Both genes showed statistically significant differences among controls and ARCND. **Two-tailed p=0.0094; *one-tailed p=0.0250, unpaired Student’s t-test.** (C–E) Evaluation of cell cycle and cell migration in ARCND NCC. (C) Cell cycle assay, not statistically significant. (D) Bar graph depicting the rate of cell migration (cell-covered area, %) after 24 hours; data shown are representative of two independent assays and three independent measurements in each. ****Two-tailed, p=0.0009, Student’s t-test. (E) Representative phase-contrast micrographs acquired immediately after wounding at 0 and 24 hours. All values represent mean±SEM.

ARCND, auriculocondylar syndrome; AU, arbitrary unit; NCC, neural crest cells.

Furthermore, rearrangements that leave the gene intact but remove regions telomeric of TWIST1 lead to Saethre-Chotzen syndrome (SCS),58–61 which is caused by TWIST1 haploinsufficiency. SCS is characterised by craniosynostosis of the coronal sutures, which is not found in ARCND. Overall, these findings underline the importance of this region for TWIST1 regulation and suggest that duplication of these and other regulatory elements could be associated with altered TWIST1 expression during development. Importantly, transgenic enhancer assays have demonstrated in both mouse and zebrafish that some of these regulatory elements can drive Twist1 pharyngeal expression. For example, the enhancer known as eTw651 or Hs2307 (Vista Enhancer Browser62) that overlaps with Hdac9 exon 19 drives the expression of Twist1 in mouse E11.5 pharyngeal arches,49 51 providing a mechanism whereby rearrangements at this genomic locus could lead to the pharyngeal arch-related developmental abnormalities found in ARCND. Further support for a link between increased TWIST1 expression and the developmental anomalies found in our ARCND family comes from overlap with the clinical features described in cases with three copies of the 7p chromosomal region. For example, micrognathia or small mandibles have been found in patients with partial trisomy 7p63–66. We note, however, that our patients did not have large, open fontanelles, the hallmark feature of trisomy 7p which has been linked to triple dosage of TWIST1,67 which may reflect having three copies of regulatory elements rather than

Figure 3 Evaluation of osteogenic potential in ARCND-mesenchymal stem cells (MSC). (A) Quantification of ALP enzymatic activity after 9 days and (B) alizarin red staining after 21 days of osteoinduction in ARCND-MSC in comparison with controls. Measurements from differentiated cells were normalised to paired, undifferentiated negative staining controls. (A and B) Student’s t-test, ALP activity two-sided and alizarin red one-sided (*p<0.05). (C) Representative alizarin red staining micrographs showing matrix mineralisation (in dark brown) of ARCND-MSC samples versus one representative control (osteogenic differentiation for 9 days); micrographs are shown paired to respective negative controls (undifferentiated cells). (D–I) Transcriptional profiles of TWIST1 and osteogenic differentiation markers during the initial 6 days of osteoinduction. ALP was statistically significant (*p<0.05). MSX2, RUNX2, COL1A1 and BGLAP did not show statistically significant differences (ns). All values represent mean±SEM. (D–I) Two-way ANOVA with Bonferroni post-tests. (C) Scale bars: 1000 µm. ALP, alkaline phosphatase; ANOVA, analysis of variance; ARCND, auriculocondylar syndrome; AU, arbitrary units; MSC, mesenchymal stem cells (undifferentiated cells); OST, osteogenic differentiation.
three copies of the gene itself. Similarly, a significant number of DECIPHER duplications (~40%) at this locus were also associated with ARCND overlapping features, despite the variable phenotype and incomplete penetrance of ARCND.9 10 11 18 Interestingly, one of the duplications (276644) was entirely contained within the ARCND duplication, but this CNV did not span the known Twist1 regulatory element1 and the case was not associated with any features of ARCND. Altogether, these observations reinforce that the 430 kb HDAC9 duplication is pathogenic and that altered expression of TWIST1 might contribute to the ARCND phenotype.

To investigate the pathogenicity of the CNV, we used an iPSC-based approach to screen for molecular and cellular alterations associated with ARCND in the family. As the affected craniofacial structures in ARCND arise from the neural crest, we generated and analysed NCC derived from patient and control iPSCs and found upregulation of both HDAC9 and TWIST1. Only the shorter, catalytically inactive forms of HDAC967 could be transcribed from within the duplication to account for the upregulation of this gene, while the larger transcripts (the catalytic domain is encoded by multiple exons at the 3’ end of the gene) are predicted to be disrupted. HDAC9, a class II histone deacetylase enzyme, usually associated with transcriptional repression,69 has been linked to many types of cancer such as glioblastoma, breast cancer and oral squamous cell carcinoma, and chronic disorders such as diabetes and osteoporosis.70–72 There is no evidence of a role of HDAC9 in craniofacial development and it is not expressed in mouse E11.5 pharyngeal arch.49 although a role in bone development is possible as Hdc9 expression has been shown to increase osteoclastogenesis and regulate osteogenesis.73 74 and a contribution to the development of the ARCND phenotype cannot be excluded. In contrast, the relevance of TWIST1 in craniofacial development has been highlighted by the human conditions SCS75 76 and Sweeney-Cox syndrome,77 which are caused by pathogenic variants in this gene, as well as by studies of Twist1 mouse models.43 46 Moreover, Twist1 directly inhibits Runx2,78 79 the master regulator of osteogenic differentiation, as well as downstream targets of Runx2 like bone sialoprotein.80 Its overexpression leads to reduced ossification78 81 and conditional inactivation of Twist1 has demonstrated an essential role in the survival of NCC in mandibular development, as well as in ossification of the mandible leading to mandibular hypoplasia, condylar process loss and altered middle ear.81 46 Of note, a comparable mandibular phenotype is also observed in a mouse model deleted for the enhancer that regulates pharyngeal arch expression of Hand2,82 a bHLH transcription factor that dimerises with Twist1 in mandible development.83 84 Together, this supports the contention that deregulation of TWIST1, as we have shown in iNCCs, could contribute to the ARCND phenotype.

As Twist1 knockout in different mutant animal models leads to defective NCC migration,83 86 we performed in vitro scratch assays on iNCC derived from affected family members and controls and found significant reduction in ARCND-iNCCs. Reduced migration in early NCC stages has been found in RCPS, a craniofacial disorder also characterised by underdevelopment of mandibles and shown to be related to altered neural crest functions.17 The reduced migration found in ARCND-iNCCs is in contrast to previous data showing that reduced migration is associated with loss of Twist1 in mice,83 86 as opposed to increased expression, as found here. An explanation for this might be that the iNCCs modelled in this study are in an earlier developmental stage and/or lack factors necessary to activate endothelin signalling, which are secreted by tissues within the pharyngeal arches in vivo.87 This may account for the low DLX5/6 expression observed in our experiments. Although the iNCC derivation protocol used here is biased towards the cranial neural crest lineage,88 expression profiling assays will be necessary to further clarify the positional identity and developmental stage of iNCCs. Our observations suggest that upregulation or downregulation of TWIST1 levels can lead to reduced migration depending on the developmental stage. The migration defect could potentially explain ARCND features through a reduction in NCC reaching the first and second pharyngeal arches, resulting in malformed derivatives such as mandible and external ear, a mechanism that also seems to be involved in RCPS.17 In regard to HDAC9, even though its overexpression has been associated with increased proliferation and migration in cancer cells, to date this gene has not been associated with neural crest proliferation/migration or specification of craniofacial elements.

Furthermore, our osteogenesis analysis showed that ARCND-iNCCs have defects in their ability to form bone. We observed significantly decreased levels and activity of ALP in nMSC-derived from affected family members resulting in a decrease in matrix mineralisation, which may suggest a delay in the process of mineralisation. Interestingly, this is in opposition to the findings in nMSC of RCPS, which showed increased mineralisation.17 Even though reduced iNCC migration was observed in ARCND and RCPS, the underdeveloped mandible observed in patients with these disorders may depend on different molecular pathways. Notably, TWIST1 expression in ARCND-nMSC did not show significant differences during the osteogenic differentiation as compared with control cells. Decreased ossification could potentially be related to the dynamics of TWIST1 dimerisation at a previous stage of the cellular differentiation, as these cells were differentiated from iNCCs with higher TWIST1 levels in patients as compared with controls. Increased expression levels of TWIST1 in iNCCs would lead to an alteration in the ratios of TWIST1 homodimers and heterodimers (with E-proteins such as TCF3, TCF4 and TCF12). Studies of cranial sutures suggest an antagonistic relationship with TWIST1 homodimers activating FGF2 and osteogenic genes for ossification, while TWIST1 heterodimers, for example TWIST1-TCF3, promote mesenchymal expansion.80 89 We should also consider as an additional contributing factor to deregulated osteogenesis the role of TWIST1 in regulating osteogenesis by its direct interaction with RUNX2 in preosteoblasts.78 Interestingly, overexpression of Hand2, a partner of Twist1 in mandible differentiation and an inhibitor of Runx2, leads to delayed ossification, characterised by ALP low levels,90 which is comparable with our findings. We speculate that the reduced iNCC migration and the delayed ossification in nMSC differentiation could relate to altered expression of TWIST1 in early stages of NCC that would depend on the availability of the bHLH class partners or inhibitor of DNA-binding (ID) proteins.91 Further studies are necessary to test these hypotheses.

In summary, our data suggest that a unique 430 kb tandem duplication at the HDAC9/TWIST1 locus is pathogenic, causing deregulation of TWIST1 expression, which leads to the development of ARCND features through compromised neural crest migration and osteogenic differentiation, thus representing a novel mechanism to be investigated in the aetiology of ARCND.

Author affiliations

1Genética e Biologia Evolutiva, Universidade de São Paulo Instituto de Biociências, São Paulo, Brazil
2Clinical Genetics Group, MRC Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, University of Oxford, Oxford, UK
3Molecular Oncology Center, Hospital Sírio-Libanês, São Paulo, Brazil
Acknowledgements  We are grateful to the family for their participation in this study. We thank Gabriela Hisa, Claudia Ismania, Naiha Lourenço, Simone Bertola, Tatiana T Costa, Guilherme Kobayashi, Silvia Costa, Josiane Souza, Fabiana Poerner, Aimee Fenwick, Jim R Hughes, Simon MacGowan, Nils Koelling, Timothy Rostron, John Frankland and the Human Developmental Biology Resource. This study makes use of information generated by the DECIPHER community. A full list of centres that contributed to the generation of the data is available from https://decipher.sanger.ac.uk/about/stats and via email from decipher@sanger.ac.uk.

Contributors  VLT conducted Sanger sequencing, linkage analysis, whole exome sequencing analysis, gene prioritisation and CNV analysis. SLG-R performed in vitro cellular studies and real-time QPCR. YZ conducted targeted sequencing and Capture-C. CM performed linkage analysis. SE conducted whole exome sequencing analysis. DPM performed microsatellite experiment. HB conducted exome data processing. RSF identified patients, collected and analysed clinical data, and provided the biological specimens. JTD conducted whole exome sequencing. SRFT provided the biological specimens. YT conducted Sanger sequencing, linkage analysis, whole exome sequencing. SRFT and MRP-B wrote the manuscript. SRFT and MRP-B supervised and conceived the study. MRP-B is responsible for the overall content of the manuscript acting as guarantor. All the authors revised the manuscript for important intellectual content and approved the final version.

Funding  Work was supported by CEPRIFAPESP (2013/00820-1) and CNPq (MRPB/303712/2016-3) in Brazil and in Oxford: Action Medical Research (GNA2483 to SRFT), VITCT Foundation Fellowship (SRFT, Andrew Wilkie), the MRC through the WIMM Strategic Alliance (G0902148 and MC_UU_12025) and Wellcome (Investigator Award 102731 to Andrew Wilkie and Project Grant 093329 to Andrew Wilkie and SRFT). Funding for the DECIPHER project was provided by Wellcome.

Competing interests  None declared.

Patient consent for publication  Not required.

Ethics approval  Approval for this study was obtained from the Ethics Committee of the Instituto de Biociências - Universidade de São Paulo (CAAE: 06216813.4.0000.5464).

Provenance and peer review  Not commissioned; externally peer reviewed.

Data availability statement  All data relevant to the study are included in the article or uploaded as supplementary information.

Supplemental material  This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access  This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: https://creativecommons.org/licenses/by/4.0/.

ORCID iDs  Vanessa Luiza Romaneli Tavares http://orcid.org/0000-0002-8845-5952
Stephen R F Twigg http://orcid.org/0000-0001-5024-049X
Maria Rita Passos-Bueno http://orcid.org/0000-0002-9248-3008

References

Novel disease loci


45 Freer J,Commercial use requires permission from the publisher. © 2021 BMJ Publishing Group Ltd. Unauthorized reproduction of this article is prohibited. Published on 11 November 2021. doi:10.1136/jmedgenet-2021-107825.J Med Genet: first published as 10.1136/jmedgenet-2021-107825 on 8 November 2021. Downloaded from http://jmg.bmj.com/
Novel disease loci


A new locus underlying Auriculocondylar syndrome: 400kb duplication involving TWIST1 regulatory elements.


1Human Genome and Stem Cell Research Center, University of São Paulo, Institute of Bioscience, Department of Genetics and Evolutionary Biology.
2Clinical Genetics Group, MRC Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, United Kingdom.
3Molecular Oncology Center, Hospital Sírio-Libanês, São Paulo, Brazil (current location of the author).
4Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, United Kingdom (current location of the author).
5Leiden Genome Technology Center, Leiden University Medical Center, Leiden, Nederland.
6Centro de Atendimento Integral ao Fissurado Labial Palatal (CAIF), Curitiba, PR, Brasil.

* These authors contributed equally to the study.

** Corresponding authors: Maria Rita Passos-Bueno 106 Matão Street, 05508-090, São Paulo – SP, Brazil, Email: passos@ib.usp.br and Stephen R. F. Twigg, Clinical Genetics Group, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, United Kingdom, Email: stephen.twigg@imm.ox.ac.uk
SUPPLEMENTARY MATERIAL

Cell Characterization

Total DNA was extracted from iPSC cultures with the use of a NucleoSpin Tissue kit (Macherey-Nagel), following the supplier’s instructions. For PCR reactions, primers targeting the OriP gene present in the backbone of the episomal vectors were used (F 5′-TTCCACGAGGGTAGTGAACC-3′ and R 5′-TCGGGGGTGTTAGAGACAAC-3′), according to recommendations provided elsewhere (Epi5 Episomal iPSC Reprogramming Kit, Life Technologies). Multiplex ligation-dependent probe amplification (MLPA) analysis was performed with subtelomeric kits (P036 and P070; MRC-Holland), to detect chromosomal imbalances, as previously described [1].

iPSC were fixed in 4% paraformaldehyde for 20 minutes at room temperature, followed by permeabilization with PBS 0.2% Triton X-100 for 30 minutes at 4°C. Blocking was carried out with PBS 5% BSA for 1 hour at room temperature, followed by incubation with primary antibodies (Supplementary Table S3) in blocking solution overnight at 4°C. After washing 3x with PBS, cells were incubated with secondary antibodies in blocking solution at 4°C for 1 hour, in the dark, followed by two PBS washes and counterstaining with DAPI solution (Life Technologies) for 2 minutes at room temperature. After a final PBS wash, cells were analysed with a fluorescent microscope (Axiovision; Zeiss).

To assess the immunophenotype of NCCs and MSCs, cells were detached with Accutase and TrypLE Express respectively, and washed twice with 2 volumes of blocking solution (4% BSA in PBS without Ca²⁺ and Mg²⁺). Cells were incubated with the conjugated antibodies (Supplementary Table S3) in blocking solution in the dark for 1 hour at 4°C, washed twice with PBS, and fixed in 1% paraformaldehyde/PBS. Antibody concentrations followed the manufacturer’s recommendations. A minimum of 5,000 events were acquired in a FACS Aria II flow cytometer (BD Biosciences) and analyzed on FlowJo software (v2.5) and Guava Express PRO (Millipore).

Microscope imaging revealed iPSC, NCC, and MSC each exhibiting their typical cell morphology, suggesting a successful differentiation and no apparent difference between patients and controls with regard to cell morphology. After reprogramming iPSC, all the cells exhibited a typical pluripotent stem-cell morphology and positive staining for both
pluripotent-markers SSEA-4 and Oct3/4. All iPSCs showed no detectable signs of aneuploidy or genomic integration of the episomal vectors, thus being a suitable model for our study purposes (Supplementary Figure S6).

NCCs induced from iPSC populations were positively stained for NCC markers p75 and HNK1 with proportions of double-positive cells ranging between 74.6 and 95.8%, in both control and patients’ cells. Furthermore, RT-qPCR assays showed upregulation of NCC markers PAX3, SOX10, ZIC1 and TFAP2A, when compared with the originating iPSCs (Supplementary Figure S7).

MSCs also exhibited typical and homogeneous mesenchymal immunophenotype, with positive staining (>82%) for mesenchymal markers CD73, CD166 and CD90, and negative staining (<5%) for endothelial marker CD31 (Supplementary Figure S8). Also, MSCs could be differentiated into mesenchymal derivatives in vitro, such as osteoblasts, in agreement with the expected mesenchymal stem cell phenotype seen in previous works from our group [2–5].

References


**Supplementary Table S1:** Summary of clinical features of Family F1 diagnosed with ARCND and previously reported in Masotti *et al.*, 2008.

<table>
<thead>
<tr>
<th>Family F1</th>
<th>Masotti <em>et al.</em>, 2008, referred to F1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diagnosis / Inheritance</strong></td>
<td>ARCND / AD</td>
</tr>
<tr>
<td><strong>Familial recurrence</strong></td>
<td>+</td>
</tr>
<tr>
<td><strong>Consanguinity</strong></td>
<td>-</td>
</tr>
<tr>
<td><strong>Gender of affected patients</strong></td>
<td>F (8/11); M (3/11)</td>
</tr>
<tr>
<td>Neuropsychomotor development delay</td>
<td>0/10</td>
</tr>
<tr>
<td>Apnea</td>
<td>4/10</td>
</tr>
<tr>
<td>Respiratory difficulties</td>
<td>4/10</td>
</tr>
<tr>
<td><strong>FACE</strong></td>
<td></td>
</tr>
<tr>
<td>Facial asymmetry</td>
<td>6/10</td>
</tr>
<tr>
<td>Prominent cheeks</td>
<td>8/10</td>
</tr>
<tr>
<td><strong>MOUTH</strong></td>
<td></td>
</tr>
<tr>
<td>Microstomia</td>
<td>8/10</td>
</tr>
<tr>
<td>Abnormal palate; cleft palate</td>
<td>3/10 ; 1/10</td>
</tr>
<tr>
<td>Excess soft tissue / Atypical uvula</td>
<td>NE</td>
</tr>
<tr>
<td>Glossoptosis</td>
<td>4/9</td>
</tr>
<tr>
<td><strong>MANDIBLE</strong></td>
<td></td>
</tr>
<tr>
<td>Micognathia</td>
<td>6/10</td>
</tr>
<tr>
<td><strong>CONDYLE</strong></td>
<td></td>
</tr>
<tr>
<td>Mandibular condyle abnormality</td>
<td>3/4</td>
</tr>
<tr>
<td><strong>EARS</strong></td>
<td></td>
</tr>
<tr>
<td>Other auricular dysplasia</td>
<td>2/10</td>
</tr>
<tr>
<td>QME</td>
<td>8/10</td>
</tr>
<tr>
<td>Post-auricular tags</td>
<td>2/10</td>
</tr>
<tr>
<td>Hearing loss</td>
<td>6/6</td>
</tr>
</tbody>
</table>

F - female; M - male; ARCND - auriculocondylar syndrome; AD - autosomic dominant; NE - not evaluated.
**Supplementary Table S2.** Training lists used to run ENDEAVOUR prioritization genes. (A) Genes that have an important role in formation of the structures mainly affected in ARCND patients - ears, mandible, and mandibular condyle. (B) Genes involved in the embryonic developmental processes of the structures affected in ARCND patients; including those important in neural crest development and formation of the first and second pharyngeal arches.

<table>
<thead>
<tr>
<th>Training list A</th>
<th>Training list B</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLX3</td>
<td>BAPX1</td>
</tr>
<tr>
<td>DLX4</td>
<td>DLX3</td>
</tr>
<tr>
<td>DLX5</td>
<td>DLX5</td>
</tr>
<tr>
<td>DLX6</td>
<td>DLX6</td>
</tr>
<tr>
<td>EDN1</td>
<td>EDN1</td>
</tr>
<tr>
<td>EYA1</td>
<td>FGF8</td>
</tr>
<tr>
<td>FGF8</td>
<td>GSC</td>
</tr>
<tr>
<td>GSC</td>
<td>HAND2</td>
</tr>
<tr>
<td>HAND2</td>
<td>LHX6</td>
</tr>
<tr>
<td>OTX2</td>
<td>SNAI1</td>
</tr>
<tr>
<td>RUNX2</td>
<td>SOX9</td>
</tr>
</tbody>
</table>

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s).
Supplementary Table S3: Antibodies used for iPSC, NCC, and MSC characterization, through immunofluorescence and/or flow cytometry.

<table>
<thead>
<tr>
<th>Cell type</th>
<th>Technique</th>
<th>Antibody</th>
<th>Company (catalog#)</th>
</tr>
</thead>
<tbody>
<tr>
<td>iPSC</td>
<td>immunofluorescence</td>
<td>primary Anti-OCT4 antibody</td>
<td>Abcam (ab19857)</td>
</tr>
<tr>
<td>iPSC</td>
<td>immunofluorescence</td>
<td>primary Anti-SSEA4 antibody</td>
<td>Abcam (ab16287)</td>
</tr>
<tr>
<td>iPSC</td>
<td>immunofluorescence</td>
<td>secondary Goat anti-Rabbit IgG (H+L) antibody Alexa Fluor 546 conjugate</td>
<td>Life Technologies(A-11010)</td>
</tr>
<tr>
<td>iPSC</td>
<td>immunofluorescence</td>
<td>secondary Goat anti-Mouse IgG (H+L) Antibody, Alexa Fluor 488 conjugate</td>
<td>Life Technologies(A-11001)</td>
</tr>
<tr>
<td>NCC</td>
<td>flow citometry</td>
<td>IgM k FITC Mouse Anti-Human CD57</td>
<td>BD Pharmingen (561906)</td>
</tr>
<tr>
<td>NCC</td>
<td>flow citometry</td>
<td>IgG1 k Alexa Fluor 647 Mouse Anti-Human CD271</td>
<td>BD Pharmingen (560877)</td>
</tr>
<tr>
<td>NCC</td>
<td>flow citometry</td>
<td>FITC Mouse IgM k isotype control</td>
<td>BD Pharmingen (555583)</td>
</tr>
<tr>
<td>NCC</td>
<td>flow citometry</td>
<td>Alexa Fluor 647 Mouse IgG1 k isotype control</td>
<td>BD Pharmingen (557714)</td>
</tr>
<tr>
<td>MSC</td>
<td>flow citometry</td>
<td>FITC Mouse Anti- Human CD31</td>
<td>BD Pharmingen (555445)</td>
</tr>
<tr>
<td>MSC</td>
<td>flow citometry</td>
<td>APC Mouse Anti-Human CD73</td>
<td>BD Pharmingen (560847)</td>
</tr>
<tr>
<td>MSC</td>
<td>flow citometry</td>
<td>PE Mouse Anti-Human CD90</td>
<td>BD Pharmingen (555596)</td>
</tr>
<tr>
<td>MSC</td>
<td>flow citometry</td>
<td>FITC Mouse Anti-Human CD105</td>
<td>BD Pharmingen (551443)</td>
</tr>
<tr>
<td>MSC</td>
<td>flow citometry</td>
<td>PE Mouse Anti-Human CD166</td>
<td>BD Pharmingen (559263)</td>
</tr>
<tr>
<td>MSC</td>
<td>flow citometry</td>
<td>FITC Mouse IgG1 k Isotype Controls</td>
<td>BD Pharmingen (555748)</td>
</tr>
<tr>
<td>MSC</td>
<td>flow citometry</td>
<td>PE Mouse IgG1 k Isotype Control</td>
<td>BD Pharmingen (554681)</td>
</tr>
<tr>
<td>MSC</td>
<td>flow citometry</td>
<td>APC Mouse IgG1 k Isotype Controls</td>
<td>BD Pharmingen (555749)</td>
</tr>
</tbody>
</table>
### Supplementary Table S4. Real time PCR primer sequences.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward sequence 5′-3′</th>
<th>Reverse sequence 5′-3′</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALP</td>
<td>GATACAAGCACTCCCACCTTCATCT</td>
<td>CTGTTCAGCTCGTAGCATGTC</td>
</tr>
<tr>
<td>BARX1</td>
<td>GAAGTGGAAGAAAATAGTGCTG</td>
<td>CGCTCGTGGAAATTGAATGTCT</td>
</tr>
<tr>
<td>BGLAP</td>
<td>GGCGCTACCTGTATCAATGG</td>
<td>TCAGCCACCTCGTCACAGTC</td>
</tr>
<tr>
<td>COL1A1</td>
<td>TGCACCCACCACTGCTTAC</td>
<td>GCCATGGACTGTTGGATCATG</td>
</tr>
<tr>
<td>DLX3</td>
<td>CCGGTACGATCTACTCCAG</td>
<td>ATTTTCACCTGTCAGCTGTGTC</td>
</tr>
<tr>
<td>DLX5</td>
<td>ACCAGCCAGAAAGAAGTGA</td>
<td>CCTCTCTTGAATGCCGACCA</td>
</tr>
<tr>
<td>GAPDH</td>
<td>ATCCCACATCCACAGAGGCC</td>
<td>GGGCGAGATGAGACCCCTT</td>
</tr>
<tr>
<td>GSC</td>
<td>CGAGGAAGAAGTGGAGGTGT</td>
<td>AGCTGTCCAGTCCAAATCG</td>
</tr>
<tr>
<td>HAND2</td>
<td>CACACAGCTACATCGCTACC</td>
<td>GTATTTCGTCAGCTCCCTTCTCC</td>
</tr>
<tr>
<td>HDAC9*</td>
<td>GCATGAGAACTTGACCC</td>
<td>TTGTTGCTTTTTATGCTAGAAG</td>
</tr>
<tr>
<td>HMBS</td>
<td>AGCTTGCTCGATACAGAC</td>
<td>AGCTCTTGTTAAACGCTTT</td>
</tr>
<tr>
<td>MSX2</td>
<td>GAAGACGGACGACCCGTTGAGTA</td>
<td>TCCAAGGGTGGACGCTGTGTG</td>
</tr>
<tr>
<td>NKX3.2</td>
<td>ACCCTTAACAGGTTGATCCCC</td>
<td></td>
</tr>
<tr>
<td>RUNX2</td>
<td>AGTGGAGGACGGGAAAGTTTC</td>
<td>GTCCTGAGTTCACTAGCTG</td>
</tr>
<tr>
<td>TBP</td>
<td>GTGACCCAGACATCATGTTTTC</td>
<td>GCAACCAGGAACCTTGCC</td>
</tr>
<tr>
<td>TWIST1</td>
<td>GACCTAGATGTGACTTGGTTCCCCA</td>
<td>GCCTGTCTCGTCTCTTCTTCTT</td>
</tr>
</tbody>
</table>

*HDAC9* primers were designed to exons covered by duplication
**Supplementary Table S5.** Gene prioritization using the ENDEAVOUR program. *TWIST1* was the top ranked gene, with a p-value equal to 0.000527 and 0.000273, using training gene lists A and B, respectively. The table shows the first twenty genes.

<table>
<thead>
<tr>
<th>Rank position</th>
<th>Results using training list A</th>
<th>Results using training list B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gene Symbol</td>
<td>Gene main identifier</td>
</tr>
<tr>
<td>1</td>
<td>TWIST1</td>
<td>ENSG000000122691</td>
</tr>
<tr>
<td>2</td>
<td>EVX1</td>
<td>ENSG000000106038</td>
</tr>
<tr>
<td>3</td>
<td>HOXA1</td>
<td>ENSG000000105991</td>
</tr>
<tr>
<td>4</td>
<td>SP4</td>
<td>ENSG000000105866</td>
</tr>
<tr>
<td>5</td>
<td>HOXA3</td>
<td>ENSG000000105997</td>
</tr>
<tr>
<td>6</td>
<td>HOXA13</td>
<td>ENSG000000106031</td>
</tr>
<tr>
<td>7</td>
<td>MEOX2</td>
<td>ENSG000000106511</td>
</tr>
<tr>
<td>8</td>
<td>HOXA11</td>
<td>ENSG000000005073</td>
</tr>
<tr>
<td>9</td>
<td>ETV1</td>
<td>ENSG000000006468</td>
</tr>
<tr>
<td>10</td>
<td>HOXA2</td>
<td>ENSG000000105996</td>
</tr>
<tr>
<td>11</td>
<td>HOXA7</td>
<td>ENSG000000122592</td>
</tr>
<tr>
<td>12</td>
<td>HOXA5</td>
<td>ENSG000000106004</td>
</tr>
<tr>
<td>13</td>
<td>ITGB8</td>
<td>ENSG000000105855</td>
</tr>
<tr>
<td>14</td>
<td>HOXA9</td>
<td>ENSG000000078399</td>
</tr>
<tr>
<td>15</td>
<td>NPY</td>
<td>ENSG000000122585</td>
</tr>
<tr>
<td>16</td>
<td>SOSTDC1</td>
<td>ENSG000000171243</td>
</tr>
<tr>
<td>17</td>
<td>IGF2BP3</td>
<td>ENSG000000136231</td>
</tr>
<tr>
<td>18</td>
<td>AGR2</td>
<td>ENSG000000106541</td>
</tr>
<tr>
<td>19</td>
<td>GPNMB</td>
<td>ENSG000000136235</td>
</tr>
</tbody>
</table>
Supplementary Table S6. Patients from the DECIPHER database with copy-number gains that overlap the duplicated region found in the present ARCND family. ND – not described. (*) patient with a duplication within the ARCND duplicated region.Italic and underlined – non-specific auricular malformation. Bold – micrognathia, an alteration also found in ARCND. Refer to the main text and Supplementary Figure S5 for details. All data were extracted from DECIPHER database (https://decipher.sanger.ac.uk/; Firth et al. 2009).

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Phenotype</th>
<th>CRCh37/hg19</th>
<th>CRCh38/hg38</th>
<th>Size</th>
<th>Gene</th>
<th>Inheritance/Genotype</th>
<th>Pathogenicity/Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>393911</td>
<td>Brachycephaly, Camptodactyly of finger, Downslanted palpebral fissures,</td>
<td>7:10239-20826760</td>
<td>7:10239-20787141</td>
<td>20.78 Mb</td>
<td>182</td>
<td>Imbalance arising from a</td>
<td>Likely pathogenic</td>
</tr>
<tr>
<td></td>
<td>Downturned corners of mouth, Epicanthus, Hypertelorism, Intellectual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>balanced parental rearrangement</td>
<td></td>
</tr>
<tr>
<td></td>
<td>disability, Limitation of knee mobility, low-set ears, Prominent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Heterozygous</td>
<td></td>
</tr>
<tr>
<td></td>
<td>occiput, Rocker bottom foot, Short nose, Wide intermamillary distance,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wide nasal bridge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>393942</td>
<td>Abnormality of the pinna, Brachycephaly, Camptodactyly of finger,</td>
<td>7:10239-20826760</td>
<td>7:10239-20787141</td>
<td>20.78 Mb</td>
<td>182</td>
<td>Imbalance arising from a</td>
<td>Likely pathogenic</td>
</tr>
<tr>
<td></td>
<td>Convex nasal ridge, Corned ear, Deep gullum, Delayed closure of the</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>balanced parental rearrangement</td>
<td></td>
</tr>
<tr>
<td></td>
<td>anterior fontanelle, Downslanted palpebral fissures, Epicanthus, Finger</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Heterozygous</td>
<td></td>
</tr>
<tr>
<td></td>
<td>clinodactyly, Frontal bossing, High palate, Hypertelorism, low-set ears,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prominent occiput, Rocker bottom foot, Short neck, Short nose, Single</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>transverse palmar crease, Sparse hair, Wide cranial sutures,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wide intermamillary distance, Wide nasal bridge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>395511</td>
<td>Abnormality of the outer ear, Anteverted nares, Cat cry, Cryptorchidism,</td>
<td>7:10239-20826760</td>
<td>7:10239-20787141</td>
<td>20.78 Mb</td>
<td>182</td>
<td>Imbalance arising from a</td>
<td>Likely pathogenic</td>
</tr>
<tr>
<td></td>
<td>Depressed nasal bridge, Finger clinodactyly, Micrognathia, Micropenis,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>balanced parental rearrangement</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Muscular hypotonia, Patent ductus arteriosus, Prominent antihelix,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Heterozygous</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prominent occiput, Poutis, Rocker bottom foot, Short neck, Short nose,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Short stature, Single transverse palmar crease, Sloping forehead, Thin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>lower lip vermilion, Thin upper lip vermilion, Wide nasal bridge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>396512</td>
<td>Abnormality of the armpit, Brachycephaly, Broad palm, Delayed closure of</td>
<td>7:10239-20826760</td>
<td>7:10239-20787141</td>
<td>20.78 Mb</td>
<td>182</td>
<td>Imbalance arising from a</td>
<td>Likely pathogenic</td>
</tr>
<tr>
<td></td>
<td>the anterior fontanelle, Delayed speech and language development,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>balanced parental rearrangement</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Depressed nasal bridge, Downslanted palpebral fissures, Downturned</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Heterozygous</td>
<td></td>
</tr>
<tr>
<td></td>
<td>corners of mouth, Furrowed tongue, Genu valgum, Intellectual disability,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>low-set ears, Malar flutting, Muscular hypotonia, Proximal placement of</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>thumb, Talipes equinovarus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>280316</td>
<td>Absent speech, Constipation, Global developmental delay, Intellectual</td>
<td>7:10239-25112979</td>
<td>7:10239-25073360</td>
<td>25.06 Mb</td>
<td>230</td>
<td>Unknown Heterozygous</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td>disability, severe, Long fingers, Micrognathia, Muscular hypotonia,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Narrow mouth, Pancreatitis, Thoracolumbar scoliosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2363</td>
<td>Anteverted nares, Arachnoid cyst, Delayed cranial suture closure,</td>
<td>7:54215-18545043</td>
<td>7:54215-18505420</td>
<td>18.45 Mb</td>
<td>16</td>
<td>Unknown Heterozygous</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td>Feeding difficulties in infancy, Hypertelorism, Hypoplasia of the corpus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>callosum, Intellectual disability, Micrognathia, Pajoccephaly, Wide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>nasal bridge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>259691</td>
<td>Intellectual disability</td>
<td>7:2996438-26613851</td>
<td>7:2956804-26574232</td>
<td>23.62 MB</td>
<td>194</td>
<td>De novo Heterozygous</td>
<td>ND</td>
</tr>
<tr>
<td>250623</td>
<td>ND</td>
<td>7:0080957-29334149</td>
<td>7:0041327-29294533</td>
<td>20.25</td>
<td>161</td>
<td>Unknown Heterozygous</td>
<td>ND</td>
</tr>
<tr>
<td>Patient ID</td>
<td>Phenotype</td>
<td>CRCh37/hg19</td>
<td>CRCh38/hg38</td>
<td>Size</td>
<td>Gene</td>
<td>Inheritance/Genotype</td>
<td>Pathogenicity/Contribution</td>
</tr>
<tr>
<td>------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------</td>
<td>-----------------------</td>
<td>------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>383307</td>
<td>Abnormal facial shape, Intellectual disability, mild, Iris coloboma, Retinal coloboma, Right hemiplegia, Seizure, Short stature</td>
<td>7:10341312-22824687</td>
<td>7:10301685-22785068</td>
<td>12.48</td>
<td>74</td>
<td>De novo heterozygous</td>
<td>Pathogenic full</td>
</tr>
<tr>
<td>366699</td>
<td>Failure to thrive, Intellectual disability, mild, Microcephaly</td>
<td>7:10795079-19802691</td>
<td>7:10755452-19763068</td>
<td>9.01 Mb</td>
<td>49</td>
<td>Paternally inherited Heterozygous</td>
<td>Likely benign Uncertain</td>
</tr>
<tr>
<td>396373</td>
<td>Abnormal immunoglobulin level, Abnormality of cardiovascular system morphology, Abnormality of the pinna, Aplasia/Hypoplasia of the earlobe, Camptodactyly of finger, Deep palmar crease, Delayed closure of the anterior fontanelle, Delayed speech and language development, Downslanted palpebral fissures, Hip dislocation, Intellectual disability, Long philtrum, Microcephaly, Micrognathia, Muscular hypotonia, Prominent ear helix, Proportionate short stature, Small for gestational age, Telecanthus, Tricuspid regurgitation</td>
<td>7:13326761-30526760</td>
<td>7:13327813-30487144</td>
<td>17.20 Mb</td>
<td>162</td>
<td>De novo heterozygous</td>
<td>Likely pathogenic</td>
</tr>
<tr>
<td>280487</td>
<td>Global developmental delay</td>
<td>7:14775680-24813558</td>
<td>7:14736055-24773939</td>
<td>10.04 Mb</td>
<td>85</td>
<td>De novo Heterozygous</td>
<td>ND</td>
</tr>
<tr>
<td>326652</td>
<td>Intellectual disability, Macrocephaly</td>
<td>7:15614415-24597968</td>
<td>7:15574790-24558349</td>
<td>8.98 Mb</td>
<td>79</td>
<td>Paternally inherited Heterozygous</td>
<td>Uncertain Uncertain</td>
</tr>
<tr>
<td>300578</td>
<td>Atrial septal defect, Neonatal hypotonia</td>
<td>7:15614415-24621160</td>
<td>7:15574790-24581541</td>
<td>9.01 Mb</td>
<td>80</td>
<td>Unknown Heterozygous</td>
<td>Pathogenic full</td>
</tr>
<tr>
<td>394346</td>
<td>Abnormal thumb morphology, Abnormality of finger, Downslanted palpebral fissures, Genu valgum, High palate, Intellectual disability, Low-set ears, Macroglossia, Micrognathia, Micropenis, Muscular hypotonia, Premature birth, Promoting ear, Ptosis</td>
<td>7:16426760-37226760</td>
<td>7:16387155-37187155</td>
<td>20.80 Mb</td>
<td>214</td>
<td>Unknown Heterozygous</td>
<td>Likely pathogenic</td>
</tr>
<tr>
<td>288069</td>
<td>Intellectual disability</td>
<td>7:17299010-18730173</td>
<td>7:17259386-18609550</td>
<td>1.43 Mb</td>
<td>6</td>
<td>Unknown Heterozygous</td>
<td>Uncertain</td>
</tr>
<tr>
<td>288080</td>
<td>Intellectual disability, Microcephaly</td>
<td>7:17299010-18730173</td>
<td>7:17259386-18609550</td>
<td>1.43 Mb</td>
<td>6</td>
<td>Maternally inherited Heterozygous</td>
<td>Uncertain</td>
</tr>
<tr>
<td>255320</td>
<td>Abnormal facial shape, Cleft palate, Cognitive impairment</td>
<td>7:17759419-20010339</td>
<td>7:17755795-19970716</td>
<td>2.21 Mb</td>
<td>13</td>
<td>Inherited from parent with similar phenotype to child Heterozygous</td>
<td>ND</td>
</tr>
<tr>
<td>*276644</td>
<td>ND (however, there is no ear or mandible malformation – personal communication from Laurence Olivier-Faivre)</td>
<td>7:18520874-18700051</td>
<td>7:18481251-1860428</td>
<td>179.18 Kb</td>
<td>ND</td>
<td>HDAC9</td>
<td>Unknown heterozygous</td>
</tr>
</tbody>
</table>

**DECIPHER ANNOTATIONS**

Supplementary Figure S1. Multipoint linkage analysis performed with GeneHunter using sets of 100 and 60 markers, shown respectively at the top and bottom of the figure. The start/end of each marker set is shown through the blue bars along the x-axis. The highest LOD score was found on Chromosome 7 (see text for details).
Supplementary Figure S2. Microsatellite markers genotyping. **Top**, multipoint linkage analysis performed with Merlin using microsatellites genotypes. D7S513 (17.7 cM), D7S507 (28.7 cM), D7S493 (34.7 cM), D7S516 (41.7 cM), D7S484 (53.5 cM), D7S510 (59.9 cM), and D7S519 (69.0 cM). LOD score positive region from D7S513 to D7S484 at chr7:11651237-35284906 (hg19). **Bottom**, haplotypes, constructed using HaploPainter, showing segregation of microsatellites markers in the candidate region among affected patients (D7S513 to D7S516). Haplotypes for individuals I-1, I-2, II-6, III-4, III-9 e IV-5 were inferred by Merlin software once their DNA samples were unavailable. This analysis narrowed down linkage region from chr7:14395902-32017194 to chr7:14395902-28158440 (hg38).
Supplementary Figure S3 - Targeted resequencing of a 2.4 Mb region around TWIST1. A tandem duplication of 430.30 kb within HDAC9 is seen in affected patients IV-3, IV-6, and III-5. Individual II-7 is unaffected and does not have the duplication. Image captured with Samplot (https://www.biorxiv.org/content/10.1101/2020.09.23.310110v3).
Supplementary Figure S4. HDAC9 breakpoint evaluation in patients of F1 family. Sanger sequencing of ARCND patients showing the presence of the breakpoint region, in affected individuals only. The exact breakpoint was mapped using UCSC Blat.
Supplementary Figure S5. UCSC screenshot scheme of chr7:10,239-37,226,760 (hg19; above) showing the surrounding region of ARCND 430 kb and DECIPHER database CNVs. Below, there is a zoom in the ARCND 430 Kb duplication region (located at chr7:18,476,861-18,907,163 hg19), along with the HDAC9 scheme. We can see that not all isoforms would be disrupted by the ARCND duplication (refer to the main text). DECIPHER track was filtered by variant class to only display copy-number gain, duplication, and duplication/trip. After that, only duplications that overlap with the ARCND 430 kb CNV were included. Blue CNVs bars represent the ones with some type of auricular malformation described in the DECIPHER database. DECIPHER patients without described ear malformation are depicted as gray bars (CNVs). Refer to the main text and Supplementary Table S5 for details.
Supplementary Figure S6. Characterization of Induced Pluripotent Stem Cells (iPSC) reprogrammed from ARCND erythroblasts cells. (A) Immunofluorescence staining of pluripotency markers OCT3/4 (red) and SSEA-4 (green) in iPSCs (10x magnification); DAPI nuclear staining is in blue. (B) End-point PCR for vector backbone gene OriP; DNA from one iPSC line in which genomic integration had been detected was used as positive control (PC); (NC) negative control. (C) MLPA analysis with peak ratios for subtelomeric probes (blue dots) and control probes (green dots) using P036 and P070 kits showing no evidence of alterations.
Supplementary Figure S7. Characterization of iPSC-derived neural crest cells. (A) Biparametric flow cytometry dot plots for HNK-1/FITC and p75/Alexa Fluor 647 expression in ARCND and control cells. Values in upper right quadrants represent p75+/HNK-1+ events. (B) RT-qPCR assessment of neural crest markers (PAX3, SOX10, ZIC1 and TFAP2A) and NCCs from controls (gray) and ARCND patients (black) with n=number of biological samples; values represent mean ± SEM, (***) p-value< 0.001, (*) p-value < 0.05, Student’s t-test.
Supplementary Figure S8. Characterization of mesenchymal stem cells. Flow cytometry immunophenotype profile of MSCs showing positive staining for mesenchymal markers CD73, CD90 and CD166, and negative staining for endothelial marker CD31. Histograms represent event count (y-axis) vs. fluorescence (x-axis). Experimental data (red) were plotted in overlay with data from isotype controls (blue).
Supplementary Figure S9. mRNA expression analysis of mandible-related genes in NCC under EDN-1 treatment; not statistically significant differences were found in comparisons between controls and ARCND before and after EDN1 induction (two-way ANOVA). All values represent mean ± SEM. AU=arbitrary unit.