Impact of rare variants in *ARHGAP29* to the etiology of oral clefts: Role of loss-of-function versus missense variants

Clarice P. Savastano[1], Luciano A. Brito[1], Ágatha C. Faria[1], Núria Setó-Salvia[2,3,] Emma Peskett[2], Camila M. Musso[1], Lucas Alvizi[1], Suzana A. M. Ezquina[1], Chela James[2], GOSgene[2], Philip Beales[2], Melissa Lees[4], Gudrun E. Moore[2], Philip Stanier[2*], Maria Rita Passos-Bueno[1*]

1 Centro de Pesquisa sobre o Genoma Humano e Células-Tronco, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil;

2 Genetics and Genomic Medicine, UCL Institute of Child Health, London, UK.

3 Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK,

4 Department of Clinical Genetics, Great Ormond Street Children's Hospital, London, UK

The authors declare no conflict of interest

Running title

ARHGAP29 variants in Oral Clefts

*Corresponding authors:

Maria Rita Passos-Bueno,

Instituto de Biociências da USP, Rua do Matão 277, sala 200, Cidade Universitária, São Paulo, SP, 05508-090, Brasil. Tel.: +55 11 3091-7740, E-mail:passos@ib.usp.br

and Philip Stanier,

Genetics and Genomic Medicine, UCL Institute of Child Health, 30 Guilford Street, London, WC1N 1EH, UK. Tel.: +44 207 905 2867, E-mail: p.stanier@ucl.ac.uk

ABSTRACT

Non-syndromic cleft lip with or without cleft palate (NSCL/P) is a prevalent, complex congenital malformation. Genome-wide association studies (GWAS) on NSCL/P have consistently identified association for the 1p22 region, in which *ARHGAP29* has emerged as the main candidate gene. *ARHGAP29* re-sequencing studies in NSCL/P patients have identified rare variants; however their clinical impact is still unclear. In this study we identified ten rare variants in *ARHGAP29*, including five missense, one in-frame deletion, and four loss-of-function (LoF) variants, in a cohort of 188 familial NSCL/P cases. A significant mutational burden was found for LoF (Sequence Kernel Association Test, P=0.0005) but not for missense variants in *ARHGAP29*, suggesting that only LoF variants contribute to the etiology of NSCL/P. Penetrance was estimated as 59%, indicating that heterozygous LoF variants in *ARHGAP29* confer a moderate risk to NSCL/P. The GWAS hits in *IRF6* (rs642961) and 1p22 (rs560426 and rs4147811) do not seem to contribute to the penetrance of the phenotype, based on co-segregation analysis. Our data demonstrate that rare variants leading to haploinsufficiency of *ARHGAP29* represent an important etiological clefting mechanism, and genetic testing for this gene might be taken into consideration in genetic counseling of familial cases.

Keywords: Cleft lip and palate, GWAS, haploinsufficiency, IRF6, nonsense mutations, penetrance, rare variants, 1p22.

INTRODUCTION

Non-syndromic cleft lip with or without cleft palate (NSCL/P) represents one of the most common congenital human malformations, affecting about one in 700 live born children worldwide, varying according to ethnicity and socioeconomic status. Both genetic and environmental factors contribute to the etiology of NSCL/P, and most cases fit a multifactorial pattern of inheritance [1,2].

Genome-wide association studies (GWAS) have successfully detected several common susceptibility alleles for NSCL/P [3–8]. The 1p22.1 region ranks among the most frequently replicated GWAS hits, originally implicating the gene *ABCA4* [5]. However, this gene was largely excluded through its primarily retinal expression and known role in retinal disorders [9]. Single Nucleotide Variants (SNVs) in the 1p22.1 region have also been suggested to affect a cis-enhancer of *ARHGAP29* [10]. *ARHGAP29* (Online Mendelian Inheritance in Man - OMIM: 610496) is specifically expressed in the frontonasal and lateral prominences as well as the palatal shelves of murine embryos, which further reinforces it as a strong candidate underlying NSCL/P in the 1p22.1 region [11]. About eighteen possibly pathogenic rare variants (eleven missense and seven loss-of-function) in *ARHGAP29* have been reported in NSCL/P patients of European, Asian and African ancestries [11–14]. However, it is not clear if both missense and loss-of-function (LoF) variants contribute to the phenotype. In order to address this issue, a systematic analysis of *ARHGAP29* was conducted in a large cohort of 188 familial NSCL/P cases. We have also investigated if previously identified GWAS hits at 1p22.1 and in *IRF6* contribute to the penetrance of the phenotype in individuals with pathogenic variants in *ARHGAP29*.

SUBJECTS AND METHODS

Ethical Compliance

This study was approved by the Ethics Committee of the Instituto de Biociências (Universidade de São Paulo, Brazil) (CAAE: 37287314.6.0000.5464) and Great Ormond Street Hospital for Children NHS Trust Ethics Committee (REC No. 08H0713/46). Biological samples were obtained after signed informed consent by the patients, parents or legal guardians.

Samples

The NSCL/P cohort included 173 families from Brazil and fifteen families from the United Kingdom. The average number of affected individuals per family was 2.6 (ranging between two and seven), with coefficients of relationship (r) between ½ and 1/32. *ARHGAP29* sequences were screened for rare variants in 188 probands and sixteen relatives (four affected and twelve non-affected), which were included for segregation analysis. Brazilian individuals were ascertained at the Hospital das Clínicas of Universidade de São Paulo (São Paulo, Brazil), Hospital de Clínicas de Porto Alegre (Rio Grande do Sul, Brazil) or during missions of Operation Smile Brazil, in the Brazilian states of Ceará, Alagoas, Pará, Rondônia and Rio de Janeiro. British individuals were ascertained at Great Ormond Street Hospital for Children (London, United Kingdom).

DNA samples were extracted either from whole blood (following standard protocols) or saliva (collected with Oragene® DNA Collection Kits OG-500 and OG-575, and purified following prepIT-L2P manufacturer's instructions; DNA Genotek Inc., Ottawa, Canada). For controls, we used in-house whole exome sequencing data from 609 Brazilian and 601 British controls as well as public databases (1000 Genomes Project [15], Exome Variant Server/NHLBI ESP exomes [16], and Exome Aggregation Consortium – ExAC [17]).

ARHGAP29 variant screening

ARHGAP29 coding regions and exon-intron boundaries were sequenced using next-generation sequencing (NGS): twenty seven whole exome sequences from fifteen independent UK families (fifteen probands, three affected cousins and nine unaffected relatives), and 173 targeted gene sequences from unrelated probands from the Brazilian cohort were included in the analysis. Sanger sequencing was used for one affected and three non-affected relatives from the Brazilian cohort.

ARHGAP29 targeted sequencing was performed with Illumina MiSeq (Illumina, San Diego, CA, USA) sequencer, using Illumina's Nextera kits for library preparation. KAPA Library Quantification kit (KAPA Biosystems) was used to quantify the libraries by real-time quantitative PCR. Whole-exome sequencing of the British samples was conducted using Agilent Exome v4 51Mb Capture Technology and enriched libraries were sequenced on an Illumina HiSeq2000 (Illumina). Sequence alignment, data processing, variant calling, and variant annotation were performed with Burrows-Wheeler Aligner (BWA; http://bio-

bwa.sourceforge.net), Picard (http://broadinstitute.github.io/picard/), Genome Analysis Toolkit package (GATK; https://www.broadinstitute.org/gatk/) and ANNOVAR (http://www.openbioinformatics.org/annovar/), respectively.

We considered as rare variants those with a frequency below 0.5% in public databases (ExAC, Exome Variant Server, and 1000 Genomes) and in our in-house control databases. Missense variants were considered as possibly pathogenic only if predicted to be possibly/probably damaging in at least three out of four *in silico* tools (Polyphen HumDiv and HumVar [18], SIFT [19], Mutation Taster [20] and LRT [21]). Synonymous and UTR variants were excluded due to the uncertainty of their functional relevance. Splice site predictions were performed with Human Splicing Finder 3.0 [22].

All *ARHGAP29* rare variants detected by NGS were visually inspected using the Integrative Genomics Viewer software (Broad Institute of MIT and Harvard). Indels and low coverage (<50x) variants were subsequently validated by Sanger sequencing.

PCR primers for Sanger sequencing are described by Leslie et al. [12]. Capillary electrophoresis was performed on an ABI3730 DNA Analyzer (Applied Biosystems, USA) and sequences were visualized using Sequencher® 5.2 sequence analysis software (Gene Codes, USA). The genomic position of variants are based on the hg19/GRCH37 version of the human reference genome (Genome Reference Consortium - http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/), and cDNA positions refer to the sequence NM_004815.3 (NCBI Reference Sequence Database - http://www.ncbi.nlm.nih.gov/refseq/). Pathogenic variants were submitted to the ClinVar public database (http://www.ncbi.nlm.nih.gov/clinvar/).

Genotyping of GWAS hits

Genotypes for rs560426 (1p22), rs4147811 (1p22), and rs642961 (1q32) were obtained with the Illumina GoldenGate VeraCode assay, on Illumina BeadXpress platform, following manufacturer´s instructions, or by Sanger sequencing. Primers and PCR amplification conditions are available on request.

Statistical analyses

A gene-based Sequence Kernel Association Test (SKAT) was used to investigate the overall burden of *ARHGAP29* rare variants among patients in comparison to 1210 Brazilian and British controls. Two-tailed Fisher's exact test was performed to compare the proportion of patients and controls carrying rare variants in *ARHGAP29*. For these tests, rare variants were included after adjusting the sequence windows covered in all patients. Statistical significance was considered as $P \leq 0.05$. Penetrance was estimated using the PenCalc program [23].

RESULTS

Sequencing analysis of coding regions and exon-intron boundaries of *ARHGAP29* in 188 unrelated affected probands from familial cases of NSCL/P, led to the identification of ten rare variants. Five out of these ten were missense changes already described in public databases. The remaining five were unique variants, four of which were predicted to be LoF

and one was an in-frame deletion (Table 1). Among the missense variants, only c.91C>T (p.(L31F)) was predicted to be possibly pathogenic by three out of four *in silico* programs. However, segregation analysis in family BC84 excluded a likely causal role since the missense variant was absent in another affected relative. The non-frameshift deletion c.3326_3328delCAA (p.(T1109del)) was predicted to be non-pathogenic by two out of four *in silico* tools, so it was also not considered further.

The four LoF variants, including three splice site and one stopgain variant, were all predicted to disrupt the protein, based on *in silico* analysis: the splice site variants were predicted to alter canonical donor (c.2109+1G>A, intron 18 and c.1576+1G>A, intron 14) or acceptor sites (c.698-1G>C, intron 7), according to the HSF tool; meanwhile, the variant c.1475C>A was predicted to cause premature termination of protein synthesis at codon 492 (p.(S492*)) in exon 14 and lead to nonsense mediated decay (Mutation Taster; probability = 1). It was possible to investigate segregation of LoF variants in two of these families (F4118 and 314) given the availability of DNA from other affected and unaffected relatives. In both cases variants segregated with NSCL/P in accordance with an autosomal dominant pattern with incomplete penetrance (Figure 1).

A higher proportion of *ARHGAP29* rare variants were found in NSCL/P probands as compared to controls (Supplementary Table 1) with a borderline level of significance (SKAT: P=0.06; Fisher´s exact: P=0.08; patients: 11/188, 5.85%; controls: 37/1210, 3.06%). Splitting the analysis by variant type, no significant difference in the distribution of missense variants with pathogenic *in silico* predictions was observed between groups (SKAT: P=0.35; Fisher: P=0.51; patients: 1/188, 0.53%; controls: 4/1210, 0.33%). On the other hand, the number of

LoF variants was significantly higher in patients (SKAT: P=0.0005; Fisher´s test: P=0.001; patients: 4/188, 2.13%; controls: 1/1210, 0.08%).

Based on our findings, we hypothesized that the LoF variants are most likely implicated with NSCL/P. All four patients with LoF variants in *ARHGAP29* have cleft lip (bilateral or unilateral) and cleft palate (Figure 1). Phenotype expressivity among affected relatives ranged from a lip scar to a bilateral cleft lip/palate. Considering that NSCL/P segregates in an autosomal dominant model, penetrance in the four families was estimated as 0.595 (CI 95%: 0.375-0.803).

To investigate whether penetrance effects in our families could be explained by common variants, we next evaluated if the LoF variants in *ARHGAP29* were present in *trans* or in *cis* with the at-risk alleles of loci rs560426 and rs4147811, at 1p22, which were shown to be associated with NSCL/P by GWAS. We observed that all the six affected genotyped individuals were homozygous for at least one of these SNVs while one of the two non-penetrant individuals was also homozygous for one of these SNVs (Figure 1).

To take this idea further, we also tested if the at-risk allele of rs642961 [24], located in the *IRF6* regulatory region, could influence the penetrance in individuals with *ARHGAP29* LoF variants. This locus was also selected as it was previously suggested to be within the same pathway as *ARHGAP29* [11,25]. We observed that four of the nine individuals genotyped for rs642961 harbored the at-risk allele A, two were affected (Figure 1, Family 314 individual II-4 and Family F4118 individual II-2), while the two other individuals were non-penetrant mutation carriers (Figure 1, Family 314 individual II-2 and Family F4118 individual I-1).

These results suggest therefore that this SNP is unlikely to have a strong epistatic interaction with *ARHGAP29*.

DISCUSSION

Exome and genome analyses have been contributing exponentially to the description of novel rare variants in several healthy and disease populations, particularly in those of European ancestry. Consequently, one of the current challenges is to distinguish variants leading to phenotypic variability from those that do not, particularly those in the heterozygous state. These difficulties are illustrated by the finding that healthy human genomes harbor about 100-250 LoF variants per individual [26]. Studies of the distribution of rare variants in different populations with the phenotypes of interest are a possible approach to validate pathogenic variants.

In the present study, *ARHGAP29* mutation screening revealed five unique, rare heterozygous variants (one non-frameshift deletion and four LoF) and five already described rare missense variants in 188 NSCL/P families. *In silico* predictions of protein damage, segregation analysis and aggregation tests indicated that only the LoF variants, but not missense variants, in *ARHGAP29* represent a major genetic risk factor for NSCL/P in our cohort. Indeed, according to ExAC database, while *ARHGAP29* tolerates missense variants (ExAC Z-score: -0.27), it does not seem to tolerate LoF variants (ExAC Probability of LoF Intolerance: 1.00) [17]. Structural variants involving *ARHGAP29* are uncommon in the Database of Genomic Variants [27] while four of the five deletions in the Decipher database [28] involving *ARHGAP29*

are described with facial anomalies, including cleft palate. Further, down-regulation of *ARHGAP29* by genetic and epigenetic changes in both alleles have also been shown to be important in the phenotypic determination of mantle cell lymphomas [29,30]. Therefore, while there are evidence that LoF variants in *ARHGAP29* contribute to the NSCL/P phenotype, missense variants should be functionally tested before their definition as pathogenic.

The prevalence of rare LoF variants in our sample (2.1%) was higher than those observed in other NSCL/P large studies (0.2% - 0.5%) [11,12]. Possibly, these differences reflect the familial enrichment of multiplex families over population-dependent factors in our cohort. Interestingly, an enrichment of *CDH1* variants among familial cases (15%) were also observed in a previous study of our group [31].

There is no evidence of a mutational hotspot within *ARGHAP29* (Figure 2), although the modest number of known LoF variants may obscure such a finding. In our families, these variants are associated with a broad spectrum of inter and intrafamilial clinical variability, as reported by others [11–14] (Supplementary Table 2).

In all four families with LoF *ARHGAP29* variants, NSCL/P segregated according to an autosomal dominant inheritance model with incomplete penetrance (59%). We hypothesised that the NSCL/P penetrance in individuals with the LoF variants in *ARHGAP29* could be modified by the presence of at-risk alleles identified by GWAS. Even though our data is based on a small sample, these preliminary results suggest that the SNVs at 1p22 or rs642961 at *IRF6* do not significantly contribute to the penetrance and do not support an interaction between *ARHGAP29* and *IRF6*.

In summary, this study expands the mutational repertoire implicating *ARHGAP29* with NSCL/P and provides evidence that heterozygous LoF variants in this gene confer a moderate risk to the disease and may be an important genetic factor at 1p22 driving the phenotype. In addition, given the incomplete penetrance observed, additional mechanisms may be required to trigger the phenotype. The study of multiplex families has proved to be an effective strategy to identify rare variants with moderate to high effect on NSCL/P since genetic factors that contribute to the etiology of the disease are likely to be more prevalent in this group of patients. Likewise, considering the relatively high proportion of families positive for LoF in ARHGAP29 (~2%), sequencing of this gene might be taken into consideration in genetic counseling of familial cases.

REFERENCES

1. Mossey PA, Little J, Munger RG, Dixon MJ, Shaw WC. Cleft lip and palate. Lancet 2009: 374 (9703): 1773–1785.

2. Dixon MJ, Marazita ML, Beaty TH, Murray JC. Cleft lip and palate: understanding genetic and environmental influences. Nat Rev Genet 2011: 12 (3): 167–178.

3. Birnbaum S, Ludwig KU, Reutter H, et al. Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. Nat Genet 2009: 41 (4): 473–477.

4. Mangold E, Ludwig KU, Birnbaum S, et al. Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. Nat

Genet 2010: 42 (1): 24–26.

5. Beaty TH, Murray JC, Marazita ML, et al. A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. Nat Genet 2010: 42 (6): 525–529.

6. Ludwig KU, Mangold E, Herms S, et al. Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. Nat Genet 2012: 44 (9): 968–971.

7. Sun Y, Huang Y, Yin A, et al. Genome-wide association study identifies a new susceptibility locus for cleft lip with or without a cleft palate. Nat Commun 2015: 6 (76): 6414.

8. Leslie EJ, Carlson JC, Shaffer JR, et al. 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 2016: Mar 30 [Epub ahead of print].

9. Burke TR, Tsang SH. Allelic and Phenotypic Heterogeneity in ABCA4 mutations. Ophthalmic Genet 2011: 32 (3): 165–174.

10. Attanasio C, Nord AS, Zhu Y, et al. Fine Tuning of Craniofacial Morphology by Distant-Acting Developmental Enhancers. NIH Public Access 2014: 342 (6157): 1–20.

11. Leslie EJ, Mansilla MA, Biggs LC, et al. 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 Res Part A

- Clin Mol Teratol 2012: 94 (11): 934–942.

12. Leslie EJ, Taub MA, Liu H, et al. Identification of functional variants for cleft lip with or without cleft palate in or near PAX7, FGFR2, and NOG by targeted sequencing of GWAS loci. Am J Hum Genet 2015: 96 (3): 397–411.

13. Butali A, Mossey P, Adeyemo W, et al. Rare functional variants in genome-wide association identified candidate genes for nonsyndromic clefts in the African population. Am J Med Genet Part A 2014: 164 (10): 2567–2571.

14. Chandrasekharan D, Ramanathan A. Identification of a novel heterozygous truncation mutation in exon 1 of ARHGAP29 in an Indian subject with nonsyndromic cleft lip with cleft palate. Eur J Dent 2014: 8 (4): 528–532.

15. Consortium T 1000 GP. A global reference for human genetic variation. Nature 2015: 526 (7571): 68–74.

16. NHLBI GO Exome Sequencing Project (ESP). Exome Variant Server [Internet]. Available from: http://evs.gs.washington.edu/EVS/ [accessed 2015 Dec 1]

17. Lek M, Karczewski K, Minikel E, et al. Analysis of protein-coding genetic variation in 60,706 humans. bioRxiv 2015: 1–26.

18. Adzhubei I, Jordan DM, Sunyaev SR. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. In: Current Protocols in Human Genetics. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2013: 7.20.1–7.20.41.

19. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res 2001:

11 (5): 863–874.

20. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 2010: 7 (8): 575–576.

21. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. Genome Res 2009: 19 (9): 1553–1561.

22. Desmet F-O, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res 2009: 37 (9): e67–e67.

23. Horimoto ARVR, Onodera MT, Otto PA. PENCALC: A program for penetrance estimation in autosomal dominant diseases. Genet Mol Biol 2010: 33 (3): 455–459.

24. Rahimov F, Marazita ML, Visel A, et al. Disruption of an AP-2α binding site in an IRF6 enhancer is strongly associated with cleft lip. Nat Genet 2008: 40 (11): 1341–1347.

25. Biggs LC, Naridze RL, DeMali KA, et al. Interferon regulatory factor 6 regulates keratinocyte migration. J Cell Sci 2014: 127 (Pt 13): 2840–2848.

26. MacArthur DG, Tyler-Smith C. Loss-of-function variants in the genomes of healthy humans. Hum Mol Genet 2010: 19 (R2): 125–130.

27. MacDonald JR, Ziman R, Yuen RKC, Feuk L, Scherer SW. The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Res 2014: 42 (Database issue): D986–D992.

28. Firth HV, Richards SM, Bevan AP, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet 2009: 84 (4): 524–533.

29. Ripperger T, von Neuhoff N, Kamphues K, et al. Promoter methylation of PARG1, a novel candidate tumor suppressor gene in mantle cell lymphomas. Haematologica 2007: 92 (4): 460–468.

30. Schraders M, Jares P, Bea S, et al. Integrated genomic and expression profiling in mantle cell lymphoma: Identification of gene-dosage regulated candidate genes. Br J Haematol 2008: 143 (2): 210–221.

31. Brito LA, Yamamoto GL, Melo S, et al. Rare Variants in the Epithelial Cadherin Gene Underlying the Genetic Etiology of Nonsyndromic Cleft Lip with or without Cleft Palate. Hum Mutat 2015: 36 (11): 1029–1033.

TITLES AND LEGENDS TO FIGURES

Figure 1: Pedigrees of the NSCL/P families showing genotypes for *ARHGAP29* LoF variants, 1p22 SNVs (rs560426, rs4147811) and *IRF6* SNV (rs642961). At-risk alleles are given in bold; the black and gray line indicate the recombination point; unavailable genotypes are indicated with "?".

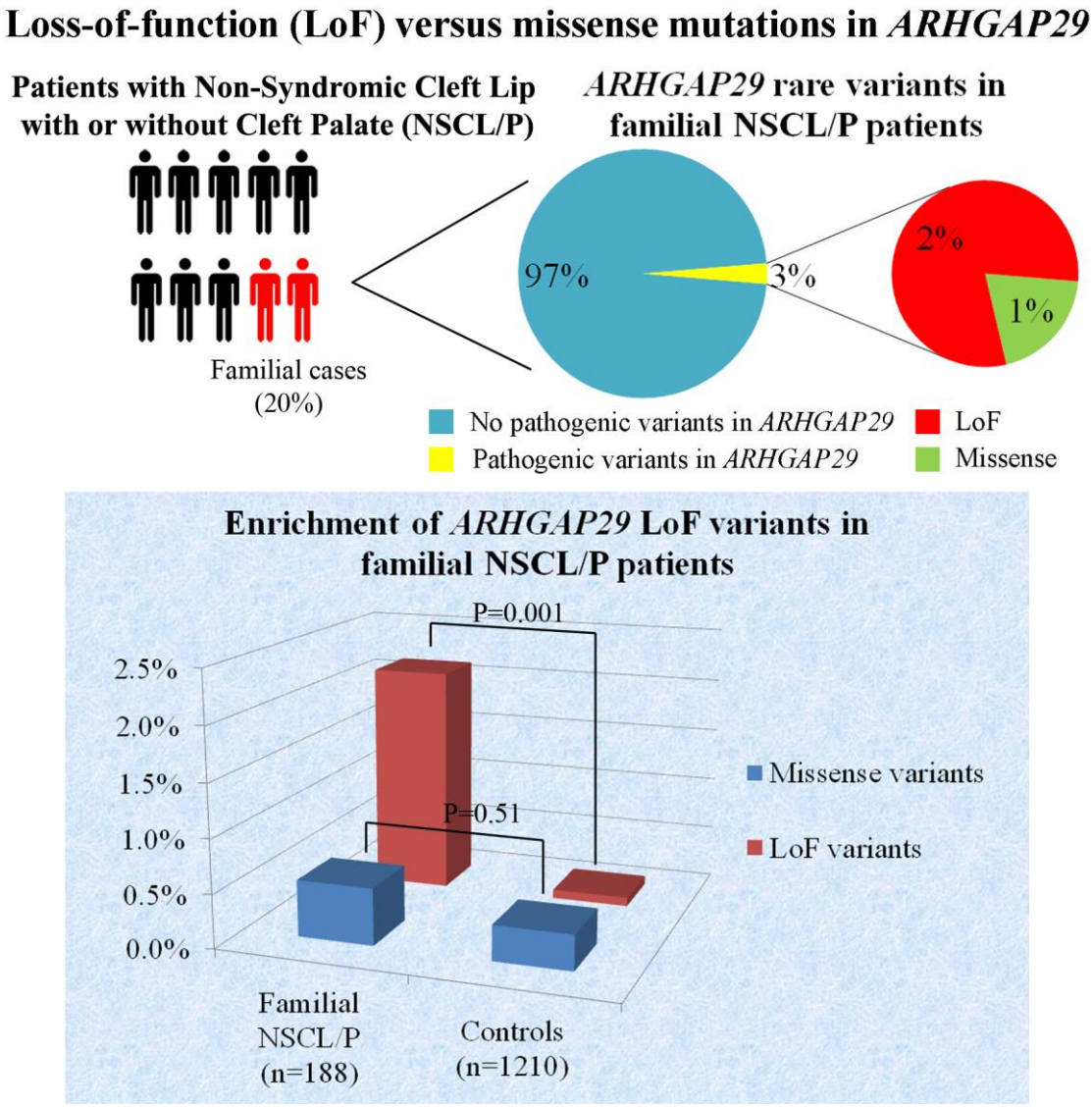
Figure 2: Scheme of the *ARHGAP29* protein showing the distribution of rare germ-line LoF variants observed in NSCL/P patients. Variants described in our study are on top of the image and the variants described in the literature are below.

ACKNOWLEDGEMENTS

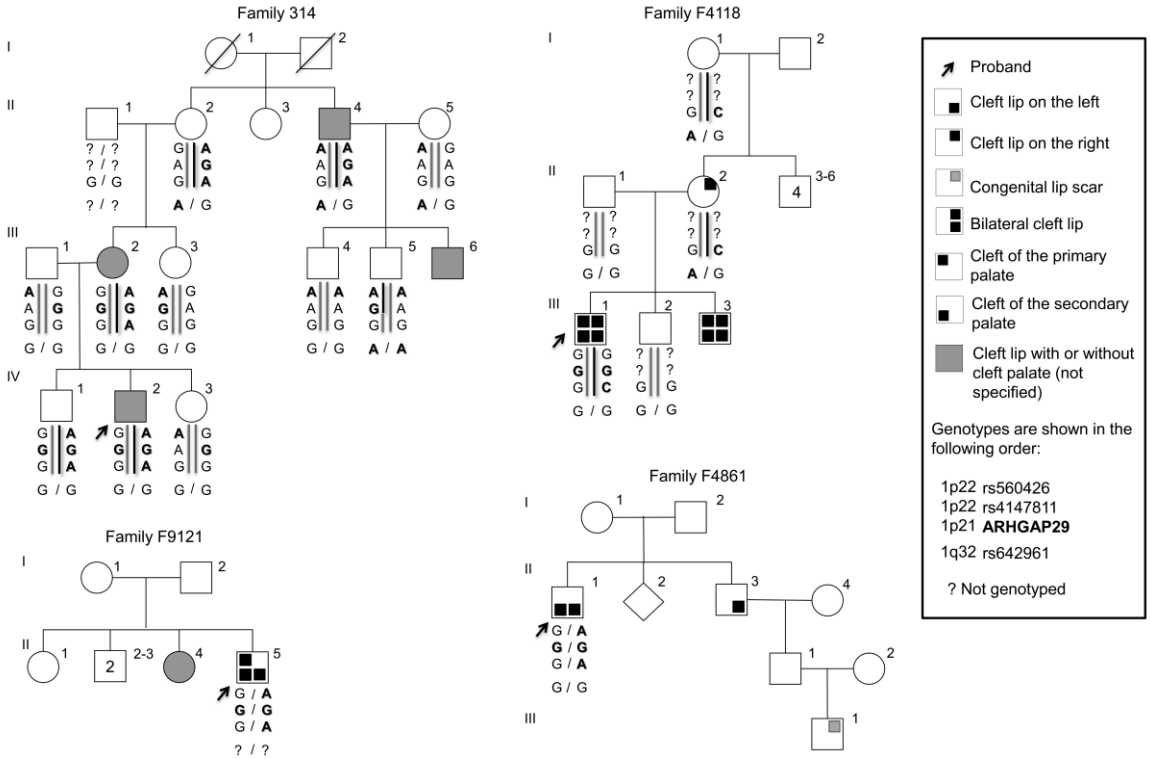
This research was supported by grants from CNPq (402425/2014-6; 457643/2014-5) and FAPESP/CEPID (2013/08028-1). We are also indebted to the CLEFT charity for providing a grant for the UK arm of the study. PS, PB, GEM and GOSgene are supported by the National Institute for Health Research Biomedical Research Centre at Great Ormond Street Hospital for Children NHS Foundation Trust and University College London. We are grateful to Osama Alsmadi and Dinu Antony for exome sequencing at the Dasman Diabetes Institute Genome Centre (Kuwait City). We also thank the Operation Smile Brazil team, Dr. Nivaldo Alonso (Hospital das Clinicas), and Dr. Daniela Bueno (Hospital Menino Jesus) for the productive collaboration.

CONFLICT OF INTEREST

The authors declare no conflict of interest



Graphical Abstract


Family 314
Family F4118
Family F9121
Family F4861
Proband
Cleft lip on the left
Cleft lip on the right
Congenital lip scar
Bilateral cleft lip
Cleft of the primary palate
Cleft of the secondary palate
Cleft lip with or without cleft palate (not specified)
Genotypes are shown in the following order:
1p22 rs560426
1p22 rs4147811
1p21 ARHGAP29
1q32 rs642961
? Not genotyped

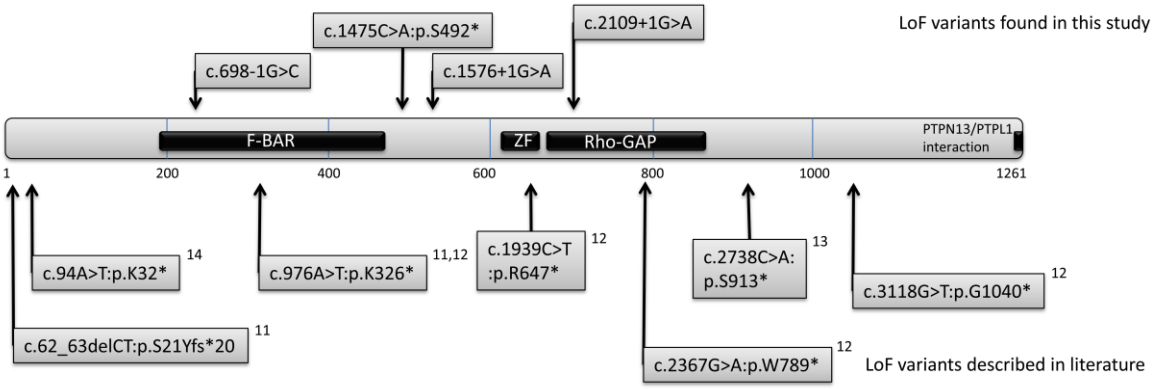

c.1475C>A:p.S492*
c.2109+1G>A
LoF variants found in this study
c.698-1G>C
c.1576+1G>A
F-BAR
ZF
Rho-GAP
PTPN13/PTPL1 interaction
1
200
400
600
800
1000
1261
c.94A>T:p.K32* 14
c.976A>T:p.K326* 11,12
c.1939C>T :p.R647* 12
c.2738C>A: p.S913* 13
c.3118G>T:p.G1040* 12
c.62_63delCT:p.S21Yfs*20 11
c.2367G>A:p.W789* 12
LoF variants described in literature

Table 1 ARHGAP29 rare variants observed in the non-syndromic cleft lip with or without cleft palate (NSCL/P) probands.

Proband ID	Chromosome position	cDNA/Protein alteration	NGS study	Read depth; allelic balance (ref, alt)	Variant type	rs	Tools predicting pathogenicity	Exac AC[1] (n. homoz)	1000G[2] AC[1]	EVS[3] AC[1]	NSCL/P probands AC[1]	Brazilian Controls AC[1]	British Controls AC[1]
F7595-1	1:94639872A>C	c.3339T>G p.(I1113M)	Panel of genes	50; 28,22	missense		1	1	0	0	1	0	0
F1843-1	1:94639883-94639885delTTG	c.3326_3328delCAA p.(T1109del)	Panel of genes	70; 30,40	nonframeshift deletion		2	0	0	0	1	0	0
F7399-1	1:94643587G>A	c.2617C>T p.(R873C)	Panel of genes	145; 63,82	missense	rs147929232	2	10	0	0	1	0	0
F6474-1; F9121-1	1:94645368C>T	c.2393G>A p.(R798Q)	Panel of genes	101; 53,49	missense	rs41311172	2	223 (1)	7	33	2	3	0
F9121-1	1:94650427C>T	c.2109+1G>A	Panel of genes	62; 28,33	splicing site		1*	0	0	0	1	0	0
314-379	1:94654771C>T	c.1576+1G>A	Exome sequencing		splicing site		1*	0	0	0	1	0	0
F4861-1	1:94654873G>T	c.1475C>A p.(S492*)	Panel of genes	80; 43,37	stopgain		NA	0	0	0	1	0	0
F8041-1	1:94667305C>T	c.G1252A p.(V418I)	Panel of genes	181; 111,70	missense	rs148959325	1	181 (1)	6	0	1	0	1
F4118-1	1:94669551C>G	c.698-1G>C	Panel of genes	202; 101,98	splicing site		1*	0	0	0	1	0	0
BC84	1:94697077G>A	c.91C>T p.(L31F)	Exome sequencing	50; 20,30	missense		3	1	0	0	1	0	1

1- Allele Count

2- 1000 Genomes Project.

3- Exome Variant Server.

* Pathogenicity evaluated only by the Human Splicing Finder 3.0 tool

Chromosome positions are based on human genome references hg19/GRCH37

cDNA position are based on RefSeq NM_004815.3

Supplementary Table 1: *ARHGAP29* rare variants found in Brazilian and UK controls.

Chromosome position	cDNA/Protein alteration	Variant type	rs	Tools predicting pathogenicity	Exac AC[1] (n. homoz)	1000 Genomes AC[1] (n. homoz)	Exome Variant Server AC[1]	CLP probands AC[1] (AN[2] = 376)	Brazilian Controls AC[1] (AN[2] = 1218)	British Controls AC[1] (AN[2] = 1201)
1:94639475G>A	c.3736C>T:p.(R1246*)	stopgain			1	0	0	0	0	1
1:94639607C>T	c.3604G>A:p.(V1202M)	missense	rs14458552	0	18	0	0	0	0	1
1:94639940T>C	c.3271A>G:p.(T1091A)	missense		0	0	0	0	0	0	2
1:94639993T>C	c.3218A>G:p.(D1073G)	missense		2	15	0	0	0	1	0
1:94640030C>T	c.3181G>A:p.(A1061T)	missense	rs14165333	0	25 (1)	0	3	0	1	0
1:94640129C>G	c.3082G>C:p.(A1028P)	missense	rs14622235	0	332 (6)	30	0	0	0	4
1:94640180G>A	c.3031C>T:p.(P1011S)	missense		2	2	0	0	0	1	0
1:94640188C>T	c.3023G>A:p.(R1008K)	missense	rs14063889	2	25	3	7	0	1	0
1:94640208C>G	c.3003G>C:p.(R1001S)	missense		1	0	0	0	0	0	6
1:94643236C>T	c.2837G>A:p.(R946Q)	missense		0	5	1	1	0	0	1
1:94643436A>G	c.2768T>C:p.(M923T)	missense		0	0	0	0	0	0	1
1:94643616G>A	c.2588C>T:p.(S863F)	missense		3	0	0	0	0	0	1
1:94643619A>G	c.2585T>C:p.(I862T)	missense	rs36860478	3	6	0	2	0	1	0
1:94645327C>T	c.2434G>A:p.(812A>T)	missense		0	0	0	0	0	1	0
1:94645368C>T	c.2393G>A:p.(R798Q)	missense	rs41311172	2	223 (1)	7	33	2	3	0
1:94654810A>G	c.1538T>C:p.(I513T)	missense	rs61758880	0	66	1	19	0	0	3
1:94667305C>T	c.G1252A:p.(V418I)	missense	rs14895932	1	181 (1)	6	0	1	0	1
1:94668724T>C	c.805A>G:p.(N269D)	missense		2	2	0	0	0	0	2
1:94668734A>C	c.795T>G:p.(N265K)	missense	rs61743723	1	25	6	9	0	1	0
1:94674437T>C	c.461A>G:p.(D154G)	missense	rs36932915	3	2	0	2	0	1	0
1:94685903C>T	c.251G>A:p.(R84H)	missense	rs18341043	0	76	12 (1)	2	0	0	1
1:94697077G>A	c.91C>T p.(L31F)	missense		3	1	0	0	1	0	1
1:94697111T>G	c.57A>C:p.(Q19H)	missense		1	5	0	0	0	1	0

1- Allele Count

2- Allele Number

Chromosome positions are based on human genome references hg19/GRCH37

cDNA position are based on RefSeq NM_004815.3

Supplementary Table 2: ATHGAP29 LOF variants in the literature

Chromosome position	cDNA/Protein alteration	Variant type	Reference	Exac MAF[1]	Exac AC[2]	NSCL/P probands	Controls[3]	Phenotype	Inheritance
1:94640093C>A	c.3118G>T p.(G1040*)	stopgain	Leslie et al., 2015	0	0	1	0	Cleft lip and palate	Unspecified parent
1:94643466G>T	c.2738C>A p.(S913*)	stopgain	Butali et al., 2014	0*	0*	1	0	Cleft lip	Not avaliable
1:94645394C>T	c.2367G>A p.(W789*)	stopgain	Leslie et al., 2015	0	0	1	0	Cleft lip and palate	Unspecified parent
1:94650427C>T	c.2109+1G>A	splicing site	This study	0	0	1	0	Unilateral cleft lip and palate	Unspecified parent
1:94650598G>A	c.1939C>T p.(R647*)	stopgain	Leslie et al., 2015	0	0	1	0	Cleft lip and palate	Unspecified parent
1:94654771C>T	c.1576+1G>A	splicing site	This study	0	0	1	0	Oral cleft not specified (proband, mother and great-uncle), unnafected grandmother and brother	Mother
1:94654873G>T	c.1475C>A p.(S492*)	stopgain	This study	0	0	1	0	Unilateral cleft lip and palate (proband), unilateral cleft lip (brother)	Unspecified parent
1:94668267T>A	c.976A>T p.(K326*)	stopgain	Leslie et al., 2012; Leslie et al., 2015	0	0	1	0	Bilateral cleft lip and palate (proband),unnaffected mother and grandfather	Mother
1:94669551C>G	c.698-1G>C	splicing site	This study	0	0	1	0	Bilateral cleft lip and palate (proband and brother), unilateral cleft lip (mother), unnafected grandmother	Mother
1:94697074T>A	c.94A>T p.(K32*)	stopgain	Chandrasekharan & Ramanathan, 2014	0	0	1	0	Cleft lip and palate	Not avaliable
1:94697105-94697106delAG	c.62_63delCT p.(S21Yfs*20)	frameshift deletion	Leslie et al., 2012	0.00003297	1	1	0	Bilateral cleft lip and palate (proband), unnaffected sibling	Unspecified parent

1- Minor Allele Frequency.
2- Allele Count.
3- Control subjects included in each study
* A different variant in the same codon is described in Exac
Chromosome positions are based on human genome references hg19/GRCH37
cDNA position are based on RefSeq NM_004815.3