

# **Craniofrontonasal syndrome caused by introduction of a novel uATG in the 5'UTR of *EFNB1***

**Authors:** Vanessa L. Romanelli Tavares<sup>a</sup>, Erika Kague<sup>a,b</sup>, Camila M. Musso<sup>a</sup>, Thiago G.P. Alegria<sup>c</sup>, Renato S. Freitas<sup>d</sup>, Debora Bertola<sup>a,e</sup>, Stephen R.F. Twigg<sup>f</sup>, Maria Rita Passos-Bueno<sup>a</sup>

<sup>a</sup>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, São Paulo, Brazil

<sup>b</sup>Department of Physiology, Pharmacology and Neuroscience, University of Bristol, Bristol, UK

<sup>c</sup>Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, São Paulo, Brazil

<sup>d</sup>Centro de Atendimento Integral ao Fissurado Lábio Palatal (CAIF), Curitiba, Paraná, Brazil

<sup>e</sup>Instituto da Criança do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, São Paulo, Brazil

<sup>f</sup>Clinical Genetics Group, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK

**Running title:** A variant-created uATG in *EFNB1*

**Corresponding author:** Maria Rita Passos-Bueno, Instituto de Biociências, Universidade de São Paulo, Rua do Matão, 277, sala 200, São Paulo - SP 05508-090, Brazil. Phone : +55 11 3091-9910. E-mail: [passos@ib.usp.br](mailto:passos@ib.usp.br)

## **Abstract**

Craniofrontonasal syndrome (CFNS) is an X-linked disorder caused by *EFNB1* mutations, in which females are more severely affected than males. Severe male phenotypes are associated with mosaicism, supporting cellular interference for sex bias in this disease. Although many variants have been found in the coding region of *EFNB1*, just two pathogenic variants have been identified in a same nucleotide in 5' UTR, disrupting the stop codon of an upstream open reading frame (uORF). uORFs are known to be part of a wide range of post-transcriptional regulation processes and only recently has their association with human diseases come to light. In the present study, we analyzed *EFNB1* in a female patient with typical features of CFNS. We identified a variant, located at c.-411, creating a new upstream ATG (uATG) in the 5'UTR of *EFNB1*, which is predicted to alter an existing uORF. Dual-luciferase reporter assays showed significant reduction in protein translation, but no difference in the mRNA levels. Our study demonstrates, for the first time, the regulatory impact of uATG creation on *EFNB1* levels and suggests that this region should be target in molecular diagnosis of CFNS cases without pathogenic variants in the coding and splice sites regions of *EFNB1*.

**Keywords:** genetic counseling, molecular testing, molecular diagnosis, upstream ORF, upstream ATG, protein translation, regulatory variant

## 1 Introduction

2  
3 Craniofrontonasal syndrome (CFNS; OMIM #304110) is a rare X-linked dominant  
4 disorder caused by loss-of-function mutations in *EFNBI* (Twigg et al., 2004; Wieland et al.,  
5 2004). Greater severity is observed in heterozygous female patients compared with  
6 hemizygous male individuals. Female patients can present with severe hypertelorism, a  
7 central nasal groove, craniofacial asymmetry and coronal craniosynostosis associated with  
8 extracranial features such as sloping shoulders, mild cutaneous syndactyly, grooved nails,  
9 duplication of the first digit, and wiry hair. On the other hand, hemizygous males present  
10 only with hypertelorism and occasional cleft lip (van den Elzen et al., 2014; Twigg et al.,  
11 2004, 2013). However, severe phenotypes have been observed in male patients with mosaic  
12 *EFNBI* mutations (Twigg et al., 2013), a finding that supports the hypothesis that cellular  
13 interference underlies clinical severity (Niethamer et al., 2017; Twigg et al., 2013; Wieacker  
14 and Wieland, 2005).

15 More than a hundred *EFNBI* mutations have been described in association with  
16 CFNS. Almost all pathogenic variants are in the coding region, with the majority (>95%)  
17 leading to premature termination codons (Apostolopoulou et al., 2012; Cui et al., 2012;  
18 Grasso et al., 2011; Hogue et al., 2010; Makarov et al., 2010; Ramirez-Garcia et al., 2013;  
19 Seven et al., 2013; Shotelersuk et al., 2006; Torii et al., 2007; Twigg et al., 2006, 2013, 2004;  
20 Wallis et al., 2008; Wieland et al., 2004, 2005, 2007). Only 2 mutations affecting the same  
21 nucleotide have been reported outside the coding and splice site sequences of *EFNBI*: c.-  
22 95T>G and c.-95T>C. These variants are present in the 5'UTR and abolish the stop codon of  
23 a small uORF, leading to an overlapped and out-of-frame ORF with the main downstream  
24 ORF (dORF, corresponding to *EFNBI*), with reduced translation of *EFNBI* as a consequence  
25 (Twigg et al., 2013).

26 uORFs regulate gene expression through multiple functional mechanisms, for  
27 example in a peptide-dependent manner, by consumption of functional pre-initiation  
28 complex, or triggering nonsense-mediated decay (Barbosa et al., 2013; Ebina et al., 2015;  
29 Hurt et al., 2013; Rahmani et al., 2009). These regulatory sequences are mostly linked to  
30 repression of translation of the main dORF, although under specific conditions, uORFs can

confer induction of protein translation (Wethmar, 2014). uORF-altering variants have been described in association with the etiology of human disorders (Barbosa et al., 2013). Such variants can introduce new uORFs, as found in Gonadal dysgenesis and Van der Woude syndrome (Calvo et al. 2009; Poulat et al. 1998; Kondo et al. 2002), as well as disrupt uORFs in diseases like Marie Unna hereditary hypotrichosis (Wen et al. 2009) and thrombocythemia (Kondo et al. 1998).

In this work we demonstrate the regulatory impact of a variant-created uATG on mRNA translation of *EFNB1*. Identification of such variants contributes to the expanding field of translation regulation and to definition of critical non-coding sequences that can be targeted in genetic diagnosis.

## Case report

The female patient (referred to as F5162-1), first evaluated 10 days after birth, presented with typical features of CFNS: unilateral coronal craniosynostosis (left), hypertelorism, asymmetric craniofacial contour, widows peak, arched eyebrows, hypoplastic supraorbital ridges, telecanthus, downslanting palpebral fissures, short nose with depressed nasal bridge, bifid nasal tip, deep philtrum, tented upper lip, widened metopic suture, and a frontal bone defect (Figure 1A, B). A fronto-orbital advancement surgery was performed at the age of 5 months. No clinical manifestations were observed in her parents and there is no additional family history.

## Materials and Methods

### *DNA samples*

Genomic DNA from the patient and her clinically unaffected parents was extracted from peripheral blood using Gentra Systems Autopure LS (AutoGen), according to the manufacturer's instructions.

### *DNA sequencing*

Standard genetic screening of the coding regions of *EFNB1* was carried out with the patient's DNA using Sanger sequencing. Primers were designed with Primer3 (Untergasser et al., 2012) (primer sequences available upon request). Sequencing reactions were performed using the BigDye Terminator v3.1 Cycle Sequencing Kit and the amplicons were sequenced using an ABI 3730 DNA Analyzer (Thermo Fisher Scientific). Data were analyzed using Sequencher v.5.1 software.

A more comprehensive analysis of the patient's DNA was performed by targeted next generation sequencing (NGS). Specific probes were used to capture 5' and 3' UTR sequences and all coding regions of *EFNB1* (RefSeq NM\_004429), as well as splice acceptor and donor sites (Nextera – Custom Target Enrichment, Illumina), using MiSeq (Illumina), according to the manufacturer's protocols. Data were aligned to the human reference sequence (hg19) with the Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009) and PCR duplicates were removed with the Picard toolkit. Variants were called by the Unified Genotyper tool from Genome Analysis Toolkit (McKenna et al., 2010) and annotated with ANNOVAR (Wang et al., 2010). Sanger sequencing was used to analyse the candidate variant in parental DNA (primer sequences available upon request).

#### *Variant analysis*

Variants present in the 1000 Genomes Database, dbSNP150, genome Aggregation Database (gnomAD), or in the Online Archive of Brazilian Mutations (ABraOM) (Naslavsky et al., 2017) were excluded from downstream analysis as unlikely to be pathogenic. Evolutionary sequence conservation was assessed using the Vertebrate Multiz Alignment & Conservation track (100 vertebrate species) and Genomic Evolutionary Rate Profiling (GERP) (35 species of mammals), both through the UCSC Genome Browser. For multiple vertebrate sequence alignment, twenty *EFNB1* orthologues were extracted from the Ensembl genome browser along with the human gene, and comprised the 5' UTR when this was annotated or ~700 bp of upstream sequence from the main ATG start codon, and the first coding exon. The sequences were aligned using MAFFT (Katoh et al., 2017) through the AliView v.1.18.1 program (Larsson, 2014); consensus sequence for the alignment was obtained using Jalview v.2.10.2b2 (Waterhouse et al., 2009). Open Reading Frame Finder

(ORFfinder, NCBI) was used to search for uORFs in the sequences. NetStart 1.0 (<http://www.cbs.dtu.dk/services/NetStart/>) was used to predict and score translation initiation sites.

#### *Microsatellite analysis*

Microsatellite genotyping was conducted using fluorescently labeled PCR primer pairs from the Linkage Mapping Set v2.5, according to the manufacturer's protocol. PCR fragments were read in an ABI 3730 DNA Analyzer (Thermo Fisher Scientific) and analyzed with GeneMapper v5.0.

#### *Plasmids*

Site-directed mutagenesis was carried out to introduce specific nucleotide substitutions within the 5'UTR of *EFNB1*, using the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies), following the manufacturer's instructions. The wild type (WT) 5'UTR of *EFNB1* was inserted into a previously modified psiCHECK-2 vector (Calvo et al., 2009; referred to as WT construct), in which the ATG of *Renilla* was modified to TTG, so that *Renilla* expression was driven from the primary *EFNB1* ATG codon. We used this vector as template to generate a construct containing the c.-411C>G variant (referred to as F5162 construct) and a second construct containing c.-227G>T and c.-225G>A in addition to c.-411C>G (referred to as F5162<sup>SHORT</sup>; further explanation below). Sanger sequencing was used to confirm the incorporated modifications. Primer sequences are available upon request.

For comparative analysis in the luciferase assays, a previously described construct used to investigate the c.-95T>G variant (Twigg et al., 2013) was also used (referred to as 1330).

#### *Luciferase Assay*

Human embryonic kidney 293T (HEK293T) cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 100 U/mL penicillin and 100 µg/mL streptomycin (all provided by Life Technologies). Cells were

seeded at  $6 \times 10^3$  cells/well in 96-well optical-bottom plates (Nunc<sup>TM</sup>, Thermo Scientific) 24 h prior to transfection. Transient transfections were performed using 100 ng of the vector constructs, in triplicate, with TurboFectin 8.0 (OriGene) according to the manufacturer's instructions. Forty-eight hours after DNA transfection, Renilla activity was measured with the Dual-Glo Luciferase Assay System (Promega), in a GloMax-Multi Detection System (Promega). All constructs also contained the Firefly luciferase gene, used as a transfection control. Renilla luminescence results were normalized by Firefly luciferase. Statistical analysis was performed by One-way ANOVA with a post-hoc Bonferroni multiple comparison test using the GraphPad Prism 6.0. Data were presented relative to the wild type construct, as mean  $\pm$  SEM.

#### *Real-time quantitative PCR (RT-qPCR)*

HEK293T cells were seeded at  $10^5$  cells per well in 6-well culture plates 24 hours prior to transfection. Cells were transfected with 1  $\mu$ g of vector construct, in duplicate, with TurboFectin 8.0. Forty-eight hours after transfection, cells were washed with PBS and total RNA was obtained using the NucleoSpin RNA II extraction kit (Macherey-Nagel), following the manufacturer's recommendations. cDNA synthesis was performed with the SuperScript IV First-Strand Synthesis System (Life Technologies) and oligo-dT primers, using 1.5  $\mu$ g of total RNA from each cell sample as starting material. RT-qPCR reactions were performed with 2X Fast SYBR Green PCR Master Mix (Applied Biosystem) and 200 nM of each primer. Primer pairs for Renilla luciferase 5'-TCCATGCTGAGAGTGTCGTG-3' (forward) and 5'-CAAGCACCATTTTCTCGCCC-3' (reverse) and Firefly luciferase 5'-TCTGGCGACATTGCCTACTG-3' (forward) and 5'-CGGCGTCGAAAATGTTAGGG-3' (reverse) were designed with the Primer-BLAST tool (NCBI) (Ye et al., 2012) and their amplification efficiencies (E) were determined by serial cDNA dilutions; primers were supplied by Exxtend. Fluorescence was detected using the QuantStudio 5 System (Applied Biosystems), under standard temperature protocol. The relative mRNA expression values were determined by dividing  $E^{-\Delta C_t}$  of the target gene (Renilla luciferase) by  $E^{-\Delta C_t}$  of the endogenous control (Firefly luciferase) (Pfaffl, 2001). Statistical significance was assessed by One-way ANOVA with a post-hoc Bonferroni multiple comparison test using the

GraphPad Prism 6.0. Data were presented relative to the wild type construct, as mean  $\pm$  SEM.

## Results

### *Detection of a uAUG codon-creating EFNBI variant*

As there was a clinical diagnosis of CFNS the patient was referred to our center for analysis of *EFNBI*. Sanger sequencing analysis of coding regions did not identify potential causative variants in *EFNBI*. Next, the NGS approach confirmed no potential pathogenic variants in coding regions or in splice sites. However, we identified a variant located within the 5'UTR of *EFNBI* (NM\_004429.4): c.-411C>G. This variant (chrX:68049209 C>G) was not inherited from the patient's parents (Figure 1C) and analysis of microsatellite markers was consistent with the correct biological relationships (correct segregation of 15/15 markers). The identified variant was novel and absent from 1000 Genomes, dbSNP150, ABraOM, and gnomAD.

The wild type 5'UTR sequence of *EFNBI* contains two uORFs on the positive strand and downstream of the c.-411C>G variant: the first is the functional four codon uORF (uORF1), studied by Twigg et al (2013); the second is a ten codon uORF (uORF2) that was predicted by ORFfinder *in silico* analysis. Substitution of C by G at position c.-411 was predicted to introduce a uAUG start codon, which is in-frame with the uUGA stop codon of uORF2, establishing the creation of a new seventy-six codon uORF (hereafter referred to as uORF3) overlapping with uORF2 and sharing the same uUGA stop codon. All uORFs1-3 are out-of-frame with the AUG of the main dORF (Figure 3A).

Additionally, we investigated whether reported variants in gnomAD would alter, create, and/or delete some uORFs in the 5'UTR of *EFNBI*. Forty-seven point variants (allele frequency < 0.09) were found. None of them were predicted to modify uORFs in the sequence, according to pre-defined parameters of the ORFfinder.

As the sequence around uAUGs has been associated with the likelihood of translation initiation, we evaluated the AUG context in uORFs1-3 and the main dORF. In the Kozak consensus the most critical bases are a purine (A or G) at position -3 and a G at position +4 (the A of AUG, is designated as +1; Kozak, 1987, 1989, 2002). The uORF3 sequence has a



match with the Kozak sequence at purine -3. The uORF2 sequence context lacks a purine at position -3 and the G base at position +4, while uORF1 and the main dORF does have a match at position -3 and +4 of the Kozak sequence. In support, the ATGs of uORF1-3 and the main dORF were scored as probable translation sites (NetStart score > 0.5) (Figure 2).

The c.-411C site was found to be evolutionary conserved (GERP score = 2.48) and no uATG was identified in any species at this position. Analysis of *EFNB1* orthologues in twenty vertebrates revealed a high degree of identity, most strongly among mammalian species. No uORFs were observed near the human c.-411C, with the exception of 24 and 16 codon uORFs found in zebra finch and zebrafish, respectively. The region comprising uORF2 shows a high degree of sequence homology in several species of mammals and, just a few bases downstream the start of human uORF2, zebra finch and zebrafish also have a uORF (Figure 2). The multi-species alignment comprising uORF1-3 and the main ORF shows more than one uORF in all species (Figure 2).

Collectively, these observations suggest that the introduction of a uATG, by the c.-411C>G variant, could potentially interfere with *EFNB1* translation.

#### *c.-411C>G reduces translation of the main EFNB1 ORF*

We performed dual luciferase reporter assays to functionally evaluate the regulatory importance of the 5' UTR and to test the effect of c.-411C>G on protein expression from the main dORF, which could explain the observed CFNS phenotype. In the presence of c.-411C>G translation of the main dORF was reduced by 62%, when compared to wild type (F5162 construct compared to WT;  $P < 0.001$ ). Next, we investigated if the variant-created uORF3 would have an independent effect from uORF2 in the translation of the main uORF. To test this hypothesis, we built a construct with a premature stop codon, located 13 codons upstream of the UGA of uORF2 (F5162<sup>SHORT</sup>), thus containing 3 separate uORFS. We found a slightly stronger reduction in the main dORF translation in F5162<sup>SHORT</sup> compared to F5162 (the main dORF translation was about 11% less expressed when compared to F5162) (Figure 3A-B).

To determine if the c.-411C>G variant in our patient has similar regulatory effects to c.-95T>G, previously described in CFNS by Twigg et al. (2013), we compared luciferase

activity using the c.-95G construct. Both c.-411C>G (F5162) and c.-95T>G (1330) variants significantly reduced the main dORF translation on average 63% compared to WT.

As an RNA sample was not available from the patient, to evaluate if the reduced protein levels could be a consequence of impaired *EFNB1* transcription, we transfected HEK293T cells with each construct (F5162, F5162<sup>SHORT</sup>, 1330, and WT) and analysed the Renilla luciferase RNA expression. No significant difference was observed between any of the constructs relative to WT (Figure 3C), suggesting that the effect of the c.-411C>G variant is most likely at the level of translation.

## Discussion

There are an increasing number of examples of diseases in which the underlying mechanisms are caused by variants at 5'UTR (Barbosa et al., 2013; Calvo et al., 2009). However, pathogenicity interpretation of such variants is usually difficult and detailed functional analyses are required to provide evidence of causality. Here, we add to this still scarce list of pathogenic 5'UTR variants an additional, novel, and *de novo* c.-411C>G variant in *EFNB1*, which is causative of craniofrontonasal syndrome.

The c.-411C>G change is a *de novo* variant located in a highly evolutionary conserved region, providing evidence that it might be functional. *In silico* analysis predicted the creation of a new uAUG, leading to a new uORF (uORF3) or representing a 5'elongation of a presumed uORF2. Favoring this hypothesis, we observed that this new uAUG of uORF3 has a good Kozak consensus sequence that scored as a probable translation site. The potential functional effect of this variant was also supported by the observation that none of the 47 single nucleotide variants (SNVs), annotated at the 5'UTR in gnomAD, seemed to disrupt, create, or alter uORFs. Therefore, all the *in silico* analysis supported a functional effect for the c.-411C>G variant, identified in a region usually not sequenced in CFNS patients.

Addressing this functional hypothesis, through *in vitro* assays of sequences containing uORF3, our findings suggest that this variant leads to translation dysregulation. Both the F5162 and F5162<sup>SHORT</sup> assays provided evidence that the c.-411C>G variant decreases protein translation of the main dORF. The F5162<sup>SHORT</sup> analysis showed a slightly

greater reduction in protein expression of the main dORF (compared to F5162 construct), in agreement with the literature for additional number of uORFs (Calvo et al., 2009; Chew et al., 2016; Iacono et al., 2005; Johnstone et al., 2016). Furthermore, the c.-411C>G uORF3-creating variant and c.-95T>G uORF1-disrupting variant (Twigg et al., 2013), showed highly similar effects on the protein levels of the main dORF, despite their different positions and sizes. Taken together, these results strongly suggest that the c.-411C>G variant has a functional effect on translation. Although one of the mechanisms thought to be triggered by uORFs is the nonsense mediated decay (Barbosa et al., 2013), this is not supported by our *in vitro* mRNA analysis.

To date, at least twenty-four diseases have been described to be caused by uORF-altering variants in the 5'UTR of different genes. It is interesting to note that the c.-411C>G variant is one of the most distantly located uORF variant compared to other pathogenic uORF-altering mutations in different disorders (Barbosa et al., 2013; Hornig et al., 2016). Therefore, in CFNS patients negative for pathogenic variants in coding and splice sites of *EFNB1*, a careful analysis of the entire 5' UTR should be carried out as causative variants in the entire region could result in decreased protein levels and disease.

About 49% of human transcripts contain uORFs that are most likely to be physiologically relevant, with their functional importance supported by conservation in vertebrate evolution. The mean length of uORFs across vertebrates was found to be <60 nucleotides (Johnstone et al 2016). The novel 228 nucleotide uORF3 described in this work is much longer than the majority of uORFs present within the *EFNB1* 5'UTR of the evaluated species, except for one predicted uORF found in the mammal hyrax (>680 bp) (data not shown in the alignment), in agreement with depletion of long ORFs in vertebrate 5' UTRs (Johnstone et al 2016).

This study contributes an additional CFNS case with a regulatory variant in the *EFNB1* 5' UTR sequence, and supports screening of this noncoding region in CFNS patients where other variants or gene copy number changes are absent. The 5' UTRs of disease-related genes are not routinely screened raising the possibility that similar pathogenic variants contribute more widely to disease. In light of this, it is interesting to note that uORFs can be targeted by antisense oligonucleotides (ASOs) to reestablish normal levels of the main

protein (Liang et al., 2016). In this context, a CRISPR-Cas9 approach might be useful to create different mutations in a disease-relevant cell type, like iPSC-derived neuroepithelial cells, to provide further insights into pathophysiology of the disease along with investigation of ASOs strategies. In summary, this report adds to the understanding of the functional effects of uAUG/uORFs on protein expression and pathophysiological mechanisms in CFNS.

## **Acknowledgements**

The authors would like to thank Monica Varela, Suzana Ezquina, Melina Rodrigues, and Francine Favaro for technical support. We are also grateful to Carlos Menck and Diogo Meireles for helpful discussions. M.R.P.B. is funded by CEPID/FAPESP (2013/08028-1) and CNPq (PQ10/2011). S.R.F.T. is funded by Action Medical Research (GN2483).

## **Statement of Ethics**

Ethics approval for this study was provided by the Ethics Committee of Instituto de Biociências at Universidade de São Paulo, São Paulo, SP, Brazil (Protocol 024/2004). Patients donated biological samples after providing signed informed consent.

## **Disclosure Statement**

The authors declare no conflicts of interest.

## References

- Apostolopoulou D, Stratoudakis A, Hatzaki A, Kaxira OS, Panagopoulos KP, et al: A novel de novo mutation within EFNB1 gene in a young girl with craniofrontonasal syndrome. *Cleft Palate Craniofac J* 49:109–113 (2012).
- Barbosa C, Peixeiro I, Romão L: Gene expression regulation by upstream open reading frames and human disease. *PLoS Genet* 9:e1003529 (2013).
- Braverman N, Chen L, Lin P, Obie C, Steel G, et al: Mutation analysis of PEX7 in 60 probands with rhizomelic chondrodysplasia punctata and functional correlations of genotype with phenotype. *Hum Mutat* 20:284–297 (2002).
- Calvo SE, Pagliarini DJ, Mootha VK: Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci U S A* 106:7507–7512 (2009).
- Chew GL, Pauli A, Schier AF: Conservation of uORF repressiveness and sequence features in mouse, human and zebrafish. *Nat Commun* 7:11663 (2016).
- Cui Y, Zhao H, Liu Z, Liu C, Luan J, et al: A systematic review of genetic skeletal disorders reported in Chinese biomedical journals between 1978 and 2012. *Orphanet J Rare Dis* 7:55 (2012).
- Ebina I, Takemoto-Tsutsumi M, Watanabe S, Koyama H, Endo Y, et al: Identification of novel *Arabidopsis thaliana* upstream open reading frames that control expression of the main coding sequences in a peptide sequence-dependent manner. *Nucleic Acids Res* 43:1562–1576 (2015).
- Edgar RC: MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797 (2004).
- van den Elzen MEP, Twigg SRF, Goos JAC, Hoogeboom AJM, van den Ouweland AMW, et al: Phenotypes of craniofrontonasal syndrome in patients with a pathogenic mutation in EFNB1. *Eur J Hum Genet* 22:995–1001 (2014).
- Grasso C, Licata F, Rossi C, Balistreri M, Sorge G: Craniofrontonasal syndrome: genetic aspects and description of a clinical case. *Minerva Pediatr* 63:431–438 (2011).
- Hogue J, Shankar S, Perry H, Patel R, Vargervik K, et al: A novel EFNB1 mutation (c.712delG) in a family with craniofrontonasal syndrome and diaphragmatic hernia. *Am J Med Genet A* 152A:2574–2577 (2010).
- Hornig NC, de Beaufort C, Denzer F, Cools M, Wabitsch M, et al: A Recurrent Germline Mutation in the 5'UTR of the Androgen Receptor Causes Complete Androgen Insensitivity by Activating Aberrant uORF Translation. *Plos One* 11(4), p. e0154158 (2016).
- Huopio H, Jääskeläinen J, Komulainen J, Miettinen R, Kärkkäinen P, et al: Acute insulin response tests for the differential diagnosis of congenital hyperinsulinism. *J Clin Endocrinol Metab* 87:4502–4507 (2002).
- Hurt JA, Robertson AD, Burge CB: Global analyses of UPF1 binding and function reveal expanded scope of nonsense-mediated mRNA decay. *Genome Res* 23:1636–1650 (2013).
- Iacono M, Mignone F, Pesole G: uAUG and uORFs in human and rodent 5' untranslated mRNAs. *Gene* 349:97–105 (2005).
- Johnstone TG, Bazzini AA, Giraldez AJ: Upstream ORFs are prevalent translational repressors in vertebrates. *EMBO J* 35:706–723 (2016).
- Kondo T, Okabe M, Sanada M, Kurosawa M, Suzuki S, et al: Familial essential thrombocythemia associated

336 with one-base deletion in the 5'-untranslated region of the thrombopoietin gene. *Blood* 92(4):1091–1096  
337 (1998).

338 Kondo S, Schutte BC, Richardson RJ, Bjork BC, Knight AS, et al: Mutations in *IRF6* cause Van der Woude  
339 and popliteal pterygium syndromes. *Nat Genet* 32:285–289 (2002).

340 Kozak M: An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res*  
341 15:8125–8148 (1987).

342 Kozak M: The scanning model for translation: an update. *J Cell Biol* 108:229–241 (1989).

343 Kozak M: Pushing the limits of the scanning mechanism for initiation of translation. *Gene* 299:1–34 (2002).

344 Krude H, Biebermann H, Luck W, Horn R, Brabant G, et al: Severe early-onset obesity, adrenal insufficiency  
345 and red hair pigmentation caused by *POMC* mutations in humans. *Nat Genet* 19:155–157 (1998).

346 Larsson A: AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*  
347 30:3276–3278 (2014).

348 Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*  
349 25:1754–1760 (2009).

350 Liang XH, Shen W, Sun H, Migawa MT, Vickers TA, et al: Translation efficiency of mRNAs is increased by  
351 antisense oligonucleotides targeting upstream open reading frames. *Nat Biotechnol* 34:875–880 (2016).

352 Makarov R, Steiner B, Gucev Z, Tasic V, Wieacker P, et al: The impact of *CFNS*-causing *EFNB1* mutations  
353 on ephrin-B1 function. *BMC Med Genet* 11:98 (2010).

354 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, et al: The Genome Analysis Toolkit: a  
355 MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303  
356 (2010).

357 Naslavsky MS, Yamamoto GL, de Almeida TF, Ezquina SAM, Sunaga DY, et al: Exomic variants of an elderly  
358 cohort of Brazilians in the ABraOM database. *Hum Mutat* 38:751–763 (2017).

359 Niethamer TK, Larson AR, O'Neill AK, Bershteyn M, Hsiao EC, et al: *EPHRIN-B1* Mosaicism Drives Cell  
360 Segregation in Craniofrontonasal Syndrome hiPSC-Derived Neuroepithelial Cells. *Stem Cell Reports* 8:529–  
361 537 (2017).

362 Pfaffl MW: A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*  
363 29:e45 (2001).

364 Poulat F, Desclozeaux M, Tuffery S, Jay P, Boizet B, et al: Mutation in the 5' noncoding region of the *SRY*  
365 gene in an XY sex-reversed patient. *Hum Mutat Suppl* 1:S192-4 (1998).

366 Rahmani F, Hummel M, Schuurmans J, Wiese-Klinkenberg A, Smeekens S, et al: Sucrose control of translation  
367 mediated by an upstream open reading frame-encoded peptide. *Plant Physiol* 1356–1367 (2009).

368 Ramirez-Garcia MA, Chacon-Camacho OF, Leyva-Hernandez C, Cardenas-Conejo A, Zenteno JC: A Novel  
369 De Novo *EFNB1* Gene Mutation in a Mexican Patient with Craniofrontonasal Syndrome. *Case Rep Genet*  
370 2013:349725 (2013).

371 Rideau A, Mangeat B, Matthes T, Trono D, Beris P: Molecular mechanism of hepcidin deficiency in a patient  
372 with juvenile hemochromatosis. *Haematologica* 92:127–128 (2007).

373 Seven M, Gezdirici A, Ulucan H, Karatas OF, Yosunkaya E, et al: A novel *EFNB1* mutation in a patient with

374 craniofrontonasal syndrome and right hallux duplication. *Gene* 527:675–678 (2013).

375 Shotelersuk V, Siriwan P, Ausavarat S: A novel mutation in EFNB1, probably with a dominant negative effect,  
376 underlying craniofrontonasal syndrome. *Cleft Palate Craniofac J* 43:152–154 (2006).

377 Tassin J, Dürr A, Bonnet AM, Gil R, Vidailhet M, et al: Levodopa-responsive dystonia. GTP cyclohydrolase I  
378 or parkin mutations? *Brain* 123(Pt 6):1112–1121 (2000).

379 Torii C, Izumi K, Nakajima H, Takahashi T, Kosaki K: EFNB1 mutation at the ephrin ligand-receptor  
380 dimerization interface in a patient with craniofrontonasal syndrome. *Congenit Anom (Kyoto)* 47:49–52 (2007).

381 Twigg SRF, Kan R, Babbs C, Bochukova EG, Robertson SP, et al: Mutations of ephrin-B1 (EFNB1), a marker  
382 of tissue boundary formation, cause craniofrontonasal syndrome. *Proc Natl Acad Sci U S A* 101:8652–8657  
383 (2004).

384 Twigg SRF, Matsumoto K, Kidd AMJ, Goriely A, Taylor IB, et al: The origin of EFNB1 mutations in  
385 craniofrontonasal syndrome: frequent somatic mosaicism and explanation of the paucity of carrier males. *Am*  
386 *J Hum Genet* 78:999–1010 (2006).

387 Twigg SRF, Babbs C, van den Elzen MEP, Goriely A, Taylor S, et al: Cellular interference in craniofrontonasal  
388 syndrome: males mosaic for mutations in the X-linked EFNB1 gene are more severely affected than true  
389 hemizygotes. *Hum Mol Genet* 22:1654–1662 (2013).

390 Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, et al: Primer3--new capabilities and interfaces.  
391 *Nucleic Acids Res* 40:e115 (2012).

392 Wallis D, Lacbawan F, Jain M, Der Kaloustian VM, Steiner CE, et al: Additional EFNB1 mutations in  
393 craniofrontonasal syndrome. *Am J Med Genet A* 146A:2008–2012 (2008).

394 Wang K, Li M, Hakonarson H: ANNOVAR: functional annotation of genetic variants from high-throughput  
395 sequencing data. *Nucleic Acids Res* 38:e164 (2010).

396 Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ: Jalview Version 2--a multiple sequence  
397 alignment editor and analysis workbench. *Bioinformatics* 25:1189–1191 (2009).

398 Wen Y, Liu Y, Xu Y, Zhao Y, Hua R, Wang K, et al: Loss-of-function mutations of an inhibitory upstream  
399 ORF in the human hairless transcript cause Marie Unna hereditary hypotrichosis. *Nature Genetics* 41(2):228–  
400 233 (2009).

401 Wethmar K: The regulatory potential of upstream open reading frames in eukaryotic gene expression. *Wiley*  
402 *Interdiscip Rev RNA* 5:765–778 (2014).

403 Wieacker P, Wieland I: Clinical and genetic aspects of craniofrontonasal syndrome: towards resolving a genetic  
404 paradox. *Mol Genet Metab* 86:110–116 (2005).

405 Wieland I, Jakubiczka S, Muschke P, Cohen M, Thiele H, et al: Mutations of the ephrin-B1 gene cause  
406 craniofrontonasal syndrome. *Am J Hum Genet* 74:1209–1215 (2004).

407 Wieland I, Reardon W, Jakubiczka S, Franco B, Kress W, et al: Twenty-six novel EFNB1 mutations in familial  
408 and sporadic craniofrontonasal syndrome (CFNS). *Hum Mutat* 26:113–118 (2005).

409 Wieland I, Weidner C, Ciccone R, Lapi E, McDonald-McGinn D, Kress W, et al: Contiguous gene deletions  
410 involving EFNB1, OPHN1, PJA1 and EDA in patients with craniofrontonasal syndrome. *Clin Genet* 72:506–  
411 516 (2007).

412 Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, et al: Primer-BLAST: a tool to design target-specific

413 primers for polymerase chain reaction. BMC Bioinformatics 13:134 (2012).



## Legends to figures

**Fig. 1.** Frontal view of the patient showing facial asymmetry, hypertelorism, telecanthus, down slanting palpebral fissures, short nose with bifid nasal tip, and tented mouth. (B) Cranial computed tomography scan showing unilateral coronal craniosynostosis (arrowhead). (C) Dideoxy sequencing of the 5'UTR of *EFNB1* in the patient and her parents. A *de novo* c.-411C>G (chrX:g.68049209C>G) variant was detected (arrow).

**Fig. 2.** Multi-species alignment of *EFNB1*. MAFFT (Katoh et al., 2017) was used, through AliView (Larsson, 2014), to perform the alignment of human *EFNB1* against 19 vertebrate orthologue sequences, all of them taken from Ensembl genome browser. uORFs were predicted by ORFfinder with a minimum of 30 nucleotides or identified manually. Indicated in the figure are just uORFs relevant to our findings. The variant-created uORF3 region is shown along with uORF2 (*in silico* predicted for human sequence, through ORFfinder). The nucleotide affected by the c.-411C>G change is indicated above the sequence by a red triangle. In the uORF3 region no uORFs are found in the majority of vertebrates, but uORFs of 24 and 16 codons are found in zebra finch and zebrafish, respectively. The first base of uORF2 is indicated above the sequence with a black triangle. uORF2 is well conserved and present in most of the mammalian species. Different uORFs in this region were found in zebra finch, zebrafish, and chicken. The main ORF is the most highly conserved, followed by uORF1. Below the consensus sequence of the multi-species alignment, the sequence surrounding the uORF ATGs (black bars) is expanded to compare the human uORFs with the Kozak consensus (boxed); underlined bases indicate start codons (AUG, at +1 to +3) and the most important bases for efficient translation; the NetStart score is indicated (*Scores* >0.5 indicate a probable *translation* start site). Note that for uORF3, this Kozak comparison is just valid when considering c.-411C>G variant (indicated with an asterisk inside the box). The vertical dashed black arrow indicates hidden columns (63 human nucleotides) to better fit the figure. The red numbers below the alignment indicate the number of bases hidden in the relevant sequence (red bar) at those positions relative to gaps in the human sequence. uORFs are boxed and colored according to nucleotide base.

**Fig. 3.** Functional analysis of the uORF1-creating variant. **(A)** Constructs generated to investigate the effect of uORFs on translation from the main dORF. WT: wild type 5'UTR sequence of *EFNB1* containing the *in silico* predicted uORF2 and the functional validated uORF1 (Twigg et al. 2013). F5162: The mutant allele found in the patient containing the new c.411C>G uAUG codon; note the new uORF3 shares the same stop codon with uORF2 (blue bars). F5162<sup>SHORT</sup>: c.411C>G variant along with an earlier stop codon in uORF3 which separates it from uORF2. 1330: c.-95T>G construct, found in a mosaic CFNS male (Twigg et al. 2013); this variant disrupts the stop codon of uORF1 that is out-of-frame with and overlaps the main dORF. Asterisks indicate nucleotides changes and blue bars represent uORFs. The AUG sequence context is shown for each uORF/ORF up to base -3 and +4. **(B)** Graph depicting the relative Renilla luciferase activity [(experimental sample ratio) – (negative control ratio) / (positive control ratio) – (negative control ratio)] for each construct. **(C)** RT-qPCR analysis of relative Renilla luciferase mRNA levels in transfected HEK293T cells. All values are relative to wild type as mean ± SEM of three experiments, which were performed in triplicate. One-way ANOVA with a post-hoc Bonferroni multiple comparison test; \*\*\* P value <0.001; \*\* P value <0.01; \* P value <0.05.