Supplementary Material

*Congenital insensitivity to pain: a novel mutation affecting a U12-type intron causes multiple aberrant splicing of SCN9A*

Margherita Marchi1,*, Ilaria D’Amato 1,*, Mirna Anđelić 1,2, Daniele Cartelli 1, Erika Salvi 1, Raffaella Lombardi 1, Evren Gumüs 3, Giuseppe Lauria 1,4.

1. Neuroalgology Unit, Fondazione IRCCS Istituto Neurologico "Carlo Besta", Milan, Italy.
2. Department of Neurology, School of Mental Health and Neuroscience, Maastricht University Medical Center, Maastricht, The Netherlands
3. Department of Medical Genetics, Faculty of Medicine, University of Harran, Sanliurfa, Turkey
4. Department of Biomedical and Clinical Sciences "Luigi Sacco", University of Milan, Milan, Italy

*These authors contributed equally to this work.

**Correspondence**

Margherita Marchi, 3rd Neurology Unit – Neuroalgology and Pain Genetics Lab. IRCCS Foundation "Carlo Besta" Neurological Institute. Via Celoria 11, 20133, Milan, Italy. Email: margherita.marchi@istituto-besta.it.
Figure S1. Investigation of residual amount of wild-type transcript in the proband. WT-selective primers, precluding the amplification of the two aberrant transcripts, have been designed to avoid quantitative bias in PCR amplification. (a) Agarose gel shows a very weak signal corresponding to the WT allele in the proband and in the mother, whereas it is well represented in the control samples. P: Proband; M: Mother; CTRL: healthy controls; NTC: No Template Control. (b) Sanger sequencing of the proband wild-type allele, cropped-out from agarose gel, confirmed the expected nucleotide sequence.
Table S1. Quantitative RT-PCR data on SCN9A transcript from skin and from blood. HCs: healthy controls; GAPDH, housekeeping gene; NTC: No Template Control.

| Sample | Target | Average Ct |
|--------|--------|------------|
|        |        | Skin       |
|        |        |            |
| HC-1   | GAPDH  | 28.92      |
| HC-1   | SCN9A  | 36.09      |
| HC-2   | GAPDH  | 28.3       |
| HC-2   | SCN9A  | 36.83      |
| mother | GAPDH  | 28.15      |
| mother | SCN9A  | 38.66      |
| proband| GAPDH  | 28.03      |
| proband| SCN9A  | 39.61      |
| NTC    | GAPDH  | Undetermined |
|        | SCN9A  | Undetermined |

| Sample | Target | Average Ct |
|--------|--------|------------|
|        |        | Blood      |
|        |        |            |
| HC-1   | GAPDH  | 21.25      |
| HC-1   | SCN9A  | 30.6       |
| HC-2   | GAPDH  | 25.52      |
| HC-2   | SCN9A  | 33.39      |
| HC-3   | GAPDH  | 25.26      |
| HC-3   | SCN9A  | 31.94      |
| HC-4   | GAPDH  | 25.21      |
| HC-4   | SCN9A  | 33.53      |
| Proband| GAPDH  | 23.17      |
| Proband| SCN9A  | 35.83      |
| NTC    | GAPDH  | Undetermined |
| NTC    | SCN9A  | Undetermined |
Supplementary Methods

Splice sites predictions

Comparison of Splice sites prediction between WT allele and the mutated c.377+7T>G with NetGene2 World Wide Web Server considering the genomic region from exon 2 to 4. The tool recognizes positions 310 and 5299 as the canonical DONORS for intron 2 and 4 respectively, but fails to identify the natural splice DONOR for intron 3 and its ACCEPTOR. The tool identifies as acceptor only the consensus of SCN9A isoform ENST00000303354.6 in position 5204.

The tool does not highlight any difference between the WT and the Mutated sequence; only the molecular analysis revealed that the consensus in position 4660 represents the strongest DONOR, resulting in a cryptic intronic donor site. The best ACCEPTOR consensus for this donor is recognized by the tool in position 4733 in both the WT and the Mutated sequence, but transcript molecular analysis revealed that, in presence of c.377+7T>G substitution, the stronger ACCEPTOR is position 4539.

NetGene2 World Wide Web Server:

| Donor splice sites, direct strand, WT allele |
|---------------------------------------------|
| pos 5'->3' | phase | strand | confidence | 5' exon intron 3' |
| 310 | 0 | + | 1.00 | AGACAAAAAG^GTGAGTTTAT |
| 333 | 2 | + | 0.31 | GACTTCAGTG^GTCAGTTTCT |
| 602 | 0 | + | 0.44 | CATTTCGGAC^GTTAGTTACA |
| 1972 | 2 | + | 0.32 | TTGAGTCAG^GTAGCTGTGAT |
| 2268 | 1 | + | 0.34 | CACATCCCTT^GTAAGTTGGA |
| 2559 | 2 | + | 0.45 | CAAGTATAAG^GTAAGTTGAT |
| 3008 | 2 | + | 0.34 | GGAGCTACAG^GTTAGTTACA |
| 3818 | 1 | + | 0.64 | CCCATACAG^GTAAGTTGGA |
| 4120 | 0 | + | 0.34 | CCTGTTACAG^GTAAGTTGGA |
| 4124 | 1 | + | 0.54 | ACCGCTTGTTA^GTAAGTTGGA |
| 4660 | 1 | + | 0.71 | ATGTGAAAG^GTAAGTTGGA |
| 5299 | 0 | + | 0.96 | AAAATGTCGA^GTAAGTTGGA |
| 5303 | 1 | + | 0.70 | TGTGTTACAG^GTAAGTTGGA |
| 5307 | 2 | + | 0.24 | GGAAGTTCAG^GTAAGTTGGA |
| 6017 | 2 | + | 0.67 | TGTGTTACAG^GTAAGTTGGA |

| Acceptor splice sites, direct strand, WT allele |
|------------------------------------------------|
| pos 5'->3' | phase | strand | confidence | 5' intron exon 3' |
| 105 | 0 | + | 0.43 | CACAAAAACAG^TCTCTTGCCC |
| 717 | 2 | + | 0.33 | CCCACCAACAG^TCCCCGTAAG |
| 1199 | 0 | + | 0.16 | TCTCTCTCAG^CACCCTGTGT |
| 1602 | 2 | + | 0.27 | GGTTGGTCTAG^ACATGAAGTC |
| 1830 | 2 | + | 0.56 | TTGTTTTCTAG^TTTTGTCAAA |
| 2008 | 2 | + | 0.43 | TGTGTTACAG^GTAAGTTGGA |
| 2071 | 2 | + | 0.43 | TTGTTTTCTAG^TTTGGACAG |
| 2081 | 1 | + | 0.07 | TTGTTTTCTAG^TTTTGTCAAA |
| 2767 | 2 | + | 0.53 | ATGTGAAAG^GTAAGTTGGA |
| 3462 | 0 | + | 0.27 | TTGTTTTCTAG^TTTGGACAG |
| 3682 | 1 | + | 0.16 | TTGTTTTCTAG^TTTTGTCAAA |
| 4516 | 0 | + | 0.17 | TTAACAAATAG^CTCAATTTT |
| 4530 | 2 | + | 0.19 | ATTTTTAAAG^TTTACTATGAA |
| pos  | phase | strand | confidence | 5' exon | intron | 3' |
|------|-------|--------|------------|---------|--------|----|
| 310  | 0     | +      | 1.00       | AGACAAAAG^GTGAGTTTAT | H     |
| 333  | 2     | +      | 0.31       | GACCTCAGTG^GTCAGTTTCT |
| 602  | 0     | +      | 0.44       | CATTTGTCAG^GTAGTTTACA |
| 1972 | 2     | +      | 0.32       | TTGAAGTCTAG^GTACGGTGAT |
| 2268 | 1     | +      | 0.34       | CACATCCCTTTG^GAAGTTGGA |
| 2559 | 2     | +      | 0.45       | CAAGTATAAG^GTAAGTTGAT |
| 3008 | 2     | +      | 0.34       | GGACTACAG^GTACCGCCA |
| 3818 | 0     | +      | 0.64       | CTGAGATTAG^GTGAAGAAAG |
| 4120 | 0     | +      | 0.34       | AATTACCCATAG^GTAAGTTGAG |
| 4124 | 1     | +      | 0.54       | ACCGTTGTTA^GTAAGTTGAG |
| 4660 | 1     | +      | 0.71       | ATGTTGAAAG^GTAAGTTGCT |

**CUTOFF values used for confidence:**

- Highly confident donor sites (H): 95.0 %
- Nearly all true donor sites: 50.0 %

**Acceptor splice sites, direct strand, c.377+7T>G allele**

| pos  | phase | strand | confidence | 5' intron | exon | 3' |
|------|-------|--------|------------|-----------|------|----|
| 105  | 0     | +      | 0.43       | CACAAAACAG^TCTCTTTGCC |
| 717  | 2     | +      | 0.33       | CCCACACAC^TCTCTTAGT |
| 1199 | 0     | +      | 0.16       | TCCTCCTCC^CACCTGTGTT |
| 1602 | 2     | +      | 0.27       | GGTTGTCTTAG^CATGAGGTC |
| 1830 | 2     | +      | 0.56       | TTTTTCTCAG^GTGGTCTCAA |
| 2008 | 2     | +      | 0.43       | TTTGGCTCTAG^GATTGACTTG |
| 2071 | 2     | +      | 0.43       | TTTTTTCCAG^TCTGTGAAAG |
| 2081 | 1     | +      | 0.07       | TTTCTGAGAG^AAAGTCTAG |
| 2767 | 2     | +      | 0.53       | ATTTTTCTAG^GAAGTTCTTT |
| 3462 | 0     | +      | 0.27       | TACCACCCAG^GTCCTCTCCC |
| 3682 | 1     | +      | 0.16       | TTCTATTTAG^AGATGACTG |
| 4516 | 0     | +      | 0.17       | TAACAAATAG^CTCAATTTC |
| 4530 | 2     | +      | 0.19       | ATTTTTAAAG^TTACTATGAA |
| 4541 | 1     | +      | 0.19       | TACTATGAAG^AGTGACTG |
| 4544 | 1     | +      | 0.19       | TAGTTGAAG^TGAGCTTG |
| 4555 | 0     | +      | 0.18       | GGACTTGAG^TCTATGGC |
| 4733 | 0     | +      | 0.95       | CCTCTTGCAG^ACTCTATAG |
| 4813 | 0     | +      | 0.43       | TTCTCTTCAG^TCTCTAAGA |
| 5121 | 1     | +      | 0.49       | TTTTTCTTAG^GAAGTTG |

**CUTOFF values used for confidence:**

- Highly confident acceptor sites (H): 95.0 %
- Nearly all true acceptor sites: 20.0 %
The splicing prediction analysis with other on-line free tools Splice Site Prediction by Neural Network (FruitFly) and Alternative Splice Site Predictor (ASSP), gave the same results.

The genomic sequence of SCN9A (RefSeq NCBI NM_002977; Ensembl ENST00000409672.1) is reported below, beginning from the START codon in exon2, showing the exonic regions in blue capital letters, and introns in black lower case. Splicing consensus sequences are highlighted in grey. Underlined nucleotides represent the retained intronic region (129 bp from intron 2 and 4bp from intron 3), caused by the mutation c.377+7T>G, in the aberrant transcript, named as SK3_INT2_ENST354 in the main text.

SCN9A gene sequence (Ensembl ENST00000409672.1)
Comparison between the SCN9A transcript isoforms

Alignment, made by Clustal W Multiple Sequence Aligner, of the two transcript isoforms of SCN9A, including the region between exon 2 and exon 5, reveals the presence of 3 supernumerary nucleotides in position c.378-380. ENST00000409672.1 transcript is annotated in NCBI with the Reference Sequence: NM_002977.3 and is classified as reference standard in the RefSeqGene project. Transcript ENST00000303354.6 is not reported in NCBI. The adjacent exons are alternately colored in blue and black letters.

Notably, the splice-junction between exon 3 and exon 4 is 1-bp shifted in the two transcripts.
Wild-type aminoacidic sequence from the first Met (exon 2) till exon 4:
MAMLPPGPQSFVHFTKQSLALIEQRIAERKSKEPKEEKDDDEEAPKPSDLEAGKQLPFIYGDIPPGMVSEPL
EDLDPPYADKKTFIVLNKGTKIFRFNAPALYMLSPFSLRRISIKILVHLSLFSLMICTILTNCIFMTMNNPPD
WTKNV [...]

Predicted consequence on translation for the transcript SK3_INT2_ENST354: Intron 2 partial retention + exon 3 skipping + Exon 4 ENST00000303354.6
ATGGCAATGTTGCCCTCCCAACAGACCTCAGACGCTTTGTCCATTTTCACAAACAGTGCTCTTGCCTCCTCATTTGAACAA
CGCATTGCTGAAAGAAATCAAAAGCAACCCAAAAGAAGAAAGATGATGATGAAGAGCACCCAAAGGCAAGC
AGTGACTTGGAAGCTGGCAACAGCTGGCCTCTCATCTATGGGACATTCCCTCCCGGCAATGGTGTCAGAGCCCTG
GAGGACTTGGAACCTCCTACTATGCAACAAAGGttactatgaaagttgcttttcatgtgcaactgacacatgacgccacatgtg
acacttgtgactgttgatgacactgtgactgtgactgtgactgtgactgtgactgtgactgtgactgtgactgtgactgtgactgtg
Predicted Consequence: p.Lys86fs2Stop

Predicted consequence on translation of the transcript SK3_ENST354: Exon 3 skipping + Exon 4 ENST00000303354.6
ATGGCAATGTTGCCCTCCCAACAGACCTCAGACGCTTTGTCCATTTTCACAAACAGTGCTCTTGCCTCCTCATTTGAACAA
CGCATTGCTGAAAGAAATCAAAAGCAACCCAAAAGAAGAAAGATGATGATGAAGAGCACCCAAAGGCAAGC
AGTGACTTGGAAGCTGGCAACAGCTGGCCTCTCATCTATGGGACATTCCCTCCCGGCAATGGTGTCAGAGCCCTG
GAGGACTTGGAACCTCCTACTATGCAACAAAG|CTACCTTATTCAGCATGCTCATCTATGTGCACATATTCTGCAAAACTGCATATTATGACCATTGAATAACCCACCGGACTGGACCAAAATATGTGCA
Predicted Consequence: p.Lys86fs12Stop

Predicted consequence on translation of the transcript SK3_INT2_ENST354: Intron 2 partial retention + exon 3 skipping + Exon 4 ENST00000303354.6
ATGGCAATGTTGCCCTCCCAACAGACCTCAGACGCTTTGTCCATTTTCACAAACAGTGCTCTTGCCTCCTCATTTGAACAA
CGCATTGCTGAAAGAAATCAAAAGCAACCCAAAAGAAGAAAGATGATGATGAAGAGCACCCAAAGGCAAGC
AGTGACTTGGAAGCTGGCAACAGCTGGCCTCTCATCTATGGGACATTCCCTCCCGGCAATGGTGTCAGAGCCCTG
GAGGACTTGGAACCTCCTACTATGCAACAAAG|CTACCTTATTCAGCATGCTCATCTATGTGCACATATTCTGCAAAACTGCATATTATGACCATTGAATAACCCACCGGACTGGACCAAAATATGTGCA
Predicted Consequence: p.Lys86fs12Stop

Predicted consequence on translation of the transcript SK3_ENST354: Exon 3 skipping + Exon 4 ENST00000303354.6
ATGGCAATGTTGCCCTCCCAACAGACCTCAGACGCTTTGTCCATTTTCACAAACAGTGCTCTTGCCTCCTCATTTGAACAA
CGCATTGCTGAAAGAAATCAAAAGCAACCCAAAAGAAGAAAGATGATGATGAAGAGCACCCAAAGGCAAGC
AGTGACTTGGAAGCTGGCAACAGCTGGCCTCTCATCTATGGGACATTCCCTCCCGGCAATGGTGTCAGAGCCCTG
GAGGACTTGGAACCTCCTACTATGCAACAAAG|CTACCTTATTCAGCATGCTCATCTATGTGCACATATTCTGCAAAACTGCATATTATGACCATTGAATAACCCACCGGACTGGACCAAAATATGTGCA
Predicted Consequence: p.Lys86fs12Stop
**Web resources**

Ensembl: http://grch37.ensembl.org/Homo_sapiens/

NetGene2: http://www.cbs.dtu.dk/services/NetGene2/

FruitFly: https://www.fruitfly.org/seq_tools/splice.html

ASSP: http://wangcomputing.com/assp/

Clustal W: https://www.genome.jp/tools-bin/clustalw

ExPASy: https://web.expasy.org/translate/