

Molecular characterization of pathogenic *OTOA* gene conversions in hearing loss patients

Sacha Laurent¹, Corinne Gehrig¹, Thierry Nospikel¹, Sami S Amr², Andrea Oza², Elissa Murphy², Anne Vannier¹, Frédérique Sloan Béné^{1,3}, Maria Teresa Carminho-Rodrigues¹, Jean-Louis Blouin^{1,3}, Hélène Cao Van⁴, Marc Abramowicz^{1,3}, Ariane Paoloni-Giacobino^{1,3}, and Michel Guipponi^{1,3}

¹Genetic Medecine Unit, Department of Diagnostic, University Hospitals of Geneva

²Laboratory for Molecular Medicine, Mass General Brigham Personalized Medicine, Cambridge, Massachusetts, USA

³Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland

⁴Pediatric Otolaryngology Unit, Department of Otorhinolaryngology, Head and Neck Surgery, University Hospitals of Geneva, Geneva, Switzerland

December 10, 2021

Abstract

Bi-allelic loss-of-function variants of *OTOA* are a well-known cause of moderate-to-severe hearing loss. Whereas non-allelic homologous recombination-mediated deletions of the gene are well known, gene conversions to pseudogene *OTOAP1* have been reported in the literature but never fully described nor their pathogenicity assessed. Here, we report two unrelated patients with moderate hearing-loss, who were compound heterozygotes for a converted allele and a deletion of *OTOA*. The conversions were initially detected through sequencing depths anomalies at the *OTOA* locus after exome sequencing, then confirmed with long range polymerase chain reactions. Both conversions lead to loss-of-function by introducing a premature stop codon in exon 22 (p.Glu787*). Using genomic alignments and long read nanopore sequencing, we found that the two probands carry stretches of converted DNA of widely different lengths (at least 9 kbp and around 900 bp, respectively).

Introduction

Mendelian genetic defects constitute a leading cause of non-syndromic hearing loss, one of the most common sensorineural defects in humans (Nance, 2003). Although dominant and recessive phenotypes linked to *GJB2/GJB6* are a frequent etiology (Yokota et al., 2019), more than 150 genes have been linked with syndromic or non-syndromic deafness (Sloan-Heggen et al., 2016; Azaiez et al., 2018). Consequently, genetic diagnosis has moved from screening of the connexin locus to interrogating simultaneously thousands of exons of the human genome, to reach an overall diagnostic rate of around 40% (Shearer and Smith, 2015; Azaiez et al., 2018).

STRC and *OTOA* bi-allelic loss-of-function are well-known contributors of mild-to-moderate autosomal recessive hearing loss in humans (Vona et al., 2015; Mehta et al., 2016; Zazo Seco et al., 2017). Both *STRC* and *OTOA* reside in repetitive regions of the genome, leading to recurrent deletions or duplications (copy number variants, CNVs) in the general population, through non-allelic homologous recom-

bination (Zhang et al., 2009). These regions also encompass paralogous pseudogenes for both genes: *CKMT1A-STRCP1-CATSPER2P1* lie directly downstream of *CKMT1B-STRC-CASTPER2* on chromosome 15, whereas *OTOAP1* lies approximately 800kbp downstream of *OTOA* on chromosome 16, with 11 interspersed genes or pseudogenes.

Pathogenic and likely pathogenic frameshift, nonsense, missense and splicing variants have been deposited to ClinVar (Landrum et al., 2016) for *OTOA* and *STRC*. Using array Comparative Genomic Hybridization (aCGH), Hoppman et al. (2013) found an allele frequency of 1.09% for the *STRC-CATSPER2* deletion, over a cohort of more than 5000 patients. The gnomAD structural variant (SV) dataset v2.1 (Collins et al., 2020), calling CNVs using read depth from more than 10000 whole genome sequencing data, reports an allele frequency of 0.90% for the deletion of *STRC* specific exons. However, deletions of *OTOA* are apparently much rarer, with estimates ranging from less than 0.1% in gnomAD SV v2.1 to less than 0.2% in the Database of Genomic Variants (MacDonald et al., 2014).

Gene conversion, defined as the replacement of a locus in the genome by a paralogous sequence, is a mechanism known to generate pathogenic alleles (Chen et al., 2007; Casola et al., 2012). For instance, a pathogenic converted allele has been described for the *TMEM231* gene, causing autosomal recessive Joubert and Meckel-Gruber syndromes (Maglic et al., 2016; Watson et al., 2020). Other examples include a benign conversion of 88 to 351 nucleotides of *BRCA1* intron 2 (Tessereau et al., 2015) mediated by the same breakpoint sites leading to a 37kbp long deletion of exon 1 and 2 of *BRCA1*.

In their cohort of 686 patients with hearing-loss, using high throughput sequencing data, Shearer et al. (2014) reported that CNVs in *STRC* were the most common of all CNVs found in deafness-related genes, followed by CNVs in *OTOA*. Interestingly, they highlighted the case of a causative homozygous gene conversion of *STRC*. The authors also mentioned 3 conversions of *OTOA*, which length was estimated between one to two unspecified exons. No pathogenic CNVs nor single nucleotide variants (SNV) were detected in trans in any of the 3 patients, and the pathogenicity of the conversion itself was not assessed. The conversions were detected by comparative analysis of sequencing depth in the paralogous regions of the genes and pseudogenes, but without confirmation with another molecular approach. Moteki et al. (2016) also reported the case of a homozygous gene conversion of *STRC* in a cohort of 40 patients, identified by high throughput sequencing and confirmed by aCGH. More recently, Rajagopalan et al. (2020) identified two single exon gene conversions of *STRC*, initially identified as single exon deletions in exome sequencing CNVs calls, confirmed with long range PCR.

Methods

DNA was extracted from whole blood samples using a Qiagen Symphony instrument (Hilden, Germany). Library preparation and exome capture were performed using the Twist Library Preparation Enzymatic Fragmentation Kit and the Twist Human Core Exome Kit (Twist Bioscience, San Francisco, USA) with the RefSeq spike-in, following the manufacturer protocol. High throughput sequencing was performed on an Illumina (San Diego, USA) NextSeq 500 sequencer.

SNVs were called using an in-house pipeline based on bwa-mem (Li and Durbin, 2009) for alignment against GRCh37 and GATK HaplotypeCaller (Poplin et al., 2018) for genotyping. CNVs were identified with an algorithm based on Conifer (Krumm et al., 2012) andXHMM (Fromer et al., 2012). Briefly, it detects variation of the z-scores of mean normalized sequencing depth values, after filtering low quality and indeterminate mapping, of each exon that are compatible with duplications or deletions. All samples previously sequenced at our facility after exome capture using the same kit are used as reference. Visual inspection of sequencing

alignments and z-scores was performed with a local instance of JBrowse (Buels et al., 2016). A panel of 187 genes involved in hearing loss and ear malformations was used to filter SNV and CNV, as well as automatic frequency and functional filters, before manual curation by three different biologists.

MLPA was performed using the Salsa Probemix for Deafness Infertility Syndrome (DIS P461, MRC Holland, Amsterdam, The Netherlands) which includes probes for the *OTOA* and *STRC/CATSPER2* regions. PCR amplification was carried out with AccuPrime Taq DNA Polymerase (ThermoFisher, Waltham, USA) and LA Taq DNA polymerase (Takara, Kusatsu, Japan) for amplicons longer than 10kbp. Sanger sequencing of purified PCR products was performed on an ABI 3500 instrument (ThermoFisher).

Nanopore sequencing of long amplicons was performed using a MinION Mk1B instrument (Oxford Nanopore Technologies, ONT, Oxford, United-Kingdom) and R.9.4.1 flowcells with MinKNOW high accuracy base calling. Libraries were prepared according to ONT’s instructions using NEBNext FFPE repair kit, NEBNext Ultra II end repair/dA-tailing module (New England Biolabs, NEB, Ipswich, USA) and SQK-LSK109 ligation sequencing kit (ONT). Reads were mapped using minimap2 against chromosome 16 of the GRCh38 reference genome (NC_000016.10) and visualized using the Integrative Genomics Viewer (Robinson et al., 2017).

Results

Blood samples of a 2 years old female child (case 1) with non-syndromic hearing loss and her parents were referred to our facility for molecular investigations, after genetic counseling and informed consent was obtained. Exome sequencing did not reveal any pathogenic or likely pathogenic SNV in our panel of 187 genes. However, the algorithm for CNV detection showed z-scores compatible with bi-allelic *OTOA* variants (Fig. 1A and 1B), consisting of a heterozygous deletion spanning at least 110kb from exon 2 of *METTL9* (NM_016025.5) to exon 22 of *OTOA* (NM_144672.4) and a smaller variant centered around exon 22 (z-score compatible with a homozygous deletion of exon 22). The MLPA analysis confirmed the large 110kb deletion and showed it was inherited from the mother.

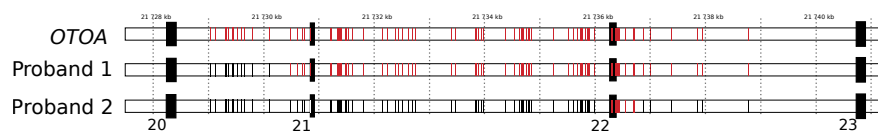
The resolution of the paternal allele was hindered by the fact that the region spanning intron 20 to exon 29 of *OTOA* displays very high sequence identity with exons 1 to 9 of the *OTOAP1* pseudogene (NR_003676.3) and therefore lacks MLPA probes. Upon visual inspection of the reads aligning to exon 22 of *OTOA*, we hypothesized the presence on the paternal chromosome of either a small deletion of exon 22 of *OTOA* or of a conversion of the wild-type sequence of *OTOA* by *OTOAP1*.

A 13kb PCR product specific to the paternally inherited *OTOA* allele was obtained using the proband's DNA and primer #1 and #2; primer #1 being specific to *OTOA* and primer #2 to both *OTOA* and *OTOAP1* (Fig. 1C). This PCR product was then Sanger sequenced using a set of internal primers listed in Supplementary Table 1 (Fig. 1C).

Taken together the sequencing results showed that the paternal *OTOA* allele was not carrying a deletion of exon 22 but underwent pseudogene-mediated gene conversion. This conversion event introduced several sequence variants contributed by *OTOAP1* into *OTOA* including among others a pathogenic premature stop codon (p.Glu787*) (Fig. 1E).

To determine the starting point of the conversion event, we sequenced the 13kbp amplicon using a MinION nanopore sequencer and assessed known differences between *OTOA* and *OTOAP1* reference genomic sequences (Fig. 2). The first *OTOAP1* specific nucleotide was found approximately 350 bp upstream of *OTOA* exon 21. All expected downstream sequence differences corresponded to *OTOAP1* sequences, at the hemizygous state, up to the end of the amplicon at exon 23. This fine mapping analysis allowed us to map the start of the conversion to a 334bp window, and to estimate its length to at least 9kbp. We reported the conversion with the following HGVS nomenclature: NC_000016.10:g.(21730155_21730489)_(21739516_?)con(22546282_22546616)_(22555638_?). Finally, we designed a 3kb PCR product around the mapped conversion start window with primers #1 and #7 (Fig. 1B, Supplementary Data) and confirmed the presence of the converted allele in the proband and her father, the conversion being absent in the mother (Fig. 1D).

Figure 2: Fine mapping of the *OTOA* conversion using nanopore sequencing of the 13kbp amplicon (Primer#1 & #2) in both probands.



Red marks on the *OTOA* locus represent every predicted difference between *OTOA* (NG_012973.2) and *OTOAP1* (NC_000016.10, coordinates 22545698-22576865) when aligning genomic sequences. Black marks indicate positions where the *OTOA* nucleotide was found in the nanopore sequencing mapping, whereas red marks indicate positions where the *OTOAP1* nucleotide was found. This analysis indicates that the conversion was at least 9kbp in proband 1 but less than 1kbp and restricted to exon 22 in proband 2.

Following the characterization of the converted allele in our proband, we identified one ClinVar submission (accession SCV000966854.1) reporting the stop codon mutation contributed by *OTOAP1*, p.Glu787*, in trans with a deletion, identified in a patient with hearing-loss (case 2). After obtaining DNA from this patient, we confirmed the deletion of *OTOA* on one chromosome using MLPA and performed the long range amplification and nanopore sequencing as described above. In contrast with the first proband, only 11 positions around exon 22 showed nucleotides specific to *OTOAP1*, including the three in the coding sequence (p.Thr785Pro, p.Glu787* and p.Tyr806Ser). This allowed us to estimate that the size of the conversion on this second allele was less than 900bp and to code the conversion with the following HGVS nomenclature: NC_000016.10:g.(21735999_21736311)_(21736714_21736886)con(22552122_225552435)_(22552837_22553008).

Discussion

Gene conversion leading to pathogenic variants has already been described in a variety of human Mendelian diseases (Chen et al., 2007, 2010). They have been hypothesized to arise after a double-strand break where the damaged chromosome is repaired, after resection, using a non-allelic homologous region of either the sister chromatid or the homologous chromosome. However, it has been assumed that, in humans, conversion events longer than 3kbp are more likely occurring via a non-allelic homologous double recombination mechanism, because most well characterized gene conversion events typically replace regions of a few hundreds of nucleotides only (Chen et al., 2007). For instance, this mechanism has been proposed to be at the origin of the 12f2 deleted lineages of the human Y chromosome (Blanco et al., 2000). A formal distinction between the two mechanisms would require characterization of the by-product chromosome, which is not possible when investigating human alleles.

Given the lengths of the two conversions we described (more than 9kbp and less than 900 bp), we assume that the mechanisms at the origin of these two alleles are respectively non-allelic homologous double recombination and double strand break repair-mediated gene conversion events between *OTOA* and *OTOAP1*. Comparison between *OTOA* and *OTOAP1* reference sequences from GRCh38 shows identity higher than 99% over more than 42 kbp of aligned nucleotides, which is sufficient to promote non-allelic homologous recombination between the two loci (Gu et al., 2008).

Heterozygous conversion of one to two exons of *OTOA* was reported in 3 patients among a cohort of 686 patients with hearing loss (Shearer et al., 2014), without direct molecular confirmation. Estimating the allelic frequencies of the conversion in the population from exome sequencing or short-reads whole genome sequencing is non-trivial. It requires coverage depth comparison between regions of *OTOA* and *OTOAP1*. Using exome sequencing, only the exons showing sufficient divergence, allowing for unique mapping to either loci, can reliably be called as converted or not. The alignments of *OTOA* and *OTOAP1* exonic sequences from GRCh38 show differences only for exons 21, 22 and 29 and their 50 flanking nucleotides, with respectively 1, 7 and 1 changes. For the evaluation of pathogenicity, only the conversion of exon 22 appears relevant, as the change in exon 21 is synonymous and the one in exon 29 lies in the 3'UTR.

It is likely that polymorphisms in regions of segmental duplication in *OTOA* and *OTOAP1* are under evaluated due to the difficulty of genotyping non-unique sequences. Interestingly, the p.(Thr785Pro), p.(Glu787*) and p.(Tyr806Ser) mutations of *OTOA* exon 22 are present in gnomAD v2.1.1 (Karczewski et al., 2020) as low frequency alleles, as well as low quality alleles for p.(Thr785Pro) and p.(Glu787*). The corresponding opposite mutations on *OTOAP1* exon 2 are also present in this database. These observed substitutions could result from errors in the mapping algorithm, leading to sequencing reads being incorrectly assigned to their paralogous regions in the genome. The alternative hypothesis would be that recurrent conversions

between the two loci create alleles that appear as a mixture of *OTOA* and *OTOAP1* wild-type specific changes. This hypothesis is consistent with the observation that gene conversions contribute to a significant part of the shared polymorphism between paralogous regions in the human genome (Dumont and Eichler, 2013; Dumont, 2015).

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Acknowledgments

We thank Alexandre Reymond for critical reading of the manuscript. This work is supported by the Fondation Privée des HUG. None of the authors has a conflict of interest to declare.

References

- Azaiez H, Booth KT, Ephraim SS, Crone B, Black-Ziegelbein EA, Marini RJ, Shearer AE, Sloan-Heggen CM, Kolbe D, Casavant T, Schnieders MJ, Nishimura C, et al. 2018. Genomic Landscape and Mutational Signatures of Deafness-Associated Genes. *Am J Hum Genet* 103:484–497.
- Blanco P, Shlumukova M, Sargent CA, Jobling MA, Affara N, Hurles ME. 2000. Divergent outcomes of intrachromosomal recombination on the human Y chromosome: male infertility and recurrent polymorphism. *J Med Genet* 37:752–758.
- Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M, Helt G, Goodstein DM, Elisk CG, Lewis SE, Stein L, Holmes IH. 2016. JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol* 17:66.
- Casola C, Zekonyte U, Phillips AD, Cooper DN, Hahn MW. 2012. Interlocus gene conversion events introduce deleterious mutations into at least 1% of human genes associated with inherited disease. *Genome Res* 22:429–435.
- Chen J-M, Cooper DN, Chuzhanova N, Férec C, Patrinos GP. 2007. Gene conversion: mechanisms, evolution and human disease. *Nat Rev Genet* 8:762–775.
- Chen J-M, Férec C, Cooper DN. 2010. Gene Conversion in Human Genetic Disease. *Genes* 1:550–563.
- Collins RL, Brand H, Karczewski KJ, Zhao X, Alföldi J, Francioli LC, Khera AV, Lowther C, Gauthier LD, Wang H, Watts NA, Solomonson M, et al. 2020. A structural variation reference for medical and population genetics. *Nature* 581:444–451.
- Dumont BL. 2015. Interlocus gene conversion explains at least 2.7 % of single nucleotide variants in human segmental duplications. *BMC Genomics* 16:456.
- Dumont BL, Eichler EE. 2013. Signals of Historical Interlocus Gene Conversion in Human Segmental Duplications. *PLOS ONE* 8:e75949.
- Fromer M, Moran JL, Chambert K, Banks E, Bergen SE, Ruderfer DM, Handsaker RE, McCarroll SA, O'Donovan MC, Owen MJ, Kirov G, Sullivan PF, et al. 2012. Discovery and Statistical Genotyping of Copy-Number Variation from Whole-Exome Sequencing Depth. *Am J Hum Genet* 91:597–607.
- Gu W, Zhang F, Lupski JR. 2008. Mechanisms for human genomic rearrangements. *PathoGenetics* 1:4.
- Hoppman N, Aypar U, Brodersen P, Brown N, Wilson J, Babovic-Vuksanovic D. 2013. Genetic testing for hearing loss in the United States should include deletion/duplication analysis for the deafness/infertility locus at 15q15.3. *Mol Cytogenet* 6:19.
- Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, et al. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581:434–443.
- Krumm N, Sudmant PH, Ko A, O'Roak BJ, Malig M, Coe BP, NHLBI Exome Sequencing Project, Quinlan AR, Nickerson DA, Eichler EE. 2012. Copy number variation detection and genotyping from exome sequence data. *Genome Res* 22:1525–1532.
- Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, Jang W, Katz K, et al. 2016. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* 44:D862–D868.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760.
- MacDonald JR, Ziman R, Yuen RKC, Feuk L, Scherer SW. 2014. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res* 42:D986–D992.

- Maglic D, Stephen J, Malicdan MCV, Guo J, Fischer R, Konzman D, Mullikin JC, Gahl WA, Vilboux T, Gunay-Aygun M. 2016. TMEM231 Gene Conversion Associated with Joubert and Meckel-Gruber Syndromes in the Same Family. *Hum Mutat* 37:1144–1148.
- Mehta D, Noon SE, Schwartz E, Wilkens A, Bedoukian EC, Scarano I, Crenshaw EB, Krantz ID. 2016. Outcomes of evaluation and testing of 660 individuals with hearing loss in a pediatric genetics of hearing loss clinic. *Am J Med Genet A* 170:2523–2530.
- Moteki H, Azaiez H, Sloan-Heggen CM, Booth K, Nishio S, Wakui K, Yamaguchi T, Kolbe DL, Iwasa Y, Shearer AE, Fukushima Y, Smith RJH, et al. 2016. Detection and Confirmation of Deafness-Causing Copy Number Variations in the STRC Gene by Massively Parallel Sequencing and Comparative Genomic Hybridization. *Ann Otol Rhinol Laryngol* 125:918–923.
- Nance WE. 2003. The genetics of deafness. *Ment Retard Dev Disabil Res Rev* 9:109–119.
- Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Auwera GAV der, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, Shakir K, Thibault J, et al. 2018. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* 201178.
- Rajagopalan R, Murrell JR, Luo M, Conlin LK. 2020. A highly sensitive and specific workflow for detecting rare copy-number variants from exome sequencing data. *Genome Med* 12:14.
- Robinson JT, Thorvaldsdottir H, Wenger AM, Zehir A, Mesirov JP. 2017. Variant Review with the Integrative Genomics Viewer. *Cancer Res* 77:e31–e34.
- Shearer AE, Kolbe DL, Azaiez H, Sloan CM, Frees KL, Weaver AE, Clark ET, Nishimura CJ, Black-Ziegelbein EA, Smith RJH. 2014. Copy number variants are a common cause of non-syndromic hearing loss. *Genome Med* 6:37.
- Shearer AE, Smith RJH. 2015. Massively Parallel Sequencing for Genetic Diagnosis of Hearing Loss: The New Standard of Care. *Otolaryngol Neck Surg* 153:175–182.
- Sloan-Heggen CM, Bierer AO, Shearer AE, Kolbe DL, Nishimura CJ, Frees KL, Ephraim SS, Shibata SB, Booth KT, Campbell CA, Ranum PT, Weaver AE, et al. 2016. Comprehensive genetic testing in the clinical evaluation of 1119 patients with hearing loss. *Hum Genet* 135:441–450.
- Tessereau C, Leone M, Buisson M, Duret L, Sinilnikova OM, Mazoyer S. 2015. Occurrence of a non deleterious gene conversion event in the BRCA1 gene. *Genes Chromosomes Cancer* 54:646–652.
- Vona B, Hofrichter M a. H, Neuner C, Schroder J, Gehrig A, Hennermann JB, Kraus F, Shehata-Dieler W, Klopocki E, Nanda I, Haaf T. 2015. DFNB16 is a frequent cause of congenital hearing impairment: implementation of STRC mutation analysis in routine diagnostics. *Clin Genet* 87:49–55.
- Watson CM, Dean P, Camm N, Bates J, Carr IM, Gardiner CA, Bonthron DT. 2020. Long-read nanopore sequencing resolves a TMEM231 gene conversion event causing Meckel-Gruber syndrome. *Hum Mutat* 41:525–531.
- Yokota Y, Moteki H, Nishio S, Yamaguchi T, Wakui K, Kobayashi Y, Ohyama K, Miyazaki H, Matsuoka R, Abe S, Kumakawa K, Takahashi M, et al. 2019. Frequency and clinical features of hearing loss caused by STRC deletions. *Sci Rep* 9:.
- Zazo Seco C, Wesdorp M, Feenstra I, Pfundt R, Hehir-Kwa JY, Lelieveld SH, Castelein S, Gilissen C, Wijs IJ de, Admiraal RJ, Pennings RJ, Kunst HP, et al. 2017. The diagnostic yield of whole-exome sequencing targeting a gene panel for hearing impairment in The Netherlands. *Eur J Hum Genet* 25:308–314.
- Zhang F, Gu W, Hurles ME, Lupski JR. 2009. Copy Number Variation in Human Health, Disease, and Evolution. *Annu Rev Genomics Hum Genet* 10:451–481.