

The Genetics of Cholesteatoma Study. Loss-of-function variants in an affected family.

Running Title:

The Genetics of Cholesteatoma.

Key Points

1. The aetiology of cholesteatoma remains elusive. In a recent systematic review, we discussed reports of multiple cases of cholesteatoma within families, which suggests a genetic predisposition in some cases (1).
2. We have established a U.K. database and DNA sample bank that can be used to identify genetic variants that co-segregate with cholesteatoma in multiply-affected families. Recruitment to this Genetics of Cholesteatoma (GOC) Study is via the U.K. National Institute of Health Research Clinical Research Network.
3. This preliminary communication describes the results of whole exome sequencing (WES) of DNA extracted from participants in the first fully sequenced family recruited to the study.
4. Rare variants were filtered for co-segregation with the cholesteatoma phenotype, and for their putative functional impact.
5. We have identified loss of function variants in the genes *EGFL8* and *BTNL9* as candidate variants of interest. These are preliminary observations and the variants are of unknown significance to the disease pathology without replication or further investigation.

Introduction

The aetiology of cholesteatoma is uncertain. Abnormality of drum skin migration and desquamation is thought to predispose to the accumulation of cholesteatoma within the middle ear, which may be locally invasive and capable of causing bone destruction and associated inflammation (2, 3). Reports of familial clustering of disease suggest genetic predisposition, but to date no genetic susceptibility loci have been identified (1).

The lead author (PP) previously reported the largest series of families affected by cholesteatoma (4). One of the families reported in that study participated in the present

study to investigate susceptibility loci. This family has several affected members (figure 1), and we were able to obtain DNA from five members, enabling whole exome sequencing (WES) and analysis of variants to identify those segregating with disease. The affected identical twin girls presented with glue ear at 5 years of age and were treated with bilateral grommets. At 7 years of age, they presented with otorrhoea and hearing loss. A clinical diagnosis of cholesteatoma was made. In the unilateral case in one of the twins, an extensive cholesteatoma was revealed with disease filling the middle ear and extending towards the petrous apex with erosion of the horizontal facial nerve canal. In the other twin, more limited disease was revealed in the left ear with cholesteatoma in the anterior middle ear and antrum. In the right ear, the disease was limited to the attic region. Tympanomastoidectomy was performed resulting in small dry cavities and hearing preservation, as had been the case with the older brother. CT scans were not performed.

Sequencing the exome of affected and unaffected individuals in families affected by cholesteatoma has the potential to identify variants in coding DNA that co-segregate with the phenotype. This is a preliminary report describing the first use of a whole exome sequencing (WES) strategy to study the genetics of cholesteatoma in a family with three affected siblings and a wider family history of cholesteatoma and severe ear disease.

Methods

Ethics and Research Governance

The GOC study was granted ethical approval by East of England Cambridge Research Ethics Committee (reference REC 16/EE/01311, IRAS ID: 186786). The study is sponsored by the University of East Anglia.

Biological samples and DNA extraction

Blood samples from the five family members indicated in figure 1 were collected in 3ml EDTA tubes by the research nurse and DNA extraction was completed using the Gentra Puregene Blood Kit (Qiagen, UK). All DNA samples were quantitated, checked for purity, and subsequently stored in the sample bank.

Whole Exome Sequencing (WES): library preparation, target capture and sequencing method

The whole-exome capture and library construction was performed using an amended v5.1 protocol from NimbleGen (NimbleGen 2015). The Illumina HiSeq4000 platform was used to

generate paired-end 150bp reads from the libraries. Next-generation sequencing, library construction and primary bioinformatics analysis was delivered by the BBSRC National Capability in Genomics (BB/J010375/1) at Earlham Institute by members of the Genomics Pipelines Group.

Bioinformatics and variant filtering

Sequencing reads were aligned to the GRCh37 (hg19) version of the human genome reference sequence, using the *Bowtie 2* read aligner. Variants were called with *FreeBayes* (a Bayesian genetic variant detector) and variant call format (VCF) files were produced for each WES sample (5). *VCFtools* were used to intersect the genotype files to identify shared variants. Variants were annotated with functional information from Ensembl's *Variant Effect Predictor* (VEP) (6).

Variants that segregated with cholesteatoma in the siblings and in the mother were stratified to generate a list of candidate variants. For this preliminary analysis of the case series, we have considered a dominant model with incomplete penetrance. That is, we filtered for heterozygous variants in the affected individuals (and in the mother, because phenotypes of interest were recorded for her paternal uncle and for her nephew), and removed variants that are heterozygous or homozygous for an alternative (non-reference sequence) allele in the unaffected father.

Coding DNA variants were filtered to identify those predicted to have a moderate or high functional effect on the expressed protein, by VEP. The variants were filtered in two additional steps to identify the candidate variants of most interest:

1. To exclude common variants that are less likely to be pathogenic, the variants were sorted by the minor allele frequency recorded for European populations (7). This step was to exclude those variants with a frequency > 0.0119 .
2. The final filtering step excluded any remaining missense variants that were not predicted to have a damaging or deleterious potential by either the SIFT or Polyphen software tools (8, 9). These tools use different algorithms to predict the effect of amino acid changes on protein function.

Results

Two of the variants identified are predicted to have a high functional impact with the Ensembl VEP software. These are summarised in table 1. They are both rare, loss-of-function variants. A premature stop codon was identified in *EGFL8* and a frameshift mutation was identified in *BTNL9*.

Any variants predicted to have a moderate functional effect are listed in table 2. These are non-synonymous (amino acid changing) missense mutations.

The SIFT and Polyphen scores are shown in Table 2, alongside their qualitative prediction. Both tools use a range of zero to one, but with opposite meanings. A SIFT score of 1.0 is predicted to be benign, whereas a Polyphen score of 1.0 is predicted to be damaging.

Discussion and Conclusion

New Findings and Previous Studies

In a recent systematic review, we proposed that the pattern of cholesteatoma observed in this and some other families is typical of a monogenic or oligogenic disorder with incomplete penetrance. The mutations described in Table 1 are candidate variants of most interest, from this WES study, because they are predicted to be functionally important, high impact mutations. However, neither variant is rare enough to be presented as an exclusive underlying cause of cholesteatoma. As with most disease traits, we can predict that any genetic architecture (which is the number and effect size of any contributing variants) will be complex for cholesteatoma.

EGFL8 and *BTNL9*

EGFL8 is a highly conserved gene that encodes the EGF like domain multiple 8 protein, which is highly expressed in the skin (10). Interestingly, the stop-gain variant in *EGFL8*, rs141826798, has recently been reported to be significantly associated with psoriasis in UK Biobank participants (11). For both psoriasis and cholesteatoma, there is altered keratinocyte proliferation and differentiation. It is possible that *EGFL8* variants have pleiotropic effects and/or that there is a common biological driver for these pathologies.

BTNL9 encodes the butyrophilin like 9 protein. Butyrophilins and butyrophilin-like proteins are part of the immunoglobulin superfamily, and have roles in adaptive and innate immunity (12). There are no published disease-associations for the *BTNL9* variant, rs367635312, to our knowledge.

Missense variants

We present the rare variants in table 2 because missense variants can also contribute to the pathophysiology of complex diseases.

All deleterious rare variants that co-segregate with the disease phenotype will be of interest to researchers investigating familial cholesteatoma.

Implications for Research and Clinical Practice

These are interesting preliminary findings from the first WES study of cholesteatoma to date. Further study will be required to determine whether *EGFL8* or *BTNL9* variants, or indeed the missense variants, are of any significance to cholesteatoma pathology. We also recognise that our analysis for deleterious variants was not exhaustive for this study because other mutation-effect prediction algorithms could be included in our variant filtering protocol.

Additional families will now be analysed as part of the GOC study to identify and document other co-segregating variants with a putative functional impact. For families with similar vertical inheritance patterns, we will focus on a dominant model of inheritance, considering haploinsufficiency and gain of function variants. We acknowledge that this approach can only identify those variants with major or moderate effects on the disease phenotypes.

If we identify recurrent candidate genes and gene pathways through WES studies of multiply-affected families, this could increase our understanding of the cause and pathophysiology of cholesteatoma.

References

1. Jennings BA, Prinsley P, Philpott C, Willis G, Bhutta MF. The genetics of cholesteatoma. A systematic review using narrative synthesis. *Clin Otolaryngol*. 2018;43(1):55-67.
2. Semaan MT, Megerian CA. The pathophysiology of cholesteatoma. *Otolaryngol Clin North Am*. 2006;39(6):1143-59.
3. Bhutta MF, Williamson IG, Sudhoff HH. Cholesteatoma. *BMJ*. 2011;342:d1088.
4. Prinsley P. Familial cholesteatoma in East Anglia, UK. *J Laryngol Otol*. 2009;123(3):294-7.
5. Haplotype-based variant detection from short-read sequencing: Cornell University Library; 2012 [Available from: <https://arxiv.org/abs/1207.3907>].
6. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17(1):122.
7. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536(7616):285-91.
8. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*. 2009;4:1073.

9. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nature Methods*. 2010;7:248.
10. Epidermal Growth Factor-Like Protein 8: The Human Protein Atlas; 2018 [Available from: <https://www.proteinatlas.org/ENSG00000241404-EGFL8/tissue>].
11. Emdin CA, Khera AV, Chaffin M, Klarin D, Natarajan P, Aragam K, et al. Analysis of predicted loss-of-function variants in UK Biobank identifies variants protective for disease. *Nat Commun*. 2018;9(1):1613.
12. Malinowska M, Tokarz-Deptula B, Deptula W. Butyrophilins: an important new element of resistance. *Cent Eur J Immunol*. 2017;42(4):399-403.