Eliot Kmiec

# **23 & Me Variant Analysis**

**Methods**

Data for this analysis was pulled from the Personal Genome Project for a volunteer hu1A8C1E, whom I will refer to as Jane for narrative purposes throughout this analysis. Jane is a 43-year-old female of European descent with an unremarkable medical history.

The raw data file was processed using the ‘Annotate23andMe’ script provided by Dr. Edwards to annotate each variant using the SeattleSeq database. The resulting data was then explored using python ([github](https://github.com/ek775/Bioinformatics_MS_Notes/tree/main/homework/23%26Me)) and 3 variants were selected based on their predicted and reported pathogenicity. These variants were then assessed for predicted functional impact using AlphaMissense and PolyPhen-2, as well as observed clinical impact using ClinVar and OMIM. Additional literature was consulted as needed to understand gaps in predicted impact versus observed clinical evidence.

**APOE**

* **Accessions:** NM\_000041.4, NM\_001302688.2, NM\_001302689.2, NM\_001302690.1, NM\_001302691.2
* **dbSNP:** rs429358

According to the 23&Me report on the Personal Genome Project [page](https://my.pgp-hms.org/profile/hu1A8C1E) for Jane, this was the most significant gene variant that they found, and it is easy to see why. After cleaning null values from the annotated raw data, five entries were found in Jane’s data as being associated with APOE. All of the annotated entries indicated observation of homozygous substitution of T🡪C leading to the C130R variant frequently referred to as the “E4” variant. This variant is well described in the literature and in Uniprot/SwissProt, and has been classified as pathogenic by the ACMG1,2,3,4,6,8.

Apolipoprotein E (ApoE) is the lipid transport protein encoded by the APOE gene. It contains an LDL receptor domain, lipid-binding domain, and VLDL-binding domain1,7. The translated peptide is 317 residues in length, eventually being processed down to 299 residues in its mature form along with several glycosylations1,8.

Although this variant has been recognized as pathogenic for many years now, PolyPhen-2 does not predict this variant to be deleterious. Each of the entries in the annotated data registered a PolyPhen-2 score approaching zero, however, they also registered a Grantham score of 180.0 which does indicate that this variant may represent a significant evolutionary change from the wild-type. AlphaMissense prediction ([github](https://github.com/ek775/Bioinformatics_MS_Notes/tree/main/homework/23%26Me)) supports this, indicating that some structural instability may occur, although it seems that the protein continues to perform its primary function adequately for most of an afflicted individual’s lifespan4. However, this instability may also contribute to findings related to this gene’s involvement in late-stage Alzheimer’s Dementia5,6,8. For example, it is well documented that the E4 variant can be found sequestered with beta-amyloid plaques in brain tissue samples from deceased Alzheimer’s patients, and lab studies have shown that the E4 variant can be used to precipitate beta-amyloid fibrils in a matter of hours while the wild-type requires up to several days or weeks3,4. Despite all of this, the exact structural changes induced by this mutation remain unconfirmed and there is some debate over the impact of glycosylation on the E4 variant’s ability to bind to beta-amyloid particles3.

While perhaps not overly comforting to Jane, knowledge of her genetic risk does provide some opportunity to be proactive about treatment, and possibly preserve her quality of life. Early screening may be able to detect the neuronal changes prior to symptom development, and cholinesterase inhibitors and memantine may be able to slow the progression of disease6.

**GNB3**

* **Accessions:** NM\_001297571.2, NM\_002075.4, XM\_011520953.3
* **dbSNP:** rs5442

This GNB3 variant was selected for analysis based on its high “am\_pathogenicity” score (0.9813), with an associated PolyPhen-2 score of 0.997, and the listed “am\_class” of “likely pathogenic”. The am\_pathogenicity column and am\_class columns are curiously not described in the SeattleSeq documentation, but do seem to correlate loosely with other metrics of pathogenicity.

Each entry corresponds to a missense mutation involving a G🡪A substitution leading to a glycine🡪serine change at amino acid 272 in the translated protein. This gene encodes a key subunit of a G-protein coupled receptor (GPCR) called Guanine nucleotide binding-protein, subunit beta-3 (GBB3)9. Although there are a few variants in this gene that are reported to cause congenital night blindness, such damage is caused by SNPs located further toward the N-terminus, and this particular variant has not yet been reported to cause deleterious effects10. Interestingly, the AlphaMissense prediction seems to support this variant’s benign designation in ClinVar, as it predicts relatively minor damage from replacing this particular glycine with a serine despite its interior location10. It is worth noting, however, that the PDB file obtained from Uniprot for this protein is an AlphaFold prediction as there were no experimentally derived options available. Ultimately, the actual protein may fold differently than this prediction in both the wild-type and this variant.

Although variants in GBB3 tend to affect the optical nervous system, it is worth pointing out that this variant is expressed in many other tissues and is involved in several GPCRs interacting with different signaling pathways. This is partly due to its nature as a subunit in a trimeric protein complex, as well as its role as a common alternative splicing target9.

**PLET1**

* **Accession:** NM\_001145024.1
* **dbSNP:** rs2564872

PLET1 was chosen

**References**

(1) Uniprot/Swiss-Prot, Apolipoprotein E, APOE, P02649 (Accessed 11/30/2023) <https://www.uniprot.org/uniprotkb/P02649/entry>

(2) Clinvar: RCV001195807.10 (Accessed 11/30/2023) <https://www.ncbi.nlm.nih.gov/clinvar/RCV001195807.1/>

(3) Forero, DA, López-León, S, González-Giraldo, Y, Dries, DR, Pereira-Morales, AJ, Jiménez, KM, Franco-Restrepo, JE. 2018. APOE Gene and Neuropsychiatric Disorders and Endophenotypes: A Comprehensive Review. Am J Med Genet Part B 177B: 126–142.

(4) OMIM, 107741, Apolipoprotein E; APOE (Accessed 11/30/2023) <https://www.omim.org/entry/107741?search=APOE&highlight=apoe>

(5) Lanfranco, M.F.; Ng, C.A.; Rebeck, G.W. ApoE Lipidation as a Therapeutic Target in Alzheimer’s Disease. Int. J. Mol. Sci. 2020, 21, 6336. <https://doi.org/10.3390/ijms21176336>

(6) Mayo Clinic, Alzheimer’s Disease (Accessed: 11/30/2023) <https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/diagnosis-treatment/drc-20350453>

(7) Zannis VI, McPherson J, Goldberger G, Karathanasis SK, Breslow JL. Synthesis, intracellular processing, and signal peptide of human apolipoprotein E. J Biol Chem. 1984 May 10;259(9):5495-9. PMID: 6325438.

(8) Paik YK, Chang DJ, Reardon CA, Davies GE, Mahley RW, Taylor JM. Nucleotide sequence and structure of the human apolipoprotein E gene. Proc Natl Acad Sci U S A. 1985 May;82(10):3445-9. doi: 10.1073/pnas.82.10.3445. PMID: 2987927; PMCID: PMC397792.

(9) Uniprot/Swiss-Prot, Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-3, GNB3, P16520|GBB3\_Human (Accessed 11/30/2023) <https://www.uniprot.org/uniprotkb/P16520/entry>

(10) National Center for Biotechnology Information. ClinVar; [VCV001170122.8], <https://www.ncbi.nlm.nih.gov/clinvar/variation/VCV001170122.8> (accessed Dec. 11, 2023).