S1 Appendix. Detailed description for PM1, PP3, BP4, BP7 criteria implementation and the pathogenicity probability computation

PM1
The precise regions, used for PM1 criterion, are the pore-forming domain of KCNQ4 gene and the three-stranded helices of the collagen genes COL11A2, COL4A3, COL4A4 and COL4A5. PM1 is applied to missense variants overlapping any of the annotated genomic regions. More specifically, if the variant overlaps on the three-stranded motifs of the collagen genes, it accepts only the matches which affect the Glycine residues contained in a Gly-X-Y motif.

PP3, BP4 and BP7
The used thresholds by prediction follow. To decide upon pathogenicity, we aggregate CADD and REVEL in the following scheme: if CADD score is greater than 20, then we set CADD_{vote} = 1, otherwise CADD_{vote} = 0. For REVEL, if REVEL score is greater or equal to 0.7, then REVEL_{vote} = 1, alternatively if REVEL score is lower or equal to 0.15, then REVEL_{vote} = 0, otherwise we set REVEL_{vote} = 0.5. Finally, if the average voting of CADD_{vote} and REVEL_{vote} is greater or equal to 1, GenOtoScope predicts the variant as pathogenic.

For splicing impact, we aggregate the predictors MaxEntScan and dbscSNV in the following scheme: if \(|\frac{\text{observed score}}{\text{reference score}} - \text{reference score}|\) is greater than 0.15, then MaxEntScan\(_{vote} = 1\), otherwise MaxEntScan\(_{vote} = 0\). For dbscSNV, if either ADA score or RF score is greater than 0.6 then dbscSNV\(_{vote} = 1\), otherwise dbscSNV\(_{vote} = 0\). We aggregate the votes similarly to pathogenicity. That is, if the average voting of MaxEntScan\(_{vote}\) and dbscSNV\(_{vote}\) is greater or equal to 1, GenOtoScope decides that the variant has a splicing impact.

For conservation prediction, we use PhyloP score, as follows: if PhyloP score is greater of 1.6 then GenOtoScope decides that this is a conserved site, otherwise that the site is not a conserved site.

Computation of pathogenicity probability
GenOtoScope applies the naive Bayes model to calculate the posterior probability of pathogenicity given the triggered ACMG evidence rules using the following equations:

\[
\text{Pathogenicity}_{\text{posterior}} = \frac{\text{Pathogenicity}_{\text{likelihood}} \cdot \text{Pathogenicity}_{\text{prior}}}{(\text{Pathogenicity}_{\text{likelihood}} - 1) \cdot \text{Pathogenicity}_{\text{prior}} + 1} \tag{1}
\]

\[
\text{Pathogenicity}_{\text{likelihood}} = O_{\text{PVST}} \left( \frac{N_{\text{refu}} + N_{\text{refm}} + N_{\text{refst}} + N_{\text{refvst}} - N_{\text{refu}} - N_{\text{refm}}}{X} \right), \tag{2}
\]

where the default parameters are used: Pathogenicity\(_{\text{prior}} = 0.1\), \(O_{\text{PVST}} = 350\) and \(X = 2\).

The calculation of the pathogenicity probability is calculated for all input variants automatically.