

## ESTABLISHING THE PRECISE EVOLUTIONARY HISTORY OF A GENE IMPROVES PREDICTION OF DISEASE-CAUSING MISSENSE MUTATIONS

**Ogun Adebali<sup>1,2</sup>, Alexander O. Reznik<sup>2</sup>, Daniel S. Ory<sup>3</sup>, Igor B. Zhulin<sup>1,2\*</sup>**

<sup>1</sup>Oak Ridge National Laboratory, Oak Ridge, USA, <sup>2</sup>University of Tennessee, Knoxville, USA, <sup>3</sup>Washington University, St. Louis, USA

Email: [ijouline@utk.edu](mailto:ijouline@utk.edu)

\*Corresponding author

Predicting the phenotypic effects of mutations has become an important application in clinical genetic diagnostics. Computational tools evaluate the behavior of the variant over evolutionary time and assume that variations seen during the course of evolution are probably benign in humans. However, current tools do not take into account orthologous/paralogous relationships. Paralogs have dramatically different roles in Mendelian diseases. For example, whereas inactivating mutations in the *NPC1* gene cause the neurodegenerative disorder Niemann-Pick C, inactivating mutations in its paralog *NPC1L1* are not disease-causing and, moreover, are implicated in protection from coronary heart disease. We identified major events in *NPC1* evolution and revealed and compared orthologs and paralogs of the human *NPC1* gene through phylogenetic and protein sequence analyses involving more than 300 eukaryotic genomes. We predicted whether an amino acid substitution affects protein function by reducing the organism's fitness. Removing the paralogs and distant homologs improved the overall performance of categorizing disease-causing and benign amino acid substitutions. The results show that a thorough evolutionary analysis followed by identification of orthologs improves the accuracy in predicting disease-causing missense mutations. We anticipate that this approach will be used as a reference in the interpretation of variants in other genetic diseases as well.