Evolutionary analysis of NPC1 improves prediction of disease causing missense mutations

Ogun Adebali[1,.2], Alexander O. Reznik[3] and Igor B. Zhulin[1,.2]

[1]*Computer Science & Mathematics Division, Oak Ridge National Laboratory, Oak Ridge TN 37831 USA*

[2]*Department of Microbiology, University of Tennessee, Knoxville TN 37966 USA*

[3] *Pavlov First St. Petersburg State Medical University, St. Petersburg, Russian Federation,*
ijouline@utk.edu

Niemann-Pick type C (NPC) is a rare and fatal lysosomal storage disease, caused primarily (~95%) by mutations in the *npc1* gene, which encodes a large membrane-bound protein that is involved in sterol transport. Impaired protein results in the accumulation of unesterified cholesterol and glycosphingolipids in endosomes, causing a lipid trafficking problem. This autosomal recessive disease is a remarkably variable condition and therefore may not be easily diagnosed by patient symptoms. Hence, there have been attempts to diagnose the disease using sequence information. However, distinguishing between benign and damaging variants is one of the major challenges in sequence-based diagnoses. Here, we show that a thorough evolutionary analysis of the NPC1 protein enables more accurate risk estimation for single amino acid variants (SAVs). Our phylogenetic and protein sequence analysis revealed true NPC1 orthologs and paralogs, and distinguished them from similar proteins with different functions. Consequently, we were able to identify conserved amino acid positions in the NPC1 protein that are function specific. This approach overcomes many inconsistencies introduced by automated tools that are unable to discriminate between similar sequences of proteins with different functions. This work provides useful guidelines for NPC diagnosis based on NPC1 sequence. Furthermore, we anticipate that our approach will be applicable to sequence-based diagnosis of other genetic diseases.