

## **BrainGrab: Capturing Curator Expertise as Reusable Annotation Rules**

Daniel H. Haft and Roland A. Richter

Experienced biocurators can outperform automated systems on specific genes once they determine which pieces of evidence should drive annotation, and which annotations should be spread. The annotation logic may invoke both homology evidence (BLAST matches or HMM hits) and non-homology evidence (neighboring genes, metabolic context, taxonomic group). Biocurators routinely build temporary expertise that, regrettably, tends to fall away when the logic driving the annotation is not captured along with the annotation itself. We report here the development of BrainGrab, an interface added to the MANATEE manual annotation tool for prokaryotic genomes. The curator can specify evidence scenarios that should always lead to equivalent annotation for similar genes in similar contexts, and thus create new annotation rules while the expertise is fresh. No special knowledge of programming or protein family construction is required. BrainGrab rules can mix and match evidence types from the large array of existing protein family definitions such as Pfam families, sequence analyses such as SignalP, and contextual clues, that is, the same types of evidence already familiar to experienced biocurators. We have now created an infrastructure for collecting, distributing, interpreting, and applying BrainGrab rules for automated annotation. A rules interpreter combines queries of existing evidence with specified new searches to determine if a rule must fire. If so, the interpreter writes a new piece of rule-based evidence. Once deposited, BrainGrab/RuleBase evidence can provide automated annotation, pathway reconstruction, and even input data for other rules. We demonstrate the system with sets of rules for annotating proteins and pathways of siderophore biosynthesis in human pathogens, for annotating common fusion proteins, and for applying the proper nomenclature to bacterial ribosomal proteins. The chance to harness curatorial expertise for building rules creates a promising avenue for broad community contribution to improved annotation pipelines.