

The 'unknome': A database of human genes we know almost nothing about

August 8 2023



Credit: Pixabay/CC0 Public Domain

Researchers from the United Kingdom hope that a new, publicly available database they have created will shrink, not grow, over time. That's because it is a compendium of the thousands of understudied

proteins encoded by genes in the human genome, whose existence is known but whose functions are mostly not.

The [database](#), dubbed the "unknome," is the work of Matthew Freeman of the Dunn School of Pathology, University of Oxford, England, and Sean Munro of MRC Laboratory of Molecular Biology in Cambridge, England, and colleagues, and is described in the open access journal *PLOS Biology*. Their own investigations of a subset of proteins in the database reveal that a majority contribute to important cellular functions, including development and resilience to stress.

The sequencing of the [human genome](#) has made it clear that it encodes thousands of likely [protein](#) sequences whose identities and functions are still unknown. There are multiple reasons for this, including the tendency to focus scarce research dollars on already-known targets, and the lack of tools, including antibodies, to interrogate cells about the function of these proteins. But the risks of ignoring these proteins are significant, the authors argue, since it is likely that some, perhaps many, play important roles in critical cell processes, and may both provide insight and targets for therapeutic intervention.

To promote more rapid exploration of such proteins, the authors created the unknome database (www.unknome.org), that assigns to every protein a "knownness" score, reflecting the information in the [scientific literature](#) about function, conservation across species, subcellular compartmentalization, and other elements.

Based on this system, there are many thousands of proteins whose knownness is near-zero. Proteins from model organisms are included, along with those from the human genome. The database is open to all and is customizable, allowing the user to provide their own weights to different elements, thereby generating their own set of knownness scores to prioritize their own research.

To test the utility of the database, the authors chose 260 genes in humans for which there were comparable genes in flies, and which had knownness scores of 1 or less in both species, indicating that almost nothing was known about them. For many of them, a complete knockout of the gene was incompatible with life in the fly; partial knockdowns or tissue-specific knockdowns led to the discovery that a large fraction contributed to essential functions influencing fertility, development, tissue growth, protein quality control, or stress resistance.

The results suggest that, despite decades of detailed study, there are thousands of fly genes that remain to be understood at even the most basic level, and the same is clearly true for the human genome. "These uncharacterized genes have not deserved their neglect," Munro said. "Our database provides a powerful, versatile and efficient platform to identify and select important [genes](#) of unknown function for analysis, thereby accelerating the closure of the gap in biological knowledge that the unknown represents."

Munro adds, "The role of thousands of human proteins remains unclear and yet research tends to focus on those that are already well understood. To help address this we created an Unknown database that ranks proteins based on how little is known about them, and then performed functional screens on a selection of these mystery proteins to demonstrate how ignorance can drive biological discovery."

More information: Rocha JJ, Jayaram SA, Stevens TJ, Muschalik N, Shah RD, Emran S, et al. (2023) Functional unknowns: Systematic screening of conserved genes of unknown function, *PLoS Biology* (2023). [DOI: 10.1371/journal.pbio.3002222](https://doi.org/10.1371/journal.pbio.3002222)

Provided by Public Library of Science

Citation: The 'unknome': A database of human genes we know almost nothing about (2023, August 8) retrieved 26 June 2024 from <https://phys.org/news/2023-08-unknome-database-human-genes.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.