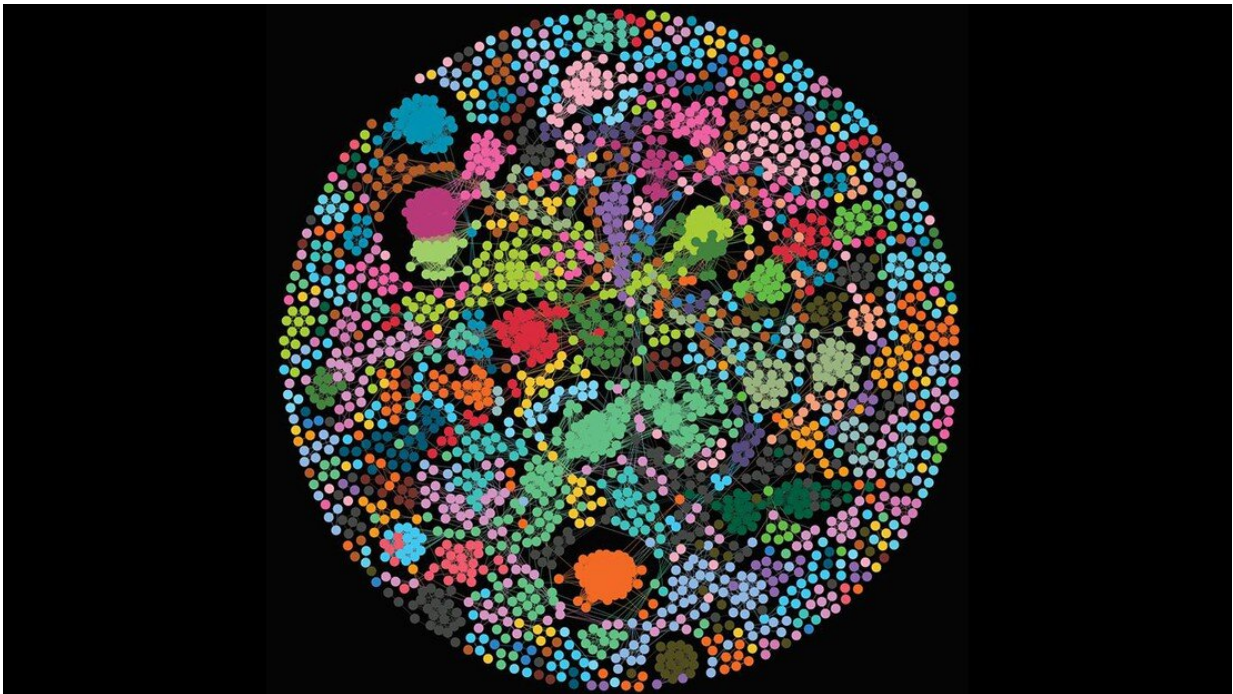


# Big data toolkit to mine the dark genome for precision medicine

December 4 2019

---



Gene networks. Credit: J. Auwerx, EPFL

EPFL researchers have developed Big Data tools for identifying new gene functions. The work identifies millions of connections between genes and their functions, and can facilitate the development of precision medicine.

Genes are the functional units of heredity, and the understanding of gene

function is the major focus of biomedical research, serving as the basis of precision [medicine](#). However, most research efforts have been devoted to only a small part of the genes, neglecting the larger "dark genome." This impedes our understanding of the underlying mechanisms of complex traits and diseases, which is necessary for the advancement of precision medicine.

"Most of the research are gene-oriented and largely influenced by our [prior knowledge](#), therefore many potentially important genes are ignored," says Johan Auwerx, whose lab at EPFL led the study, along with colleagues from University of Lausanne and University of Tennessee, and EPFL professors Kristina Schoonjans and Stephan Morgenthaler.

In an article published in *Genome Research*, the scientists address the issue of the "dark genome" by developing novel approaches based on systems genetics. "Genes with similar functions tend to have similar expression patterns," explains first author Hao Li. "We used this feature to predict the function of unknown genes by learning from those of the known ones."

The researchers collected large-scale gene-expression datasets containing more than 300,000 samples from six different species. They then used these to develop a toolkit termed "GeneBridge" that can identify potential gene functions. The [toolkit](#) was then used by the team to identify hundreds of thousands of novel functions of [genes](#), many of which have been verified by Auwerx's group as well as by other research groups.

"We have deposited GeneBridge and its seven billion [data points](#) on [systems-genetics.org](#) along with the already existing 300 million data points," says Auwerx. "This resource will undoubtedly improve our knowledge of the 'dark [genome](#),' and promote the development of

precision medicine."

**More information:** Hao Li et al. Identifying gene function and module connections by the integration of multispecies expression compendia, *Genome Research* (2019). [DOI: 10.1101/gr.251983.119](https://doi.org/10.1101/gr.251983.119)

Provided by Ecole Polytechnique Federale de Lausanne

Citation: Big data toolkit to mine the dark genome for precision medicine (2019, December 4) retrieved 11 May 2024 from <https://phys.org/news/2019-12-big-toolkit-dark-genome-precision.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.