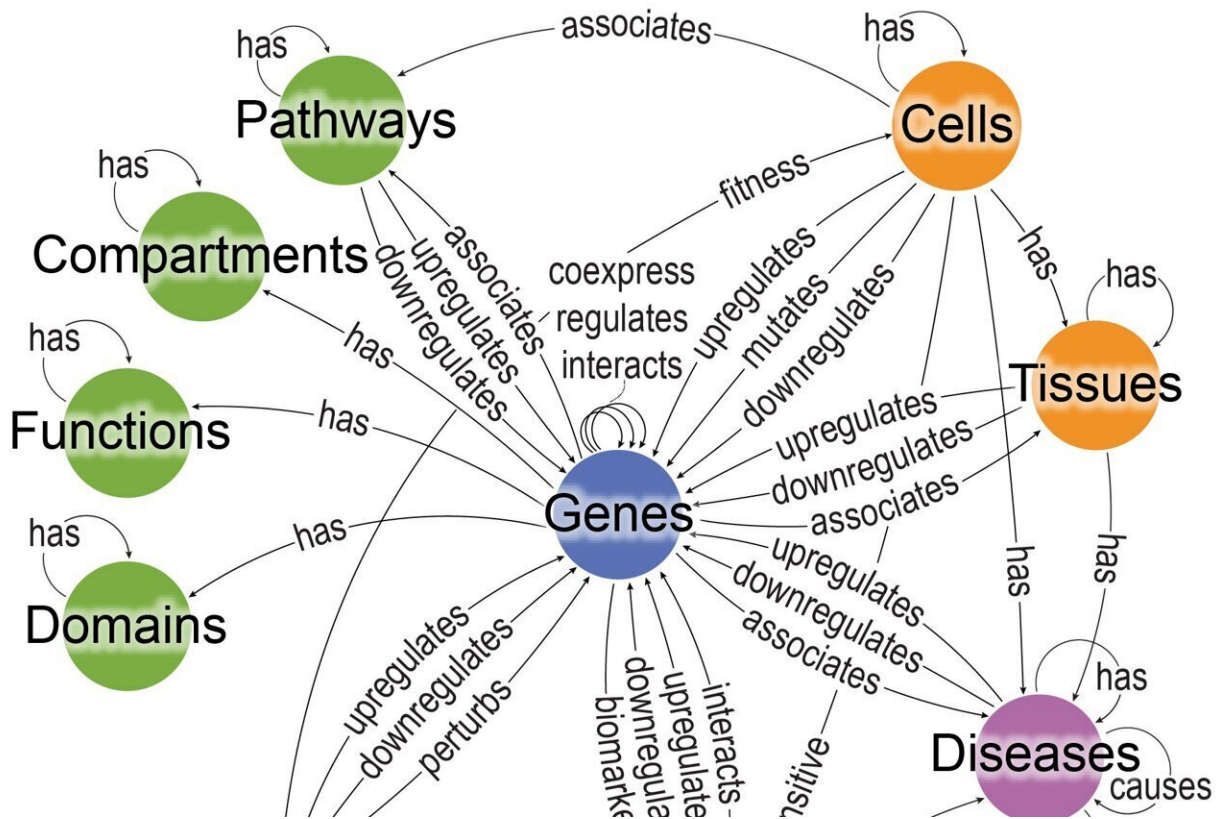


# The Bioteque: A computational tool to harmonize biological knowledge

September 15 2022



Bioteque is a resource of descriptors for different biological entities. By traversing this knowledge graph through specific entities and relationships we explored more than 1000 paths (aka metapaths) which were encoded into numerical vectors and made available for the community. Credit: IRB Barcelona

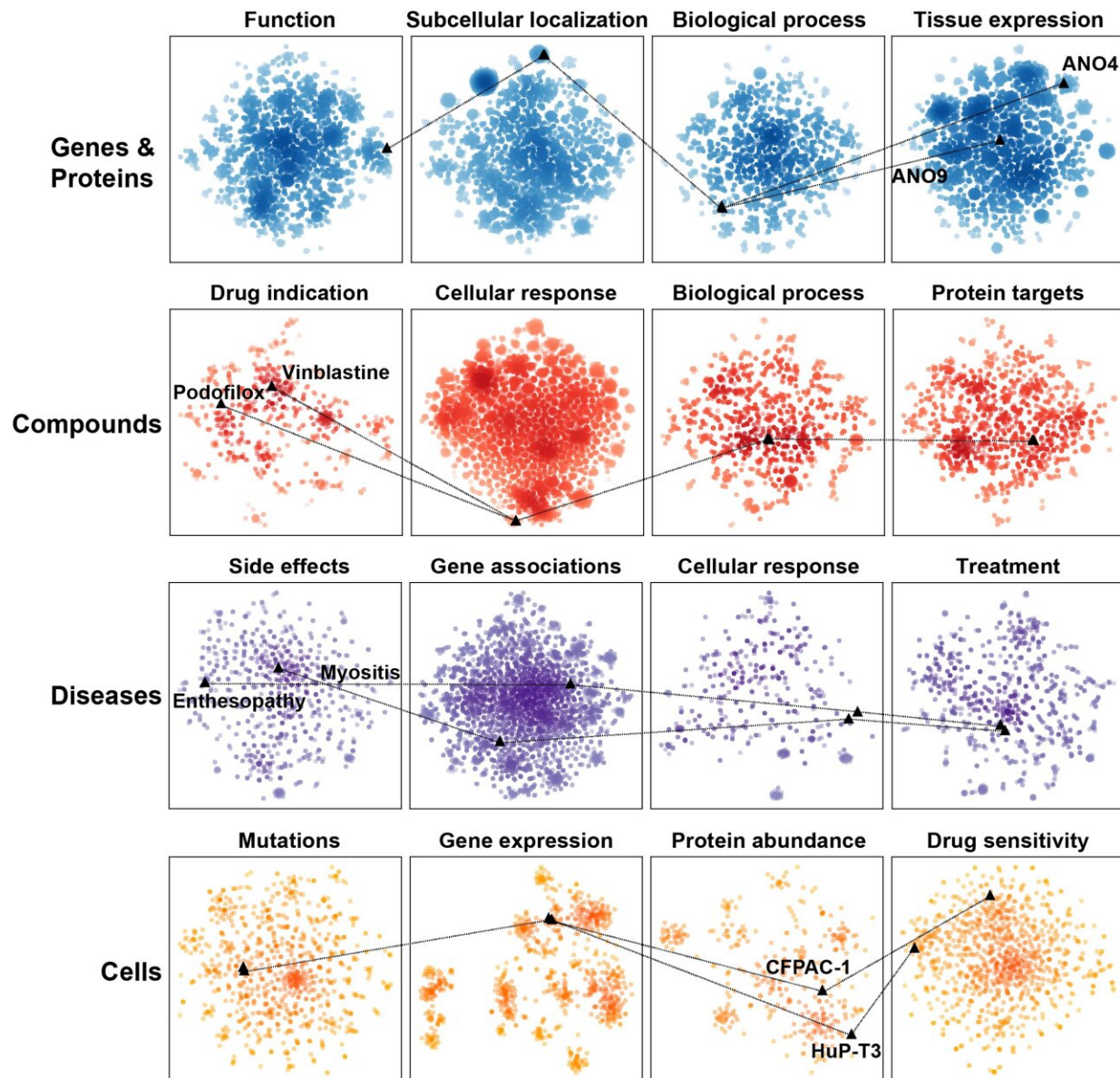
The rapid development of the different disciplines in the fields of

biological and biomedical research (such as genomics, proteomics, and transcriptomics) in recent decades has led to exponential growth in the amount of biological data available. For example, at the European Bioinformatics Institute (EMBL-EBI), they have gone from managing a volume of 40 petabytes to working with 250 petabytes in just 6 years.

Scientists led by Dr. Patrick Aloy, ICREA researcher and head of the Structural Bioinformatics and Network Biology laboratory at IRB Barcelona, have developed a [computational tool](#) to harmonize, integrate and simplify these data. The result is a [knowledge graph](#) that provides information on how different biological entities are related to each other, including more than 30 million functional interactions.

The Bioteque works by integrating different levels of biological complexity and thus can report, for example, on two [genes](#) that are related, whether they physically interact, whether they are active in the same type of [cells](#), and whether they are related to the same disease. It can also predict the sensitivity or resistance of a type of cell to a specific drug.

"This computational resource that we've developed is one of the first aimed at unifying biological information and it's the only one to address such diversity and amount of data. It allows access, in an easy and harmonized way, to practically all the biological knowledge currently available, and it has enormous potential to accelerate [biomedical research](#)," explains Aloy.



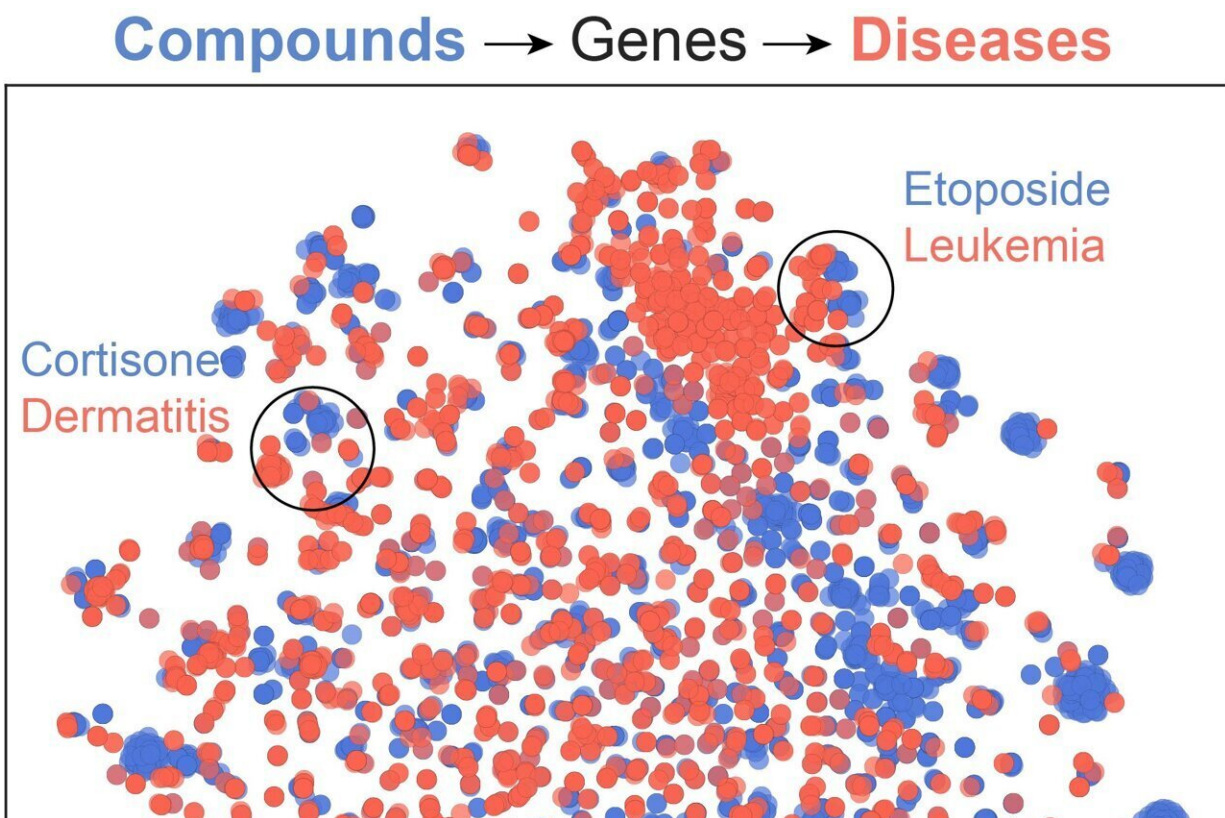
Illustrating 4 different descriptors for 4 types of biological entities. Credit: IRB Barcelona

### Almost 1,000 descriptors for 12 biological entities

The information held in the Bioteque is structured into 12 types of biological entities, such as gene, disease, tissue, cell, etc. For each of

these entities, the tool considers a series of descriptors or characteristics, for example, the pattern of mutations of a gene, the profile of physical interactions of the resulting proteins, the expression of said gene in different cell types, or its relationship with different diseases. Among the 12 biological entities, the system covers around 1,000 types of descriptors.

"We have worked with information from 150 different databases, so first we had to integrate them, that is, put them all in the same 'language'. And then we converted that knowledge into numerical descriptors that could be interpreted by algorithms, and that way we could computationally exploit these networks and connections," concludes Adrià Fernández, the first author of the article and a doctoral student in the same laboratory.





Three groups are highlighted where diseases and their treatments are associated.  
Credit: IRB Barcelona

The Bioteque will be expanded periodically with new databases, as they are made public. Both the tool and the databases and algorithms are [open access](#) and are available online.

The research was published in *Nature Communications*.

**More information:** Adrià Fernández-Torras et al, Integrating and formatting biomedical data as pre-calculated knowledge graph embeddings in the Bioteque, *Nature Communications* (2022). [DOI: 10.1038/s41467-022-33026-0](#)

Tool/database: [bioteque.irbbarcelona.org/](https://bioteque.irbbarcelona.org/)

Provided by Institute for Research in Biomedicine (IRB Barcelona)

Citation: The Bioteque: A computational tool to harmonize biological knowledge (2022, September 15) retrieved 22 June 2024 from <https://phys.org/news/2022-09-bioteque-tool-harmonize-biological-knowledge.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.