Combining pathway data resources gives a clearer view on complex biological interactions
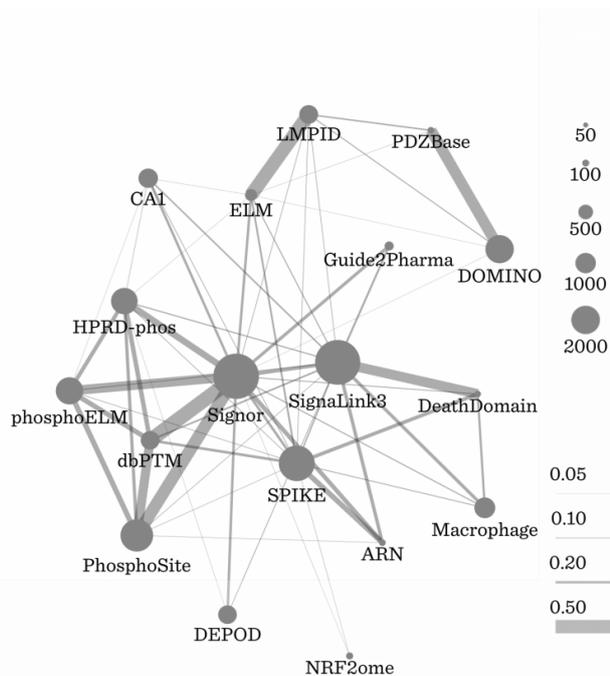
29 November 2016

Fig. 1: Overlap of interactions across causal resources. Sizes of circles represent the number of interactions per resource, and connection line widths show the overlap of interactions between them, measured by Simpson index (Equation S3).

Combining the power of 27 data resources, Omnipath helps researchers see biological signalling pathways with unprecedented accuracy. Developed by researchers in the UK and Germany and published in *Nature Methods*, OmniPath offers a comprehensive, unified collection of literature-curated signalling pathways based on an analysis of 41,000 scientific papers.

All the functions happening in our cells are controlled by groups of molecules working together through signalling pathways. Once the first

molecule receives a signal, the next one is activated, and so on. When things go wrong in these pathways, cancer can develop. Many cancer drugs work by putting up roadblocks in a pathway, stopping the signal and hopefully the growth of cancerous tissue.

To figure out how signalling pathways work, molecular biologists carry out and validate experiments, sometimes over many years, to characterise the exact interactions taking place between proteins.

Researchers can share the results of these pathway studies in public databases, to build knowledge collectively. The data are put together with the results of thousands of published studies on molecular interactions. These are organised by expert 'curators' so they are discoverable, and can help researchers shape new experiments or analyse new results.

There are now over 27 public databases on signalling interactions, each of which offers something different and many of which offer custom formats. OmniPath, developed by researchers at EMBL-EBI, RWTH Aachen University and the Earlham Institute, gives a unified view of all the 'literature-curated' signalling interactions in these databases.
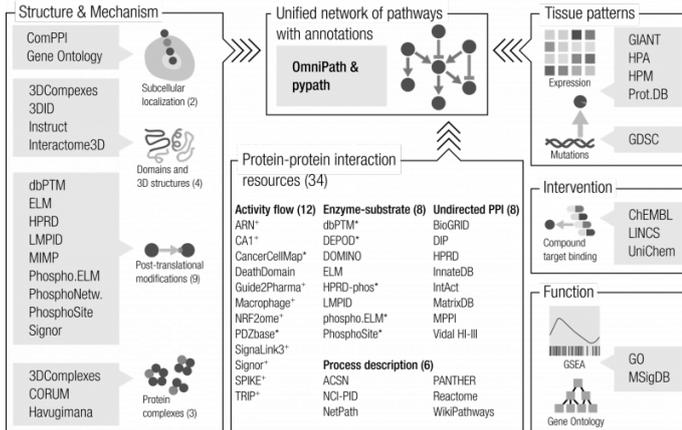
Fig. 2: Overview of OmniPath and pypath.

The data in OmniPath are primarily based on small-scale experiments, but its Pypath software makes it possible to add datasets obtained from large screening experiments or converted from reactions. Pypath (a Python module) lets users build custom signalling networks and combine them with other data. It is a powerful tool for incorporating pathways into bioinformatics workflows and makes the analysis behind OmniPath fully open source, transparent and easily reproducible.

"We compared all manner of signalling data resources and clarified the properties of different datasets, which helps researchers make better-informed decisions in their analyses," says Julio Saez-Rodriguez, visiting group leader at EMBL-EBI and professor at RWTH Aachen. "It has already proved very valuable for the research within our groups, and we hope others will find it valuable as well."

More information: Turei D, Korcsmaros T and Saez-Rodriguez J (2016) Omnipath: guidelines and gateway for literature-curated signaling pathway resources. *Nature Methods* 13(12); published online 29 November 2016. DOI: 10.1038/nmeth.4077

At its launch, OmniPath has references to more than 41,000 original studies, with data representing 36,557 interactions between 7,984 proteins. The interactome, which describes all the biological interactions in an organism, could include anywhere from 100,000 to 250,000 interactions in a human. That is a huge amount of information to piece together, so accuracy and consistency are paramount.

"The work of data curators is invaluable because without them the data would never come together with the kind of precision you need in biology," says Dénes Türei, EIPOD postdoctoral fellow at EMBL-EBI. "It has been exciting to work together with people from so many disciplines, and produce this concise view into the collective, current knowledge of signalling pathways."

"Researchers tend to trust the accuracy of curated resources, without looking too deeply into their actual content and methods," says Tamás Korcsmáros, Fellow of the Earlham Institute and Institute of Food Research. "Benchmarking studies have mainly focused on resources with interactions from high-throughput experiments, and even these have been few and far between."

The new study provides comprehensive guidelines, based on an extensive examination of more than 50 data resources, to help researchers select the most appropriate data resource for their work.

Provided by Earlham Institute