SignaFish: a zebrafish-specific signaling pathway resource

Kitti Csályi[1,2,*], Dávid Fazekas[1,3*], Tamás Kadlecsik[1], Dénes Türei[1,4], Leila Gul[1], Balázs Horváth[1], Dezső Módos[1,5,6], Amanda Demeter[1,7], Nóra Pápai[1], Katalin Lenti[5], Péter Csermely[8], Tibor Vellai[1], Tamás Korcsmáros[1,3,7,§], Máté Varga[1,9]

[*] These authors contributed equally to this work.

[1] Department of Genetics, Eötvös Loránd University, Budapest, Hungary

[2] Max Delbrück Center for Molecular Medicine, Berlin, Germany

[3] TGAC, The Genome Analysis Centre, Norwich Research Park, Norwich, UK

[4] European Molecular Biology Laboratory – European Bioinformatics Institute; Wellcome Trust Genome Campus; Hinxton, UK

[5] Department of Morphology and Physiology, Faculty of Health Sciences, Semmelweis University, Budapest, Hungary

[6] Centre for Molecular Informatics, University of Cambridge, Cambridge, UK

[7] Gut Health and Food Safety Programme, Institute of Food Research, Norwich Research Park, Norwich, UK

[8] Department of Medical Chemistry, Semmelweis University, Budapest, Hungary

[9] MTA-SE Lendület Nephrogenetic Laboratory, Budapest, Hungary

[§] corresponding author: Tamas.Korcsmaros@tgac.ac.uk

Abstract

nderstanding living systems requires an in depth knowledge of the signaling etworks that drive cellular homeostasis, regulate intercellular communication and contribute to cell fates during development. Several resources exist to provide high-throughput datasets or manually curated interaction information from human or invertebrate model organisms. We previously developed SignaLink, a uniformly curated, multi-layered signaling resource containing information for human and for the model organisms nematode *Caenorhabditis elegans* and fruit fly *Drosophila melanogaster*. Until now, the use of the SignaLink atabase for zebrafish pathway analysis was limited. To overcome this limitation we created SignaFish (http://signafish.org), a fish-specific signaling resource, built using the concept of SignaLink. SignaFish contains more than 200 curation based signaling interactions, 132 further interactions listed in other resources, and it also lists potential miRNA based regulatory connections for 7 major signaling pathways. From the SignaFish website, users can reach other web resources, such as ZFIN. SignaFish provides signaling or signaling-related interactions that can be examined for each gene, or downloaded for each signaling pathway. We believe that the SignaFish resource will serve as a novel navigating point for experimental design and evaluation for the zebrafish community and for researchers focusing on non-model fish species, such as cyclids.

The biology of living organisms cannot be interpreted without an in depth knowledge of the signaling networks that drive cellular homeostasis, regulate intercellular communication and contribute to establish cell fates, which have an indispensable role both during and after development. While early research concentrated on deciphering individual signaling pathways, later results demonstrated that these pathways are often unexpectedly interwoven [1, 2]. This particular feature of signaling networks can often contribute to seemingly confusing results upon the analysis of interactions between signaling proteins. Without accounting for the details of an interaction (direction or sign), the interpretation of experimental data will be often erroneous.

A better understanding of signaling network topology allows for a better accounting of the observed phenotypes by recognizing robustness within the examined network or identifying key proteins, which could become potential targets in future treatments or interventions [3]. Hence the need for well annotated protein-protein interaction (PPI) and signaling databases for biological and biomedical research.

SignaLink (http://signalink.org/) [4, 5], a uniformly curated, multi-layered signaling resource has become over the years a widely used signaling resource for the model organisms nematode *Caenorhabditis elegans* and fruit fly *Drosophila melanogaster*, as well as a gap-filling human database with high coverage. With an easy-to-use interface, customizable download site and integrated datasets, it had been used in several high profile studies (for example [6]) and also integrated in model-organism resources, such as FlyBase and WormBase.

However, the use of SignaLink for the zebrafish community was limited. To overcome this limitation, we developed SignaFish (http://signafish.org), a fish-specific signaling resource, built using the concept of SignaLink.

With SignaFish, we have created a manually curated database of seven major signaling pathways, which are biochemically and evolutionary defined, and encompass all major developmental signaling mechanisms [1]: Rtk (receptor tyrosine kinase), Tgf-β (transforming growth factor beta), Wnt/Wingless, Hedgehog, Jak/Stat, Notch and NHR (Nuclear hormone receptor). The pathway interactions were coming from (1) review-based literature searches, which identified the primary zebrafish-related paper that described the interactions; or (2) we used the signaling interactions of *C. elegans, D. melanogaster* and *H. sapiens* from SignaLink2 [5], and based on the interolog and signalog concepts [7, 8], we predicted potential zebrafish signaling interactions. We included a predicted interaction to the SignaFish database if (1) we found no direct evidence that the given interaction has been verified in zebrafish, (2) the orthologs of potentially interacting proteins were present and interacted in at least one species in SignaLink, (3) and, to minimize the false positive interactions, a zebrafish study showed that genetic modification of both potentially interacting proteins produces a phenotype related to the annotated pathway.

SignaFish contains altogether 217 signaling interactions found with manual curation or with high-confident predictions (as described above). These interactions were extended with protein-protein interactions (imported from [9, 10]), and potential miRNA-mRNA connections (from TargetScanFish [11]). Altogether, SignaFish contains pathway and interaction information for 389 proteins and 178 miRNAs (Figure 1). The number of manually curated interactions directly found in zebrafish studies is low (86) compared to the total number of papers available for zebrafish, because we applied a strict curation protocol [4], which allowed the inclusion of only those studies where the identified interaction was investigated between two zebrafish proteins, and the experiments were carried out in zebrafish. Due to the lack of zebrafish specific regulatory resources listing transcription factor – target gene connections, we were not able to include transcriptional regulations to this version of SignaFish. The database structure of SignaFish already allows the incorporation of such transcriptional data upon a large dataset will be available.

To facilitate the exploration of this complex resource, we implemented it to a website with a user-friendly graphical interface, available at http://signafish.org. The website is linked to main web resources including UniProt, ENSEMBL and ZFIN.

We clearly recognize that the list of the included data is far from complete and we are committed in updating SignaFish annually. We welcome inputs from the wider community of zebrafish researchers, therefore, we included an online form (available from the website) that could be used to submit novel interactions, which – after manual verification – will be included in the next update of SignaFish.

Since its publication the SignaLink database has proven its usefulness for invertebrate model organism, therefore we believe SignaFish will also become an important resource for zebrafish research. The signaling interactions that can be examined for each gene, or downloaded for each signaling pathway will serve as a novel navigating point for experimental design and evaluation for the zebrafish community and for researchers focusing on non-model fish species.

Acknowledgement

We thank the technical help of Kata Ferenc. This work was supported by research grants from the Hungarian Science Foundation [grant numbers: OTKA K115378, K109349, K115378, NK78012], by the Institute of Advanced Studies Kőszeg (iASK, Kőszeg, Hungary), a Technology Innovation Fund grant from the Hungarian National Developmental Agency (KTIA-AIK-2012-12-1-0010). TD was supported by the EMBL Interdisciplinary Postdoc Programme (EIPOD) under Marie Skłodowska-Curie COFUND Actions (Grant number 291772). TK is a Computational Biology Fellow at The Genome Analysis Centre in partnership with the Institute of Food Research (Norwich, UK), and strategically supported by Biotechnological and Biosciences Research Council, UK. MV was supported by János Bolyai and MedInProt fellowships of the Hungarian Academy of Sciences.

Reference List

1. Pires-daSilva A, Sommer RJ: **The evolution of signalling pathways in animal development**. *Nat Rev Genet* 2003, **4:**39-49.

2. Barabasi AL, Oltvai ZN: **Network biology: understanding the cell's functional organization**. *Nat Rev Genet* 2004, **5:**101-113.

3. Csermely P, Korcsmaros T, Kiss HJ, London G, Nussinov R: **Structure and dynamics of molecular networks: A novel paradigm of drug discovery: A comprehensive review**. *Pharmacol Ther* 2013, **138:**333-408.

4. Korcsmaros T, Farkas IJ, Szalay MS, Rovo P, Fazekas D, Spiro Z, Bode C, Lenti K, Vellai T, Csermely P: **Uniformly curated signaling pathways reveal tissue-specific cross-talks and support drug target discovery**. *Bioinformatics* 2010, **26:**2042-2050.

5. Fazekas D, Koltai M, Turei D, Modos D, Palfy M, Dul Z, Zsakai L, Szalay-Bek M, Lenti K, Farkas IJ, Vellai T, Csermely P, Korcsmaros T: **SignaLink 2 -- a signaling pathway resource with multi-layered regulatory networks**. *BMC Syst Biol* 2013, **7:**7.

6. Vinayagam A, Zirin J, Roesel C, Hu Y, Yilmazel B, Samsonova AA, Neumuller RA, Mohr SE, Perrimon N: **Integrating protein-protein interaction networks with phenotypes reveals signs of interactions**. *Nat Methods* 2014, **11:**94-99.

7. Yu H, Luscombe NM, Lu HX, Zhu X, Xia Y, Han JD, Bertin N, Chung S, Vidal M, Gerstein M: **Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs**. *Genome Res* 2004, **14:**1107-1118.

8. Korcsmaros T, Szalay MS, Rovo P, Palotai R, Fazekas D, Lenti K, Farkas IJ, Csermely P, Vellai T: **Signalogs: orthology-based identification of novel signaling pathway components in three metazoans**. *PLoS One* 2011, **6:**e19240.

9. Chatr-Aryamontri A, Breitkreutz BJ, Heinicke S, Boucher L, Winter A, Stark C, Nixon J, Ramage L, Kolas N, O'Donnell L, Reguly T, Breitkreutz A, Sellam A, Chen D, Chang C, Rust J, Livstone M, Oughtred R, Dolinski K, Tyers M: **The BioGRID interaction database: 2013 update**. *Nucleic Acids Res* 2013, **41:**D816-D823.

10. Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, Duesbury M, Dumousseau M, Feuermann M, Hinz U, Jandrasits C, Jimenez RC, Khadake J, Mahadevan U, Masson P, Pedruzzi I, Pfeiffenberger E, Porras P, Raghunath A, Roechert B, Orchard S, Hermjakob H: **The IntAct molecular interaction database in 2012**. *Nucleic Acids Res* 2012, **40:**D841-D846.

11. Lewis BP, Burge CB, Bartel DP: **Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets**. *Cell* 2005, **120:**15-20.

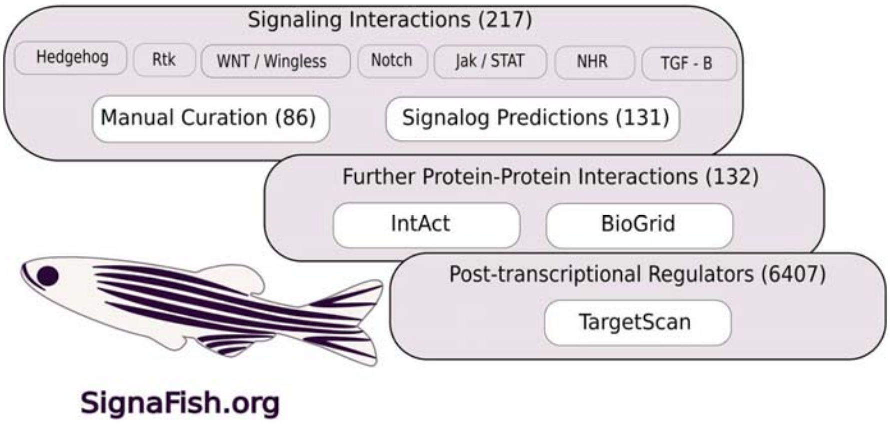

Figure 1. The structure of SignaFish and its sources. The number of interactions acquired from each sources is shown in parenthesis. All parts of the SignaFish database can be browsed and downloaded through the website, where users can filter to specific signaling pathways.