PrOnto database: GO term functional dissimilarity inferred from biological data.

Charles E Chapple, Carl Herrmann, Christine Brun

frontiers in Genetics

TECHNOLOGY REPORT
published: 03 June 2015
doi: 10.3389/fgene.2015.00200

PrOnto database : GO term functional dissimilarity inferred from biological data

Charles E. Chapple[1,2], *Carl Herrmann*[1,2†] *and Christine Brun*[1,2,3*]

[1] Inserm, UMR_S1090 TAGC, Marseille, France, [2] Aix-Marseille Université, UMR_S1090 TAGC, Marseille, France, [3] Centre National de la Recherche Scientifique, Marseille, France

Moonlighting proteins are defined by their involvement in multiple, unrelated functions. The computational prediction of such proteins requires a formal method of assessing the similarity of cellular processes, for example, by identifying dissimilar Gene Ontology terms. While many measures of Gene Ontology term similarity exist, most depend on abstract mathematical analyses of the structure of the GO tree and do not necessarily represent the underlying biology. Here, we propose two metrics of GO term *functional dissimilarity* derived from biological information, one based on the protein annotations and the other on the interactions between proteins. They have been collected in the PrOnto database, a novel tool which can be of particular use for the identification of moonlighting proteins. The database can be queried via an web-based interface which is freely available at http://tagc.univ-mrs.fr/pronto.

Keywords: moonlighting protein, gene ontology, functional similarity, protein-protein interactions, database

OPEN ACCESS

Edited by:
Constance J. Jeffery, University of Illinois at Chicago, USA

Reviewed by:
Mikhail P. Ponomarenko, Institute of Cytology and Genetics of Siberian Branch of Russian Academy of Sciences, Russia
Daniele Merico, The Hospital for Sick Children, Canada

****Correspondence:***
Christine Brun, Inserm, UMR_S1090 TAGC, Parc Scientifique de Luminy, Case 928, Marseille F-13009, France
brun@tagc.univ-mrs.fr

†Present Address:
Carl Herrmann, IPMB - University Heidelberg and DKFZ - Department of Theoretical Bioinformatics, Heidelberg, Germany

Specialty section:
This article was submitted to Bioinformatics and Computational Biology, a section of the journal Frontiers in Genetics

Received: *20 February 2015*
Accepted: *21 May 2015*
Published: *03 June 2015*

Citation:
Chapple CE, Herrmann C and Brun C (2015) PrOnto database : GO term functional dissimilarity inferred from biological data. Front. Genet. 6:200. doi: 10.3389/fgene.2015.00200

1. Introduction

Moonlighting proteins are a subset of multifunctional proteins involved in several, unrelated biological functions. Because of the growing importance of this functional singularity (Copley, 2012) for the understanding of cellular regulations and human diseases (Jeffery, 2011), computational methods for the large scale prediction of moonlighting proteins have long been awaited (Khan and Kihara, 2014). Yet, so far, most of the known moonlighting proteins were serendipitous discoveries (Mani et al., 2015). One of the major hurdles that need to be overcome in order to tackle such a task is defining the notion of "unrelated functions." What are biologically "unrelated functions" in the context of moonlighting? How can they be defined according to the current gene/protein functional annotations in way that computers can understand?

The Gene Ontology (GO) (Ashburner et al., 2000) is a controlled vocabulary of terms to describe gene product functions. Over the last decade, it has become the *de facto* standard ontology used to formalize gene annotation data. It is organized as three independent directed acyclic graphs (DAGs), one for each of the sub-ontologies Biological Process (BP), Molecular Function (MF), and Cellular Component (CC).

The structure of the GO DAG means that many GO terms are related, either because they describe related functions or because one term is the child of another. Therefore, proteins annotated to similar GO terms are assumed to perform similar functions and can be categorized as such. This has led to various methods of evaluating the semantic similarity of GO annotations (reviewed in Gan et al., 2013). Most of these depend on the relationships between the terms in the DAG, either by measuring their distance as the number of edges connecting them, or by evaluating

their information content. Such methods can therefore identify semantically similar GO terms, cases where the terms are linked in the structure of the DAG. The identification of moonlighting proteins requires defining *dissimilar functions*. However, *semantically dissimilar* GO terms are often clearly connected from a biological perspective, and therefore semantic similarity measures are not the best option for implementation in a moonlighting discovery pipeline. For instance, the terms "response to tumor necrosis factor" (GO:0034612) and "positive regulation of apoptotic process" (GO:0043065) share no parent terms apart from the root of the ontology although they are descriptions of tightly linked biological processes. Indeed, TNF is a well known inducer of apoptosis (see Gaur and Aggarwal, 2003 for a review) and "positive regulation of apoptotic process" describes one of the cellular "responses to tumor necrosis factor." These *semantically dissimilar* terms can, therefore, be considered *functionally similar* since they are different descriptions of the same or obviously connected biological processes. This similarity is reflected in the fact that the two terms co-occur in the annotations of multiple proteins (e.g., 19 and 25 in the mouse and human proteomes, respectively).

We have therefore developed PrOnto, a web-based tool that provides two metrics of GO *functional dissimilarity* based on gene product GO annotations and protein-protein interactions (PPI). We use the frequency of co-occurrence of GO term pairs in (i) protein annotations (Annotation Probabilities, *APs*) and in (ii) the annotations of interacting protein pairs (Interaction Probabilities, *IPs*) to compute probabilities reflecting biases toward infrequent GO terms associations implying *functional dissimilarity* (**Figure 1**). In this paper, we present the metrics, the webtool we provide to the community as well as different usage examples among which our recent characterization of potential moonlighting proteins from the human PPI network (Chapple et al., 2015).

The current version of the PrOnto database contains probabilities for human, mouse, fly, worm and yeast (see **Supplementary Table 1** for database statistics). The database will be regularly updated to keep up with annotation and interaction data. PrOnto is free and accessible through a simple web-based interface (see **Figure 2** available at http://tagc.univ-mrs.fr/pronto).

2. Materials and Methods

2.1. Annotation and Interaction Probabilities

We express *functional dissimilarity* as a probability of GO term pair co-occurrence modeled by a hypergeometric distribution. Let X be the variable "number of proteins annotated to both terms" which follows a hypergeometric law. The probability of observing this or smaller values of X by chance is given by

$$P(X \leq k_0) = \sum_{k=0}^{k_0} P(X = k) = \sum_{k=0}^{k_0} \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \quad (1)$$

with

$$P(X = 0) = \frac{\binom{N-K}{n}}{\binom{N}{n}} \quad (2)$$

where, for *AnnotationProbabilities* (*APs*), N is the number of proteins annotated directly to at least two different GO terms, K is the number of proteins annotated to GO_1, n is the number of proteins annotated to GO_2 and k is the number of proteins annotated to both terms.

For *InteractionProbabilities* (*IPs*), N is the number of interactions in the PPI network between proteins annotated directly to at least two different GO terms. K is the number of interactions involving proteins annotated to GO_1, n the number of interactions involving proteins annotated to GO_2 and k the number of interactions between a protein annotated to GO_1 and one annotated to GO_2.

In both cases, N is the size of the event pool. $P(X)$ is the probability of observing as large a co-occurence as X in a set of size N. Therefore, when computing N, only proteins with at least 2 direct annotations (i.e., explicit annotations, not including the implicit parent terms) are considered since co-occurence is only relevant for proteins annotated with at least two different GO terms. When calculating cross-ontology probabilities, only proteins with at least one explicit annotation in each ontology of interest are considered. When calculating K, n, and k, all annotations per protein are counted, both direct and inherited. All GO annotations have been included, irrespective of their evidence codes. Indeed, electronic annotations have greatly improved in recent years and their reliability now rivals that of manual annotations (Skunca et al., 2012). Pairs whose low-tail p-value is below a user-defined threshold (default: 0.05) are listed as "Dissimilar" and all others as "Not dissimilar."

In addition, the Jaccard index and Cohen's kappa have been computed for all pairs.

2.2. Building High Quality PPI Networks

To calculate the *IPs*, a high quality interactome was compiled for each target species. Interaction data were retrieved using the PSIQUICK (Aranda et al., 2011) interfaces of the APID (Prieto and Rivas, 2006), BioGrid (Chatr-Aryamontri et al., 2013), IntAct (Kerrien et al., 2012), DIP (Salwinski et al., 2004), MINT (Ceol et al., 2010), MatrixDB (Chautard et al., 2009), Reactome (Croft et al., 2011), InnateDB (Lynn et al., 2008), MolCon, Spike (Elkon et al., 2008), and TopFind (Lange and Overall, 2011) databases. Interaction data were filtered by identification method and only binary interactions between proteins were kept. A full list of the PSI-MI IDs used to build out networks is provided as **Supplementary Table 2**.

Protein names were mapped to UniProt IDs, and sequences, downloaded from UniProt, clustered using CD-HIT (Fu et al., 2012). TrEMBL/SwissProt protein pairs sharing ≥95% similarity were considered to be the same protein: interactions of the TrEMBL protein were then inherited by the Swiss-Prot protein. Self interactions were discarded.

The final result was high quality interactomes consisting entirely of experimentally verified, direct, binary interaction pairs

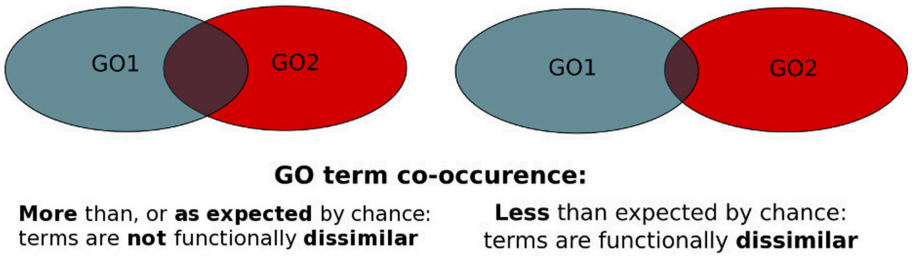

FIGURE 1 | PrOnto probabilities principle. The blue set represents the proteins annotated to GO1, the red set represents proteins annotated to GO2. The intersection corresponds to either proteins annotated to both terms (*APs*) or interactions involving proteins annotated to both terms (*IPs*).

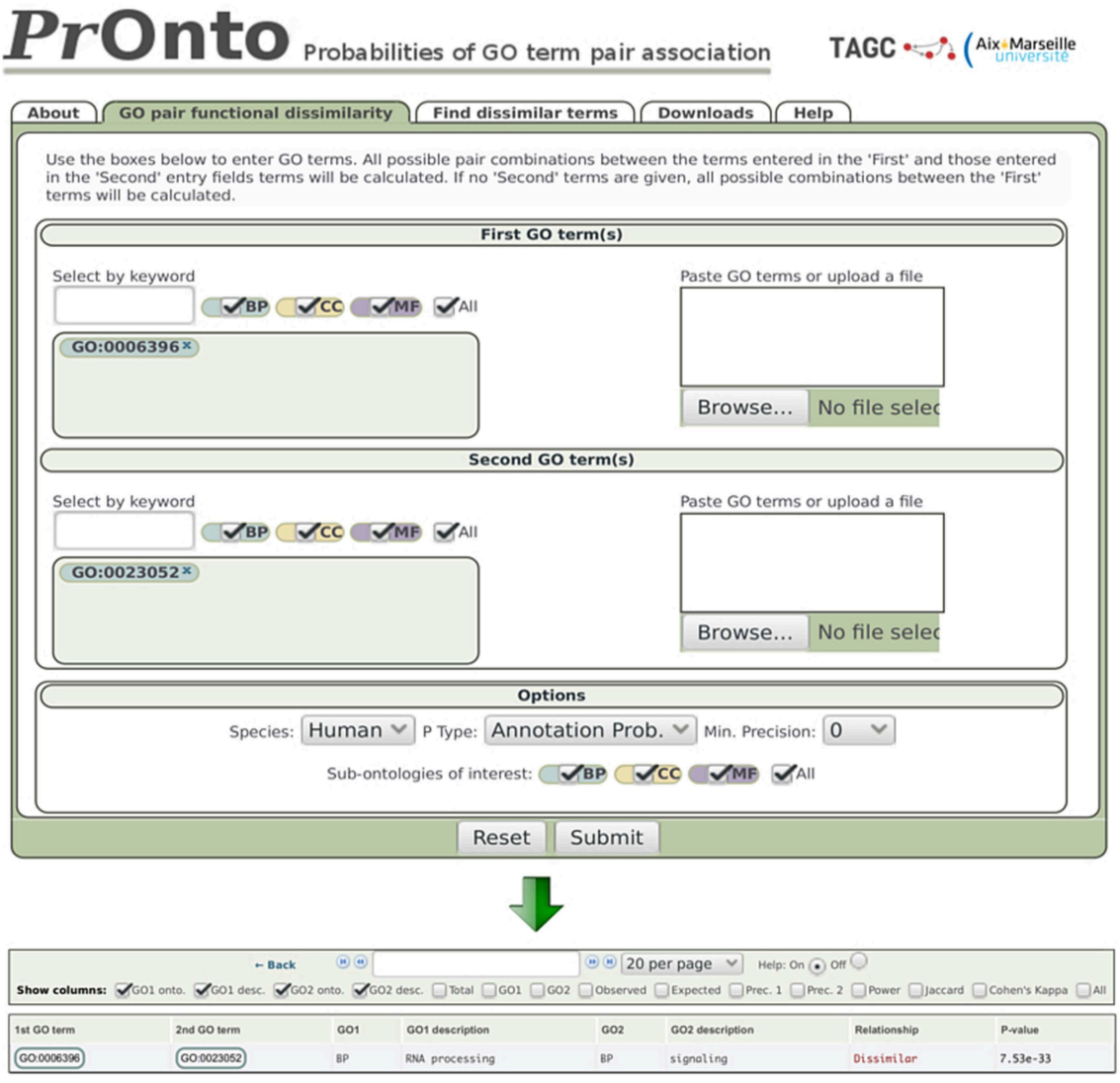

FIGURE 2 | PrOnto web interface. In the example shown, two terms have been submitted. The lower panel shows the results for *APs*.

for each species studied. These networks will be regularly updated with each update of PrOnto. Current interactomes are available on the downloads page of the PrOnto database (http://tagc.univ-mrs.fr/pronto/index.php?id=downloads).

2.3. Tools and Resources Used

The probabilities were calculated using the `phyper` function of the R statistical environment (R Core Team, 2012) and protein annotations were taken from the EBI's QuickGO

service (Barrell et al., 2009) (ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/). The power of the test was calculated using the `power.fischer.test` function of the `statmod` R package and 1000 simulations per pair. Jaccard indeces and Cohen's kappa were calculated using a simple `awk` script. The PrOnto webpage is written using a combination of HTML, PHP, and Javascript, the data are stored in a MySQL database which is queried using a Perl script.

3. Results and Discussion

3.1. PrOnto Content: The Different PrOnto Categories

PrOnto probabilities have been computed from the proteomes and interactomes of several species (human, mouse, fly, worm, yeast) to assess the relationships between GO terms of the same sub-ontology as well as between sub-ontologies.

Dissimilar - When the low-tail $p < 0.05$, the GO terms are very rarely associated among protein annotations and are therefore qualified as "dissimilar" by PrOnto. For example, the probability of finding a lower co-occurrence than observed for the GO terms "RNA processing" (GO:0006396) and "signaling" (GO:0023052) is very low in human (therefore highly significant), as indicated by their *AP*, $p = 7.5e\text{-}33$ (**Figure 2**). The co-occurrence is even less probable between interacting proteins since for *IPs*, $p = 1.4e\text{-}230$. These GO terms are then functionally "dissimilar" and very rarely linked through protein-protein interactions. Interestingly, these processes have been shown to be instead linked through protein-RNA interactions (Hogan et al., 2008). To demonstrate the validity of our approach when identifying dissimilar pairs, the power of the hypergeometric test was calculated for all pairs for which the null hypothesis was not rejected (dissimilar pairs). The results are shown in **Supplementary Figure 1**. Notably, the power was very high for the overwhelming majority of pairs (mean = 0.90 and median = 0.99).

Not Dissimilar - When the $p \geq 0.05$, PrOnto returns *Not Dissimilar*. For instance, in the human proteome, the probability of finding a greater co-occurrence than observed for the terms "response to tumor necrosis factor" (GO:0034612) and "positive regulation of apoptosis" (GO:0043065) is low (*AP* $p = 1$, therefore highly significant), clearly indicating that the terms are functionally "similar." This is explained by the following: as 116 human proteins are annotated to "response to tumor necrosis factor" and 459 to "positive regulation of apoptosis," 3 are expected to be annotated to both given the human proteome size (17866 annotated proteins), but 25 co-annotated proteins are observed. The same is observed for *IP* as proteins that respond to TNF signaling will often interact with those that promote apoptosis.

NA - Additionally, a *NA* category exists when no score could be computed. PrOnto produces *NA* when at least one of the GO terms of the pair under consideration does not annotate any protein in the target species. For instance, PrOnto returns "NA" for the GO terms GO:0030326 ("embryonic limb morphogenesis") and GO:0048736 ("appendage development") in yeast. This makes perfect sense biologically speaking since, for obvious reasons, no yeast proteins will be annotated to either of those terms.

Globally, as shown in **Table 1**, where the percentages of dissimilar and not dissimilar GO term pairs are reported for each species, most GO term pairs are not dissimilar according to PrOnto (*AP*, 84.1–87.1% and *IP*, 54.2–72.3%). Since the cell is a complex system whose constituent parts are very often interlinked (Schwikowski et al., 2000), many GO terms co-occur more often than expected by chance among gene/protein annotations or between interacting proteins. This link between processes in the cell is thus captured by PrOnto which is based on functional data.

As expected, since PrOnto is based on existing protein annotations, dissimilar terms according to PrOnto are rare (*AP*, 0.2–0.6%). That they are more numerous in the *IPs* (0.5–4.9%) shows functions that are rarely carried out by interacting proteins, therefore suggesting that they are performed by different functional modules, known as groups of interacting proteins involved in the same biological process (Spirin and Mirny, 2003). Interestingly, a larger proportion of dissimilar pairs has been identified in the organisms for which interaction data are more complete (human and yeast), highlighting the necessity of deciphering protein-protein networks in other organisms to gain a deeper understanding of the links between functions.

Table 2 shows the percentage of dissimilar GO term pairs per ontology for human. Interestingly, the fact that dissimilar

TABLE 1 | Percentage of GO term pairs from all ontologies (including cross-ontology pairs) that are Dissimilar or Not Dissimilar for all species.

	AP		IP	
	Dissimilar	**Not dissimilar**	**Dissimilar**	**Not dissimilar**
Human	0.3	99.6	4.0	96
Mouse	0.2	99.8	0.5	99.6
Fly	0.5	99.5	2.2	97.8
Worm	0.6	99.4	2.2	97.8
Yeast	0.5	99.5	4.9	95.2

"Dissimilar" corresponds to low-tail $p < 0.05$ and "Not Dissimilar" to all other cases.

TABLE 2 | Percentage GO term pairs from each ontology, excluding cross-ontology, for human (MF: Molecular Function, CC: Cellular Component, BP: Biological Process) that are Dissimilar or Not Dissimilar.

	AP		IP	
	Dissimilar	**Not dissimilar**	**Dissimilar**	**Not dissimilar**
MF	0.4	99.6	2.0	98
CC	1.3	98.6	7.0	93.1
BP	0.3	99.8	4.2	95.7

"Dissimilar" corresponds to low-tail $p < 0.05$ and "Not Dissimilar" to all other cases.

terms reach 7% for *IP* CC pairs reflect the shuttling of proteins throughout different cell compartments.

3.2. PrOnto Usage to Predict Moonlighting Protein Candidates

Unlike most tools that provide a measure of GO term functional association, PrOnto was conceived in order to identify dissimilar terms. Indeed, whereas most comparable approaches are geared toward the identification of similar terms, PrOnto has the advantage of being able to identify terms that very rarely co-occur. We therefore used PrOnto probabilities in a protein-protein network analysis dedicated to the discovery of moonlighting candidates and implemented them in our MoonGO pipeline (Chapple et al., 2015). The pipeline first extracts overlapping clusters from a PPI network using the OCG algorithm (Becker et al., 2012). These clusters are formed by highly interconnected proteins which tend to be involved in the same cellular processes. The cellular process(es) in which the clusters are involved, are identified based on the BP GO annotations of their constituent proteins, following a majority rule. Potential moonlighting proteins are then identified at the intersection of clusters involved in unrelated biological processes according to PrOnto GO term association probabilities. Using both APs and IPs ensures that the multiple functions in which the candidate protein is found to be involved, are very rarely performed (i) by a single protein and (ii) by interacting proteins. APs and IPs are therefore used here as two proxies, indicators of unrelated cellular functions.

Using this approach, we have identified 430 moonlighting candidates that form a distinct sub-group of proteins displaying specific features, distinguishing them from non-candidates proteins and constituting a signature of extreme multifunctionality. Among the striking features, candidates are more connected in the network, enriched in short linear motifs and in disease-related proteins compared to non-candidates, and are less intrinsically disordered than network hubs (Chapple et al., 2015; Zanzoni et al., 2015). These results therefore underline that PrOnto is particularly well suited to identify moonlighting candidates from biological data since it is especially stringent when determining term dissimilarity (see **Tables 1**, **2**).

As we provide PrOnto probabilities for multiple species, we predict that it will be soon used for the identification of moonlighting candidates in these other species. Finally, it should be noted that a recent analysis of GO terms has been proposed to identify moonlighting candidates in *E. coli* (Khan et al., 2014). Unlike our approach which uses PrOnto probabilities to assess the dissimilarity of functions of functional modules, this work is comparing the extent to which each function of the moonlighting candidates is described by the GO terms annotating the proteins, using a semantic similarity measure. The approach is therefore also using GO term annotations but in a completely different context.

3.3. Other PrOnto uses

Overall, *APs* can be used to identify cellular processes that are rarely carried out by the same proteins (i.e., "dissimilar"), whereas *IPs* can offer insights into the links between different cellular processes mediated by protein interactions. In addition their use for the investigation of the links between the functional modules formed by interacting proteins in PPI networks as described above, *APs* and *IPs* could also be used to score protein interaction predictions. In some cases, in the absence of experimental data, protein-protein interactions are predicted using bioinformatics approaches. One may then want to assign a confidence score based on real biological data to these predictions. Because PrOnto probabilities are derived from biological knowledge and experimental data, they can be used for this purpose and a lower score can be assigned to predicted interactions between proteins annotated to dissimilar terms.

They can also be used when constructing ontologies from -omics data as recently proposed by Dutkowski et al. (2013) and Kramer et al. (2014), the latter of which uses semantic similarity for assessment purposes. PrOnto can also help guide protein annotations as in the annotation tool GOAT (Bada et al., 2004) which uses term functional association scores to annotate proteins of unknown function. Once one term has been assigned to a protein, dissimilar terms are less likely to be added.

3.4. PrOnto compared to similar tools

Using term co-occurence as a proxy for functional similarity has already been done by other methods and tools. For example, the EBI's QuickGO service provides the top 100 most co-occuring terms for the query GO term. However, a way of getting more than those 100 co-occuring terms is not provided, nor is any way of getting terms that do not co-occur (the "dissimilar" terms of PrOnto). As already mentioned, the later can be very useful in studies of protein multifunctionality.

The FAM (Function Association Matrix) which is part of the PFP protein function predictor (Hawkins et al., 2009) uses a very similar approach. However, the FAM scores (available at http://dragon.bio.purdue.edu/FAM/), unlike PrOnto, are asymmetric, meaning that $P(GO1|GO2) \neq P(GO2|GO1)$ which is something that should be taken into account when choosing which tool to use. Depending on the analysis, either symmetric or asymmetric probabilities might be preferred.

In addition, both methods (QuickGO and FAM) provide co-occurrences calculated from the entirety of the UniProt database whereas PrOnto considers a specific species' proteome or interactome. A species-specific measure can indeed be considered an asset depending on the analysis being undertaken since cellular and physiological functions often differ between species, essentially due to tissue-specificity. Term co-occurences are therefore expected to be different across species.

Moreover, neither method provides the user with the ability to input a list of terms and obtain a list of probabilities nor the possibility of querying for specific pairs. The web-based interface of PrOnto is designed with that in mind and is a simple and practical way of quantifying GO term functional dissimilarity.

Finally, PrOnto is, to our knowledge, the only tool that offers a measure of GO term functional dissimilarity based on a species' interactome. A similar approach was undertaken by Dotan-Cohen et al. (2009) to identify "Process Linkage Networks" rather than to assess GO term functional dissimilarity. While they used

interactome data of a single species to build these networks, they did not provide a tool to the community.

4. Conclusion

We have presented PrOnto, a novel tool for quantifying GO term *functional dissimilariy* based on two species-dependent metrics of GO term association derived from either the annotations or the interactome data of the species in question. This tool was developed for the specific goal of identifying moonlighting proteins from interactome data. As such, the emphasis has been on the identification of dissimilar functions.

The current version of PrOnto is using a relatively simple statistical approach which, nevertheless provides robust results (see Chapple et al., 2015 and **Supplementary Figure 1**). In the future, in addition to the hypergeometric probability, power and Jaccard indeces, we plan to implement other statistical measures and combine them into a single metric of GO term similarity. In addition, we plan to expand the tool to also address "similar," as opposed to merely "not dissimilar," GO terms.

Funding

The work was funded by the ANR (Agence Nationale pour la Recherche) as a PIRIbio grant 09- PIRI-0028, Moonlight project, and ITMO Cancer-Plan Cancer (System Biology call, A12171AS) to CB, CC was a postdoctoral fellow of the Fondation pour la Recherche Médicale.

Acknowledgments

Thanks are due to Lionel Spinelli, Emmanuelle Becker, Anaïs Baudot, Elisabeth Remy, Jacques van Helden, and Benoit Robisson for helpful discussions and advice.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fgene.2015.00200/abstract

Supplementary Figure 1 | Density plot of the power of the hypergeometric test for all GO pairs tested. The power is high for the vast majority of pairs tested (mean = 0.90 and median = 0.99).

Supplementary Table 1 | The number of GO pairs for each of the possible sub-ontology combinations. CC, Cellular Compartment; BP, Biological Process; MF, Molecular Function.

Supplementary Table 2 | List of experimental methods identifying binary protein-protein interactions.

References

Aranda, B., Blankenburg, H., Kerrien, S., Brinkman, F. S. L., Ceol, A., Chautard, E., et al. (2011). Psicquic and psiscore: accessing and scoring molecular interactions. *Nat. Methods* 8, 528–529. doi: 10.1038/nmeth.1637

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Bada, M., Turi, D., McEntire, R., and Stevens, R. (2004). Using reasoning to guide annotation with gene ontology terms in goat. *ACM Sigmod Rec.* 33, 27–32. doi: 10.1145/1024694.1024699

Barrell, D., Dimmer, E., Huntley, R. P., Binns, D., O'Donovan, C., and Apweiler, R. (2009). The goa database in 2009–an integrated gene ontology annotation resource. *Nucleic Acids Res.* 37, D396–D403. doi: 10.1093/nar/gkn803

Becker, E., Robisson, B., Chapple, C. E., Guénoche, A., and Brun, C. (2012). Multifunctional proteins revealed by overlapping clustering in protein interaction network. *Bioinformatics* 28, 84–90. doi: 10.1093/bioinformatics/btr621

Ceol, A., Aryamontri, A. C., Licata, L., Peluso, D., Briganti, L., Perfetto, L., et al. (2010). Mint, the molecular interaction database: 2009 update. *Nucleic Acids Res.* 38, D532–D539. doi: 10.1093/nar/gkp983

Chapple, C., Robisson, B., Spinelli, L., Guien, C., Becker, E., and Brun, C. (2015). Extreme multifunctional proteins identified from a human protein interaction network. *Nat. Commun.* 6:7412. doi: 10.1038/ncomms8412

Chatr-Aryamontri, A., Breitkreutz, B.-J., Heinicke, S., Boucher, L., Winter, A., Stark, C., et al. (2013). The biogrid interaction database: 2013 update. *Nucleic Acids Res.* 41, D816–D823. doi: 10.1093/nar/gks1158

Chautard, E., Ballut, L., Thierry-Mieg, N., and Ricard-Blum, S. (2009). Matrixdb, a database focused on extracellular protein-protein and protein-carbohydrate interactions. *Bioinformatics* 25, 690–691. doi: 10.1093/bioinformatics/btp025

Copley, S. D. (2012). Moonlighting is mainstream: paradigm adjustment required. *Bioessays* 34, 578–588. doi: 10.1002/bies.201100191

Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., et al. (2011). Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* 39, D691–D697. doi: 10.1093/nar/gkq1018

Dotan-Cohen, D., Letovsky, S., Melkman, A. A., and Kasif, S. (2009). Biological process linkage networks. *PLoS ONE* 4:e5313. doi: 10.1371/journal.pone.0005313

Dutkowski, J., Kramer, M., Surma, M. A., Balakrishnan, R., Cherry, J. M., Krogan, N. J., et al. (2013). A gene ontology inferred from molecular networks. *Nat. Biotechnol.* 31, 38–45. doi: 10.1038/nbt.2463

Elkon, R., Vesterman, R., Amit, N., Ulitsky, I., Zohar, I., Weisz, M., et al. (2008). Spike–a database, visualization and analysis tool of cellular signaling pathways. *BMC Bioinform.* 9:110. doi: 10.1186/1471-2105-9-110

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). Cd-hit: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565

Gan, M., Dou, X., and Jiang, R. (2013). From ontology to semantic similarity: calculation of ontology-based semantic similarity. *Sci. World J.* 2013:793091. doi: 10.1155/2013/793091

Gaur, U., and Aggarwal, B. B. (2003). Regulation of proliferation, survival and apoptosis by members of the tnf superfamily. *Biochem. Pharmacol.* 66, 1403–1408. doi: 10.1016/S0006-2952(03)00490-8

Hawkins, T., Chitale, M., Luban, S., and Kihara, D. (2009). Pfp: automated prediction of gene ontology functional annotations with confidence scores using protein sequence data. *Proteins* 74, 566–582. doi: 10.1002/prot.22172

Hogan, D. J., Riordan, D. P., Gerber, A. P., Herschlag, D., and Brown, P. O. (2008). Diverse rna-binding proteins interact with functionally related sets of rnas, suggesting an extensive regulatory system. *PLoS Biol.* 6:e255. doi: 10.1371/journal.pbio.0060255

Jeffery, C. J. (2011). Proteins with neomorphic moonlighting functions in disease. *IUBMB Life* 63, 489–494. doi: 10.1002/iub.504

Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., et al. (2012). The intact molecular interaction database in 2012. *Nucleic Acids Res.* 40, D841–D846. doi: 10.1093/nar/gkr1088

Khan, I., Chen, Y., Dong, T., Hong, X., Takeuchi, R., Mori, H., et al. (2014). Genome-scale identification and characterization of moonlighting proteins. *Biol. Dir.* 9, 30. doi: 10.1186/s13062-014-0030-9

Khan, I., and Kihara, D. (2014). Computational characterization of moonlighting proteins. *Biochem. Soc. Trans.*42, 1780–1785. doi: 10.1042/BST20140214

Kramer, M., Dutkowski, J., Yu, M., Bafna, V., and Ideker, T. (2014). Inferring gene ontologies from pairwise similarity data. *Bioinformatics* 30, i34–i42. doi: 10.1093/bioinformatics/btu282

Lange, P. F., and Overall, C. M. (2011). Topfind, a knowledgebase linking protein termini with function. *Nat. Methods* 8, 703–704. doi: 10.1038/nmeth.1669

Lynn, D. J., Winsor, G. L., Chan, C., Richard, N., Laird, M. R., Barsky, A., et al. (2008). Innatedb: facilitating systems-level analyses of the mammalian innate immune response. *Mol. Syst. Biol.* 4, 218. doi: 10.1038/msb.2008.55

Mani, M., Chen, C., Amblee, V., Liu, H., Mathur, T., Zwicke, G., et al. (2015). Moonprot: a database for proteins that are known to moonlight. *Nucleic Acids Res.* 43, D277–D282. doi: 10.1093/nar/gku954

Prieto, C., and Rivas, J. D. L. (2006). Apid: agile protein interaction dataanalyzer. *Nucleic Acids Res.* 34, W298–W302. doi: 10.1093/nar/gkl128

R Core Team. (2012). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*. Vienna.

Salwinski, L., Miller, C. S., Smith, A. J., Pettit, F. K., Bowie, J. U., and Eisenberg, D. (2004). The database of interacting proteins: 2004 update. *Nucleic Acids Res.* 32, D449–D451. doi: 10.1093/nar/gkh086

Schwikowski, B., Uetz, P., and Fields, S. (2000). A network of protein-protein interactions in yeast. *Nat. Biotechnol.* 18, 1257–1261. doi: 10.1038/82360

Skunca, N., Altenhoff, A., and Dessimoz, C. (2012). Quality of computationally inferred gene ontology annotations. *PLoS Comput. Biol.* 8:e1002533. doi: 10.1371/journal.pcbi.1002533

Spirin, V., and Mirny, L. A. (2003). Protein complexes and functional modules in molecular networks. *Proc. Natl. Acad. Sci. U.S.A.* 100, 12123–12128. doi: 10.1073/pnas.2032324100

Zanzoni, A., Chapple, C., and Brun, C. (2015). Relationship between extreme multifunctional proteins, human diseases and comorbidities from a network perspective. *Front. Physiol.* 6:171. doi: 10.3389/fphys.2015.00171

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.