

SOFTWARE

Open Access



Protein aggregation, structural disorder and RNA-binding ability: a new approach for physico-chemical and gene ontology classification of multiple datasets

Petr Klus^{1,2}, Riccardo Delli Ponti^{1,2}, Carmen Maria Livi^{1,2} and Gian Gaetano Tartaglia^{1,2,3*}

Abstract

Background: Comparison between multiple protein datasets requires the choice of an appropriate reference system and a number of variables to describe their differences. Here we introduce an innovative approach to discriminate multiple protein datasets (*multiCM*) and to measure enrichments in gene ontology terms (*cleverGO*) using semantic similarities.

Results: We illustrate the powerfulness of our approach by investigating the links between RNA-binding ability and other protein features, such as structural disorder and aggregation, in *S. cerevisiae*, *C. elegans*, *M. musculus* and *H. sapiens*. Our results are in striking agreement with available experimental evidence and unravel features that are key to understand the mechanisms regulating cellular homeostasis.

Conclusions: In an intuitive way, *multiCM* and *cleverGO* provide accurate classifications of physico-chemical features and annotations of biological processes, molecular functions and cellular components, which is extremely useful for the discovery and characterization of new trends in protein datasets. The *multiCM* and *cleverGO* can be freely accessed on the Web at http://www.tartaglialab.com/cs_multi/submission and http://www.tartaglialab.com/GO_analyser/universal. Each of the pages contains links to the corresponding documentation and tutorial.

Keywords: Protein classification, Physico-chemical properties, Gene ontology, Solubility, RNA-binding ability

Background

There is a growing gap between amount of proteomic data and availability of tools for their analysis [1]. While several application programming interfaces are available to analyse computational and experimental results [2], a simple and intuitive interface is currently lacking or missing. Our goal is to start bridging this gap by providing algorithms for analysis of protein sets and discovery of mechanisms that regulate protein function and interactions.

The first method presented here, the *multiCleverMachine* (*multiCM*), is an extension of the *cleverMachine* approach (CM [3]) to classify multiple protein datasets

using physico-chemical properties. The second algorithm, the *cleverGO*, is inspired by the need to simplify Gene Ontology (GO) annotation output. While GO statistics are important to characterize the functional role of proteins, their interpretation is difficult without further downstream processing [2, 4]. Current tools do not provide a unique interface that combines GO term analysis with intuitive interpretation and visualization. For instance, *GORilla* [5] calculates GO terms enrichments, but other tools are needed to summarize the results (e.g. *REVIGO* [6]). *cleverGO* integrates multiple analyses in one platform and facilitates GO processing through an interactive analysis accessible via web browser.

We demonstrate the usefulness of our methods by investigating the RNA-binding abilities of *S. cerevisiae* chaperones and their substrates, the physico-chemical determinants of protein insolubility in *S. cerevisiae*,

* Correspondence: gian.tartaglia@crg.es

¹Gene Function and Evolution, Centre for Genomic Regulation (CRG), Dr. Aiguader 88, 08003 Barcelona, Spain

²Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain

Full list of author information is available at the end of the article

M. musculus and *H. sapiens*, and the relationship between aggregation and longevity in *C. elegans*. The purpose of our analysis is twofold: to provide examples that can be used as a reference in other studies and to shed light on the link between nucleic-acid binding abilities and protein features, such as structural disorder and aggregation, that are increasingly recognized as key factors for cellular function and homeostasis [7–9].

Implementation

The *multiCM* accepts multiple protein sets in FASTA format. Individual sets are classified as positive or negative for binary comparison (the assignment is only needed to create two groups and does not influence the calculations). In each list, the CM screens physico-chemical properties encoded by protein sequences [3] to identify those that best discriminate positive and negative classes (currently supported physico-chemical properties are: nucleic acid binding propensity, membrane propensity, alpha-helix propensity, aggregation propensity, beta sheet propensity, burial propensity and hydrophobicity, but custom properties can be included, as explained in the online Tutorial). For a detailed description of CM performances, we refer to our previous publication [3].

In each *multiCM* run, the information is compiled together from individual models into a high-level overview:

- The user can glean what trend is detected in the data using different physico-chemical features. The indicators collate 10 predictors for each selected feature and represent their consensus with a colour, akin to a *micro-array* slide (Fig. 1a). The colour of each array-spot represents differential states of enrichment for the dataset pairs and allows easy interpretation of increase, decrease or insufficient signal.

The analysis is not restricted to the consensus information only - a link to a full CM view is provided in the main panel (with details on p-value, cross-validation performances, ROC curves and other statistics). The detail view contains ID number of the CM run providing the ability to use it in creation of a *cleverClassifier* to study new datasets [3], as well as a link to perform Gene Ontology analysis using the second part of our toolkit, the *cleverGO*.

The *cleverGO* webserver provides two ways to explore data:

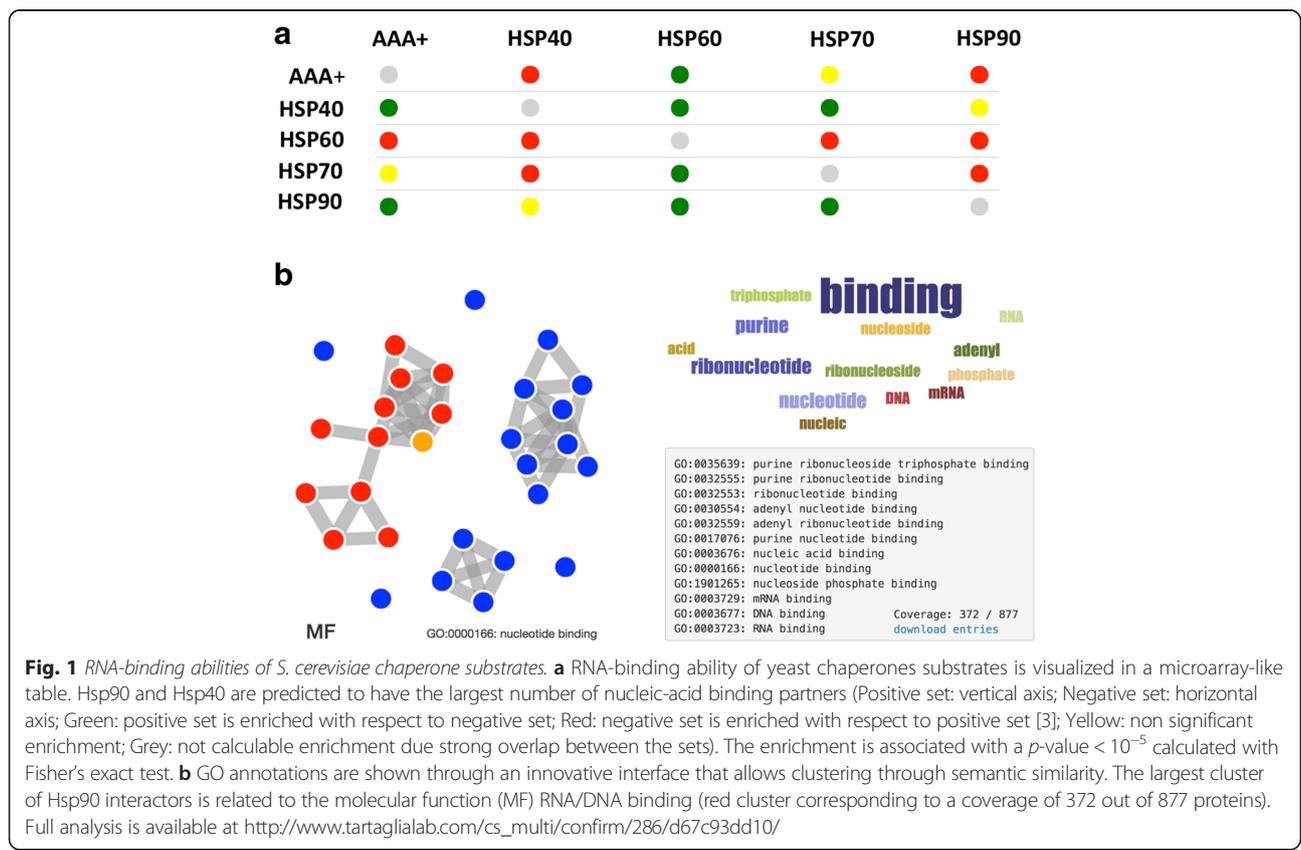


Fig. 1 RNA-binding abilities of *S. cerevisiae* chaperone substrates. **a** RNA-binding ability of yeast chaperones substrates is visualized in a microarray-like table. Hsp90 and Hsp40 are predicted to have the largest number of nucleic-acid binding partners (Positive set: vertical axis; Negative set: horizontal axis; Green: positive set is enriched with respect to negative set; Red: negative set is enriched with respect to positive set [3]; Yellow: non significant enrichment; Grey: not calculable enrichment due strong overlap between the sets). The enrichment is associated with a p -value $< 10^{-5}$ calculated with Fisher's exact test. **b** GO annotations are shown through an innovative interface that allows clustering through semantic similarity. The largest cluster of Hsp90 interactors is related to the molecular function (MF) RNA/DNA binding (red cluster corresponding to a coverage of 372 out of 877 proteins). Full analysis is available at http://www.tartagliolab.com/cs_multi/confirm/286/d67c93dd10/

- The first view of the *cleverGO* tool is a classic enrichment table. Enriched GO terms are showed along with coverage, significance and additional information such as the term depth taken from the acyclic GO graph [4]. The enrichment employs interactive filters - users can match text in the description field, sort by significance or exclude terms based on their term depth or precision [10]. Each GO term is linked to AmiGO [11].

In the second *cleverGO* visualisation, a force-layout is used to dynamically organize the graph depending on the strength of the connections and separate analyses are generated for biological process, molecular function and cellular component ontologies (Fig. 1b). To illustrate relationships between GO terms and to perform functional clustering, we use semantic similarity [12]. The user can interact with the graph: hover over each node with the cursor yields information about the node, clicking it activates an information panel about the cluster the node belongs to (Fig. 1b). For each of the clusters, *cleverGO* shows a list of GO terms that can be individually interrogated, as well as the description of the cluster content. We also provide cluster coverage, i.e. how many of the entries in the user's submission are annotated with GO terms found in the cluster (the list of entries is also available for the user to download). Each of the operations above is based on the current state of the graph - if the signal strength threshold is changed, the graph's connectedness changes. If the user applies the minimal term level or precision cut-off, nodes from the graph are filtered. The same principle applies for the *p*-value cut-off (Bonferroni test). Making the graph behaviour dynamic significantly reduces the time needed to perform analysis - the user does not need to re-run any calculation to see the result of a parameter change.

Additional features:

- Upon activation of the detail view on the *multiCM* output page, the user can access the *Boxplotter*. The *Boxplotter* takes the input datasets with best-performing features (passed automatically from the detail view) and shows the distribution of associated propensity scores. On top of the physico-chemical scale information, the *Boxplotter* matches protein IDs with protein abundance databases [13] to provide information on the distribution of expression values. In addition, the *Boxplotter* performs discrimination analysis with the data, showing *p*-values for the statistics and Receiver Operating Characteristic (ROC) curves.

Results and discussion

To illustrate the performances of both *multiCM* and *cleverGO*, we studied the RNA-binding abilities of *S.*

cerevisiae chaperone substrates [14], the physico-chemical determinants of protein insolubility in *S. cerevisiae*, *M. musculus* and *H. sapiens* [15], and the link between protein aggregation and longevity in *C. elegans* [16].

RNA-binding abilities of *S. cerevisiae* chaperone substrates

Systematic analysis of physical TAP-tag based protein-protein interactions revealed individual networks of *S. cerevisiae* chaperones [14]. In agreement with experimental evidence, the *multiCM* predicts that Hsp90 (Hsp82) [17] and Hsp40 (Cwc23) [18] are prone to associate with RNA-binding proteins (RBPs; Fig. 1a; red dots indicate enrichment over other chaperones). By contrast, Hsp60 shows the lowest propensity to interact with RBPs, which is consistent with its main role of guiding hydrophobic proteins to fold into the native state [19] (Fig. 1a; green dots indicate depletion over other chaperones). Moreover, Hsp70 (Ssb1) binds directly with transcripts and is predicted to have more RBP partners than Hsp60 [20]. AAA+ (Hsp78) shows similar pattern as Hsp70, in agreement with the fact that the two chaperones work together [21]. As for other physico-chemical features, *multiCM* reports that both Hsp40 and Hsp78 associate with structurally disordered (and hydrophilic [22]) proteins, which is in line with previous experimental studies on prion propagation [23], while Hsp60, Hsp70 and Hsp90 are predicted to bind to hydrophobic proteins [3, 19]. To further investigate Hsp90 features, we performed *cleverGO* analysis of its substrates. Looking at the molecular function (Fig. 1b), we observe an enrichment in GO terms related to RBPs (e.g., class "RNA-binding" shows *p*-value < 10⁻⁵; Bonferroni test), which very well complements our predictions of physico-chemical features. Importantly, the nucleic-acid cluster is the largest in terms of dataset coverage (>40% of the substrates list; Fig. 1b).

Physico-chemical determinants of protein insolubility

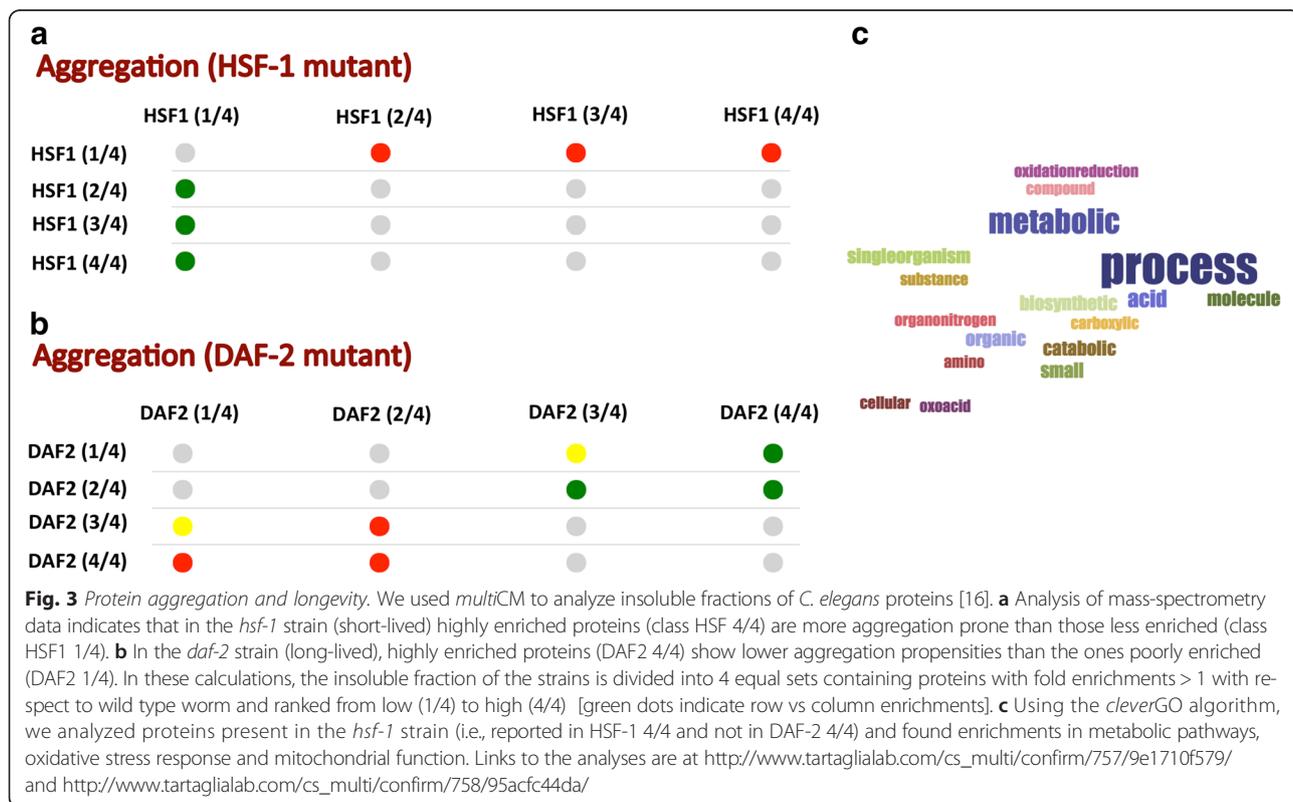
A recent mass-spectrometry study investigated protein precipitates formed upon centrifugation of *S. cerevisiae*, *M. musculus* and *H. sapiens* cells [15]. Two major determinants have been reported to promote insolubility: structural disorder in *H. sapiens* and *M. musculus*, which induces aberrant interactions promoting precipitation of protein complexes [24], and aggregation propensity [25] in *S. cerevisiae* cells, which is linked to the presence of hydrophobic residues exposed on protein surfaces [22]. Using the *multiCM* approach to compare low-solubility (LS) and high-solubility (HS) proteins, we observed that *H. sapiens* and *M. musculus* have a larger fraction of structurally disordered regions in the LS group, while non-significant enrichments were found in yeast (Fig. 2a). Differently from *H. sapiens* and *M. musculus* cells, *S. cerevisiae* shows high intrinsic aggregation propensity

extension of *C. elegans* strains carrying mutation in the *daf-2* receptor and that transcription factor *hsf-1* is essential for longevity [16]. Mass-spectrometry analysis of long-lived *daf-2* and short-lived *hsf-1* mutant strains revealed two major types of deposits that accumulate during aging: *hsf-1* mutant proteins have high aggregation propensities, while *daf-2* mutant proteins show decreased structural content [16]. Thus, decrease in longevity can be associated with accumulation of aggregation-prone proteins, whereas lower hydrophobicity is linked to different type of deposits and significantly reduced toxicity. Using the *multiCM* approach to compare the insoluble fraction of *hsf-1* mutant strain with wild type worm (WT), we found that proteins showing high enrichment in mass-spectrometry analysis (class HSF-1 4/4) are more aggregation-prone than those with low enrichment (class HSF-1 1/4) [Fig. 3a]. By contrast, proteins enriched in *daf-2* mutant worms (DAF-2 4/4) have lower aggregation propensities than those showing low enrichment (DAF-2 1/4). In the *daf-2* mutant strain (DAF-2 3/4 and DAF-2 4/4) enrichments are associated with decrease in beta-sheet content (Additional file 1: Figure S2A), while in *hsf-1* mutant worms (HSF-1 3/4 and HSF-1 4/4) we observe depletion of structural disorder (Additional file 1: Figure S2B). Proteins present in the *hsf-1* strain (i.e., listed in HSF-1

4/4 and not included in DAF-2 4/4) are involved in several metabolic processes (e.g., class “oxidative stress response” with p -value $< 6 \cdot 10^{-4}$; Bonferroni test), oxidative stress response (e.g., class “metabolic process” shows p -value $< 10^{-7}$) and mitochondrial function (e.g., class “mitochondrion” with p -value $< 10^{-7}$), as reported in the original study (Fig. 3c) [16]. In addition, and in line with the work on *S. cerevisiae*, *M. musculus* and *H. sapiens* proteomes [15], we found an enrichment of RBPs (e.g., class “RNA-binding” shows p -value $= 7 \cdot 10^{-3}$), which reinforces the link between protein deposition and nucleic acid binding [34].

Conclusions

In this work, we introduced two innovative approaches to compare multiple protein datasets using physico-chemical properties and GO annotations: the *multiCM* allows feature classification and the *cleverGO* provides clustering through semantic relationships. We illustrated the performances of both *multiCM* and *cleverGO* using examples related to RNA-binding abilities of *S. cerevisiae* chaperone substrates [14], physico-chemical determinants of protein insolubility in *S. cerevisiae*, *M. musculus* and *H. sapiens* [15] and the link between aggregation and life-span in *C. elegans* [16]. In all cases, the results are in agreement with available evidence on protein functions and



interactions, providing a clear indication on the flexibility and broad applicability of our algorithms.

As shown in the examples, we are particularly interested in understanding the relationship between nucleic-acid binding ability and structural disorder and aggregation. Indeed, previous studies indicate that RNA secondary structures [35], especially when enriched in GC content [36], contribute to spatial rearrangement of disordered regions, promoting the formation of protein-RNA complexes. In agreement with these observations, it has been reported that intrinsically disordered proteins interact with RNA [8, 37], which influences protein aggregation [38] and, in turn, toxicity [39]. The involvement of nucleic acid molecules in protein aggregation [40] is compatible with the findings discussed in our examples and provides an intriguing working hypothesis [7, 41] to study neurodegenerative events [42] that are characterized by aggregation [43] and structural disorder [44]. As a matter of fact, previous work indicates that presence of polyanions lead to reduction of protein stability [45] and nucleic acids have a strong tendency to accumulate in neurofibrillary tangles and senile plaques [46]. Recent evidence also shows that aggregation-related mutations in the RBPs Tar DNA-binding protein 43 TDP-43 and Translocated in liposarcoma protein FUS are associated with the formation of RNA granules [47, 48] that are phase separated, non-membrane-bound ribonucleoprotein aggregates [49, 50].

In conclusion, theoretical approaches for prediction of protein features, such as those integrated in the *multiCM* for prediction of structural disorder, aggregation and nucleic-acid binding ability [51–53], will be useful to provide insights into functional networks. We hope that our tools will be useful for the discovery of trends in protein datasets, complementing experimental [54, 55] and theoretical analyses [31, 56–58].

Availability and requirements

The *multiCM* and *cleverGO* are available at http://www.tartagliolab.com/cs_multi/submission and http://www.tartagliolab.com/GO_analyser/universal.

Tutorials can be accessed at http://www.tartagliolab.com/cs_multi/tutorial and http://www.tartagliolab.com/GO_analyser/tutorial. Documentation files are deposited at http://service.tartagliolab.com/static_files/algorithms/clever_machine/documentation.html.

Additional file

Additional file 1: Figure S1. Physico-chemical determinants of protein insolubility. High-solubility (HS) proteins show A) higher burial in human and mouse, in agreement with the observations reported in the original study. **Figure S2.** Physico-chemical of *C. elegans* mutant strains. A) In the *hsf-1* strain, highly enriched proteins (HSF 4/4) are less structurally

disordered than those poorly enriched (HSF1 1/4). B) In the *daf-2* strain (long-lived), highly enriched proteins (DAF2 4/4) show lower beta-sheet propensities than those poorly enriched (DAF2 1/4), in agreement with observations reported in the original experimental study. (DOCX 412 kb)

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

PK implemented the webserver. PK and GGT designed the algorithm. CM and RDP tested the method. CM, RDP, PK and GGT wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank B. Lehner, Gianni de Fabritiis and Roderic Guigó for stimulating discussions.

Funding

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013), through the European Research Council, under grant agreement RIBOMYLOME_309545 (Gian Gaetano Tartaglia), and from the Spanish Ministry of Economy and Competitiveness (BFU2014-55054-P). We also acknowledge support from AGAUR (2014 SGR 00685), the Spanish Ministry of Economy and Competitiveness, 'Centro de Excelencia Severo Ochoa 2013–2017' (SEV-2012-0208). PK and RDP are recipients of "La Caixa" and "Severo Ochoa" studentships, respectively.

Author details

¹Gene Function and Evolution, Centre for Genomic Regulation (CRG), Dr. Aiguader 88, 08003 Barcelona, Spain. ²Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain. ³Institució Catalana de Recerca i Estudis Avançats (ICREA), 23 Passeig Lluís Companys, 08010 Barcelona, Spain.

Received: 28 May 2015 Accepted: 8 December 2015

Published online: 16 December 2015

References

- Vizcaíno JA, Côté RG, Csordas A, Dianas JA, Fabregat A, Foster JM, et al. The Proteomics Identifications (PRIDE) database and associated tools: status in 2013. *Nucl Acids Res.* 2013;41:D1063–9.
- Mi H, Dong Q, Muruganujan A, Gaudet P, Lewis S, Thomas PD. PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res.* 2010;38(Database issue):D204–10.
- Klus P, Bolognesi B, Agostini F, Marchese D, Zanzoni A, Tartaglia GG. The cleverSuite Approach for Protein Characterization: Predictions of Structural Properties, Solubility, Chaperone Requirements and RNA-Binding Abilities. *Bioinformatics.* 2014;30(11):1601–8.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000;25:25–9.
- Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics.* 2009;10:48.
- Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE.* 2011;6:e21800.
- Wolozin B. Regulated protein aggregation: stress granules and neurodegeneration. *Molecular Neurodegeneration.* 2012;7:56.
- Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, et al. Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell.* 2012;149:1393–406.
- Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. Methods of integrating data to uncover genotype-phenotype interactions. *Nat Rev Genet.* 2015;16:85–97.
- Herrmann C, Bérard S, Tichit L. SimCT: a generic tool to visualize ontology-based relationships for biological objects. *Bioinformatics.* 2009;25:3197–8.
- Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, et al. AmiGO: online access to ontology and annotation data. *Bioinformatics.* 2009;25:288–9.

12. Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics*. 2010;26:976–8.
13. Wang M, Weiss M, Simonovic M, Haertinger G, Schrimpf SP, Hengartner MO, et al. PaxDb, a database of protein abundance averages across all three domains of life. *Mol Cell Proteomics*. 2012;11(8):492–500.
14. Gong Y, Kakiyama Y, Krogan N, Greenblatt J, Emili A, Zhang Z, et al. An atlas of chaperone-protein interactions in *Saccharomyces cerevisiae*: implications to protein folding pathways in the cell. *Mol Syst Biol*. 2009;5:275.
15. Albu RF, Chan GT, Zhu M, Wong ETC, Taghizadeh F, Hu X, et al. A feature analysis of lower solubility proteins in three eukaryotic systems. *J Proteomics*. 2015;118:21–38.
16. Walther DM, Kasturi P, Zheng M, Pinkert S, Vecchi G, Ciryam P, et al. Widespread Proteome Remodeling and Aggregation in Aging *C. elegans*. *Cell*. 2015;161:919–32.
17. Sawarkar R, Sievers C, Paro R. Hsp90 globally targets paused RNA polymerase to regulate gene expression in response to environmental stimuli. *Cell*. 2012;149:807–18.
18. Ohi MD, Link AJ, Ren L, Jennings JL, McDonald WH, Gould KL. Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel pre-mRNA splicing factors, and snRNAs. *Mol Cell Biol*. 2002;22:2011–24.
19. Tartaglia GG, Dobson CM, Hartl FU, Vendruscolo M. Physicochemical determinants of chaperone requirements. *Journal of Molecular Biology*. 2010;400:579–88.
20. Zimmer C, von Gabain A, Henics T. Analysis of sequence-specific binding of RNA to Hsp70 and its various homologs indicates the involvement of N- and C-terminal interactions. *RNA*. 2001;7:1628–37.
21. von Janowsky B, Major T, Knapp K, Voos W. The disaggregation activity of the mitochondrial ClpB homolog Hsp78 maintains Hsp70 function during heat stress. *J Mol Biol*. 2006;357:793–807.
22. Tartaglia GG, Caflich A. Computational analysis of the *S. cerevisiae* proteome reveals the function and cellular localization of the least and most amyloidogenic proteins. *Proteins*. 2007;68:273–8.
23. Doyle SM, Genest O, Wickner S. Protein rescue from aggregates by powerful molecular chaperone machines. *Nat Rev Mol Cell Biol*. 2013;14:617–29.
24. Babu MM, van der Lee R, de Groot NS, Gsponer J. Intrinsically disordered proteins: regulation and disease. *Curr Opin Struct Biol*. 2011;21:432–40.
25. Tartaglia GG, Vendruscolo M. The Zyggregator method for predicting protein aggregation propensities. *Chem Soc Rev*. 2008;37:1395–401.
26. Fernandez-Escamilla A-M, Rousseau F, Schymkowitz J, Serrano L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol*. 2004;22:1302–6.
27. Conchillo-Solé O, de Groot NS, Avilés FX, Vendrell J, Daura X, Ventura S. AGGRESKAN: a server for the prediction and evaluation of “hot spots” of aggregation in polypeptides. *BMC Bioinformatics*. 2007;8:65.
28. Tartaglia GG, Pawar AP, Campioni S, Dobson CM, Chiti F, Vendruscolo M. Prediction of aggregation-prone regions in structured proteins. *J Mol Biol*. 2008;380:425–36.
29. Bolognesi B, Tartaglia GG. Physicochemical principles of protein aggregation. *Prog Mol Biol Transl Sci*. 2013;117:53–72.
30. Tartaglia GG, Pechmann S, Dobson CM, Vendruscolo M. Life on the edge: a link between gene expression levels and aggregation rates of human proteins. *Trends Biochem Sci*. 2007;32:204–6.
31. Tartaglia GG, Vendruscolo M. Correlation between mRNA expression levels and protein aggregation propensities in subcellular localisations. *Mol Biosyst*. 2009;5:1873–6.
32. Ciryam P, Tartaglia GG, Morimoto RI, Dobson CM, Vendruscolo M. Widespread aggregation and neurodegenerative diseases are associated with supersaturated proteins. *Cell Rep*. 2013;5:781–90.
33. Tsai M-C, Manor O, Wan Y, Mosammamaparast N, Wang JK, Lan F, et al. Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. *Science*. 2010;329:689–93.
34. Cirillo D, Livi CM, Agostini F, Tartaglia GG. Discovery of protein-RNA networks. *Mol Biosyst*. 2014;10:1632–42.
35. Gray DA, Woulfe J. Structural disorder and the loss of RNA homeostasis in aging and neurodegenerative disease. *Front Genet*. 2013;4:149.
36. Zanzoni A, Marchese D, Agostini F, Bolognesi B, Cirillo D, Botta-Orfila M, et al. Principles of self-organization in biological pathways: a hypothesis on the autogenous association of alpha-synuclein. *Nucl Acids Res*. 2013;41(22):9987–98.
37. Cirillo D, Marchese D, Agostini F, Livi CM, Botta-Orfila T, Tartaglia GG. Constitutive patterns of gene expression regulated by RNA-binding proteins. *Genome Biol*. 2014;15:R13.
38. Olzscha H, Schermann SM, Woerner AC, Pinkert S, Hecht MH, Tartaglia GG, et al. Amyloid-like Aggregates Sequester Numerous Metastable Proteins with Essential Cellular Functions. *Cell*. 2011;144:67–78.
39. Vavouri T, Semple JJ, Garcia-Verdugo R, Lehner B. Intrinsic Protein Disorder and Interaction Promiscuity Are Widely Associated with Dosage Sensitivity. *Cell*. 2009;138:198–208.
40. Kampers T, Friedhoff P, Biernat J, Mandelkow EM, Mandelkow E. RNA stimulates aggregation of microtubule-associated protein tau into Alzheimer-like paired helical filaments. *FEBS Lett*. 1996;399:344–9.
41. Papatriantafyllou M. Protein aggregation: The secret recipe for RNA granules. *Nat Rev Mol Cell Biol*. 2012;13:405.
42. Cirillo D, Agostini F, Klus P, Marchese D, Rodriguez S, Bolognesi B, et al. Neurodegenerative diseases: Quantitative predictions of protein-RNA interactions. *RNA*. 2013;19:129–40.
43. Vendruscolo M, Tartaglia GG. Towards quantitative predictions in cell biology using chemical properties of proteins. *Mol Biosyst*. 2008;4:1170–5.
44. Porcari R, Proukakis C, Waudby CA, Bolognesi B, Mangione PP, Paton JF, et al. The H50Q Mutation Induces a 10-fold Decrease in the Solubility of α -Synuclein. *J Biol Chem*. 2015;290:2395–404.
45. Sedláček E, Fedunová D, Veselá V, Sedláčková D, Antalík M. Polyanion hydrophobicity and protein basicity affect protein stability in protein-polyanion complexes. *Biomacromolecules*. 2009;10:2533–8.
46. Ginsberg SD, Crino PB, Lee VM, Eberwine JH, Trojanowski JQ. Sequestration of RNA in Alzheimer’s disease neurofibrillary tangles and senile plaques. *Ann Neurol*. 1997;41:200–9.
47. Bentmann E, Haass C, Dormann D. Stress granules in neurodegeneration – lessons learnt from TAR DNA binding protein of 43 kDa and fused in sarcoma. *FEBS J*. 2013;280:4348–70.
48. Baron DM, Kaushansky LJ, Ward CL, Sama RRR, Chian R-J, Boggio KJ, et al. Amyotrophic lateral sclerosis-linked FUS/TLS alters stress granule assembly and dynamics. *Mol Neurodegener*. 2013;8:30.
49. Malinowska L, Kroschwald S, Alberti S. Protein disorder, prion propensities, and self-organizing macromolecular collectives. *Biochim Biophys Acta*. 1834:2013:918–31.
50. Ramaswami M, Taylor JP, Parker R. Altered Ribostasis: RNA-Protein Granules in Degenerative Disorders. *Cell*. 2013;154:727–36.
51. Tartaglia GG, Cavalli A, Pellarin R, Caflich A. Prediction of aggregation rate and aggregation-prone segments in polypeptide sequences. *Protein Sci*. 2005;14:2723–34.
52. Agostini F, Vendruscolo M, Tartaglia GG. Sequence-based prediction of protein solubility. *J Mol Biol*. 2012;421:237–41.
53. Terribilini M, Lee J-H, Yan C, Jernigan RL, Honavar V, Dobbs D. Prediction of RNA binding sites in proteins from amino acid sequence. *RNA*. 2006;12:1450–62.
54. Calloni G, Chen T, Schermann SM, Chang H, Genevax P, Agostini F, et al. DnaK Functions as a Central Hub in the *E. coli* Chaperone Network. *Cell Reports*. 2012;1:251–64.
55. Mossuto MF, Bolognesi B, Guixer B, Dhulesia A, Agostini F, Kumita JR, et al. Disulfide Bonds Reduce the Toxicity of the Amyloid Fibrils Formed by an Extracellular Protein. *Angew Chem Int Ed Engl*. 2011;50:7048–51.
56. Agostini F, Zanzoni A, Klus P, Marchese D, Cirillo D, Tartaglia GG. catRAPID omics: a web server for large-scale prediction of protein-RNA interactions. *Bioinformatics*. 2013;29:2928–30.
57. Livi CM, Klus P, Delli Ponti R, Tartaglia GG. catRAPID signature: identification of ribonucleoproteins and RNA-binding regions. *Bioinformatics*. 2015 Oct 31. pii: btv629. [Epub ahead of print].
58. Klus P, Cirillo D, Botta Orfila T, Tartaglia GG. Neurodegeneration and Cancer: Where the Disorder Prevails. *Sci Rep*. 2015 Oct 23;5:15390. doi: 10.1038/srep15390.