Article

Prediction and Ranking of Biomarkers Using *multiple* UniReD

Ismini Baltsavia [1], Theodosios Theodosiou [1,2], Nikolas Papanikolaou [3], Georgios A. Pavlopoulos [4], Grigorios D. Amoutzias [5], Maria Panagopoulou [2], Ekaterini Chatzaki [2,6], Evangelos Andreakos [7] and Ioannis Iliopoulos [1,*]

1   Department of Basic Sciences, School of Medicine, University of Crete, 71003 Heraklion, Greece
2   Laboratory of Pharmacology, Medical School, Democritus University of Thrace,
    68100 Alexandroupolis, Greece
3   EnzyQuest PC, Science and Technology Park of Crete, 100 Nikolaou Plastira Str., Vassilika Vouton,
    70013 Heraklion, Greece
4   Institute for Fundamental Biomedical Research, Biomedical Sciences Research Center "Alexander Fleming",
    16672 Vari, Greece
5   Bioinformatics Laboratory, Department of Biochemistry and Biotechnology, University of Thessaly,
    41500 Larisa, Greece
6   Institute of Agri-Food and Life Sciences, University Research Centre, Hellenic Mediterranean University,
    71410 Heraklion, Greece
7   Laboratory of Immunobiology, Center for Clinical, Experimental Surgery and Translational Research,
    Biomedical Research Foundation of the Academy of Athens, 11527 Athens, Greece
*   Correspondence: ioannis@uoc.gr

Abstract: Protein–protein interactions (PPIs) are of key importance for understanding how cells and organisms function. Thus, in recent decades, many approaches have been developed for the identification and discovery of such interactions. These approaches addressed the problem of PPI identification either by an experimental point of view or by a computational one. Here, we present an updated version of UniReD, a computational prediction tool which takes advantage of biomedical literature aiming to extract documented, already published protein associations and predict undocumented ones. The usefulness of this computational tool has been previously evaluated by experimentally validating predicted interactions and by benchmarking it against public databases of experimentally validated PPIs. In its updated form, UniReD allows the user to provide a list of proteins of known implication in, e.g., a particular disease, as well as another list of proteins that are potentially associated with the proteins of the first list. UniReD then automatically analyzes both lists and ranks the proteins of the second list by their association with the proteins of the first list, thus serving as a potential biomarker discovery/validation tool.

Keywords: biomarker validation and ranking; protein–protein interaction prediction

1. Introduction

Protein–protein interactions (PPIs) play an important role in the proper function of living organisms. Proteins can interact directly in a binary form or as members of a complex, or can be functionally related without physical interaction (e.g., when they are involved in the same biological pathway). Due to the importance of PPIs in understanding how complex biological networks function, in the past 30 years, high-throughput technologies such as Y2H, protein arrays, co-immunoprecipitation, mass spectrometry and others [1] have been used for the generation and understanding of PPI networks. While these large-scale experimental approaches come with many advantages, they often become time-consuming, are expensive to run and report findings with high false positive rates, low accuracy and low reproducibility. On the other hand, in silico PPI prediction approaches are often faster and of low cost and can be roughly summarized in two categories. In the first category, computational approaches are dedicated to analyzing the genome context and

suggest interactions based on information related to structural and evolutionary features of genomes (phylogenetic profiling [2], gene fusion detection [3] and conservation of gene clusters on chromosomes of different species [4]. In the second category, computational methods utilize biomedical literature mining to exploit the information hidden in millions of abstracts/articles stored in various biomedical-related databases such as Medline [1].

Text mining (most particularly biomedical literature mining) methods have been utilized in a variety of challenging problems in biomedical research such as concept discovery in biomedical literature (BioTextQuset+) [5], drug discovery (DrugQuest) [6], disease–gene associations from biomedical abstracts (DISEASES) [7] and biomedical term co-occurrence (CoPub) [8]. The Darling application, for example [9], mines disease-related databases such as OMIM, DisGeNET and Human Phenotype Ontology (HPO). OnTheFly [10] performs named entity recognition in texts and spreadsheets. EXTRACT [11] finds associations between biomedical terms in web pages and consequently the literature.

In the last 20 years, many methods have been developed for PPI extraction from biomedical literature [1]. In particular, PubGene can detect associations between genes/proteins using terms from the MeSH index (Medical Subject Headings) as well as terms from the gene ontology (GO) database [12], HIPPIE provides human PPI networks through PPI network scoring, integration of experimental information and basic graph algorithms [13], iHOP [14] uses genes and proteins as hyperlinks between sentences and abstracts derived from PubMed to provide a protein association network, etc.

In addition to the aforementioned methods, there are databases/tools such as STRING that integrate information from various sources (literature, experiments, databases and genome context) in order to provide protein associations from a large number of species [15] as well as a large number of manually curated databases that contain experimentally validated PPIs, either extracted from the literature or from repositories of large-scale omics experiments [1] (Papanikolaou et al., 2015) (e.g., IntAct [16], DIP [17], MINT [18]).

UniReD is a computational tool that is used to predict functional associations between proteins based on the Markov cluster algorithm (MCL). UniReD parses the references section of UniProt entries for *Homo sapiens* and *Mus musculus* organisms and by using the "similar articles" feature of PubMed, it creates clusters of functionally associated proteins [19]. UniReD performs quite well after following benchmarks against public databases of experimentally verified PPIs and has been used by experimental biologists to identify known and undocumented interactors of protein(s) of interest [19–21]. In addition, UniReD has been used previously [22,23] for the ranking of biomarkers extracted from high-throughput datasets. The whole process was performed manually, and the user had to analyze two lists of genes/proteins, using UniReD results. The first one was a list of proteins/genes known to be associated with the disease/pathway of interest, whereas the second one was a list of proteins/genes that were extracted from high-throughput datasets and had to be validated for relatedness against the first list of genes. The whole process was laborious since the user had to run UniReD for each pair of proteins (see Materials and Methods section for details).
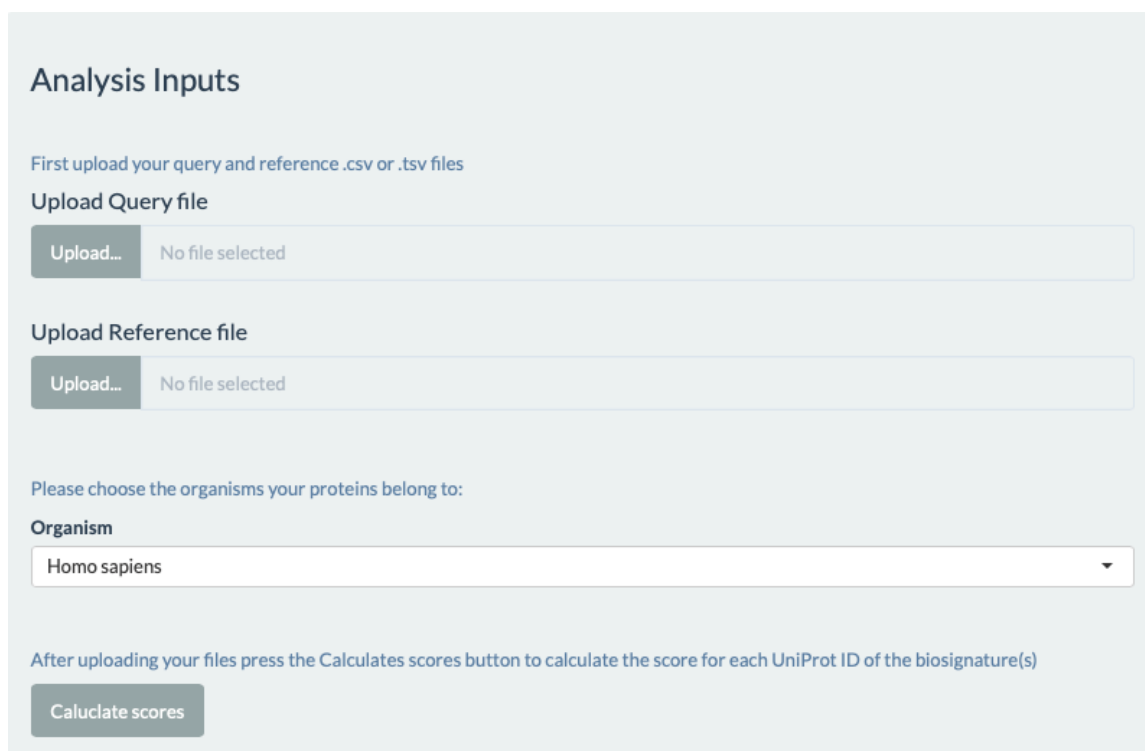
Herein, we present a new powerful feature of the UniReD tool (*multiple* UniReD) which automates the aforementioned process, in an effortless way. Users may provide the two lists of proteins of interest and *multiple* UniReD will return the validated results ranked within a few seconds. This functionality is of particular usefulness in biomarker discovery/validation pipelines, as demonstrated by a test case. UniReD is available at http://bioinformatics.med.uoc.gr/shiny/multiunired/, accessed on 16 September 2022.

2. Results and Discussion

We have developed a user friendly and self-explanatory web interface for *multiple* UniReD to assist wet lab biologists with limited to no programming skills.

The multiple search feature of UniReD can be accessed from the main page of UniReD's web interface and users are redirected to the new dedicated interface for this pipeline. By choosing the analysis button, the user now can view the main interface page of *multiple* UniReD analysis. In order to run the analysis, a two-step procedure has to be followed.

Firstly, two separate lists of proteins have to be uploaded, in a comma- or tab-separated format (.csv or .tsv files). The first list (query file) represents proteins whose functional associations the user wants to explore in relation to the second list of proteins (reference file) that are selected by the user and are related to the research topic under investigation (e.g., a specific disease). When uploaded, an overview of these lists is displayed. Then, the organism from which these proteins derive has to be defined by the user. In its present form, UniReD (as well as *multiple* UniReD) can analyze human or mouse proteins but we are planning to expand the analysis to other species as well (Figure 1).

Figure 1. The input files required for the analysis using the web interface. The query input requires UniProt accession IDs for the proteins under investigation. The reference input requires UniProt accession IDs as well for the proteins known to be implicated in a specific biological process or disease. The organism input refers to the organism that the UniProt accession IDs belong to. The "Calculate scores" button calculates the scores for each of the UniProt accession IDs or the query.

In order to test the *multiple* UniReD tool, we selected two gene lists that we used in a previous work [22], in which functional associations were searched manually one by one, using UniReD's web interface. The reference list contains proteins whose implication in breast cancer is well documented in the literature [22], whereas the query list contains proteins that were proposed as biosignatures for breast cancer by a machine learning approach. It should be mentioned that the latter analysis was conducted using the UniProt version of March 2017, whereas for the new analysis, we used the updated version of UniProt (October 2021). This may be a reason why the results of the analysis are slightly different when compared to those of [22], because the two versions of the UniProt database vary and the difference has an impact on UniReD results.

The analysis via *multiple* UniReD is conducted within a few seconds and the results are available in three distinct tables. The first one reports the scores for each pairwise association between proteins of the two lists, accompanied by the overall score for each one of the proteins under investigation. The overall score is reported in the last column, shown in a descending order. In this way, the user can view a ranking of the proteins of interest sorted by their relatedness to the proteins in the reference list. UniProt accession numbers and gene names in parentheses are provided as identifiers for each protein (Figure 2).

Signature scores

Show 20 entries   Search:

	P12830 (CDH1)	P42336 (PIK3CA)	Q9HCU9 (BRMS1)	P60484 (PTEN)	P03372 (ESR1)	P04637 (TP53)	P51587 (BRCA2)	Q9NS23 (RASSF1)	P38398 (BRCA1)	Q15828 (CST6)	Overall_Score
P35625 (TIMP3)	1	1	1	1	1	1	1	0.5	1	1	9.5
P09619 (PDGFRB)	1	1	0.5	1	1	1	1	0.5	1	0.5	8.5
Q9Y5T5 (USP16)	1	1	0.5	1	1	1	1	0.5	0.5	0.5	8
P50570 (DNM2)	1	1	0	1	1	1	1	0.5	1	0	7.5
O14862 (AIM2)	1	1	0	1	1	1	0.5	0.5	1	0	7
Q8WYL5 (SSH1)	1	1	0	1	1	1	1	0	0.5	0.5	7
Q9H4L7 (SMARCAD1)	0.5	0.5	0.5	0.5	1	1	1	0.5	1	0.5	7
Q7Z628 (NET1)	1	1	0	1	1	1	0.5	0.5	0.5	0.5	7
Q96CW1 (AP2M1)	1	1	0	1	1	1	0.5	0	0	0	5.5
Q8NFR9 (IL17RE)	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0	0	0	3.5
Q9H7B2 (RPF2)	0	0	0	1	0	1	0	0	0	0	2
Q9BZH6 (WDR11)	0	0	0	0	0	0	0	0	0	0	0

Showing 1 to 12 of 12 entries   Copy Excel CSV   Previous 1 Next

Figure 2. *multiple* UniReD output of the breast cancer biomarker analysis. The scoring table for each UniProt accession ID. The rows correspond to the query UniProt accession IDs whereas the columns to the reference UniProt accession IDs. Inside the parentheses is the corresponding gene name. The last column contains the overall (sum) score for each of the query UniProt accession IDs.

The second table describes every pairwise association type for which a score has been assigned. In this table, the first column corresponds to the query proteins whereas the second one holds the reference proteins. In the third column, the user can view the type of association between the query and the reference protein that *multiple* UniReD was able to record. In total, four different cases can be recorded. *InCluster* specifies that both query and reference proteins coexist in a protein cluster of UniReD, i.e., a functional association has been derived from the literature. *Paralogue* states if a paralogue of the query protein has been found in the same cluster with a reference protein. *Orthologue* applies when an association has been found between the orthologues of the query and reference proteins in a different species. In the case where another protein component of a complex that the query protein is part of is associated with a reference protein, it is labeled *Complex* in the third column. The fourth column is used to display the pair of proteins found to be associated. For example, if a paralogue of the query protein has been found to be related to a reference protein, the UniProt accession number and gene name in parentheses of this paralogue are provided. In the case of an orthologue search, the pair of orthologues to the query and reference proteins is shown.

Proteins that are not part of any protein cluster formed by UniReD are excluded from the analysis and become entries of the third table. Tables are available for download.

3. Materials and Methods

3.1. Workflow and Scoring Scheme of multiple UniReD

UniReD is a computational tool used to predict functional associations between proteins based on a machine learning algorithm called MCL [19]. The workflow and the scoring scheme of the new version of UniReD, called *multiple* UniReD, are summarized below and presented analytically in Figure 3.
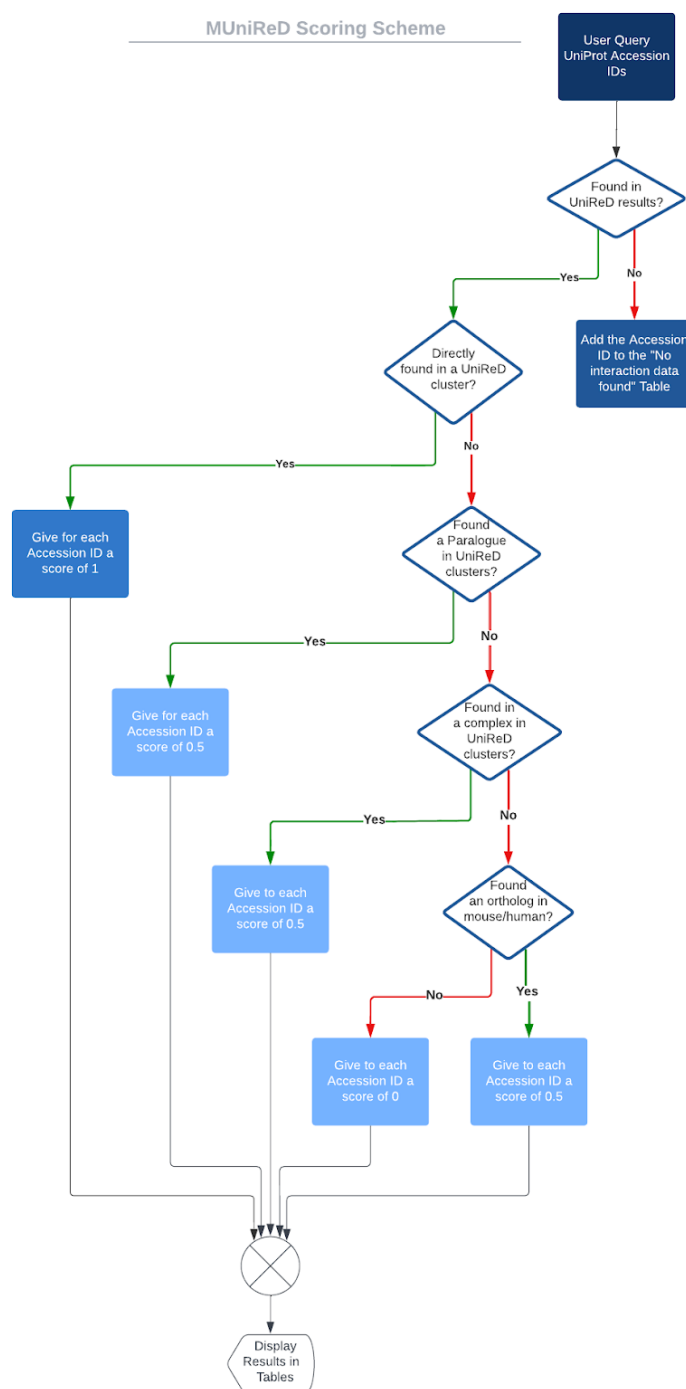
Figure 3. Workflow and scoring scheme of *multiple* UniReD.

The user uploads two lists in .csv or .tsv format. These are: (*i*) a reference list (of proteins that are known to be implicated in a specific disease or biological pathway) and (*ii*) a query list (list of proteins whose associations with the proteins of the reference list the user would like to identify). Proteins in each input list should be in comma- or tab-separated format using UniProt/SwissProt accession numbers.

The user must select the relevant species to be analyzed (currently *Homo sapiens* or *Mus musculus*).

multiple UniReD searches for the existence of any association in UniReD clusters between the protein pairs of the two lists (reference and query list). We followed the same scoring scheme as in [22,23]. If the pair is present in a UniReD cluster, the association will obtain a score of 1 and UniReD will continue to the next pair of reference and query

proteins. If a protein has not been analyzed by UniReD, no results will be retrieved. These proteins will be reported in a separate table upon analysis completion.

If *multiple* UniReD cannot find any association within the clusters, it will proceed to the next step in order to search for paralogues. More specifically, *multiple* UniReD will search for a co-occurrence of a paralogue of the query protein with the reference protein in a UniReD cluster. If such a pair is present, the association will obtain a score of 0.5.

In the case that UniReD does not predict any association within a query paralogue and the reference protein, then *multiple* UniReD will investigate whether the query protein is part of a complex. In this case, it will search for co-occurrences of the reference protein in the complex. If such an association is confirmed, the association will obtain a score of 0.5.

Finally, if none of the above applies, UniReD will search for a relation between the orthologues of reference and query proteins. If such an association is documented, it will obtain a score of 0.5.

When none of the aforementioned cases are applicable, then no score will be assigned to the specific protein pair.

If, in any of the aforementioned steps, a score is assigned to a protein pair, the procedure continues to the next protein pair without searching for any further association (demonstrated in Figure 3). The overall score for each query protein is calculated as the sum of the scores of the query protein with each of the reference ones. The results are returned as a table in a descending order by the query protein's overall score, and can be downloaded in a comma/tab-separated format.

3.2. Data Integration from Various Resources Used by multiple UniReD

multiple UniReD uses protein clusters and PPIs predicted by UniReD version October 2021 using an inflation value of 2.0 for the creation of protein clusters. For each protein found in any cluster formed by UniReD's methodology, we integrated information regarding paralogues, human–mouse orthologues and complexes. Paralogue protein pairs were obtained from Ensembl [24] for both *H. sapiens* and *M. musculus* species (data retrieved on 25 May 2022 (*H. sapiens*) and 21 June 2022 (*M. musculus*)). Orthologues were also obtained from the same source on 25 May 2022. Complex information originates from the ComplexPortal [25] resource at EBI (data retrieved on 25 May 2022 (*H. sapiens*) and 21 June 2022 (*M. musculus*)). Furthermore, gene names, NCBI gene names and Ensembl gene IDs were retrieved from UniProt and Ensembl (data retrieved on 25 May 2022 (*H. sapiens*) and 21 June 2022 (*M. musculus*)). The result of this procedure was a dictionary (JSON format) summarizing information for each protein found to appear in clusters by UniReD. Each protein entry in this dictionary is represented by the UniProt/SwissProt accession number of proteins. For each one of them, fields for the gene name, NCBI gene name, Ensembl gene ID, paralogues, ComplexPortal complex id and mouse–human orthologues are provided. Finally, labels are used to highlight the existence of reviewed mappings between a UniProt accession number and all aforementioned identifiers.

3.3. Implementation and Running Time

The pipeline was developed in a Python 3 environment [25] (Python 3.7). The web interface was built using R and Shiny. The reticulate package was used in order to integrate the Python script with R [26] (R Core Team, 2022) and Shiny [27] (Winston et al., 2021). Regarding runtime, it takes just a few seconds for an average reference and query list of 100 proteins.

3.4. Statistical Analysis of the Results

We produced 1000 random lists of query proteins, and we calculated the sum of the overall score with the same list of reference proteins. The score differs significantly when compared to the sum of the overall score of the test case (p value $< 10^{-3}$).

4. Conclusions

There are several experimental or in silico approaches to confront the important problem of PPI discovery/prediction. Living in the -omics era, there are different large-scale experimental approaches which produce a lot of results that can sometimes be confusing and difficult to interpret. Therefore, emerging computationally predicted PPIs can provide strong indications and hints about putative PPIs and thus complement the wet lab findings to maximum exploitation. UniReD is a tool that, among other things, predicts PPIs using biomedical literature mining and, with the updated feature of *multiple* UniReD, assists in the identification/validation of putative biomarkers. It accepts two lists of genes, one with genes known to participate in a specific pathway/disease and a second one with genes suspected to be involved in this particular pathway/disease. UniReD ranks the latter in relation to the first one, thus helping in steering the experimental verification, which is required for the confirmation of the PPI prediction, in the right direction. Biomarker validation can be of great importance, because, amongst others, can shed light on disease biology by revealing new unknown interactions, unveil potential novel therapeutic targets and contribute to the understanding of the pathogenic processes.

Author Contributions: Conceptualization, I.I.; methodology, I.B., T.T. and I.I.; software, I.B., T.T. and I.I.; validation, M.P., E.C. and E.A.; formal analysis, I.B., T.T., N.P., G.A.P., G.D.A., M.P., E.C., E.A. and I.I.; writing—original draft preparation, I.B., T.T. and I.I.; writing—review and editing, I.B., T.T., N.P., G.A.P., G.D.A., M.P., E.C., E.A. and I.I.; visualization, I.B., T.T. and I.I.; supervision, I.I.; funding acquisition, I.I. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data analyzed in this study are publicly available.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Papanikolaou, N.; Pavlopoulos, G.A.; Theodosiou, T.; Iliopoulos, I. Protein-protein interaction predictions using text mining methods. *Methods* **2015**, *74*, 47–53. [CrossRef]
2. Pellegrini, M.; Marcotte, E.M.; Thompson, M.J.; Eisenberg, D.; Yeates, T.O. Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 4285–4288. [CrossRef] [PubMed]
3. Promponas, V.J.; Ouzounis, C.A.; Iliopoulos, I. Experimental evidence validating the computational inference of functional associations from gene fusion events: A critical survey. *Brief. Bioinform.* **2014**, *15*, 443–454. [CrossRef] [PubMed]
4. Dandekar, T.; Snel, B.; Huynen, M.; Bork, P. Conservation of gene order: A fingerprint of proteins that physically interact. *Trends Biochem. Sci.* **1998**, *23*, 324–328. [CrossRef]
5. Papanikolaou, N.; Pavlopoulos, G.A.; Pafilis, E.; Theodosiou, T.; Schneider, R.; Satagopam, V.P.; Ouzounis, C.; Eliopoulos, A.G.; Promponas, V.J.; Iliopoulos, I. BioTextQuest(+): A knowledge integration platform for literature mining and concept discovery. *Bioinformatics* **2014**, *30*, 3249–3256. [CrossRef]
6. Papanikolaou, N.; Pavlopoulos, G.A.; Theodosiou, T.; Vizirianakis, I.S.; Iliopoulos, I. DrugQuest—A text mining workflow for drug association discovery. *BMC Bioinform.* **2016**, *17* (Suppl. S5), 182. [CrossRef]
7. Pletscher-Frankild, S.; Pallejà, A.; Tsafou, K.; Binder, J.X.; Jensen, L.J. DISEASES: Text mining and data integration of disease-gene associations. *Methods* **2015**, *74*, 83–89. [CrossRef]
8. Fleuren, W.W.M.; Verhoeven, S.; Frijters, R.; Heupers, B.; Polman, J.; Van Schaik, R.; De Vlieg, J.; Alkema, W. CoPub update: CoPub 5.0 a text mining system to answer biological questions. *Nucleic Acids Res.* **2011**, *39*, W450–W454. [CrossRef]
9. Karatzas, E.; Baltoumas, F.A.; Kasionis, I.; Sanoudou, D.; Eliopoulos, A.G.; Theodosiou, T.; Iliopoulos, I.; Pavlopoulos, G.A. Darling: A Web Application for Detecting Disease-Related Biomedical Entity Associations with Literature Mining. *Biomolecules* **2022**, *12*, 520. [CrossRef]

10. Baltoumas, F.A.; Zafeiropoulou, S.; Karatzas, E.; Paragkamian, S.; Thanati, F.; Iliopoulos, I.; Eliopoulos, A.G.; Schneider, R.; Jensen, L.J.; Pafilis, E.; et al. OnTheFly2.0: A text-mining web application for automated biomedical entity recognition, document annotation, network and functional enrichment analysis. *NAR Genom. Bioinform.* **2021**, *3*, lqab090. [CrossRef]

11. Pafilis, E.; Buttigieg, P.L.; Ferrell, B.; Pereira, E.; Schnetzer, J.; Arvanitidis, C.; Jensen, L.J. EXTRACT: Interactive extraction of environment metadata and term suggestion for metagenomic sample annotation. *Database* **2016**, *2016*, baw005. [CrossRef]

12. Jenssen, T.-K.; Lægreid, A.; Komorowski, J.; Hovig, E. A literature network of human genes for high-throughput analysis of gene expression. *Nat. Genet.* **2001**, *28*, 21–28. [CrossRef]

13. Alanis-Lobato, G.; Andrade-Navarro, M.A.; Schaefer, M.H. HIPPIE v2.0: Enhancing meaningfulness and reliability of protein-protein interaction networks. *Nucleic Acids Res.* **2017**, *45*, D408–D414. [CrossRef] [PubMed]

14. Hoffmann, R.; Valencia, A. A gene network for navigating the literature. *Nat. Genet.* **2004**, *36*, 664. [CrossRef] [PubMed]

15. Szklarczyk, D.; Gable, A.L.; Nastou, K.C.; Lyon, D.; Kirsch, R.; Pyysalo, S.; Doncheva, N.T.; Legeay, M.; Fang, T.; Bork, P.; et al. The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **2021**, *49*, D605–D612. [CrossRef] [PubMed]

16. Aranda, B.; Achuthan, P.; Alam-Faruque, Y.; Armean, I.; Bridge, A.; Derow, C.; Feuermann, M.; Ghanbarian, A.T.; Kerrien, S.; Khadake, J.; et al. The IntAct molecular interaction database in 2010. *Nucleic Acids Res.* **2010**, *38*, D525–D531. [CrossRef]

17. Salwinski, L.; Miller, C.S.; Smith, A.J.; Pettit, F.K.; Bowie, J.U.; Eisenberg, D. The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.* **2004**, *32*, D449–D451. [CrossRef]

18. Chatr-Aryamontri, A.; Ceol, A.; Palazzi, L.M.; Nardelli, G.; Schneider, M.V.; Castagnoli, L.; Cesareni, G. MINT: The Molecular INTeraction database. *Nucleic Acids Res.* **2007**, *35*, D572–D574. [CrossRef]

19. Theodosiou, T.; Papanikolaou, N.; Savvaki, M.; Bonetto, G.; Maxouri, S.; Fakoureli, E.; Eliopoulos, A.G.; Tavernarakis, N.; Amoutzias, G.D.; Pavlopoulos, G.A.; et al. UniProt-Related Documents (UniReD): Assisting wet lab biologists in their quest on finding novel counterparts in a protein network. *NAR Genom. Bioinform.* **2020**, *2*, lqaa005. [CrossRef]

20. Savvaki, M.; Kafetzis, G.; Kaplanis, S.I.; Ktena, N.; Theodorakis, K.; Karagogeos, D. Neuronal, but not glial, Contactin 2 negatively regulates axon regeneration in the injured adult optic nerve. *Eur. J. Neurosci.* **2021**, *53*, 1705–1721. [CrossRef]

21. Kalafatakis, I.; Kalafatakis, K.; Tsimpolis, A.; Giannakeas, N.; Tsipouras, M.; Tzallas, A.; Karagogeos, D. Using the Allen gene expression atlas of the adult mouse brain to gain further insight into the physiological significance of TAG-1/Contactin-2. *Brain Struct. Funct.* **2020**, *225*, 2045–2056. [CrossRef] [PubMed]

22. Panagopoulou, M.; Karaglani, M.; Manolopoulos, V.G.; Iliopoulos, I.; Tsamardinos, I.; Chatzaki, E. Deciphering the Methylation Landscape in Breast Cancer: Diagnostic and Prognostic Biosignatures through Automated Machine Learning. *Cancers* **2021**, *13*, 1677. [CrossRef] [PubMed]

23. Karaglani, M.; Panagopoulou, M.; Baltsavia, I.; Apalaki, P.; Theodosiou, T.; Iliopoulos, I.; Tsamardinos, I.; Chatzaki, E. Tissue-Specific Methylation Biosignatures for Monitoring Diseases: An in Silico Approach. *Int. J. Mol. Sci.* **2022**, *23*, 2959. [CrossRef] [PubMed]

24. Cunningham, F.; Allen, J.E.; Allen, J.; Alvarez-Jarreta, J.; Amode, M.R.; Armean, I.M.; Austine-Orimoloye, O.; Azov, A.G.; Barnes, I.; Bennett, R.; et al. Ensembl 2022. *Nucleic Acids Res.* **2022**, *50*, D988–D995. [CrossRef] [PubMed]

25. Meldal, B.H.M.; Bye-A-Jee, H.; Gajdoš, L.; Hammerová, Z.; Horáčková, A.; Melicher, F.; Perfetto, L.; Pokorný, D.; Lopez, M.R.; Türková, A.; et al. Complex Portal 2018: Extended content and enhanced visualization tools for macromolecular complexes. *Nucleic Acids Res.* **2019**, *47*, D550–D558. [CrossRef] [PubMed]

26. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022.

27. Winston, C.; Cheng, J.; Allaire, J.J.; Sievert, C.; Schloerke, B.; Xie, Y.; Allen, J.; McPherson, J.; Dipert, A.; Borges, B. Shiny: Web Application Framework for R. Available online: https://shiny.rstudio.com (accessed on 16 September 2022).