

# *PINOT: an intuitive resource for integrating protein-protein interactions*

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Tomkins, J. E., Ferrari, R., Vavouraki, N., Hardy, J., Lovering, R. C., Lewis, P. A., McGuffin, L. J. and Manzoni, C. (2020) PINOT: an intuitive resource for integrating protein-protein interactions. *Cell Communication and Signaling*, 18 (92). ISSN 1478-811X doi: <https://doi.org/10.1186/s12964-020-00554-5>  
Available at <http://centaur.reading.ac.uk/90595/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1186/s12964-020-00554-5>

Publisher: BMC

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

## **CentAUR**

Central Archive at the University of Reading


Reading's research outputs online

METHODOLOGY

Open Access

# PINOT: an intuitive resource for integrating protein-protein interactions



James E. Tomkins<sup>1</sup>, Raffaele Ferrari<sup>2</sup>, Nikoleta Vavouraki<sup>1</sup>, John Hardy<sup>2,3,4,5,6</sup>, Ruth C. Lovering<sup>7</sup>, Patrick A. Lewis<sup>1,2,8</sup>, Liam J. McGuffin<sup>9\*</sup> and Claudia Manzoni<sup>1,10\*</sup> 

## Abstract

**Background:** The past decade has seen the rise of omics data for the understanding of biological systems in health and disease. This wealth of information includes protein-protein interaction (PPI) data derived from both low- and high-throughput assays, which are curated into multiple databases that capture the extent of available information from the peer-reviewed literature. Although these curation efforts are extremely useful, reliably downloading and integrating PPI data from the variety of available repositories is challenging and time consuming.

**Methods:** We here present a novel user-friendly web-resource called PINOT (Protein Interaction Network Online Tool; available at [http://www.reading.ac.uk/bioinf/PINOT/PINOT\\_form.html](http://www.reading.ac.uk/bioinf/PINOT/PINOT_form.html)) to optimise the collection and processing of PPI data from IMEx consortium associated repositories (members and observers) and WormBase, for constructing, respectively, human and *Caenorhabditis elegans* PPI networks.

**Results:** Users submit a query containing a list of proteins of interest for which PINOT extracts data describing PPIs. At every query submission PPI data are downloaded, merged and quality assessed. Then each PPI is confidence scored based on the number of distinct methods used for interaction detection and the number of publications that report the specific interaction. Examples of how PINOT can be applied are provided to highlight the performance, ease of use and potential utility of this tool.

(Continued on next page)

\* Correspondence: [l.j.mcguffin@reading.ac.uk](mailto:l.j.mcguffin@reading.ac.uk); [c.manzoni@ucl.ac.uk](mailto:c.manzoni@ucl.ac.uk)

<sup>9</sup>School of Biological Sciences, University of Reading, Whiteknights, Reading RG6 6AS, UK

<sup>1</sup>School of Pharmacy, University of Reading, Whiteknights, Reading RG6 6AP, UK

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



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

**Conclusions:** PINOT is a tool that allows users to survey the curated literature, extracting PPI data in relation to a list of proteins of interest. PINOT extracts a similar numbers of PPIs as other, analogous, tools and incorporates a set of innovative features. PINOT is able to process large queries, it downloads human PPIs live through PSICQUIC and it applies quality control filters on the downloaded PPI data (i.e. removing the need for manual inspection by the user). PINOT provides the user with information on detection methods and publication history for each downloaded interaction data entry and outputs the results in a table format that can be straightforwardly further customised and/or directly uploaded into network visualization software.

**Keywords:** Protein interaction, Protein network, Network, Data mining, Protein database, Data integration

## Background

During the past two decades the use of omics data to understand biological systems has become an increasingly valued approach [1]. This includes extensive efforts to detect protein-protein interactions (PPIs) on an almost proteome-wide scale [2, 3]. The utility of such data has been greatly supported by primary database curation and the International Molecular Exchange (IMEx) Consortium, which promotes collaborative efforts in standardising and maintaining high quality data curation across the major molecular interaction data repositories [4]. The primary databases, such as IntAct [5] and BioGRID [6], are rich data resources providing a comprehensive record of published PPI literature. PPI data are critical to describe connections among proteins, which in turn supports both inference of new functions for proteins (based on the guilt by association principle [7]) and visualization of protein connectivity via shared interactors. This enables identification of potential communal pathways involving proteins of interest [8–10]. Additionally, literature extracted PPI data can support the prioritization of interactions from high-throughput experiments (which generate large lists of potential PPI hits), assisting the selection of candidates for further analysis/validation [11].

However, the process of collating PPI data is currently hampered by the fact that no single data source encompasses the full extent of PPIs reported in literature; hence, users are required to merge (partial) information mined from multiple different primary databases. Merging such data is not straightforward due to inconsistencies in data format and differences in data curation across the PPI databases (e.g. IMEx member vs non-member databases).

To optimize the use of PPI data within the public domain, we developed a user-friendly tool that assists PPI data extraction and processing: the Protein Interaction Network Online Tool (PINOT). This tool represents the development (and automation) of our previous PPI analysis framework, termed Weighted Protein-Protein Interaction Network Analysis (WPPINA) [9, 11–15]. Through PINOT, PPI data are downloaded directly (i.e.

downloaded “live” at the time of the query) from seven databases using the Proteomics Standard Initiative Common Query Interface (PSICQUIC) and integrated to ensure a wide coverage of the PPIs available from these repositories [16]. These data are scored through a simple and transparent procedure based on ‘method detection’ and ‘publication records’ and allows the user to further apply customized confidence thresholds. PINOT is fully automated and available online as an open access resource. Output data are provided as a summary table (directly online or emailed to the user), which summarizes the most comprehensive current knowledge of the PPI landscape for the protein(s)-of-interest submitted in the query list. Additionally, and of note, the R scripts that underlie PINOT can be freely downloaded from the help-page.

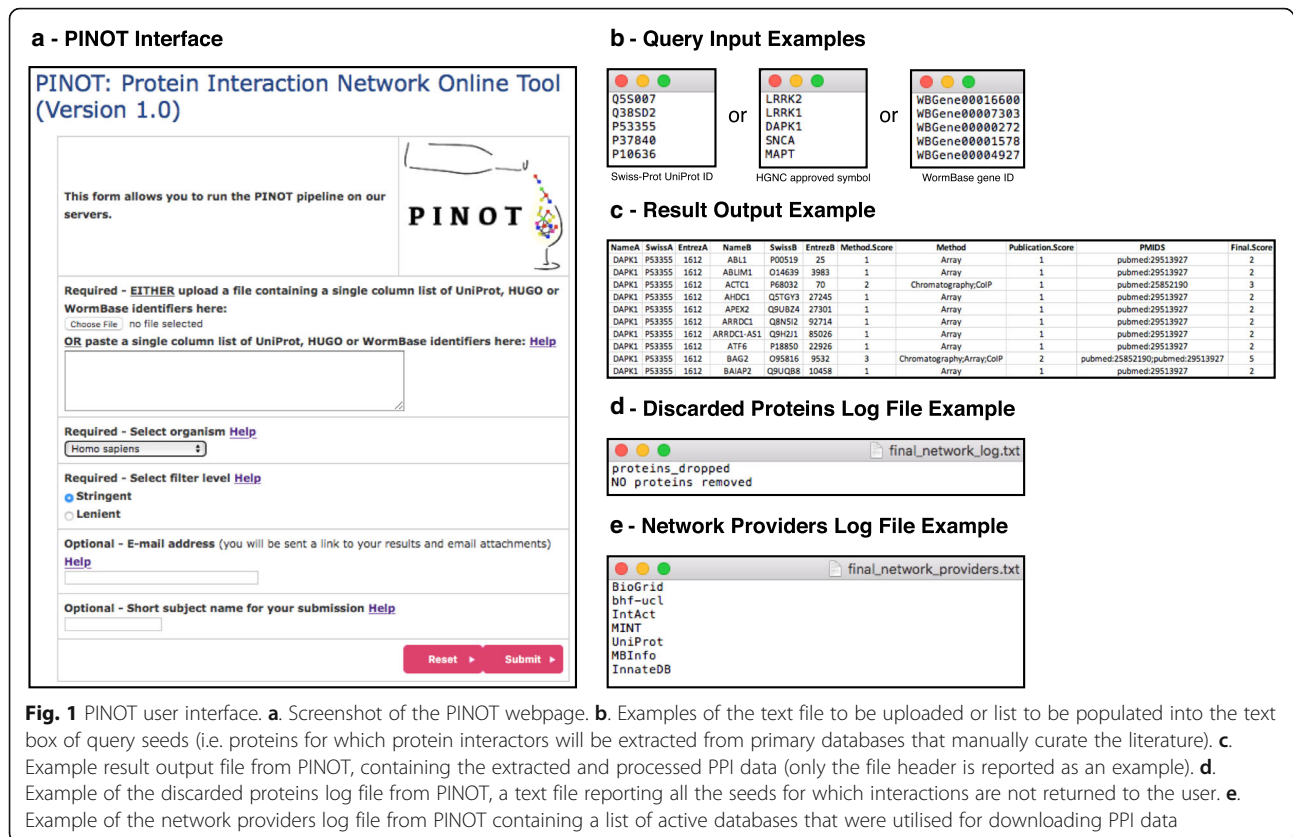
## Methods

### Protein interaction network online tool (PINOT)

PINOT can be run automatically at [http://www.reading.ac.uk/bioinf/PINOT/PINOT\\_form.html](http://www.reading.ac.uk/bioinf/PINOT/PINOT_form.html) (hereafter referred to as “web tool”). A choice of parameters is integrated by default as explained further below and in Supplementary Materials S1. Alternatively, R scripts can be downloaded from the help-page and in this way PINOT can be used as a “standalone tool” whereby parameters can be modified as *per* user choice.

A list of proteins of interest (seeds) can be queried to identify their literature-reported interactors that have been curated into PPI databases (Fig. 1).

For *Homo sapiens* (taxonomy ID: 9606) the seed identifiers submitted into the query field must be in an approved HUGO Gene Nomenclature Committee (HGNC) gene symbol [17] or valid Swiss-Prot UniProt ID format. Upon query submission, PPI data are extracted directly (*via* API: Shannon, P. (2020) PSICQUIC R package, DOI: <https://doi.org/10.18129/B9.bioc.PSICQUIC> [18]) from seven primary databases, all of which directly annotate PPI data from peer-reviewed literature: bhf-ucl, BioGRID [6], InnateDB [19], IntAct [5], MBIInfo (<https://www.mechanobio.info>), MINT [20] and UniProt [21]. The downloaded protein interaction data are then



parsed, merged, filtered and scored (Fig. 2) automatically by PINOT. A detailed description of the PINOT pipeline can be found in the Supplementary Materials S1. The user can select to run PINOT with lenient or stringent filter parameters. The output of PINOT (Fig. 1c-e) consists of: i) a network file (final\_network.txt), which is a tab-spaced text file containing the processed PPI data in relation to the seeds in the initial query list; ii) a log file (final\_network\_log.txt) reporting proteins that have been discarded within the data processing procedure, and; iii) a further log file (final\_network\_providers.txt) indicating the PPI databases used by the API for data acquisition. The output dataset is available for download and/or emailed to the user.

For *Caenorhabditis elegans* (taxonomy ID: 6239) the seed identifiers must be in an approved WormBase gene ID [22, 23] format, “WBGene” followed by 8 numerical digits. Upon submission, PPI data are downloaded from an internal network stored within PINOT and created (following similar criteria applied for the human PPIs - details in Supplementary Materials S1) based on the WormBase PPI catalogue (Alliance\_molecular\_interactions.tar file downloaded from the Alliance of Genome Resources [version 2.1] on 15th April 2019) [22, 24]. The user can apply stringent or lenient filtering options. The output of PINOT for a *C. elegans* query consists of:

i) a network file (final\_network.txt), which is a tab-spaced text file containing the processed PPIs for the seeds in the initial query list; and ii) a log file (final\_network\_log.txt) reporting proteins that have been discarded within the data processing procedure.

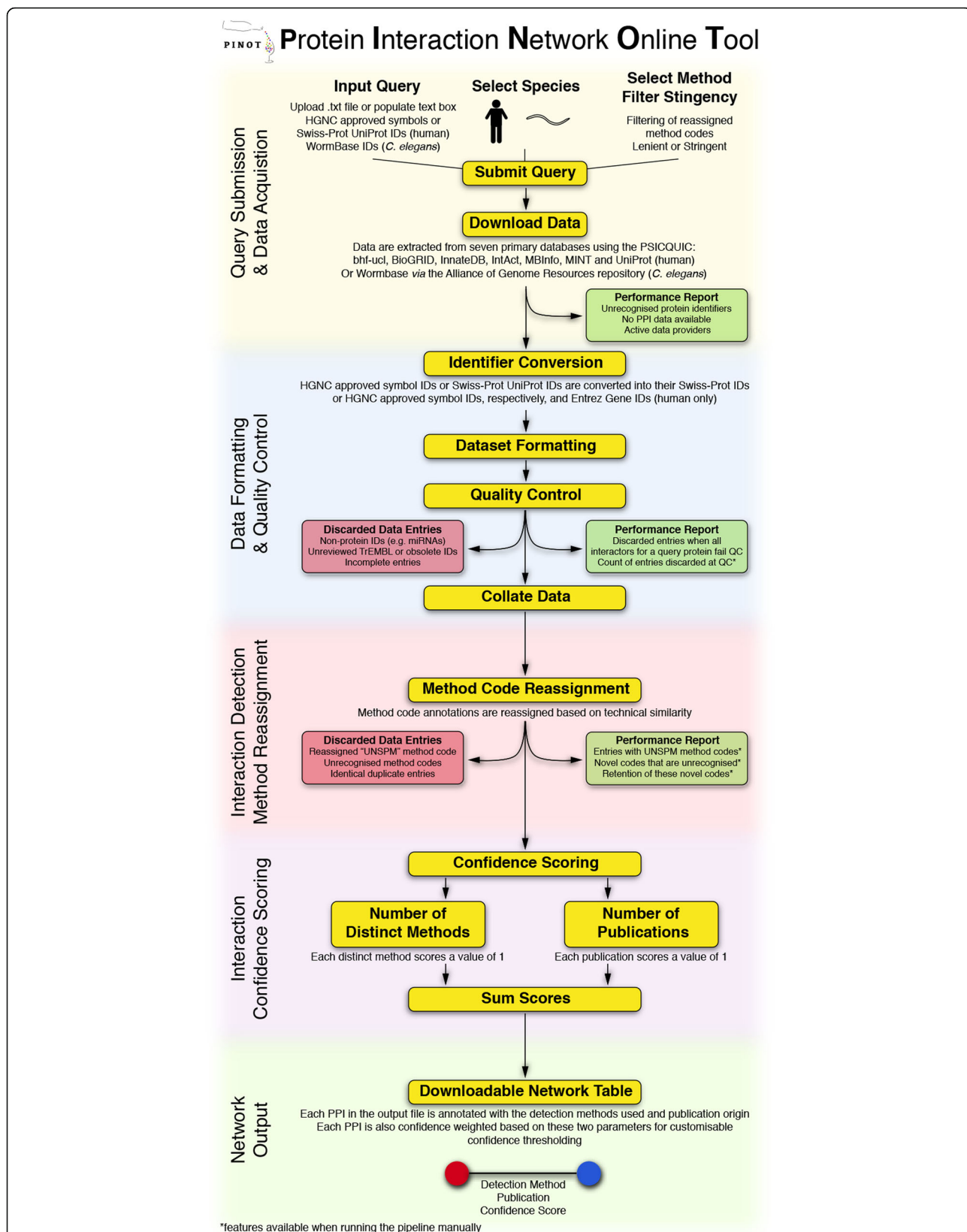
### Software

The PINOT pipeline is coded in R and runs on a Linux server at the University of Reading, with java servlets processing user’s submissions via the web interface.

### PINOT quality control

We have tested the PINOT pipeline using multiple input query lists structured as follows: i) small input lists (less than 20 seed proteins), selected randomly or in association with typical processes suspected to be functionally relevant for Parkinson’s Disease (PD); and ii) a large input list = 941 proteins, the human mitochondrial proteome as reported by MitoCarta2.0 [25].

PINOT was compared to two alternative yet related online PPI query tools, Human Integrated Protein-Protein Interaction Reference (HIPPIE; for human data only) and Molecular Interaction Search Tool (MIST; for human and *C. elegans* data). For this analysis, query parameters were selected (where possible) to maximize the extraction of protein interactions: HIPPIE was used with



**Fig. 2** PINOT pipeline. A stepwise insight into the process which underlies the PINOT pipeline. Performance reports (green boxes) are generated and data are discarded (red boxes) at numerous stages within the pipeline to ensure high quality and transparent data processing

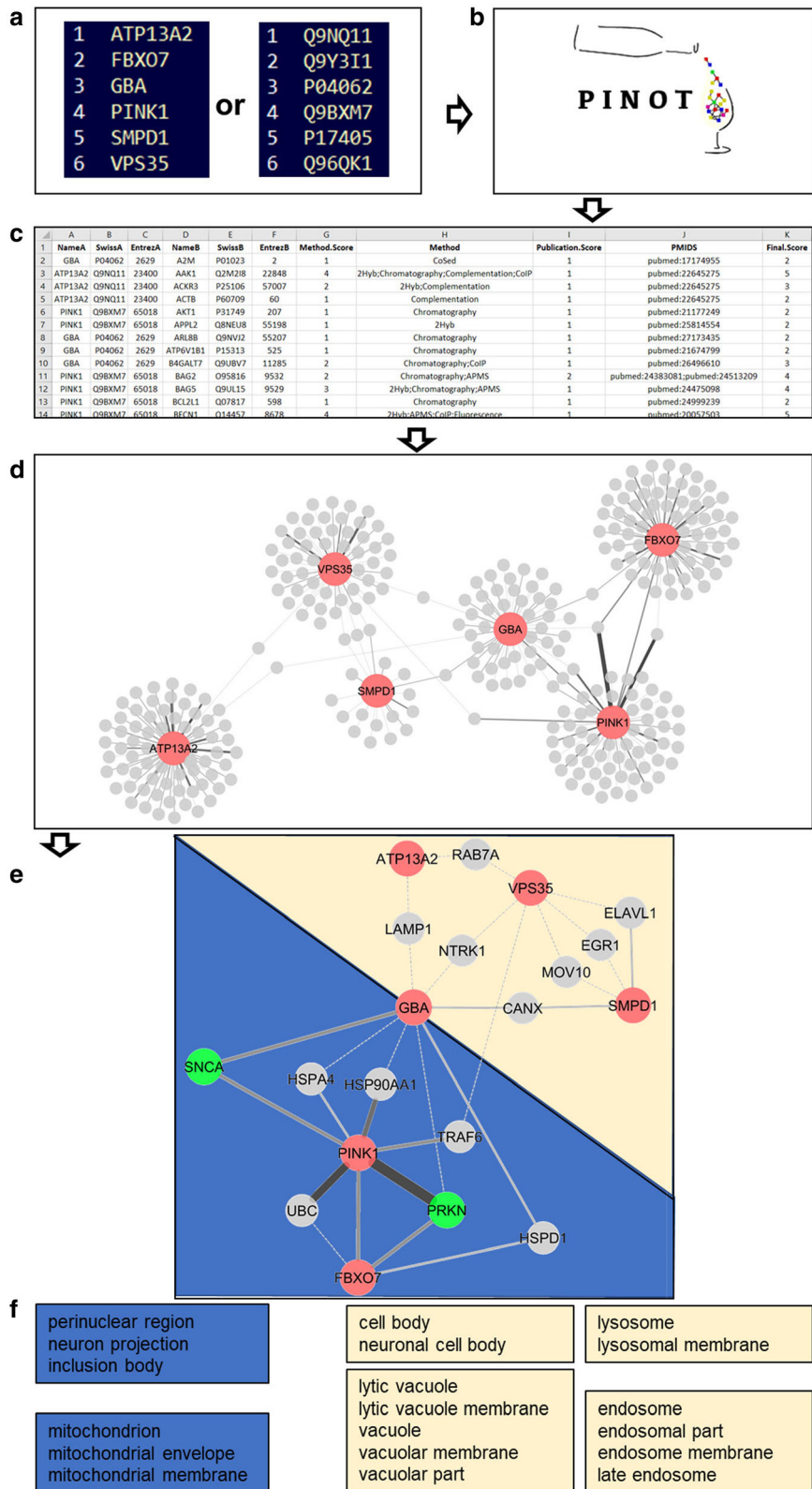


Fig. 3 (See legend on next page.)

(See figure on previous page.)

**Fig. 3** PINOT: An example application. A stepwise insight into the potential use of PINOT. **a.** A submission list is created as a text file using gene names as per HGNC approved symbols or Swiss-Prot UniProt IDs; the submission list can be uploaded as a file or pasted into the text box within the PINOT interface. **b.** PINOT downloads, from PSICQUIC, the human PPIs (in this example, stringent filter applied) **c.** PPIs are provided back to the user via email or from the webpage; results are in a parsable file that can be opened by a text reader application and/or imported into spreadsheet software such as Microsoft Excel. **d.** The interactions can be visualized in a network format by opening the PINOT output in network visualisation software, such as Cytoscape. Connections between nodes (edges) are coded with increased line width based on the final score that interaction was assigned by PINOT. The wider the edge – the more confident PINOT is about the interactions. **e.** The interactions can be further processed according to the user's research question, in this case, only interactors that are communal to at least 2 of the initial query proteins have been retained, generating a core network (in dark-red the initial seeds; in bright-green the identified common interactors that are proteins mutated in PD). Based on the network topology the seeds and their interactors can be visually clustered into group 1 (depicted in gold) and group 2 (depicted in blue). **f.** Specific functional enrichment (GO CC terms) for groups 1 and 2 after filtering out the less represented terms. Analyses performed on the 22nd August 2019 (g:Profiler)

confidence score = 0 and no filters on confidence level, interaction type or tissue expression; and MIST was used with no filter by rank parameter to download all PPIs regardless of the assigned confidence score. It is noteworthy to highlight that files from HIPPIE and MIST required manual parsing after download to remove entries that do not include a PMID and/or conversion method code (incomplete entries). Data were downloaded on 18th September 2019 (*H. sapiens*) and on 24th September 2019 (*C. elegans*).

## Results

PINOT is a web tool that takes a list of proteins/genes (seeds) as input, queries PSICQUIC at submission and returns an up-to-date table containing a comprehensive list of PPIs - published in peer-reviewed literature - centred upon the seeds. This table consists of a variable number of rows and 11 columns (Fig. 1c and Fig. 3c). Each row represents a binary interaction between one of the seeds (interactor A) and one of its specific protein interactors (interactor B). The 11 columns contain: the gene name, the Swiss-Prot protein ID and the Entrez gene ID for interactor A and B ("NameA", "SwissA", "EntrezA", "NameB", "SwissB", "EntrezB"); the number and type of different methods through which the interaction has been identified ("Method.Score", "Method"), and; the number of different publications reporting the interaction and the corresponding PubMed IDs ("Publication.Score", "PMIDS"). The final column ("Final.Score") contains a confidence score based on the number of different methods + the number of different publications reporting the interaction. PPIs with a final score of 2 are reported in literature by 1 publication and detected by 1 technique; these PPIs should be considered with caution since the available data suggest they are not robustly replicated. They might be either: i) false positives, ii) true novel interactions that have not yet been replicated in additional studies, or iii) true interactions that have been replicated in additional studies but which have not yet been incorporated into any of the seven primary databases used for data acquisition. A

final score > 2 suggests a degree of replication that can be either or both: multiple publications reporting the PPI and/or multiple distinct techniques used to detect the interaction. It is not possible to obtain a final score < 2 since every PPI annotation - to be retained in PINOT - has to be supported by at least 1 interaction detection method and 1 PMID; if this condition is not met, the PPI is discarded by PINOT and not shown in the output file.

The PINOT output can be imported into Cytoscape [26] for network visualization by selecting the "NameA" and "NameB" columns as source and target nodes, respectively.

## PINOT: example of application

In Fig. 3, PINOT has been used to download PPIs for a limited selection of human protein products of genes mutated in familial PD: ATP13A2, FBXO7, GBA, PINK1, SMPD1 and VPS35 (seeds). PINOT quickly retrieved a table containing 327 interactions from peer-reviewed curated literature (with associated PMIDs) thus supporting and simplifying otherwise time-consuming classical literature mining. The PINOT output was imported into Cytoscape and PPIs were visualized in a network ("NameA" = source and "NameB" = target), the seeds were highlighted in dark-red and the edges (interactions between each protein pair) were coded based on the "Final.Score" field, thus highlighting the confidence (number of methods + number of publications) of the interaction, which positively correlated with the thickness of the edge. Since we were interested in interactors that were common to the seeds - and not exclusive interactors of just one seed - the network was filtered retaining only the nodes (interactors) that bridged two or more seeds. The obtained core-network revealed that among the common interactors of the seeds (6 PD proteins) there were 2 proteins (SNCA and PRKN) which are products of 2 additional genes known to be mutated in familial PD. Thus, the analysis pointed towards the involvement of SNCA and PRKN in PD even if they were initially excluded from the list of seeds. Additionally, topological analysis (based on the number and thickness of the edges) suggested that the core network could be subdivided into 2



distinct clusters: PINK1, FBXO7 and the newly identified PRKN and SNCA in the first cluster, while ATP13A2, VPS35 and SMPD1 were more closely associated in the second cluster, with GBA a bridge seed between the 2 clusters. This observation suggested a dichotomy, based on the protein interactomes of the seeds included in the initial input list. Based on the guilt-by-association principle we hypothesised that the proteins contributing to these clusters could be associated with different cellular functions and components. We therefore performed functional enrichment analysis (based on Gene Ontology (GO) Cellular Component (CC) annotations) [27, 28] using g:Profiler [29] revealing that indeed, clusters 1 and 2 are associated with mitochondria and vacuoles/lysosomes/endosomes, respectively.

#### H. sapiens - PINOT performance

The performance of PINOT was compared to that of alternative resources for both small and large lists of seeds. Regarding the former, five different small seed lists were used as input for PPI query in HIPPIE [30] and MIST [31], two alternative online and freely available resources. It should be noted that, despite apparent similarities, each of these tools has been developed differently. All three resources (PINOT, HIPPIE and MIST) have distinguishing features for addressing different research questions (Table 1). The results of the different online-platforms queries have been compared, evaluating the total number of interactors provided in the output (Fig. 4a).

PINOT, HIPPIE and MIST retrieved a similar number of PPIs. PINOT with stringent filtering applied, was

always extracting fewer interactions; this is an expected outcome since this filter option is built with the purpose of retaining only annotations that have survived stringent screening, largely based on completeness of curated data entries.

Results for the large input list query were compared, PINOT vs HIPPIE. These two web tools allowed for easy processing of more than 900 seeds within the submission list. The number of retrieved interactors was slightly higher for HIPPIE in comparison with PINOT when the stringent quality control (QC) filter was applied; however PINOT retrieved more interactions than HIPPIE when the lenient filter option was selected (Fig. 4b). Furthermore, the vast majority of downloaded interactions were similar when using the two resources, suggesting that PINOT is able to confidently extract specific interactions from the curated PPI literature (Fig. 4c).

#### C. elegans - PINOT performance

The performance of PINOT for querying *C. elegans* PPI data was tested alongside the *C. elegans* query option in MIST, assessing interaction networks of different dimensions (Fig. 5). The data acquisition strategy underlying these two resources differs slightly, PINOT extracts data from the latest release of WormBase molecular interaction data, whereas MIST utilises data from numerous sources, including WormBase, BioGRID and IMEx associated repositories.

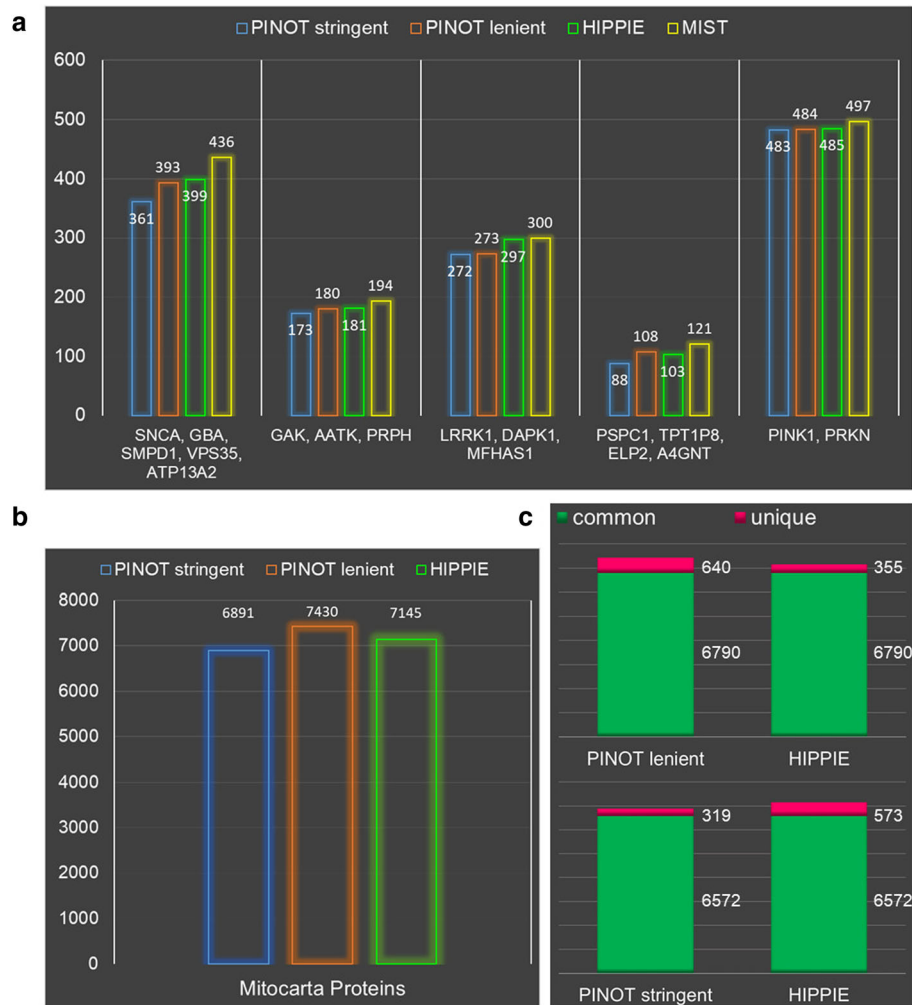
Similar to comparisons across the human PPI query capacity, PINOT and MIST performed comparably in terms of the number of PPI data entries extracted. More specifically and as previously described with human data, PINOT extracting slightly fewer across these test query cases. However, upon assessing the completeness of these extracted data entries, in terms of interaction detection method and/or PMID annotations, there was a striking difference in performance. Since the PINOT pipeline focusses on the QC of data, all data entries within the output dataset were complete, whereas incomplete data entries persisted in the MIST output dataset thus requiring manual inspection. In the more abundant PPI data pools, for example when querying the ATP and CED *C. elegans* proteins, incomplete data entries accounted for the majority of the output dataset in MIST.

#### Discussion

PINOT can be used as a tool to quickly and effectively survey the curated literature and download the most up-to-date PPI data available for a given set of proteins/genes of interest. This is particularly useful to assist anyone attempting to mine overwhelmingly abundant literature targeting certain proteins/genes, in relation to identifying reported PPIs.

**Table 1** Comparison of features available within the PINOT, HIPPIE, MIST and STRING online resources

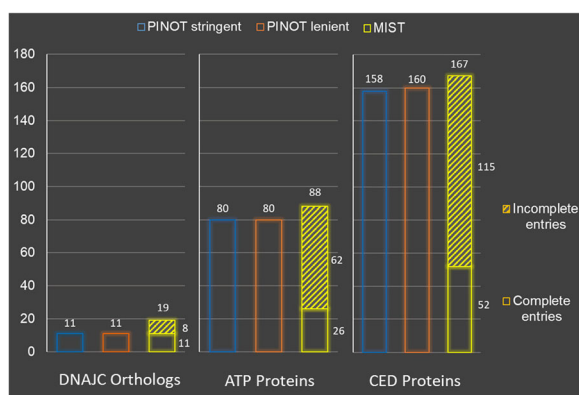
	PINOT	HIPPIE	MIST	STRING
Live PPI data	yes	no	no	no
Large submission	yes	yes	yes	no
Easily parsable result table	yes	yes	no	yes
PPIs for seeds only	yes	yes	yes	no
Visualization app	no	yes	yes	yes
Multiple species PPIs	yes	no	yes	yes
Score	yes	yes	yes	yes
Pubmed ID (PMID)	yes	yes	yes	no
Detection method	yes	yes	yes	no
QC on method and PMID	yes	no	no	–
Entrez ID in output	yes	yes	yes	no
Swiss-Prot ID in output	yes	yes	no	no
Codes available	yes	command line version	API	API



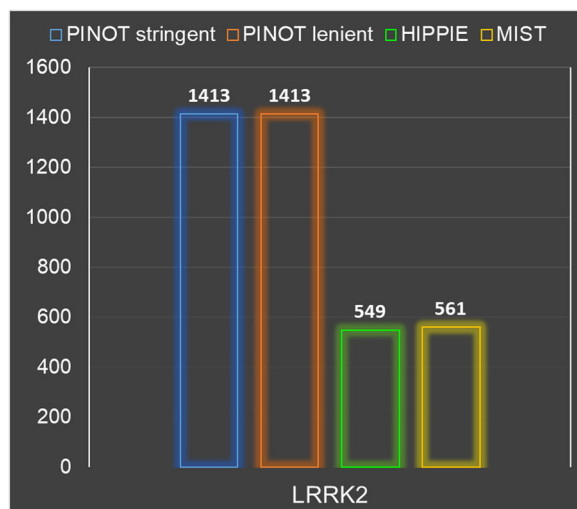
**Fig. 4** PINOT: Performance & Sensitivity. **a.** The performance of PINOT was evaluated by assessing the number of interactors retrieved (gene names) upon submission of the reported query lists to PINOT (with stringent and lenient filtering), HIPPIE and MIST (on 18th September 2019). The databases were set to retrieve the maximum number of interactions (by removing all possible filters). HIPPIE and MIST were queried online and outputs were manually cleaned to remove interactions with i) no interaction detection method; ii) no PMID; iii) multiple Entrez IDs. The number of retained interactions retrieved is reported on top of each bar within the graphs. **b.** The number of interactions PINOT (with stringent and lenient filtering) and HIPPIE obtained when querying a seed list of 941 protein from Mitocarta 2.0. **c.** Comparison between PINOT and HIPPIE indicated that the vast majority of interactors (Entrez IDs) downloaded by the two tools was identical: 6790 common interactors for PINOT lenient (640 unique interactors) vs HIPPIE (355 unique interactors); 6572 common interactors for PINOT stringent (319 unique interactors) vs HIPPIE (573 unique interactors)

The PPI data downloaded through PINOT can be used as a literature reference list for experimental PPI data resulting from high-throughput experiments (protein microarrays, yeast two-hybrid screens, etc) facilitating the prioritisation of experimental results for validation. PINOT is also useful to evaluate interactors of different proteins/genes of interest within an input seed list simultaneously. The analysis of the combined interactomes of such seeds can reveal the existence of communal interactors, can provide a base to cluster the seeds into groups and can support further functional analysis to better characterize the functional landscape of seeds of interest.

Alternative tools that appear to be similar to PINOT are HIPPIE and MIST. STRING [32] is a conceptually different tool; it does not report 'interaction detection methods' nor 'Publication IDs' for PPIs, which are crucial pieces of information for the evaluation and interpretation of PPI data. Additionally, the reported interactions are not focused only on the proteins in the input list; interactions of interactors are also reported. Distinguishing features of PINOT, HIPPIE and MIST include the implementation of a tailored confidence score for different methodological approaches (as well as a tissue filtering option) in HIPPIE; MIST provides a valuable resource for users interested in mapping PPIs



**Fig. 5** PINOT and MIST performance comparison for *C. elegans* PPI data. The performance of PINOT (with stringent and lenient filter options) and MIST was assessed in terms of the number of PPI data entries extracted upon querying different protein lists (on 24th September 2019). The output datasets were evaluated in relation to the number of complete and incomplete (lacking interaction detection method and/or PubMed ID annotations) data entries extracted. The query lists were PD-associated DNAJC orthologs: DNJ-14, DNJ-25, DNJ-27, Y73B6BL.12, K07F5.16, RME-8 and GAKH-1; ATP proteins: ATP-1, ATP-2, ATP-3, ATP-4, ATP-5 and ATP-6; and CED proteins: CED-1, CED-2, CED-3, CED-4, CED-5, CED-6, CED-7, CED-8, CED-9, CED-10, CED-11, CED-12 and CED-13. The input format used for PINOT was the WormBase gene ID, the common gene name (as listed here) was used for MIST querying and no filter by rank parameter was set



**Fig. 6** LRRK2 interactome. PINOT performance was evaluated by counting the number of interactors retrieved (gene names) for LRRK2 by using PINOT (with stringent and lenient filtering), HIPPIE and MIST. The databases have been set to retrieve the maximum number of interactions (by removing all possible filters). HIPPIE and MIST output were manually cleaned to remove interactions with i) no interaction detection method; ii) no PMID; iii) multiple Entrez IDs. The number of retained interactions retrieved is reported on top of each bar (queries submitted 18th September 2019)

across species (i.e. interologs); PINOT focusses on high quality PPI data output by implementing multiple QC steps to remove problematic or non-univocal annotations. The performance of PINOT was comparable to that of HIPPIE and MIST both in terms of number and identity of downloaded interactions. However, there are some unique features of PINOT that are not, at the moment, integrated within the other data resources. Human PPIs in PINOT are directly downloaded from PSICQUIC at every query submission. In contrast, PPIs in HIPPIE and MIST are recovered from a central built-in repository within the servers. This difference is clearly demonstrated by searching for interactors of LRRK2, where (at the time of analysis) one high-throughput publication was updated in PSICQUIC, while both HIPPIE and MIST did not contain this full annotation yet (Fig. 6).

PINOT has access to the most up-to-date interactions that could be retrieved at a given time from PSICQUIC (however, it has to be considered that each database is responsible for updating their PSICQUIC service and therefore discrepancies might exist with the central databases).

PINOT implements QC filtering, which involves discarding PPI data entries that are curated without a PMID or with multiple PMIDs and/or without an interaction detection method annotation. Therefore the output file from PINOT does not require any further QC by the user, while output data from HIPPIE and MIST require manual parsing and inspection before analysis to remove incomplete data entries through a time consuming, post-hoc processing procedure.

Another distinctive feature of the PINOT pipeline is the implementation of a unique interaction detection method conversion step. During this step, the interaction detection method annotation for each downloaded interaction data entry is converted based on a conversion table (Supplementary Materials S2) that is available from the PINOT web-portal. During this conversion, technically similar methods are grouped together. For example: “Two Hybrid - MI:0018”, “Two Hybrid Array - MI:0397” and “Two Hybrid Pooling Approach - MI:0398” are grouped together into the “Two Hybrid (2Hyb)” method category. This step of ‘method clustering and reassignment’ is critical to assess the actual number of distinct methods used to detect a particular interaction and to dilute the bias caused in the event of the same technique being annotated under slightly different method codes in different PPI databases.

Interaction scores are provided in different formats by HIPPIE, MIST and PINOT. HIPPIE incorporates a filtering system based on a confidence score between 0 and 1 that can be set either before or after the analysis. This is a complex scoring system, which takes into consideration multiple parameters, such as the number of

publications that report a specific interaction and a semi-computational quality score based on the experimental approach (for example, imaging techniques would score different than direct interactions etc.) [33]. MIST similarly has an option for filtering interactions pre- or post-analysis; however, this is based on fixed ranking values defined as low, medium (interaction supported by other species), or high (supported by multiple experimental methods and/or reported in multiple publications). In the case of PINOT, two different scores are provided: the interaction detection method score (MS) which reports the number of different methods used (after method annotation reassignment), while the publication score (PS) counts the number of different publications which report the interaction. Finally, the coding scripts which underlie the human PPI data PINOT pipeline are fully available for download. They are coded in R to make them accessible to a large research audience and a read-me text file helps customization of the scripts according to the users' needs. Some of the divergent features across PINOT, HIPPIE, MIST and STRING are reported in Table 1.

### Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12964-020-00554-5>.

**Additional file 1: S1.** = supplementary materials and methods.

**Additional file 2: S2.** = supplementary file 1: 'interaction detection method conversion'.

### Abbreviations

HIPPIE: Human integrated protein-protein interaction reference; MIST: Molecular interaction search tool; MS: Method score; PPI: Protein protein interaction; PD: Parkinson's disease; PINOT: Protein interaction network online tool; PMID: Pubmed identifier; PS: Publication score; PSICQUIC: Proteomics standard initiative common query interface; QC: Quality control; WPPINA: Weighted protein-protein interaction network analysis

### Acknowledgements

We thank WormBase for providing the list of PPIs used for the generation of the internal *C. elegans* network used when *C. elegans* genes/proteins are queried by PINOT users.

### Authors' contributions

JET, CM and RF, elaborated the pipeline and wrote the manuscript; JET, CM and NV wrote the scripts and tested the pipeline, LJM implemented the pipeline for the website, JH, RCL, PAL offered critical advice for the implementation of the pipeline, critically reviewed the manuscript and obtained financial support. All authors read and approved the final manuscript.

### Funding

This work was supported by the: Biomarkers Across Neurodegenerative Diseases Grant Program 2019, BAND3 (Michael J Fox Foundation, Alzheimer's Association, Alzheimer's Research UK and Weston Brain Institute [grant number 18063 to CM and PAL]); Biotechnology and Biological Sciences Research Council CASE studentship with BC Platforms to PAL [grant number BB/M017222/1]; Engineering and Physical Sciences Research Council to PAL [grant number EP/M508123/1]; the Medical Research Council [grant numbers MR/N026004/1 to JH and PAL; MR/L010933/1 to PAL]; the Wellcome Trust to

JH [grant number 202903/Z/16/Z]; Alzheimer's Society to RF [grant number 284]; Alzheimer Research UK to RCL [grant number ARUK-NAS2017A-1]; the National Institute for Health Research University College London Hospitals Biomedical Research Centre to RCL and JH. This work was also supported: by the UK Dementia Research Institute (which receives its funding from DRI Ltd, funded by the UK Medical Research Council, Alzheimer's Society and Alzheimer's Research UK); by the Dolby Family Fund; and by the BRCNIHR Biomedical Research Centre at University College London Hospitals NHS Foundation Trust and University College London.

### Availability of data and materials

This resource is available as a fully automated web-server at: [http://www.reading.ac.uk/bioinf/PINOT/PINOT\\_form.html](http://www.reading.ac.uk/bioinf/PINOT/PINOT_form.html); R scripts, which underlie this bioinformatics pipeline, are free to download at the help-page.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>School of Pharmacy, University of Reading, Whiteknights, Reading RG6 6AP, UK. <sup>2</sup>Department of Neurodegenerative Disease, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK. <sup>3</sup>UK Dementia Research Institute at UCL and Department of Neurodegenerative Disease, UCL IoN, UCL London, UK. <sup>4</sup>Reta Lila Weston Institute, UCL IoN, 1 Wakefield Street, London WC1N 1PJ, UK. <sup>5</sup>UCL Movement Disorders Centre, UCL, London, UK. <sup>6</sup>Institute for Advanced Study, The Hong Kong University of Science and Technology, Hong Kong SAR, China. <sup>7</sup>Functional Gene Annotation, UCL Institute of Cardiovascular Science, 5 University Street, London WC1E 6JF, UK. <sup>8</sup>Royal Veterinary College, Royal College Street, London NW1 0TU, UK. <sup>9</sup>School of Biological Sciences, University of Reading, Whiteknights, Reading RG6 6AS, UK. <sup>10</sup>School of Pharmacy, UCL, 29-39 Brunswick Square, London WC1N 1AX, UK.

Received: 5 October 2019 Accepted: 17 March 2020

Published online: 11 June 2020

### References

- Manzoni C, Kia DA, Vandrovicova J, Hardy J, Wood NW, Lewis PA, et al. Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. *Brief Bioinform.* 2018;19(2):286–302.
- Hein MY, Hubner NC, Poser I, Cox J, Nagaraj N, Toyoda Y, et al. A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell.* 2015;163(3):712–23.
- Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K, et al. Architecture of the human interactome defines protein communities and disease networks. *Nature.* 2017;545(7655):505–9.
- Orchard S, Kerrien S, Abbani S, Aranda B, Bhate J, Bidwell S, et al. Protein interaction data curation: the international molecular exchange (IMEx) consortium. *Nat Methods.* 2012;9(4):345–50.
- Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, et al. The IntAct molecular interaction database in 2012. *Nucleic Acids Res.* 2012;40(Database issue):D841–6.
- Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, Chen D, et al. The BioGRID interaction database: 2015 update. *Nucleic Acids Res.* 2015;43(Database issue):D470–8.
- Bonetta L. Protein-protein interactions: Interactome under construction. *Nature.* 2010;468(7325):851–4.
- Blokhuis AM, Koppers M, Groen EJ, van den Heuvel DM, Dini Modigliani S, Anink JJ, et al. Comparative interactomics analysis of different ALS-associated proteins identifies converging molecular pathways. *Acta Neuropathol.* 2016;132(2):175–96.
- Ferrari R, Kia DA, Tomkins JE, Hardy J, Wood NW, Lovering RC, et al. Stratification of candidate genes for Parkinson's disease using weighted protein-protein interaction network analysis. *BMC Genomics.* 2018;19(1):452.
- Hwang S, Son SW, Kim SC, Kim YJ, Jeong H, Lee D. A protein interaction network associated with asthma. *J Theor Biol.* 2008;252(4):722–31.
- Tomkins JE, Dihanich S, Beilina A, Ferrari R, Ilacqua N, Cookson MR, et al. Comparative protein interaction network analysis identifies shared and

- distinct functions for the human ROCO proteins. *Proteomics*. 2018;18(10): e1700444.
12. Bonham LW, Steele NZR, Karch CM, Manzoni C, Geier EG, Wen N, et al. Protein network analysis reveals selectively vulnerable regions and biological processes in FTD. *Neurol Genet*. 2018;4(5):e266.
  13. Ferrari R, Lovering RC, Hardy J, Lewis PA, Manzoni C. Weighted protein interaction network analysis of Frontotemporal dementia. *J Proteome Res*. 2017;16(2):999–1013.
  14. Rossi G, Redaelli V, Contiero P, Fabiano S, Tagliabue G, Perego P, et al. Tau mutations serve as a novel risk factor for Cancer. *Cancer Res*. 2018;78(13): 3731–9.
  15. Taymans JM, Chartier-Harlin MC. In silico and wet bench Interactomics sheds light on the similarities and differences between human ROCO proteins. *Proteomics*. 2018;18(13):e1800103.
  16. Aranda B, Blankenburg H, Kerrien S, Brinkman FS, Ceol A, Chautard E, et al. PSICQUIC and PSIScore: accessing and scoring molecular interactions. *Nat Methods*. 2011;8(7):528–9.
  17. Braschi B, Denny P, Gray K, Jones T, Seal R, Tweedie S, et al. Genenames.org: the HGNC and VGNC resources in 2019. *Nucleic Acids Res*. 2019;47(D1): D786–D92.
  18. Shannon P. PSICQUIC R package. <https://doi.org/10.18129/B9.bioc.PSICQUIC>. 2020.
  19. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, et al. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Res*. 2013;41(Database issue):D1228–33.
  20. Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, Galeota E, et al. MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res*. 2012; 40(Database issue):D857–61.
  21. UniProt C. UniProt: a hub for protein information. *Nucleic Acids Res*. 2015; 43(Database issue):D204–12.
  22. Harris TW, Arnaboldi V, Cain S, Chan J, Chen WJ, Cho J, et al. WormBase: a modern model organism information resource. *Nucleic Acids Res*. 2020; 48(D1):D762–D7.
  23. Lee RYN, Howe KL, Harris TW, Arnaboldi V, Cain S, Chan J, et al. WormBase 2017: molting into a new stage. *Nucleic Acids Res*. 2018;46(D1):D869–D74.
  24. Alliance of Genome Resources C. Alliance of genome resources portal: unified model organism research platform. *Nucleic Acids Res* 2020;48(D1): D650–D6D8.
  25. Calvo SE, Clauser KR, Mootha VK. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res*. 2016;44(D1):D1251–7.
  26. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–504.
  27. 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(1):25–9.
  28. The Gene Ontology C. The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res*. 2019;47(D1):D330–D8.
  29. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. G:profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res*. 2019;47(W1):W191–W8.
  30. Alanis-Lobato G, Andrade-Navarro MA, Schaefer MH. HIPPIE v2.0: enhancing meaningfulness and reliability of protein-protein interaction networks. *Nucleic Acids Res*. 2017;45(D1):D408–D14.
  31. Hu Y, Vinayagam A, Nand A, Comjean A, Chung V, Hao T, et al. Molecular interaction search tool (MIST): an integrated resource for mining gene and protein interaction data. *Nucleic Acids Res*. 2018;46(D1):D567–D74.
  32. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res*. 2017;45(D1):D362–D8.
  33. Schaefer MH, Fontaine JF, Vinayagam A, Porras P, Wanker EE, Andrade-Navarro MA. HIPPIE: integrating protein interaction networks with experiment based quality scores. *PLoS One*. 2012;7(2):e31826.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

