Patterns

Contextualized Protein-Protein Interactions

Highlights

- We present PPI Context: contextualization of existing literature-curated PPIs

- A resource for filtering PPIs by cell-line information mined from reporting studies

- A fast and flexible pipeline implementing the presented data mining method

Authors

Anthony Federico, Stefano Monti

Correspondence

smonti@bu.edu

In Brief

Literature-curated protein-protein interactions (PPIs) are an essential bioinformatics resource. A major challenge is determining the biological context in which an interaction was observed. Here, we present a data mining method for extracting cell-line information for existing PPIs from reporting studies and make the resulting data available for building disease-centric interaction networks.

CelPress

Descriptor

Contextualized Protein-Protein Interactions

Anthony Federico[1,2,3] and Stefano Monti[1,2,*]
[1]Section of Computational Biomedicine, Boston University School of Medicine, Boston, MA 02118, USA
[2]Bioinformatics Program, Boston University, Boston, MA 02215, USA
[3]Lead Contact
*Correspondence: smonti@bu.edu
https://doi.org/10.1016/j.patter.2020.100153

THE BIGGER PICTURE Existing literature-curated protein-protein interaction (PPI) databases usually aggregate cell-type-agnostic interactions, yet PPIs are dependent on environmental conditions. Thus, new methods and resources for inferring the context in which a PPI is reported will extend their application and use in disease-centric modeling. We expect the resource presented in this article to be of high interest to those querying known interactions of proteins of interest, reconstruction and analyses of molecular interaction networks, and multi-omics data integration approaches.

1 2 3 4 5   **Proof-of-Concept:** Data science output has been formulated, implemented, and tested for one domain/problem

---

SUMMARY

Protein-protein interaction (PPI) databases are an important bioinformatics resource, yet existing literature-curated databases usually represent cell-type-agnostic interactions, which is at variance with our understanding that protein dynamics are context specific and highly dependent on their environment. Here, we provide a resource derived through data mining to infer disease- and tissue-relevant interactions by annotating existing PPI databases with cell-contextual information extracted from reporting studies. This resource is applicable to the reconstruction and analysis of disease-centric molecular interaction networks. We have made the data and method publicly available and plan to release scheduled updates in the future. We expect these resources to be of interest to a wide audience of researchers in the life sciences.

INTRODUCTION

Network biology is an emerging trend in biomedical research that takes a systems-based approach to understanding biological processes and modeling complex disease, whereby interacting molecules—rather than individual genes—are mapped to phenotypic outcomes.[1] An accurate reconstruction of the interactions of the proteome would allow for a detailed understanding of how interacting proteins carry out cellular functions, explain biological phenomena, and predict the consequences of interventions. There is currently a large selection of repositories for protein-protein interactions (PPIs) and an ever-growing number of experimentally observed or computationally predicted interactions. Efforts have emerged, such as the Proteomics Standard Initiative Common Query Interface (PSICQUIC), to aggregate these interactions across various providers, enabling querying of millions of interactions based on a subset of interactors or detection methods.[2,3] This has proved to be an essential resource for network-based analyses, mapping interactome networks, and seeding advanced network models.[4]

Whereas literature-curated interactions often include the experimental assays supporting the interaction, there is less emphasis on describing the biological context (e.g., the cell lines) in which an interaction was assayed *in vitro*. Protein dynamics are context specific and highly dependent on their environment. For example, it was reported that a majority of protein complexes measured in yeast were dependent on environmental conditions.[5] Without context, one ignores the dynamic rewiring of biological networks and assumes PPIs measured across heterogeneous cellular contexts are uniformly relevant to a given biological system under study.[6] This assumption has been shown to be false and that local conditions of observed interactions are important considerations when reconstructing networks for exploring specific biological subsystems.[7] Thus, researchers should consider querying reported interactions relevant to the model under investigation. Previous efforts to infer environmental specificity of PPIs include integration of tissue-specific gene expression information—such as GTEx—whereby two proteins in a PPI that are both expressed above a certain threshold in a given tissue are deemed available for
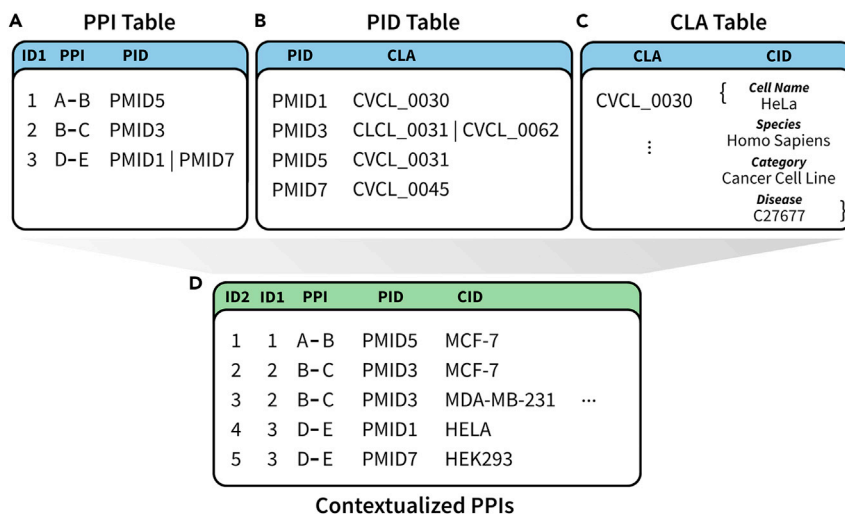
Figure 1. A Graphical Overview

The schematic describes the organization of existing bioinformatics resources to create three mapping tables—(A) the PPI table which maps interactions to reporting publications, (B) the PID table which maps publications to extracted cell lines, and (C) the CLA table which maps cell-line accessions to official cell-line names and associated cell-type information—to generate (D) the presented dataset of contextualized PPIs.

interaction.[8,9] Here we present an alternative approach of utilizing associated cell lines within the original publication of literature-curated PPIs to infer environmental context.

In summary, multiple lines of evidence suggest that the cellular context of reported PPIs is an important factor in determining the relevance of their use toward other biological research efforts. Here we present a data mining method for annotating existing PPI databases with contextual information in an attempt to determine their biological relevance.

RESULTS

To demonstrate our method, we start with interactions from the Human Integrated Protein-Protein Interaction Reference (HIPPIE), a manually curated subset of experimentally detected PPIs from PSICQUIC. To date, HIPPIE contains 391,410 interactions from 41,330 publications sourced from various providers, including IntAct, MINT, BioGRID, HPRD, DIP, BIND, and MIPS.[8] Each entry takes the form of two protein interactors—identified by their encoding gene symbol—and zero, one, or multiple PubMed identifiers (PIDs), in addition to other relevant information such as the types of experimental evidence supporting the interaction. The PIDs link to the original studies in which an interaction was reported. With these PIDs, one can reference the reporting studies to further understand the context in which an interaction was observed, such as the cell lines used to conduct the experiment. This information is valuable in determining which interactions are relevant to the context of one's own biological study, yet the manual curation of cellular context for hundreds of thousands of interactions is time intensive.[10]

Here we present an approach to automating this process through the use of two additional resources: NCBI's PubTator and ExPASy's Cellosaurus. PubTator is a text mining tool for literature curation that extracts bioconcepts (e.g., gene, disease, chemical, mutation, species, and cell line) from text and has pre-processed and annotated roughly 3 million full-text PubMed Central articles.[11,12] Cellosaurus describes many of the available cell lines used in biomedical research. It provides unique cell-line accessions (CLA) for more than 100,000 cell lines, which can be

mapped through a controlled vocabulary to various cell-line attributes, including an official cell identifier or name, category, species of origin, etc., as well as synonyms and common spelling variations or misspellings.[13]

Using these resources, we created a simple method for annotating a large collection of PPIs with cell-type contextual information. The basic workflow consists of mapping PPIs and their supporting publications reported in HIPPIE to cell lines reported in these studies extracted via PubTator to official cell-line identifiers and various cell-line attributes described in Cellosaurus. The idea is that the originating article for an interaction will likely describe one or more cell lines used in the study (e.g., in the methods section) and that these cell lines may have been used to carry out the experiment itself or are at least relevant to the interactions reported. By extracting this information to annotate existing interactions, we can filter interactions by cell-line context based on the biological system/state we are interested in.

We developed a fast and reproducible pipeline for annotating literature-curated PPIs with associated PIDs. The pipeline can efficiently annotate hundreds of thousands of interactions in a few minutes. It does so by fetching and processing the raw bulk data of HIPPIE, PubTator, and Cellosaurus and generating three mapping tables (Figures 1A–1C). The first table (PPI table) maps interactions to reporting publications (one to many), the second (PID table) maps publications to extracted bioconcepts (cell line) and cell-line accession numbers (one to many), and the third (CLA table) maps cell-line accessions to official cell-line names and associated cell-type information (one to one). Due to the multi-mapping nature of the data, original interactions can be supported by multiple studies, each of which could report multiple cell lines. Therefore, we create an entry in the contextualized dataset for each combination observed. Using these tables, the pipeline executes the routine described in Figure 2 to create the dataset of contextualized PPIs (Figure 1D).

Interactions are ignored if they do not have supporting publications or have publications where cell lines are not reported or cannot be extracted. The result is a data frame of original PPIs with additional columns, including cell name, category, species, etc., for all annotatable PPIs (contextualized PPIs). This format is compatible with the primary use case envisioned for the data: building interaction networks by filtering on one or more cell types relevant to a biological setting or question of interest. Application of this routine to the latest versions (as of June

```
   Data: PPI, PID, CLA tables
   Result: Contextualized PPIs
1  initialization;
2  for ppi in PPI do
3      match ppi to pids in PPI;
4      if pids exist then
5          for pid in pids do
6              match pid to clas in PID;
7              if clas exist then
8                  for cla in clas do
9                      match cla to cid in CLA;
10                     add ppi, pid, cid to Contextualized PPIs;
11                 end
12             end
13         end
14     end
15 end
```

Figure 2. The Main Routine Behind PPI Context

The pseudocode includes the main routine executed in the data pre-processing pipeline for creating contextualized PPI entries from the three mapping tables. The tool can be downloaded from GitHub, which includes example commands for installing the required Python dependencies and fetching the raw data.

2020) of the previously described resources started with 391,410 original interactions and found at least one publication for 385,740 interactions. This resulted in a final contextualized dataset of 1,016,726 unique interaction/cell line pairs across 2,012 unique cell lines, originating from 247,065 interactions. We found that a majority of the contextualized interactions were derived from papers reporting human-derived cancer cell lines (Figure 3B). A majority of the reported interactions indeed come from commonly used cell lines such as HeLa and HEK293[14] (Figure 3A).

Despite a bias toward popular cell lines, there still remain sufficient interactions for many less common cell lines to perform

disease-centric modeling through filtered PPIs. For example, we reconstructed a molecular interaction network using PPIs from the breast cancer cell lines MCF-7 and MDA-MB-231,[15] resulting in a breast cancer-centric network of 4,645 nodes and 9,015 edges. By deriving PPIs annotated with breast cancer cell lines, we would expect these interactions to be experimentally validated in said cell lines or at least reported in a context relevant to breast cancer. Thus, this network should exhibit known properties of a breast cancer model better than non-breast cancer networks. To test this hypothesis, we assessed the network's ability to rediscover known disease genes through network propagation and compared it with the results from networks generated with other top cell lines in the dataset.[16]

To this end, we used the random walk with restart (RWR) algorithm, a popular method for network propagation.[17] RWR measures the proximity of nodes in a graph to a given seed or set of seed nodes. The algorithm randomly traverses the graph starting from seed nodes and moving with a given restart probability.[18] It exploits the disease module hypothesis, which postulates that disease genes are likely to be close to one another in a given network.[19] Hence, highly traversed nodes (other than the disease-gene seeds) are classified as disease genes with high probability. Using this algorithm, we tested if the breast cancer network was more efficient at recovering known breast cancer disease genes. We queried 538 breast cancer genes from DisGeNET[20] and adopted a standard random resampling approach, whereby the 538-gene set was randomly split in half, with half used as the seed set and recovery scored on the left-out half as the area under the receiver operating characteristic, with the process repeated 100 times.

We compared the recovery scores of the breast cancer network with those of networks built from interactions annotated with the other most frequent 30 cell lines. For each network, we ranked and compared the mean recovery score across the 100 iterations. We found the breast cancer network to outperform

A

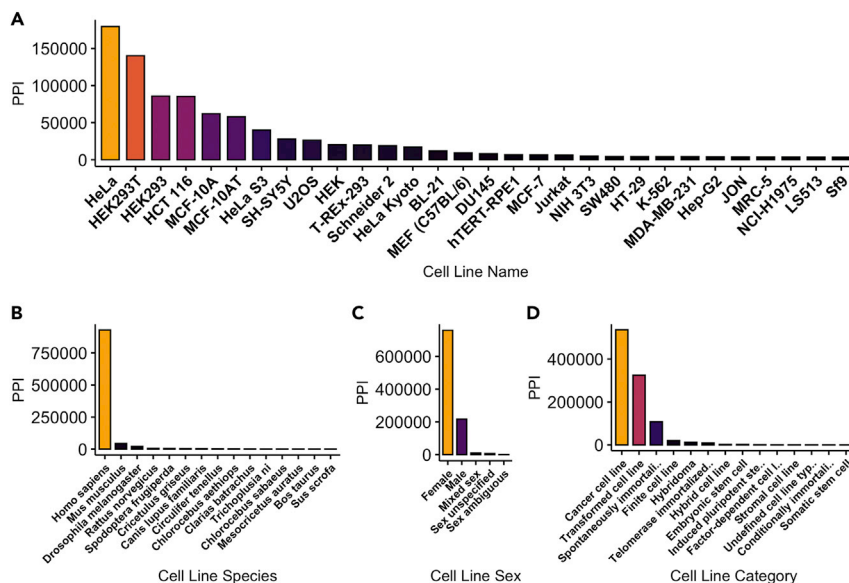**Figure 3. Summary of Contextualized PPIs**

The processed dataset provides cell-line information for each contextualized PPI. The summary plots compare the frequencies of annotations for contextualized PPIs, including (A) the most frequently annotated cell-line names, (B) cell-line species of origin, (C) cell-line sex, and (D) cell-line category. The majority of annotations were human cancer-derived cell lines.

Table 1. Network Propagation of Disease Genes

Cell name	Nodes	Edges	Density	Assortivity	Mean AUROC	SD	Delta
BRCA	4,645	9,015	8.4×10^{-4}	-1.8×10^{-1}	0.693	0.028	
Breast Expressed	10,850	180,342	3.1×10^{-3}	-6.8×10^{-2}	0.691	0.014	0.002
HEK293	11,069	79,207	1.3×10^{-3}	-7.4×10^{-2}	0.664	0.017	0.029
HEK293T	13,884	116,709	1.2×10^{-3}	-8.2×10^{-2}	0.659	0.014	0.034
HeLa	13,824	149,136	1.6×10^{-3}	-5.4×10^{-2}	0.658	0.015	0.034
MEF (C57BL/6)	4,189	8,689	9.9×10^{-4}	-2.2×10^{-1}	0.643	0.023	0.049
DU145	3,219	8,075	1.6×10^{-3}	-5.5×10^{-1}	0.637	0.031	0.055
Jurkat	2,467	5,953	2.0×10^{-3}	-4.1×10^{-1}	0.637	0.034	0.056
HCT 116	11,936	82,956	1.2×10^{-3}	1.1×10^{-2}	0.633	0.018	0.060
Schneider 2	4,228	18,745	2.1×10^{-3}	-7.2×10^{-2}	0.632	0.028	0.060
U2OS	6,667	25,309	1.1×10^{-3}	-2.7×10^{-1}	0.631	0.020	0.061
SW480	1,763	4,316	2.8×10^{-3}	-4.0×10^{-1}	0.629	0.029	0.063
MCF-10A	11,469	61,722	9.4×10^{-4}	-3.7×10^{-2}	0.627	0.018	0.066
Hep-G2	1,304	3,728	4.4×10^{-3}	-3.0×10^{-1}	0.626	0.046	0.066
BL-21	5,135	11,433	8.7×10^{-4}	-2.0×10^{-1}	0.625	0.021	0.068
NCI-H1975	1,295	3,546	4.2×10^{-3}	-4.3×10^{-1}	0.618	0.041	0.074
LS513	1,246	3,486	4.5×10^{-3}	-4.4×10^{-1}	0.603	0.038	0.089
NIH 3T3	2,914	4,806	1.1×10^{-3}	-2.9×10^{-1}	0.603	0.030	0.090
HT-29	1,693	4,219	2.9×10^{-3}	-3.7×10^{-1}	0.601	0.034	0.092
HeLa Kyoto	4,992	16,901	1.4×10^{-3}	-1.1×10^{-1}	0.601	0.022	0.092
MCF-10AT	11,112	57,754	9.4×10^{-4}	-1.4×10^{-1}	0.597	0.018	0.096
K-562	1,922	3,687	2.0×10^{-3}	-3.8×10^{-1}	0.596	0.036	0.096
MRC-5	2,015	3,538	1.7×10^{-3}	-3.7×10^{-1}	0.583	0.022	0.109
T-REx-293	5,395	19,558	1.3×10^{-3}	-4.1×10^{-1}	0.581	0.022	0.111
HeLa S3	8,756	39,157	1.0×10^{-3}	1.0×10^{-1}	0.580	0.019	0.112
Sf9	1,819	3,159	1.9×10^{-3}	-2.0×10^{-1}	0.575	0.029	0.118
JON	1,354	3,629	4.0×10^{-3}	-4.1×10^{-1}	0.574	0.036	0.119
SH-SY5Y	8,422	27,864	7.9×10^{-4}	-1.5×10^{-1}	0.571	0.023	0.122
HEK	6,161	19,569	1.0×10^{-3}	-2.2×10^{-1}	0.546	0.022	0.147
293T/AT1	1,994	3,315	1.7×10^{-3}	-3.1×10^{-1}	0.518	0.028	0.174
hTERT-RPE1	2,553	6,577	2.0×10^{-3}	-4.2×10^{-1}	0.486	0.035	0.206

A comparison of the recovery of breast cancer disease genes in a breast cancer-centric network and networks built from non-breast cancer interactions, in addition to measured graph properties, including nodes, edges, density, and assortivity. Delta values measure the difference in mean AUROC (area under the receiver operating characteristic) of 100 repeats between the BRCA network and the rest.

networks built from non-breast cancer interactions at rediscovering known breast cancer genes (Table 1). There are many cell lines compared with which BRCA performs significantly better, and these would be primary candidates for removal when reconstructing a breast cancer-centric interaction network. Although encouraging that BRCA outranks other networks, it performs only marginally better than networks built from commonly used non-breast cancer cell lines such as HEK293 and HeLa. This is likely due to inspection bias toward well-studied disease genes known to play a role in multiple cancers (e.g., TP53) and commonly assayed in these well-established and widely adopted cell lines.[21–23] In addition, we tested a network based on PPIs filtered for breast tissue expression (Breast Expressed) and found it had a recovery score roughly equal to that of BRCA, suggesting that both methods—literature mining and tissue expression—of inferring context are similar and could be used in complementary ways.

We performed an additional test to determine if interactions were relevant to their derived cell-line annotations. In particular, we selected two genes known to be highly specific to breast cancer, BRCA1 and BRCA2, and counted the number of PPIs (interactions) involving one or both of these genes and annotated with breast cancer cell lines (MCF-7 and MDA-MB-231) compared with those annotated with one of the other cell lines. The expectation was that breast cancer annotated interactions should have a significantly higher proportion of interactions involving BRCA1 and/or BRCA2 than non-breast cancer annotated interactions. Indeed, relative to the total number of interactions in each network, we found BRCA1/2 interactions much more likely to be annotated with breast cancer cell lines, supporting our assumption that the method is extracting relevant cell-line interactions and that other PPIs annotated with these cell lines are also likely relevant to breast cancer (Table 2).

Table 2. Targeted Enrichment by Cell Line

Cell name	Interactions	BRCA1/2	%	p	FDR
MDA-MB-231	4,185	152	0.036	2.6×10^{-133}	7.9×10^{-132}
MCF-7	6,577	67	0.010	7.0×10^{-26}	1.0×10^{-24}
MCF-10A	62,019	174	0.003	1.9×10^{-5}	1.9×10^{-4}
MCF-10AT	57,832	163	0.003	2.8×10^{-5}	2.1×10^{-4}
U2OS	26,221	83	0.003	8.5×10^{-5}	5.1×10^{-4}
BL-21	11,981	27	0.002	3.3×10^{-1}	1.0
MEF (C57BL/6)	9,265	21	0.002	3.4×10^{-1}	1.0
NIH 3T3	5,031	10	0.002	5.7×10^{-1}	1.0
K-562	4,197	7	0.002	7.5×10^{-1}	1.0
JON	3,641	6	0.002	7.5×10^{-1}	1.0
HCT 116	85,366	164	0.002	8.0×10^{-1}	1.0
Sf9	3,414	4	0.001	9.2×10^{-1}	1.0
Hep-G2	3,854	2	0.001	1.0	1.0
SW480	4,351	1	0.000	1.0	1.0
DU145	8,123	4	0.000	1.0	1.0
Jurkat	6,245	1	0.000	1.0	1.0
HeLa	179,407	285	0.002	1.0	1.0
HeLa S3	39,911	35	0.001	1.0	1.0
SH-SY5Y	27,964	17	0.001	1.0	1.0
T-REx-293	19,912	6	0.000	1.0	1.0
HEK	20,382	4	0.000	1.0	1.0
HeLa Kyoto	17,093	1	0.000	1.0	1.0
HEK293T	140,112	132	0.001	1.0	1.0
HEK293	85,737	45	0.001	1.0	1.0
Schneider 2	18,789	0	0.000	1.0	1.0
hTERT-RPE1	6,739	0	0.000	1.0	1.0
HT-29	4,271	0	0.000	1.0	1.0
MRC-5	3,597	0	0.000	1.0	1.0
NCI-H1975	3,550	0	0.000	1.0	1.0
LS513	3,487	0	0.000	1.0	1.0

A comparison of the annotation of BRCA1/2-interactions across the most frequent cell lines. The significance was computed with a hyper-geometric test for over-representation and p values were adjusted for multiple comparisons using the Benjamini-Hochberg method (FDR).

Reproducibility and Extension

In addition to hosting the presented data online, we also developed a command line interface utility for downloading and processing the raw data for reproducing our results and extending the method. The only prerequisite is access to a machine with Python installed. The repository can be cloned from GitHub to any local directory in addition to installation of the required Python dependencies through the following commands:

    $ git clone https://github.com/montilab/ppi-context
    $ pip install -r requirements.txt

The full pipeline, which includes downloading and processing of raw data, can be run through a single command:

    $ python ppictx.py –download –run

Given the constantly evolving nature of the repositories our approach uses as its input this pipeline is an essential contribution. The pipeline can optionally take as arguments file paths to the expected raw files locally stored if users wish to process alternative versions of the data. This pipeline is readily extensible to annotating interactions with additional cell-line information available on Cellosaurus as well as text mining methods alternative to PubTator.

DISCUSSION

PPI databases are an important bioinformatics resource. Existing literature-curated databases usually represent cell-type-agnostic interactions that are not sufficiently specific to a domain of study to significantly improve the predictive accuracy and specificity of the learned models.[24,25] Due to the dynamic rewiring of biological networks in different cellular states and environments, an ability to pre-filter interactions by individual cell lines and types will increase confidence that a given interaction is present in a given biological context and will enhance our ability to model these systems. Here we present a method for annotating existing and future literature-curated PPI databases with cell-contextual information. We also generated a cleaned dataset for general use, immediately applicable to support typical PPI-

based analyses with additional context, such as querying known interactions of proteins of interest, reconstruction and analyses of molecular interaction networks, and multi-omics data integration approaches.

Our approach assumes that cell lines extracted from reporting articles can be used to infer the biological context in which an interaction was detected, rather than identifying the natural cellular context in which a PPI would take place. More specifically, we expect extracted cell lines to have been used for experimental assays that either directly observe or are relevant to the reported interaction. Under this assumption, extracted cell lines can be used to infer the disease or tissue relevance of annotated interactions. We use breast cancer as a primary example to support these assumptions in finding that breast cancer-centric PPI networks are enriched for breast cancer-relevant interactions and display expected network properties such as the proximity of known breast cancer-disease genes.

A limitation of the method was the availability of interaction-associated publications on PubMed pre-mined with PubTator. We were able to extract at least one cell line for 6,146 of potentially 41,329 articles, leaving room for improvement. However, since many of these articles report multiple interactions, the majority of original interactions were still annotated with at least one cell line. For example, the most frequent article (PubMed: 28514442)[26] was associated with 56,297 interactions. In addition, this study assayed the observed interactions in HEK293T cells, exemplifying the disproportionate frequencies at which interactions are annotated with popular cell lines such as HEK. This relates to a third limitation, which is that many PPIs are tested in cell lines such as HEK, due to a high transfection efficiency rather than their relevance to the interrogated interaction.[27] However, the primary purpose of the presented dataset is to provide researchers with an additional tool to make informed decisions about which literature-curated PPIs are relevant to their research needs. Some PPIs (e.g., those from high-throughput assays in HEK-related cells) may not provide contextual information researchers can leverage, while others (e.g., PPIs from many small-scale studies annotated with cell lines not primarily used as expression vectors) are likely to be more applicable. Last, direct comparisons of distinct context-specific networks are limited by the unequal and unknown sets of tested interactions per cell line (e.g., some cell lines are overstudied while some are understudied). Therefore, the absence of PPI_x in a given cell line could be because those proteins were interrogated and found not to interact, or their interaction was never tested under those conditions; we cannot distinguish between these cases.

Despite these limitations, the major use case envisioned for the presented contextualized PPIs—building interaction networks relevant to a biological system of interest—will serve as an important resource to researchers. Furthermore, the contextualized dataset contains over 100 cell lines with at least 500 interactions each, facilitating an important filtering of non-relevant interactions and the application toward meaningful analyses—as exemplified by our breast cancer network—for a variety of research domains. As PPI resources continue to grow in size, so too will the contextualized dataset, as we plan to release scheduled updates of the data. In addition, we expect these annotations to improve as more full-text articles become available on PubMed Central and text mining resources such as PubTator improve and grow in coverage of available articles.

Conclusion

We use existing literature-curated PPI databases and available text mining resources to annotate interactions with cell-contextual information. The contextualized dataset is freely available and ready for use immediately in network-based analyses.

EXPERIMENTAL PROCEDURES

Resource Availability

Lead Contact

Further information and requests for data or additional code should be directed to and will be fulfilled by the lead contact, Anthony Federico (anfed@bu.edu).

Materials Availability

This study did not generate any reagents or materials.

Data and Code Availability

The presented data and method are freely available online. The processed data are hosted on GitHub in addition to the source code for raw data fetching and pre-processing, which is implemented in Python and compatible with all major operating systems. We have also provided comprehensive documentation with code examples for working with the processed data.

  Repository: github.com/montilab/ppi-context
  Commit: 81e31020e6e4244ec23065c72d1fe614256b6391
  Documentation: montilab.github.io/ppi-context
  Operating systems: Linux, OS X, Windows
  Programming languages: R, Python
  License: GNU GPLv3

AUTHOR CONTRIBUTIONS

Conceptualization, A.F. and S.M.; Methodology, A.F. and S.M.; Data Curation, A.F.; Software, A.F.; Formal Analysis, A.F. and S.M.; Visualization, A.F.; Writing – Original Draft, A.F.; Writing – Review & Editing, A.F. and S.M.; Funding Acquisition, S.M.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Yadav, A., Vidal, M., and Luck, K. (2020). Precision medicine — networks to the rescue. Curr. Opin. Biotechnol. *63*, 177–189.

2. Aranda, B., Blankenburg, H., Kerrien, S., Brinkman, F.S.L., Ceol, A., Chautard, E., Dana, J.M., De Las Rivas, J., Dumousseau, M., Galeota, E., et al. (2011). PSICQUIC and PSISCORE: accessing and scoring molecular interactions. Nat. Methods *8*, 528–529.

3. del-Toro, N., Dumousseau, M., Orchard, S., Jimenez, R.C., Galeota, E., Launay, G., Goll, J., Breuer, K., Ono, K., Salwinski, L., et al. (2013). A

Patterns
Descriptor

CellPress
OPEN ACCESS

new reference implementation of the PSICQUIC web service. Nucleic Acids Res. *41*, 601–606.

4. Vidal, M., Cusick, M.E., and Barabási, A.L. (2011). Interactome networks and human disease. Cell *144*, 986–998.

5. Celaj, A., Schlecht, U., Smith, J.D., Xu, W., Suresh, S., Miranda, M., Aparicio, A.M., Proctor, M., Davis, R.W., Roth, F.P., et al. (2017). Quantitative analysis of protein interaction network dynamics in yeast. Mol. Syst. Biol. *13*, 934.

6. Califano, A. (2011). Rewiring makes the difference. Mol. Syst. Biol. *7*, 463.

7. Stacey, R.G., Skinnider, M.A., Chik, J.H.L., and Foster, L.J. (2018). Context-specific interactions in literature-curated protein interaction databases. BMC Genomics *19*, 758.

8. Alanis-Lobato, G., Andrade-Navarro, M.A., and Schaefer, M.H. (2017). HIPPIE v2.0: enhancing meaningfulness and reliability of protein-protein interaction networks. Nucleic Acids Res. *4*, 45.

9. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods *5*, 621–628.

10. Baumgartner, W.A., Cohen, K.B., Fox, L.M., Acquaah-Mensah, G., and Hunter, L. (2007). Manual curation is not sufficient for annotation of genomic databases. Bioinformatics *23*, 41–48.

11. Wei, C.H., Kao, H.Y., and Lu, Z. (2013). PubTator: a web-based text mining tool for assisting biocuration. Nucleic Acids Res. *41*, 518–522.

12. Wei, C.H., Allot, A., Leaman, R., and Lu, Z. (2019). PubTator central: automated concept annotation for biomedical full text articles. Nucleic Acids Res. *47*, 587–593.

13. Bairoch, A. (2018). The cellosaurus, a cell-line knowledge resource. J. Biomol. Tech. *29*, 25–38.

14. Hyman, A.H., and Simons, K. (2011). The new cell biology: beyond HeLa cells. Nature *480*, 34.

15. Neve, R.M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F.L., Fevr, T., Clark, L., Bayani, N., Coppe, J.P., Tong, F., et al. (2006). A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell *10*, 515–527.

16. Cowen, L., Ideker, T., Raphael, B.J., and Sharan, R. (2017). Network propagation: a universal amplifier of genetic associations. Nat. Rev. Genet. *18*, 551–562.

17. Tong, H., Faloutsos, C., and Pan, J.Y. (2006). Fast random walk with restart and its applications. ICDM, 613–622.

18. Macropol, K., Can, T., and Singh, A.K. (2009). RRW: repeated random walks on genome-scale protein networks for local cluster discovery. BMC Bioinformatics *10*, 283.

19. Barabási, A.L., Gulbahce, N., and Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. Nat. Rev. Genet. *12*, 56–58.

20. Piñero, J., Queralt-Rosinach, N., Bravo, À., Deu-Pons, J., Bauer-Mehren, A., Baron, M., Sanz, F., and Furlong, L.I. (2015). DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. Database (Oxford) *2015*, bav028.

21. Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N., et al. (2005). Towards a proteome-scale map of the human protein-protein interaction network. Nature *437*, 1173–1178.

22. Edwards, A.M., Isserlin, R., Bader, G.D., Frye, S.V., Willson, T.M., and Yu, F.H. (2011). Too many roads not taken. Nature *470*, 163–165.

23. Skinnider, M.A., Stacey, R.G., and Foster, L.J. (2018). Genomic data integration systematically biases interactome mapping. PLoS Comput. Biol. *14*, e1006474.

24. Marbach, D., Lamparter, D., Quon, G., Kellis, M., Kutalik, Z., and Bergmann, S. (2016). Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. Nat. Methods *13*, 366–370.

25. Schneider, G., Schmidt-Supprian, M., Rad, R., and Saur, D. (2017). Tissue-specific tumorigenesis: context matters. Nat. Rev. Cancer *17*, 239–253.

26. Huttlin, E.L., Bruckner, R.J., Paulo, J.A., Cannon, J.R., Ting, L., Baltier, K., Colby, G., Gebreab, F., Gygi, M.P., Parzen, H., et al. (2017). Architecture of the human interactome defines protein communities and disease networks. Nature *545*, 505–509.

27. Ooi, A., Wong, A., Esau, L., Lemtiri-Chlieh, F., and Gehring, C. (2016). A guide to transient expression of membrane proteins in HEK-293 cells for functional characterization. Front Physiol. *7*, 300.