

OncoPPi Portal

A Cancer Protein Interaction Network to Inform
Therapeutic Strategies



EMORY
UNIVERSITY

**Emory
Chemical Biology
Discovery Center**

2017

Contents

Datasets	2
Protein-protein interaction dataset.....	2
Set of known PPIs.....	3
Domain-domain interactions	3
Mutual exclusivity of genomic alterations.....	3
PPI Essentiality	3
Protein-drug connectivity	3
User Interface	4
OncoPPI Network	4
Interactive Hub protein bar graph	5
The PPI hub panel	5
Gene query panel.....	6
Gene information.....	6
Data export	7
PPI Essentiality	7
Therapeutic connectivity	8
References	9

Recently, we have developed a PPI high-throughput screening platform to detect PPIs between cancer-associated proteins in the context of cancer cells. Here, we present the OncoPPI Portal, an interactive web resource that allows investigators to access, manipulate and interpret a high-quality cancer-focused network of PPIs experimentally detected in cancer cell lines. To facilitate prioritization of PPIs for further biological studies, this resource combines network connectivity analysis, mutual exclusivity analysis of genomic alterations, cellular co-localization of interacting proteins, and domain-domain interactions. Estimates of PPI essentiality allow users to evaluate the functional impact of PPI disruption on cancer cell proliferation. Furthermore, connecting the OncoPPI network with FDA approved drugs and compounds in clinical trials enables discovery of new tumor dependencies to inform strategies to interrogate undruggable targets like tumor suppressors. The OncoPPI Portal serves as a resource for the cancer research community to facilitate discovery of cancer targets and therapeutic strategies.

Datasets

Protein-protein interaction dataset

A set of 83 lung cancer-associated genes was curated based on the analysis of genomic alterations in lung cancer patients and literature searches (1). It includes the major oncogenes (e.g. EGFR, KRAS, BRAF, MYC, PIK3CA, ERBB2) and tumor suppressors (e.g. TP53, STK11, CDKN2A, SMARCA4, or RB1). In addition to major tumor drivers, the protein library is populated with key regulators of oncogenic pathways, such as 14-3-3, MAPK14 (p38), or Beclin1. Structurally, the various classes of proteins (such as transmembrane receptors, kinases, GTPases, transcription factors, or adaptor proteins) with different cellular localization are included in the library.

The network of lung cancer-associated PPIs was generated based on the Time-Resolved Fluorescence Energy Transfer (TR-FRET) screening performed in lung cancer H1299 cells using the cell lysates derived from systematic, pairwise transfection of GST- and Venus-fusion expression vectors containing the genes included in the OncoPPI gene library. Each PPI was tested in triplicate, with both fusion tags along with the corresponding empty vector negative controls. The whole PPI screening was repeated in three independent experiments (1). The positive and negative experimental data for a total of 3,486 protein-protein pairs are available through the OncoPPI Portal.

For each experiment the average TR-FRET signals for the PPI ($SPPI_{G1V2}$, $SPPI_{G2V1}$), GST empty vector control ($SVec_{G1,G2}$), and Venus empty vector control ($SVec_{V1,V2}$) were calculated over the triplicates for each of two tested fusions (GST, Venus). Then, for each fusion the Fold Over Control (FOC) values were calculated with the following equation:

$$FOC = \text{Max}\left(\frac{(SPPI_{G1V2})}{\text{Max}(SVec_{G1}, SVec_{V2})}, \frac{(SPPI_{G2V1})}{\text{Max}(SVec_{G2}, SVec_{V1})}\right)$$

The maximum of the FOC values obtained for the two fusions (GST, Venus) was considered as the final FOC value for a given PPI in a given experiment. Then, the average FOC values were calculated by averaging the FOC values for individual experiments. These values were considered as the final FOC value for a given PPI.

A statistical significance of positive PPIs is estimated in terms of p-values calculated with the permutation test performed according to the following procedure. First, for a given PPI pair, all PPI signals and control signals obtained in all experiments were ranked. The sum of the ranks of the PPI signals was calculated and was used as the test statistic for the permutation test. The permutation tests were repeated 10,000 times. The p-value was calculated with the following equation: $p = (Ns + 1) / 10,001$, where Ns is the number of cases where total ranks of shuffled labels exceed or are equal to that of true label. The p-values adjusted for the multiple comparisons (q-values) were calculated with the Benjamini-Hochberg procedure.

Set of known PPIs

PPI databases included: String (2), BioGrid(3), Intact (4) and GeneMania (5) datasets were used to identify PPIs with reported physical association.

Domain-domain interactions

Structural domains for the proteins were extracted from Pfam (6).

Data for domains involved in physical associations in co-crystallized proteins was extracted from the 3DID database (7).

Mutual exclusivity of genomic alterations

Mutual exclusivity analysis was performed in MatLab with the Cancer Genomics Data Server tool box (8, 9). The complete tumor samples from lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) TCGA Provisional datasets (downloaded 1/2016) were used to analyze the mutual exclusivity of genomic alterations in lung cancer patient samples. Mutations, DNA amplifications and deletions were taken into account. Mutual exclusivity was evaluated in terms of log odds ratio (OD) values calculated as described previously (9). The alterations of two genes were considered as mutually exclusive if $\text{Log(OD)} < 0$.

PPI Essentiality

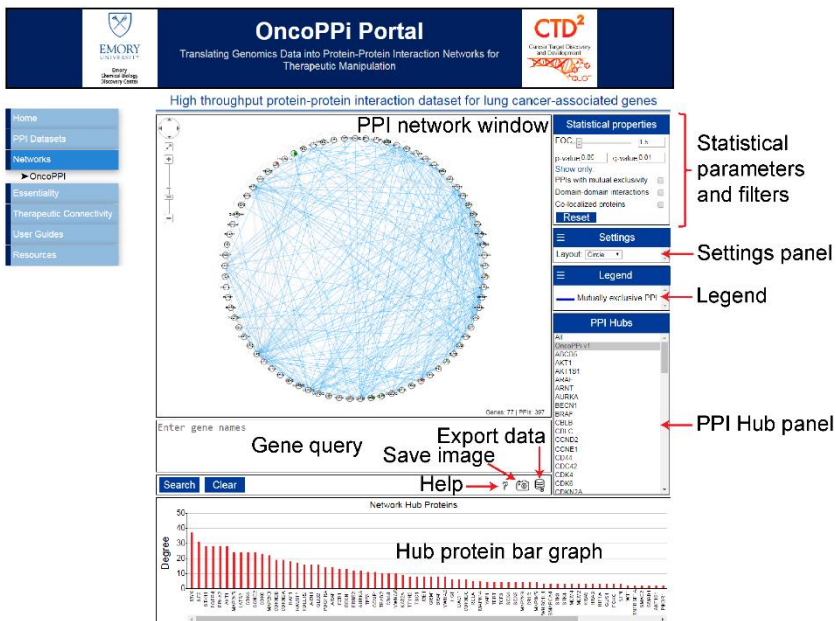
The PPI essentiality values for over 7,900 PPIs in 206 single cell lines were calculated using the MEDICI algorithm as described by Harati et al. (2). 1,548 proteins are available for the analysis. If multiple cell lines selected, the averaged essentiality values are calculated for the given set of cell lines.

Protein-drug connectivity

The OncoPPI network was integrated with the FDA-approved drugs. The drug names and corresponding target proteins were extracted from the Drug Bank (3)

User Interface

OncoPPI Network



The PPI network for a set of selected genes is shown in the **PPI network window**. The network is visualized using a Cytoscape plug-in (4) that allows a user to zoom it in or out as well as to rearrange (move) the nodes. A number of preset network layouts is available through the **Settings panel**. The interacting proteins are connected by grey lines. In the case of mutual exclusivity of corresponding genes in lung cancer the lines are highlighted with blue color.

The **Statistical properties panel** allows to adjust the values of thresholds for statistical significance, including:

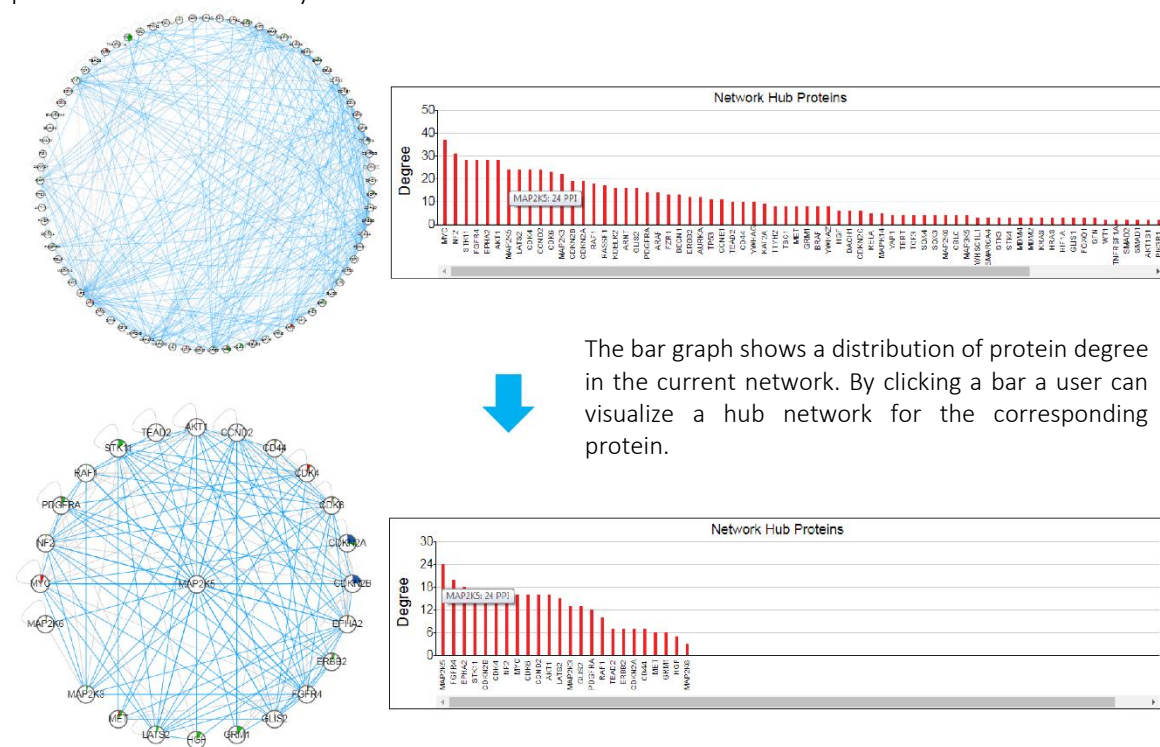
- Fold Over Control (FOC): The $\text{FOC} \geq 1.2$ is considered as a minimum cut-off for positive PPI in original TR-FRET-based PPI screening; $\text{FOC} \geq 1.5$ is used to define high-confident PPIs.
- The permutation test p-value: p-values ≤ 0.05 are considered as statistically significant
- The q-value (p-value, adjusted for multiple comparisons): q-values ≤ 0.01 are used as a threshold for high-confident PPIs.

The panel also provides a set of PPI properties to filter the data, including:

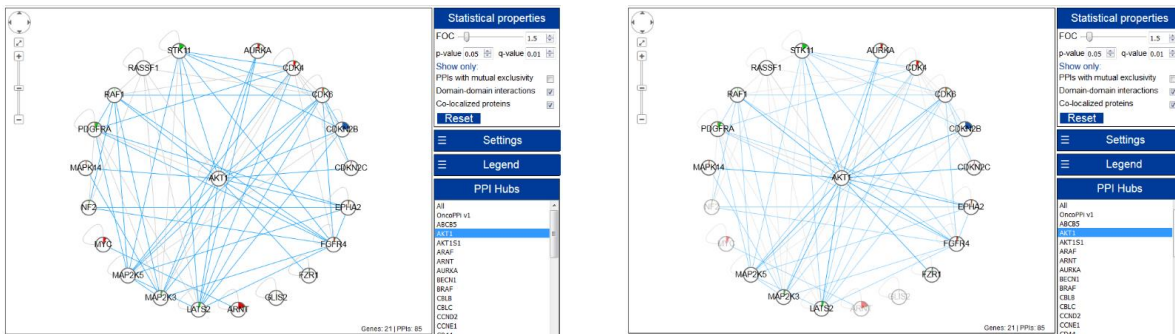
- PPIs reported by other groups: if checked (default) show PPIs identified in String, BioGrid, IntAct, or GeneMania databases. If unchecked, the PPIs detected with the Emory PPI TR-FRET screening are shown.
- PPIs with mutual exclusivity (ME): if unchecked (default) all PPIs included in the dataset are shown. If checked, only the PPIs between the genes with genomic alterations mutually exclusive in lung adenocarcinoma or lung squamous cell carcinoma are shown. PPIs for the genes with ME are indicated with blue lines.
- Domain-domain interactions (DDI): if unchecked (default) all PPIs included in the dataset are shown. If checked, only the PPIs that share structural domains known to interact in crystallized protein-protein complexes (based on 3DID database) are shown.

- Co-localized proteins: if unchecked (default) all PPIs included in the dataset are shown. If checked, only the PPIs for proteins that can co-localize to the same cellular compartments are shown.

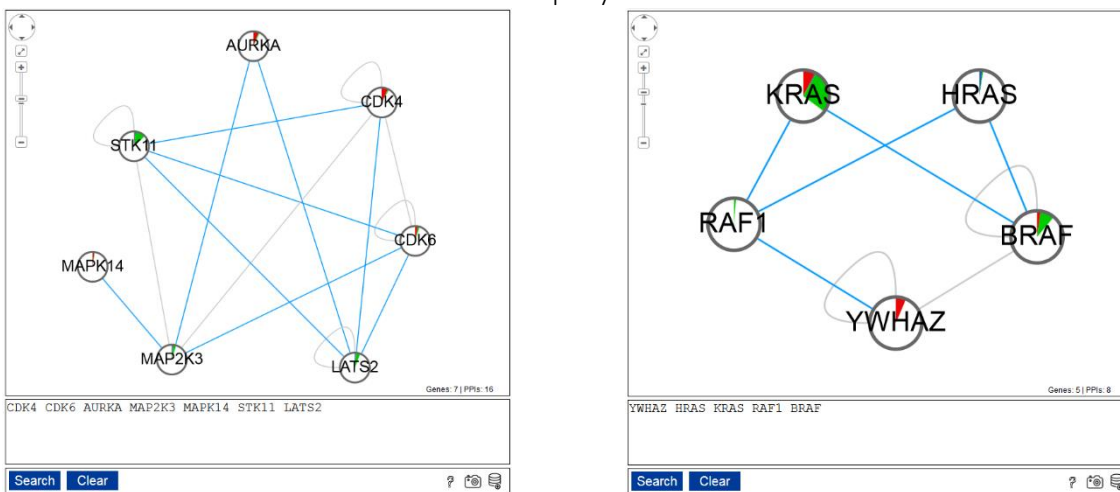
Interactive Hub protein bar graph is shown under the PPI network. The X-axis of the graph shows the proteins included in the given network, and the Y-axis shows the number of protein binding partners (node degree) identified with the selected statistical parameters. The protein order and data shown on the graph changes dynamically when the network and statistical thresholds are changed (e.g. FOC or p-values). Furthermore, all the binding partners of a hub protein and their interconnections are shown by clicking on the corresponding bar on the graph. The graphical representation of protein connectivity allows immediate visualization and ranking of the major hub proteins, and identification of their binding partners for further analysis.



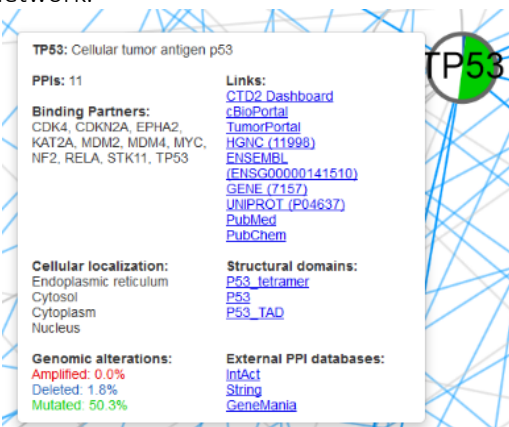
The **PPI hub panel** on the right side contains a list of all proteins available for the current version of the network, including the OncoPPI v1 set defined in ref. 1. On a left-click on a hub protein a sub-network constructed for the corresponding protein and its binding partners will be shown. By left-clicking a protein node in the network the interactions for that particular protein will be highlighted.



Gene query panel at the bottom allows a search for PPIs within specified group of genes. The HUGO gene names separated by a space, comma, or new line can be used to query the dataset. Search button will start the search. Clear button will clear the query.



Gene information window can be accessed with a right-click on the corresponding node in the network.



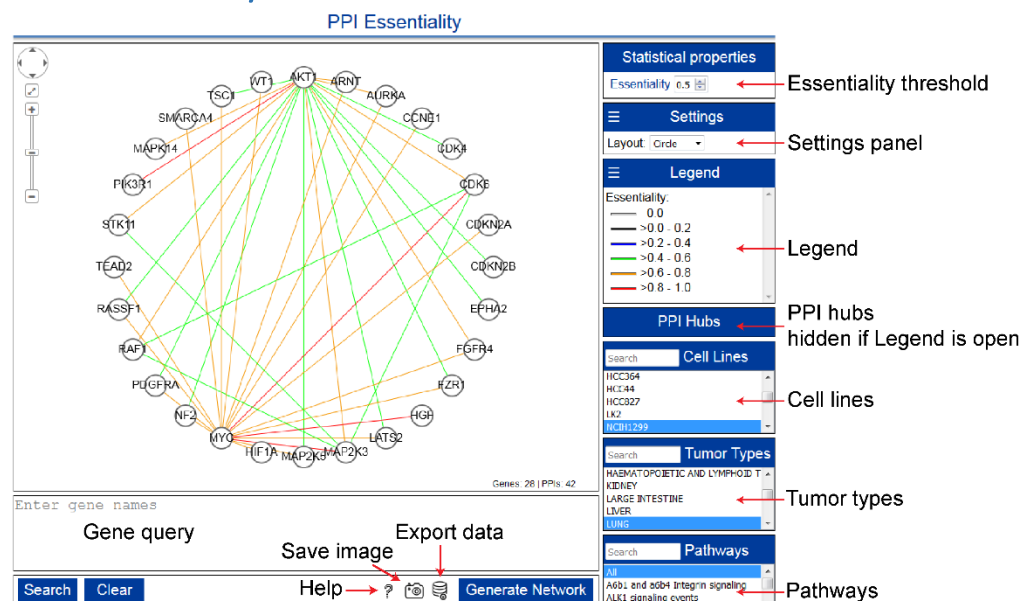
The gene info includes the gene name, the number of PPIs, and a list of the binding partners. Each gene is also annotated with the cellular localization data extracted from Gene Ontology (GO) annotations, structural domains, and with the percent of cases with gene amplification, deletion, or mutation in lung adenocarcinoma (based on TCGA LUAD provisional dataset). The percent of genomic alterations are also visualized with red (amplification), blue (deletion), and green (mutation) sectors inside the gene node circle. To simplify the data mining of structural and biological information associated with the proteins, each protein in the dataset is linked to external databases such as the

CTD2 Dashboard, cBioPortal, TumorPortal, HGCN, Gene, Ensembl, UniProt, PubMed, and PubChem. The PPIs reported in external databases can be easily extracted from the IntAct, String, and GeneMania databases through the direct links to the corresponding protein pages.

Data export. The data associated with selected (shown) PPIs can be exported as a comma separated file (csv) by clicking on the Export PPI data icon. The data includes statistical characteristics of PPI, namely the FOC, p-value, and q-value, indication of mutual exclusivity of alterations of corresponding genes in lung adenocarcinoma or lung squamous carcinoma, predicted interacting domains, co-localized proteins, as well as the indication whether a PPI was reported in String, BioGrid, IntAct, or GeneMania databases.



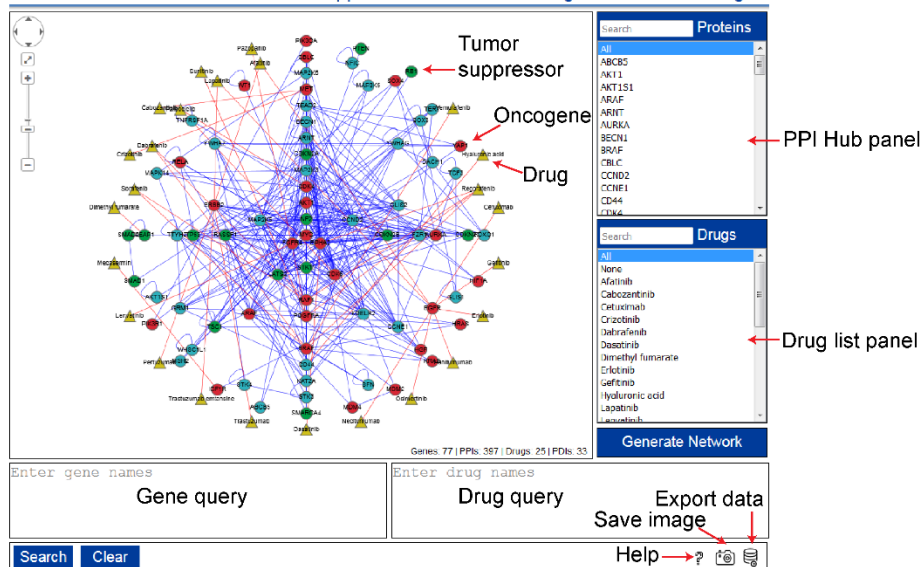
PPI Essentiality



The PPI Essentiality page provides a graphical user interface for the MEDICI algorithm designed to predict the role of individual PPIs in cancer cell death and survival (2). The essentiality values range from 0 (less essential) to 1 (most essential PPIs). Accordingly, the subsets of PPIs with a desired level of essentiality can be visualized by adjusting the essentiality threshold in the Statistical Properties panel. The essentiality values can be calculated either for all 206 available cell lines or for a specific set of cell lines as well as for the individual cell lines. The cell lines can be filtered based on the corresponding tumor type or through the Search option. Similar to the OncoPPI Network page, all proteins included in the dataset are listed in the PPI Hubs panel. In addition, the proteins can be filtered based on their involvement in defined 196 biological pathways identified by the Pathway Interaction Database (5).

Therapeutic connectivity

The OncoPPI network links tumor suppressors to actionable targets and cancer drugs



Therapeutic connectivity page of the OncoPPI Portal provides the interface to explore the connectivity between the FDA approved drugs and PPIs included in the OncoPPI network. The tumor suppressor genes are colored in green, while the oncogenes are highlighted in red. The drugs are shown as yellow triangles. The proteins are connected by blue lines, and the protein-drug connections are highlighted with red lines. All proteins included in the network are listed in the PPI Hub panel. The Gene query and Drug query panels allow a user to search for a subset of proteins and drugs included in the network.

References

1. Li Z, *et al.* (2017) The OncoPPi network of cancer-focused protein-protein interactions to inform biological insights and therapeutic strategies. *Nature Communications* 8:14356, 14310.11038/ncomms14356.
2. Harati S, *et al.* (2017) MEDICI: Mining Essentiality Data to Identify Critical Interactions for Cancer Drug Target Discovery and Development. *PloS one* 12(1):e0170339.
3. Wishart DS, *et al.* (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic acids research* 34(Database issue):D668-672.
4. Franz M, *et al.* (2016) Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics (Oxford, England)* 32(2):309-311.
5. Schaefer CF, *et al.* (2009) PID: the Pathway Interaction Database. *Nucleic acids research* 37(Database issue):D674-679.