Case Studies – 3 – EIF4A2 interactions; input file prepared by ProHits

The dataset was taken from "A cost-benefit analysis of multidimensional fractionation of affinity purification-mass spectrometry samples", Wade H Dunham, et al., Proteomics, 2011 July; 11(13): 2603–2612.

The study aimed at understanding what are the benefits associated with multidimensional fractionation (MudPIT, RP-RP, GeLC) for samples of low to moderate complexity such as AP-MS samples. Here, one of the baits (EIF4A2, a translation initiation factor) was analyzed by first fractionating the sample offline on base-stable reversed phase chromatography followed by online LC-MS/MS. This type of fractionation is referred to by the authors as RP-RP. Essentially, the bait protein was stably expressed in parallel to a FLAG alone empty vector control in clones generated from Flp-In T-REx HEK293 cells (Invitrogen), enabling tetracycline-inducible expression. Expressed proteins were purified on anti-FLAG M2-agarose beads, and eluted with a volatile base; tryptic digests were performed and the resulting peptides were separated on a C18 at high pH. Mass spectrometric analysis was conducted in a data-dependent mode on a Thermo LTQ mass spectrometer equipped with a Proxeon Nanosource and an Agilent capillary pump using a standard low pH acetonitrile-based elution.

The subset of the dataset used as a case study here consisted of two replicates of EIF4A2 protein purifications and two negative controls. In this particular worked example, we are exploring the following points:

1) How the user can generate their own input files for running the workflow 3 of the CRAPome
2) How to select CRAPome controls when there are no perfect matches in the database

The data was processed as described in the original publication (Dunham et al., 2011), using the ProHits LIMS for interaction proteomics (Liu et al., Nature Biotech, 2010; Liu et al., Current Protocols in Bioinformatics, 2012). Briefly, RAW files were converted to mzXML using ProteoWizard. The spectra were searched using the Mascot search engine (one fraction per file) against the human complement of the RefSeq database (V37), with trypsin as the selected enzyme. One missed cleavage site was allowed, and methionine oxidation as well as asparagine/glutamine deamidation were allowed as variable modifications. The results for an entire fractionation experiment were combined and the data further analyzed using the PeptideProphet and ProteinProphet modules of the Trans Proteomics Pipeline. TPP search results were parsed to the relational ("Analyst") of ProHits.

1) Generation of the input files for CRAPome analysis

The ProHits graphical user interface was used to select the desired experiments and associated controls (Figure 1), and view the results as a color-coded matrix in the "Comparison" page (Figure 2). Unfiltered results were downloaded by selecting the "Export(select)" function, which enables the user to create customized lists of bait-prey interactions in a tabular format. As shown in Figure 4, the four columns required by the CRAPome can be selected. Mapping is as follows:

CRAPome "**Bait Name**": ProHits "Bait Gene Name", which is the Official Gene Symbol, followed by the epitope tag in parenthesis (this has to be removed in the final CRAPome input sheet for proper mapping)

CRAPome "**AP Name**": ProHits "Sample ID", which in ProHits is a unique identifier describing a sample. Note that many other identifiers from ProHits (e.g. Sample Name, Experiment ID, etc.) can also be used.

CRAPome "**Prey Name**": ProHits "Protein Gene Name", which is the mapping to an official gene symbol of the protein identifiers for the "hits" or "preys". Note that other identifiers are also available, e.g. "Protein Gene ID", which maps to Entrez Gene ID at NCBI, and "Protein Acc" which is here RefSeq Protein.

CRAPome "**Spectral Counts**": ProHits "Total Number Peptide".

From ProHits, the downloading in csv format generates a table as in Figure 4. From there, the only modifications required to create the CRAPome input are: 1) remove the header and other information above the table; 2) rename the bait so that it matches exactly the Official Gene Symbol (here, EIF4A2(N-Flag) becomes EIF4A2); 3) Rename the FLAG controls "CONTROL" in the first column. This led to the format in Figure 5: the file is now ready for uploading to the CRAPome.

2) Selection of CRAPome controls and data analysis

After logging into the CRAPome database, and selecting "*H. sapiens*" as an organism, workflow 3 was selected. First, matching controls were narrowed down using the CRAPome filters which include the cell line (HEK293), epitope tag (FLAG), affinity approach (M2 Anti FLAG), affinity support (agarose). There were multiple samples which matched these criteria. Unfortunately, no controls exactly matched the experimental procedures used to generate this data, which included RP-RP fractionation. However, we reasoned that we could select from these prefiltered samples two sets of controls from the CRAPome as follows: the first set selected involves MudPIT fractionation of the samples, which according to the Dunham et al., manuscript provided results similar to those of the RP-RP fractionation in terms of spectral counts. 11 such samples were available (10 with protocol 56, one with protocol 40), and were all selected. We then selected 9 additional controls (for a total of 20 CRAPome control) by sorting the remaining controls by number of identified spectra, and selecting the 9 highest counts (7/9 are from protocol 26), thereby decreasing the chances to miss some of the lower abundance contaminants from our analysis (Figure 6). Once the controls were selected, we pressed "Next" to proceed to the upload of the user data, prepared as described above.

Uploading the user data (by browsing through our files and pressing "submit" once the desired file was found) led to the appearance of a left panel "User data" alongside the right panel "CRAPome controls" in which our files were correctly uploaded (Figure 7). We selected to proceed without excluding any of the data for subsequent analysis. Clicking "Next" enabled us to select the filtering parameters, and first elected to perform the quick empirical Fold Change calculations (FC-A and FC-B), by using all controls

available (CRAPome and user controls; see Figure 8). Under these conditions, FC-A and FC-B perform quite similarly, as can be seen in the analysis page (Figures 9, 10). In particular, most of the interactions present in iRefIndex are captured by both scoring schemes. We next tested the effects of including or not the user controls in the analysis, or omitting the CRAPome controls. To do so, we applied FC-B as shown in Figure 8, but used only CRAPome controls or only user controls for FC-A calculation. Omitting the user controls had essentially no effect on the scoring (not shown), but including only the user controls did result in a moderate spreading of the scores, as shown in Figure 11.

Note that most of the high scoring protein pairs not annotated in iRefWeb are in fact known interaction partners for the eIF4A paralogous protein family that have simply escaped direct interaction assessment with EIF4A2 and/or annotation in the repositories. These predominantly comprise several eIF3 components (e.g. EIF3B, EIF3C, EIF3E, EIF3G, EIF3J). Other very high scoring interaction partners which have not been characterized for their interaction with EIF4A2 include the poorly characterized proteins PRRC2A, PRRC2B and PRRC2C. While they have no established role in translation initiation, it is noteworthy that PRRC2A was identified as an interaction partner with eIF3 components in a previous study (Sowa et al., Cell, 2009), suggesting that they may be part of functional translation initiation complexes.

Figures:

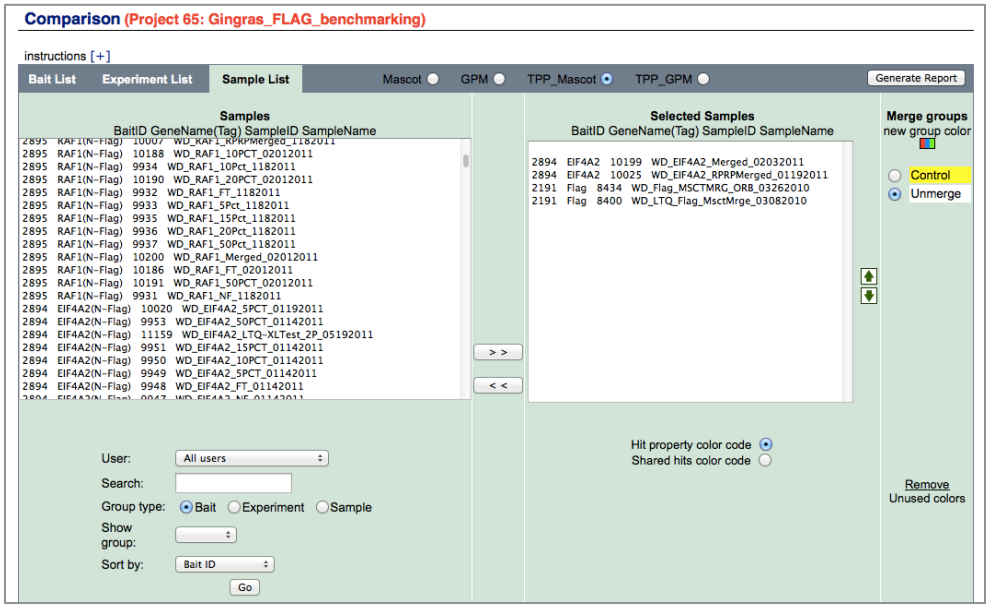


Figure 1. Selection of the Samples from the ProHits Analyst module.

Figure 2. Visualization of the spectral counts matrix (shades of blue) for the selected samples. Columns are the different samples, rows are identified proteins. At the top of the table are links to different export options. Here we will choose the “Export(select)” button.

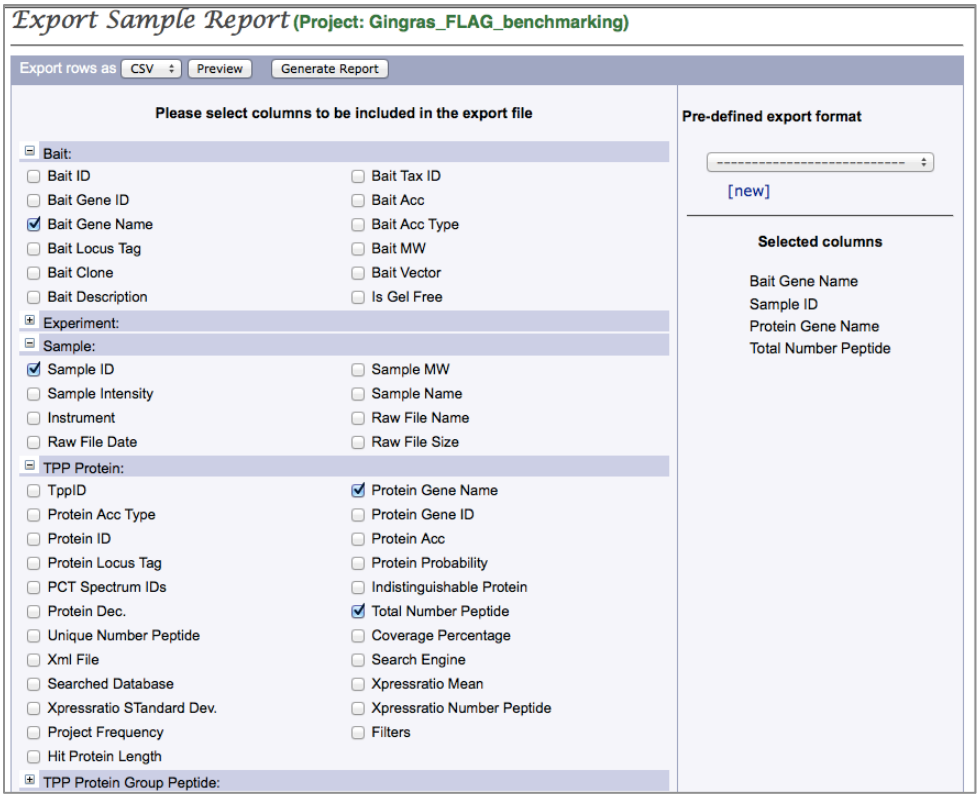


Figure 3. Parameter selection from the "Export(select)" pop-up window. Select the columns in the same ordered desired in the output. Pressing "Generate Report" will download the selected data in the format desired (here, a comma separated values table).

	A	B	C	D
1				
2	Sample report			
3	SearchEngine: TPP_Mascot			
4	Project name: Gingras_FLAG_benchmarking			
5	Created by: Anne-Claude Gingras			
6	Creation date: 2013-03-31			
7				
8	Bait Gene Name	Sample ID	Protein Gene Name	Total Number Peptide
9	EIF4A2 (N-Flag)	10199	DHX9	7
10	EIF4A2 (N-Flag)	10199	EIF4G3	57
11	EIF4A2 (N-Flag)	10199	NPM1	6
12	EIF4A2 (N-Flag)	10199	XRCC5	2
13	EIF4A2 (N-Flag)	10199	EIF4G2	24
14	EIF4A2 (N-Flag)	10199	PRRC2C	12
15	EIF4A2 (N-Flag)	10199	SRRM2	3
16	EIF4A2 (N-Flag)	10199	RPS18	27
17	EIF4A2 (N-Flag)	10199	WDR77	4
18	EIF4A2 (N-Flag)	10199	HNRNPA1	5

Figure 4: Unmodified view of the downloaded table from ProHits. A few modifications are needed prior CRAPome upload. 1) The header and additional information at the top of the page should be deleted; 2) The Bait Gene Name should be matched exactly to the Official Gene Name; 3) the user controls, if any, should be labeled simply CONTROL in the first column.

	A	B	C	D
1	EIF4A2	10199	ABCC1	2
2	CONTROL	8400	ABCC5	4
3	EIF4A2	10025	ABCC5	1
4	EIF4A2	10199	ABCE1	3
5	EIF4A2	10199	ABCF1	2
6	EIF4A2	10199	ACTA2	16
7	CONTROL	8434	ACTA2	1
8	CONTROL	8434	ACTB	48
9	CONTROL	8400	ACTB	47
10	EIF4A2	10025	ACTB	6

Figure 5. User data table ready for upload to the CRAPome.

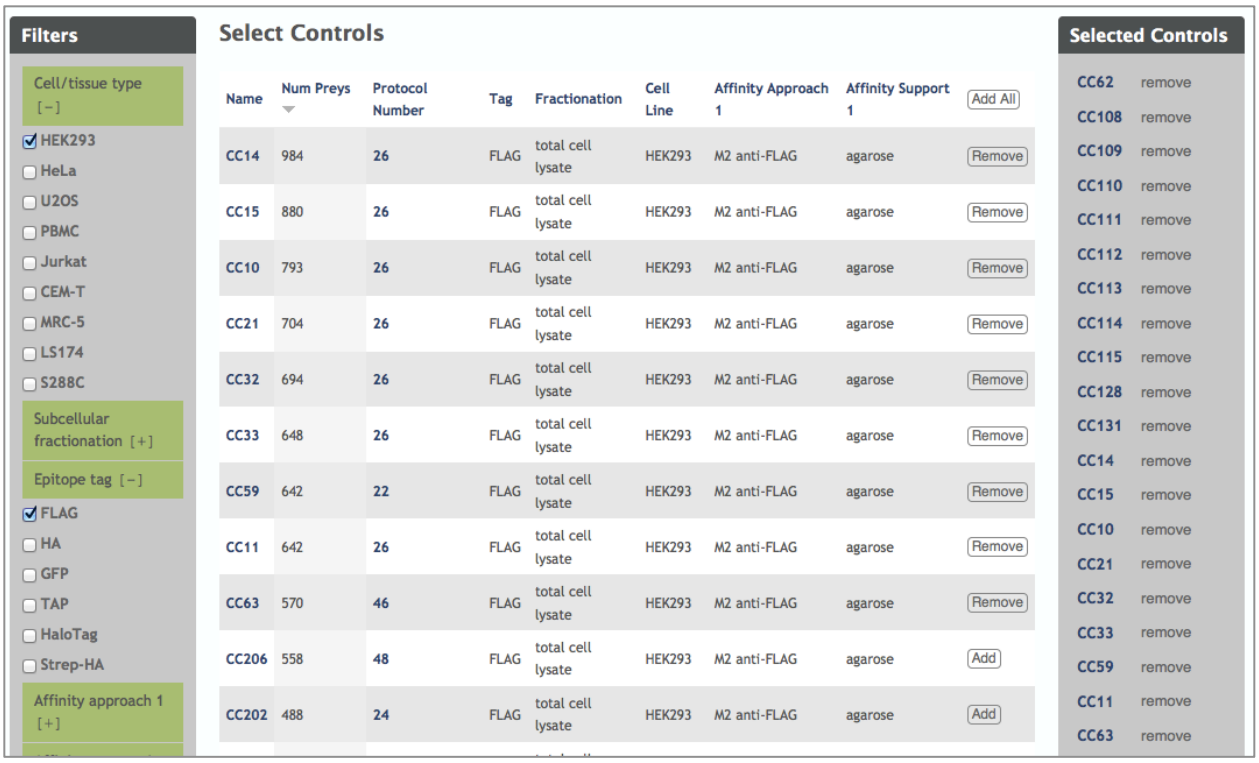


Figure 6. Selection of the CRAPome controls for filtering. Not shown: MudPIT experiments from the same cell line, and affinity purification system were first added (n = 11); here, we selected 9 additional controls by sorting the preselected CRAP controls by decreasing prey abundance and choosing the 9 most abundant.

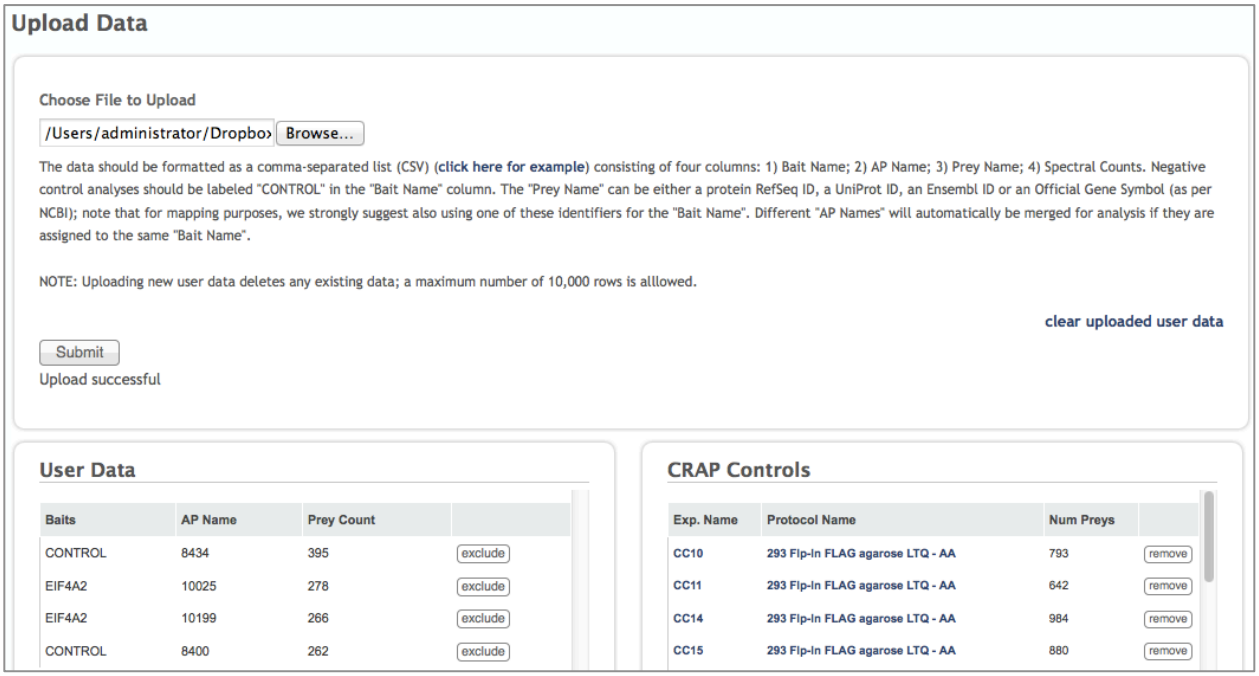


Figure 7. Upload of the user data for analysis.

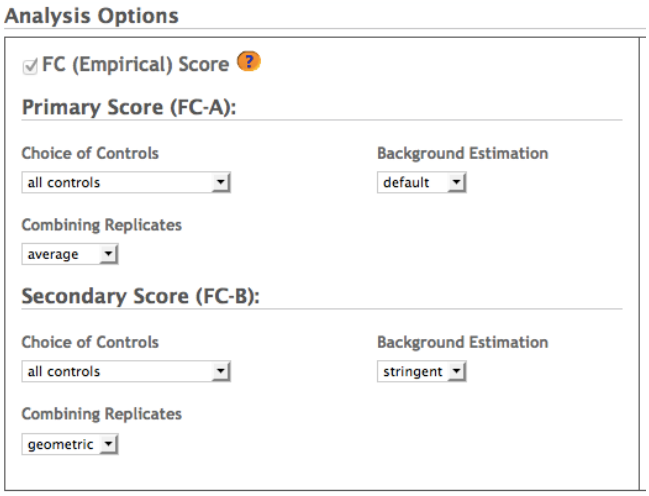


Figure 8. Fold Change parameters for rapid analysis of the data.

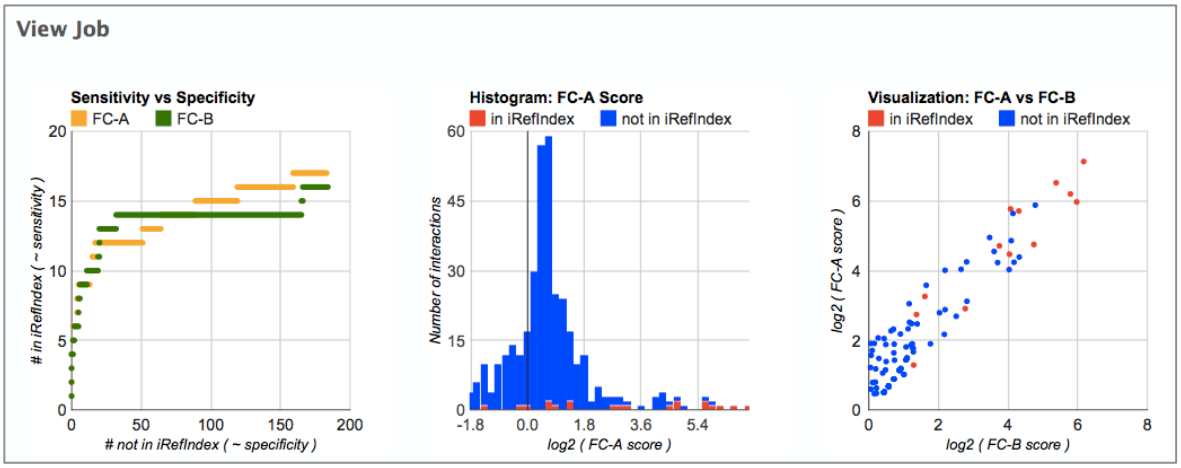
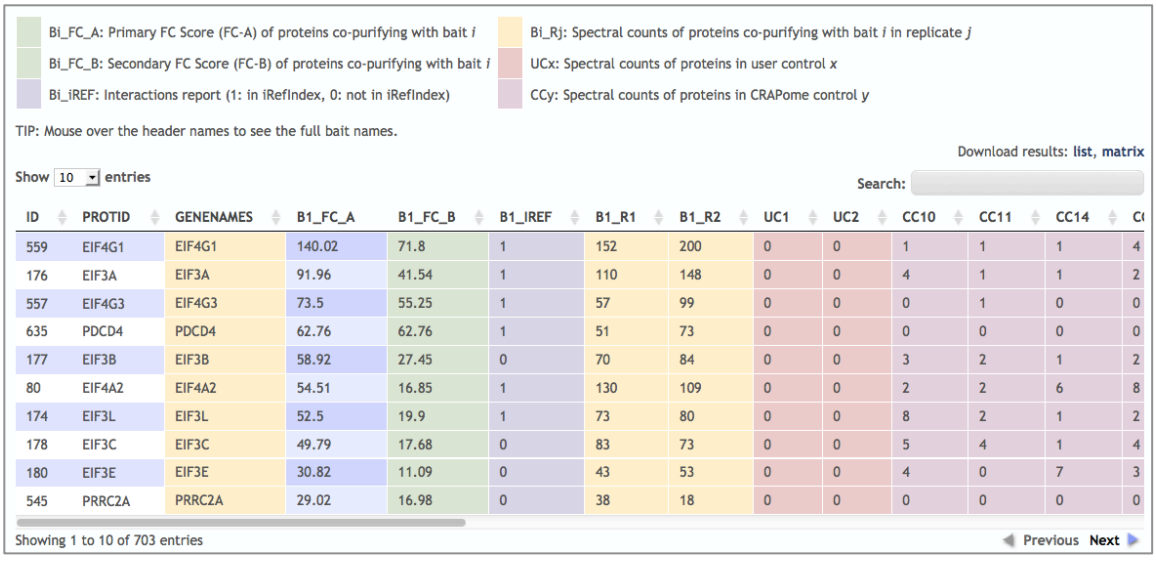


Figure 9. Comparison of the standard and stringent Fold Change results on the EIF4A2 dataset.



Bi_FC_A: Primary FC Score (FC-A) of proteins co-purifying with bait *i*

Bi_FC_B: Secondary FC Score (FC-B) of proteins co-purifying with bait *i*

Bi_iREF: Interactions report (1: in iRefIndex, 0: not in iRefIndex)

Bi_Rj: Spectral counts of proteins co-purifying with bait *i* in replicate *j*

UCx: Spectral counts of proteins in user control *x*

CCy: Spectral counts of proteins in CRAPome control *y*

TIP: Mouse over the header names to see the full bait names.

Download results: **list**, **matrix**

Show 10 entries

Search:

ID	PROTID	GENENAMES	B1_FC_A	B1_FC_B	B1_IREF	B1_R1	B1_R2	UC1	UC2	CC10	CC11	CC14	C(
559	EIF4G1	EIF4G1	140.02	71.8	1	152	200	0	0	1	1	1	4
176	EIF3A	EIF3A	91.96	41.54	1	110	148	0	0	4	1	1	2
557	EIF4G3	EIF4G3	73.5	55.25	1	57	99	0	0	0	1	0	0
635	PDCD4	PDCD4	62.76	62.76	1	51	73	0	0	0	0	0	0
177	EIF3B	EIF3B	58.92	27.45	0	70	84	0	0	3	2	1	2
80	EIF4A2	EIF4A2	54.51	16.85	1	130	109	0	0	2	2	6	8
174	EIF3L	EIF3L	52.5	19.9	1	73	80	0	0	8	2	1	2
178	EIF3C	EIF3C	49.79	17.68	0	83	73	0	0	5	4	1	4
180	EIF3E	EIF3E	30.82	11.09	0	43	53	0	0	4	0	7	3
545	PRRC2A	PRRC2A	29.02	16.98	0	38	18	0	0	0	0	0	0

Showing 1 to 10 of 703 entries

Previous Next

Figure 10. List of the results from the Fold Change calculation.

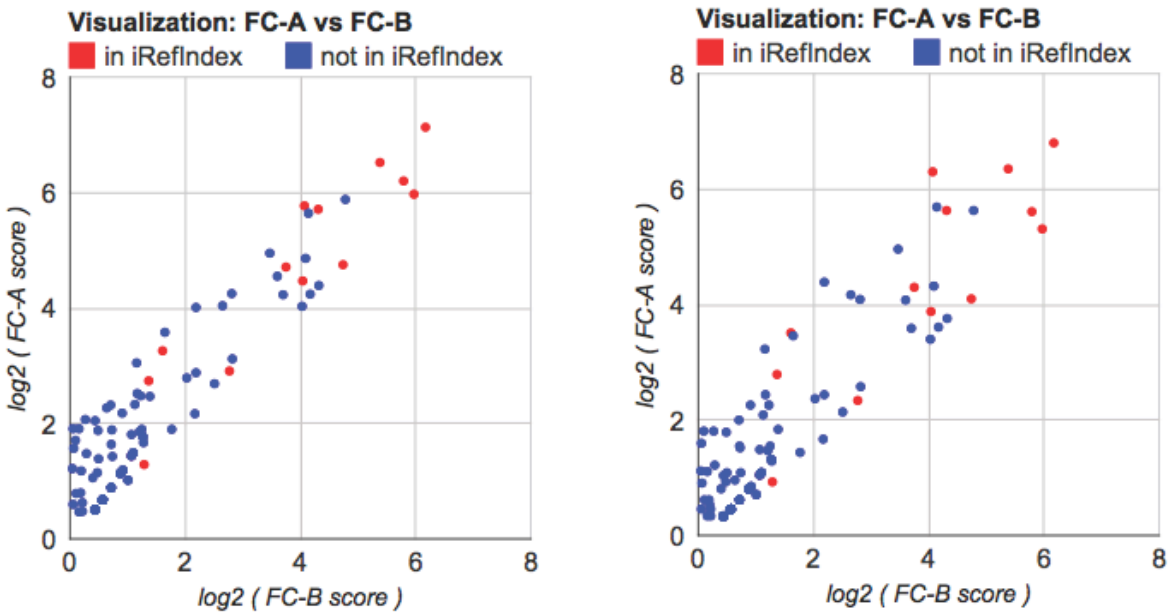


Figure 11. Relative robustness of the EIF4A2 dataset to control selection. Left: all controls included in both FC-A and FC-B analysis. Right: only the user controls are included in the analysis.