ProHits: an integrated software platform for mass spectrometry-based interaction proteomics

Guomin Liu, Jianping Zhang, Brett Larsen, Chris Stark, Ashton Breitkreutz, Zhen-Yuan Lin, Bobby-Joe Breitkreutz, Yongmei Ding, Karen Colwill, Adrian Pasculescu, Tony Pawson, Jeffrey L. Wrana, Alexey Nesvizhskii, Brian Raught, Mike Tyers*, Anne-Claude Gingras*

Supplementary Figures:

Other Supplementary information (also available at www.ProHitsMS.com):

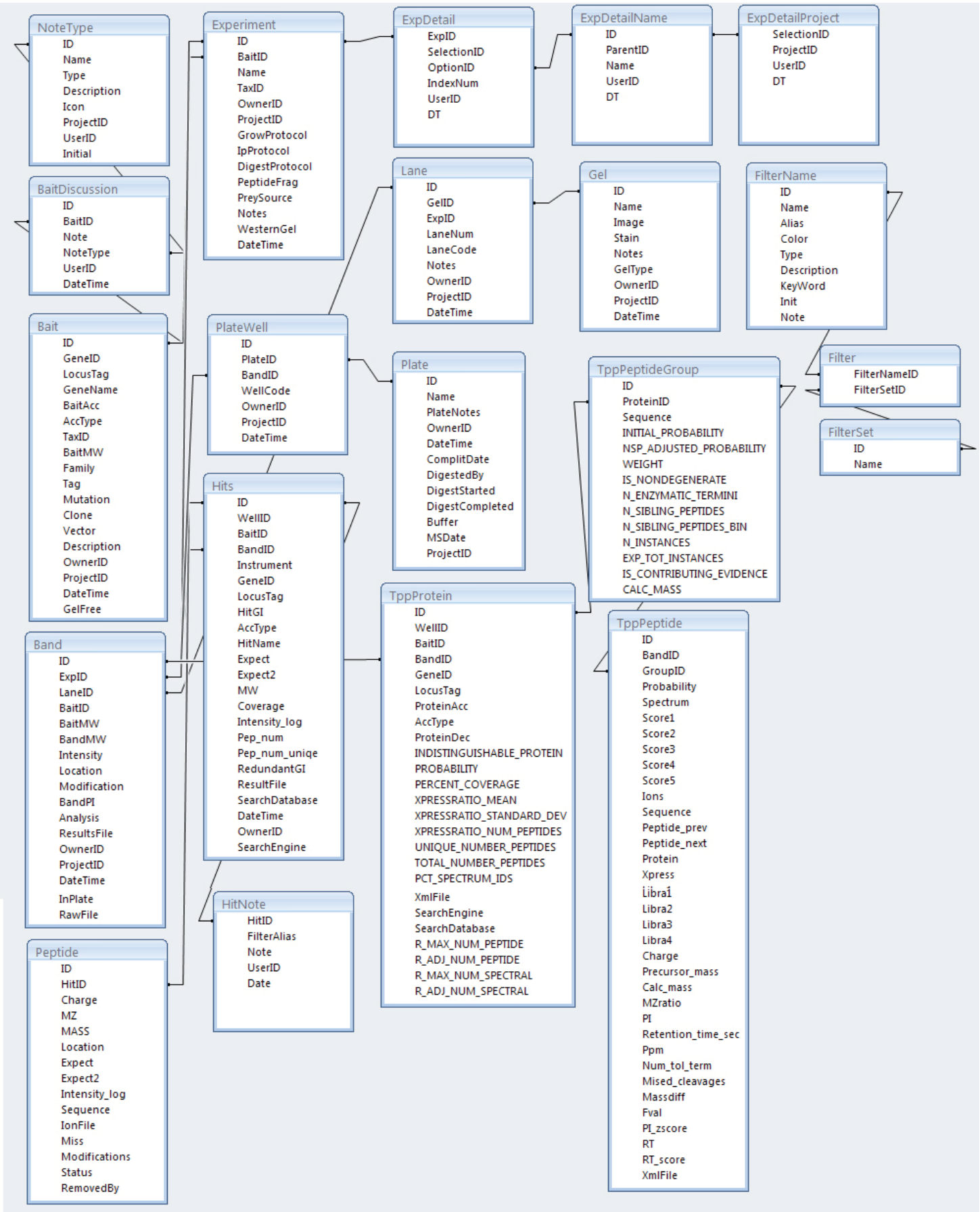
Supplementary information only available at www.ProHitsMS.com:

Demo site for navigation purposes

ProHits download package (contains the demo data and source code)

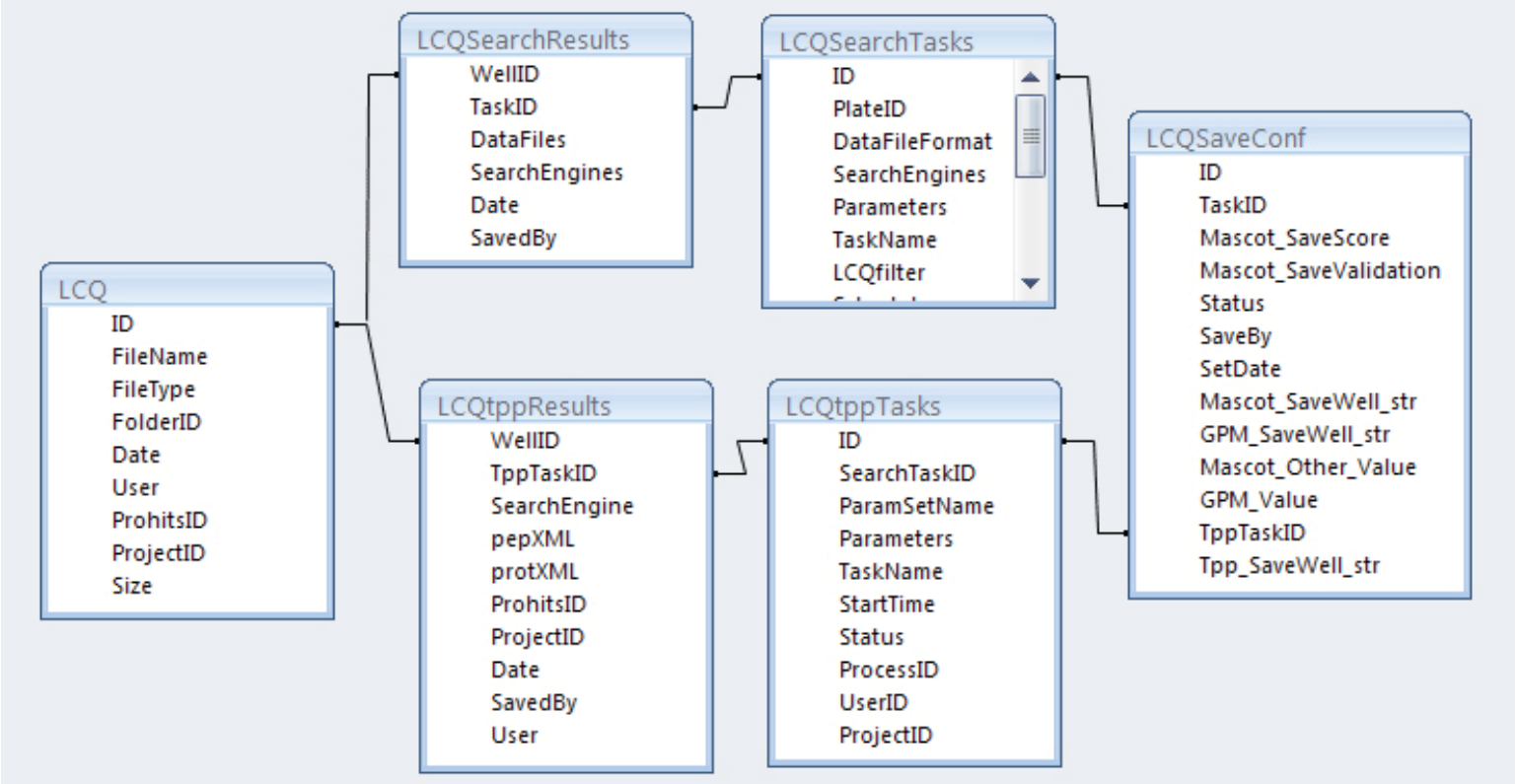
ProHits video tutorials

ProHits user manuals



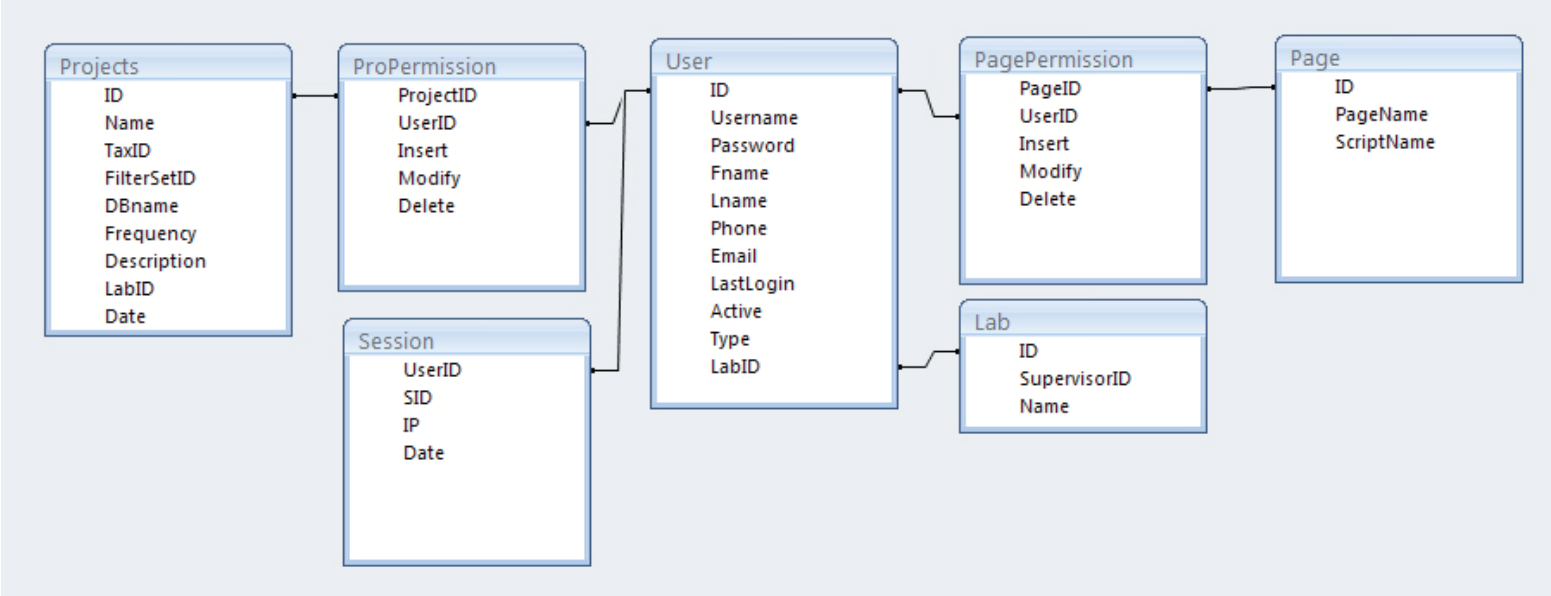
Supplementary Figure 1. Database Tables. A) Analyst module.

B



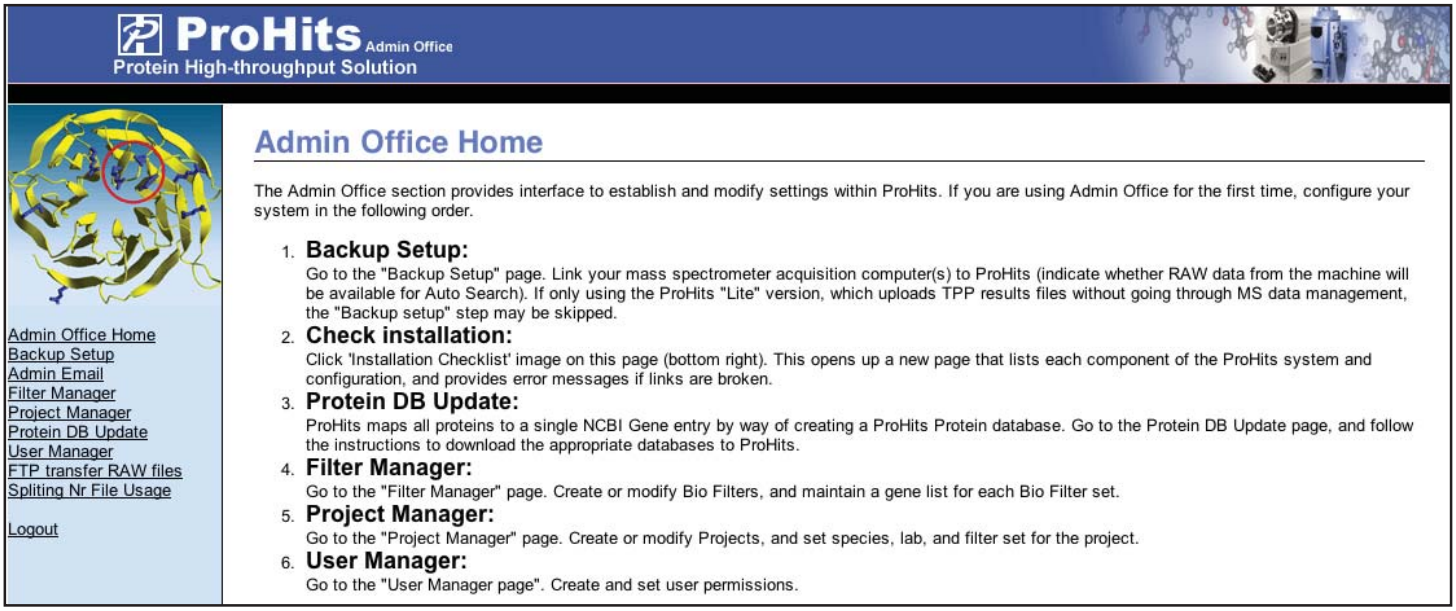
Supplementary Figure 1. Database Tables. B) Data Management module.

C

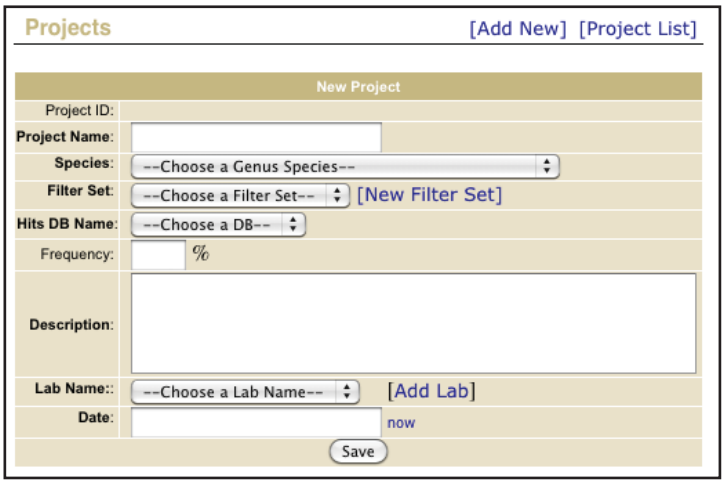


Supplementary Figure 1. Database Tables. C) Administration module.

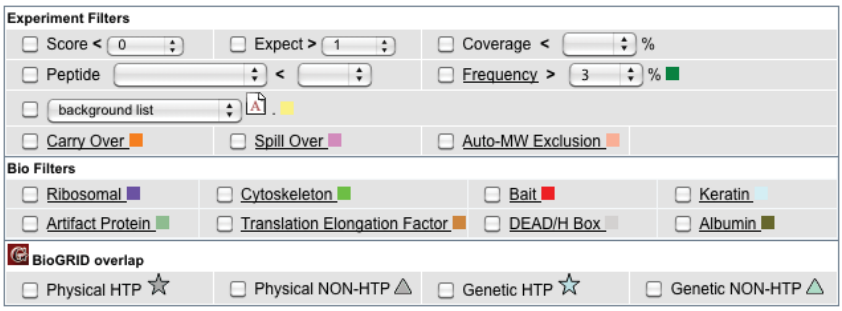
A



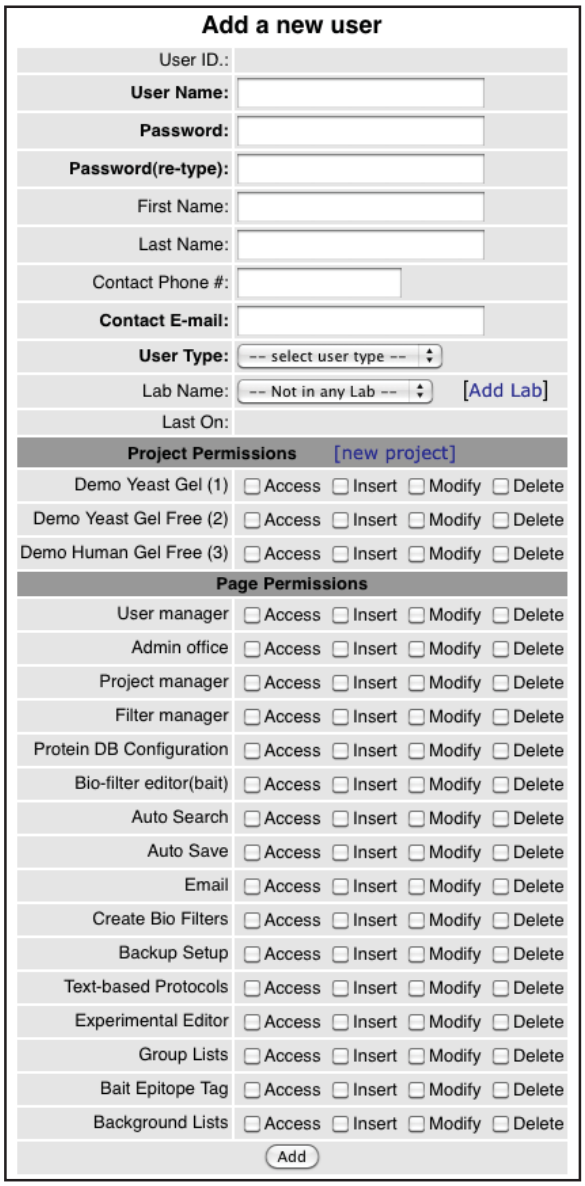
B



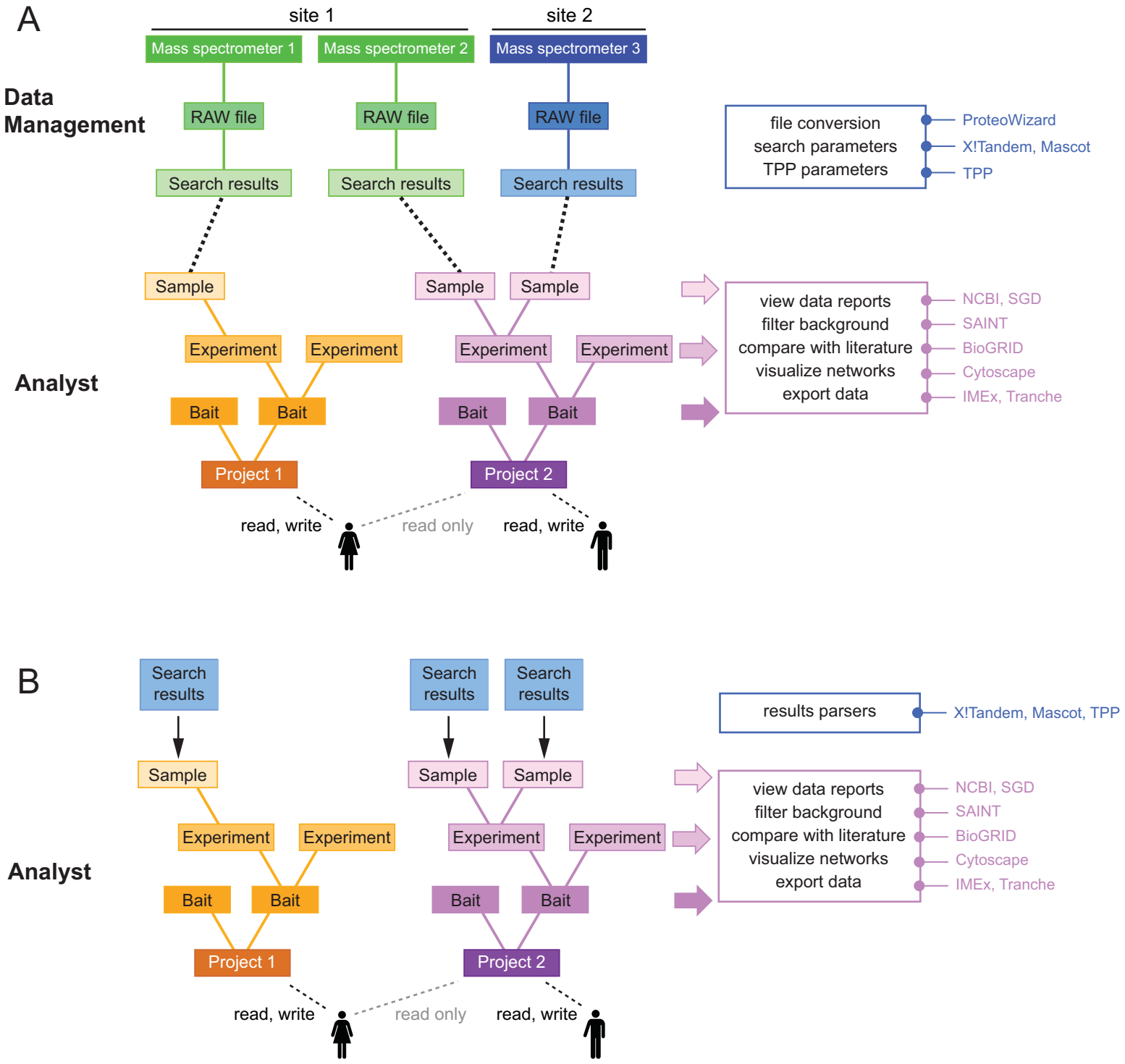
C



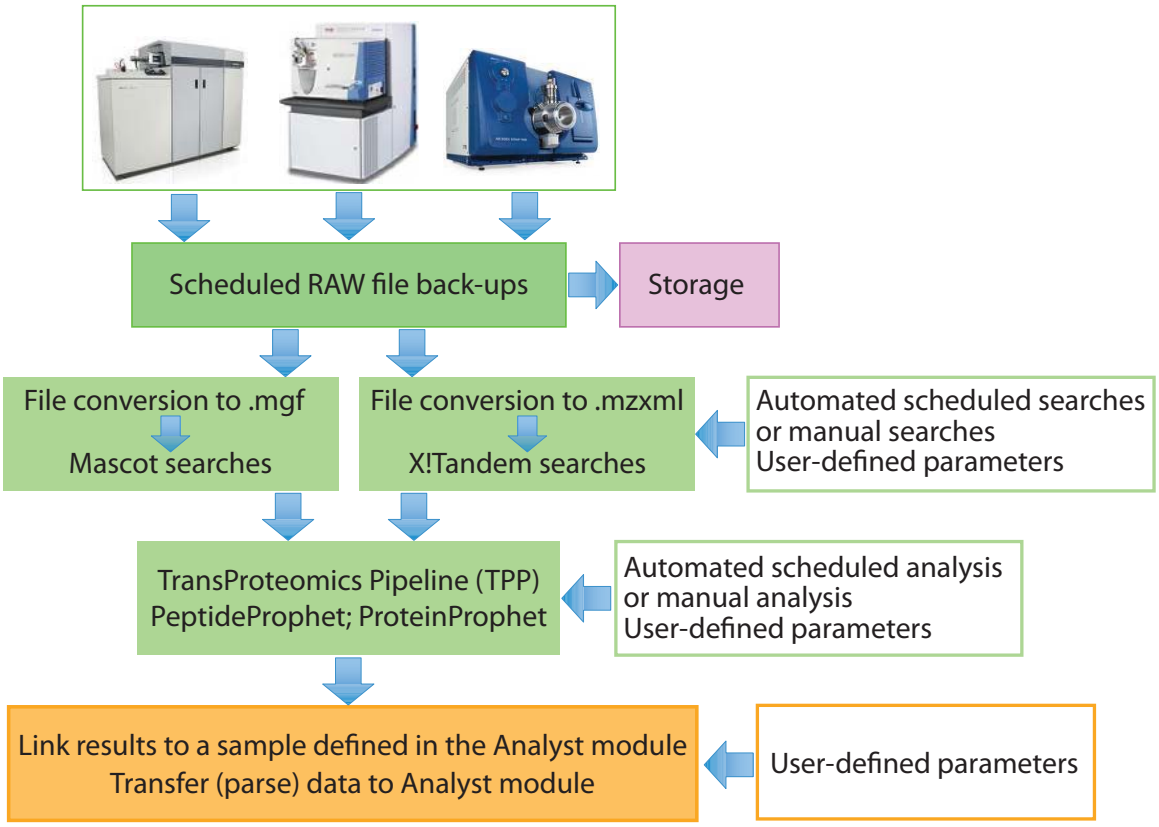
D



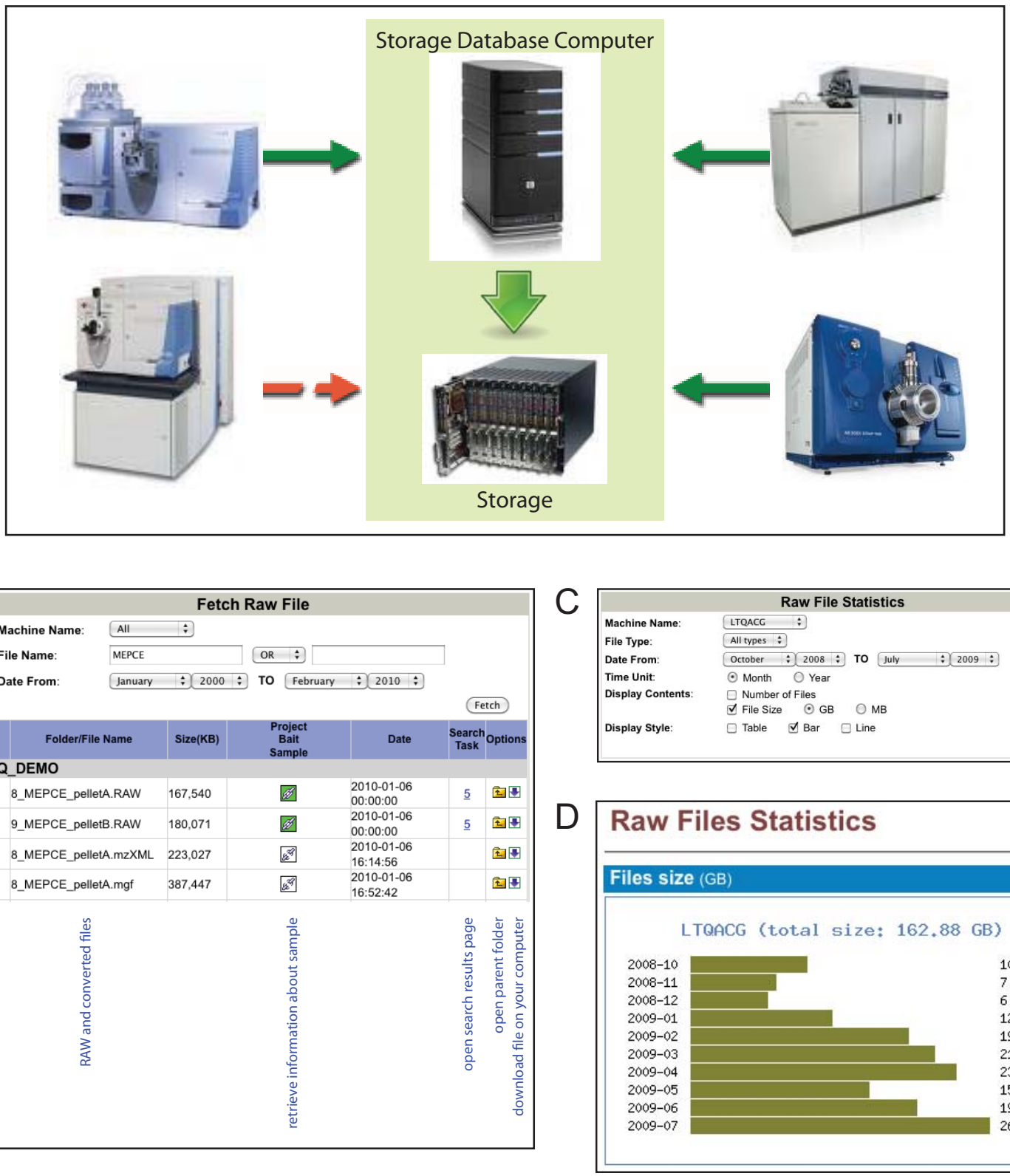
Supplementary Figure 2. Administration module. A) Home page of the Admin Office, detailing available options. B) Creation of a new project in Project Manager. Projects are associated with a specific research laboratory or working group, and generally contains results from a single species to which filter sets may be applied. "Frequency" refers here to the desired threshold for flagging a given hit as a likely contaminant based on detection across multiple purifications within the same project. C) Filters applicable to a specific project ("Report" page from the Analyst module). Filters may be "experimental", and related to quantitative information from the mass spectrometry results; these are not controlled in the Admin office. "Bio Filters" are defined in the "Filter Manager" of the Admin Office via text mining of the ProHits protein DB. D) User manager page. Each user is assigned access and modification privileges for specific projects.



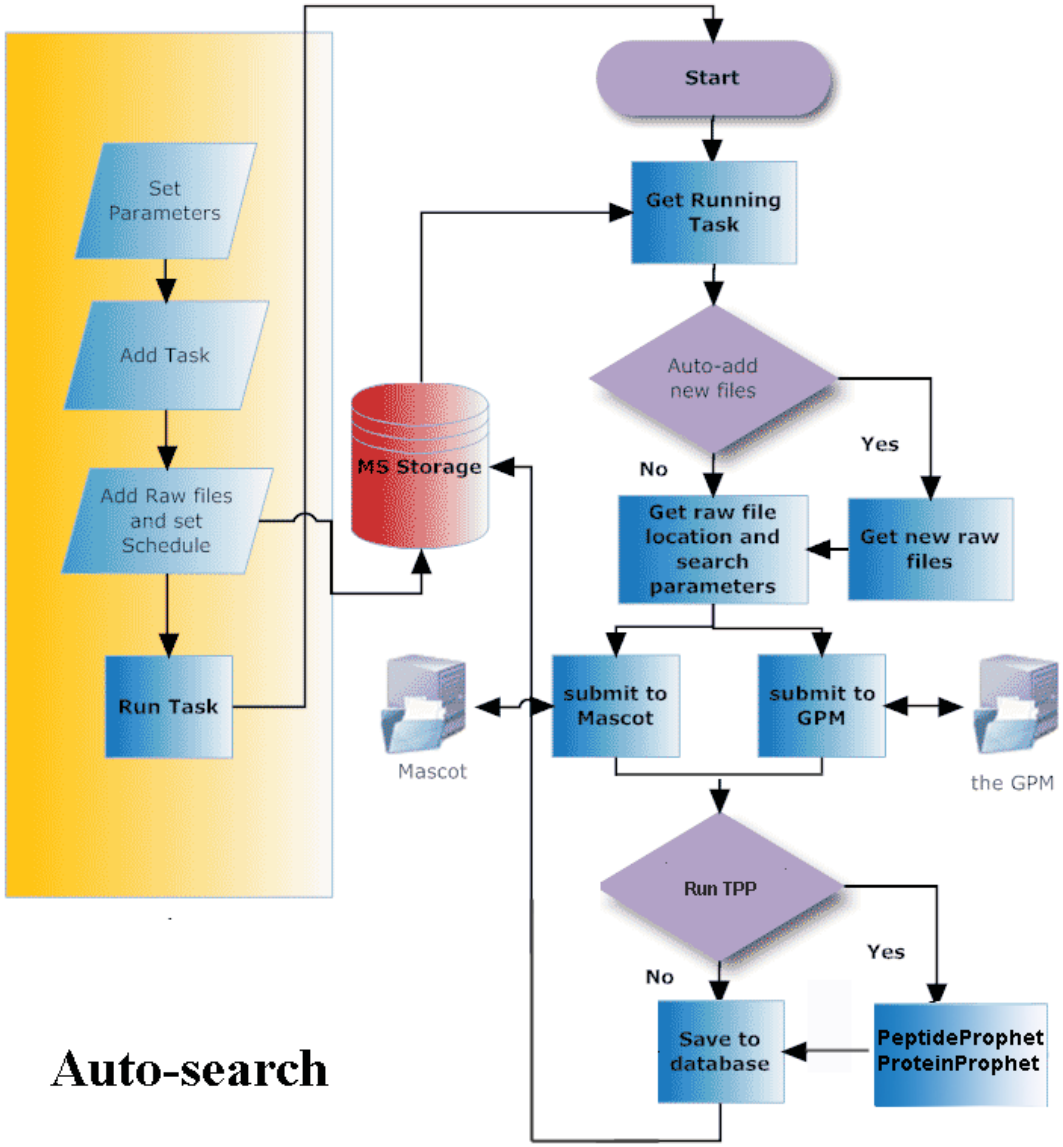
Supplementary Figure 3. Complete and Lite versions of ProHits. A) Complete version with the Data Management module: in addition to the ProHits server, specific servers for search engines and data conversion are mounted. The acquisition computers for each of the MS instruments are also linked to the ProHits server in the Data Management module. B) The ProHits lite version essentially consists of the Analyst module. Search results (obtained externally) are directly uploaded to a sample. Currently, uploads of search results from Mascot, X!Tandem and the TransProteomics pipeline are supported. ProHits Lite only requires setting up the ProHits server.



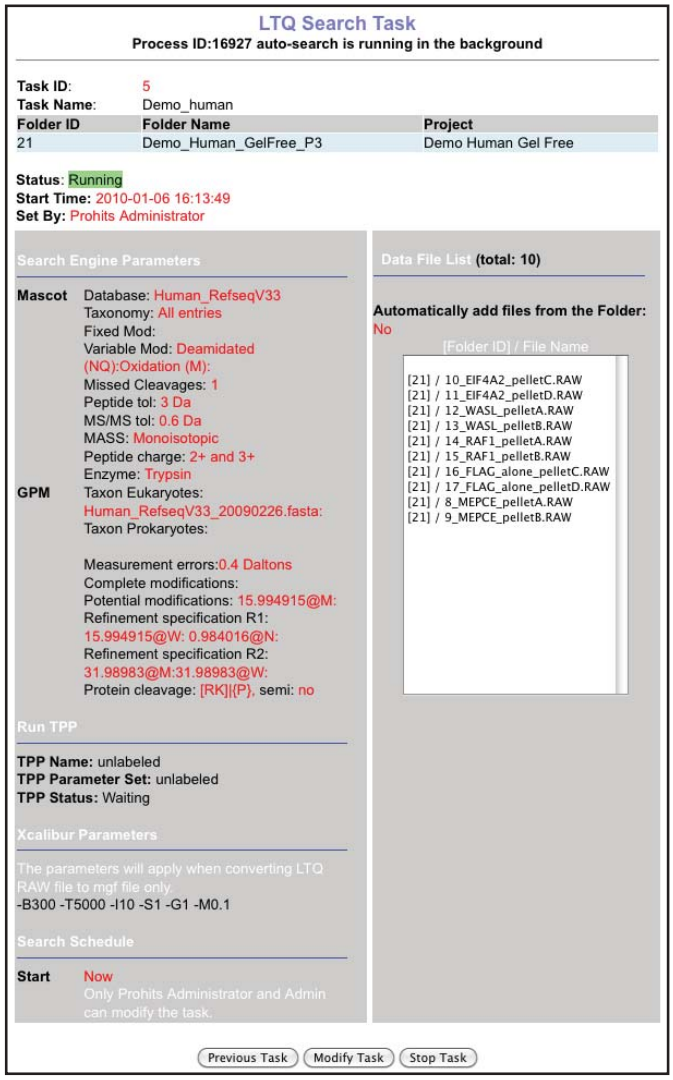
Supplementary Figure 4. Data Management module: Information flow details.
Nature Biotechnology, vol. 28, no. 10 Gingras, A-C et al.

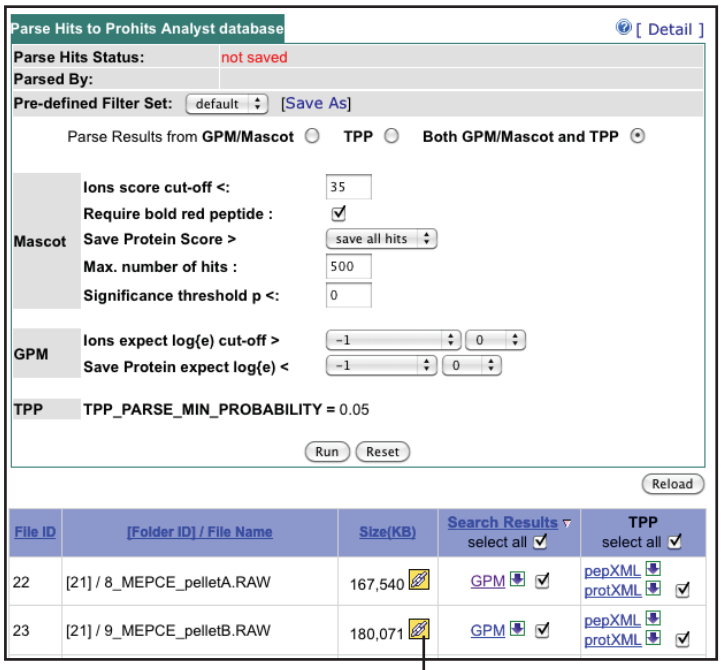


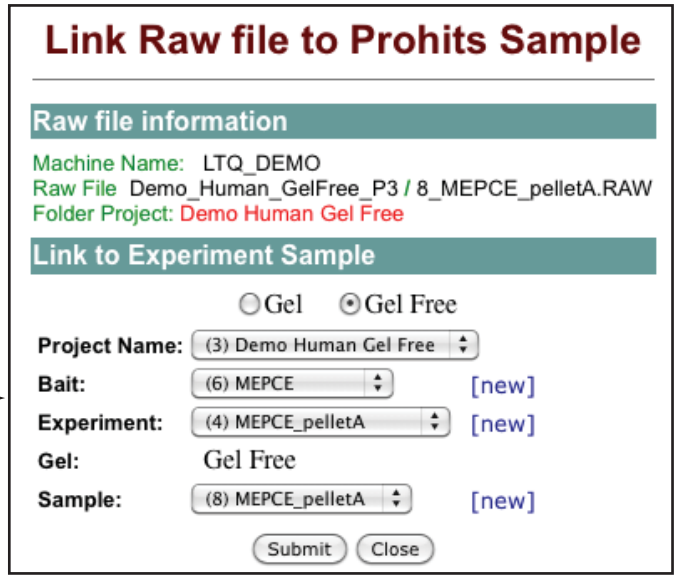
Supplementary Figure 5. Data Management module: Storage. A) Backup status. Each acquisition computer within the group or facility is automatically backed up every night on ProHits (files and folders are mirrored in the same organization as on the acquisition computer). Arrows indicate connection status: broken red arrows represent lost connections. B) Raw files retrieval. In Data Management Storage, the user can search for specific raw files (if desired, the searches can be limited by date or instrument), and retrieve all information related to these files. C) Raw file statistics. A function of the Data Management Storage is the reporting of the backed-up files on each of the instruments (or globally for the entire facility) for a given time period, and expressed as total file size or total file numbers; this facilitates reporting of the activities of the facility, and helps to identify bottlenecks. D) Example of a RAW file statistics report.



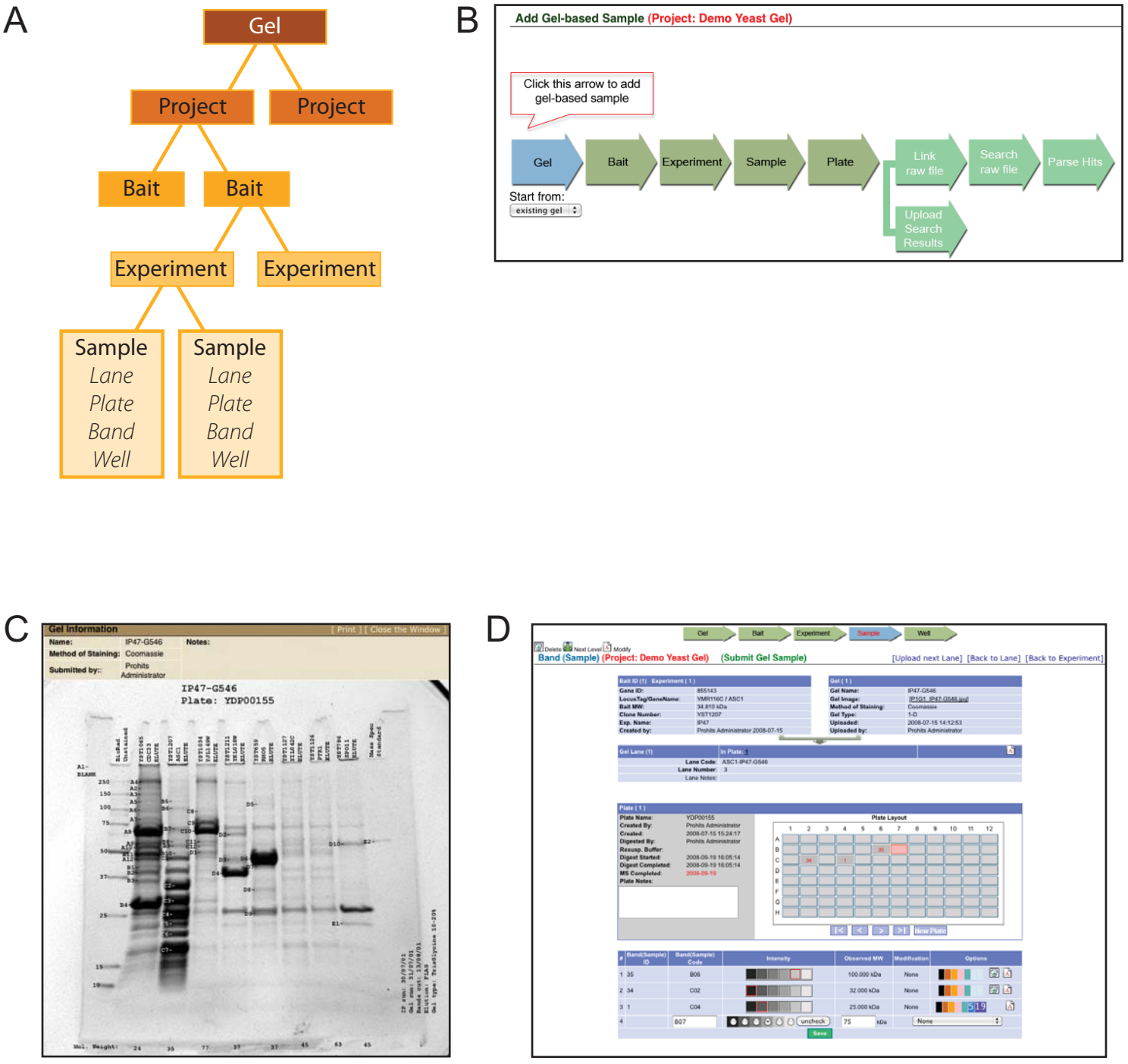
Supplementary Figure 6. Data Management module: Autosearch Flowchart.





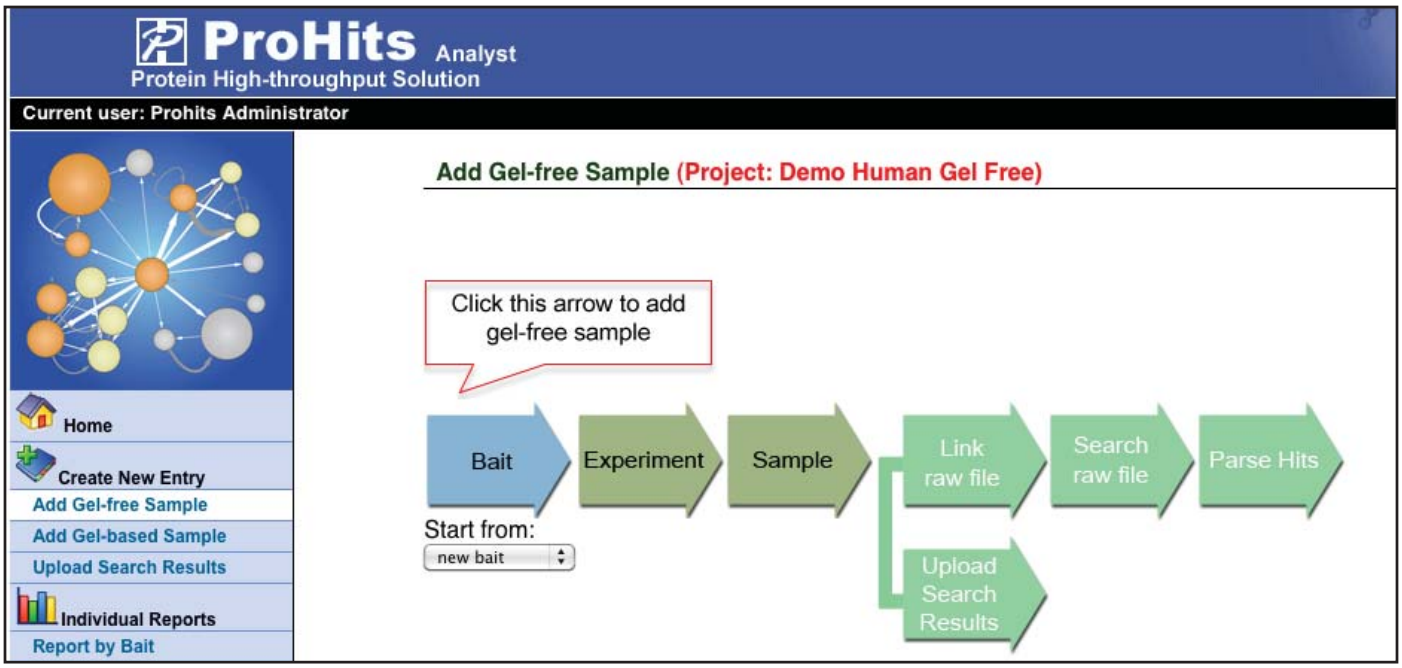


Supplementary Figure 7. Data Management module: Autosearch functions. A) Initiating a search task. Each authorized user can select search engines and search parameters, and initiate searches using these parameters. Searches can be manually initiated, or conducted in a scheduled manner, as data is backed-up onto ProHits. B) Search task details. Each search task is assigned a unique identifier, and the conversion parameters, search parameters and protein sequence database and version are recorded. C) Search results page. Search results can be directly visualized from the search engine pages. Search engine results can also be transferred (parsed) into the ProHits Analyst module, after they have been linked to the appropriate Project, Bait, Experiment and Sample entries. The yellow chain link indicates that a sample has been manually linked to the Analyst module. D) Creating links to the Analyst module. Links to the Analyst module are created by clicking on the chain link icon and selecting a pre-existing sample from the Analyst module (or by creating a [new] sample).

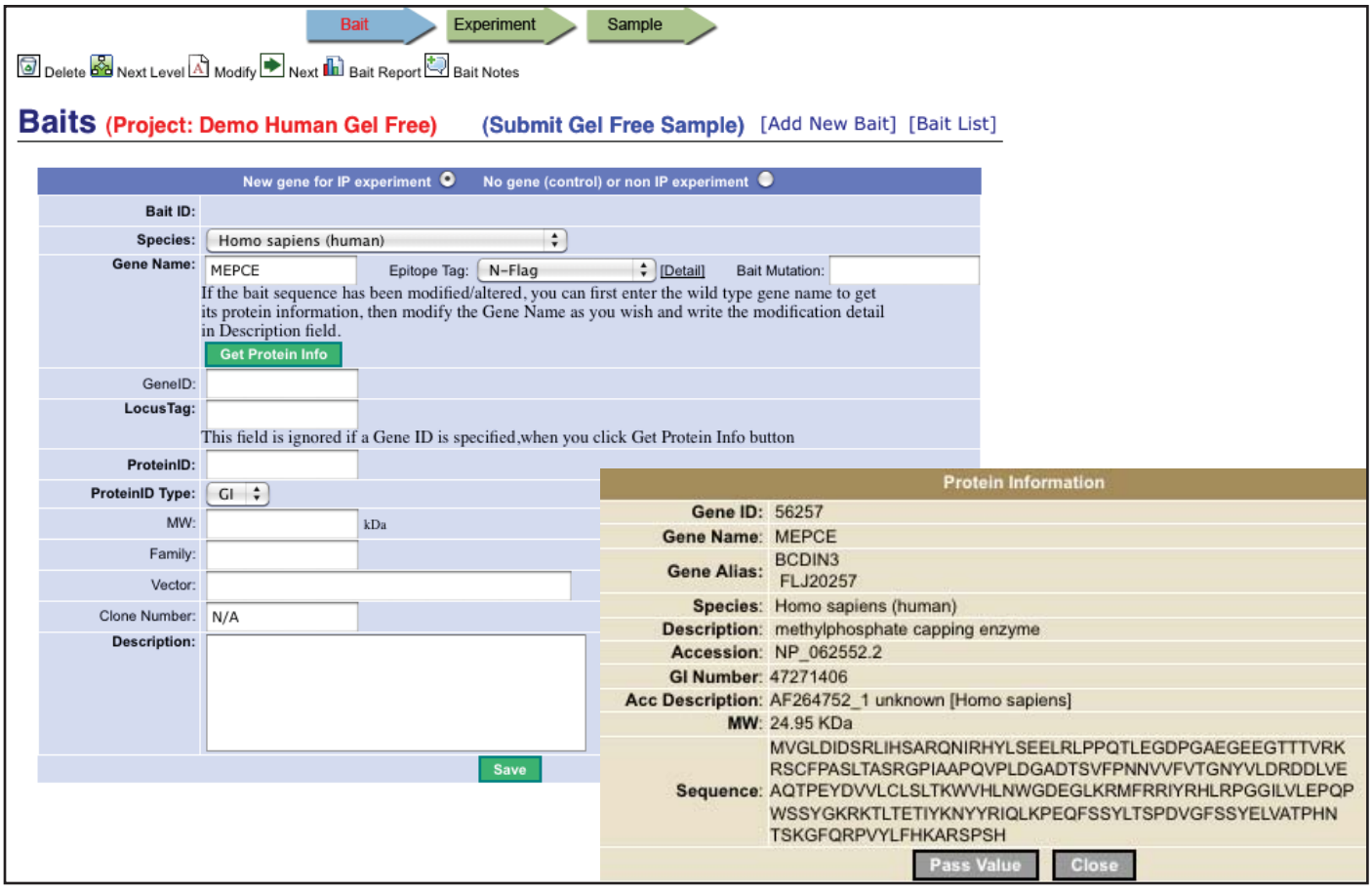


Supplementary Figure 8. Gel-Based Samples. A) Data structure for Gel-based entries. B) ProHits Analyst interface for adding Gel-based samples. C) Example of an SDS-PAGE gel recorded in ProHits; the user can access the images from the Report by Gel option. D) Detail of gel-based sample entry.

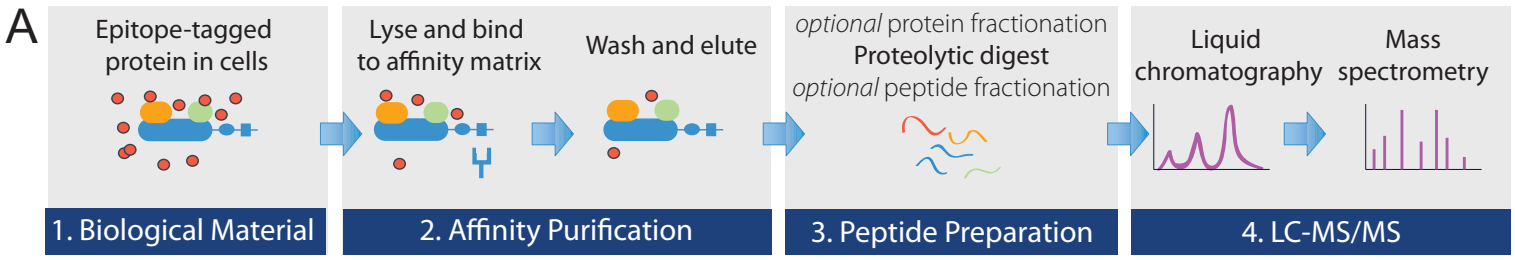
A

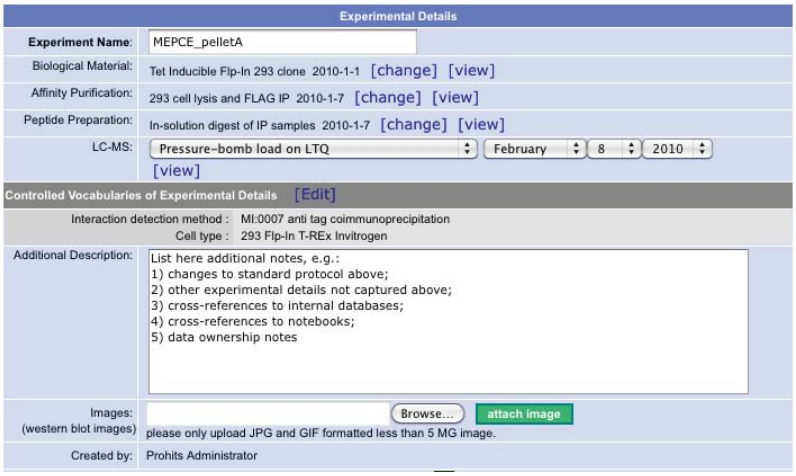


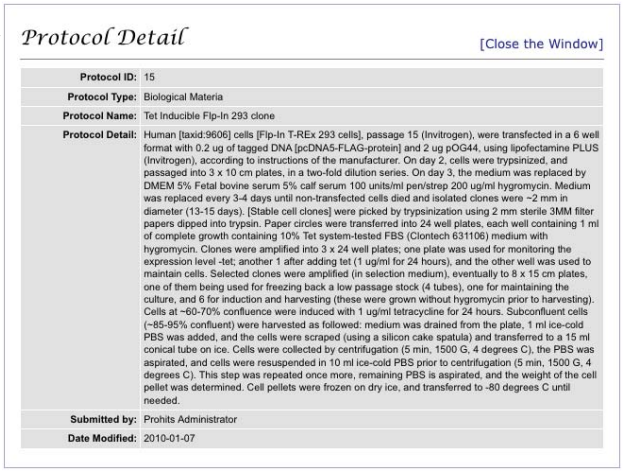
B

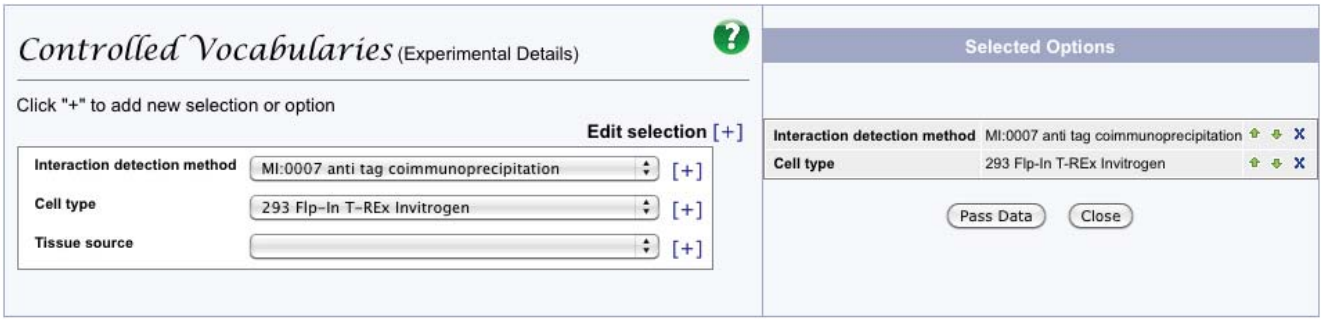


Supplementary Figure 9. Creating a new bait in ProHits Analyst. A) Screenshot of the entry page from Create New Entry > Add Gel-free sample. B) Bait description page. To add a new bait, type the official Gene Name, select the appropriate species, and click "Get Protein Info". The protein information (derived from the ProHits protein DB) is displayed. Select "Pass Value" to populate the Bait list. Additional information may be added at this step, including epitope-tag, mutations in the bait protein, cross-references to internal cDNA databases, and additional descriptive notes.









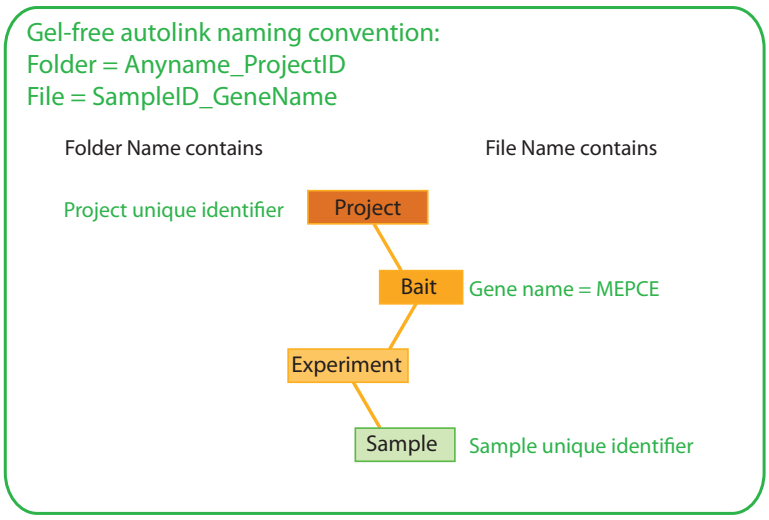
Supplementary Figure 10. Experimental annotation in the Analyst module. A) Typical AP-MS workflow. Four levels of experimental annotation (as indicated) are used to describe each experiment. B) Experimental annotation page. Annotation is entered in three areas: at the top of the page, the user selects appropriate text-based protocols via drop-down menus, reflecting the four levels of experimental annotation indicated in A. In the grey area in the middle of the page, the user enters controlled vocabularies, as selected by the administrator, to describe the interaction. At the bottom of the page, the user enters free text details not captured in the previous two sections, and links to any desired images. C) Example of a text-based protocol. D) Example of controlled vocabulary.

A

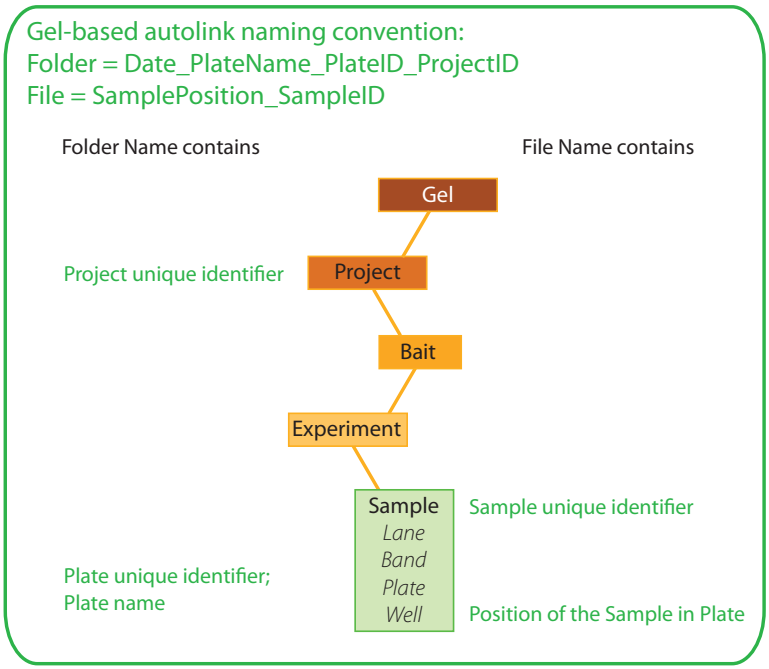
ID	File Name	Size	Project Bait sample	Date	Search Task	Download	Convert
22	8_MEPCE_pelletA.RAW	167,540		2010-01-06 00:00:00	5		
31	16_FLAG_alone_pelletC.RAW	102,264		2010-01-06 00:00:00	5		
30	17_FLAG_alone_pelletD.RAW	143,728		2010-01-06 00:00:00	5		
29	15_RAF1_pelletB.RAW	152,160		2010-01-06 00:00:00	5		
28	14_RAF1_pelletA.RAW	166,849		2010-01-06 00:00:00	5		

non-linked
auto-linked
manual-linked

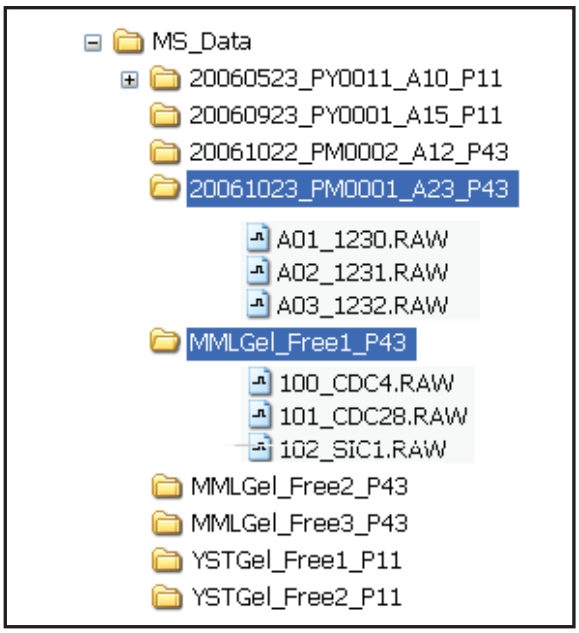
B



C

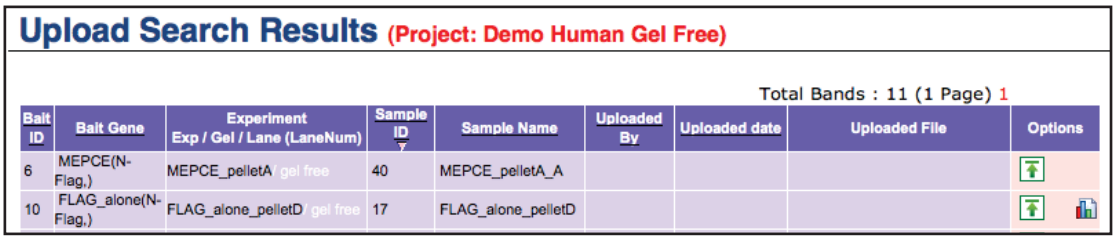


D



Supplementary Figure 11. Links between the Analyst and Data Management modules. A) View of existing liks in the Data Management module. In the Data Management module, files that have not been linked are associated with a broken white chain icon. To manually link the file to an entry created in the Analyst module, the user can click on any white icon and select the desired bait, experiment and sample. The icon color will change to yellow, indicating that the link is now functional. Note that new baits, experiments and samples can also be defined when creating manual links. B) Naming convention for automatic links in gel-free projects. To create Automatic links (indicated by green chain link icons), a convention for the naming of both folders and files on the acquisition computer must be respected. Sample entries must first be created in the Analyst module. For gel-free samples, the folders can have any name, followed by underscore and project identifier. The files are named by a sample unique identifier, underscore, the four first characters of the sample name as defined in Analyst (usually = gene name). C) Naming convention for automatic links in gel-based projects. For gel-based samples, the folders are named by date, underscore, autosampler plate name, underscore, autosampler plate identifier, underscore, project identifier. Sample names are: plate position, underscore, sample identifier. D) Examples of file names on the acquisition computer..

A

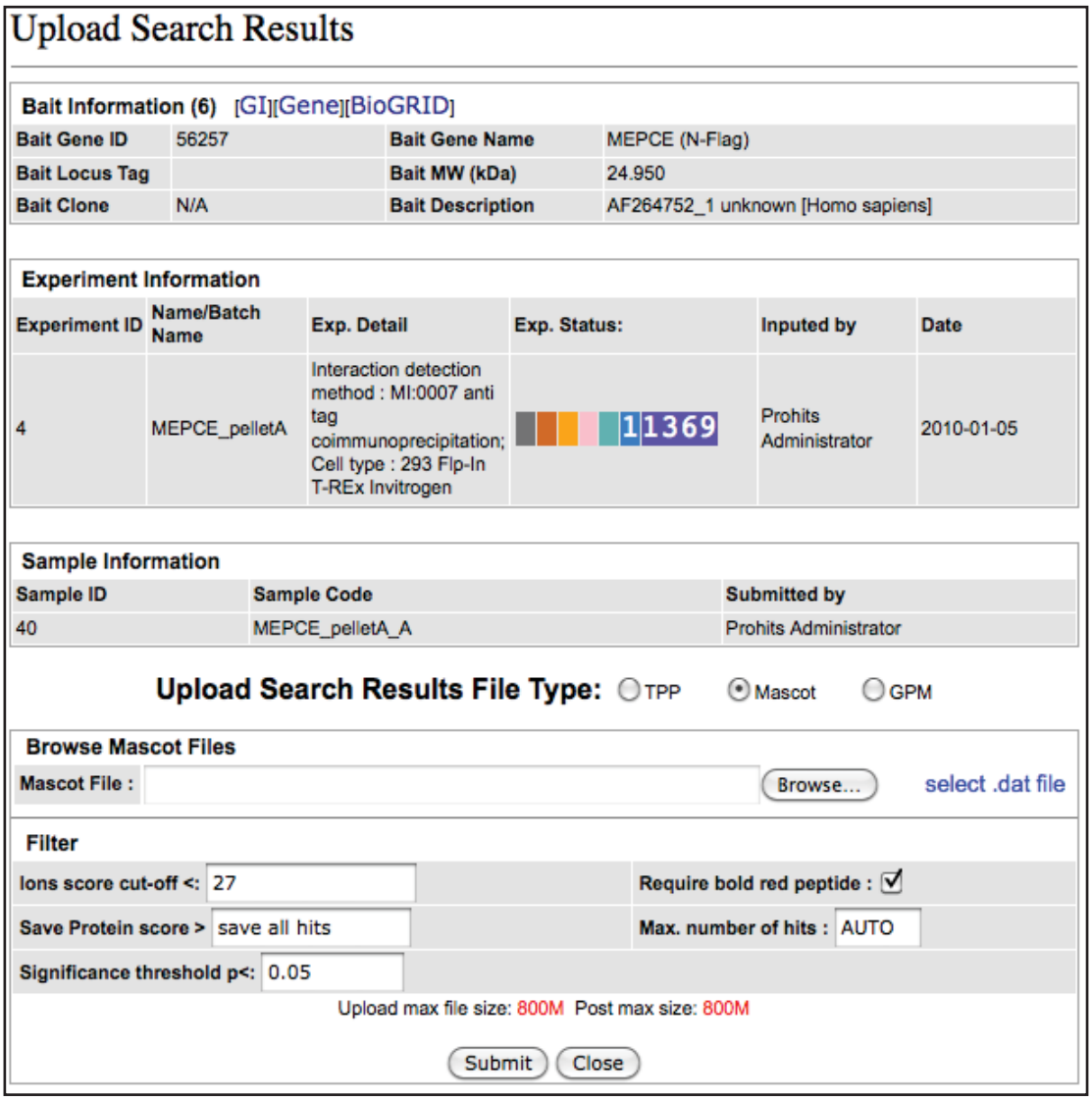


Upload Search Results (Project: Demo Human Gel Free)

Total Bands : 11 (1 Page) 1

Bait ID	Bait Gene	Experiment Exp / Gel / Lane (LaneNum)	Sample ID	Sample Name	Uploaded By	Uploaded date	Uploaded File	Options
6	MEPCE(N-Flag,)	MEPCE_pelletA / gel free	40	MEPCE_pelletA_A				
10	FLAG_alone(N-Flag,)	FLAG_alone_pelletD / gel free	17	FLAG_alone_pelletD				

B



Upload Search Results

Bait Information (6) [GI][Gene][BioGRID]

Bait Gene ID	56257	Bait Gene Name	MEPCE (N-Flag)
Bait Locus Tag		Bait MW (kDa)	24.950
Bait Clone	N/A	Bait Description	AF264752_1 unknown [Homo sapiens]

Experiment Information

Experiment ID	Name/Batch Name	Exp. Detail	Exp. Status:	Inputed by	Date
4	MEPCE_pelletA	Interaction detection method : MI:0007 anti tag coimmunoprecipitation; Cell type : 293 Flp-In T-REx Invitrogen	11369	Prohits Administrator	2010-01-05

Sample Information

Sample ID	Sample Code	Submitted by
40	MEPCE_pelletA_A	Prohits Administrator

Upload Search Results File Type: TPP Mascot GPM

Browse Mascot Files

Mascot File : Browse... select .dat file

Filter

Ions score cut-off <: 27 Require bold red peptide :

Save Protein score > save all hits Max. number of hits : AUTO

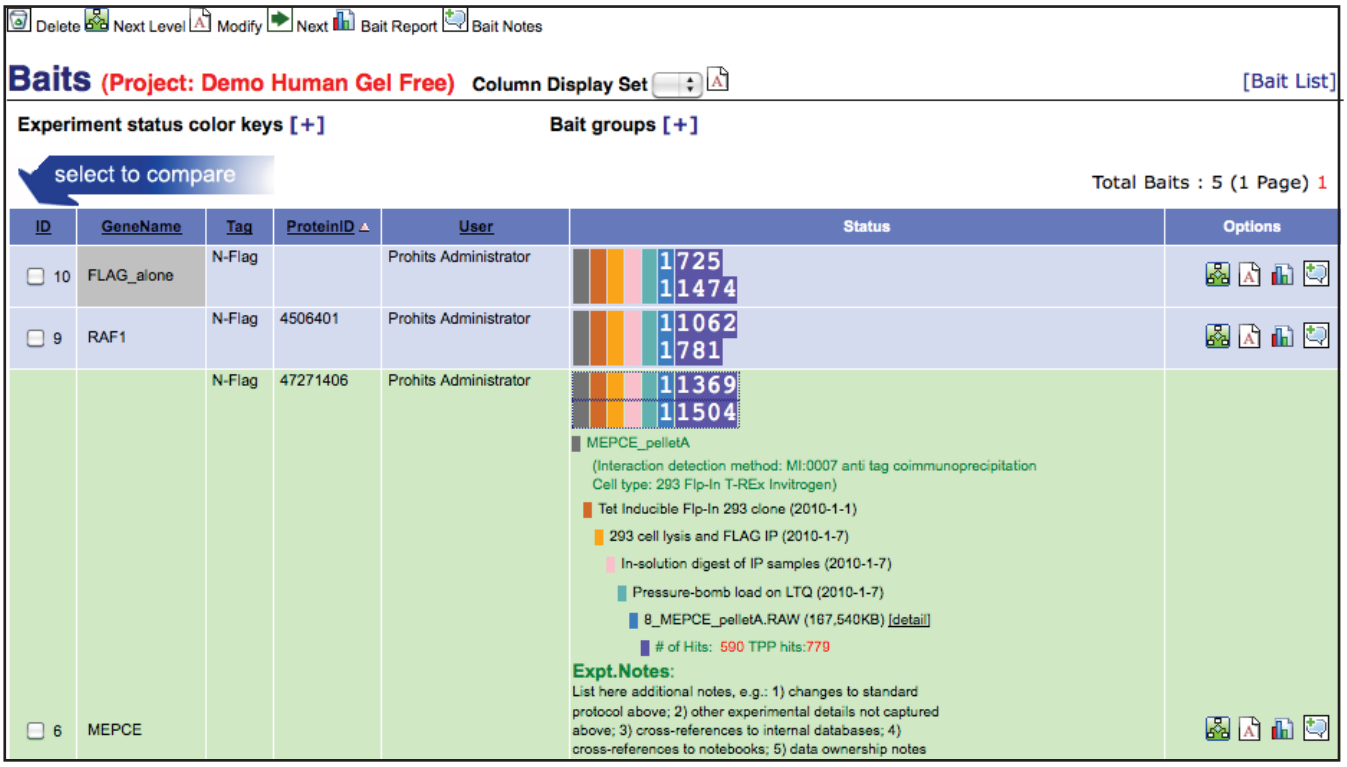
Significance threshold p<: 0.05

Upload max file size: 800M Post max size: 800M

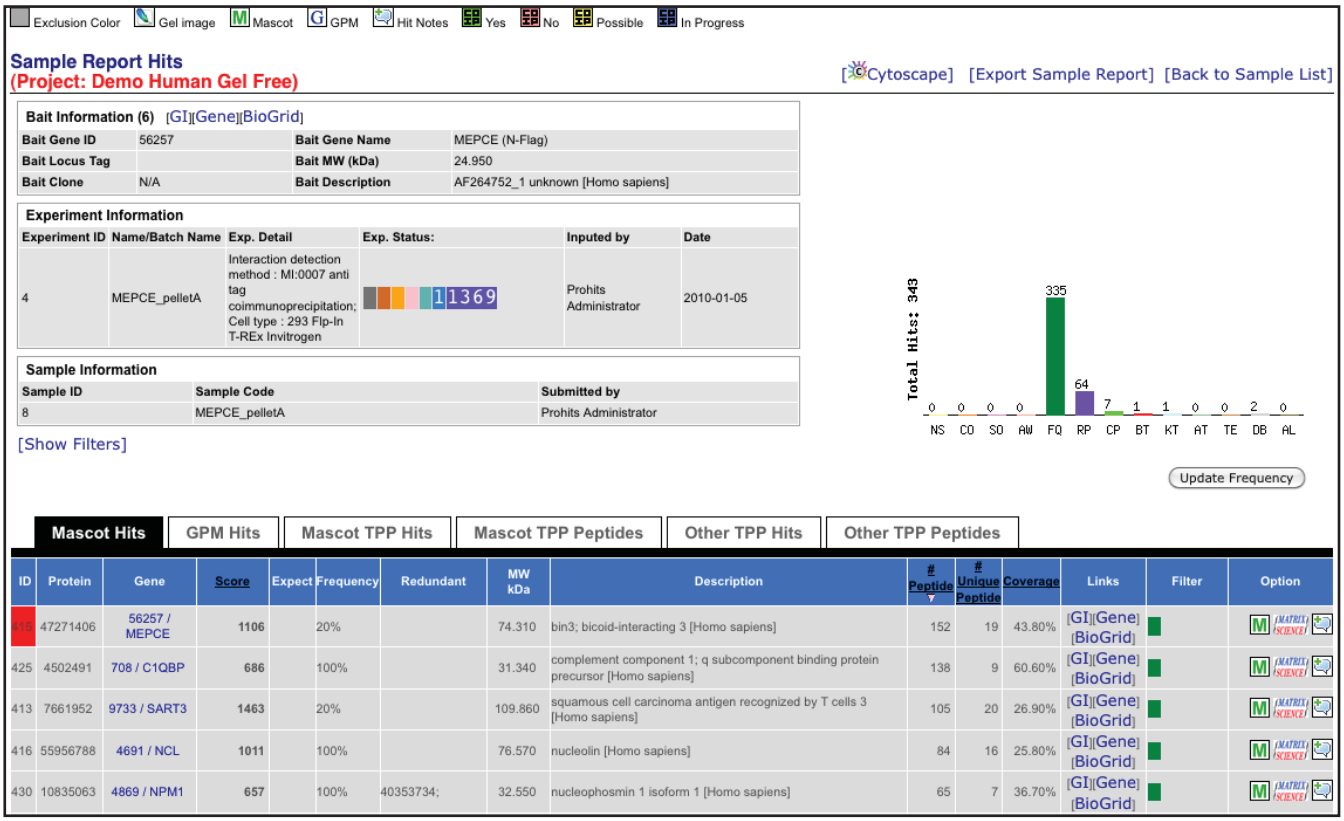
Submit Close

Supplementary Figure 12. Uploading search results generated outside of ProHits. A) Entry page for uploading search results - Bait, Experiments and Samples must first be specified. Clicking the "upload" icon opens a new window. B) Upload search result window: currently, uploads from the TransProteomics pipeline, or the search engines X!Tandem or Mascot are supported. To upload search results, simply select the appropriate radio button. For example, to process results from the Mascot search engine, select the desired files (.dat format) and appropriate parsing options. Note that while the full ProHits version allows you to monitor search parameters and parsing options, this functionality is not present when search results are uploaded from this page. We strongly suggest monitoring search parameters, database versions and parsing options throughout the course of the project.

A

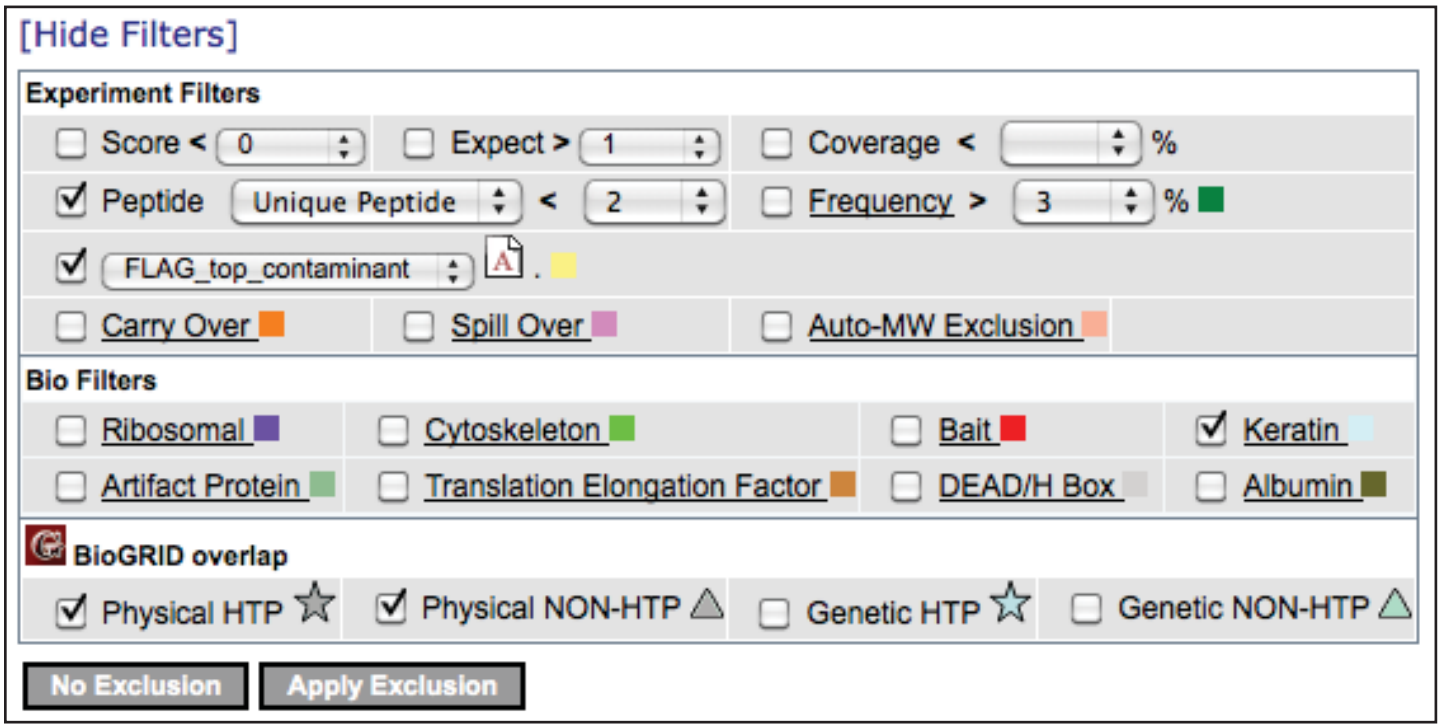


B



Supplementary Figure 13. Analyst module: Bait, Experiment and Sample Reports. A) Entry page of the Bait Report displaying the list of all baits analyzed. Clicking on any of the colored "Status" bars expands experimental details associated with a given bait. Clicking on the graph in the Options column opens up an unfiltered bait report. Note that the Bait, Experiment and Sample Report interfaces are identical: the sole difference is that all samples associated with a common bait are grouped in the Bait Report whereas they are viewed individually in a Sample Report (Experimental Reports provide an intermediate view). B) Sample Report view. Mass spectrometry results are shown at the bottom of the page (here Mascot search results are displayed). Multiple parameters, including the number of total and unique peptides, and sequence coverage, are displayed for each hit, alongside the frequency of detection for each hit within the entire project. Links to external databases (NCBI, BioGRID) and original search results pages are also provided. The Filter column flags hits that may be removed after application of one or more Bio-Filters and Experimental Filters. In the upper right corner is a bar graph indicating the numbers of hits that would be removed by applying each of the indicated filters. To view available filters and apply filtering options, select the "Show Filters" button above the results table.

A



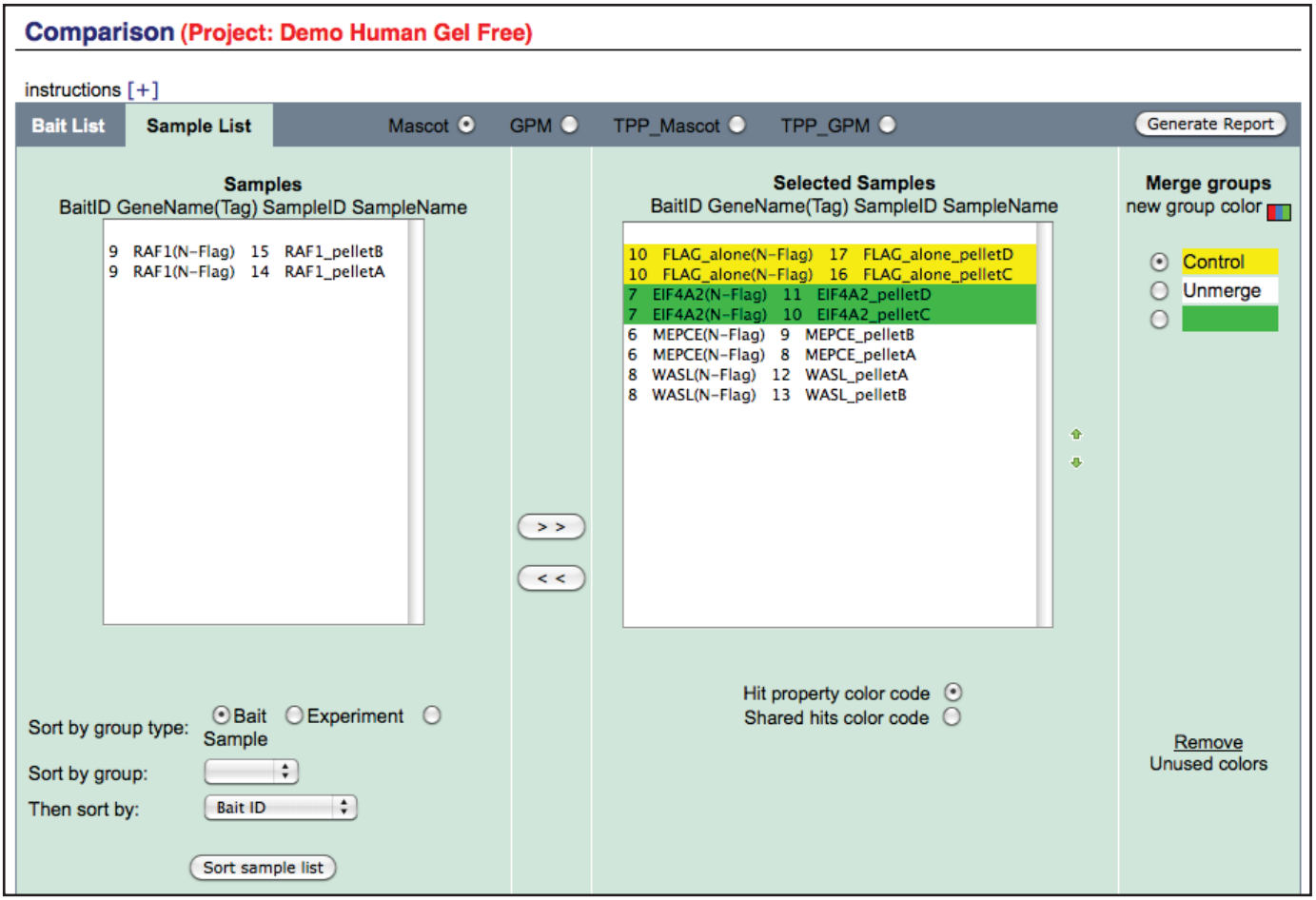
B

Mascot Hits | GPM Hits | Mascot TPP Hits | Mascot TPP Peptides | GPM TPP Hits | GPM TPP Peptides

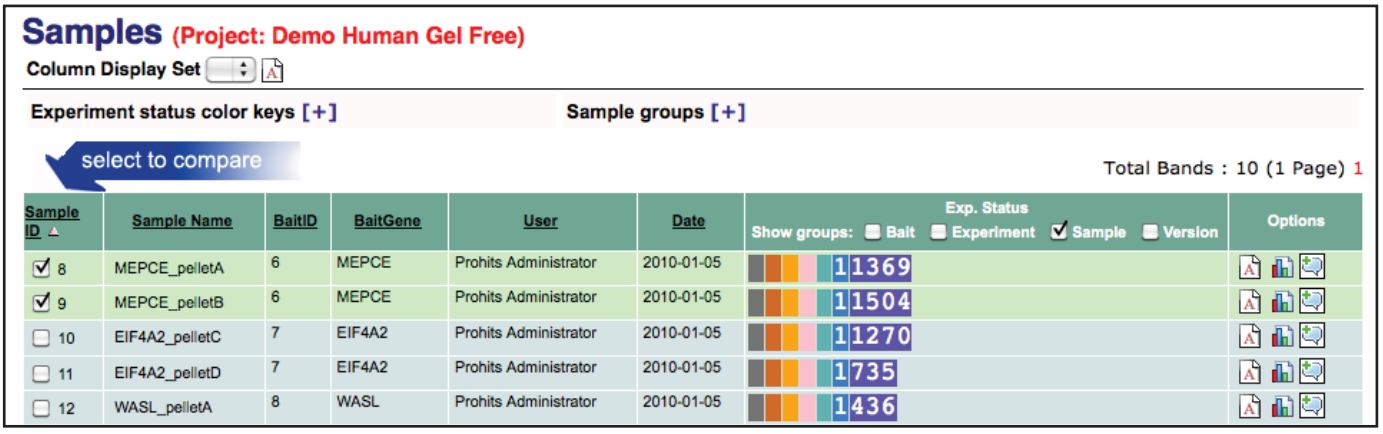
ID	Protein	Gene	Score	Expect	Frequency	Redundant	MW kDa	Description	# Peptide	# Unique Peptide	Coverage	Links	Filter	Option
415	47271406	56257 / MEPCE	1106		20%		74.310	bin3; bicoid-interacting 3 [Homo sapiens]	152	19	43.80%	[GI][Gene] [BioGRID]		
413	7661952	9733 / SART3	1463		20%		109.860	squamous cell carcinoma antigen recognized by T cells 3 [Homo sapiens]	105	20	26.90%	[GI][Gene] [BioGRID]	☆	
419	109809739	51574 / LARP7	955		20%	109809741;	66.860	La ribonucleoprotein domain family; member 7 [Homo sapiens]	53	14	25.80%	[GI][Gene] [BioGRID]	☆	
468	7706425	51691 / LSM8	305		20%		10.400	U6 snRNA-associated Sm-like protein LSm8 [Homo sapiens]	49	3	61.50%	[GI][Gene] [BioGRID]		
441	10863889	9092 / SART1	556		20%		90.200	squamous cell carcinoma antigen recognized by T cells 1 [Homo sapiens]	28	9	17.60%	[GI][Gene] [BioGRID]	☆	

Supplementary Figure 14. Analyst module: Filtering Search Results. A) List of the available filtering options for a project. Experimental Filters are based upon the mass spectrometric data or contaminants related to a specific project, while Bio-Filters are maintained by the administrator in the Admin Office. "BioGRID overlap" is not a filter per se, but this option puts a note in the "Filter" column to indicate those bait-to-hit relationships that have been previously reported. B) Filtered mass spectrometry results. This is the same list as in Figure 9, after likely contaminants have been removed. Filtered or unfiltered lists can be exported or displayed in the Cytoscape viewer.

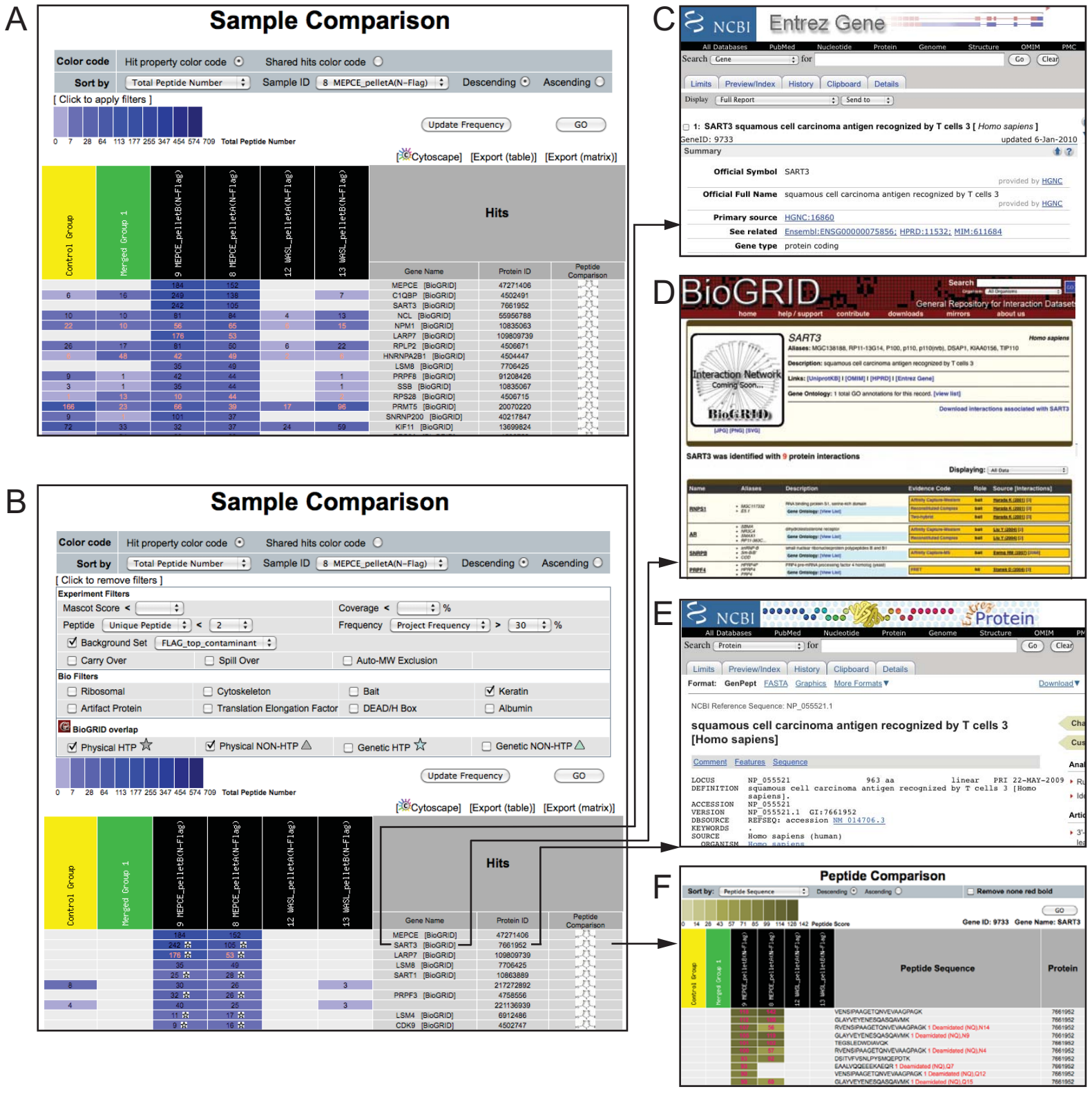
A



B



Supplementary Figure 15. Analyst module: Selecting Baits or Samples for Comparison. A) Selecting the Comparison option opens an entry page, where multiple Sample or Bait search results can be selected for comparison. Selecting Baits will lead to the merging of all Samples defined under that bait (Note that ProHits does not recalculate any of the search results after merging: only the highest search engine score, unique peptide count, or total peptide count for each hit across all samples associated with the given bait will be reported). The user can compare results from Mascot, X!Tandem or TPP by selecting desired Baits/Samples and pressing "Generate Report". Note that the user can also decide to manually combine the results of several baits and/or samples by clicking on the "merge group" icon; the selected samples will be combined into a single entry and the maximal value for each parameter displayed. Control runs can similarly be combined. B) An alternative approach to select files for comparison is to click on the bait or sample ID in the report pages. This is very useful in combination with the search functions. The selected files will be transferred to the Comparison entry page shown in A); files will remain selected until the user logs out or manually removes them.



Supplementary Figure 16. Analyst module: Comparison page. A) An unfiltered Comparison page. Individual Baits, Samples, or user-defined groups are listed in the columns, while hits are displayed in rows. In this case, the color-coding in each cell reflects the total number of peptides detected for each hit. Other parameters (e.g. search engine scores, unique peptides, sequence coverage, etc.) may also be displayed. Each attribute is assigned a different color scale. The numbers in each cell are the numerical value for the parameter displayed. Mousing over each cell opens a pop-up window detailing additional information. B) The same dataset as in (A), after filtering. Filters applied here were: (i) removal of proteins present on an internal list of "FLAG_top_contaminants", (ii) removal of proteins detected in >30% of all samples analyzed within the project, or (iii) detected with less than 2 unique peptides. The list is sorted to highlight the MEPCE sample. In addition to filtering, the "BioGRID Overlap" function was applied to mark hits previously reported in the BioGRID interaction database; these are represented by stars or triangles in each cell, depending on the experimental scale for the BioGRID deposition. C-E) External links for a given hit. F) Peptide Comparison allows the user to view all peptides identified for a given hit across all experiments loaded into the comparison report.

BaitID Comparison

Color code Hit property color code ◉ Shared hits color code ○

Sort by Total Peptide Number | BaitID ID 8 WASL(N-Flag) | Descending ◉ Ascending ○

[Click to apply filters]

0 7 28 64 113 177 255 347 454 574 709 **Total Peptide Number**

Update Frequency GO

[Cytoscape] [Export (table)] [Export (matrix)]

Control Group	8 WASL(N-Flag)	9 RAF1(N-Flag)	7 EIF4H2(N-Flag)	6 MEPCE(N-Flag)	Hits		
					Gene Name	Protein ID	Peptide Comparison
	303				WASL [BioGRID]	51702526	
13	131	6		3	ACTB [BioGRID]	4501885	
166	96	99	23	66	PRMT5 [BioGRID]	20070220	
	82				ACTC1 [BioGRID]	4885049	
72	59	41	33	37	KIF11 [BioGRID]	13699824	
	52	5		3	POTEE [BioGRID]	134133226	
	51	5	2	3	ACTA1 [BioGRID]	4501881	
	40				WIPF3 [BioGRID]	122937496	
161	37	52	8	39	WDR77 [BioGRID]	13129110	
15	34	24	6	47	PPM1B [BioGRID]	4505995	
41	33	274	26	27	TUBB [BioGRID]	29788785	
21	32	104	13	17	HSPA8 [BioGRID]	5729877	
	32				WIPF2 [BioGRID]	18959210	
65	32	38	3	7	KRT1 [BioGRID]	119395750	
36	28	275	17	15	TUBB2C [BioGRID]	5174735	
31	25	23	13	19	CLNS1A [BioGRID]	4502891	
48	23	150	20	29	HSPA1B [BioGRID]	167466173	
26	22	17	17	81	RPLP2 [BioGRID]	4506671	
26	19		13	17	TUBA1C [BioGRID]	14389309	
65	18	13	4	9	FLNA [BioGRID]	116063573	
8	18	17	8	8	HSPA5 [BioGRID]	16507237	
35	18	8	8	6	LOC651751 [BioGRID]	169218253	
12	16	4	2	8	SF3B3 [BioGRID]	54112121	
19	16	5	2	8	RBM10 [BioGRID]	20127479	
22	15	15	10	65	NPM1 [BioGRID]	10835063	
14	14	5	33	44	RPS3 [BioGRID]	15718687	
18	13	9	18	30	HNRNPM [BioGRID]	14141152	
10	13	14	10	84	NCL [BioGRID]	55956788	
12	13	4	67	16	RPSA [BioGRID]	9845502	
20	12	16	3	5	KRT9 [BioGRID]	55956899	
22	11	15	9	10	STK38 [BioGRID]	6005814	
11	10	2	5	22	ILF3 [BioGRID]	24234753	
	10	2	40	17	RPS21 [BioGRID]	4506699	
18	10	28				195972866	
20	9	14	7	44	RPLP0 [BioGRID]	4506667	
10	9	7	3	6	TRIM28 [BioGRID]	5032179	
4	9	3	1	16	RPL5 [BioGRID]	14591909	
1	8	35	3	5	HSPA9 [BioGRID]	24234688	
7	8	8	36	61	HNRNPA1 [BioGRID]	4504445	
11	8	3	14	26	DHX9 [BioGRID]	100913206	
	8				WIPF1 [BioGRID]	38373695	
4	8	1	2	1	PTS [BioGRID]	4506331	
23	7	3		1	THRAP3 [BioGRID]	167234419	
2	7	2	8	2	HNRNPL [BioGRID]	52632383	
6	7	11	16	249	C1QBP [BioGRID]	4502491	
2	7	4		3	PPM1A [BioGRID]	10337595	
4	7	3	4	6	RPL11 [BioGRID]	15431290	
2	6			9	SF3B2 [BioGRID]	55749531	
13	6	1		1	IARS [BioGRID]	94721239	

Supplementary Figure 17. Unfiltered data for main text Figure 1b. Page 1 of 13.

12	6	2		2	SF3B1 [BioGRID]	54112117	
2	6	1	2	5	DHX15 [BioGRID]	68509926	
6	6	3	48	49	HNRNPA2B1 [BioGRID]	4504447	
11	6	6	7	25	HNRNPH1 [BioGRID]	5031753	
6	6		6	8	DDX5 [BioGRID]	4758138	
24	6	3			CLTC [BioGRID]	4758012	
8	6	4	12	15	DDX17 [BioGRID]	38201710	
7	6	12			KRT2 [BioGRID]	47132620	
	6				KRT6A [BioGRID]	5031839	
2	6	1		1	PRSS1 [BioGRID]	4506145	
	5				NCK2 [BioGRID]	52630423	
10	5	2			RARS [BioGRID]	15149476	
8	5	1	5	19	RPL9 [BioGRID]	15431303	
3	5	1	7	10	HNRNPA3 [BioGRID]	34740329	
5	5	3	2	4	EEF1B2 [BioGRID]	4503477	
2	5		13	15	RPS25 [BioGRID]	4506707	
4	5	1		1	SPIN1 [BioGRID]	112293285	
6	5	5	6	24	RPLP1 [BioGRID]	4506669	
2	5	5	10	23	HIST1H1C [BioGRID]	4885375	
	5	2	2		SNRPB [BioGRID]	4507125	
4	5	5	1	28	RPL7A [BioGRID]	4506661	
6	4	1			EEF1D [BioGRID]	25453472	
28	4	6	1	1	CCT8 [BioGRID]	48762932	
15	4		3	25	XRCC6 [BioGRID]	4503841	
6	4		1	20	RPL10A [BioGRID]	15431288	
17	4	1		6	MAP1B [BioGRID]	153945728	
6	4	1		2	DCTN2 [BioGRID]	5453629	
15	4	3	10	6	LARS [BioGRID]	108773810	
13	4	7			EEF1A1 [BioGRID]	4503471	
1	4	1	1	5		217330646	
7	4	6	3	22	HNRNPD [BioGRID]	14110414	
4	4		6	20	HNRNPR [BioGRID]	5031755	
1	4		4	12	HNRNPA0 [BioGRID]	5803036	
26	4	274	5	4	HSP90AA1 [BioGRID]	154146191	
14	3	5			CCT2 [BioGRID]	5453603	
1	3	1	2	7	SFRS5 [BioGRID]	86991438	
8	3	1	69		EIF3F [BioGRID]	4503519	
2	3	1			WARS [BioGRID]	47419914	
4	3		4	27	RPL18 [BioGRID]	4506607	
4	3	2	25	22	RPS8 [BioGRID]	4506743	
3	3	1	5	9	ILF2 [BioGRID]	24234747	
1	3		4	4	SFRS2 [BioGRID]	47271443	
9	3	1	1	2	EEF1G [BioGRID]	4503481	
5	3	3	1	14	RPL12 [BioGRID]	4506597	
11	3	4	11	5	EIF4B [BioGRID]	50053795	
3	3			2	SR140 [BioGRID]	122937227	
3	3				PRDX1 [BioGRID]	4505591	
9	3	3			VCP [BioGRID]	6005942	
4	3	1		40		221136939	
8	3	4	1	30		217272892	
2	3		26	15	RPS18 [BioGRID]	11968182	
	3				PRDX4 [BioGRID]	5453549	
13	3	4	2		TCP1 [BioGRID]	57863257	
2	3		1	6	RPL22 [BioGRID]	4506613	
7	3		7	7	EIF2S2 [BioGRID]	29826335	
6	3	1		1	STK38L [BioGRID]	24307971	
2	3	2	2	5	RPL23A [BioGRID]	17105394	
8	3	8	15	30	HNRNPK [BioGRID]	14165435	
3	3	4	6	16	HNRNPAB [BioGRID]	55956919	
4	3	3			KRT14 [BioGRID]	15431310	
1	2	3	3	15	RPL10P16 [BioGRID]	41151097	
1	2			1	LTBP1 [BioGRID]	46249412	
5	2	14	2	2	RCN2 [BioGRID]	4506457	
2	2		2	31	DDX21 [BioGRID]	50659095	
	2			3	NOP56 [BioGRID]	32483374	
8	2	4			CCT4 [BioGRID]	38455427	
5	2	1	9	14	EIF2S1 [BioGRID]	4758256	
15	2	4	4	49	PRKDC [BioGRID]	13654237	
	2	1		2	PDHA1 [BioGRID]	4505685	
5	2	2			LOC652147 [BioGRID]	169218225	
1	2		1		EXOSC6 [BioGRID]	17402904	
1	2		1	5	RPL36 [BioGRID]	16117794	
3	2				OAT [BioGRID]	4557809	

	2				ARHGEF10L [BioGRID]	58761492	
3	2	1	1		TRIM21 [BioGRID]	15208660	
6	2	2	162		EIF3B [BioGRID]	33239445	
2	2	1	131		EIF3H [BioGRID]	4503515	
2	2				SF3A1 [BioGRID]	5032087	
	2			21	DDX23 [BioGRID]	41327771	
3	2		6	3	HNRPDL [BioGRID]	14110407	
7	2	2	66	7	PABPC1 [BioGRID]	46367787	
	2	43	1	1	IRS4 [BioGRID]	4504733	
3	2	1	8	5	RPS6 [BioGRID]	17158044	
	2				DDX3Y [BioGRID]	13514809	
11	2	1	231		EIF3A [BioGRID]	4503509	
5	2			1	SNRPA1 [BioGRID]	50593002	
	2				CAPZB [BioGRID]	4826659	
10	2	4	3	1	QPCT [BioGRID]	6912618	
2	2				NUDT21 [BioGRID]	5901926	
4	2	3	1	2	RIOK1 [BioGRID]	23510356	
5	2	9	2	5	SNRPD3 [BioGRID]	4759160	
1	2	3	10	17	RPS19 [BioGRID]	4506695	
2	2	3		1	RBBP7 [BioGRID]	4506439	
	2				NCK1 [BioGRID]	5453754	
5	2		7	27	RPS16 [BioGRID]	4506691	
12	2	1	3		EPRS [BioGRID]	62241042	
56	2	13	2	5	IVNS1ABP [BioGRID]	24475847	
15	2	3	17	22	RPS4X [BioGRID]	4506725	
1	2	1	1	8	SRP9 [BioGRID]	4507217	
1	2	2	13	44	RPS28 [BioGRID]	4506715	
15	2	1	15	12	RPS14 [BioGRID]	5032051	
	2	1	35	29	YBX1 [BioGRID]	34098946	
4	2				RPL7 [BioGRID]	15431301	
13	2	2			CCT5 [BioGRID]	24307939	
4	2		10	20	HNRNPU [BioGRID]	14141161	
	2				ENO1 [BioGRID]	4503571	
1	2				DCD [BioGRID]	16751921	
	2	1			GAPDH [BioGRID]	7669492	
	2				CKB [BioGRID]	21536286	
1	2				VIM [BioGRID]	62414289	
3	2	3			DDB1 [BioGRID]	148529014	
7	2			1	C11orf84 [BioGRID]	39930523	
	2				FBXO38 [BioGRID]	45505155	
3	2				LOC654340 [BioGRID]	169165505	
	1				P4HB [BioGRID]	20070125	
7	1		1		BCLAF1 [BioGRID]	7661958	
	1				PPIB [BioGRID]	4758950	
1	1				HTATSF1 [BioGRID]	21361437	
1	1	1		21	KPNA2 [BioGRID]	4504897	
1	1	3			HSPA4 [BioGRID]	38327039	
3	1	1	2	4	SRP14 [BioGRID]	149999611	
4	1	3	1	3		217416381	
3	1	1			PSMC1 [BioGRID]	24430151	
3	1	20			DNAJA1 [BioGRID]	4504511	
3	1	1	62		EIF3G [BioGRID]	49472822	
	1				DLST [BioGRID]	19923748	
1	1	6	1	19	RPL23 [BioGRID]	4506605	
	1				RBBP5 [BioGRID]	53759148	
4	1	2			PSMC5 [BioGRID]	24497435	
	1				P4HA1 [BioGRID]	63252886	
	1	1			DLAT [BioGRID]	31711992	
1	1	1	1		PRMT1 [BioGRID]	150456457	
2	1	1	2	15	RPL17 [BioGRID]	4506617	
6	1	2	12	13	RPS15A [BioGRID]	14165469	
2	1		1	14	RPL15 [BioGRID]	15431293	
3	1	1	1	44	SSB [BioGRID]	10835067	
	1				CDC42 [BioGRID]	4757952	
9	1	3			CCT7 [BioGRID]	5453607	
1	1	2		1	PRKAR1A [BioGRID]	4506063	
1	1			1	SNRPC [BioGRID]	4507127	
1	1	2	1	8	RPL19 [BioGRID]	4506609	
1	1			1	RBM17 [BioGRID]	14249678	
2	1		2	9	SNRPE [BioGRID]	4507129	
1	1			4	EXOSC2 [BioGRID]	19923403	
5	1				VARS [BioGRID]	5454158	
	1	1		1	NOP58 [BioGRID]	7706254	

1	1	7			ATP5A1 [BioGRID]	4757810	
4	1				DDX46 [BioGRID]	41327773	
4	1	1	21	17	RPS10 [BioGRID]	4506679	
2	1	2			PSMD6 [BioGRID]	7661914	
4	1	1	8	6	RPS9 [BioGRID]	14141193	
4	1		3	4	STRAP [BioGRID]	148727341	
16	1				CKAP5 [BioGRID]	57164942	
4	1	1	1	14	RPS2 [BioGRID]	15055539	
6	1	1				195539395	
1	1		11	3	CAPRIN1 [BioGRID]	42558250	
12	1	1	113		EIF3E [BioGRID]	4503521	
2	1	1			AIMP1 [BioGRID]	45006986	
	1			2	GRSF1 [BioGRID]	149193319	
	1				TBL3 [BioGRID]	19913369	
1	1		1	1	C14orf166 [BioGRID]	7706322	
2	1			3	CDC5L [BioGRID]	11067747	
2	1		2	6	RPL30 [BioGRID]	4506631	
	1				CFL2 [BioGRID]	14719392	
	1				ARHGEF10 [BioGRID]	62548864	
5	1		7	1	LRPPRC [BioGRID]	31621305	
4	1		3	4	STAU1 [BioGRID]	82659083	
17	1	2			SPTAN1 [BioGRID]	154759259	
1	1				PSMD4 [BioGRID]	5292161	
6	1	4	4	9	SNRPD1 [BioGRID]	5902102	
17	1	1	2	21	XRCC5 [BioGRID]	10863945	
	1				TBX4 [BioGRID]	18129690	
5	1			2	EIF5B [BioGRID]	84043963	
	1				CANX [BioGRID]	10716563	
1	1		3	6	RPL21 [BioGRID]	18104948	
	1				POLR2E [BioGRID]	14589951	
3	1	1			GDI2 [BioGRID]	6598323	
	1				MAPK6 [BioGRID]	4506091	
	1				ACTL6A [BioGRID]	4757718	
1	1	4	1		TUBAL3 [BioGRID]	13376181	
1	1	2	1	13	SNRPD2 [BioGRID]	4759158	
7	1		5	15	RPL6 [BioGRID]	16753227	
8	1	1			CCT3 [BioGRID]	58761486	
3	1		11	4	NONO [BioGRID]	34932414	
2	1	7		1	HSPD1 [BioGRID]	31542947	
	1			1	NACA [BioGRID]	5031931	
	1	2	2	1	RPS27A [BioGRID]	4506713	
	1				KDM1A [BioGRID]	58761546	
	1				GTF2H4 [BioGRID]	4504201	
	1				RYR1 [BioGRID]	113204615	
3	1	6	3	2	RPS27 [BioGRID]	4506711	
	1				PTPLAD1 [BioGRID]	117168248	
	1				ARHGEF2 [BioGRID]	15011974	
5	1		1	3	EIF6 [BioGRID]	4504771	
	1				ALDH3B1 [BioGRID]	71773324	
	1				SF3B5 [BioGRID]	13775200	
7	1	1			E2F7 [BioGRID]	145580626	
9	1	3	2	3	MARS [BioGRID]	14043022	
1	1		2	1	RPS12 [BioGRID]	14277700	
3	1		13	15	RPS17 [BioGRID]	4506693	
	1				POLR2B [BioGRID]	4505941	
8	1		1		GEMIN4 [BioGRID]	122939157	
	1	1	1		MYCBP [BioGRID]	57242777	
3	1	487	2	2	YWHAE [BioGRID]	5803225	
2	1		13	20	RPS5 [BioGRID]	13904870	
	1				TPH2 [BioGRID]	31795563	
	1				TERF2IP [BioGRID]	52627149	
	1				TNRC6B [BioGRID]	148491080	
4	1	1	1		DDX20 [BioGRID]	14251212	
	1		3			218505827	
8	1				ERC1 [BioGRID]	14149661	
	1			1	PUF60 [BioGRID]	17298690	
2	1				HDAC2 [BioGRID]	116284376	
9	1	3	1	44	PRPF8 [BioGRID]	91208426	
2	1				DCTN1 [BioGRID]	13259508	
	1	3	1		RPL10P6 [BioGRID]	169164494	
18	1	1			SPTBN1 [BioGRID]	112382250	
1	1				SF3A3 [BioGRID]	5803167	
	1				DSG1 [BioGRID]	119703744	

3	1		1		KARS [BioGRID]	5031815	
	1				HSP90B1 [BioGRID]	4507677	
	1				MIF [BioGRID]	4505185	
	1			2	HIST1H4A [BioGRID]	4504301	
	1				ZNF318 [BioGRID]	120587019	
	1					197383062	
2	1	1	22	29	CSDA [BioGRID]	20070160	
19	1				CMBL [BioGRID]	20270371	
	1				DNAH3 [BioGRID]	24308169	
	1	3			LOC389901 [BioGRID]	169216498	
	1				COL6A6 [BioGRID]	156616290	
2	1	1	2	19	RPL14 [BioGRID]	78000181	
	1				MDP1 [BioGRID]	33457311	
34			1		PRPS2 [BioGRID]	4506129	
31		1		1	PRPSAP2 [BioGRID]	4506133	
31		100	2		TUBB2A [BioGRID]	4507729	
27					PRPS1 [BioGRID]	4506127	
26		210	5		HSP90AB1 [BioGRID]	20149594	
23		4	1		DYNC1H1 [BioGRID]	33350932	
23		144			TUBA1A [BioGRID]	17986283	
18						194018537	
12		2	35	11	DDX3X [BioGRID]	87196351	
12		2		8	PRPF19 [BioGRID]	7657381	
12					KRT6B [BioGRID]	119703753	
12					ATR [BioGRID]	157266317	
10		1	101		EIF3L [BioGRID]	7705433	
10			2	26	PARP1 [BioGRID]	156523968	
10		1		1	DARS [BioGRID]	45439306	
10		158			TUBA1B [BioGRID]	57013276	
9			113		EIF3C [BioGRID]	4503525	
9		2			CCT6A [BioGRID]	4502643	
9			1	101	SNRNP200 [BioGRID]	40217847	
8		1			PSMC3 [BioGRID]	21361144	
7		5	16	37	SYNCRIP [BioGRID]	23397427	
7			26	4	GNB2L1 [BioGRID]	5174447	
7		1			TACC3 [BioGRID]	5454102	
7		2	15	6	PTBP1 [BioGRID]	4506243	
7		2	4	8	SNRPN [BioGRID]	4507135	
6		1	82		EIF3M [BioGRID]	23397429	
6		1	1	2	CPSF6 [BioGRID]	162329583	
6		1			PSMD11 [BioGRID]	28872725	
5		1		1	ACTR1A [BioGRID]	5031569	
5		2	2	2	KIAA1967 [BioGRID]	24432106	
5					PSMD2 [BioGRID]	25777602	
5			43	7	PABPC4 [BioGRID]	4504715	
5			1	2	DRG1 [BioGRID]	4758796	
5		7			KRT8 [BioGRID]	4504919	
5			7	14	IGF2BP1 [BioGRID]	56237027	
5					KRT5 [BioGRID]	119395754	
4					PSMC2 [BioGRID]	4506209	
4		2		27	KPNB1 [BioGRID]	19923142	
4			141	1	EIF3D [BioGRID]	4503523	
4					PSMD13 [BioGRID]	157502193	
4					PSMC4 [BioGRID]	5729991	
4		3	2	2	NAP1L1 [BioGRID]	4758756	
4		1	1		QARS [BioGRID]	4826960	
4		1	392	1	EIF4A1 [BioGRID]	4503529	
4			46		EIF3K [BioGRID]	10801345	
4					PSMD14 [BioGRID]	5031981	
4			11	3	RPS15 [BioGRID]	4506687	
4			4	1	C22orf28 [BioGRID]	7657015	
4					HSPA1L [BioGRID]	124256496	
4					KRT16 [BioGRID]	24430192	
4		1		3	PPP1CA [BioGRID]	4506003	
4		1	24	33	RPS3A [BioGRID]	4506723	
3		1			FASN [BioGRID]	41872631	
3			17	4	LARP1 [BioGRID]	39725634	
3				1	PPP1CC [BioGRID]	4506007	
3			1	2	LUC7L3 [BioGRID]	19923485	
3					MAP3K7IP1 [BioGRID]	5174703	
3						222136639	
3			7	25	RPL3 [BioGRID]	4506649	
3		5			STIP1 [BioGRID]	5803181	

3		1			PSMD7 [BioGRID]	25777615	
3				1	EEF1E1 [BioGRID]	4758862	
3			5	1	EIF2S3 [BioGRID]	4503507	
3		1	19	16	RPS24 [BioGRID]	4506703	
3					FRYL [BioGRID]	119874201	
3				15	PA2G4 [BioGRID]	124494254	
3			4	9	RPL13 [BioGRID]	15431295	
3			5	6	IGF2BP3 [BioGRID]	30795212	
2		1	1		GEMIN5 [BioGRID]	157739942	
2		1			PSMD12 [BioGRID]	4506221	
2					SMC4 [BioGRID]	50658063	
2					CFL1 [BioGRID]	5031635	
2					C20orf11 [BioGRID]	8923557	
2		1	1	3	GRWD1 [BioGRID]	31542862	
2		7			DNAJA2 [BioGRID]	5031741	
2					EEF2 [BioGRID]	4503483	
2			1	17	RPL4 [BioGRID]	16579885	
2		2			PSMD3 [BioGRID]	25777612	
2		1		1	SF3B14 [BioGRID]	7706326	
2				2	BCAS2 [BioGRID]	5031653	
2						207452735	
2					DYNC1I2 [BioGRID]	24307879	
2					CORO1C [BioGRID]	7656991	
2					CAPZA1 [BioGRID]	5453597	
2					ANKFY1 [BioGRID]	110815813	
2					MYL12A [BioGRID]	5453740	
2		1	2	4	SERBP1 [BioGRID]	66346679	
2			2	3	RPS23 [BioGRID]	4506701	
2				3	C1orf25 [BioGRID]	163792194	
2					SMC2 [BioGRID]	110347418	
2			1		EXOSC8 [BioGRID]	31415882	
2					TXNDC12 [BioGRID]	7705696	
2		1	2	16	RPL31 [BioGRID]	4506633	
2				5	RPL18A [BioGRID]	11415026	
2		1	2	3	CCNA1 [BioGRID]	4502611	
2					SNRPB2 [BioGRID]	4507123	
2		5	5	19	HNRNPF [BioGRID]	4826760	
2		1			SF3B4 [BioGRID]	5032069	
2				13	RPL37A [BioGRID]	4506643	
2		5		1	ATP5B [BioGRID]	32189394	
1		1	6	7	RPL24 [BioGRID]	4506619	
1		3	2	1	TXN [BioGRID]	50592994	
1			1	1	SNRNP70 [BioGRID]	29568103	
1			5		DDX1 [BioGRID]	4826686	
1				1	SLAIN2 [BioGRID]	149588928	
1					RNF219 [BioGRID]	88759348	
1				1	TSR1 [BioGRID]	39780588	
1		1			PDHB [BioGRID]	156564403	
1					PSMD1 [BioGRID]	25777600	
1		1			PSMB1 [BioGRID]	4506193	
1					GCN1L1 [BioGRID]	54607053	
1				2	ZRANB2 [BioGRID]	42741682	
1					UNC45A [BioGRID]	29725607	
1			1	1	MSI2 [BioGRID]	20373175	
1				1	KIF2A [BioGRID]	148612849	
1			1	2	LUC7L2 [BioGRID]	116812577	
1				2	USP15 [BioGRID]	14149627	
1		1	1		AKAP8 [BioGRID]	5031579	
1				1	SMN1 [BioGRID]	4507091	
1			13		ASCC3 [BioGRID]	76880486	
1					DIABLO [BioGRID]	9845297	
1		1			RUVBL2 [BioGRID]	5730023	
1		1		11	SNRPG [BioGRID]	4507133	
1					MYL6 [BioGRID]	17986258	
1			1		RBM3 [BioGRID]	5803137	
1				1	NSUN2 [BioGRID]	39995082	
1				3	CMAS [BioGRID]	8923900	
1					EIF2B1 [BioGRID]	4503503	
1					HOMER1 [BioGRID]	4759198	
1			2			222352151	
1			3	5	CIRBP [BioGRID]	4502847	
1			22	9	G3BP1 [BioGRID]	5031703	
1			205		EIF4G1 [BioGRID]	38201619	

1					NMT1 [BioGRID]	10835073	
1					GTF3C4 [BioGRID]	156119605	
1		197			YWHAB [BioGRID]	4507949	
1					AIMP2 [BioGRID]	11125770	
1					NCAPD3 [BioGRID]	45356151	
1					DPY30 [BioGRID]	14211889	
1					NASP [BioGRID]	27262628	
1		1	18	16	RPS7 [BioGRID]	4506741	
1					OGT [BioGRID]	32307148	
1					C21orf45 [BioGRID]	9506437	
1					PSMA1 [BioGRID]	4506179	
1			3	2	U2AF2 [BioGRID]	6005926	
1					OTUD4 [BioGRID]	40807496	
1						212276070	
1					HDLBP [BioGRID]	4885409	
1					CCDC124 [BioGRID]	19923969	
1					SFRS11 [BioGRID]	4759100	
1				1	RPL36P8 [BioGRID]	88976416	
1					RPL38 [BioGRID]	4506645	
1				4	RPL27A [BioGRID]	4506625	
1			1		NCBP1 [BioGRID]	4505343	
1					SIP1 [BioGRID]	4506961	
1			1	1	THOC4 [BioGRID]	55770864	
1					ZC3H15 [BioGRID]	118150660	
1				1	PSMB4 [BioGRID]	22538467	
1					PHB2 [BioGRID]	6005854	
1				6	ADAR [BioGRID]	70166852	
1					NCAPG [BioGRID]	21359945	
1					COPB2 [BioGRID]	4758032	
1			1	3	MRPL12 [BioGRID]	27436901	
1				14	MYBBP1A [BioGRID]	157694492	
1					TCOF1 [BioGRID]	57164975	
1					RFC5 [BioGRID]	6677723	
1					WAPAL [BioGRID]	42734325	
1					ZYX [BioGRID]	4508047	
1					AP2B1 [BioGRID]	4557469	
1					CETN2 [BioGRID]	4757902	
1		30			CDC37 [BioGRID]	5901922	
1					MYH9 [BioGRID]	12667788	
1			3		HNRNPC [BioGRID]	117189975	
1				4	C7orf50 [BioGRID]	14150149	
1				11	SNRPF [BioGRID]	4507131	
1			1	16	USP39 [BioGRID]	56550051	
1					WDR26 [BioGRID]	55743153	
1					ARL6IP4 [BioGRID]	50409691	
1					BTF3 [BioGRID]	20070130	
1					LRRK1 [BioGRID]	90991702	
1					CNOT7 [BioGRID]	85067505	
1			9	1	SFPQ [BioGRID]	4826998	
1				2	RFC4 [BioGRID]	4506491	
1					CCDC25 [BioGRID]	108936950	
1					NKRF [BioGRID]	63003897	
1			13	7	FUS [BioGRID]	4826734	
1			1	4	ABCF1 [BioGRID]	10947135	
1					ADD1 [BioGRID]	29826319	
1					PSMA6 [BioGRID]	23110944	
1					CSTB [BioGRID]	4503117	
1					LAMA2 [BioGRID]	28559088	
1			3		SKP1 [BioGRID]	25777711	
1		2			BAG2 [BioGRID]	4757834	
1					ZBTB24 [BioGRID]	7662128	
1				1	WDR61 [BioGRID]	13376840	
1					SNRPA [BioGRID]	4759156	
1					KPNA4 [BioGRID]	4504901	
1				5	PDCD11 [BioGRID]	70980549	
1					CSE1L [BioGRID]	29029559	
1					STRBP [BioGRID]	21361745	
1					PSMD10 [BioGRID]	4506217	
1			79		EIF3I [BioGRID]	4503513	
1					PSMB5 [BioGRID]	4506201	
1					RANBP2 [BioGRID]	150418007	
1					LOC100130283 [BioGRID]	169210373	
1					CAP1 [BioGRID]	5453595	

1					PSMB7 [BioGRID]	4506203	
1					TMEM188 [BioGRID]	31542781	
1					ACTBL2 [BioGRID]	63055057	
1					TMOD2 [BioGRID]	7657647	
		353			RAF1 [BioGRID]	4506401	
		240			TUBB2B [BioGRID]	29788768	
		236	13		TUBB4 [BioGRID]	21361322	
		221			YWHAZ [BioGRID]	4507953	
		210			TUBB3 [BioGRID]	50592996	
		191			TUBB6 [BioGRID]	14210536	
		170			YWHAG [BioGRID]	21464101	
		168			YWHAQ [BioGRID]	5803227	
		166			YWHAH [BioGRID]	4507951	
		131			RP11-631M21.2 [BioGRID]	42558279	
		52			NRAS [BioGRID]	4505451	
		36			FKBP5 [BioGRID]	4758384	
		30			KRAS [BioGRID]	15718761	
		23			TIMM50 [BioGRID]	48526509	
		20			KRT19 [BioGRID]	24234699	
		17			AIFM1 [BioGRID]	4757732	
		11			DIS3L [BioGRID]	19115966	
		7		2	TIMM13 [BioGRID]	11024700	
		6			AKAP8L [BioGRID]	49472841	
		5			CAD [BioGRID]	18105007	
		5			RCN1 [BioGRID]	4506455	
		5			SEH1L [BioGRID]	61743969	
		4			ATP1A1 [BioGRID]	21361181	
		4				205360838	
		4			ATP2A2 [BioGRID]	4502285	
		4			HAX1 [BioGRID]	13435356	
		4			STUB1 [BioGRID]	56181387	
		3			CLPB [BioGRID]	13540606	
		3				223278387	
		3			ALDH3A2 [BioGRID]	4557303	
		3			KRT18 [BioGRID]	4557888	
		3			MLLT11 [BioGRID]	5802968	
		3			RAP1A [BioGRID]	4506413	
		2	1		HLA-C [BioGRID]	52630342	
		2			FKBP8 [BioGRID]	52630440	
		2			SEC16A [BioGRID]	124378039	
		2				197927448	
		2			GRAMD1A [BioGRID]	92110010	
		2			EMD [BioGRID]	4557553	
		2			LGALS3BP [BioGRID]	5031863	
		2			MAD2L1 [BioGRID]	4505067	
		2			ARL1 [BioGRID]	4502227	
		2			SLC25A3 [BioGRID]	4505775	
		2			DNAJC2 [BioGRID]	94538370	
		2			CHCHD4 [BioGRID]	21389469	
		2				217272887	
		2			CALU [BioGRID]	4502551	
		2			TMX3 [BioGRID]	38505222	
		2			FDFT1 [BioGRID]	67089147	
		1			ASPH [BioGRID]	9910364	
		1			NME2 [BioGRID]	4505409	
		1			IPO8 [BioGRID]	53759103	
		1			GLUD1 [BioGRID]	4885281	
		1			SDF4 [BioGRID]	7706573	
		1			PPT1 [BioGRID]	4506031	
		1			FBL [BioGRID]	12056465	
		1			ESYT1 [BioGRID]	14149680	
		1			ARF1 [BioGRID]	4502201	
		1		1	MYL6B [BioGRID]	4505303	
		1			TRIM41 [BioGRID]	42516572	
		1			AK2 [BioGRID]	4502013	
		1			DCHS1 [BioGRID]	16933557	
		1				194306640	
		1			MRPS22 [BioGRID]	9910244	
		1	6	1	TRA2B [BioGRID]	4759098	
		1			MYO3B [BioGRID]	134288892	
		1			ESYT2 [BioGRID]	45387945	
		1			HR [BioGRID]	22547204	
		1			CNOT1 [BioGRID]	42716275	

		1			PCDH7 [BioGRID]	14589931	
		1			RAB1A [BioGRID]	4758988	
		1			DCAF7 [BioGRID]	108936958	
		1			SLC25A25 [BioGRID]	56699401	
		1			ZNF507 [BioGRID]	55925468	
		1			CCNB3 [BioGRID]	90669307	
		1			DNAH8 [BioGRID]	126012497	
		1			SLC25A5 [BioGRID]	156071459	
		1			FAR1 [BioGRID]	24308324	
		1			BAG5 [BioGRID]	6631077	
		1			SRP19 [BioGRID]	4507213	
		1			TCEB1 [BioGRID]	5032161	
		1			RRAS [BioGRID]	5454028	
		1			SLC39A7 [BioGRID]	117553608	
		1			SEC22B [BioGRID]	94429050	
		1			CDIPT [BioGRID]	5453906	
		1			ROCK1 [BioGRID]	4885583	
		1	4	9	HIST1H2BG [BioGRID]	4504257	
		1			CKMT1B [BioGRID]	10334859	
		1	2	1	YTHDF2 [BioGRID]	116812575	
		1			FARSA [BioGRID]	4758340	
		1			GLA [BioGRID]	4504009	
		1			DBT [BioGRID]	110671329	
		1			PTGES3 [BioGRID]	23308579	
		1			NDUFS3 [BioGRID]	4758788	
		1			KIAA0913 [BioGRID]	122114658	
		1			TPP1 [BioGRID]	5729770	
		1	1		MSI1 [BioGRID]	4505255	
		1			BCL11B [BioGRID]	12597635	
		1			FANCD2 [BioGRID]	21361861	
		1			RUVBL1 [BioGRID]	4506753	
		1			SAR1B [BioGRID]	7705827	
		1			DDOST [BioGRID]	20070197	
		1	1		TARDBP [BioGRID]	6678271	
		1			CTSB [BioGRID]	4503139	
		1			MET [BioGRID]	42741655	
		1			THOC3 [BioGRID]	14150171	
		1			PITRM1 [BioGRID]	41352061	
		1	6	9	SFRS1 [BioGRID]	5902076	
		1			QPCTL [BioGRID]	92110027	
		1		2	RPL35A [BioGRID]	16117791	
		1			TUFM [BioGRID]	34147630	
		1			ATP1B3 [BioGRID]	4502281	
		1			SCN1B [BioGRID]	4506805	
		1			PKD1L2 [BioGRID]	116006953	
		1			KIAA1274 [BioGRID]	166235184	
		1			FANCI [BioGRID]	82830440	
		1			C1orf57 [BioGRID]	14150100	
		1				218777837	
		1	4		MATR3 [BioGRID]	21626466	
		1			CAND2 [BioGRID]	112420977	
		1			FAM92A2 [BioGRID]	169208597	
		1			HAUS5 [BioGRID]	149944680	
		1			NT5DC2 [BioGRID]	12597653	
		1			ADCK4 [BioGRID]	27363457	
		1			XPO1 [BioGRID]	4507943	
		1			C6orf201 [BioGRID]	145966784	
		1	1	3	LRRC59 [BioGRID]	40254924	
		1			PPP2CA [BioGRID]	4506017	
			709		EIF4A2 [BioGRID]	83700235	
			76		EIF4G3 [BioGRID]	10092601	
			65		PDCD4 [BioGRID]	21735596	
			61		EIF4G2 [BioGRID]	4503539	
			32		EIF3J [BioGRID]	83281438	
			29		EIF4A3 [BioGRID]	7661920	
			26		BAT2 [BioGRID]	149158690	
			23		LOC100131720 [BioGRID]	169188690	
			21		CSNK2A1 [BioGRID]	4503095	
			20		BAT2L2 [BioGRID]	115298682	
			13		BAT2L1 [BioGRID]	149192855	
			12	8	RPS2P5 [BioGRID]	169204454	
			12	12	RPS13 [BioGRID]	4506685	
			10		IGF2BP2 [BioGRID]	56118219	

			10		G3BP2 [BioGRID]	19923399	
			9		DDX6 [BioGRID]	164664518	
			9		SFRS7 [BioGRID]	72534660	
			8		NUFIP2 [BioGRID]	32698730	
			8		C10orf137 [BioGRID]	31742484	
			8		ASCC2 [BioGRID]	20270253	
			7		ABCE1 [BioGRID]	108773782	
			7	8	SFRS3 [BioGRID]	4506901	
			6		EID2 [BioGRID]	63175652	
			6	1	SRRM2 [BioGRID]	118572613	
			6		TAF15 [BioGRID]	4507353	
			6		HNRNPH2 [BioGRID]	9624998	
			5		LSM12 [BioGRID]	22748747	
			5		UPF1 [BioGRID]	18375673	
			5		DHX29 [BioGRID]	67782362	
			4	3	RPS11 [BioGRID]	4506681	
			4		GATAD2B [BioGRID]	21218438	
			4		ATXN2L [BioGRID]	27262645	
			4		CDV3 [BioGRID]	8923710	
			4		ATXN2 [BioGRID]	171543895	
			4		CSNK2A2 [BioGRID]	4503097	
			4		ELAVL1 [BioGRID]	38201714	
			4		SFRS13A [BioGRID]	5730079	
			4		EIF1B [BioGRID]	5031711	
			4		FXR2 [BioGRID]	4758410	
			4		LARP4 [BioGRID]	40353736	
			4		EIF4E [BioGRID]	4503535	
			4		CSNK2B [BioGRID]	23503295	
			3		OBSL1 [BioGRID]	144226847	
			3		LGTN [BioGRID]	56699485	
			3	4	DHX30 [BioGRID]	20336290	
			3	3	RBMXL1 [BioGRID]	21361809	
			3	6	EWSR1 [BioGRID]	4885225	
			3		USP10 [BioGRID]	119220605	
			2		RBM14 [BioGRID]	5454064	
			2	11	RPS20 [BioGRID]	4506697	
			2		MACF1 [BioGRID]	33188443	
			2		FAM98A [BioGRID]	56699482	
			2		FAM134B [BioGRID]	77917617	
			2		CLTCL1 [BioGRID]	108860681	
			2		PLEKHH1 [BioGRID]	55741447	
			2		TLE3 [BioGRID]	157384982	
			2		LARP4B [BioGRID]	57634536	
			2	1	PCBP2 [BioGRID]	14141166	
			2		DAZAP1 [BioGRID]	25470886	
			2	4	PURA [BioGRID]	5032007	
			2		ASCC1 [BioGRID]	56676391	
			2		KHDRBS1 [BioGRID]	5730027	
			2		DDX49 [BioGRID]	31542656	
			2		PAIP1 [BioGRID]	17511255	
			2		PAIP2 [BioGRID]	75677349	
			2	1	PURB [BioGRID]	15147219	
			2		DDX50 [BioGRID]	13129006	
			2	6	HIST1H2AI [BioGRID]	4504239	
			2		ZCCHC3 [BioGRID]	29648305	
			2	5	SFRS6 [BioGRID]	20127499	
			2		RPS7P10 [BioGRID]	169205846	
			2		SMO [BioGRID]	5032099	
			2	26	RPL13A [BioGRID]	6912634	
			2		U2AF1 [BioGRID]	5803207	
			1		PAIP2B [BioGRID]	148235699	
			1	18	RPS26 [BioGRID]	15011936	
			1	1	PABPN1 [BioGRID]	4758876	
			1		FXR1 [BioGRID]	61835148	
			1		RBMS3 [BioGRID]	51317351	
			1		TAGAP [BioGRID]	21314774	
			1		LARGE [BioGRID]	4758664	
			1		TRPC1 [BioGRID]	4507685	
			1		RPL29 [BioGRID]	4506629	
			1		RBM9 [BioGRID]	133925803	
			1		NALCN [BioGRID]	24119274	
			1		RALGAPA1 [BioGRID]	51226124	
			1		RPL6P19 [BioGRID]	169171527	

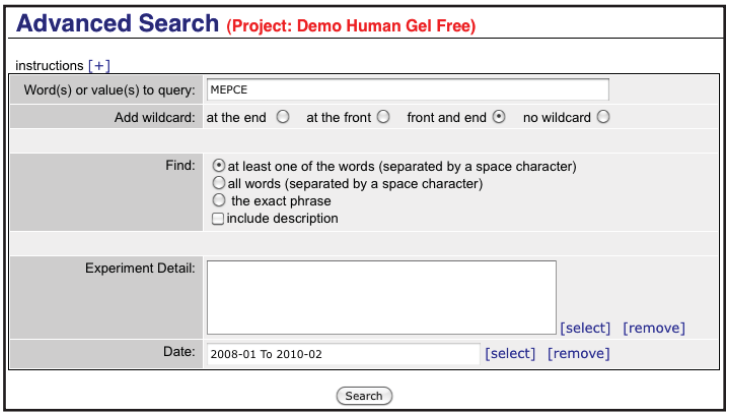
			1		GULP1 [BioGRID]	7705318	
			1	32	RPL35 [BioGRID]	6005860	
			1		POU5F1 [BioGRID]	42560248	
			1		PYCR2 [BioGRID]	21361454	
			1		FAM195A [BioGRID]	33300670	
			1	3	RPL26 [BioGRID]	4506621	
			1		AP2A1 [BioGRID]	19913414	
			1		KCTD5 [BioGRID]	9506651	
			1		TRIP4 [BioGRID]	32189376	
			1		HADHA [BioGRID]	20127408	
			1		UBAP2L [BioGRID]	188497756	
			1		FUBP3 [BioGRID]	100816392	
			1	1	TRA2A [BioGRID]	9558733	
			1	1	SFRS9 [BioGRID]	4506903	
			1		PUM1 [BioGRID]	13491166	
			1		BTRC [BioGRID]	4502477	
			1	5	SFRS4 [BioGRID]	21361282	
			1		FAM195B [BioGRID]	126722894	
			1		PTCH1 [BioGRID]	134254432	
			1		FAU [BioGRID]	4503659	
			1	2	POLDIP2 [BioGRID]	7661672	
			1	1	BAT1 [BioGRID]	4758112	
			1		YTHDF1 [BioGRID]	31377750	
			1		CAPS2 [BioGRID]	166795238	
			1			209862881	
			1		SOS2 [BioGRID]	154689780	
			1		SMCHD1 [BioGRID]	148839305	
			1	24	RPL8 [BioGRID]	4506663	
			1		ELP3 [BioGRID]	23510283	
			1		XPO7 [BioGRID]	154448892	
			1		RPS6KA2 [BioGRID]	19923570	
			1		PTGFR [BioGRID]	4506261	
			1		LOC100129791 [BioGRID]	169167619	
			1		HELZ [BioGRID]	54792138	
				242	SART3 [BioGRID]	7661952	
				184	MEPCE [BioGRID]	47271406	
				176	LARP7 [BioGRID]	109809739	
				49	LSM8 [BioGRID]	7706425	
				35	PRPF4 [BioGRID]	24431950	
				34	ENDOG [BioGRID]	53759134	
				32	PRPF3 [BioGRID]	4758556	
				28	SART1 [BioGRID]	10863889	
				24	RPL7P23 [BioGRID]	88988289	
				22	PPIH [BioGRID]	5454154	
				17	LSM4 [BioGRID]	6912486	
				16	CDK9 [BioGRID]	4502747	
				16	METT10D [BioGRID]	122114654	
				15	CCNT1 [BioGRID]	17978466	
				13	PRPF6 [BioGRID]	40807485	
				12	HEXIM1 [BioGRID]	5453682	
				12	HNRNPUL1 [BioGRID]	21536320	
				9	SNRNP40 [BioGRID]	115298668	
				9	LSM2 [BioGRID]	10863977	
				8	RRP1B [BioGRID]	57863269	
				7	H1FX [BioGRID]	5174449	
				7	RPL27 [BioGRID]	4506623	
				7	LSM6 [BioGRID]	5901998	
				7	NOP2 [BioGRID]	76150623	
				7	TXNL4A [BioGRID]	5729802	
				6	LSM7 [BioGRID]	7706423	
				6	CCNT2 [BioGRID]	4502629	
				6	AFF4 [BioGRID]	7656879	
				6	HP1BP3 [BioGRID]	56676330	
				6	DDX18 [BioGRID]	38327634	
				6	ZNF511 [BioGRID]	28274701	
				5	LSM5 [BioGRID]	6912488	
				5	NHP2L1 [BioGRID]	4826860	
				5	MKI67 [BioGRID]	103472005	
				5	EBNA1BP2 [BioGRID]	5803111	
				5	SNRNP27 [BioGRID]	24307919	
				5	TLN2 [BioGRID]	156938343	
				4	RSL1D1 [BioGRID]	118498359	
				4	LSM3 [BioGRID]	7657315	

				4	SRPK1 [BioGRID]	47419936	
				4	KPNA3 [BioGRID]	34485722	
				4	RBM39 [BioGRID]	4757926	
				4	CDK12 [BioGRID]	157817023	
				4	DDX56 [BioGRID]	9506931	
				3	PRPF4B [BioGRID]	89276756	
				3	PPAN [BioGRID]	47174861	
				3		194097365	
				3	YTHDC2 [BioGRID]	38505213	
				3	RBM42 [BioGRID]	21359951	
				3	CD2BP2 [BioGRID]	5174409	
				3	TUT1 [BioGRID]	12383074	
				3	PWP1 [BioGRID]	5902034	
				3	POPDC2 [BioGRID]	22209004	
				3	TRMU [BioGRID]	31542641	
				3	GLB1 [BioGRID]	119372308	
				3	ZC3HAV1 [BioGRID]	27477138	
				3	ZFR [BioGRID]	34101286	
				3	RPL34 [BioGRID]	16117787	
				3	TPM3 [BioGRID]	114155140	
				3	RPL28 [BioGRID]	13904866	
				2	EXOSC7 [BioGRID]	189083688	
				2	EXOSC5 [BioGRID]	47174864	
				2	PLRG1 [BioGRID]	4505895	
				2	LTV1 [BioGRID]	21361875	
				2		223555917	
				2	BRIX1 [BioGRID]	55770900	
				2	CHERP [BioGRID]	119226260	
				2	GLYR1 [BioGRID]	40556376	
				2	GTPBP4 [BioGRID]	55953087	
				2		222352111	
				2	HNRNPUL2 [BioGRID]	118601081	
				2	CLASP1 [BioGRID]	31563537	
				2	RNMTL1 [BioGRID]	8922534	
				2	RPL32 [BioGRID]	4506635	
				2	EIF2AK2 [BioGRID]	4506103	
				2	NOC3L [BioGRID]	20806097	
				2	BMS1 [BioGRID]	41281483	
				2	DDX54 [BioGRID]	51094101	
				2	AKAP7 [BioGRID]	7706527	
				2	PTBP2 [BioGRID]	10863997	
				2	EXOSC10 [BioGRID]	4505917	
				2	CDK17 [BioGRID]	37595545	
				2	EDC4 [BioGRID]	45827771	
				2	FAM107B [BioGRID]	40254982	
				1	DDX55 [BioGRID]	41327779	
				1	DNAJC9 [BioGRID]	27597059	
				1	ERI1 [BioGRID]	31543184	
				1	KRR1 [BioGRID]	117676403	
				1	SND1 [BioGRID]	77404397	
				1	TMCO6 [BioGRID]	94536840	
				1	RFC1 [BioGRID]	32528306	
				1	NOP16 [BioGRID]	148747209	
				1	RBM34 [BioGRID]	38016127	
				1	TROVE2 [BioGRID]	31377800	
				1	H1F0 [BioGRID]	4885371	
				1	EXOSC1 [BioGRID]	7705612	
				1	XRN2 [BioGRID]	18860916	
				1	MLLT1 [BioGRID]	21361272	
				1	NOC2L [BioGRID]	157694511	
				1	RRP15 [BioGRID]	126362983	
				1	C20orf26 [BioGRID]	34222104	
				1	DYRK1B [BioGRID]	4758222	
				1		194328699	
				1	PAPD5 [BioGRID]	94536832	
				1	HEXIM2 [BioGRID]	21389407	
				1	IYD [BioGRID]	42794271	
				1		193211480	
				1	GNL2 [BioGRID]	7019419	
				1	APOA2 [BioGRID]	4502149	
				1	DDX52 [BioGRID]	38569503	
				1	RIMBP2 [BioGRID]	140561070	
				1	OR4L1 [BioGRID]	52317259	

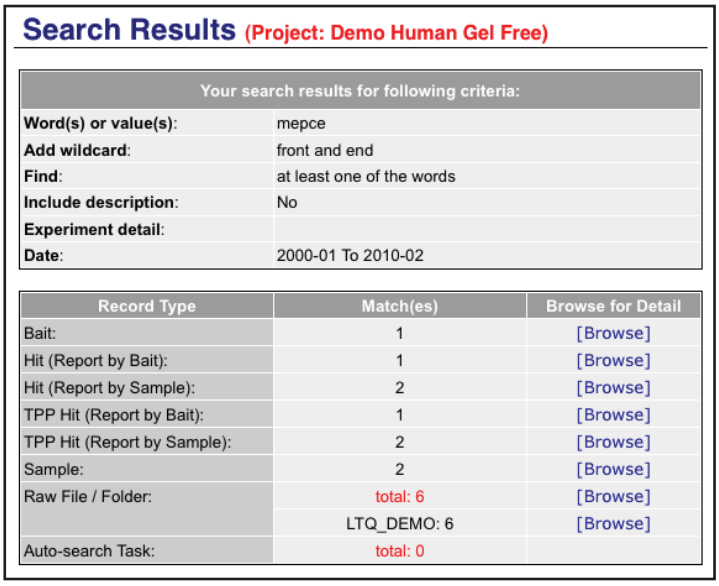
				1	LYAR [BioGRID]	8923398	
				1	RCL1 [BioGRID]	157426877	
				1	SYNE2 [BioGRID]	118918403	
				1	ANKRD34A [BioGRID]	89941470	
				1	RNF168 [BioGRID]	31377566	
				1	TAOK3 [BioGRID]	148746218	
				1	KIF23 [BioGRID]	6754472	
				1	RPL7L1 [BioGRID]	50053872	
				1	NOP14 [BioGRID]	55769587	
				1	BRI3BP [BioGRID]	19923665	
				1	NOL6 [BioGRID]	18644728	
				1	PARN [BioGRID]	4505611	
				1	CLASP2 [BioGRID]	57863301	
				1	RANGAP1 [BioGRID]	4506411	
				1	HNRNPH3 [BioGRID]	14141157	
				1	MOV10 [BioGRID]	14211540	
				1	DDX47 [BioGRID]	20149629	
				1	CHCHD2 [BioGRID]	7705851	
				1	COX17 [BioGRID]	5031645	
				1	PAF1 [BioGRID]	42476169	
				1	GNL3 [BioGRID]	45593130	
				1	GTPBP1 [BioGRID]	82546879	
				1	SPATS2 [BioGRID]	148746220	
				1	MYEF2 [BioGRID]	33620747	
				1	EIF4E1B [BioGRID]	150378456	
				1	LEO1 [BioGRID]	20270337	
				1	MAPKAPK2 [BioGRID]	32481209	
				1	TRIM25 [BioGRID]	68160937	
				1	COIL [BioGRID]	4758024	
				1	EIF2A [BioGRID]	54873624	
				1	PWP2 [BioGRID]	48762926	
				1	PGAM5 [BioGRID]	20070384	
				1	LLPH [BioGRID]	14150124	
				1	CTSA [BioGRID]	119395729	
				1	DIMT1L [BioGRID]	7657198	
				1	PRKRA [BioGRID]	4505581	
				1	ZNF346 [BioGRID]	6912440	
				1	RFC2 [BioGRID]	28882049	
				1	CCNB1 [BioGRID]	14327896	
				1	EXOSC3 [BioGRID]	50511939	
				1	CPNE3 [BioGRID]	4503015	
				1	PES1 [BioGRID]	7657455	
				1	TM9SF1 [BioGRID]	21361315	
				1	KNTC1 [BioGRID]	7661960	

Note that a red number indicates that all peptides assigned to this entry are shared with at least one additional entry in the database. Mouse over the number for details.

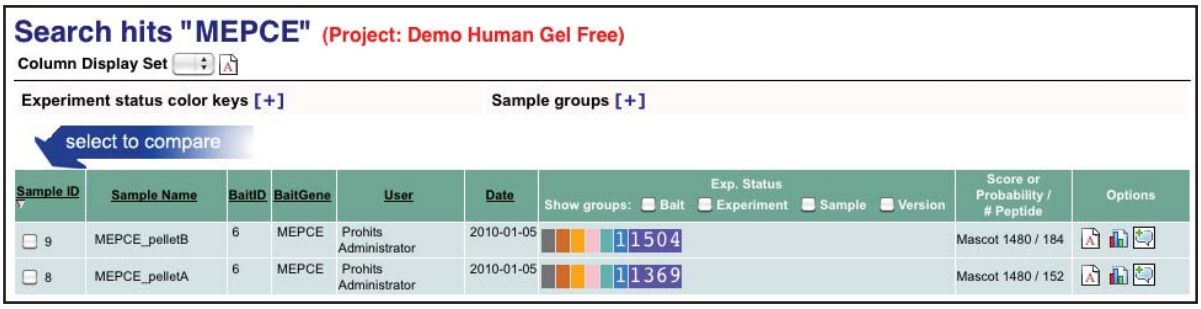
A



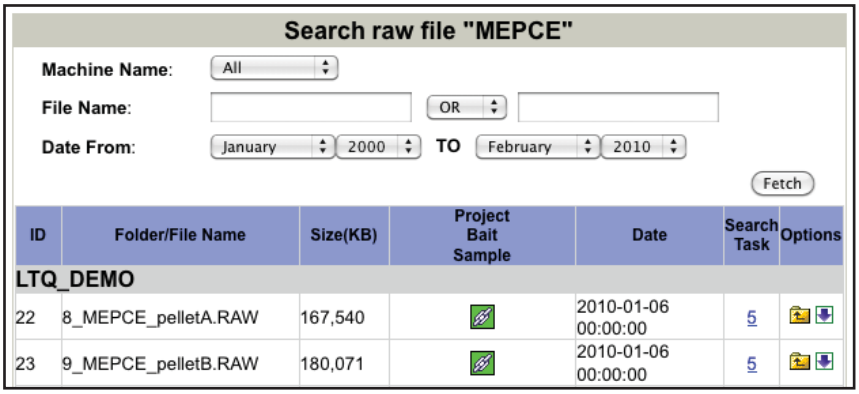
B



C

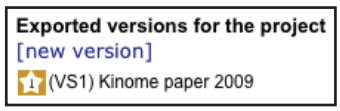


D

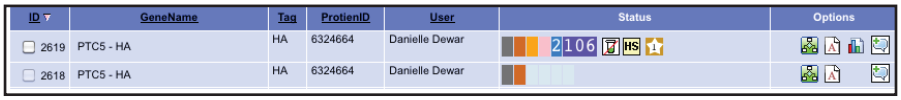


Supplementary Figure 18. Advanced Search Function in the Analyst module. A) Entry page for the Advanced Search Function. Gene Name or other keywords may be used for searches; wildcards may be employed. Searches may be restricted by Experimental Detail (controlled vocabulary) and by date, if desired (here we are restricting by date). B) Summary page of the search results. MEPCE was analyzed once as a bait (2 samples), and recovered 2 times as a hit across all samples in the project. Clicking on the [Browse] button will list all instances. C) Detail of the recovery of MEPCE as a hit. A column is added that lists the search engine score (here Mascot = 1480) and the number of spectra (184 and 182). D) Detail of the raw files (top of the list only) that have "MEPCE" in their name.

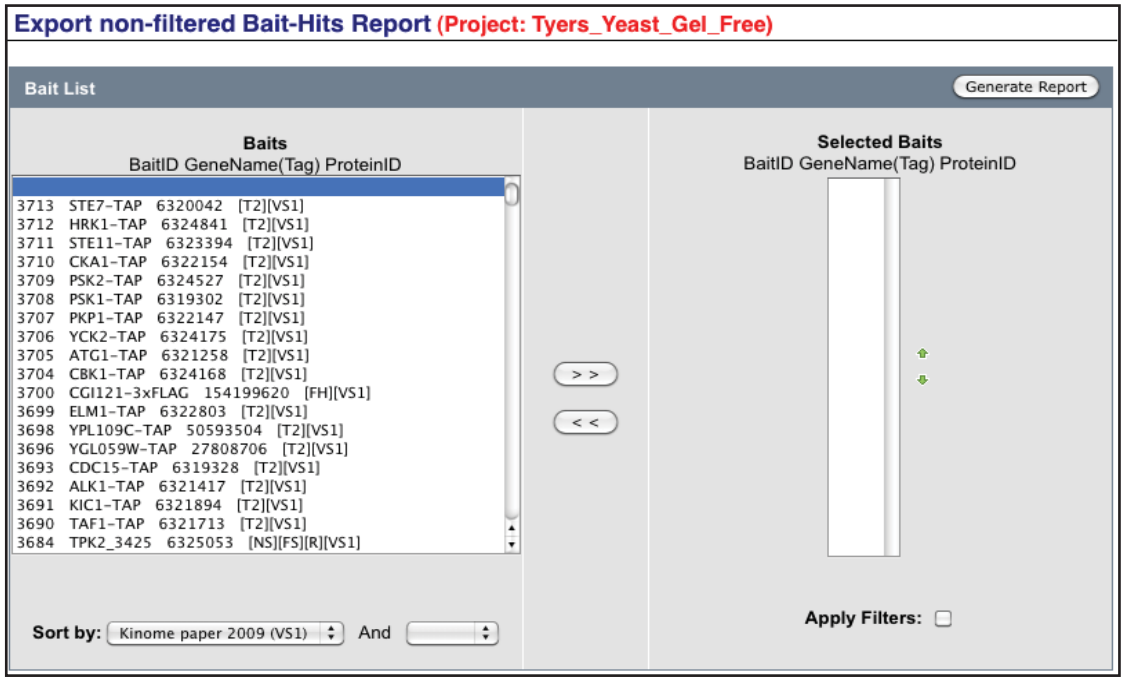
A



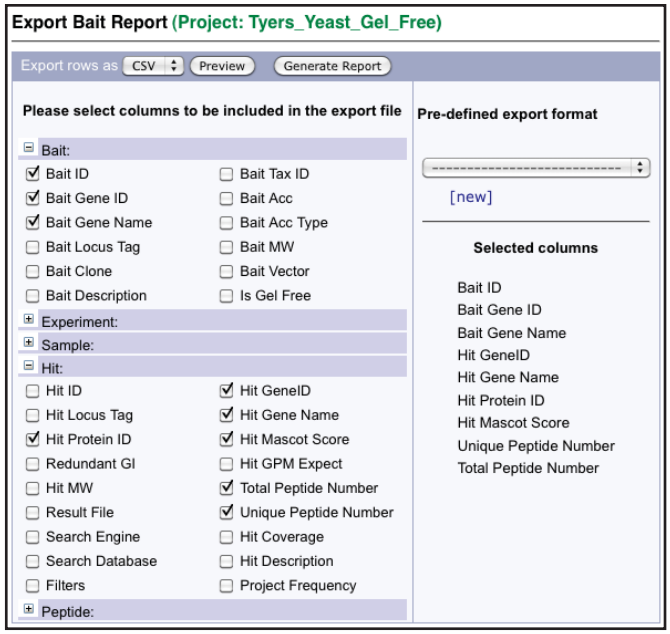
B



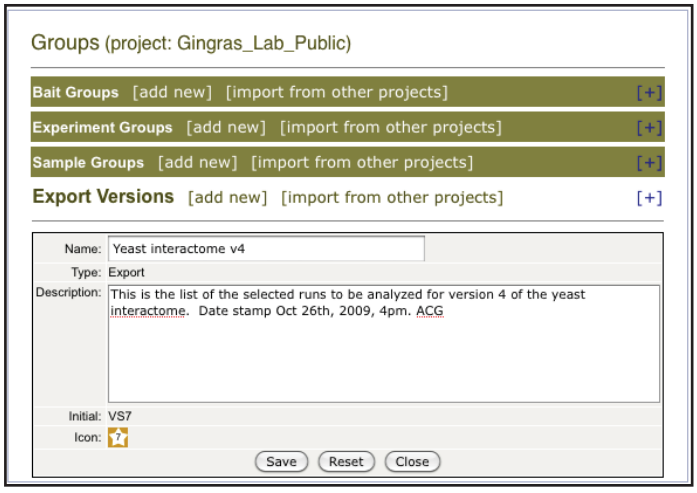
C



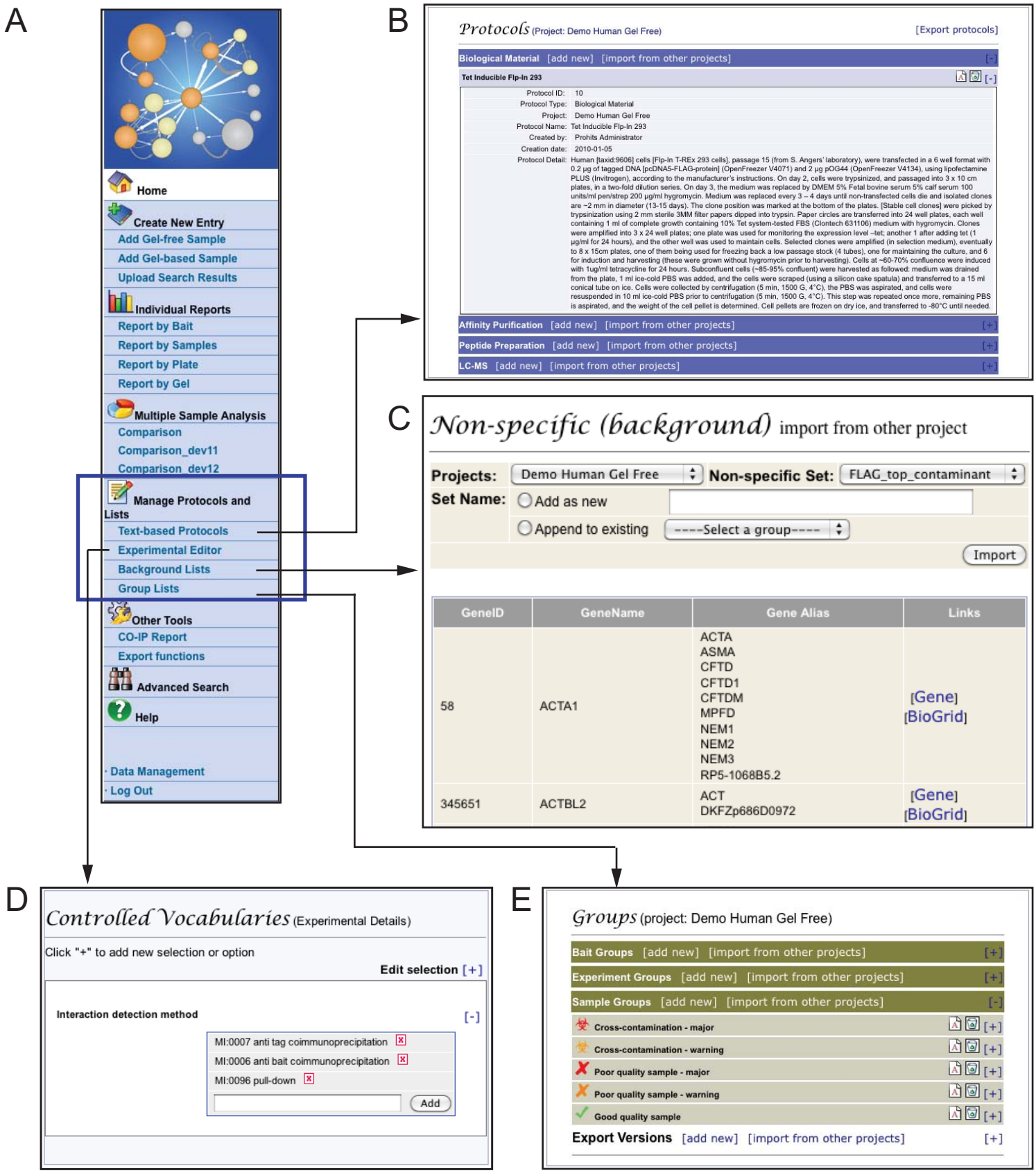
D



E

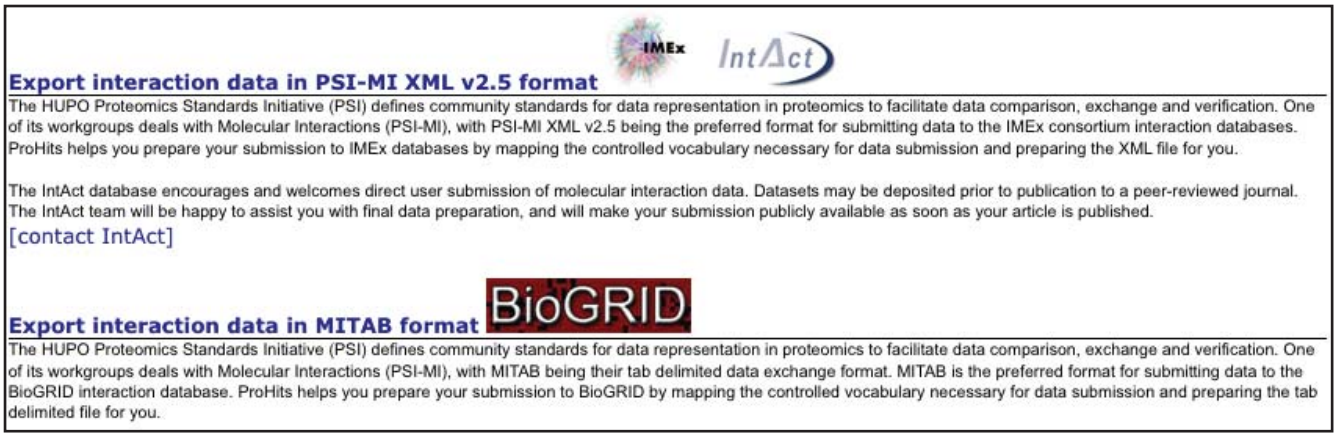


Supplementary Figure 19. Exporting hits lists for a selected set of Samples. A) ProHits enables the creation of lists of samples that can be flagged for Export. This is particularly useful when preparing data for publication. The icon(s) corresponding to the exported group associated with a project will be found at the top of the Bait or Sample Report page. B) To flag a new bait or sample for inclusion in an Export group, click on the "Notes" icon, and select the appropriate export version. After refreshing the page, the new icon will appear in the Status column. C) To export all hits associated with an Export group, under the "Other tools" options in the main page of Analyst, select Export Functions> Export non-filtered Bait-Hits Report. Select desired Export version (here: Kinome paper 2009, VS1). Only the baits specifically tagged with [VS1] will be displayed. Select all (if desired), and generate report. D) The Export Bait Report page allows you to select the parameters to be included in the Report. By default, the report will be generated as a *.csv file. E) Additional Export versions can also be created by selecting "Manage Protocols and Lists">Group Lists>Export Versions.

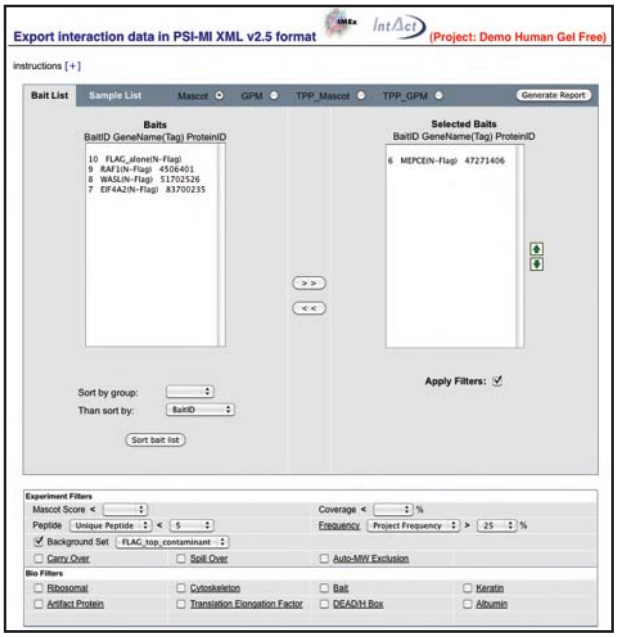


Supplementary Figure 20. Managing Protocols and Lists. A) In the Analyst Module menu, the user finds a series of tools to assist in managing protocols and lists. B-E) Entry pages of the various editors: B) Text-based Protocols; C) Background Lists; D) Controlled Vocabularies of Experimental Details; E) Groups (Bait, Experiment or Sample; this also includes another type of grouping, called "Export Versions", detailed in Supplementary Fig. 16).

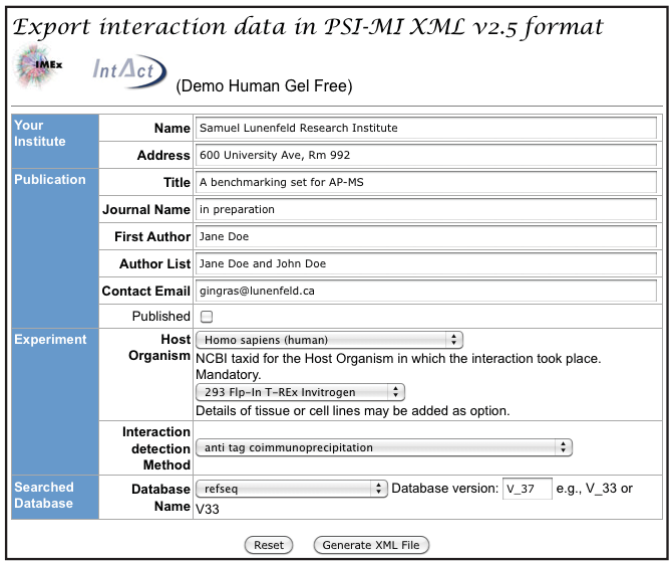
A



B

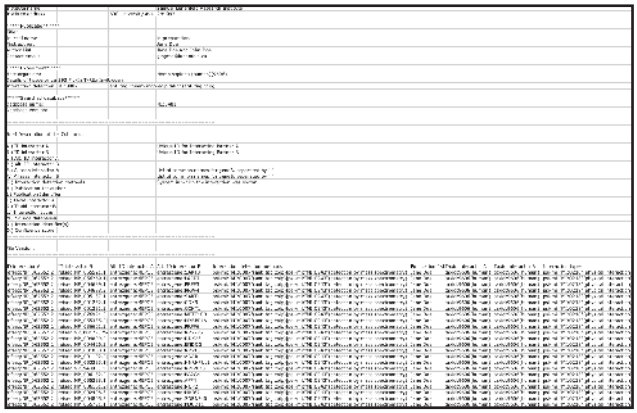


C



D

E



Supplementary Figure 21. Preparing interaction reports for submission to interaction databases. A) ProHits tracks all controlled vocabularies (CVs) necessary for the submission of interaction data to standards-compliant databases and assists the experimentalist in the preparation of PSI-MI 2.5 interaction data reports; both the XML format, preferred by the IMEx consortium, (including IntAct), and the MITAB format used by BioGRID are supported. B) To begin the preparation of the reports, select the baits or samples to be included and apply the desired set of filters. C) When you press the [Generate Report] button, a new window will pop up that will prompt you to add missing information for generation of the PSI-MI XML or MITAB files. D) XML file generated from ProHits E) MITAB file generated from ProHits; both types of files can be downloaded onto your desktop and are ready to be submitted to interaction databases.

Welcome to ProHits

ProHits is an open source software package (distributed under an Apache 2.0 license) designed to help scientists store, search and analyze mass spectrometry data. Although the platform is flexible and easily amenable to different types of projects, the system is designed to maximize the biological information from high-throughput protein-protein interaction experiments. The platform provides secure storage of mass spectrometry data, integration with search engines and mass spectrometry analytical tools (including Mascot, and the open source X!Tandem and TPP pipeline), and web-based queries of the results. An analysis module allows easy visualization of data, comparison of multiple experiments, and permits export to third-party software. A Cytoscape link provides added functionality and allows exploration of the results in a network view, with quantitative mass spectrometric information encoded and visualized as an edge attribute. ProHits also tracks all information required for deposition of the data into standards compliant public repositories, as well as for submission of manuscripts to standards compliant journals, and facilitate data export.

The software is modular and can be readily adapted by programmers to customize data flows. Written in PHP, the main database can be installed on MySQL. Users within a site easily access their results through a web interface, without the need to install additional software. The project is still under active development, and additional modules and functionality will be integrated in the core database. Programmers at other institutions are encouraged to develop tools and adds-in to ProHits (more details about this will be posted on our website, www.ProHitsMS.com).

ProHits distribution

The source code is available at www.ProHitsMS.com; demo data is included in the downloadable version. Note that an active public site hosted at the Samuel Lunenfeld Research Institute also contains the same demo data, and is available for navigation via a web interface (without the need for software installation). Because of firewall issues, however, the pages displaying the search results (Mascot, X!Tandem and TPP) are not available from outside of our Institute.

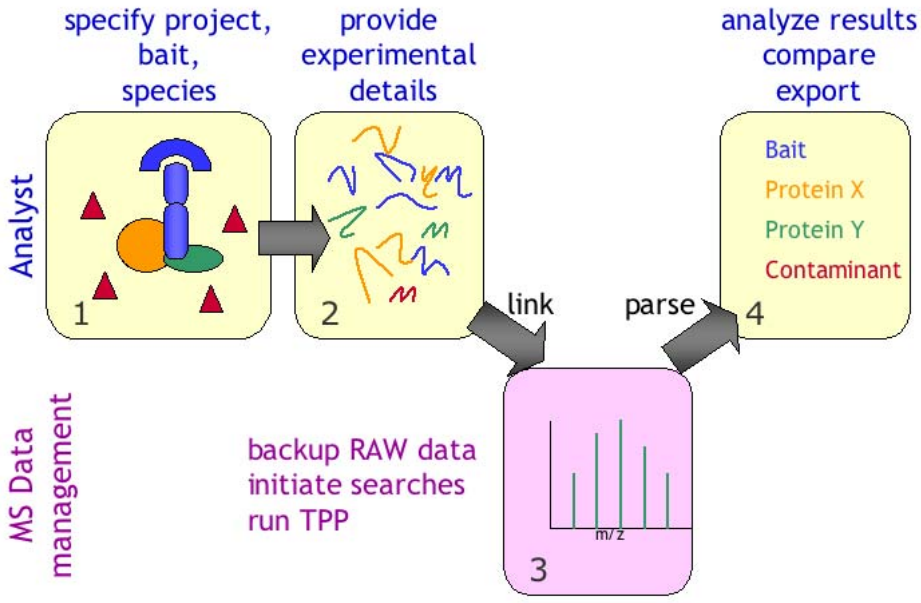
Documentation

1) Instruction manual for users
    a. ProHits MS Data Management
    b. ProHits Analyst
2) Installation instructions for database managers
3) Short video tutorials (guided tours)
4) Online help file (accessed through clicking on or [help] on the website).

At a glance:

ProHits was designed to help manage protein-protein interaction projects. Projects are created in the "Admin Office" module, and several levels of security exist for users and projects. Data is entered in each stage of the projects in a Bait-dependent manner.

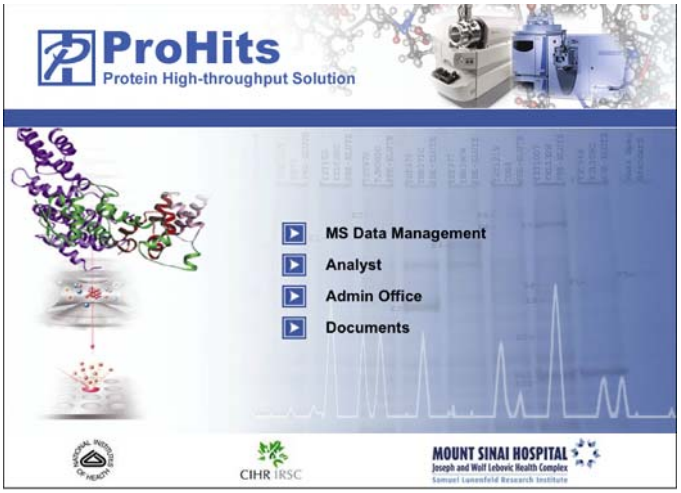
In a typical affinity purification coupled to mass spectrometry (AP/MS) experiment (shown below), the bait protein of interest is epitope-tagged, expressed in a relevant cell, and the bait and associated proteins are recovered. Gel-free or gel-based samples are then digested, and the peptides are submitted to liquid chromatography coupled to tandem mass spectrometry. CID spectra are searched using database matching algorithms and statistical analysis software, and the list of interactors is returned. Major challenges in the analysis of protein-protein interaction data include the difficulty of comparing many experiments simultaneously and the presence of background contaminants within the list of identified proteins. Prohits allows the user to compare multiple AP/MS analyses and to identify "frequent flyers" or other contaminants. Note that the same logic also applies for other types of bait purifications, providing that appropriate controls are generated.



ProHits handles the information in two linked databases: in the first database, the Analyst, information regarding the identity of the bait as well as experimental details (e.g. tissue or cell line, AP conditions, MS protocols, band position on the gel) is entered. This module interfaces with the MS Data Management module, where the raw data are automatically backed up, analyzed with user-specified database search engines including Mascot or X!Tandem, and further scored via the TransProteomics pipeline (TPP). Every raw MS datafile is linked to a bait and experimental details defined in the Analyst module, and the results of the database searches are transferred (parsed) into the Analyst database. As shown in detail in this instruction manual, once the data are searched and parsed, the scientist can visualize the data for each of the files, and export this data to Excel files or Cytoscape. The Analyst module also allows for the comparisons of hundreds of samples simultaneously. Several filters allow for removal of proteins that are identified in negative control runs, that are identified at a high frequency, that have lower scores, etc., enabling efficient biological inference from the data.

Instruction manuals for users
These manuals are designed to help end-users navigate a pre-installed ProHits database. For installation, refer to "installation instructions for database managers".

ProHits modules overview
Access your local ProHits web database (the address will be specified by your database manageer). Alternatively, access the demo version hosted by the Samuel Lunenfeld Research Institute at www.ProHitsMS.com and register as a user (registration is free). When you access ProHits, you will see the following screen that provides you links to the different sections of the ProHits database.



These sections are briefly described below:

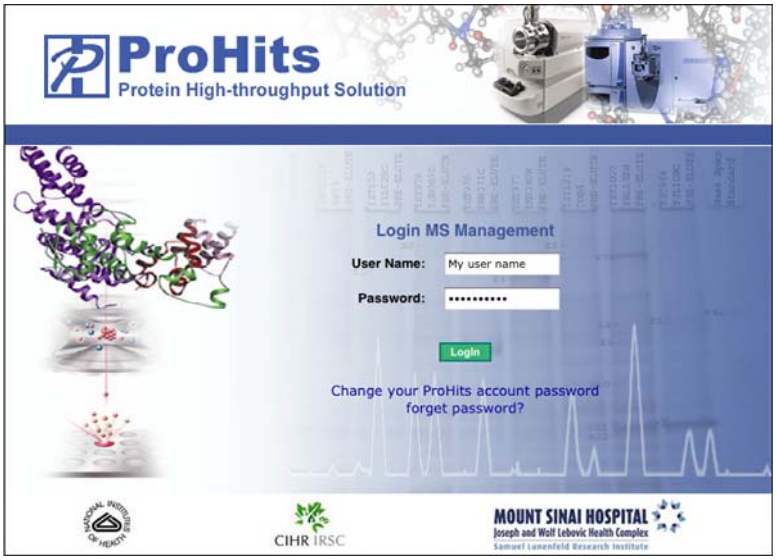
MS Data Management allows you to store your RAW mass spectrometry data from multiple instruments and to initiate database searches using either the commercial Mascot search engine (license from Matrix Science version 2.2 is necessary) or the free Open Source search engine X!Tandem. Search results can be further analyzed using the TransProteomic Pipeline (TPP, an Open Source software suite, version 3.4) and viewed directly within the MS Data Management module. Alternatively, search engine (or TPP) results can be transferred (parsed) into a bait-centric relational database, the Analyst module.

Analyst is a relational database that allows users to store, annotate, filter, compare and export protein and peptide identification results generated by search engines and/or the TPP in the MS Data Management module. The system is optimized for use in interaction proteomic analyses, in which a given protein (bait) is recovered in association with its binding partners (hits). Analyst supports both gel-based proteomic experiments and gel-free experiments.

Admin Office allows mass spectrometry specialists and database administrators to manage instrument backups, protein databases, users, projects and filters.

Documents contains the latest versions of the user manuals and other relevant information (this section is not password protected).

Accessing ProHits: To access ProHits functions, the database administrator first configures a profile for each user and provides them with the required level of privileges (refer to installation and guide for database managers). The user then enters his/her information when prompted.



ProHits MS Data Management

User manual - demo

Version demo 1, prepared on Feb 09, 2010 by Anne-Claude Gingras, with help from Brian Raught, Wade Dunham and Brett Larsen (earlier versions prepared with input from Frank Liu, JP Zhang, Brian Raught, Brett Larsen, Karen Colwill, Zhen Lin and Lisa D'Ambrosio).

Contents

Using the "MS Data Management" module

Once you have selected the MS Data Management module from the ProHits access page and have logged in, the following screen will appear.

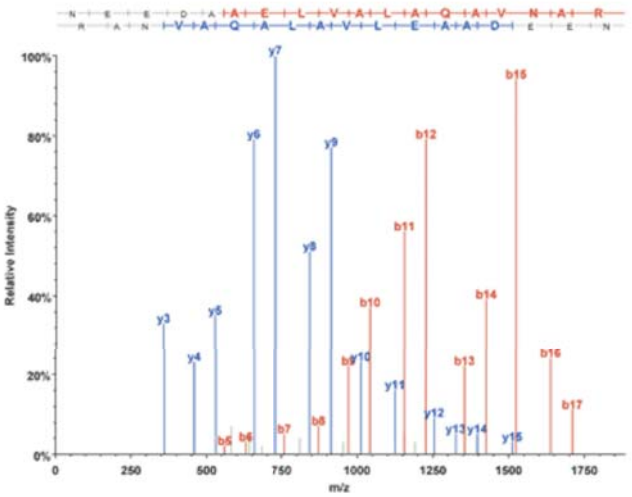
At the top of the page (left), you will find the tabs "Home", "Storage" and "Auto Search" which will allow you to navigate between the storage and the search areas of the MS Data Management Module. On the right is a link to the Analyst module which can be accessed from every page of the MS Data Management module.

Overview

Welcome to ProHits

ProHits is an open source software tool designed to help scientists manage, search and analyze mass spectrometry data.

MS Data Management allows you to store raw mass spectrometry data from multiple instruments, and to initiate database searches using the commercial Mascot search engine (licence from Matrix Science is necessary) and/or the free Open Source search engine X!Tandem. Search results can be further analyzed using the TransProteomic Pipeline (TPP, an Open Source software suite), and viewed directly within the MS Data Management module. Alternatively, search engine (and/or TPP) results can be transferred (parsed) into a bait-centric relational database, the **Analyst** module.



The **Storage** section allows you to monitor the transfer of data from each of the acquisition computers to the ProHits backup system. It also allows you to search, browse and download files, convert RAW files to other formats, and manually upload RAW data.

The **Auto Search** section allows you to perform database searching on specified files using user-defined search engines and parameters, to explore the results, and to transfer search results to the Analyst module. It also allows for database searches to be pre-scheduled for data files that will be acquired at a later time.

The "Storage" section allows you to monitor the transfer of the data from each of the acquisition computers to the ProHits backup system. It also allows you to search, browse and download files, convert RAW files to other formats, and manually upload raw data.

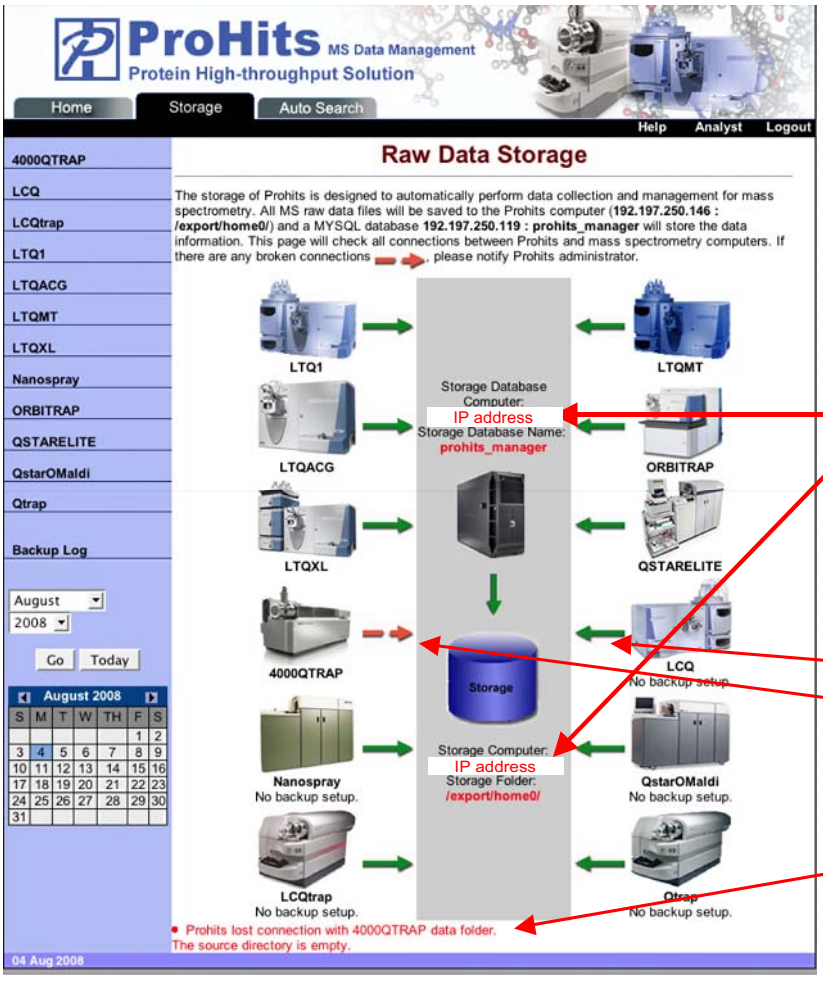
The "Auto Search" section allows you to schedule and perform database search tasks on specified files using user-defined search engines and parameters, in addition to results exploration and linking to experimental information in the Analyst module.

Storage

⇒ Select the "Storage" tab

ProHits manages the backup and storage of data files in an instrument-dependent manner. The left of this screen provides links to all available instruments in the mass spectrometry facility (also shown as pictures). Here, you can also view the Backup log, and the location of the database and data storage. The central part of the page details the status of each of the connections. A green arrow indicates a functional connection while a broken red arrow denotes a non-functional connection.

Note that the storage tab example shown below is for the Samuel Lunenfeld Research Institute facility; in the demo version only a single instrument (LTQ_demo) is listed. For the purpose of this user manual, we will toggle between the demo version and the live site at the Samuel Lunenfeld Research Institute.



This links to individual instruments backed up on Prohits in your facility

Select to monitor file transfer from your acquisition computer to Prohits and view the log for the RAW file converters and the RAW file merge functions

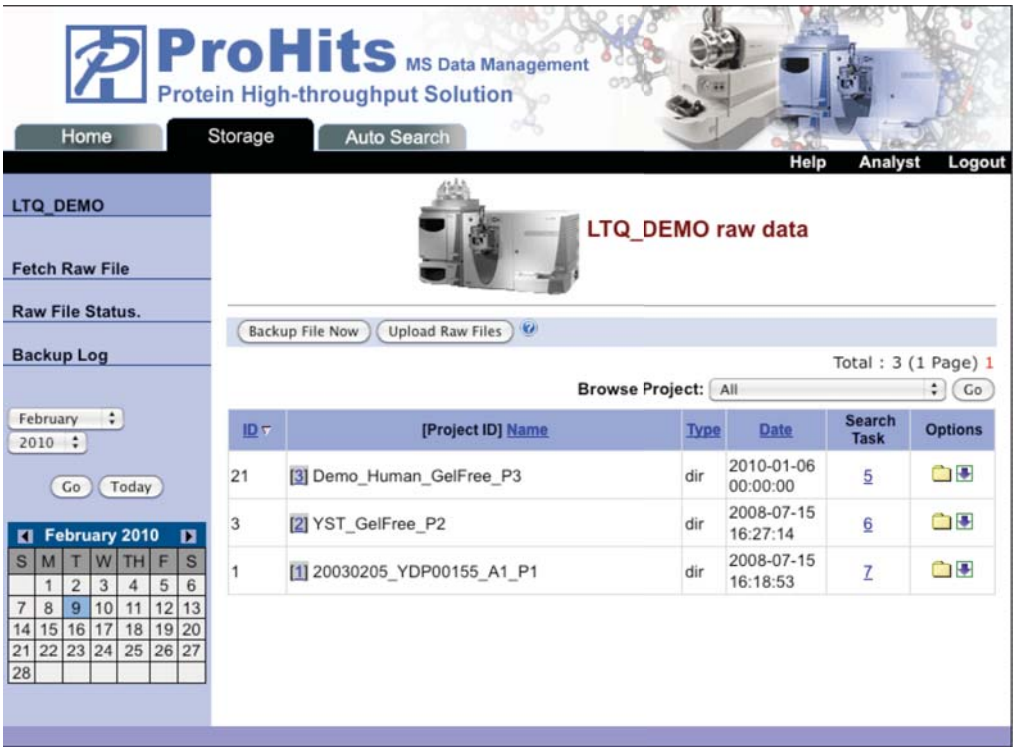
Location of storage; this is set-up by your administrator

Status notification. Successful links are shown by green arrows. When Prohits is unable to backup data from one mass spectrometer, an orange broken arrow will appear. You will also see a note appearing at the bottom of the page detailing the problem..

Selecting instruments and folders

The files backup organization mirrors the set-up on each acquisition computer (the computer linked to the MS instrument). Subfolders are allowed on the acquisition computer. Selecting the instrument will open a new page allowing you to browse files and projects. In this demo version, only one instrument (LTQ_demo) is available.

⇒ Select an instrument (LTQ_demo) by clicking on a link on the left of the page



At the bottom of the page, the folders associated with this instrument are listed (along a unique identifier assigned by ProHits). Folders are associated to individual "Analyst" projects defined in the Admin Office. Note the suffixes "_P1", "_P2" or "_P3" appended to the end of the folder names (e.g. "Demo_Human_GelFree_P3"). These suffixes will allow the creation of automatic links (autolinks) between the mass spectrometry data files and samples defined in the Analyst module, and will be discussed later.

The Search Task column allows you to navigate to the search results page(s) associated with files in this folder. In the Options column, clicking will open the folder, while selecting will download the files associated with this folder onyo your local computer.

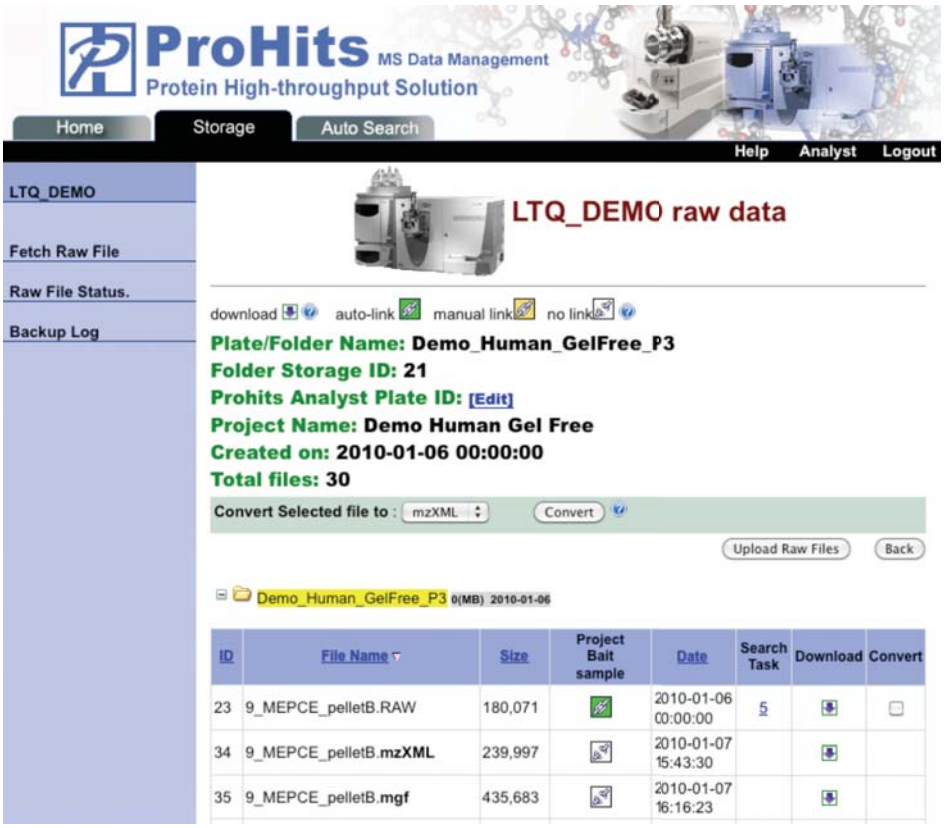
ProHits automatically backs up any new files from the acquisition computer at a time specified by the administrator.

The user may also initiate a manual back up by clicking [Backup File Now] (towards the top of the page). Please note that to prevent file corruption, ProHits has a timed delay for file transfer (delay specified by administrator – we use 2.5 hours after the last file modifications).

To upload raw files from an instrument not connected to the ProHits backup system, select [Upload Raw Files] and navigate through the options. This function can also be used to upload converted files to be searched in ProHits.

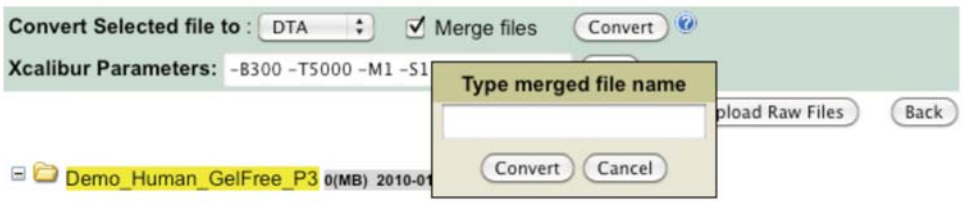
⇒ Select a folder (Demo_Human_GelFree_P3) to open

In addition to the files located on the acquisition computer, the selected folder will also contain any files that you have converted from original raw files to other formats. This folder will also contain any file manually uploaded through the link on the previous page. Here, we show the raw file 9_MEPCE_pelletB.RAW (acquired on a Thermo LTQ instrument) already converted to .mzXML (for searches using X!Tandem) and .mgf (for Mascot searches).



Data file conversion

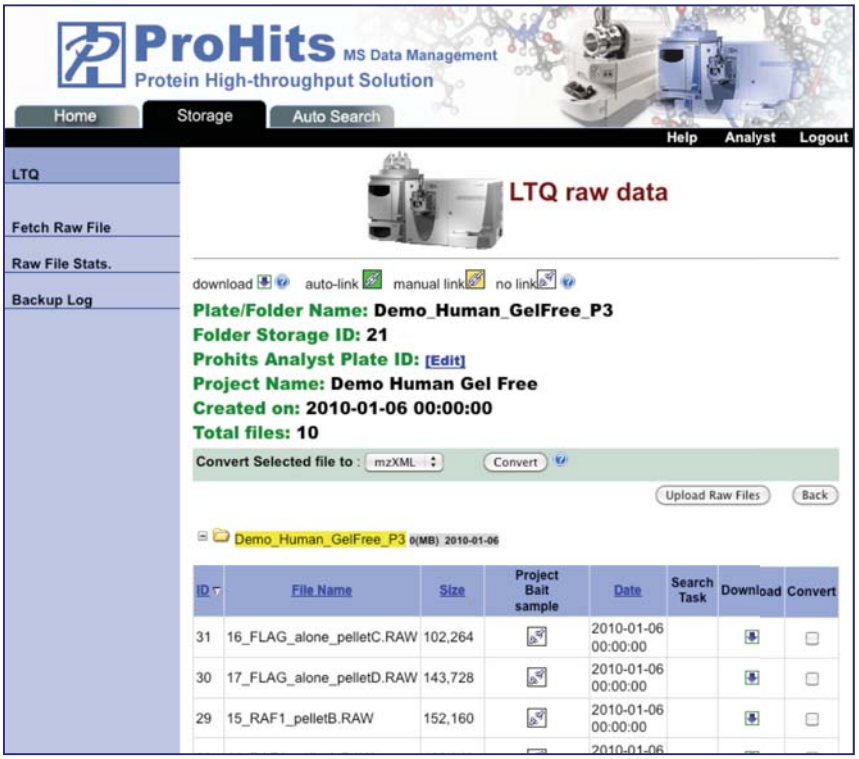
ProHits can automatically convert Thermo RAW files to the database search engine preferred file types as part of the AutoSearch pipeline. Alternatively, RAW files can easily be converted here. To convert RAW files to mgf, mzXML or dta format, simply choose the desired files by clicking the associated boxes in the "Convert" column, select the desired format (and conversion parameters if available) and hit "Convert". If you are converting to either .mgf or .dta files, you will be given the option to combine (or merge) several files. By selecting to convert to wither of these file types, the dialog box will be expanded allowing for the selection of files to be merged and for manual curation of the merged file name. This option is especially useful for combining files from fractionation of one sample (e.g. gel bands from the same lane).



Note that at the present time the publically released version of ProHits can only convert data from Thermo instruments. Please refer to http://tools.proteomecenter.org/wiki/index.php?title=Formats:mzXML, http://psidev.info/index.php?q=node/257, or http://proteowizard.sourceforge.net/ for converters for additional instruments proprietary formats.

Linking files to the Analyst module

As mentioned in the general introduction, RAW files located in the Data Management module can be linked to Baits>Experiments>Samples defined in the Analyst module and the Search Results can be transferred (parsed) from the Data Management module to the Analyst module.



The chain link icons in each table of the Data Management module indicate the linking status:

 (broken white chain link) indicates that the file is not linked to any sample in the Analyst module,

 (intact yellow chain) indicates that a link to a sample in the Analyst module has been manually created,

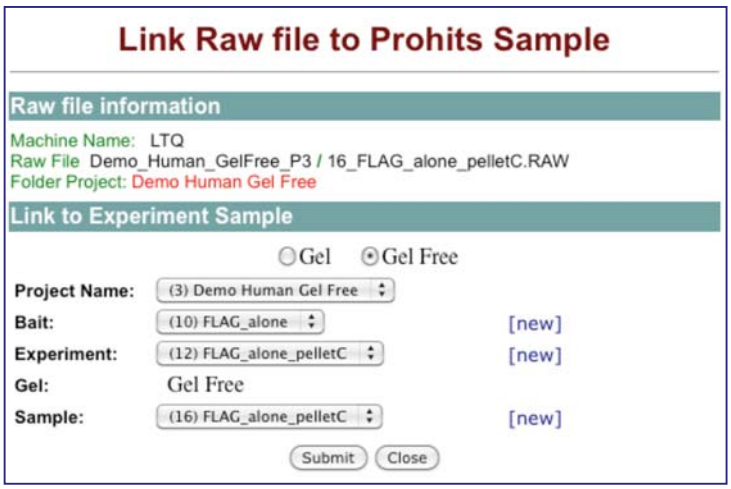
 (intact green chain) indicates that a link has been created automatically (also called "Autolink").

⇒ To create manual links, click  in the "project Bait/sample" column.

This opens up a new dialog box allowing you to select the desired Project, Bait, and Sample. Upon closing the box, the yellow chain link icon  will appear.

Creating automatic links requires interfacing with the Analyst module and using a standardized naming scheme. This will be described in a separate section.

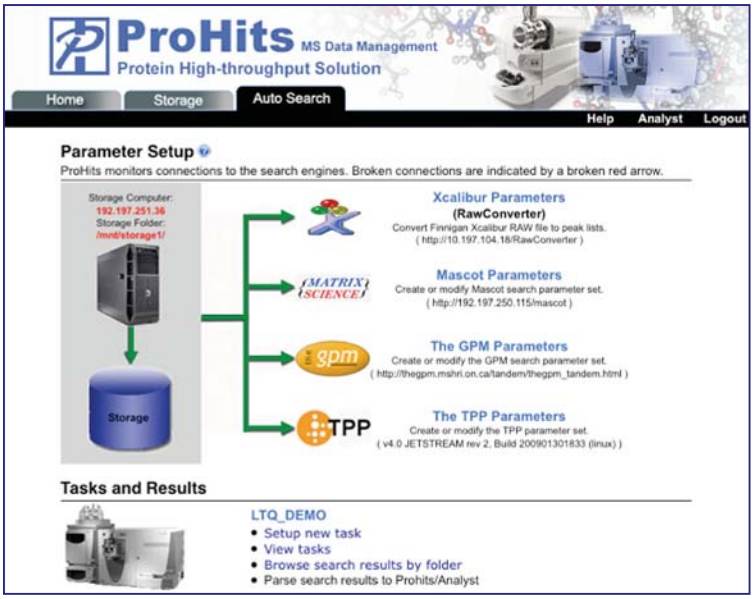
Creation of New Baits and Samples will be described in the instructions for the Analyst module module.



This completes the overview of the "Storage" part of the Data Management module of ProHits. In the next section, we will navigate through the searching and parsing functions.

Using AutoSearch for database searching

⇒ Select the "AutoSearch" tab at the top of the page



This view displays all the search engines and other tools which have been linked to your ProHits database in the top portion of the page. As with the link to different mass spectrometers in the Storage area, successful links to search engines are indicated by green arrows and broken links by a broken orange arrow.

You can modify the general search parameters from this page; you will also be able to modify search parameters when initiating searches. ProHits simply employs the standard interfaces provided by the search engines and allows you to create several standard search parameters sets.

The bottom part of the page allows you to access individual instruments for search purposes, by simply clicking on the links or on the instruments.

⇒ Select an instrument for searching (for this tutorial, we will select LTQ_Demo)

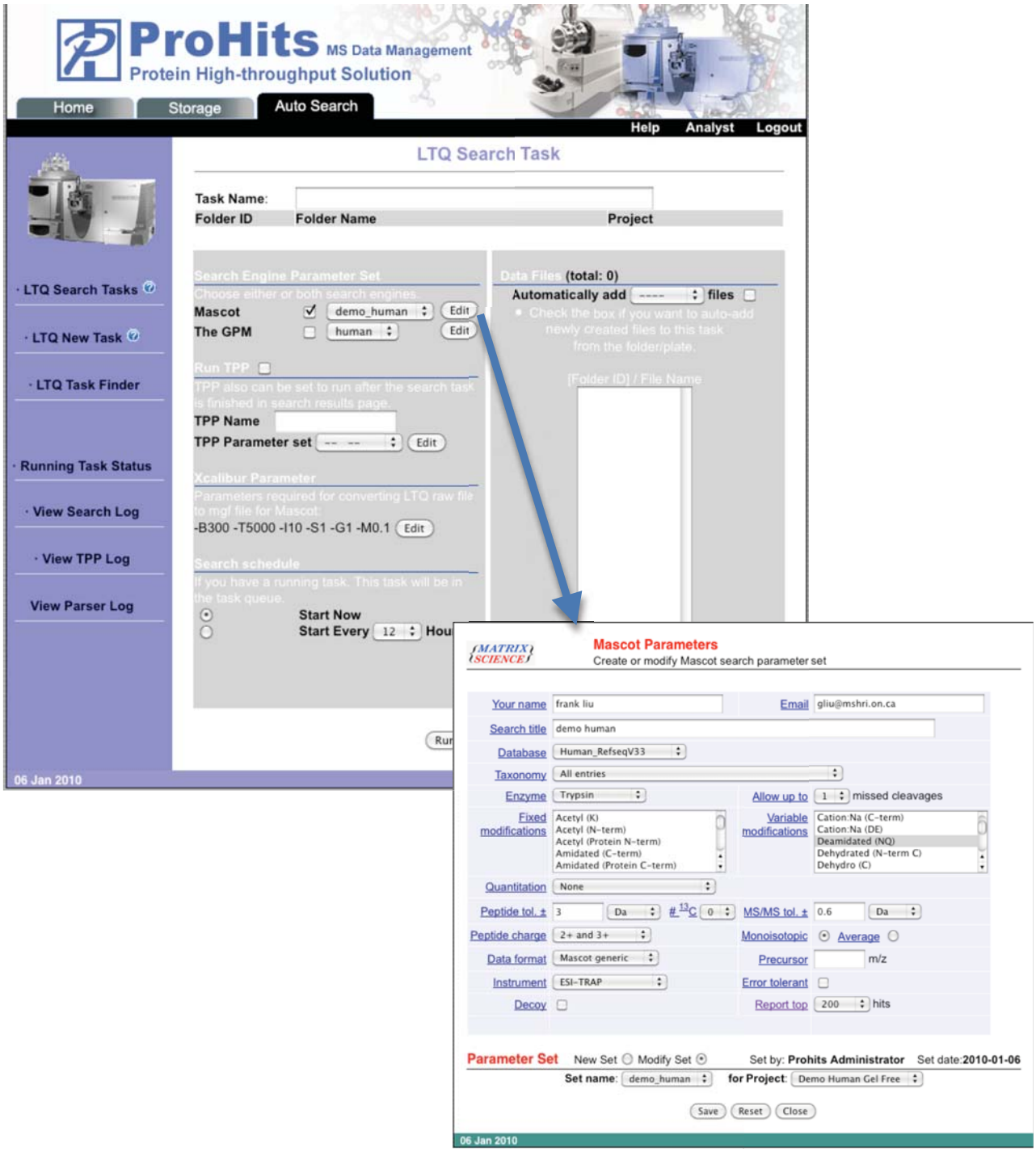
The entry page lists all of the search tasks that were performed for files collected on a given instrument. A given task may be applied to several files not necessarily located within the same subfolder or in the same format. We will return to this list later after we have created a New Task.

The table lists the current Tasks. To view the search parameters and a list of the searched files, press . To view the search results, press: (Results details).

The left menu bar allows you to add or manage Tasks, as well as to view the logs.

⇒ Select "LTQ_Demo New Task"

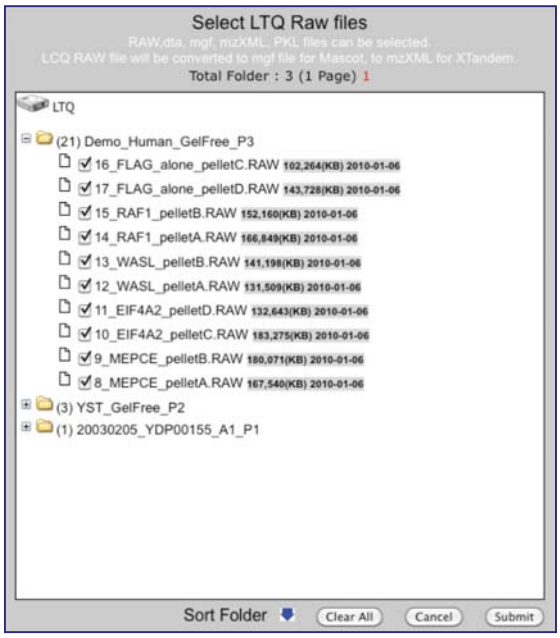
This opens up a new page. Enter a Task Name of your choice, select the search engine(s) and parameter set(s) to be utilized (these can be edited). If needed, an automatic conversion of RAW files to the mgf format used by Mascot or the mzXML format used by X!Tandem will be applied by default. The search can be initiated manually and immediately ("Start Now"), or automatically as files are added every X hours (see below). You have an option of automatically running the TPP statistical software tools on the search results. If you choose to do so, select the "Run TPP" box. Note that you can also run the TPP tools after you get the search results. (Running TPP manually post-acquisition is a more flexible option, because it allows you to combine several files into a single TPP analysis).



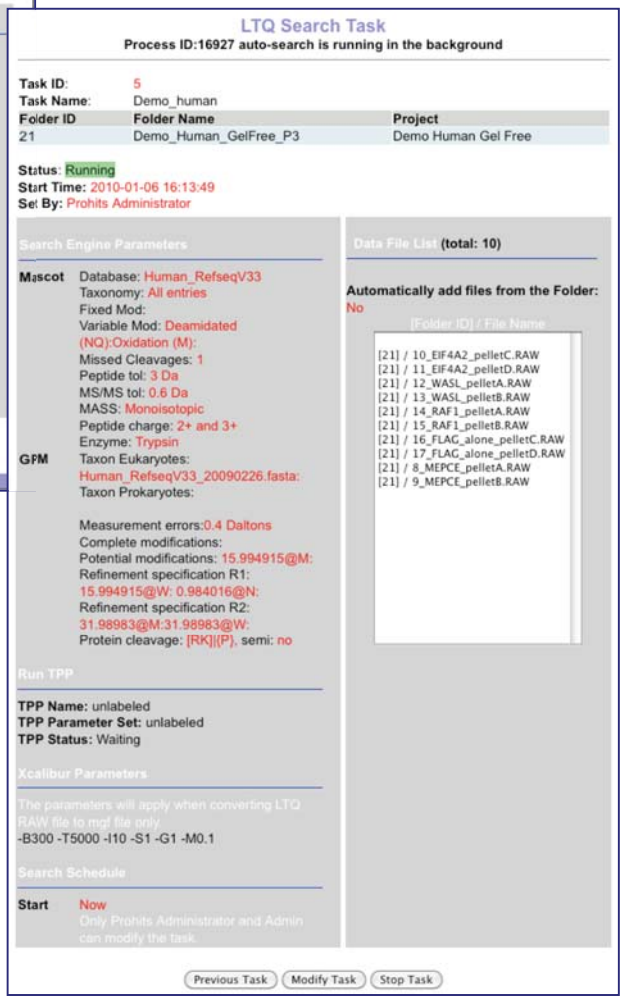
Manually initiate searches

⇒ To manually initiate a search, select the "Start Now" option, and click the "Add Files" button located on the bottom right

This will open the folders for the specified instrument, allowing you to select as many files as desired for searching.



Once all parameters are selected, and the files transferred, click the "Run Task" button at the bottom of the screen.



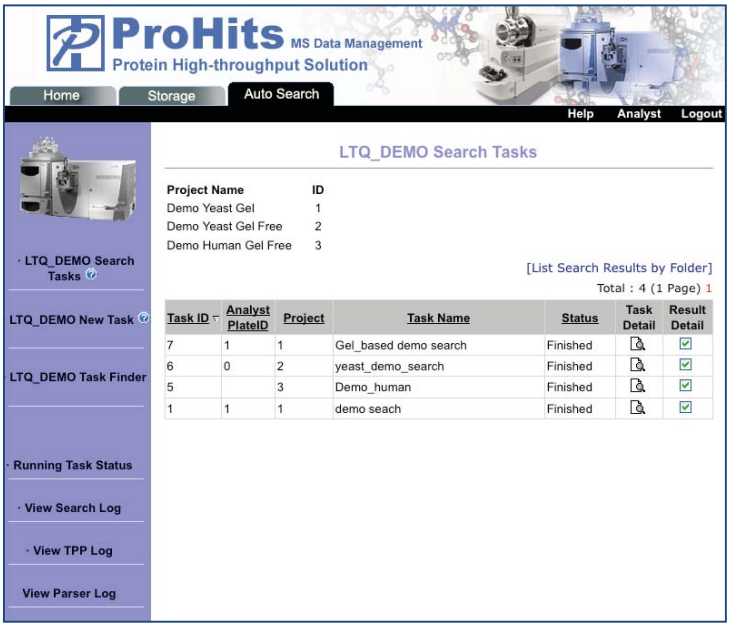
Once the task is running, the search parameters will be locked. Tip: instead of creating a new search task for each file you analyze using the same search parameters, use the "Modify Task" option at the bottom of the page: this will group all your searches within the same search task folder, making it easier to retrieve, and will also ensure that the same search parameters are used for each file. Clicking [Modify Task] allows you to add files to be searched with the same parameters, but not to change the search parameters. To research the same raw files with different search parameters, a New Task must be created.

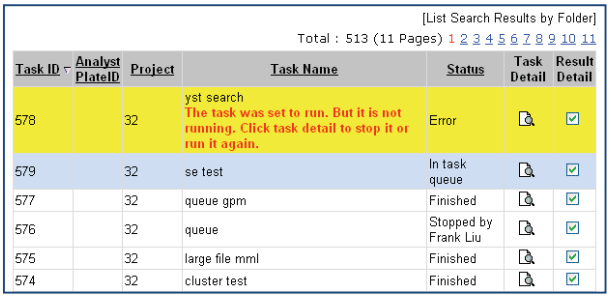
Automatically initiate searches

For some high-throughput projects, users may want to **automatically initiate searches** on every new file added to the folder(s) that are initially selected for the search. Before you can select to automatically add files and search them, you have to point to at least one file in a folder manually, as indicated above. By selecting the "Start every X hours" (left hand side), and "Automatically Add Files" (right side) options, every file of the selected format subsequently acquired within the same folder(s) or subfolder(s) will be automatically searched using the same parameters.

Search Task view

⇒ Select the LTQ_DEMO "Search Tasks" option on the left side
This opens the page listing the searches performed on this instrument. Searches still running will be highlighted in green.





Other important status information can also be obtained in Search Tasks view. While green indicates a file being actively searched, a blue colour indicates that a sample is currently in the queue. Yellow highlights an error with the search. No other searches can be initiated until this problem is resolved, either through successful running of the problematic search, or by stopping the task.
IMPORTANT: If an error is encountered while tasks are in the queue, one of these tasks must be manually re-started (the other Tasks will then be searched in queue).

⇒ Clicking on the Task Detail icon in the column will open up the same status page as shown on page 9.

⇒ To obtain the results, click on the "Result Detail" icon at the extreme right of the table.

View Search Results

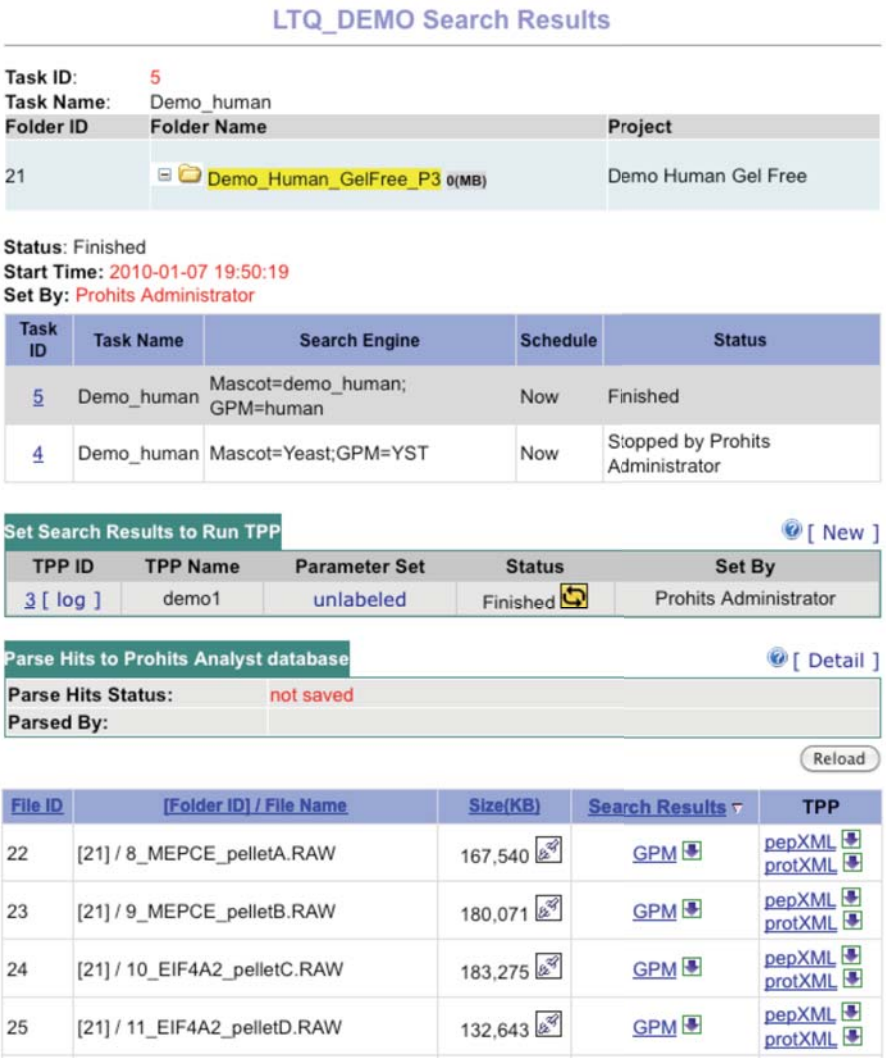
⇒ Select the "Result Detail" for one search task

This opens up a new page.

The search results are displayed at the bottom of the page. The blue link will connect to the search engine page.

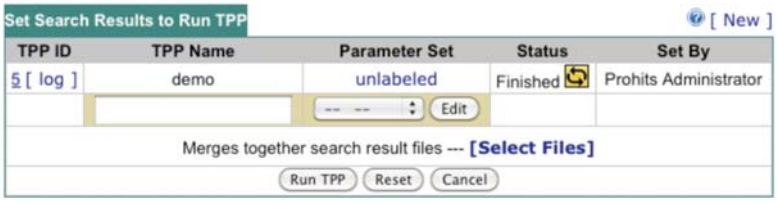
Note that if using the Demo sites from external computers, you will not be able to view the results from the search engines (firewall protection).

This page also provides you with
1) list of the other Task IDs associated with this folder
2) the option of analyzing your search results using the TPP (either for single files or for merged files)
3) the possibility to link files to the Analyst module
4) the file parsing tool that allow you to transfer your search results to the Analyst module.



Running the TPP after completion of searches and merging samples prior to TPP analysis

To add new results files to be analyzed with the TPP to an existing TPP task, click the  icon in the "Status" column of the box [Set Search Results to Run TPP]. You can simply click on the boxes that will appear in the TPP column of the results tables. When desired files are selected, press [Run TPP]. To create a completely new TPP task (e.g. if the TPP parameters have changed), click [New], select desired parameter set and desired files and press [Run TPP].



There are cases where you may want to merge files prior to running the TPP (examples include fractionated samples, gel-based or otherwise). This is a simple process in ProHits. The two requirements are that the searches be performed using the same search engine (Mascot and X!Tandem files cannot be combined) and that the results are located within the same Search Task folder. Select the "Merge" option and the files to be combined.

Selecting the "Merge" option will create a new entry at the bottom of the page. The TPP can be run on this entry in the same fashion as on individual files, by selecting the desired parameters and pressing "Run TPP".

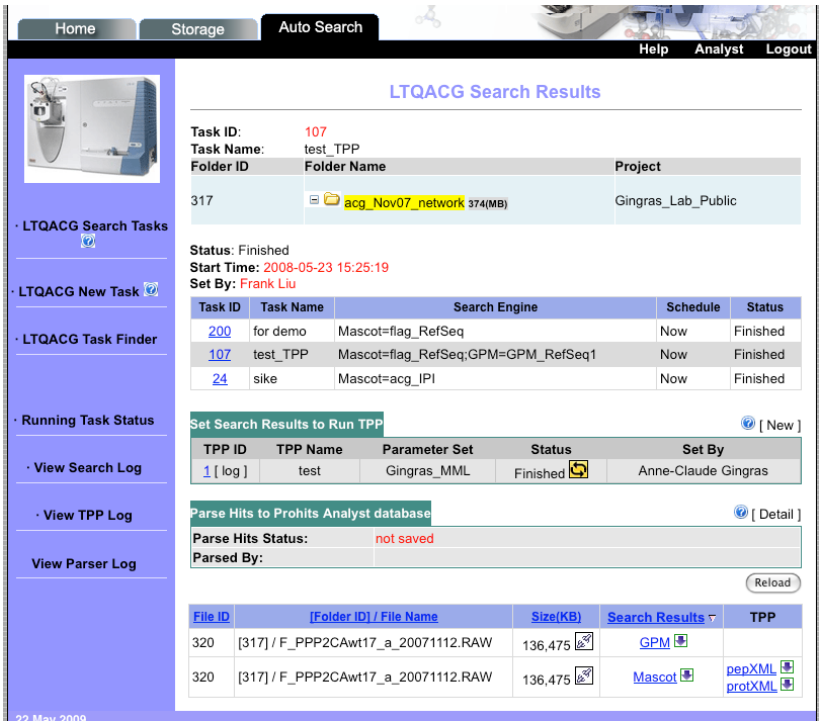
At this point, we have performed database searching and have obtained results directly from the search engines, as well as results from the TPP pipeline. We are now ready to transfer these search results into the "Analyst" module.

The transfer involves two steps: 1) linking the initial file to an entry created in the Analyst module; and 2) parsing the search results.

Linking files to Analyst

⇒ link the file

You may have already linked the native file (not searched) to the Analyst modules through the "Storage" area. If so, the "link" icon by the file size will be coloured.



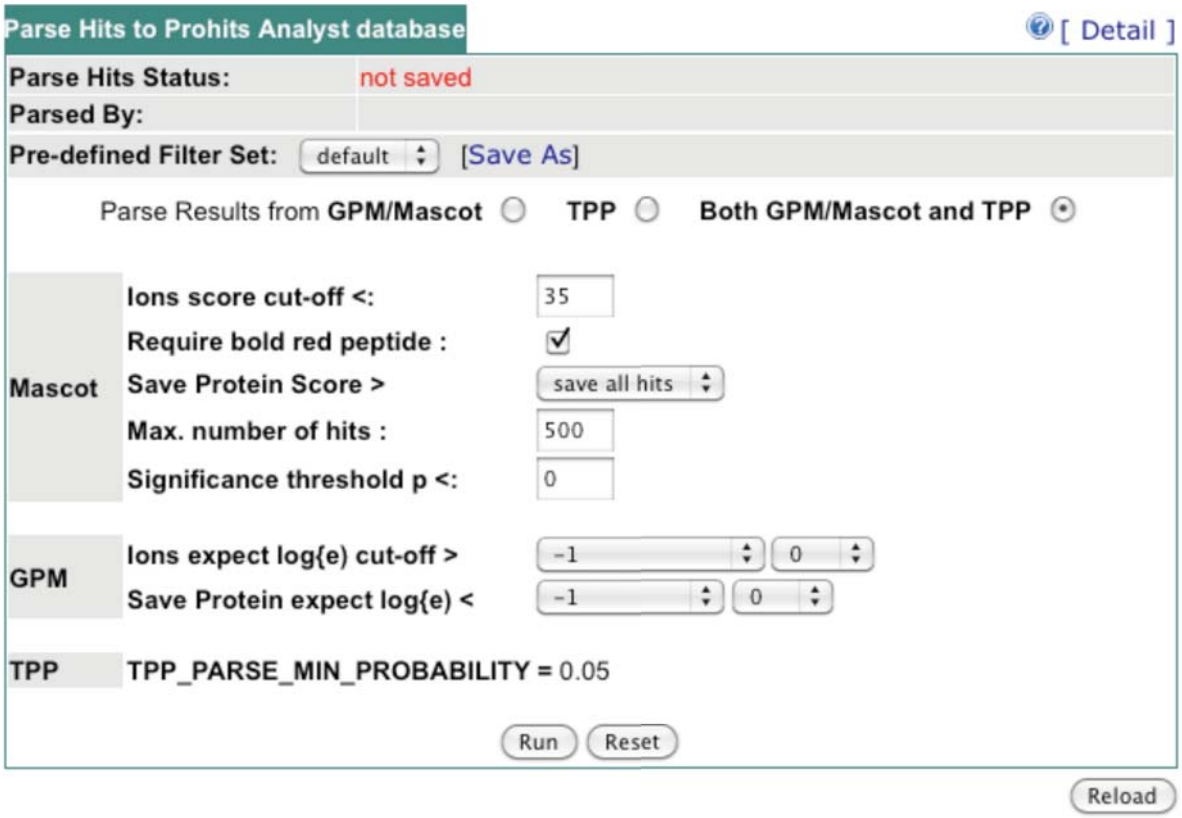
A white link  indicates that no link has been established at this point, as in the example above.

See page 6 for details about manual and automatic link creation.

Parsing files to the Analyst module

You are now ready for parsing (transferring results to the Analyst module), as you have a linked file () and search results. In this case you also have TPP results. Both types of results can be parsed.

⇒ Select the "Detail" link in the parsing area



File ID	[Folder ID] / File Name	Size(KB)	Search Results select all	TPP select all
22	[21] / 8_MEPCE_pelletA.RAW	167,540	GPM	pepXML protXML
23	[21] / 9_MEPCE_pelletB.RAW	180,071	GPM	pepXML protXML

You can transfer (parse) results from the search engines (GPM/Mascot), from the TPP or both. You can select the parameters for the parsing cut-off for Mascot and X!Tandem/GPM search engines. For the TPP, all hits with a probability greater than the cut-off selected by the administrator in the Prohits configuration file (we are using P > 0.05) are automatically parsed. Check the [select all] box at the top of the appropriate column or manually select files to be transferred to the Analyst module.

⇒ Press "Run" to initiate parsing

The "processing" status notification will appear.

When parsing is successfully completed, a green check mark will appear in the results column.

In the event that you decide to link or parse different file(s) to the same Analyst entry, you can remove the parsed files or the link: Alongside the tick box, the icon can be selected if you want to remove the hits from the Analyst module. Note that you can also unlink a sample by clicking on the or and selecting "remove link".

Other options

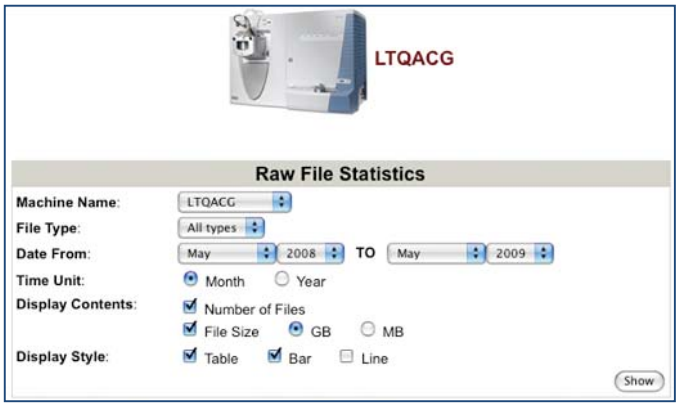
You have now completed the basic tour of the Data Management module. The following few pages will explore a few other options within the Data Management module: 1) Creating activity reports; 2) Searching for ("fetching") files; 3) Basic troubleshooting of the Data Management module.

The first few options are accessed from the "Storage" page:

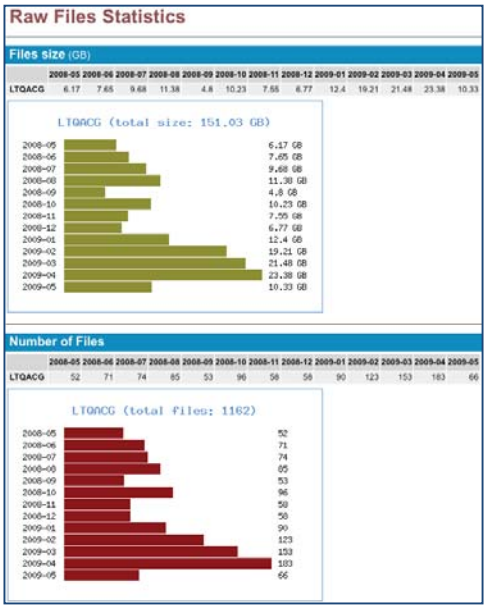
Creating activity reports for the RAW files

⇒ From the storage tab, select the "Raw File Status" option
This opens up a new window (here we are simply showing one of the SLRI instruments as an example).

⇒ Select the instrument you want the report on, the dates of the report and time units, as well as the contents and style of the display
You could also choose to get the report for all instruments linked to ProHits.



⇒ Press "show" to visualize selected display

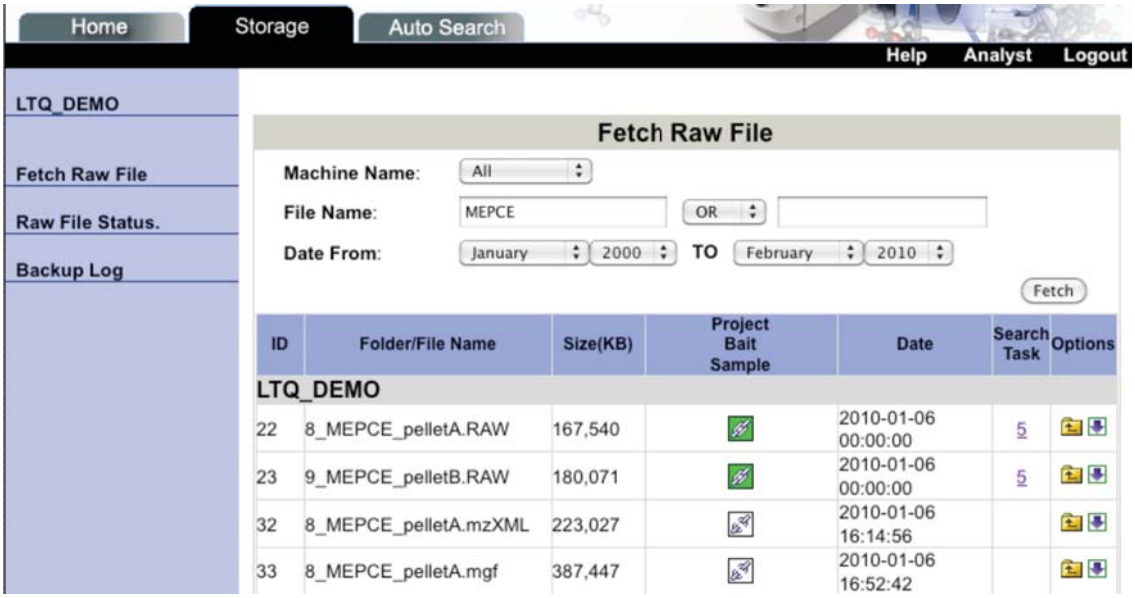


Searching for files and retrieving RAW data

⇒ Select "Fetch Raw File"

This will open a new window. Select the desired instrument (or all instruments), dates if applicable, as well as keywords that are part of the file name (note that you can perform logical operations). Press [Fetch] to retrieve results.

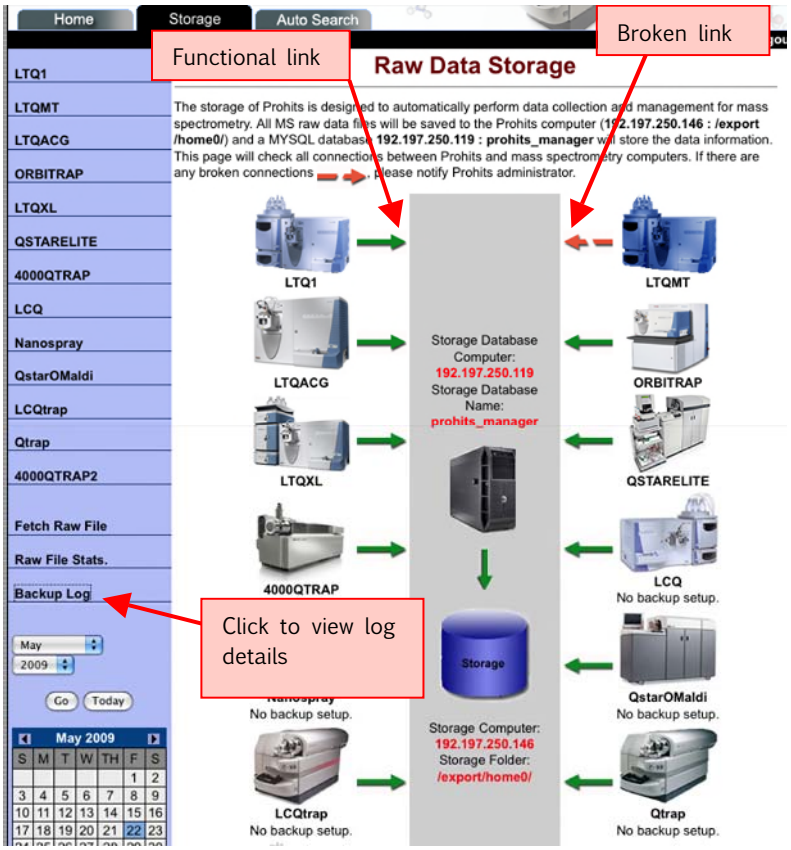
The "Options" in the last column allow you to download ( ) the data (.RAW , .mzXML or .mgf) onto your computer, and to open the parent folder ( ).



⇒ To retrieve the searched results, you can click the search task number in the "Search Task" column.

Basic Troubleshooting of the Data Management module

ProHits Data Management requires a connection between all of your acquisition computers, the search engine computers(s) and the storage computer. To facilitate the detection of broken links between these computers, ProHits has implemented an easy visual guide, both in the storage and auto search modules.



⇒ From the Data Management entry page, select the "Storage" tab

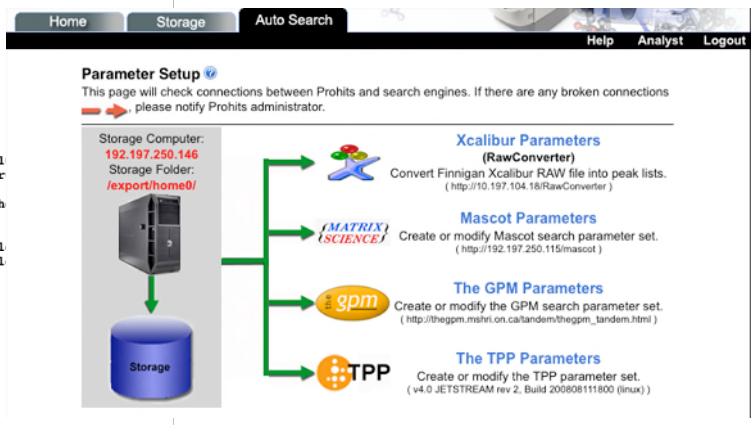
As before, all of the instruments in the facility are listed. The green arrows indicate that the connection between each instrument and the storage area is functional. If an automated backup has been selected, this also indicates that the backup was performed on schedule. Note that there is a broken connection between one of the instruments and the storage computer, easily identified by a broken orange arrow (in this case, the computer was offline for maintenance). Notify the ProHits administrator when you detect such arrows.

⇒ Click on the "Backup" log (left side of the screen, toward bottom) to monitor the transfer of data from the acquisition computers to the storage computers and to read any error messages

Log File 'raw_back.log'

Display last : 100 lines submit Refresh

```
10089   end of QSTARELITE 2009-05-22 8:04:27
10088   copied 6: /mnt/QSTARELITE/1_AJ/2009_05_20_Grb2_iTRAQ_op/Data
10087   process QSTARELITE 2009-05-22 8:03:36
10086   Backup QSTARELITE start at 2009-05-22 8:03:36
10085   PSID:3771 QSTARELITE
10084   end of 4000QTRAP2 2009-05-21 23:07:01
10083   copied 2: /mnt/4000QTRAP2/Andrew_J/2009_05_20/Data
10082   copied 1: /mnt/4000QTRAP2/Andrew_J/2009_05_20/Batch
10081   process 4000QTRAP2 2009-05-21 23:07:00
10080   end of 4000QTRAP 2009-05-21 23:07:00
10079   copy 6: /mnt/4000QTRAP/CJ_Yong/2009_05_07_0430ShcA/Data/0430ShcA_R_21
10078   copied 20: /mnt/4000QTRAP/CJ_Yong/2009_05_07_0430ShcA/Data/0430ShcA_r
10077   copied 1: /mnt/4000QTRAP/CJ_Yong/2009_05_07_0430ShcA/Batch
10076   copied 1: /mnt/4000QTRAP/CJ_Yong/2009_05_07_0430ShcA/Acquisition Meth
10075   process 4000QTRAP 2009-05-21 23:06:51
10074   end of QSTARELITE 2009-05-21 23:06:51
10073   copy 0: /mnt/QSTARELITE/Tempo Method Maker/Project Information/eventl
10072   copy 1: /mnt/QSTARELITE/Tempo Method Maker/Project Information/eventl
10071   copy 1: /mnt/QSTARELITE/Tempo Method Maker/BioAnalyst
10070   copied 10: /mnt/QSTARELITE/1_AJ/2009_05_20_Grb2_iTRAQ_op/Data
10069   copied 2: /mnt/QSTARELITE/1_AJ/2009_05_20_Grb2_iTRAQ_op/Batch
10068   process QSTARELITE 2009-05-21 23:05:15
10067   end of LTQXL 2009-05-21 23:05:15
10066   copied 1: /mnt/LTQXL/Yasmina_P36
10065   process LTQXL 2009-05-21 23:05:05
10064   end of ORBITRAP 2009-05-21 23:05:05
10063   copied 10: /mnt/ORBITRAP/Fall08dev
10062   copied 1: /mnt/ORBITRAP/Ek Install Data
10061   process ORBITRAP 2009-05-21 23:04:27
10060   end of LTQACG 2009-05-21 23:04:27
10059   copied 3: /mnt/LTQACG/MM_IP_LCMS
10058   process LTQACG 2009-05-21 23:04:01
10057   end of LTQMT 2009-05-21 23:04:01
10056   process LTQMT 2009-05-21 23:04:01
```



The same visual display and log details are also found in the AutoSearch module.

ProHits Analyst

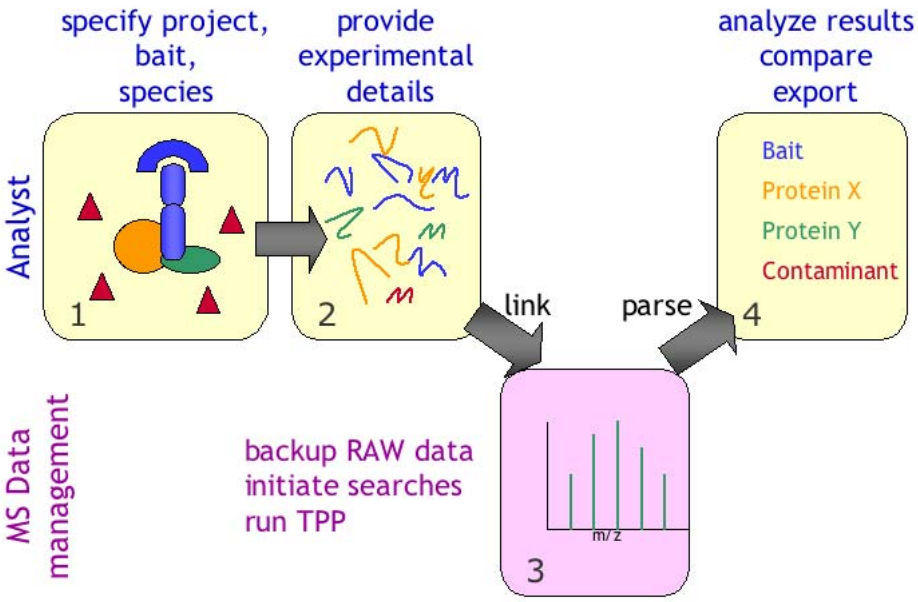
User manual - demo

Version demo1, prepared on Jan 15, 2010, by Anne-Claude Gingras, with the help of Frank Liu, JP Zhang, Brian Raught, Brett Larsen, Wade Dunham, Marilyn Goudreault and Karen Colwill.

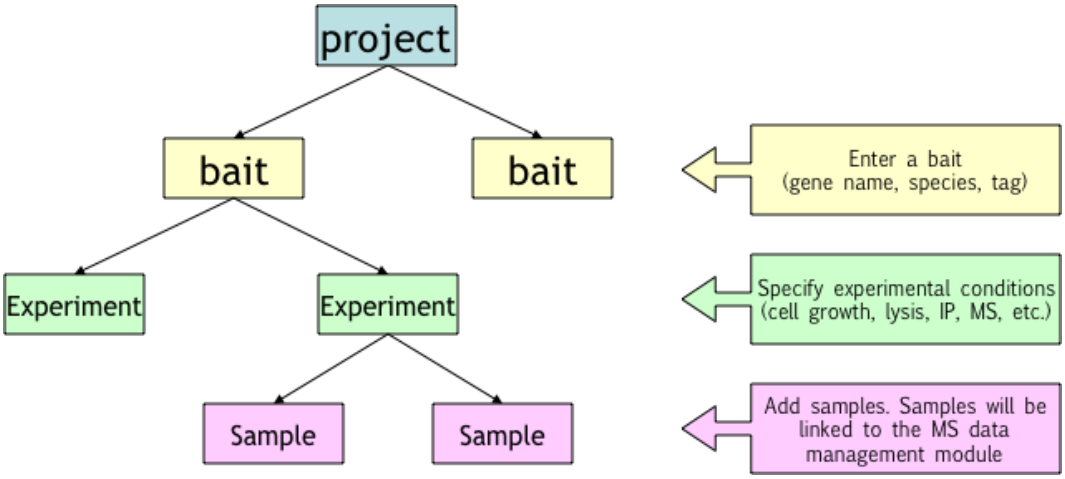
Contents

Overview

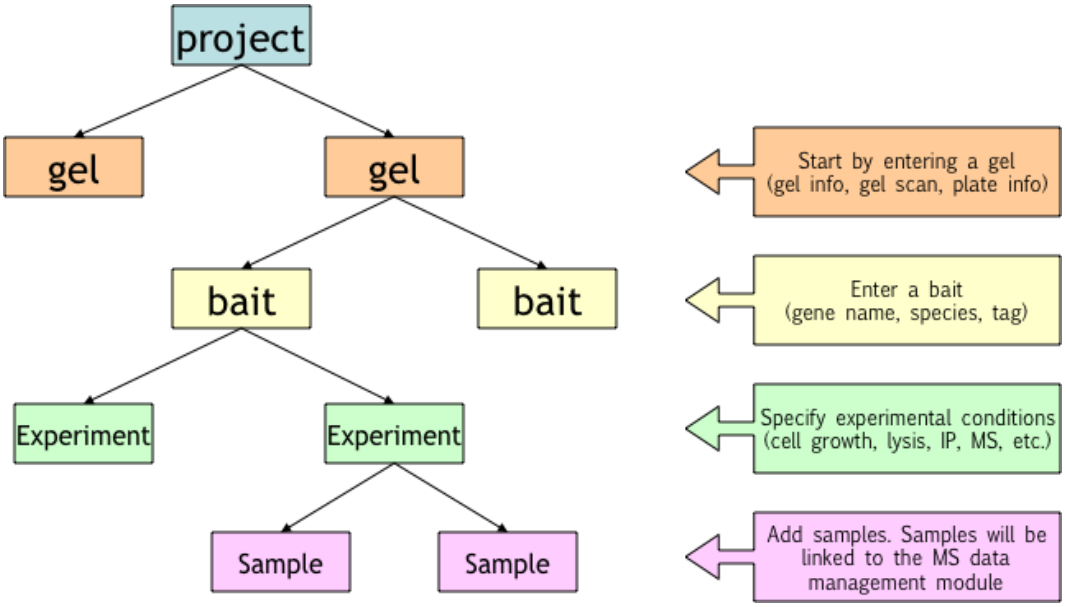
The Analyst module allows you to visualise, analyze, compare, search and export your MS results.



In order to analyze and compare data, each MS file in the MS data management module must be linked to a sample created in the Analyst module. For example, to create a sample for a gel-free experiment, you must first specify a project, create an entry for the protein of interest (bait), and define experimental conditions. Typical gel-free samples are eluates from an affinity purification.



Sample entry for gel-based projects is similar, with the exception that a gel is specified prior to the selection of a bait. Typically, samples are gel bands, and all bands from the same lane are entered under the same "Experiment".

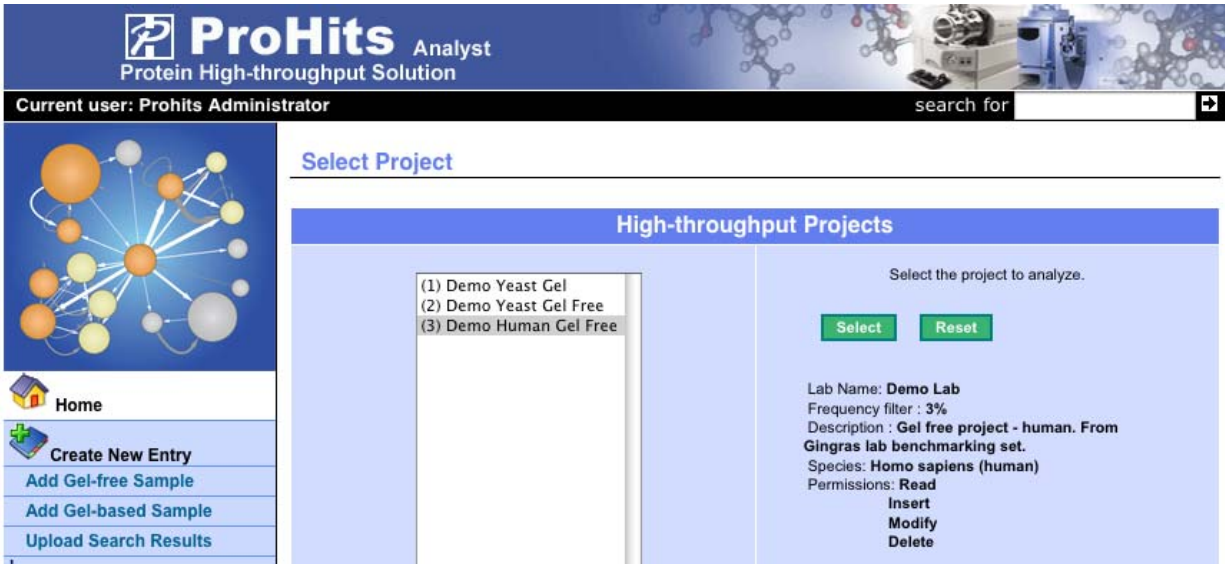


Access to projects

Projects are created by your administrator in the "Admin Office" ProHits module, and access is granted to users. Projects can be specific to a research group or an individual, to a given organism or specific methodology, etc. The creation of a new project is defined in the "Admin Office" manual.

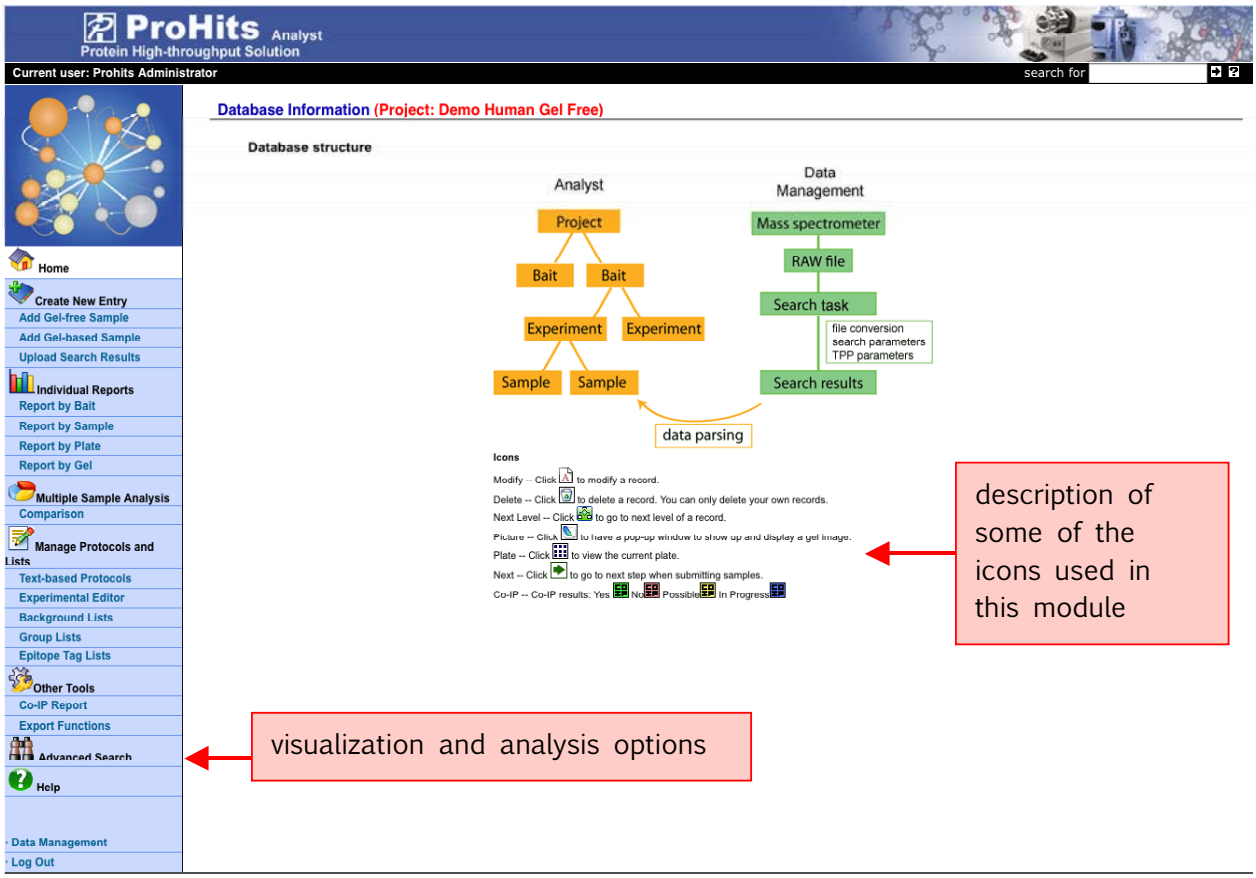
When you log into ProHits with your user name, you can see the list of all of the projects that you have access to. You may have different privileges for each project.

fi Highlight the desired project, then hit "Select"



Analyst main page

When you enter a project within the Analyst module, you will see the data workflow and a summary of the icons used in this module. The navigator bar on the left lists various visualization and analysis options.



Description of the navigator bar options:

1- **Create New Entry** allows you to define a bait, experiment, sample, and to link mass spectrometry data to this entry. These entries can then be linked to specific files in the MS Data Management module. Alternatively, you can upload search results created by external software.

2- **Individual Reports** allows you to explore your mass spectrometry results. *Report by Bait*: provides a list of all baits entered in the database for this project. *Report by Samples*: lists all samples entered for this project (a bait may be linked to multiple samples, especially in gel-based projects; we also use this nomenclature for technical replicates). *Report by Plate*: sample tracking for high-throughput projects, typically gel-based. *Report by Gel*: allows you to visualize results for each gel (gel-based projects only).

3- **Multiple Sample Analysis** (Comparison): allows you to simultaneously visualize multiple result pages.

4- **Manage Protocols and Lists** allows you to create and maintain experimental protocols, controlled vocabularies, background lists, group lists and epitope tag lists. Access to these pages is defined by the ProHits Administrator.

5- **Other Tools** provides additional functionality. *Co-IP Report*: allows you to input results from follow-up experiments aimed at confirming interaction pairs by immunoprecipitation/immunoblotting. *Export Functions*: allows you to export filtered or unfiltered lists of mass spectrometry results. Note that export functions are also available within each of the Individual Report or Comparison pages.

6- **Advanced search** allows you to query your project for genes, keywords and/or controlled vocabularies.

Creating samples and viewing individual reports

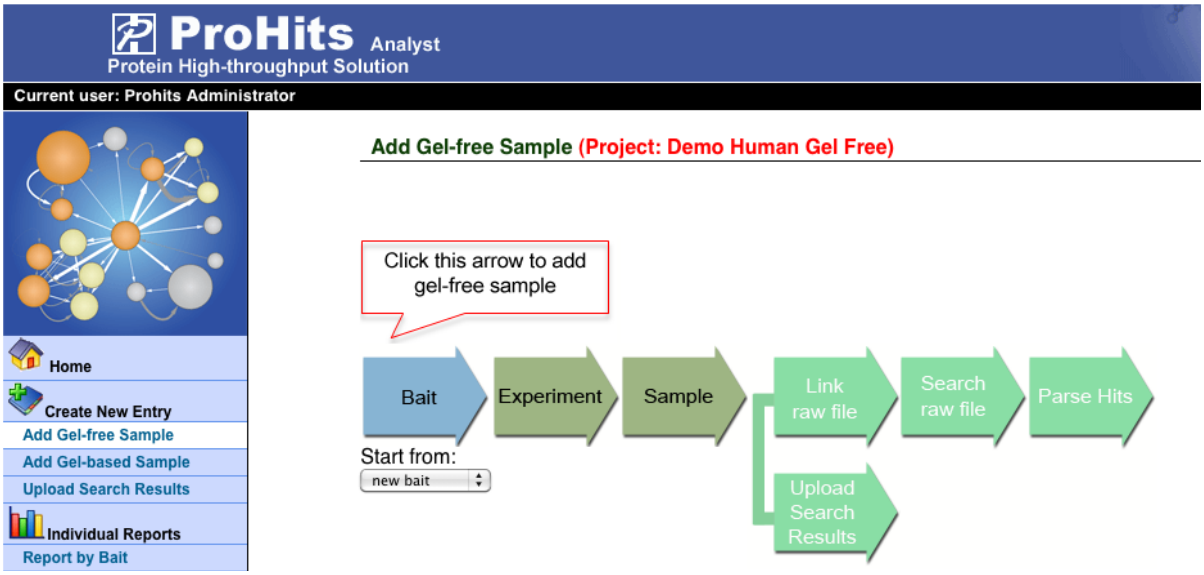
To learn more about the different functions of ProHits Analyst, we will navigate through the Analyst module by creating new baits and linking them to entries from the Data management system. We will then explore the functions available in the Analyst module. We will go through the process of adding a gel-free sample and explore the results for this type of project. We will then briefly review the differences between submitting gel-free and gel-based samples.

Adding a "Gel-Free" sample

To create a new sample to be linked to a search result file, you will first specify a bait, then an experiment, and then a sample. To submit a sample, you have two options: 1) create a new sample from an existing bait; or 2) create a new bait. Here we will start by creating 5 new baits for this project.

Creating a bait

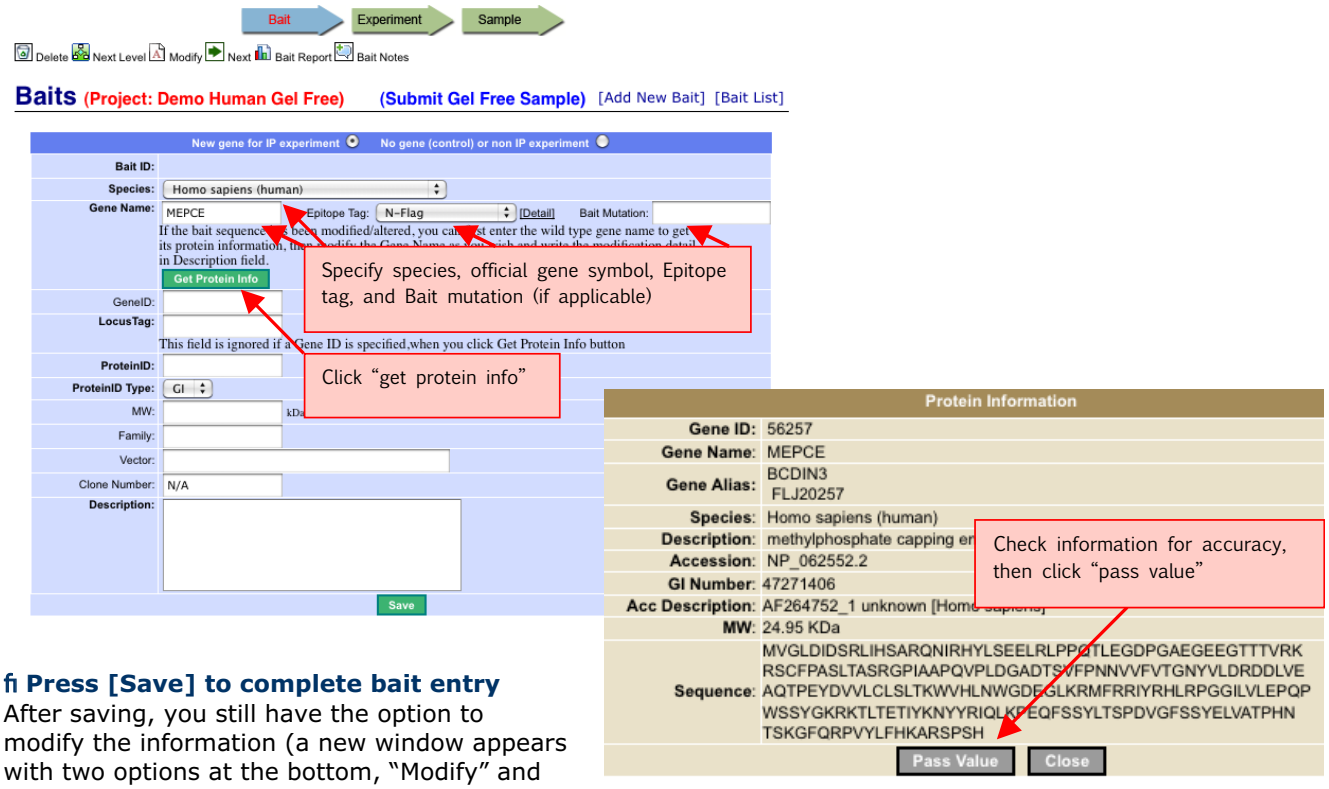
fi Select the "Add Gel-free Sample" link under "Create New Entry". Select "new bait" from the dropdown menu, then click on the "Bait" Blue arrow.



This will open a new page. Note at the top of the page the data structure; the Bait is highlighted, indicating that you are adding bait level entries. Note that each of the baits is automatically assigned a unique numeric identifier. The fields highlighted in bold indicate that the information is mandatory, but many of these can be filled automatically.

The easiest way to enter a new bait is to simply 1) select the desired species (here we have selected *Homo sapiens*); 2) enter an official Gene Name (HUGO for human; here we selected MEPCE); 3) click the "Get Protein Info" green button. Clicking "Get Protein Info" automatically retrieves the protein information which is displayed in a new window. Verify this information and hit [Pass Value] if correct – the information will automatically be transferred. Note that if there is more than one entry mapped to a given gene, the user can select which one is to be parsed into ProHits.

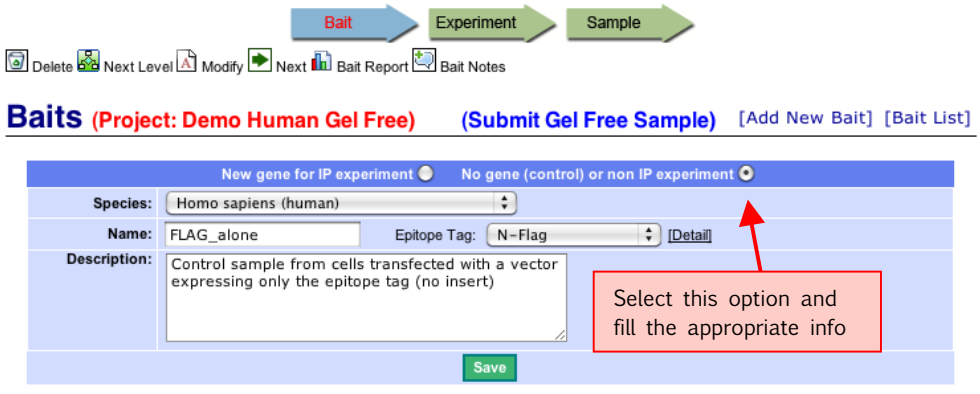
ProHits also allows you to indicate which epitope-tag you are using, by selecting from options in the "Epitope Tag" menu; you can also add new tags using the "Manage Protocols and Lists" option. If the sequence of the bait is mutated relative to the HUGO sequence, you can also enter this in the "Bait mutation" box.



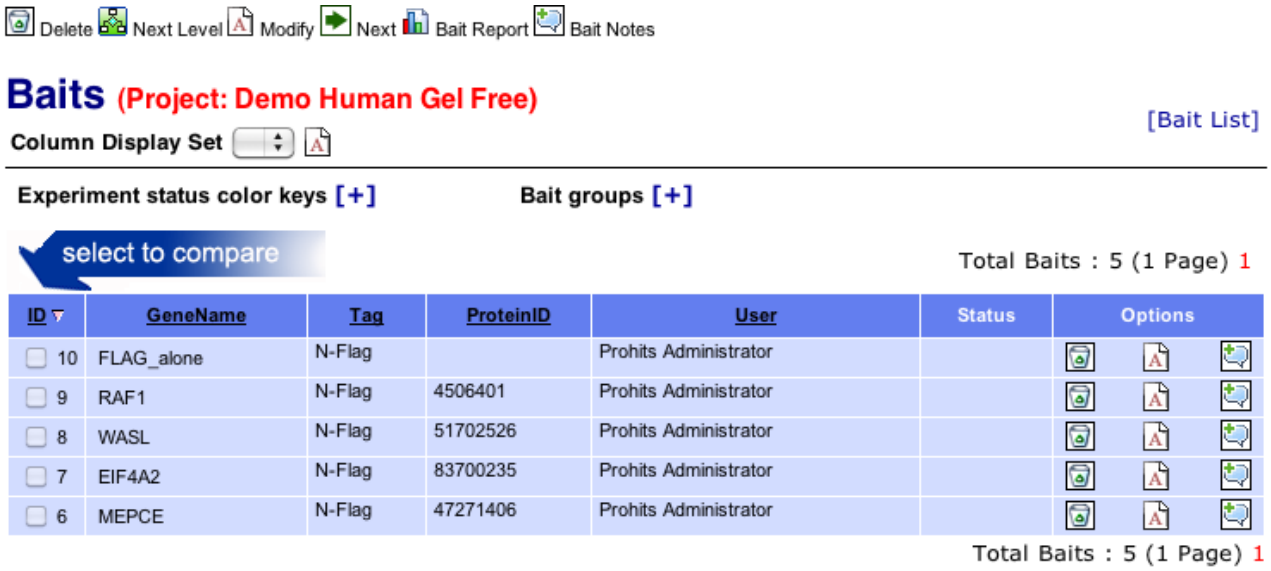
fi Press [Save] to complete bait entry
After saving, you still have the option to modify the information (a new window appears with two options at the bottom, "Modify" and "Next"). You can add additional information, e.g. in the "Description" field, or modify existing information. Hitting [Next] would bring up the Experimental detail page (for this demonstration, we will not do this yet).

Also note that you can create baits for sequences that are not in the database by manually filling in all bold fields (species, gene name, locus tag, protein ID, protein ID type). ProHits does not check for accuracy in these entries. You may wish to use this option, for example, for recombinant or chimeric proteins not corresponding to any of the entries in the database.

fi Use the [Add New Bait] button at the top of the page, and continue defining baits in the same manner as for MEPCE. Note that in the bait entry page, you can also define an experiment in which no gene/protein was tagged. To do so, simply select the "No gene (control) or non IP experiment" button at the top of the page, and manually enter information. Here we are adding a "FLAG alone" bait.



fi To visualize the entry of your new baits in the database, go back to the left bar menu and select [Report by Bait]



The Bait Report now lists the baits we have created (MEPCE, FLAG alone, and 3 additional baits that we will use for the demonstration of the functions of ProHits), along with some relevant information. The "ID" column lists a unique identifier for the bait that is automatically assigned by ProHits. The Gene Name and Tag are indicated, and the Protein ID is the accession number from the selected database (e.g. NCBI-GI). The "User" column is automatically assigned to the user who created the sample (i.e. the person who has signed up in ProHits).

Note that, on many of the ProHits pages, you will find standard icons (as seen at the top of the Bait Report page).

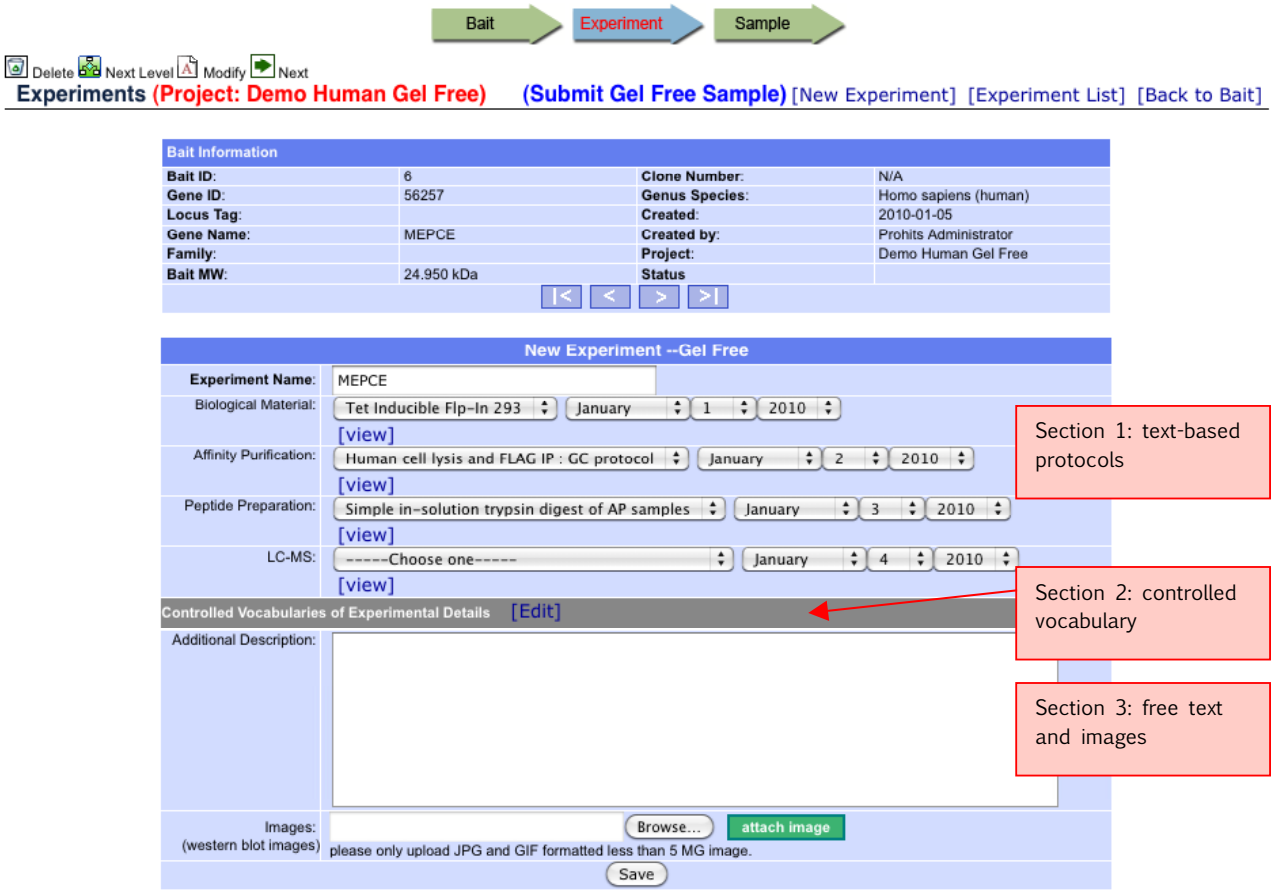
1- To remove unused material, press the "delete" icon. The "delete" function may be used to remove baits, experiments or samples, but *only if no information has been entered*. If you wish to delete a bait, experiment or sample for which information has been entered, start by deleting the information at a lower level, and work your way up. (Note that there is an Admin control for the permissions to insert, modify and delete entries, and you can only delete your own entries).
2- The next level (tree) icon allows you to navigate down in the data structure (i.e. from bait to experiment to sample).
3- The Modify icon allows you to change the information you entered for a bait, experiment or sample.
4- The green arrow (Next) icon allows you to submit information and/or exit a page after data has been entered.
5- The "Bait report" (graph) icon shows you the mass spectrometry results for the selected bait. We will review this in detail later.
6- Finally, the "Bait Notes" (callout) icon allows you to enter specific notes/information for baits or samples. Such notes can be a manually entered discussion point. Other types of notes include assignment of a project to a user-defined "bait group".

Now that you have created baits, you are ready to define your experiments. Note that in many cases, you will be seamlessly going from bait to experiment to sample when entering real samples. Here, we have simply separated these modules for ease of teaching.

Creating an experiment

fi Return to [Add Gel-free sample], and select [start from existing bait].
This will bring up essentially the same page as shown above, but with an additional option (green arrow) at the extreme right of each row.

fi Select this green arrow to enter the experimental details for a given bait
The experimental detail page allows you to specify experimental conditions and protocols used for the experiment. The top of the page states the bait information: below, the definition of an experiment can be separated into three sections.



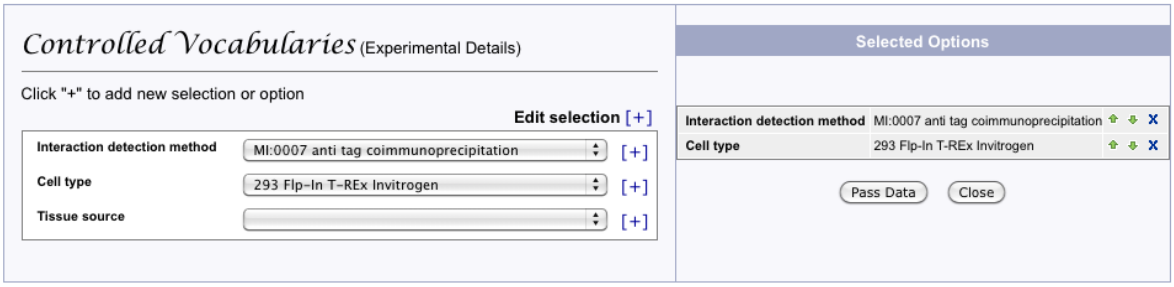
User-defined free-text protocols
In section 1, drop-down menus allow for the selection of user-specified protocols for each experiment. We suggest describing generic protocols in detail (in a manner similar to the Methods section of an article). The protocols can be entered and managed using the "Manage Protocols and Lists" option (more on this later).

Controlled vocabulary
Section 2 offers (via the Experimental Detail Editor) the possibility to specify controlled vocabulary to describe the experiment. The controlled vocabulary is specified for each project by using the "Experimental Editor" option within "Manage Protocols and Lists". Note that this vocabulary can facilitate compliance to community guidelines, such as HUPO Proteomics Standard Initiative (e.g. PSI-MI 2.5). This controlled vocabulary (drop-down keywords) can be used for searching and structuring the data using the "Advanced Search" option.

Additional annotation
Section 3 allows for additional free-text annotation in the form of notes. Here you can cross-reference to notebook page numbers, add specifics of the experiment not captured in sections 1 and 2, or describe any problem or deviation from the reference protocols. It also allows you to link image files (e.g. Western blots or silver stained gels).

fi Navigate through the dropdown menus to select appropriate protocols associated with the experiment.

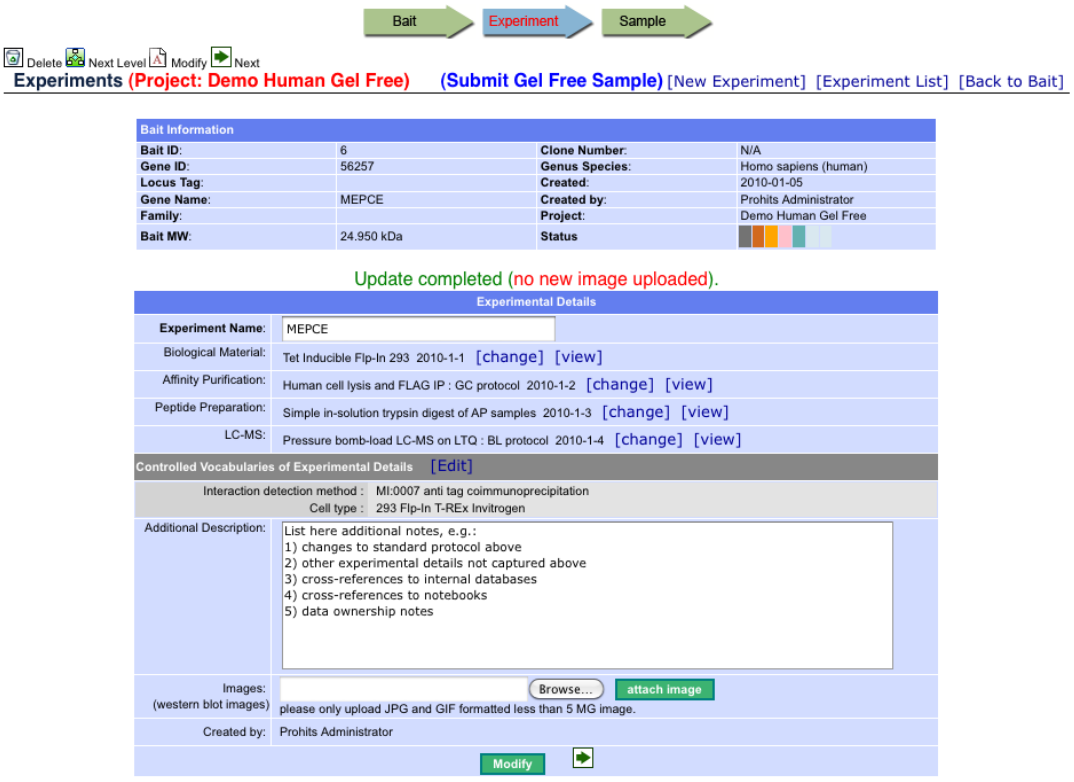
Note that selecting the option "Edit" within Section 2: Controlled Vocabularies of Experimental Details will open up a new window with dropdown menus.



fi Select all desired fields to capture using the dropdown menus.
The selected options will be displayed on the right hand side in the order that they were selected. Use the Up/Down green arrows to change the order, or click on the **x** to remove the entry.

fi Select [Pass Data] to transfer selection to the Experimental Detail page or [Close] to exit without saving the data.

fi Continue filling experimental details, link any desired image, and press [Save].



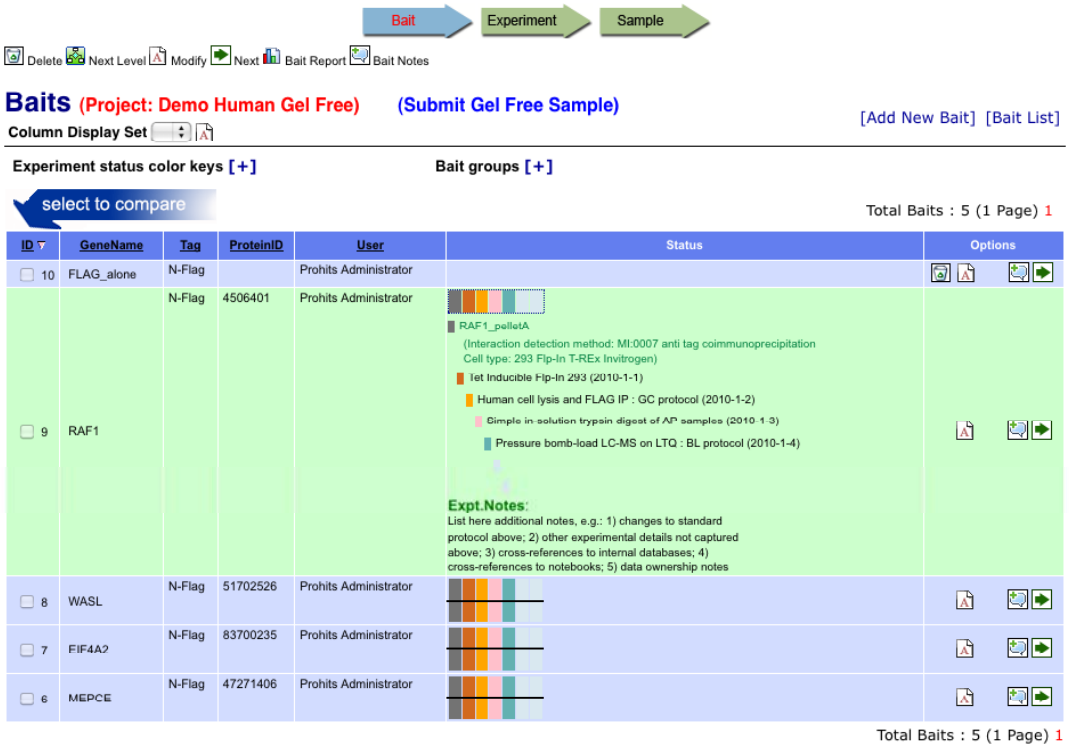
Upon saving, you will be given the option to "Modify" the entry or follow the green arrow to the next page to enter specific samples. Additionally, you can continue creating experiments by toggling between the [New Experiment], [Experiment List] and [Back to Bait] buttons at the top of the page to enter biological replicates for each of the baits.

fi Return periodically to the [Back to Bait] list to monitor your progress.

Note the colour-coded experimental status bars in the table. This view shows our five baits, with experiments defined for four of them (MEPCE, EIF4A2, WASL and RAF1). The status column displays experimental details, experimental status and bait groups (see below). The colour-coding in the "Status" column indicates that information has been entered for each of the specified fields.

fi Click on the colour-coded status bar to obtain additional experimental details

In the Bait view, experiments (and samples) defined under the same bait will be combined in the same row; multiple experiments will be shown by stacked colour bars. Note that you cannot delete baits for which experiments have been defined (note in the picture below that the FLAG_alone bait can still be deleted, since no experimental details have been entered yet). Start by deleting the Experimental Details, and work your way up as previously described.

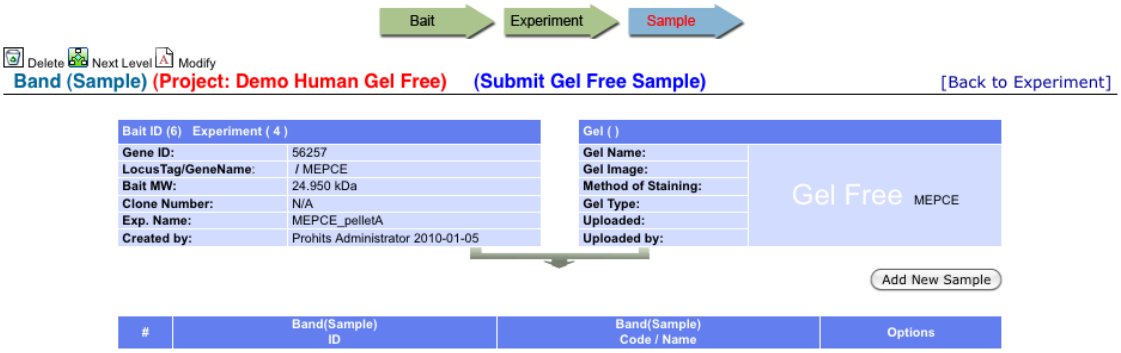


Once your baits and experiments are entered, you can create one or multiple samples to be linked to the bait and experiment. The number of samples you create for a given experiment depends on your experimental set-up. We tend to use different samples from a single experiment to represent technical replicates (i.e. different MS runs from the same biological sample), where all conditions are the same. Alternatively, multiple samples from one experiment may be created when the sample has been fractionated (e.g. by strong cation exchange) prior to the analysis. Each of the fractions is then assigned a different sample name within the same experiment. The "Notes" sections from the Experimental Details page should explain the sample-naming scheme. Note that we enter biological replicates as different experiments from the same bait.

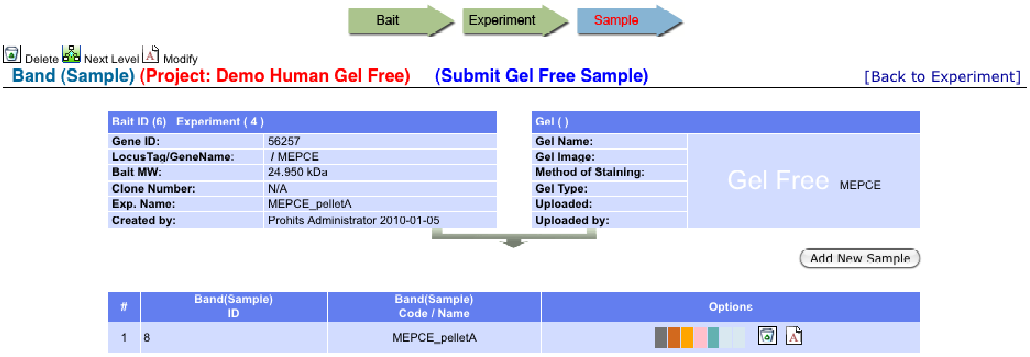
Creating a new sample

Following the green arrow on any of the Experimental Details pages will open a new window, allowing you to create one or many samples for a given experiment.

fi In the Sample page, select the [Add New Sample] button to create a sample entry for this bait and set of experimental conditions.



By default, ProHits will use the experiment name to name the first sample created from the relevant experiment. ProHits will also assign a unique Sample ID. The sample name can be modified if necessary (in this case, just type the desired sample name in the text box). In our group, we reserve the creation of duplicate samples from the same bait/experiment for technical replicates (e.g. if we split the final sample in half, and run each half separately). Note that creating multiple samples from a single bait/experiment results in an automatic appending of _A, _B, etc. at the end of the sample name. As long as a sample is not linked to any RAW file, it can be deleted by the owner.



Now that you have created a new sample entry, you are ready to link it to a mass spectrometry raw data file from the Data Management module. Links can be created automatically if the nomenclature indicated in the notice below for file naming is respected, and ProHits Data Management module is connected to the acquisition computers. Alternatively, links can be created manually either from the Data Management or the Analyst modules.

Notice:
In order to link a raw file to a gel free sample automatically, name the folder and raw file as follows.

- raw file(s) is in a folder, the folder name ends with the Project ID.
- start with sample ID and first 4 characters of sample code.

AnyNameNoSpace_P3 (project ID is 3)
8_MEPC_AnyOtherWordNoSpace.RAW

Linking raw files to a created sample

Linking raw files from the Data Management module

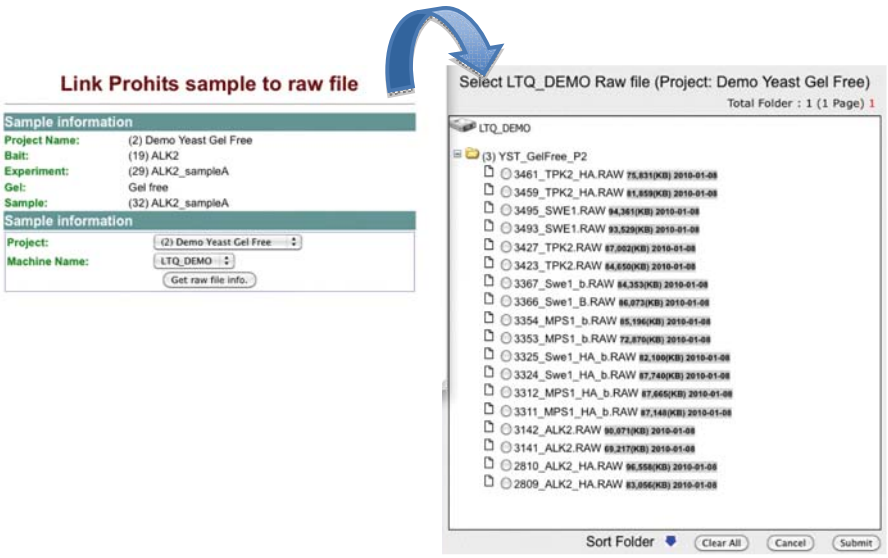
ﬁ From any page in the Analyst module, click "Data Management" on the left menu bar (shown by orange arrow on the right), link the desired file (as described in the Data Management section), parse the hits and return to the Analyst module.

Linking raw files directly from the Analyst module (alternative)
For this alternative example, we are linking files from the Demo Yeast Gel free project, which you can access by going back to the home page of the Analyst module.

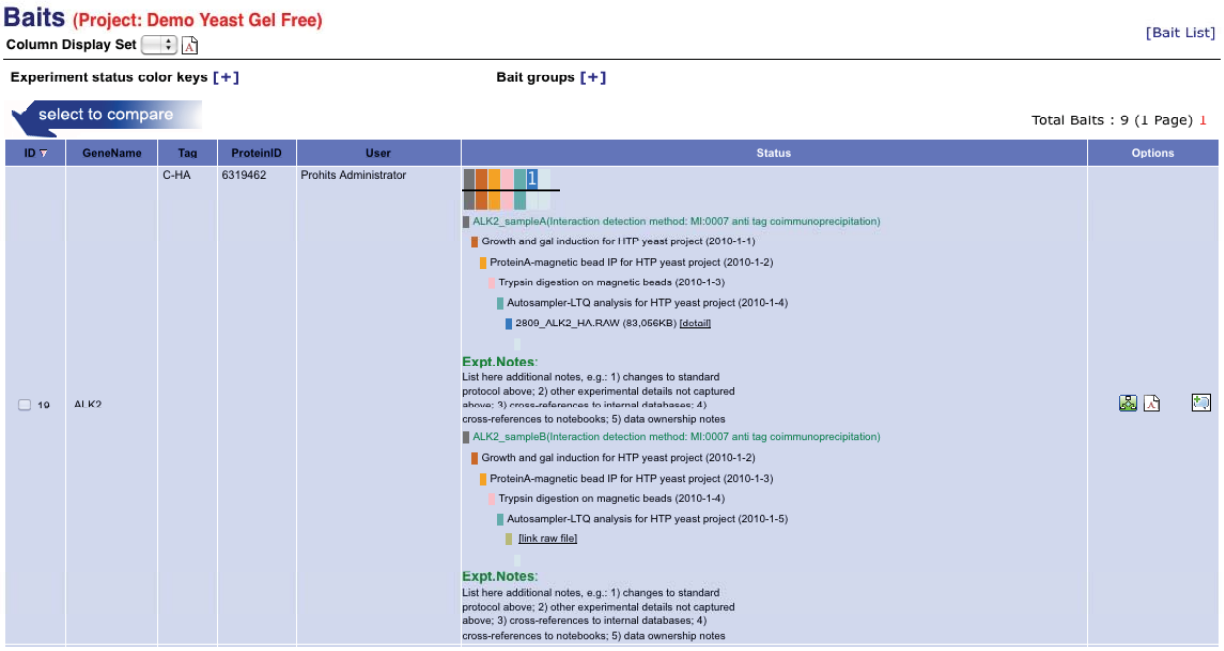
ﬁ Go to the "Report by Bait" or "Report by Sample" page of the Analyst module and click on the Status column of the desired file to display experimental details. Select [Link raw file].

This brings up a new page that allows you to select the file to be linked to the given entry.

Note that when you link files from the Analyst module, only those files not previously linked to another entry will be displayed. To modify an existing link, you need to go back to the Data Management module, remove the link to the initial file, so that it can be made available to link to an entry either through the Analyst or Data Management modules.

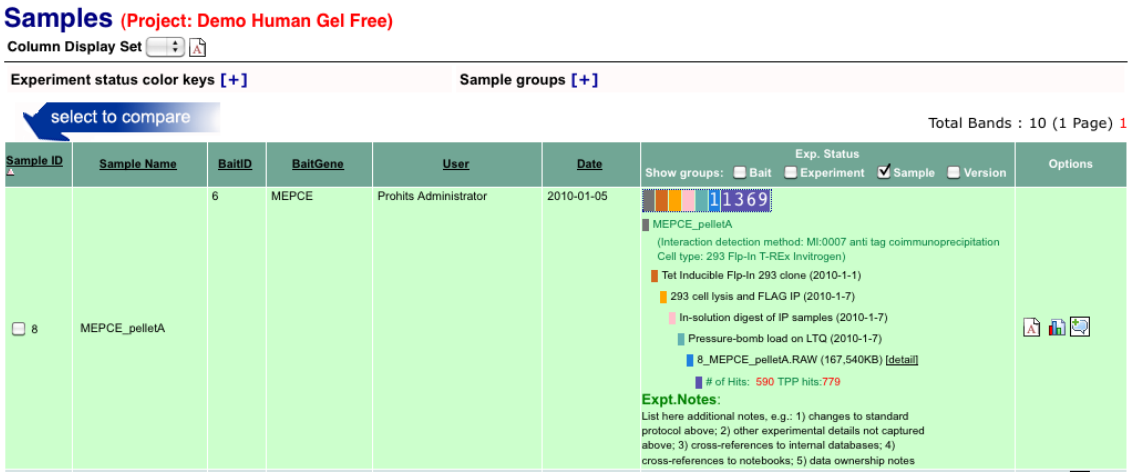


Once a raw file has been linked, the status bar will display an additional blue icon; the number indicates the number of files linked to that entry.



fi Click the "Data Management" tab from any page of the Analyst module, parse the hits (as described in the Data Management section) and return to the Analyst module.

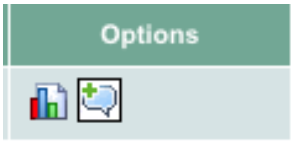
Once hits are parsed (either from the Data Management or the Analyst module), a new purple coloured tab will appear in the status bar (in either Bait Report or Sample Report pages), indicating the total number of hits identified (sum of hits if more than one search engine was used). In the "Options" column, a new graph icon appears; clicking this link brings up the search results for each sample. Here we are showing MEPCE_pellet A in the sample report view.



You are now ready to explore your results. Use the left-hand side of the ProHits Analyst main page to view "Report by Bait" and "Report by Sample". The interface for the Bait and Sample reports is very similar. Here we provide an example for the Sample Report. *Bait versus Sample view*: For some projects you may have a one-to-one correspondence between bait and sample. For other projects, you will have multiple samples linked to the same bait. Opening the Bait Report when two or more samples are linked to the bait will generate sequential protein hit lists for each of the samples linked to the bait. ProHits does not recalculate scores or peptide numbers, but indicates (in bold) proteins detected in more than one sample (mousing over bolded names activates a pop-up window that provides details about the samples and hit scores). If you wish to explore each sample individually, use the "Report by Sample" link instead.

Navigating through the results

Now that we have entered baits, linked and parsed search results, it is time to look at search results. In this example, we will start from the "Report by Sample" page for MEPCE_pelletA.



From the sample list page, under "Options", select the graph icon from one of the samples to see the results.

The following page appears, displaying the results from your search engine (Mascot in this example), alongside links to initial search results and biological databases. Additional export and viewing functions, as well as options to filter the hits are also available from this page. Over the next several pages, we will explore the Results page.

Search results

Towards the bottom of the page are the search results – by default, these are not filtered. The red colour in the ID field indicates the bait (as defined by the user when entering the experimental description). There are several tabs at the top of the search results table available for navigation. The exact tabs displayed depend on the search engines used. For the demonstration project, we have used the search engines Mascot and X!Tandem (GPM), and have analysed the results using the TransProteomics Pipeline. We will first explore the "Mascot Hits" tab.

The columns list the following parameters:

A) ID: Unique identifier assigned by ProHits (for database purposes)
B) Protein: Protein accession number from original database used by the search engine
C) Gene: NCBI Gene ID/ Gene Symbol, mapped by ProHits from Protein accession
D) Score: Mascot score (if applicable)
E) Expect value: GPM / X!Tandem Expect value (if applicable)
F) Frequency: The frequency that this protein hit is detected across all samples analyzed for this project
G) Redundant: Other protein accession numbers matching the same set of peptides
H) MW kDa: Calculated MW for the protein
I) Description: Definition field from the NCBI protein entry
J) # Peptide: Spectral counts (or total peptides), as calculated by the search engine
K) # Unique Peptide: Number of unique peptides, as calculated by the search engine

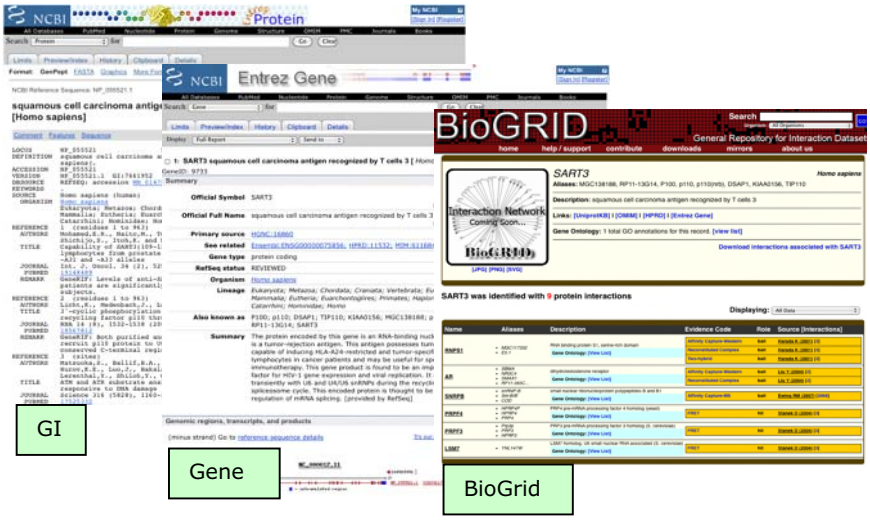
L) Coverage: Percentage of the indicated amino acid sequence identified by your search engine
M) Links: External links to the NCBI Entrez Protein page [GI], the NCBI Gene Page [Gene] and the BioGrid [BioGrid].
N) Filter: provides a colour-coded view of the Experimental Filters or Bio Filters that could be applied to remove each hit
O) Option: Provides the list of peptides belonging to this hit (green M icon), opens up the original search engine search results (here Matrix Science icon for Mascot search results), and allows for the addition of Notes (call-out icons; includes manual exclusion)

Sorting options

You can sort the results from any of the black underlined columns (Score, #Peptide, #UniquePeptide and Coverage); sorting can be in ascending or descending value.

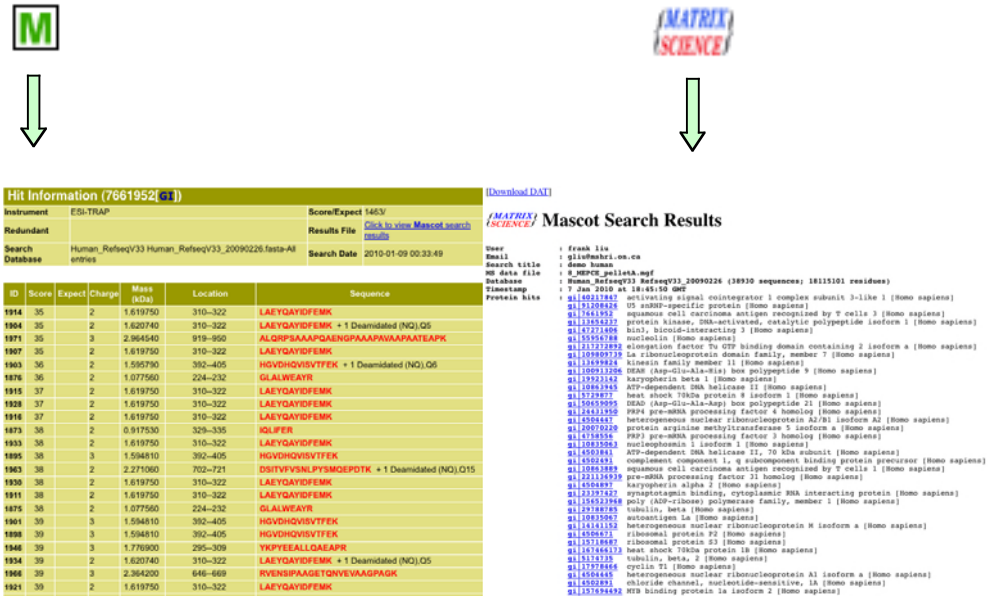
Links details

The following pages can be obtained from each of the items in the "Links" column.



Option details

Pressing the following icons in the Option column will retrieve the peptide list (from the search engine) for each hit, or the entire search results file.



Hit Information (7661952 GI)

Instrument	ESI-TRAP	Score/Expect	1463
Redundant		Results File	Click to view Mascot search results
Search Database	Human_RefseqV33 Human_RefseqV33_20090226.fasta-All entries	Search Date	2010-01-09 00:33:49

ID	Score	Expect	Charge	Mass (kDa)	Location	Sequence
1914	35		2	1.619750	310–322	LAEYQAYIDFEMK
1904	35		2	1.620740	310–322	LAEYQAYIDFEMK + 1 Deamidated (NQ).Q5
1971	35		3	2.964540	919–950	ALQRPSAAAPQAENGPAAAPAVAAPAATEAPK
1907	35		2	1.619750	310–322	LAEYQAYIDFEMK
1903	36		2	1.595790	392–405	HGVDHQVI5VTFEK + 1 Deamidated (NQ).Q6
1876	36		2	1.077560	224–232	GLALWEAYR
1915	37		2	1.619750	310–322	LAEYQAYIDFEMK
1928	37		2	1.619750	310–322	LAEYQAYIDFEMK
1916	37		2	1.619750	310–322	LAEYQAYIDFEMK
1873	38		2	0.917530	329–335	IQLIFER
1933	38		2	1.619750	310–322	LAEYQAYIDFEMK
1895	38		3	1.594810	392–405	HGVDHQVI5VTFEK
1963	38		2	2.271060	702–721	DSITVFVSNLPYSMQEPDTK + 1 Deamidated (NQ).Q15
1930	38		2	1.619750	310–322	LAEYQAYIDFEMK
1911	38		2	1.619750	310–322	LAEYQAYIDFEMK
1875	38		2	1.077560	224–232	GLALWEAYR
1901	39		3	1.594810	392–405	HGVDHQVI5VTFEK
1898	39		3	1.594810	392–405	HGVDHQVI5VTFEK
1948	39		3	1.776900	295–309	YKPYEEALLQAEAPR
1934	39		2	1.620740	310–322	LAEYQAYIDFEMK + 1 Deamidated (NQ).Q5
1966	39		3	2.364200	646–669	RVENSIPAAGETQNVEVAAGPAGK
1921	39		2	1.619750	310–322	LAEYQAYIDFEMK

[Download DAT]

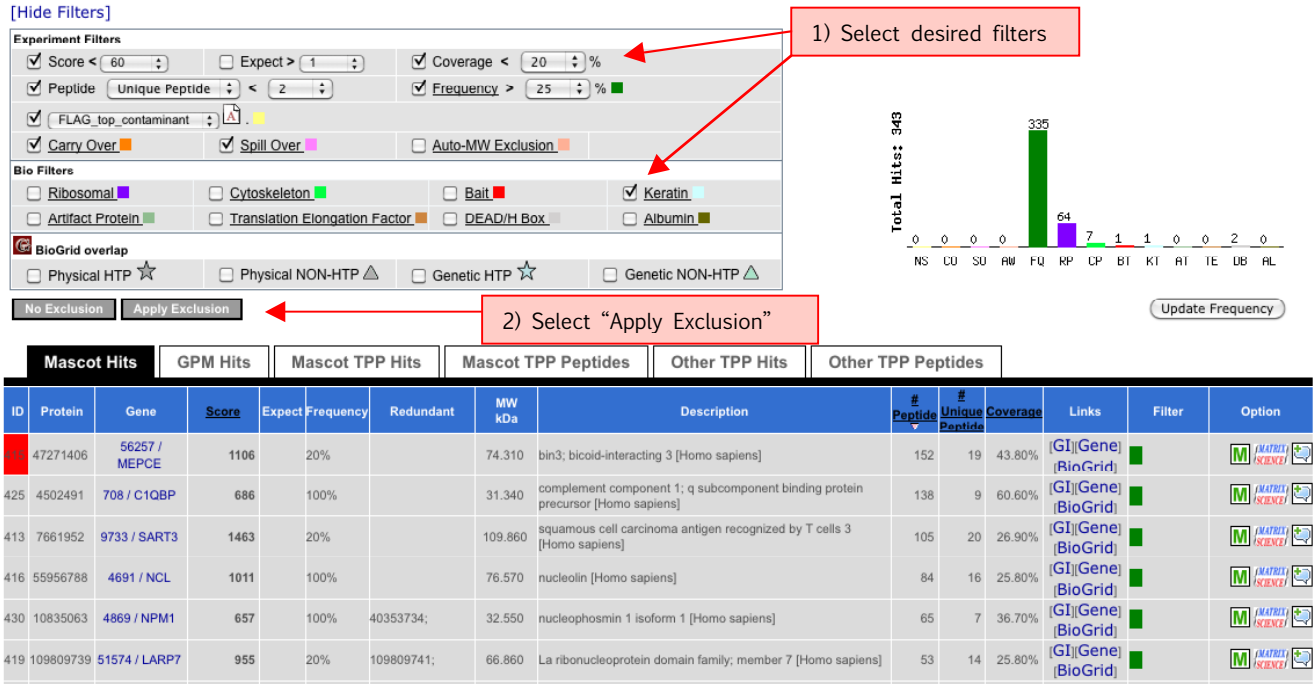
Mascot Search Results

User : frank liu
Email : gliu@mshri.on.ca
Search title : demo human
MS data file : 8_HEPCH_pelletA.mgf
Database : Human_RefseqV33 RefseqV33_20090226 (38930 sequences; 18115101 residues)
Timestamp : 7 Jan 2010 at 18:45:50 GMT
Protein hits : gi|48217847 activating signal cointegrator 1 complex subunit 3-like 1 [Homo sapiens]
gi|91208426 U5 snRNP-specific protein [Homo sapiens]
gi|7661952 squamous cell carcinoma antigen recognized by T cells 3 [Homo sapiens]
gi|113454237 protein kinase, DNA-activated, catalytic polypeptide isoform 1 [Homo sapiens]
gi|47271406 bin3, bicoid-interacting 3 [Homo sapiens]
gi|55956788 nucleolin [Homo sapiens]
gi|21727892 elongation factor Tu GTP binding domain containing 2 isoform a [Homo sapiens]
gi|109809739 La ribonucleoprotein domain family, member 7 [Homo sapiens]
gi|13699824 kinesin family member 11 [Homo sapiens]
gi|190911206 DEAH (Asp-Glu-Ala-His) box polypeptide 9 [Homo sapiens]
gi|19923162 karyopherin beta 1 [Homo sapiens]
gi|1086965 ATP-dependent DNA helicase II [Homo sapiens]
gi|5729877 heat shock 70kDa protein 8 isoform 1 [Homo sapiens]
gi|50659995 DEAD (Asp-Glu-Ala-Asp) box polypeptide 21 [Homo sapiens]
gi|24631958 PRP4 pre-mRNA processing factor 4 homolog [Homo sapiens]
gi|4504447 heterogeneous nuclear ribonucleoprotein A2/B1 isoform A2 [Homo sapiens]
gi|20070220 protein arginine methyltransferase 5 isoform a [Homo sapiens]
gi|4758556 PRP3 pre-mRNA processing factor 3 homolog [Homo sapiens]
gi|10835063 nucleophosmin 1 isoform 1 [Homo sapiens]
gi|4503841 ATP-dependent DNA helicase II, 70 kDa subunit [Homo sapiens]
gi|4502491 complement component 1, q subcomponent binding protein precursor [Homo sapiens]
gi|10863883 squamous cell carcinoma antigen recognized by T cells 1 [Homo sapiens]
gi|22113939 pre-mRNA processing factor 31 homolog [Homo sapiens]
gi|4504897 karyopherin alpha 2 [Homo sapiens]
gi|21397327 synaptotagmin binding, cytoplasmic RNA interacting protein [Homo sapiens]
gi|156523968 poly (ADP-ribose) polymerase family, member 1 [Homo sapiens]
gi|29788785 tubulin, beta [Homo sapiens]
gi|10835067 autoantigen La [Homo sapiens]
gi|14141152 heterogeneous nuclear ribonucleoprotein M isoform a [Homo sapiens]
gi|4506671 ribosomal protein P2 [Homo sapiens]
gi|15718887 ribosomal protein S3 [Homo sapiens]
gi|167466173 heat shock 70kDa protein 1b [Homo sapiens]
gi|5174735 tubulin, beta, 2 [Homo sapiens]
gi|17978466 cyclin T1 [Homo sapiens]
gi|4504445 heterogeneous nuclear ribonucleoprotein A1 isoform a [Homo sapiens]
gi|4502891 chloride channel, nucleotide-sensitive, 1A [Homo sapiens]
gi|157694492 MTB binding protein 1a isoform 2 [Homo sapiens]

We have now navigated through the table listing the search results. However, the initial list is not filtered; that is, all hits, including likely contaminants, are listed. ProHits has a built-in filter set that can be applied to the data to help identify *bona fide* interactors.

Using filters

Click on the [Show Filters] button within the results page to display the administrator-defined Bio and Experimental filters (see admin office for details of the filtering options) and background lists (see Manage Protocols and Lists) that can be applied to the data in this project. On the left is the filter list and the graph on the right indicates the number of proteins that would be removed by activating each of the filters. Filters are activated or de-activated by clicking their associated checkbox. Once the desired filters are selected, press "Apply exclusion" to remove associated proteins from the search results list. Note that the default frequency filter is set in the admin office module when creating the project, and that this value is listed when you select the project from the home page (see page 3). In the case of the "Demo Human Gel Free" project shown here, the frequency filter was set at 3%, meaning that a protein detected in >3% of samples within the project is flagged (as shown by the dark green icon in the results table). You do not need to use the default filter, and can modify this frequency cut-off as needed. Also note that the frequency is not automatically recalculated every time you add a search result to ProHits: to recalculate the frequency, use the "Update Frequency" button on any "Report" page.

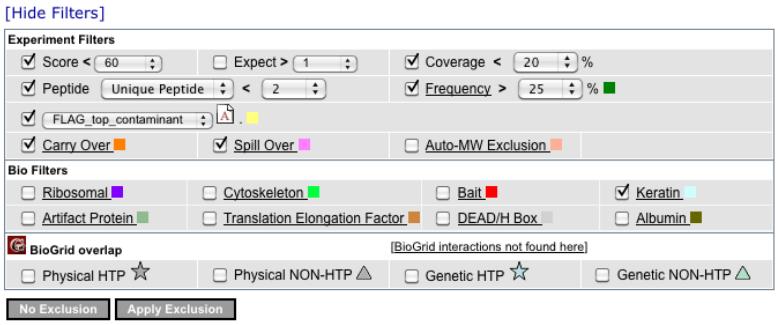


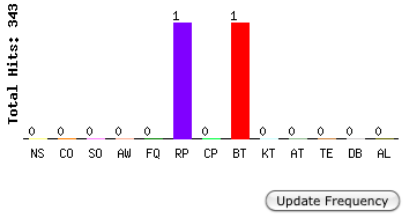
ID	Protein	Gene	Score	Expect	Frequency	Redundant	MW kDa	Description	# Peptide	# Unique Peptide	Coverage	Links	Filter	Option
[illegible]	47271406	56257 / MEPCE	1106		20%		74.310	bin3; bicoid-interacting 3 [Homo sapiens]	152	19	43.80%	[GI][Gene][BioGrid]		
425	4502491	708 / C1QBP	686		100%		31.340	complement component 1; q subcomponent binding protein precursor [Homo sapiens]	138	9	60.60%	[GI][Gene][BioGrid]		
413	7661952	9733 / SART3	1463		20%		109.860	squamous cell carcinoma antigen recognized by T cells 3 [Homo sapiens]	105	20	26.90%	[GI][Gene][BioGrid]		
416	55956788	4691 / NCL	1011		100%		76.570	nucleolin [Homo sapiens]	84	16	25.80%	[GI][Gene][BioGrid]		
430	10835063	4869 / NPM1	657		100%	40353734;	32.550	nucleophosmin 1 isoform 1 [Homo sapiens]	65	7	36.70%	[GI][Gene][BioGrid]		
419	109809739	51574 / LARP7	955		20%	109809741;	66.860	La ribonucleoprotein domain family; member 7 [Homo sapiens]	53	14	25.80%	[GI][Gene][BioGrid]		

In this example, we will filter the data shown above by applying the following filters:

1) click the "background" button, and select the "FLAG_top_contaminants" list from the dropdown menu. The background lists are user-defined, and controlled via the "Manage Lists and Protocols" option.
2) proteins detected with a Mascot score <60 will be removed
3) proteins with <20% sequence coverage will be removed
4) proteins detected with a single unique peptide will be removed
5) proteins detected in >25% of the samples in this project will be removed

After applying filters, the list of hits is reduced (see the disappearance of C1QBP, NCL and NPM1 – which are common contaminants - while SART3, LARP7 and LSM8 remain). The filters can be modified and sorting repeated: ProHits does not remove any data from the dataset, but only displays filtered lists.





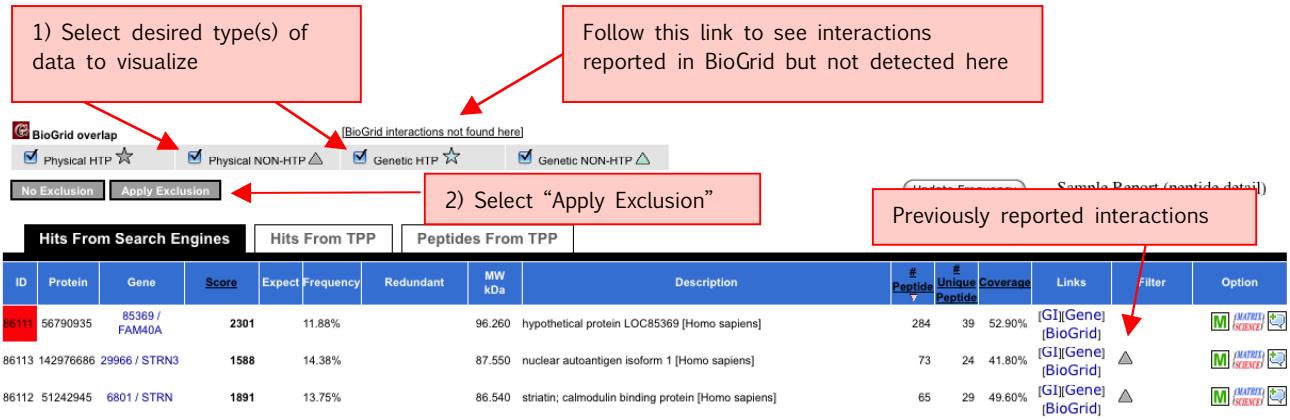
Mascot Hits | GPM Hits | Mascot TPP Hits | Mascot TPP Peptides | Other TPP Hits | Other TPP Peptides

ID	Protein	Gene	Score	Expect	Frequency	Redundant	MW kDa	Description	# Peptide	# Unique Peptide	Coverage	Links	Filter	Option
415	47271406	56257 / MEPCE	1106		20%		74.310	bin3; bicoid-interacting 3 [Homo sapiens]	152	19	43.80%	[GI][Gene] [BioGrid]		
413	7661952	9733 / SART3	1463		20%		109.860	squamous cell carcinoma antigen recognized by T cells 3 [Homo sapiens]	105	20	26.90%	[GI][Gene] [BioGrid]		
419	109809739	51574 / LARP7	955		20%	109809741;	66.860	La ribonucleoprotein domain family; member 7 [Homo sapiens]	53	14	25.80%	[GI][Gene] [BioGrid]		
468	7706425	51691 / LSM8	305		20%		10.400	U6 snRNA-associated Sm-like protein LSm8 [Homo sapiens]	49	3	61.50%	[GI][Gene] [BioGrid]		

Note that the graph on the right indicates the number of hits that have not been filtered out, but belong to the different categories that could be filtered out. In this example, after filtering, only 1 RP (Ribosomal Protein) remains, as compared to 64 in the unfiltered example.

Comparing your data to literature interactions

Prohits allows you to automatically query the BioGrid interaction database for previously-reported interactions specific to your bait. To do so, select the type of interactions desired (physical interactions from high-throughput (HTP) studies, physical interactions not from HTP studies (non-HTP), genetic interactions of both types), and press "Apply exclusion". The interactions that overlap with the literature will be highlighted in the "filter" column. (the next few figures will be replaced by MEPCE as soon as the new version of BioGrid comes online). Note that the definition of HTP and non-HTP is from BioGrid: high-throughput papers are identified as such by BioGrid curators; as a default, publications reporting >100 interactions are also identified as HTP.



Selecting [BioGrid interactions not found here] opens a new window with the details of the "missed interactions", as shown below.

Note that the overlap is performed after data filtering is applied, thus care should be taken when analyzing apparent lack of overlap. The example bellow shows the effect of the application of a stringent filter on "missed interactions".

Missing interactions: STRINGENT FILTER

BioGrid interactions not found Show found hits

Gene ID	Gene Name	Links	BioGrid Type
Bait ID: 389 Bait Gene ID: 85369			
Bait Gene Name: FAM40A [Gene][BioGrid]			
4163	MCC	[Gene][BioGrid]	☆
5515	PPP2CA	[Gene][BioGrid]	△
5516	PPP2CB	[Gene][BioGrid]	△
5518	PPP2R1A	[Gene][BioGrid]	△
10494	STK25	[Gene][BioGrid]	△
11235	PDCD10	[Gene][BioGrid]	△
29888	STRN4	[Gene][BioGrid]	△
57464	FAM40B	[Gene][BioGrid]	△
80342	TRAF3IP3	[Gene][BioGrid]	△

Missing interactions: NO FILTER

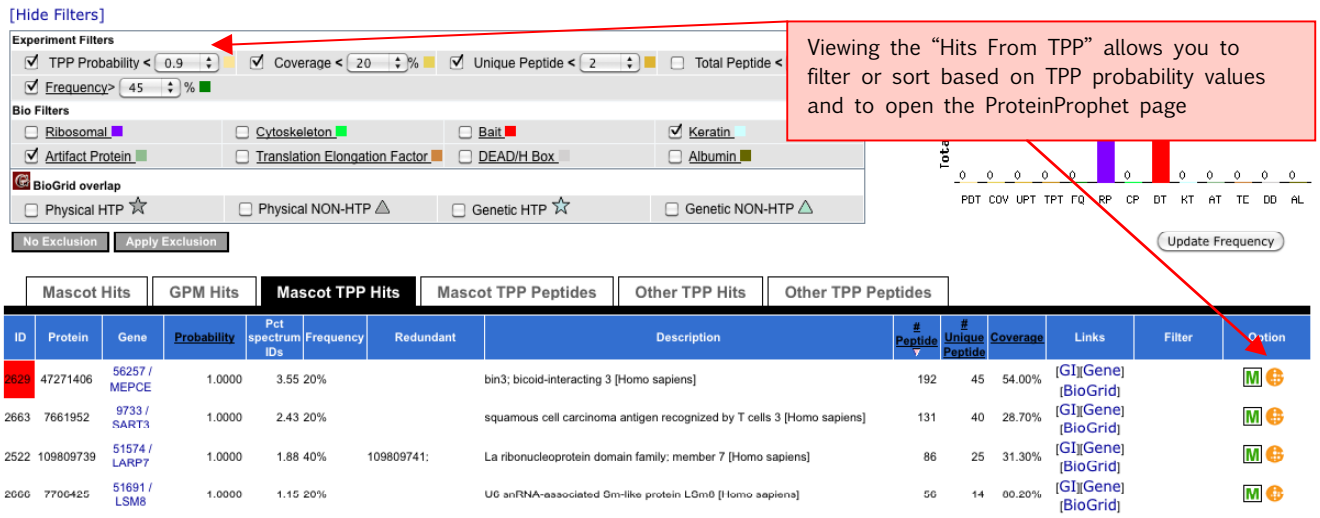
BioGrid interactions not found Show found hits

Gene ID	Gene Name	Links	BioGrid Type
Bait ID: 389 Bait Gene ID: 85369			
Bait Gene Name: FAM40A [Gene][BioGrid]			
4163	MCC	[Gene][BioGrid]	☆
5516	PPP2CB	[Gene][BioGrid]	△
10494	STK25	[Gene][BioGrid]	△
80342	TRAF3IP3	[Gene][BioGrid]	△

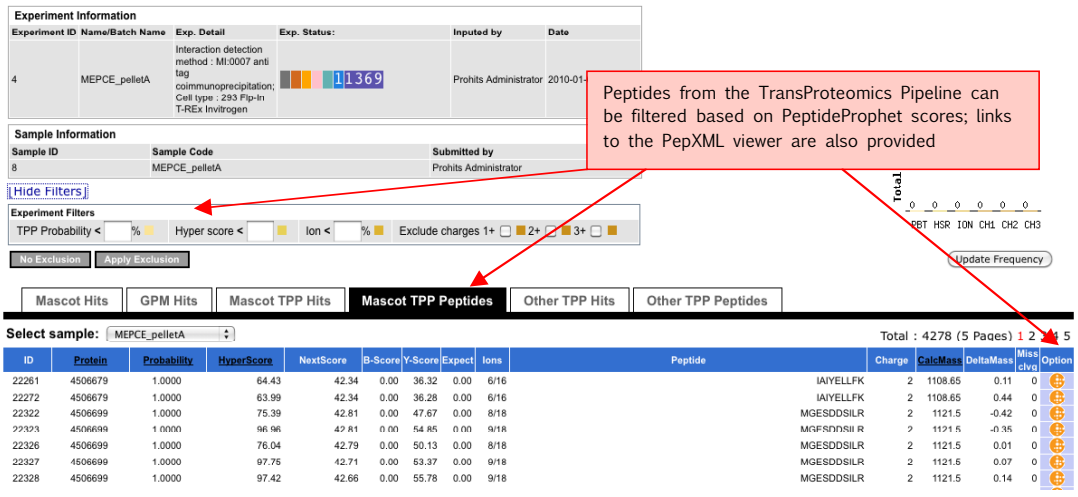
View and navigate hits from the TransProteomics Pipeline

The tabs located immediately above the search results table allow you to explore search results that have been parsed from the PeptideProphet and ProteinProphet components of the TPP.

In the page "Mascot TPP hits", different filtering options based on the number of unique or total peptides, as well as the probability values for the TPP have been implemented. A link to the TPP search result viewer is provided in the Option column of the table (orange Institute for Systems Biology icon): this opens up the standard ProteinProphet view, allowing further exploration of the data.



The "Mascot TPP Peptides" tab lists all of the parsed parameters at the peptide level, and provides some basic filtering options, as well as a link to the PepXML viewer.

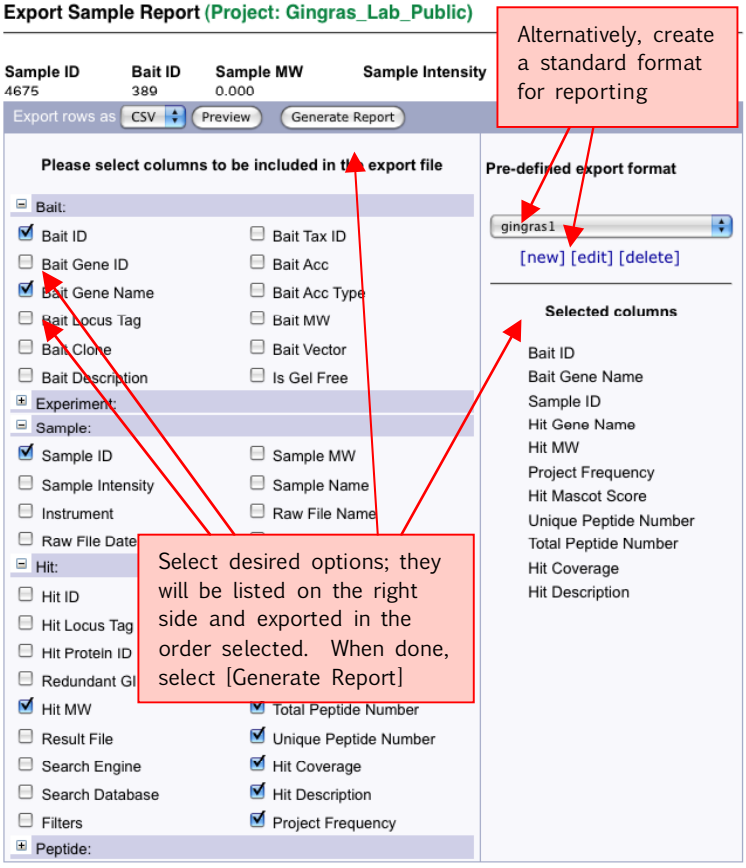


Viewing results using Cytoscape

At the top right corner on the Report page is a link to the molecular interaction visualization program Cytoscape. Clicking this link will upload the filtered data (with BioGrid interactions if this option is selected). Note that all mass spectrometry data will also be uploaded (you can use these parameters as attributes of the "edges" in Cytoscape). We will review Cytoscape requirements and basic information in the discussion of the "Comparison" function.

Export Sample report

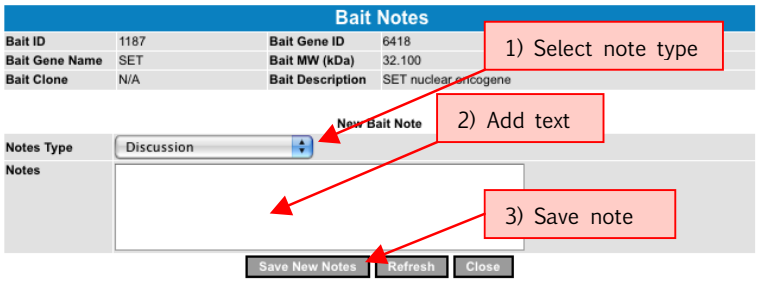
Selecting the "Export Sample Report" on the top right corner allows the user to export text (comma-separated values (CSV) or tab separated values (TSV)) files. Fields to be exported are user-defined and will be exported in the order selected. The user can also create pre-defined export formats that can be further modified. Note that this exports NON-FILTERED hits (filtered hits can be exported via the comparison tool).



The exported file can be opened with Excel or similar software.

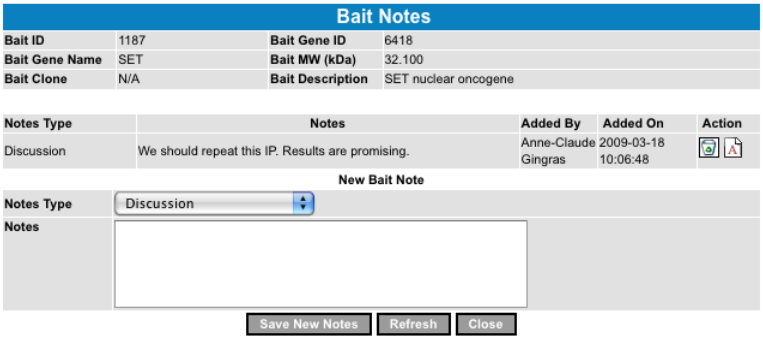
Using the Notes option

fi Click on the "callout" icon at the end of any bait row
This brings up the following window:



fi Add desired text, and press [Save New Notes]
The following screen can then be seen:

Only the person who entered the note is allowed to modify or delete it. Additional users can create additional comments on the same bait or sample.



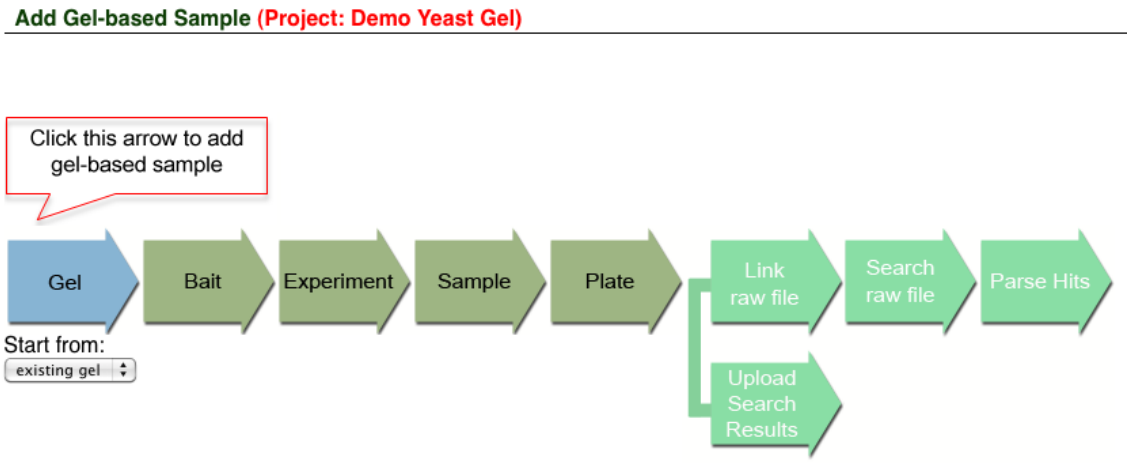
In addition to adding free text annotation (default "Discussion" note type), "Bait groups", "Experiment groups" or "Sample Groups" can be created for each project and are managed via the "Manage Protocols and Lists" option. Use the dropdown box to select the desired "Notes Types".

Creating gel-based samples

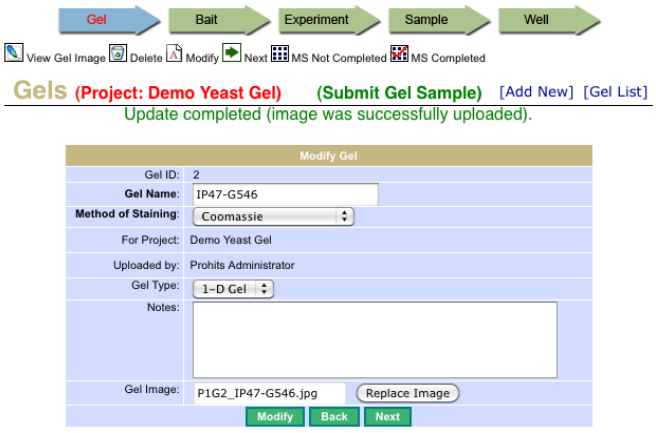
ProHits has functionality designed to track samples analyzed in a high-throughput manner from gel-based proteomics. Several of the steps are identical to the steps required to create samples for gel-free projects. Here we will briefly outline the major differences when entering gel-based samples. Note that you can add samples from in-gel digestion as "gel-free" – especially if you are only analyzing a few samples without the use of an autosampler.

Adding a "Gel-based" sample

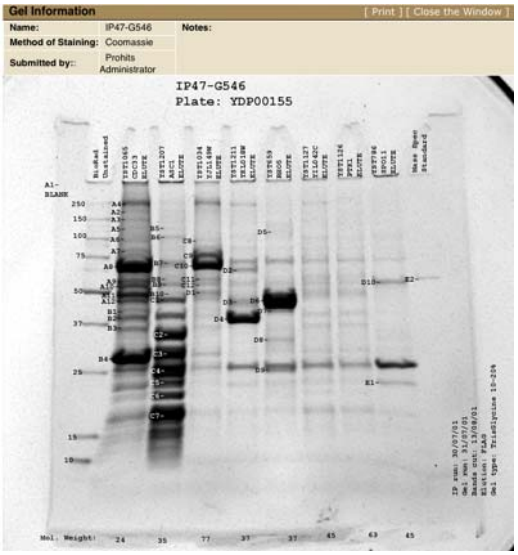
fi Select "Add Gel-based" sample from the left menu, and choose whether you will be starting from an existing gel, or create a new gel.



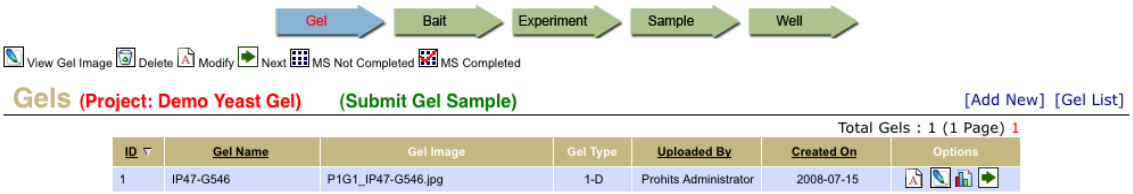
fi To create a new gel, add information required in bold, and upload the image of the gel.



While the image is not mandatory, it is highly recommended to link a well-annotated image of the gel.



After a gel is created, you can see the information via the "Report by Gel" function on the left menu.



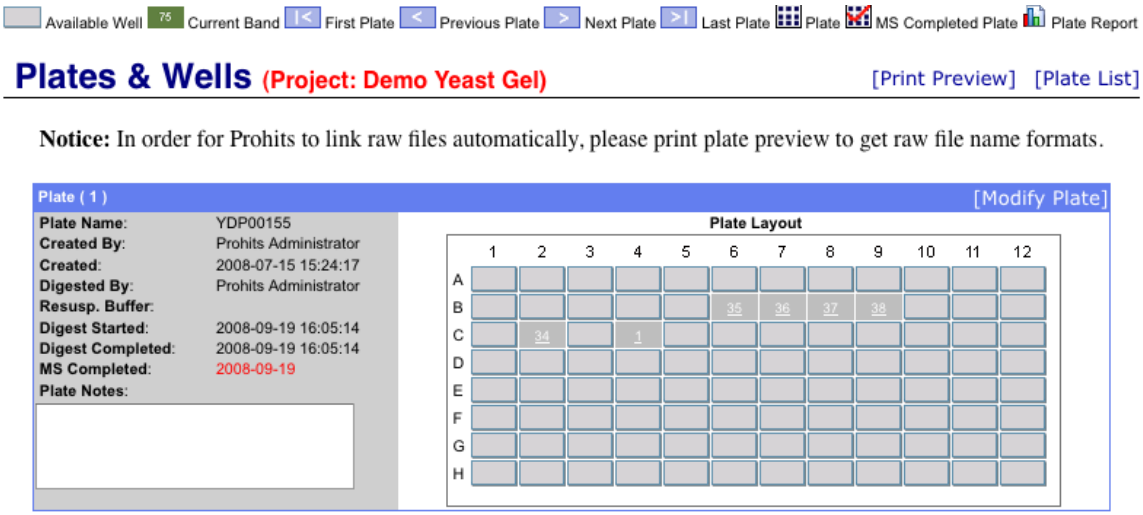
fi Use the green arrows in the "Options" field to enter baits from that gel (as shown in the gel-free sample section).

fi From each bait, define the Experimental Details, as shown in the gel-free section.

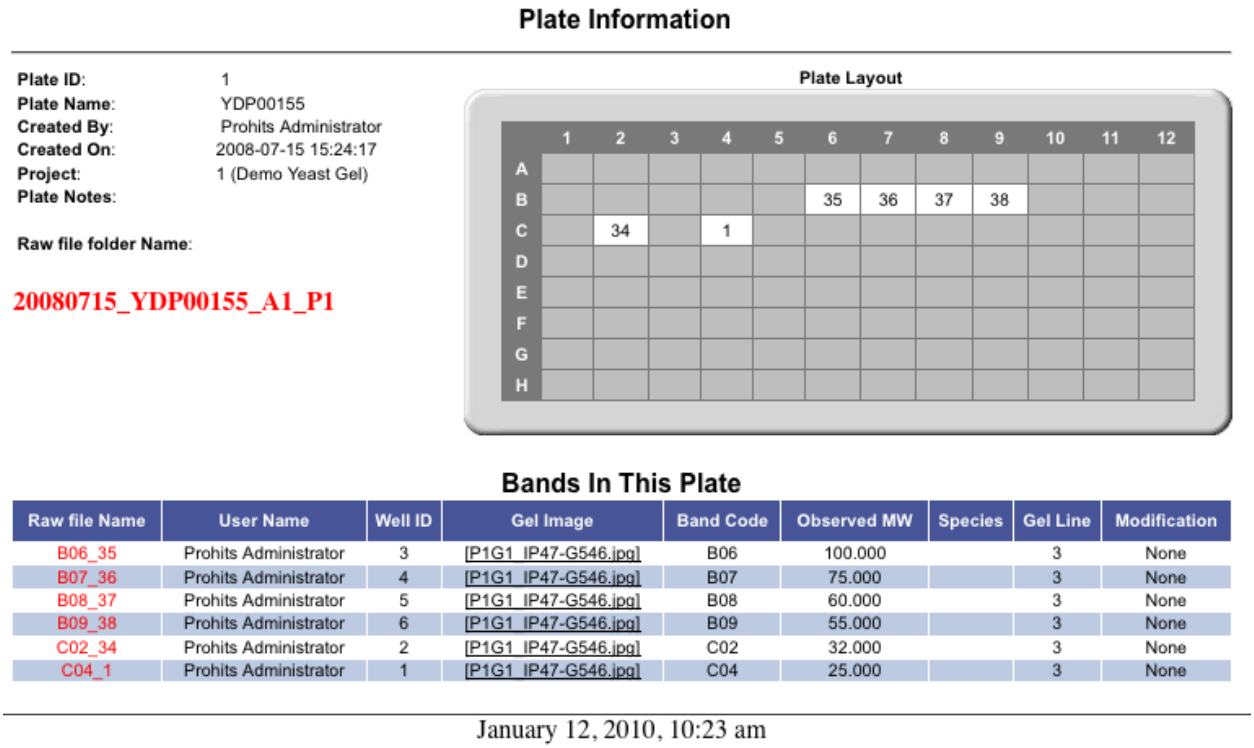
Clicking on the green arrow in the experimental details section will by default prompt you to define a lane on the gel, and guide you through the entry of individual band samples in the autosampler plate that you will use for data acquisition. Simply clicking on a plate well will create an associated sample – you can add the intensity of each band on the stained gel, as well as the approximate molecular weight.

Continue entering all desired bands from the selected lane, or use the navigation options at the top of the page to upload samples from the next lane, return to the list of all lanes, or return to the experimental description.

Opening the "Report by plate" and clicking the plate icon in the "Options" field, allows you to view your plate layout.



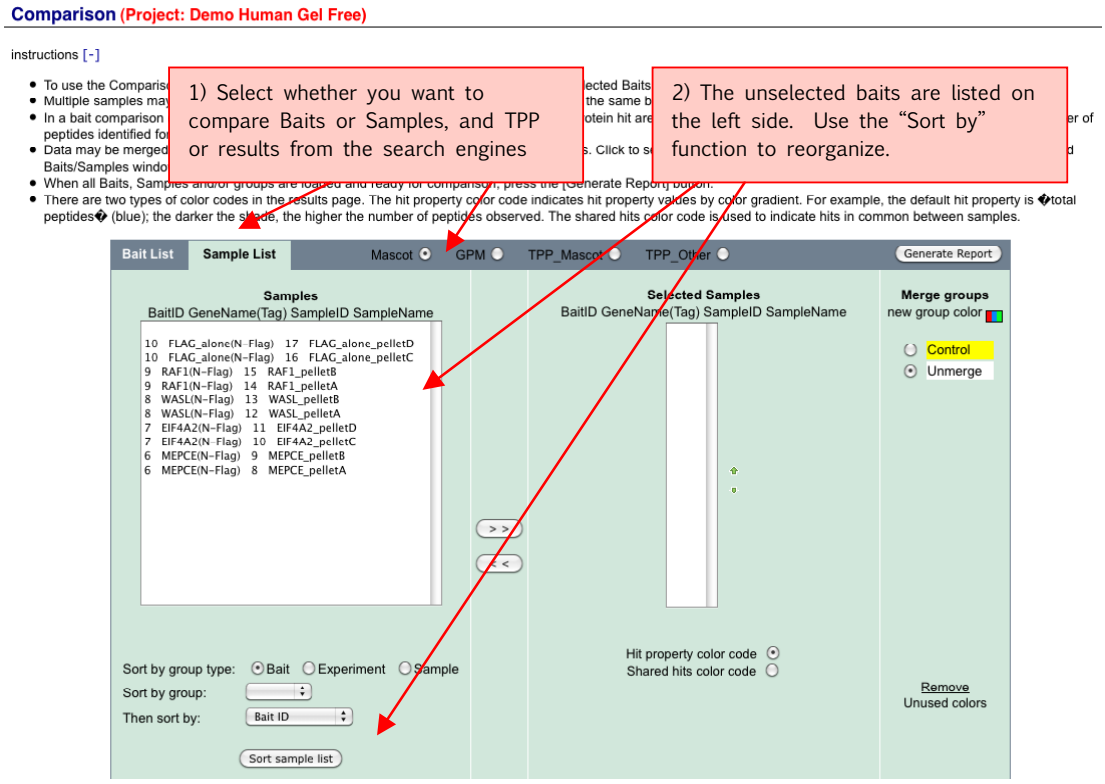
If you wish to use the "Auto-link" option to link your raw files from the Data Management module to the samples in Analyst, select [Print Preview].



When setting up the acquisition on the mass spectrometer, the folder name (here **20080715_YDP00155_A1_P1**) as well as the Raw file names (e.g. **B06_35**) need to match these above.

The Comparison tool

ProHits has a built-in comparison tool that allows you to look at the results of several experiments side-by-side. You can perform comparisons at the bait level or at the sample level, and compare the results from the search engines (e.g. Mascot or X!Tandem) or the TPP. For this demonstration, we will perform a comparison at the sample level, using the Mascot search engine.



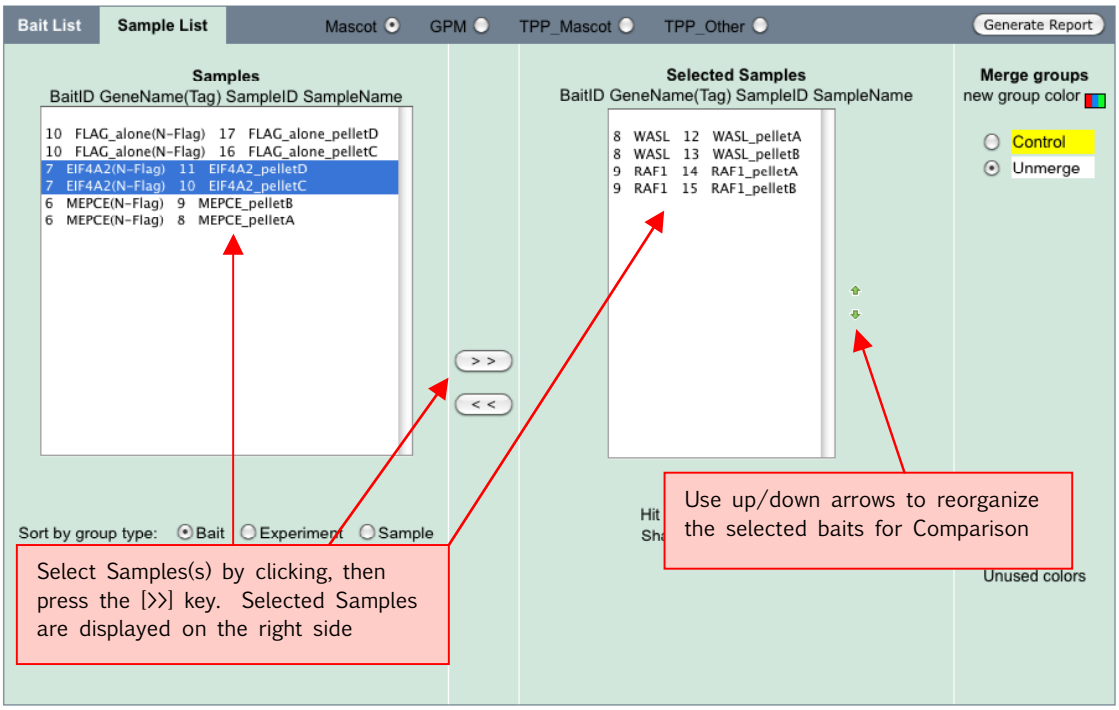
Select the desired baits to be compared

You can sort by Bait ID, Gene name, Protein ID, or by any of the user-defined flags that were used for the project.

Press the >> arrow button to transfer the baits to the "Selected Baits" window

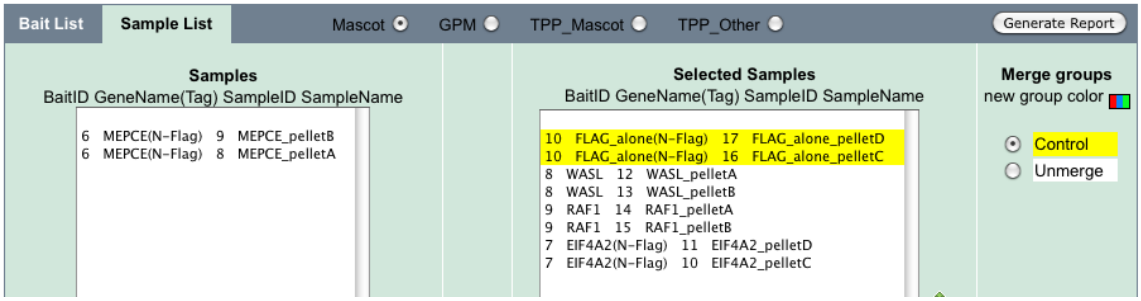
You can transfer files one at the time, or by large groups. The files are added to the list in the order selected. This will also be the order of the columns in the Comparison View.

Use the green up/down arrows on the right hand side to reorganize the sort order. Individual Baits or Groups of Baits can be reorganized.

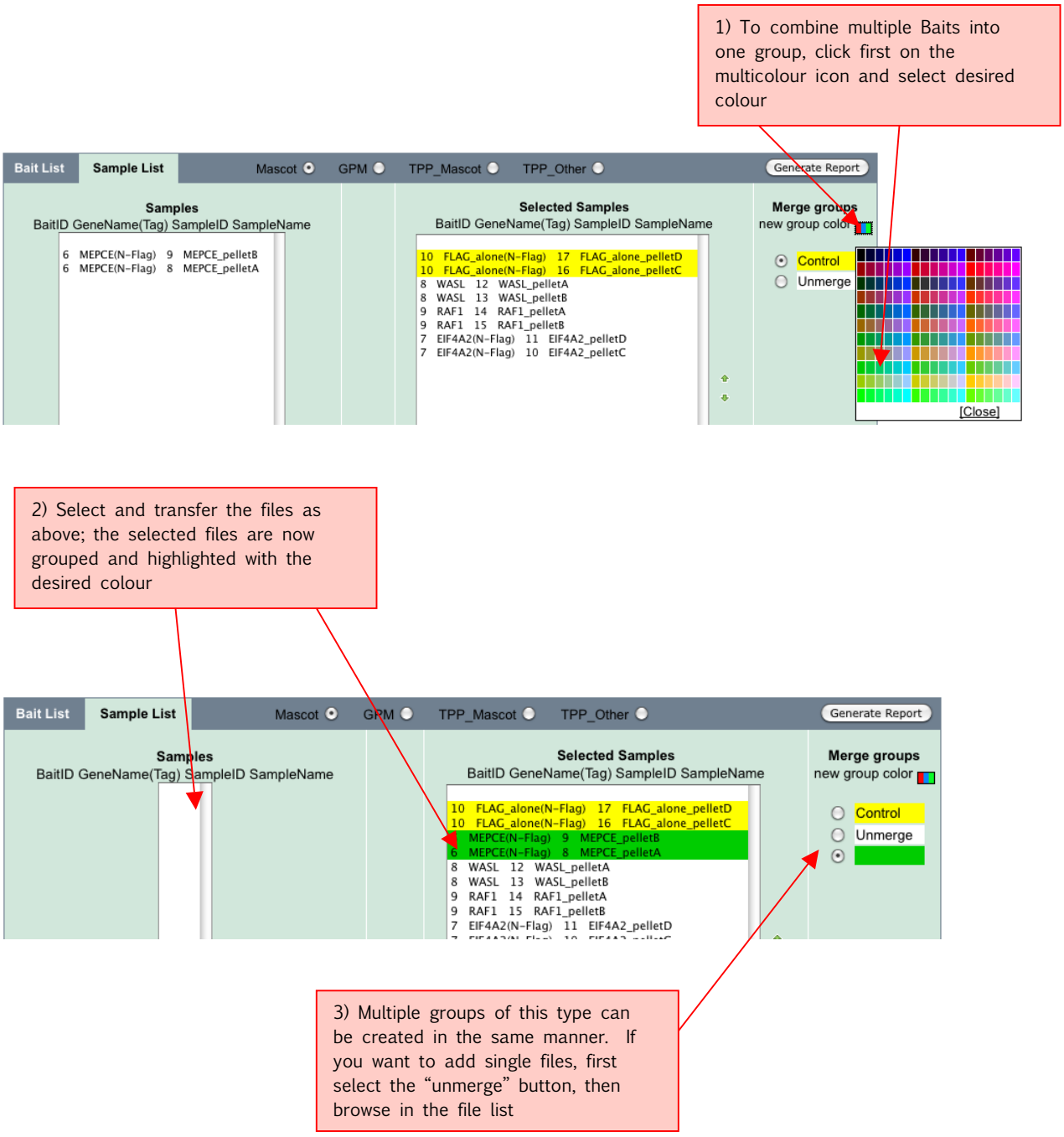


Merging files prior to Comparison

Additional options are available that provide merging options for two or more files. *Please note that the merging is a very simple process that simply reports the best hits for the item but does not do any recalculation. If the selected display option in the report is the Mascot score, the best scoring hit will be listed; if the selected display option is based on spectral counts, the hit with the highest spectral counts will be reported*. The merging function allows you to group two or more control runs (click on the "Control" button before transferring the selected files).



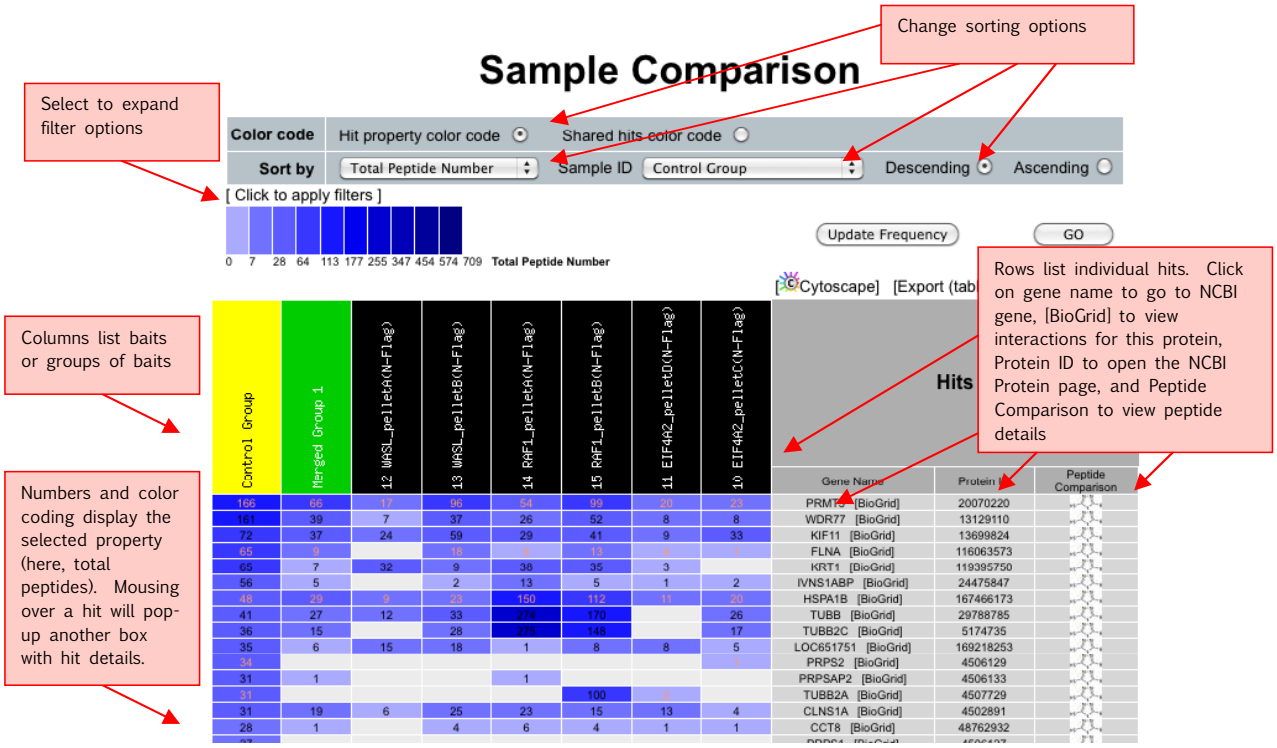
You can group any set of additional files by first clicking on the multicolour icon to select a new group, then transferring the given files to the right side. The listing order will be as follows: The control group will be listed first, followed by all other groups in the order selected by the user, followed by all individual entries in the order selected by the user. Note that within the same group, hits will be combined, and only the maximal value for each of the properties will be reported.



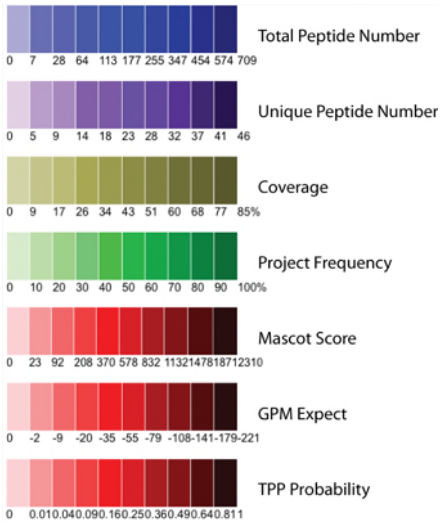
fi When you are done adding all desired baits and/or bait groups, press [Generate Report]
This will open a new window, the Comparison page.

Comparison page

When you open the Bait Comparison page, you will see an unfiltered view of the hits. Each column represents a different sample or bait (or group of samples or baits if the "merge" function was used). The rows represent each of the hits detected across the *n* samples or baits. Clicking on the Gene Name will take you to NCBI Gene; selecting [BioGrid] will open the BioGrid entry for the given protein; clicking on the number in the Protein ID field will bring you to the Entrez Protein page. The last column allows you to compare the peptides identified across the bait purifications.



The default display is with Total Peptide Numbers (spectral counts), and the default sorting option is by descending number of spectra, starting by the left-most bait or group. Note that these sorting options can be modified. In particular, ProHits recalculates and sorts using the following parameters:



In addition to the sorting options, ProHits Comparison allows you to filter your data in a manner similar to the filtering options in the Bait Report page.

fi To access the filtering option, select [Click to apply filters].

An expanded menu allows you to select criteria for removal of proteins from the Comparison list.

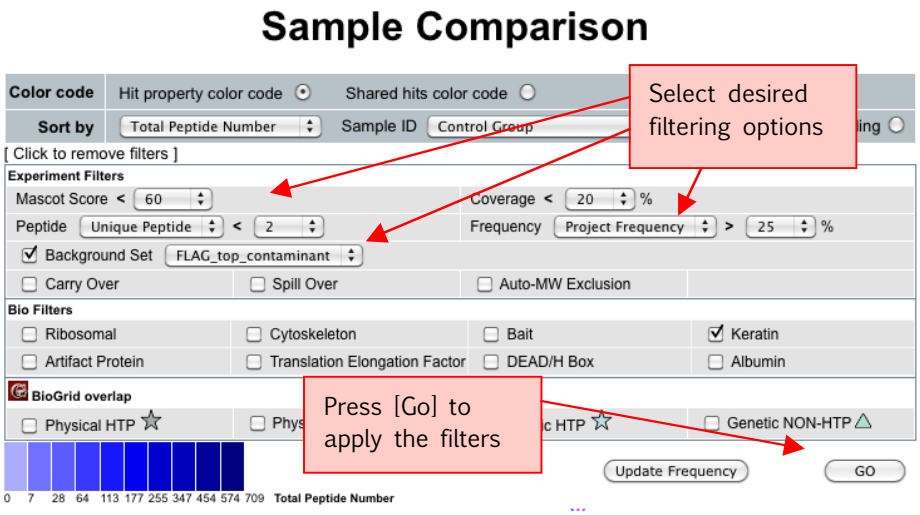
fi Select desired parameters

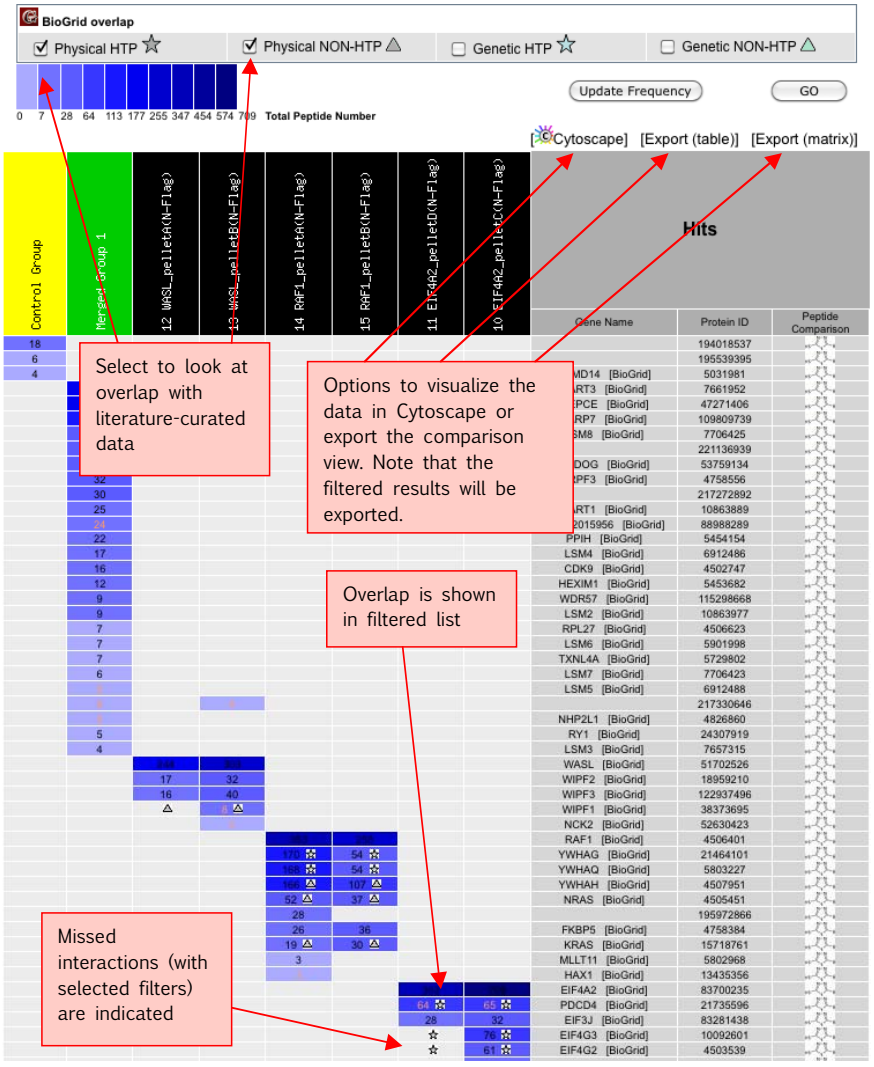
fi Select to highlight the BioGrid overlap if desired

fi To apply filters, press [Go]

This generates a modified list, similar to the process described in the Bait report section. If selected, the overlap with BioGrid is indicated by stars or triangles in the list below.

Note that mousing over any of the entries shown below will pop up a menu box listing the scoring details.

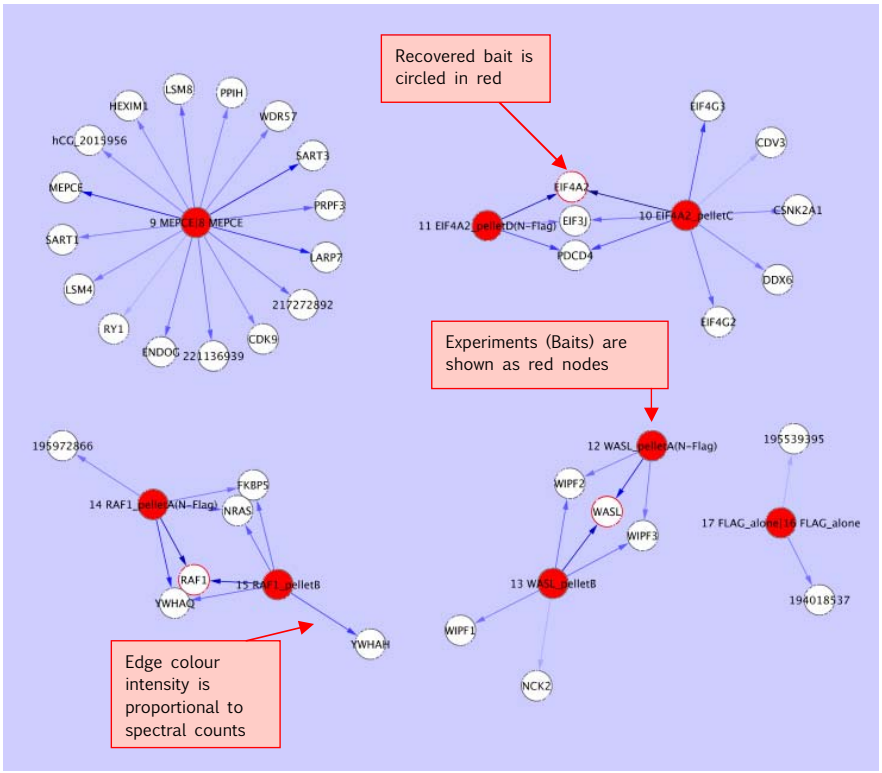




Using Cytoscape directly from ProHits comparison

ProHits allows you to visualize your data using Cytoscape. If using the ProHits filters, the data post-filtering will be displayed (changing the filter will modify the display). If the BioGrid overlap function has been selected, the resulting Cytoscape view will incorporate both your mass spectrometry data, the overlap between your mass spectrometry data and data in BioGrid, and data detected only in BioGrid (including interactions amongst first neighbours of the hits). The colour-coding (see below) allows you to identify the source of the data.

Before you can use the Cytoscape plug-in, you need to have the Runtime Environment (JRE) installed on your local computer (you can use the following URL to test whether your computer has a functional JRE: http:///www.java.com/en/download/help/testum.xml). The first time that you click the "Cytoscape" icon, Cytoscape will be installed on your local computer. Press the [Cytoscape] link immediately above the table to open the current interaction file in Cytoscape. The baits are indicated by red nodes (alongside the unique bait identifier), and the recovery of baits in a purification is indicated by circling the white baits in red. The colour-coding of the arrows is mapped to the spectral counts, as shown above, and all peptide annotation is encoded as an edge attribute. Note that if the "Overlap with BioGrid" function has been selected, interactions specific to your dataset will be still shown in blue, interactions that overlap between your dataset and BioGrid will be shown in green, while BioGrid-only interactions will be displayed in white.



The original image is a circular layout; in the example shown here, this has simply been converted to a spring-embedded layout, with weight on the edge (unique peptide).

Note that all of the standard Cytoscape tools are available.

Other export options

You may also wish to launch Cytoscape (or additional network viewers) from an Excel Table, in which you can add annotation or other mapping options. To do so, use the **[Export (table)]** option, also located at the top of the table. This will create a .csv file that can be opened and modified in Excel. The file will be displayed as a bait>hit list with each subsequent column listing a separate parameter. These lists are easily opened using a stand-alone Cytoscape version.

	A	B	C	D	E	F	G	H	I	J	K
1	Sample ID	Bait Gene Name	Hit Gene Name	Hit Gene ID	Hit Protein ID	Hit Score	Peptide Number	Unique Peptide Number	Coverage	Frequency	Shared Frequency
2	10	EIF4A2_pelletC(N-Flag)	EIF4A2	1974	83700235	1759	709	24	71.5	60	20
3	11	EIF4A2_pelletD(N-Flag)	EIF4A2	1974	83700235	1720	363	25	64.6	60	20
4	14	RAF1_pelletA(N-Flag)	RAF1	5894	4506401	1620	353	20	50.9	20	20
5	13	WASL_pelletB(N-Flag)	WASL	8976	51702526	1320	303	21	60.4	20	20
6	15	RAF1_pelletB(N-Flag)	RAF1	5894	4506401	1560	258	20	53.9	20	20
7	12	WASL_pelletA(N-Flag)	WASL	8976	51702526	1457	244	23	57.2	20	20
8	9	MEPCE_pelletB(N-Flag)	SART3	9733	7661952	1683	242	24	35.8	20	20
9	9	MEPCE_pelletB(N-Flag)	MEPCE	56257	47271406	1480	184	24	53.7	20	20
10	9	MEPCE_pelletB(N-Flag)	LARP7	51574	109809739	1121	176	19	33.5	20	20
11	14	RAF1_pelletA(N-Flag)	YWHAG	7532	21464101	321	170	3	27.1	20	20
12	14	RAF1_pelletA(N-Flag)	YWHAQ	10971	5803227	455	168	4	26.5	20	20
13	14	RAF1_pelletA(N-Flag)	YWHAH	7533	4507951	295	166	3	22.8	20	20
14	8	MEPCE_pelletA(N-Flag)	MEPCE	56257	47271406	1106	152	19	43.8	20	20
15	15	RAF1_pelletB(N-Flag)	YWHAH	7533	4507951	472	107	4	33.3	20	20
16	8	MEPCE_pelletA(N-Flag)	SART3	9733	7661952	1463	105	20	26.9	20	20
17	10	EIF4A2_pelletC(N-Flag)	EIF4G3	8672	10092601	1598	76	25	23.3	20	10

[Export (matrix)] provides a view similar to that displayed in the Comparison page, with the option to export only the parameter currently displayed (e.g. spectral counts), or the option to list all parameters inside each cell. Again, a .csv file that can be opened and modified in Excel will be created.

View only the displayed value (here = total peptide counts):

	A	B	F	G	H
1	Generated date: 2010-January-15		Project Name: Demo Human Gel Free		
2	PID:SC(PT-PU-C%-F%-SF%)		Peptide ID:Score(Total Peptide Number-Unique Peptide Numbe-Cover		
3					
4			17 FLAG_alone_pelletD	8 MEPCE_pelletA(N-Fl	9 MEPCE_pelletB(N-Flag
5					
6	Gene ID	Gene Name	Total Peptide Number	Total Peptide Number	Total Peptide Number
7	194018537		18		
8	195539395		6		
9	10213	PSMD14	4		
10	56257	MEPCE		152	184
11	9733	SART3		105	242
12	51574	LARP7		53	176
13	51691	LSM8		49	35
14	217272892			26	30
15	9129	PRPF3		26	32
16	221136939			25	40
17	25804	LSM4		17	11
18	1025	CDK9		16	
19	2021	ENDOG		15	34
20	10614	HEXIM1		12	
21	57819	LSM2		8	9

View all parameters:

	A	B	D	E	F	G	H
1	Generated date: 2010-January-15		Project Name: Demo Human Gel Free				
2	PID:SC(PT-PU-C%-F%-SF%)		Peptide ID:Score(Total Peptide Number-Unique Peptide Numbe-Coverage-Frequency-Sub Frequency)				
3							
4				16 FLAG_alone_pelletC(N-Flag)	17 FLAG_alone_pelletD(N-Flag)	8 MEPCE_pelletA(N-Flag)	9 MEPCE_pelletB(N-Flag)
5							
6	Gene ID	Gene Name	Protein ID	Total Peptide Number	Total Peptide Number	Total Peptide Number	Total Peptide Number
7	194018537		194018537		18		
8	195539395		195539395		6		
9	10213	PSMD14	5031981		4		
10	56257	MEPCE	47271406			152	184
11	9733	SART3	7661952			105	242
12	51574	LARP7	109809739			53	176
13	51691	LSM8	7706425			49	35
14	217272892		217272892			26	30
15	9129	PRPF3	4758556			26	32
16	221136939		221136939			25	40
17	25804	LSM4	6912486			17	11
18	1025	CDK9	4502747			16	
19	2021	ENDOG	53759134			15	34
20	10614	HEXIM1	5453682			12	
21	57819	LSM2	10863977			8	9

Zoom of the details inside each cell:

PID:SC(PT-PU-C%-F%-SF%)

56790935:500(16-10-20.70-11.88-75)

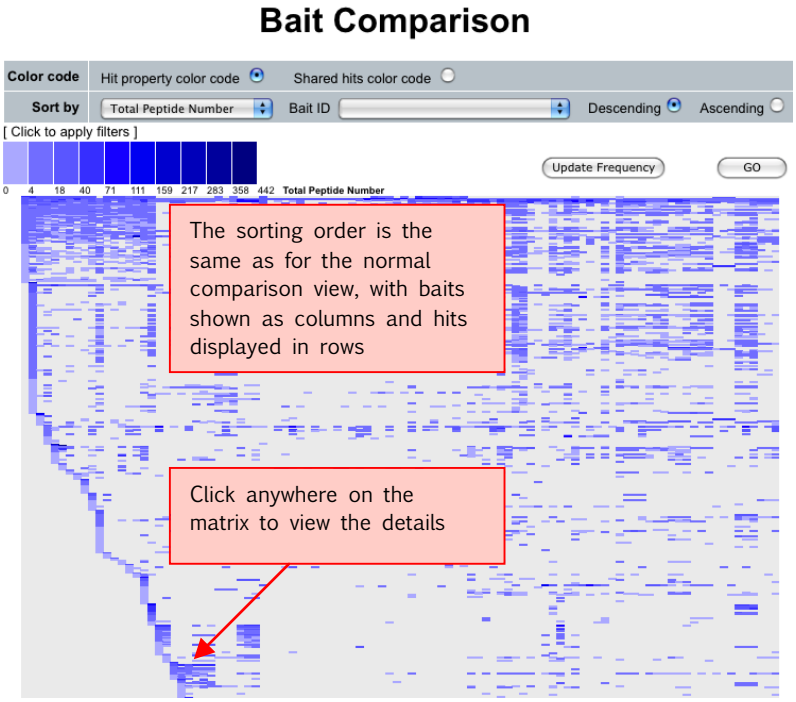
Legend:
PID: Protein ID (NCBI Entrez Protein)
SC: Mascot Score
PT: Total number of peptides
PU: Number of Unique peptides
C%: Percentage of the protein sequenced
F%: Frequency of occurrence of the protein in the entire dataset
SF%: Frequency of occurrence of the protein amongst compared baits/samples

Comparing larger numbers of baits

ProHits also allows you to visualize larger numbers of experiments.

fi Select the baits or samples to be compared and press [Generate Report]

A heat-map view of the data will be generated.



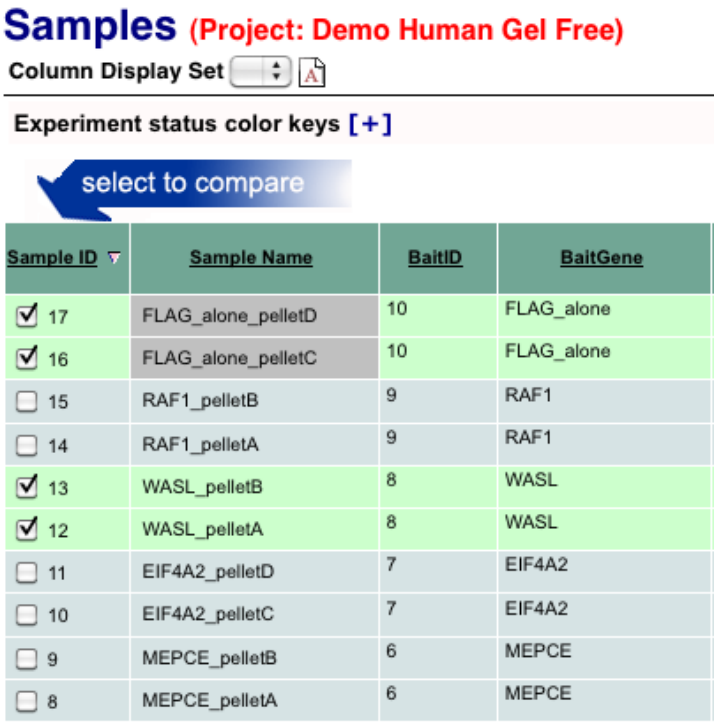
fi Click anywhere on the map to expand and view names and other details

Note, however, that due to file size, the [Cytoscape] option is not available with this heat map view. The [Export(table)] option is still available, however, and can allow you to upload data into a stand-alone Cytoscape session (the [Export(matrix)] function is also available). Note that due to large file sizes, these export functions may run slowly.

For additional export functionalities, you can go back to the main Analyst module, and select the "Export Hits" option from the left-hand menu.

Automatically adding baits for comparison from the baits or sample report list pages

ProHits allows you to select baits or samples to be added to the comparison page while working on other pages. To use this option, simply click the box located to the left side of each sample in the sample list or by the bait in the bait list.



Sample ID	Sample Name	BaitID	BaitGene
17	FLAG_alone_pelletD	10	FLAG_alone
16	FLAG_alone_pelletC	10	FLAG_alone
15	RAF1_pelletB	9	RAF1
14	RAF1_pelletA	9	RAF1
13	WASL_pelletB	8	WASL
12	WASL_pelletA	8	WASL
11	EIF4A2_pelletD	7	EIF4A2
10	EIF4A2_pelletC	7	EIF4A2
9	MEPCE_pelletB	6	MEPCE
8	MEPCE_pelletA	6	MEPCE

The selected sample (or baits) will be automatically added to the "Selected Samples" and "Selected Baits" pages of the Comparison view. Note that if a bait is selected, all samples corresponding to this bait will automatically be added to the comparison view.

You can keep browsing and adding baits or samples for Comparison as you go. These will stay selected for the duration of your session, or until you manually remove them from the Comparison page.

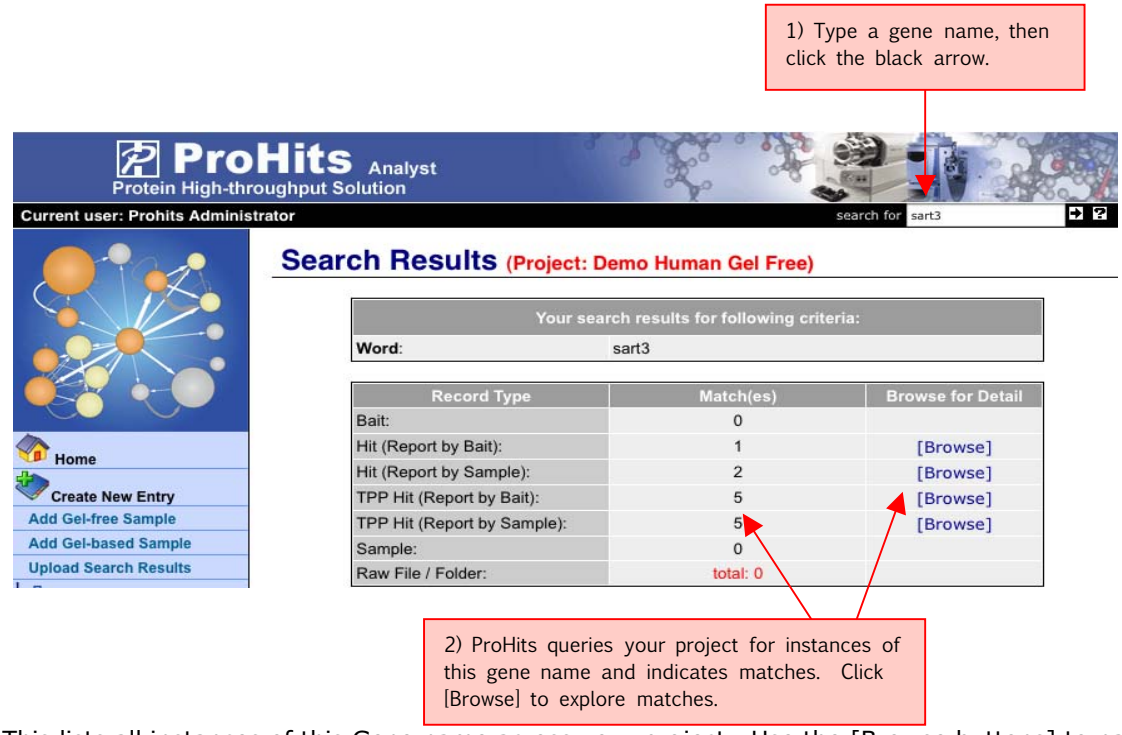
Search options

ProHits Analyst allows you to perform simple searches (for individual Gene Names) or Advanced searches (for multiple gene names or keywords in the protein description field or controlled vocabulary). Here, we will briefly review these options:

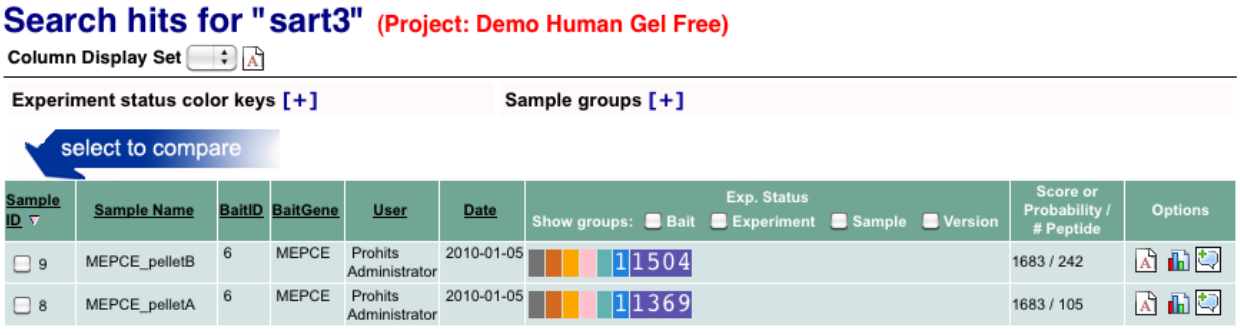
Simple Search (Gene name)

ProHits has a simple search function that is located at the upper corner of the Analyst module main page.

fi Enter an official Gene Name, then press the right pointing arrow.



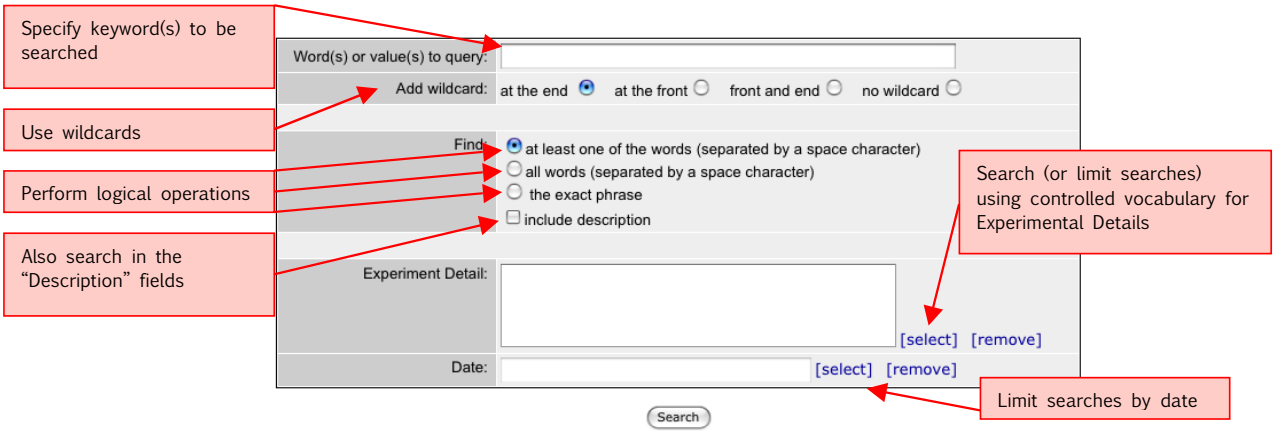
This lists all instances of this Gene name across your project. Use the [Browse buttons] to navigate through the data. Below, we have expanded the "Hit (Report by Sample)" option. The gene SART3 was identified in both of the MEPCE biological replicates. Note that the column "Score of Probability/ # Peptides" refers to the score from the search engines (or TPP) and the total number of peptides identified for SART3 in the MEPCE runs.



Advanced Search

The Advanced Search function can be accessed from the menu bar. This function allows you to search for keywords (or combinations of keywords) and retrieve entries across the following categories: **Baits**, **Hits**, **Samples**, **Gels**, **Raw Files** and **Auto Search**.

In the simplest sense, you can use the Advanced search in a manner similar to the Simple search, i.e. to retrieve entries associated with a gene name. You can use "wildcards", either at the front, at the end, or both at the front and end of your query. Note that using wildcards (especially at the front) decreases search speed.

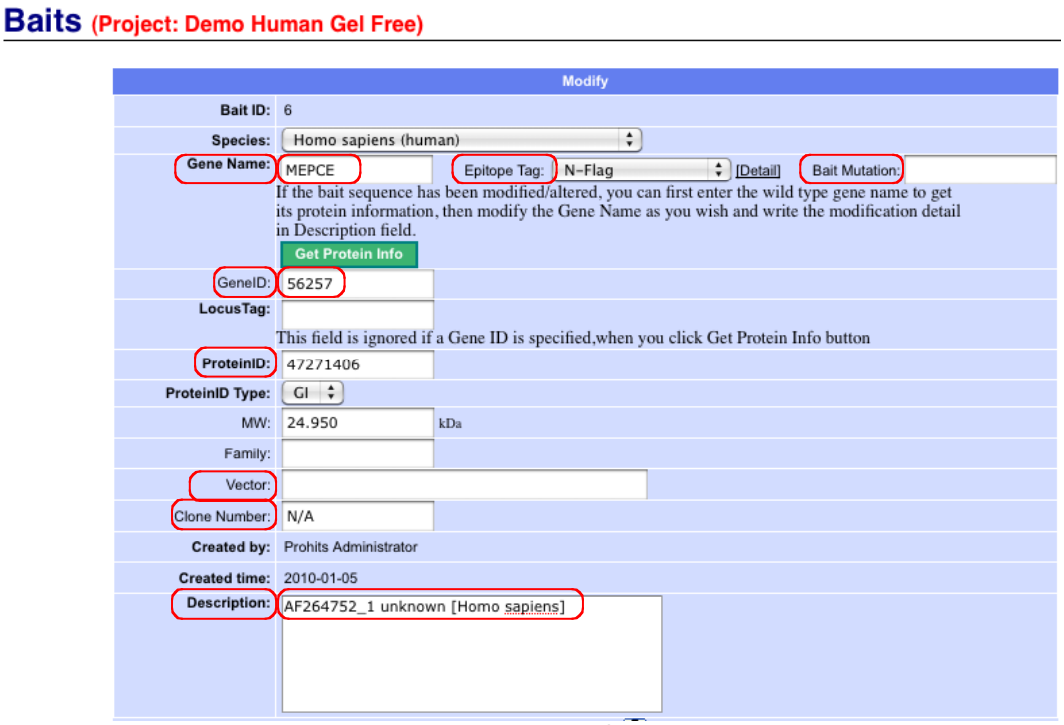


This will return a list of results that you can then explore further by selecting the [Browse] option for each of the categories, as for the simple search.

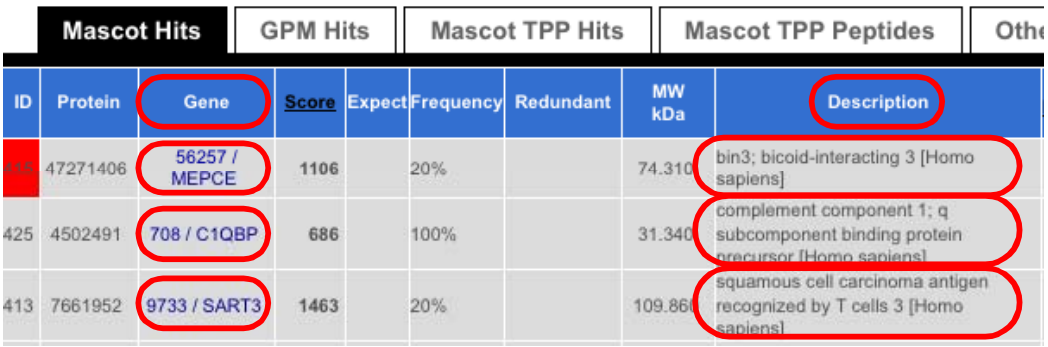
Other keywords that can be searched:

In addition to the Gene Name, different keywords can be searched. The fields searched depend upon the category, as defined below:

1) **Bait** (the keywords were detected in the entry for a bait – fields searched are "Gene Name", "Gene ID", "Locus Tag", "Protein ID", "Epitope Tag", "Bait Mutation", "Clone Number", "Vector", with optionally, Bait "Description"). The searched fields are indicated by red ovals below:



2) **Hits** (the keywords were detected in the hits list – field searched is "Gene" Name, with, optionally, Protein "Description"). *You can similarly see the hits across TPP results.* The searched fields are indicated by red ovals below.



3) **Sample** (the keywords were detected in the user-defined "Sample Name")
4) **Gel** (the keywords were detected in the fields "Gene Name", "Gene Image", and "Lane Code")
5) **Raw files** (the keywords were detected in "File Name" or "Folder Name"). This brings you to the "Data management" module, and lists the folders / files bearing the selected keywords.
6) **Auto Search** (the keywords were detected in "Search Task Name"). This brings you to the "Data management" module, and lists the search tasks bearing the selected keywords.

Searching Bait/Protein Description: You can search for a keyword inside the Description field (e.g. "squamous" in the example above), by allowing wildcards on both sides. In other words, the entire field is captured (not individual words), and any partial field (e.g. "squamous" or "carcinoma") must be preceded and/or followed by wildcards. Note again that such searches may be very slow.

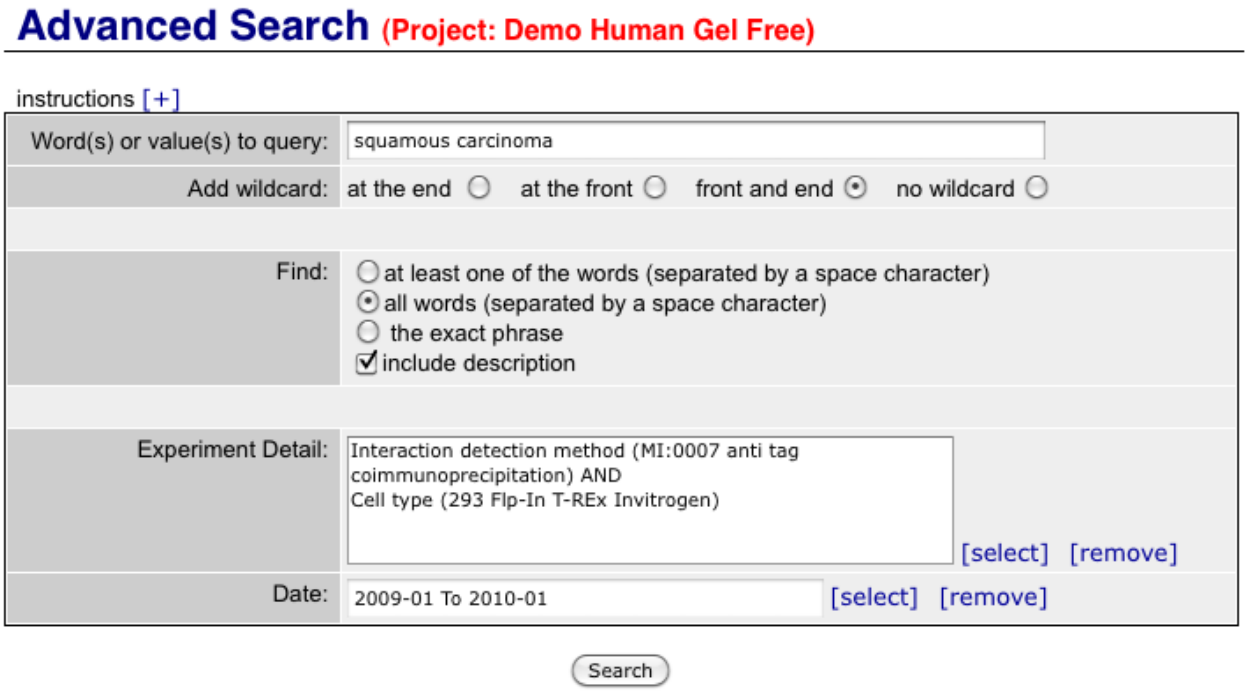
Searching in Experimental Details (controlled vocabularies): The search function also allows you to search (or limit your searches) based on selected controlled vocabulary. Simply press [Select] (bottom right corner of the Experimental Detail section). This will take you to the Experimental Details/controlled vocabulary section where you can select categories/values to be passed to the Advanced search page.

Restricting searches by date: You can restrict search results by date. Simply press the [select] button in the Date field to open a drop-menu.

Using logical operations: You can combine several keywords (simply separate them by spaces), to search for "at least one of the words", "all words" (in any order), or "the exact phrase" within a field, such as "Description". Note that the "all words" and "exact phrase" operations only apply within a field. Alternatively, you can use the "at least one of the words" option to search for different keywords even across different fields. This will generate a list of results that will be the union of the separate lists.

Hits searches returning too many results: Note that there is a limit of 3000 to search results. Try narrowing down your search parameters and try again.

Example: Searching for squamous AND carcinoma in 293 Flp-In T-REx cells and in anti tag coimmunoprecipitation; date restricted to January 2009 – January 2010.



Uploading search results

The Analyst module allows you to import search results from the TransProteomics Pipeline (TPP), Mascot or GPM/X!Tandem. This function is very useful for laboratories that are not interested in the Data Management module of ProHits, e.g. if they are using a third party analysis solution. All that is needed for this section are the search results files or both TPP ProteinProphet and TPP PeptideProphet XML files.

fi Select the [Upload Search Results] link on the left hand side of the Analyst module.

This opens up the list of all the baits that you have created in the Analyst module.

Upload Search Results (Project: Demo Human Gel Free)

Total Bands : 11 (1 Page) 1

Bait ID	Bait Gene	Experiment Exp / Gel / Lane (LaneNum)	Sample ID	Sample Name	Uploaded By	Uploaded date	Uploaded File	Options
20	COPS5(N-Flag,)	COPS5 gel free	39	COPS5				
10	FLAG_alone(N-Flag,)	FLAG_alone_pelletC gel free	16	FLAG_alone_pelletC				
		FLAG_alone_pelletD gel free	17	FLAG_alone_pelletD				
9	RAF1(N-Flag,)	RAF1_pelletA gel free	14	RAF1_pelletA				
		RAF1_pelletB gel free	15	RAF1_pelletB				
8	WASL(N-Flag,)	WASL_pelletA gel free	12	WASL_pelletA				
		WASL_pelletB gel free	13	WASL_pelletB				
7	EIF4A2(N-Flag,)	EIF4A2_pelletC gel free	10	EIF4A2_pelletC				
		EIF4A2_pelletD gel free	11	EIF4A2_pelletD				
6	MEPCE(N-Flag,)	MEPCE_pelletA gel free	8	MEPCE_pelletA				
		MEPCE_pelletB gel free	9	MEPCE_pelletB				

Total Bands : 11 (1 Page) 1

fi Select the upload option  **at the end of the desired sample**

This pops up a new page:

Upload Search Results

Bait Information (20) [GI][Gene][BioGrid]			
Bait Gene ID	10987	**Bait Gene Name**	COPS5 (N-Flag)
Bait Locus Tag		**Bait MW (kDa)**	37.560
Bait Clone	N/A	**Bait Description**	Jun activation domain binding protein [Homo sapiens]

Experiment Information

Experiment ID	Name/Batch Name	Exp. Detail	Exp. Status:	Inputed by	Date
32	COPS5	Interaction detection method : MI:0007 anti tag coimmunoprecipitation; Cell type : 293 Flp-In T-REx Invitrogen		Prohits Administrator	2010-01-15

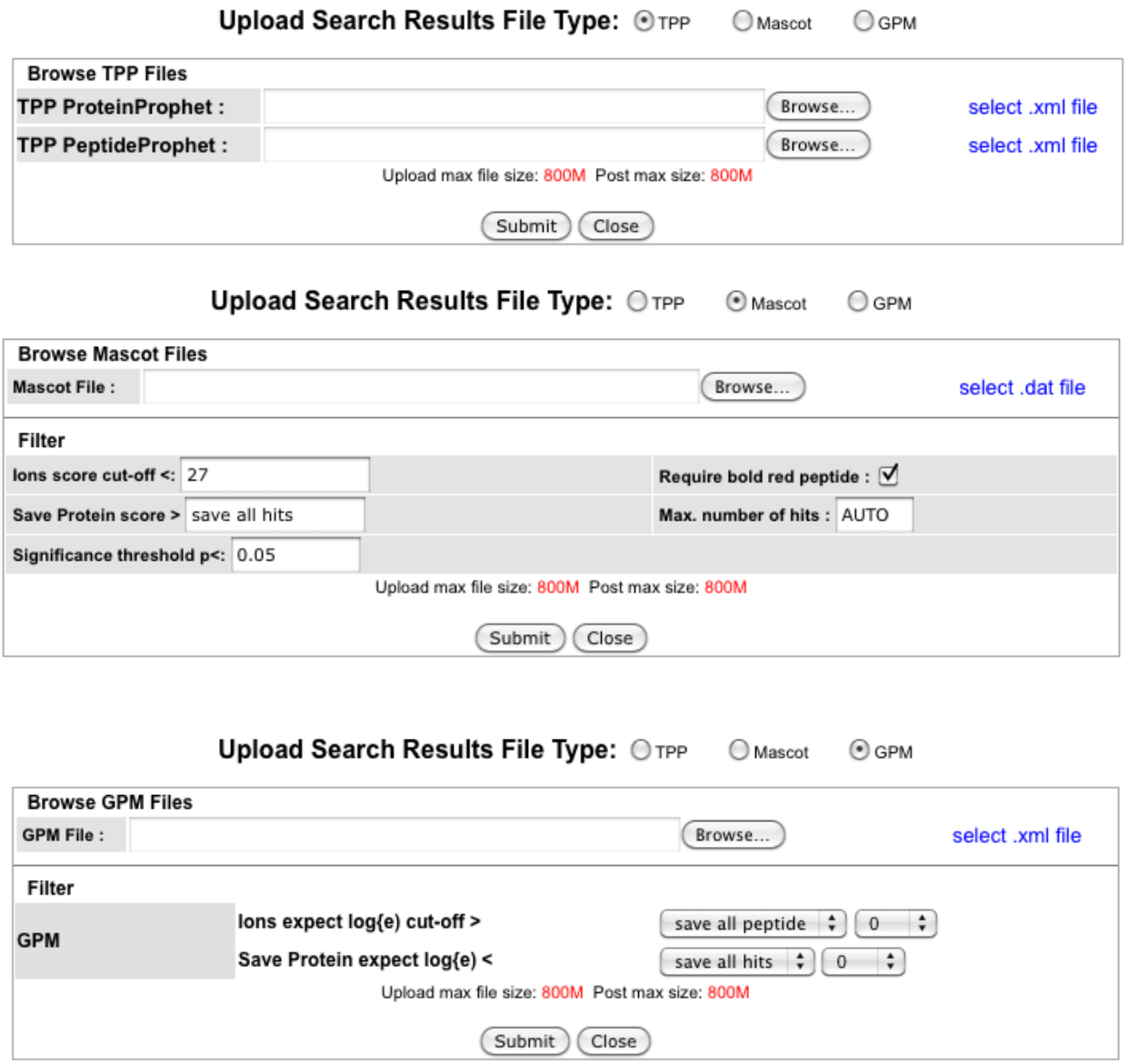
Sample Information

Sample ID	Sample Code	Submitted by
39	COPS5	Prohits Administrator

Upload Search Results File Type: ○ TPP ○ Mascot ○ GPM

fi Select the type of search results files you wish to upload (TPP, Mascot, GPM/X!Tandem), and Browse your local computer for the files in the right format.

fi Press [Submit] to upload search results.



Manage Protocols and Lists

Five types of Protocols and Lists pages are available in ProHits. With the exception of the "Epitope Tag Lists", that are applied to all projects on the local ProHits server, the other protocols and lists are only applicable to the current project. Lists and Protocols defined for a given project may be imported into a different project, so long as the user has access to both projects, and permission to modify individual lists and/or protocols. Access to individual pages of the "Manage Protocols and Lists" of the Analyst module is restricted via page permissions set in the admin office module. We suggest limiting the number of users having access to these management tools.

Here, we will briefly review the function of the different protocols and lists, then show a few examples for each category of protocol and/or list.

Text-based protocols – pages 42-43

Text-based protocols provide details on the experimental procedures. We have separated the protocols into four modules: Biological Material (i.e. what type of cells, expression system, growth conditions, etc.), Affinity Purification (from cell lysis to elution), Peptide Preparation (including separation at the protein/peptide level after elution), and LC-MS conditions. For our internal use, we attempt in providing very detailed protocols that could be used for publication with only minor modifications.

Experimental Editor – pages 44-45

The Experimental Editor allows you to create and manage the list of controlled vocabularies to be used within the Experimental Details page, in conjunction with the text-based protocols and additional notes. For our internal use, we attempt to capture information that would allow PSI MI 2.5 compliance, as well as other relevant information that would allow us to structure our data. Note that the terms entered in this section are searchable in the "Advanced Search" function.

Background Lists – pages 46-48

This function allows you to define and manage one or more lists of contaminants and/or background proteins associated with a given project. For example, you could maintain individual lists of the proteins found to associate non-specifically with different affinity matrices. The proteins on a given "Background" list can be subtracted from the list of identified proteins, in Individual Report, Comparison, or Export views.

Group Lists – pages 49-51

This function allows you to further organize and/or mark certain baits, experiments or samples by adding a colour-coded and user-defined icon that will appear in the Status bar of the Report by Bait or Report by Sample view. Useful Sample level group could include comments about the quality of the data, while Experiment level group would refer to some property of the experimental prep (e.g. phospho-enrichment), and a Bait level group could be the type of tag used. Additionally, ProHits allows you to mark (at the Sample level), samples that are to be included in publication (and/or to be exported to a third party).

Epitope Tag Lists – page 52

This is the only list that applies to the entire local ProHits database. The objects in this list are available on the Bait entry page, and define the tag (if applicable) used for tagging of the bait. N or C refer to the position of the tag relative to the bait. When available, the epitope tags have been mapped back to the standard vocabularies from the Molecular Interaction PSI MI 2.5; an automated link to the Ontology Lookup Service (OLS). We strongly suggest using this service to enter the PSI MI 2.5 terms when entering new tags.

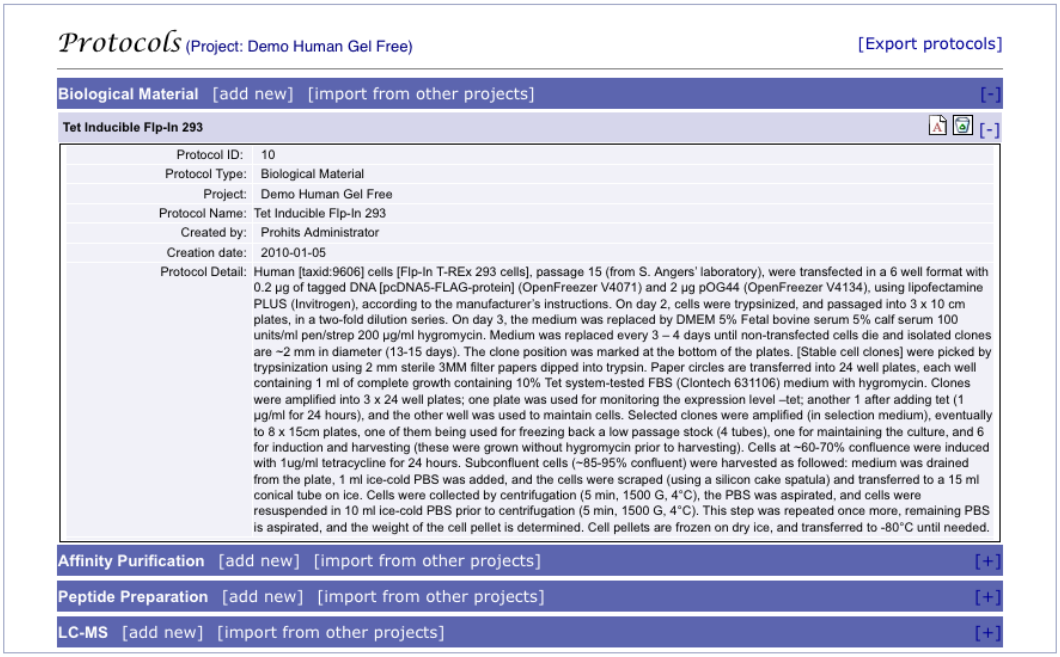
Text-based protocols

fi Select the "Text-based Protocols" entry from the Manage Protocols and Lists

fi Click [add new], and paste or type your protocol. Then press [Save].

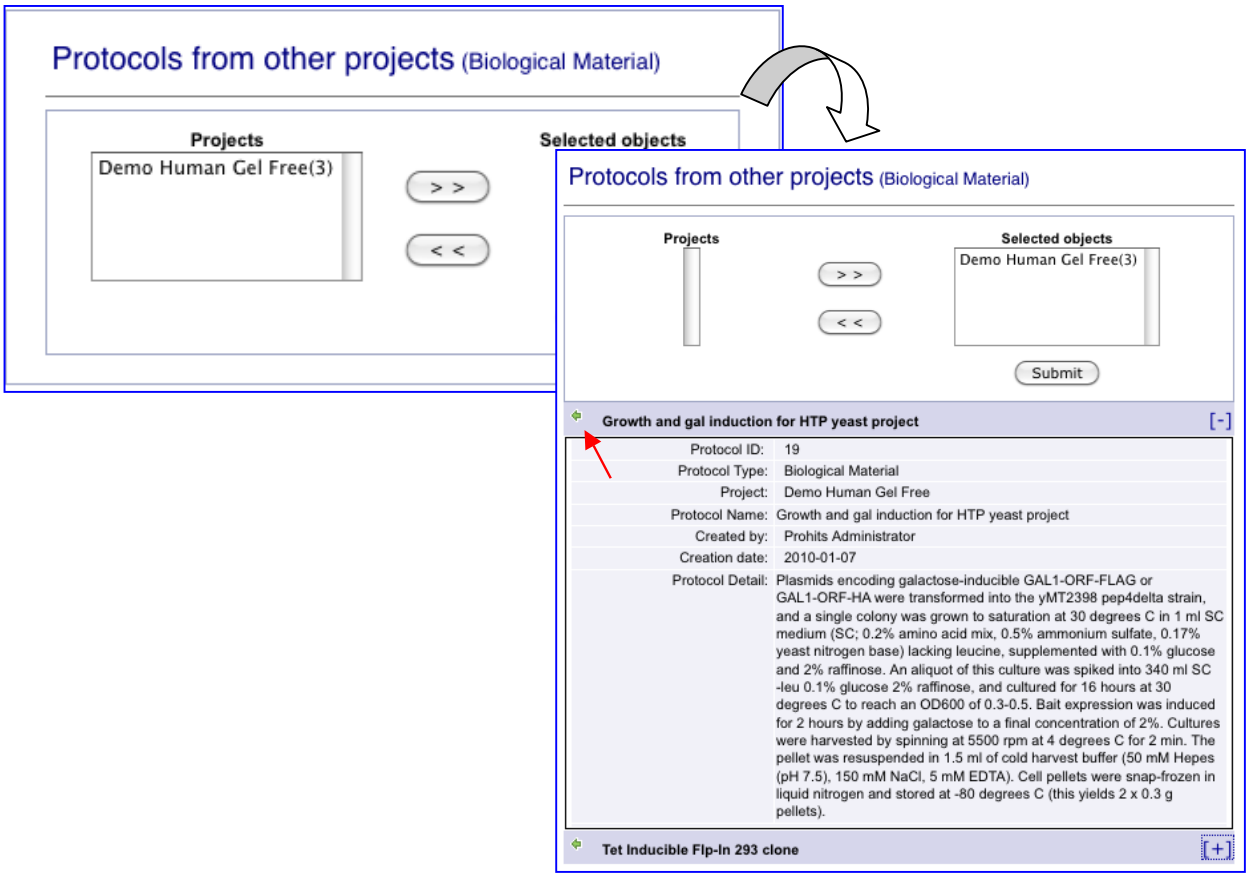
Note 1: because the protocols are displayed as html and exported as a CSV or TSV file, certain characters and symbols will not display properly, and should be spelled out. Examples are μ (u or micro), ° (degree), and ` (apostrophe).

Note 2: to each protocol is assigned a unique identifier. The protocol can be modified or even deleted as long as it has not been used. Once in use, modifications are no longer allowed and a new protocol (that will be assigned a different protocol number) will need to be created.



Continue entering protocols as above. Alternatively, if a protocol of interest already exists in another project to which you have access, you can import it directly from that project.

ﬁ Click [import from other projects], select desired project by clicking the >> button and pressing [Submit].



ﬁ Click the green arrow to transfer the protocol from the source project to the destination project, modify if needed, and press [Save].

You can export protocols linked to a project to a CSV file that can be opened in Excel or similar programs. The "Detail" column contains the full text of the protocol.

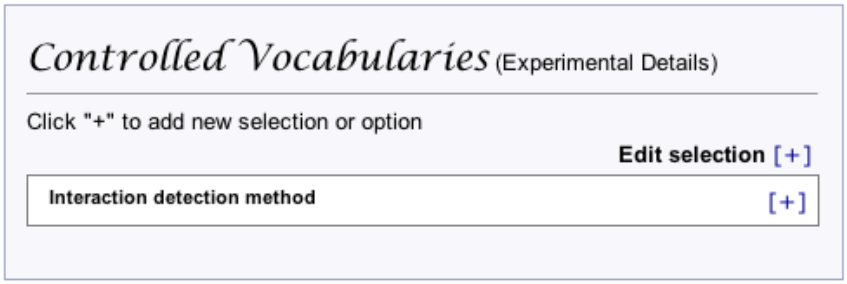
Sheets Charts SmartA

	A	B	C	D	E	F	G	H
	ID	Name	Type	Project Name	Detail	Creation Da	Creater	
	21	Trypsin digestion on magnetic beads	Peptide Preparation	Demo Human Gel Free	Proteins immobilize	1/7/10	Prohits Administrator	
	17	In-solution digest of IP samples	Peptide Preparation	Demo Human Gel Free	Trypsin (1 ug Sigma	1/7/10	Prohits Administrator	
	19	Growth and gal induction for HTP yeast project	Biological Material	Demo Human Gel Free	Plasmids encoding g	1/7/10	Prohits Administrator	
	15	Tet Inducible Flp-In 293 clone	Biological Material	Demo Human Gel Free	Human [taxid:9606	1/7/10	Prohits Administrator	

Experimental Editor

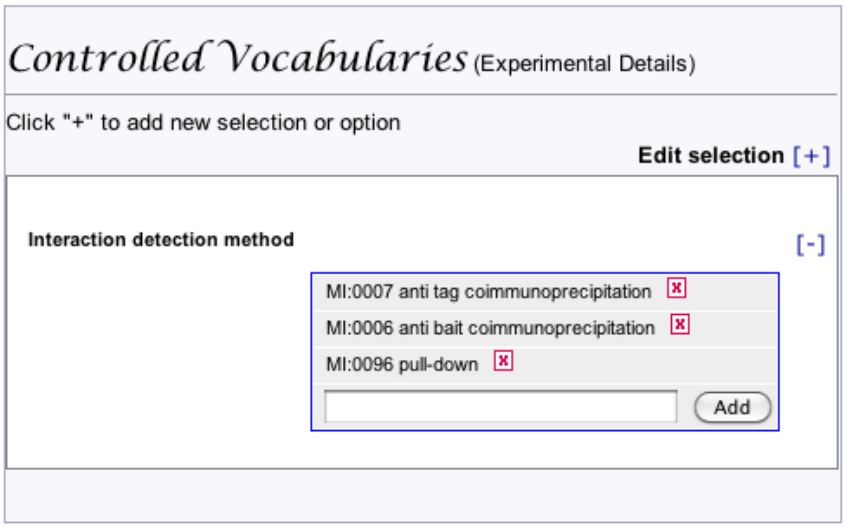
fi Select the "Experimental Editor" entry from the Manage Protocols and Lists

You will see a list of the categories already defined for your project.



fi To view the values already entered under the "interaction detection method" category, click on the [+] button to expand this category.

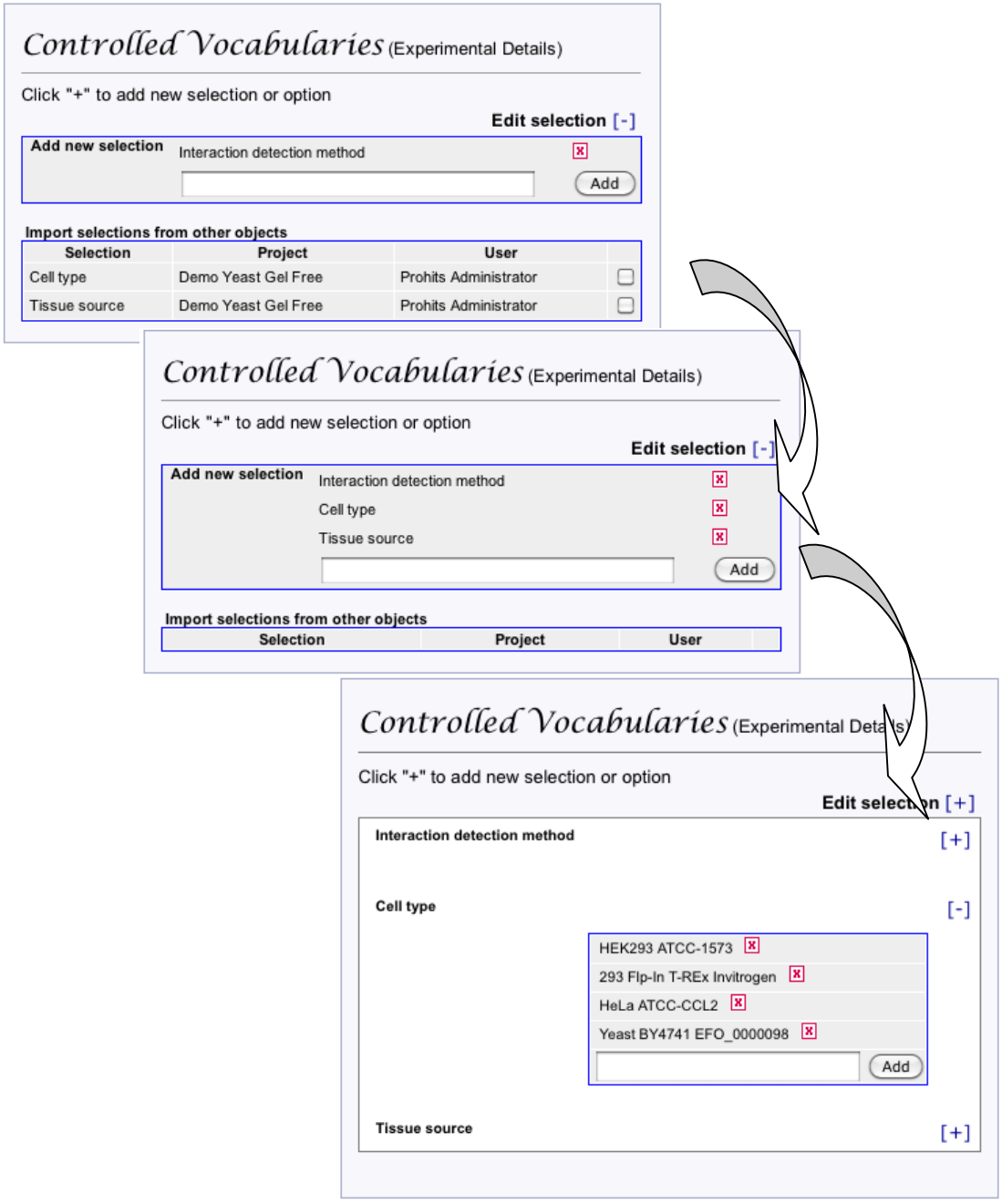
You can add additional values by typing their description and pressing [Add]. Values that are not yet linked to an entry are followed by a red **X**. Pressing **X** deletes the entry. Note that for this category, we have used PSI MI 2.5 terms, to facilitate later deposition in interaction databases.



fi To define new categories, press the [+] button next to "Edit selection".

This allows you to enter a new category.

To import a category from another project to which you have access, simply click the checkbox associated to the category under the Edit selection option to transfer the category (and associated values) to current project.

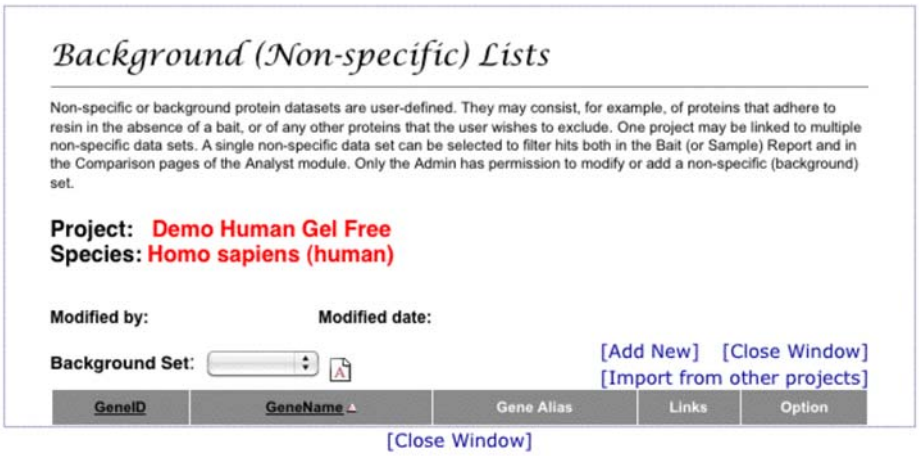


Background Lists

In addition to the Bio Filters and Experimental Filters defined in the Admin Office module, ProHits allows you to define additional filters to remove non-specific (or background) proteins. These filters are project-specific and created within a bait (or sample) report page in the Analyst module. Several different filters can be associated with the same project (e.g. corresponding to different workflows used in the project). Creation of these filters requires administrator-level privileges. The filters can be created by adding proteins manually (one-by-one) to an existing list of contaminants. The filters can also be generated by uploading a list (or table) of hits identified in control run(s), in which case the mapping only requires the Entrez Gene ID field. You can also add multiple proteins at once from any other pre-existing list (e.g. in Excel). The mapping is via the NCBI Entrez Gene ID.

fi Select the "Background Lists" entry from the Manage Protocols and Lists.

fi From the entry page, click on the (modify) icon to upload a list of contaminant proteins.



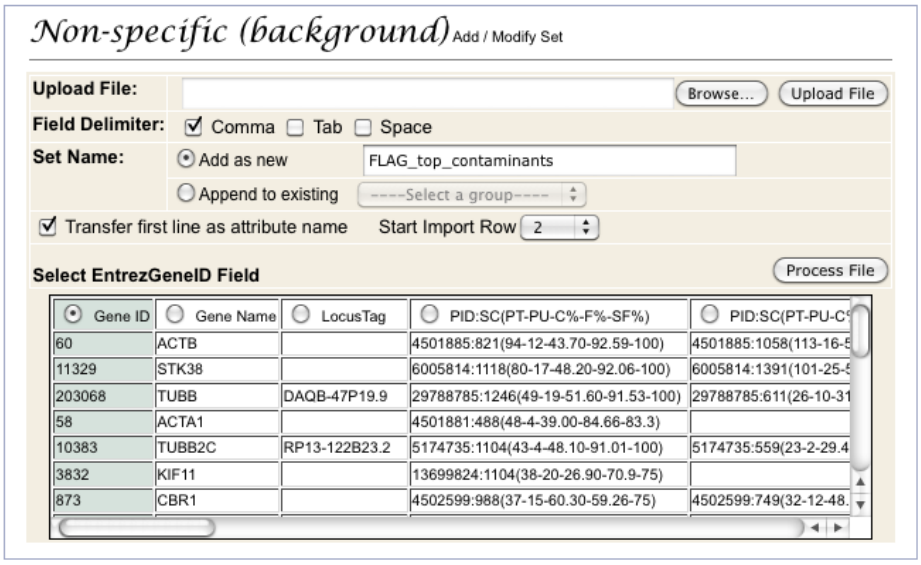
Note that an efficient method to generate a non-specific filter set utilizes the ProHits comparison tool. First, select multiple control runs and merge them into a single "Control" group. This will open up a Comparison page with a single column called "Control Group" displayed in yellow. As before, the maximal value for the parameter visualized is displayed (e.g. spectral count). Apply filters (e.g. number of unique peptides, protein coverage, etc.) desired, and select [Export(table)] to export a comma-delimited file (*.csv). Save this file on your hard drive, and go to any Bait report page. (Note that any Excel or text file that lists the NCBI Gene ID may also be used).

fi Browse the file to be uploaded, select delimiter, and press [upload file].

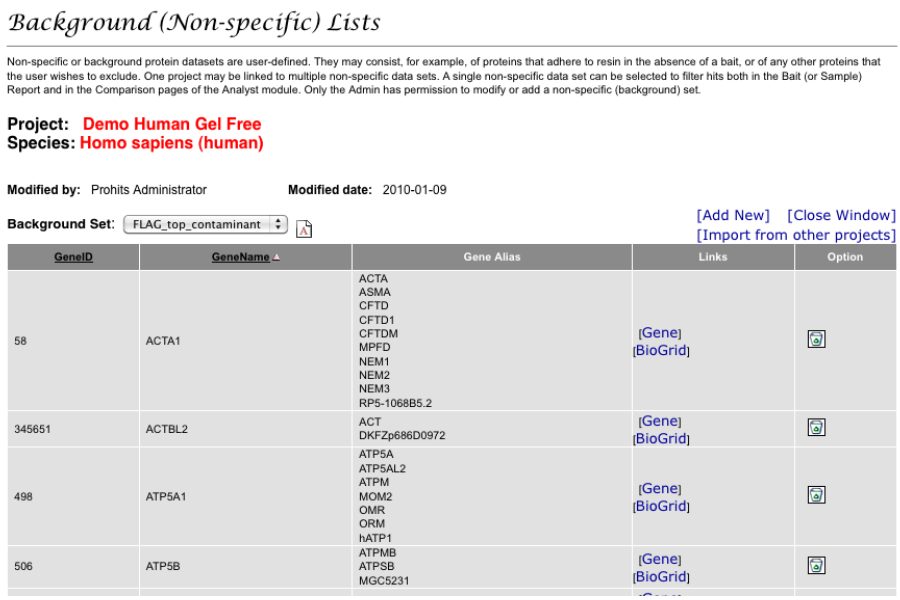
fi Select the "add as new" radio button and type a name (here: FLAG_top_contaminants).

Alternatively, append to an existing list by using the dropdown menu.

fi Select the row to start importing, and check the radio button in the GeneID field. Then, click [Process File].

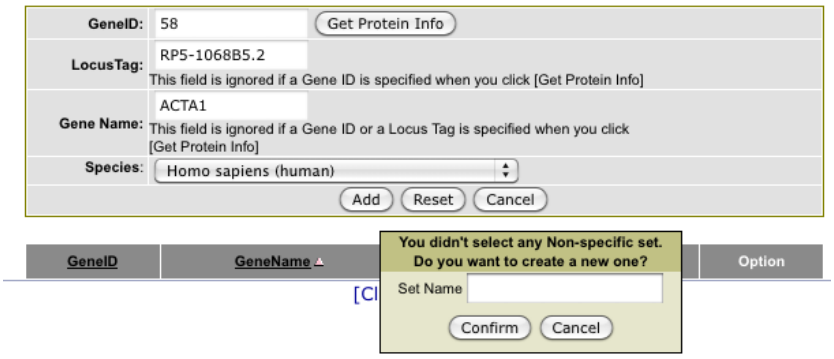


Once the file is processed, the contaminant list will be displayed (after selecting the name in the dropdown menu). You can manually remove individual entries (they will not be on the background list) by clicking the "delete" icon.



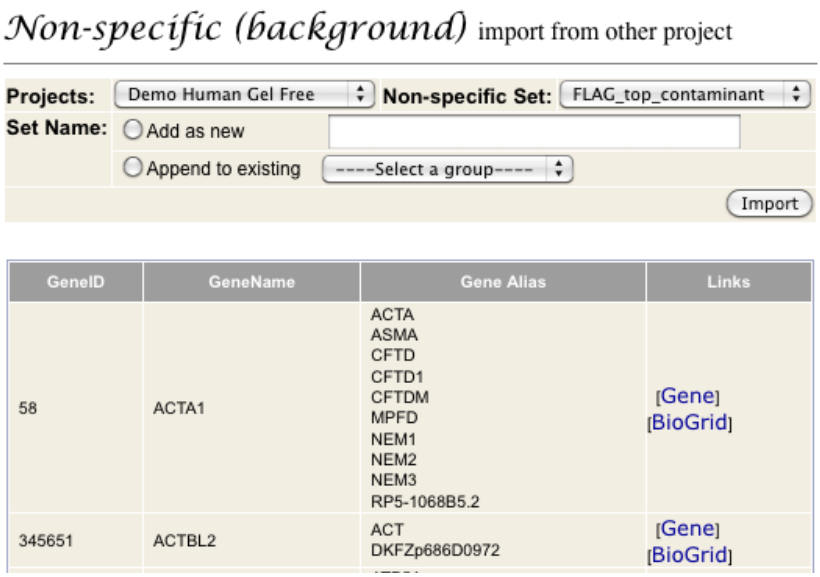
fi To manually add a protein to a background list, press [Add New].

You will then be prompted to enter a new contaminating/background protein. You can simply enter a gene name and species and press [Get Protein Info]. Press [Add] to include this protein on the background list.



If you do not specify a pre-entered non-specific set, ProHits will allow you to create a new one (press [Confirm] after entering the non-specific set name).

fi To import a contaminant list from a different project, press [Import from other projects], and navigate through the menus.



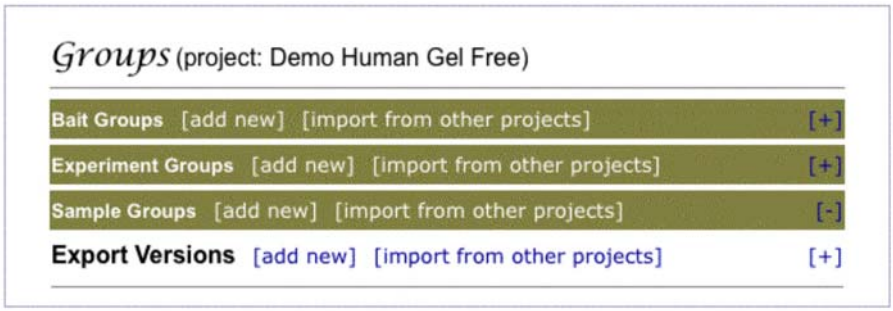
You now have your own background set that can be used for filtering both in the bait/sample report pages and in comparison. We recommend using caution when creating these sets: some proteins that are true interacting partners for a given bait may also be present (usually in lower amounts) on the background list. It may be a good idea to only include on this non-specific (background) list proteins detected across more than one control run with a high number of peptides.

Group Lists

ProHits allows the definition of new "groups" for any given project. As described earlier, groups are added to baits/samples by selecting the "Notes" Option. Groups act like flags and are displayed in the status bar in the "Report by Bait" or "Report by Sample" pages. These groups can help you organize your data.

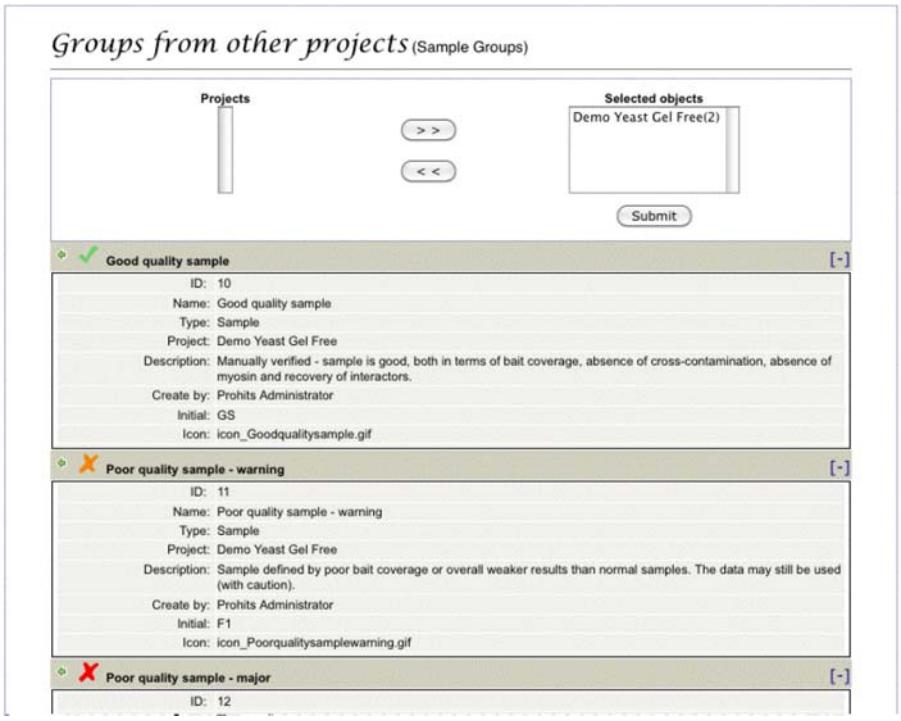
fi Select the "Groups" entry from the Manage Protocols and Lists.

As with the other Protocols and Lists, you can define new groups, or import a new group from another project. Here we will import sample groups from a different project.

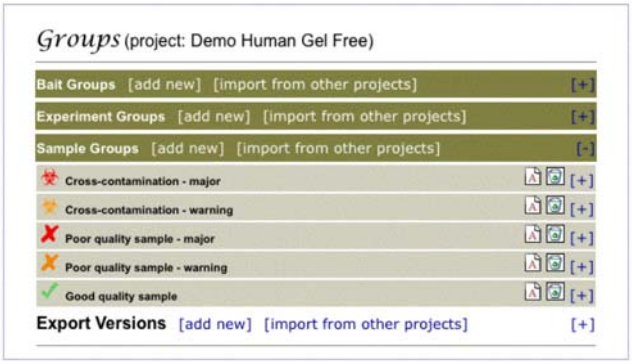


fi To import groups from a different project, press [import from other projects], and navigate through the menus.

As with the Text-based protocols, use the green arrows to transfer desired groups to the current project. You can only transfer one group at a time.

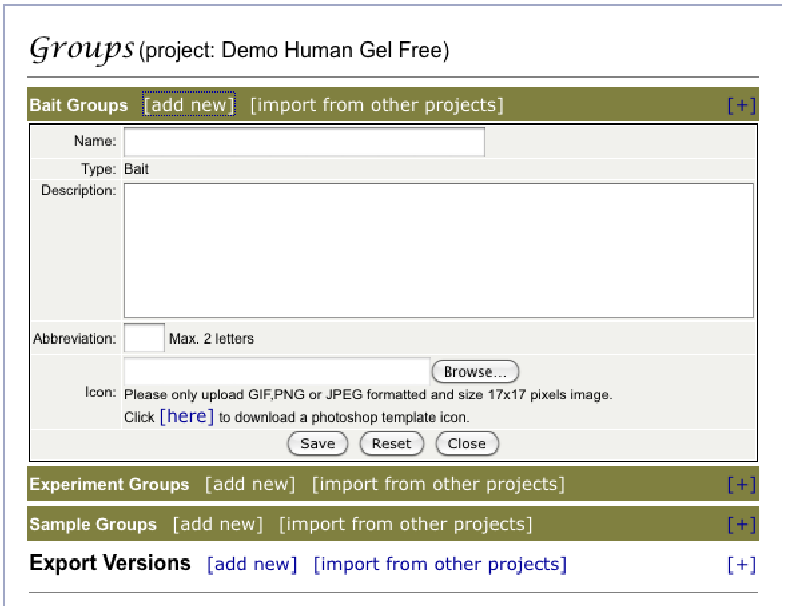


Upon transfer of a group, it will appear on your group list as shown below (the new group can be modified or deleted, unless it is used for a sample).



To create a new group, press [add new], and navigate through the menus.

Simply enter a short descriptive name for the group as well as a description, an abbreviation (that will be listed alongside the baits or samples), and an icon. Icons can easily be created in Photoshop as 17 x 17 pixel images, and saved as GIF, PNG or JPEG files. A template can be downloaded from the ProHits group page.



Export version

ProHits allows you to flag a group of samples, e.g. for inclusion in a publication or export to a third party.

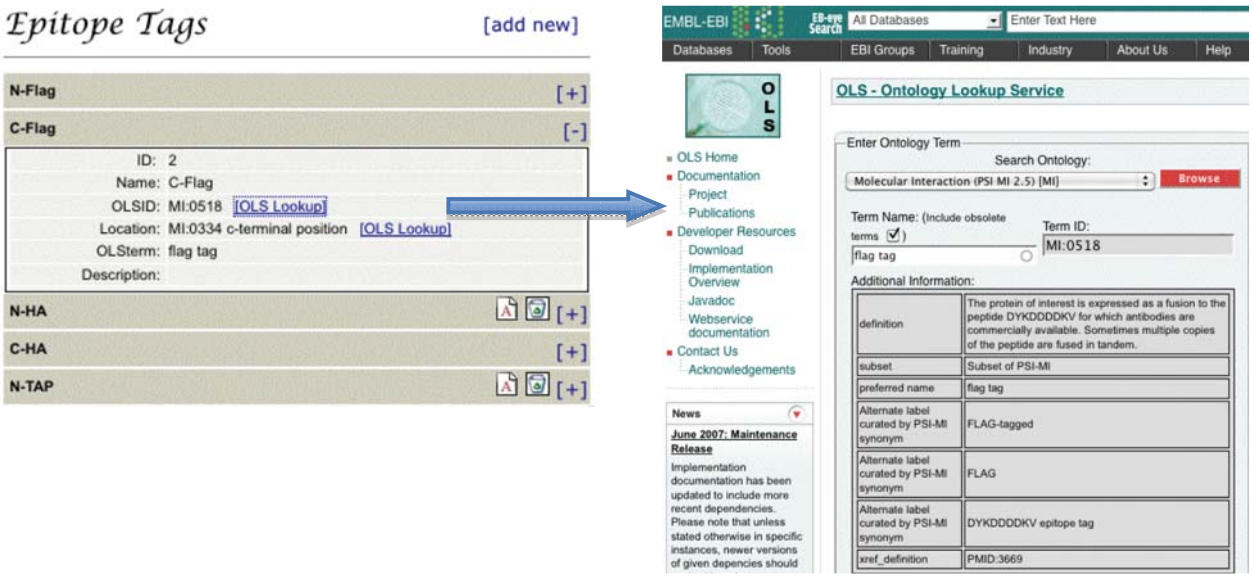
fi To create an Export Version, press [add new].

This will open a new menu with the default abbreviation (Version1, VS1), and Icon (a yellow star with the number 1). Subsequent versions will automatically be numbered VS2, VS3, etc., and the number inside the star will similarly increase. We suggest that you provide a meaningful short name and an accurate description of each "Export Version".

Epitope Tag Lists

fi Select the "Epitope Tag Lists" entry from the Manage Protocols and Lists.

This lists all tags available to the local ProHits projects. Clicking on the [+] sign expands the details of the epitope tag. We have mapped the current epitope tags in the demo database to PSI MI 2.5, using the Ontology Lookup Service (OLS) at the EBI. A link page is provided that allow retrieval of additional information.



In addition to the epitope tags currently in the system, you can create additional tags by pressing [add new] and navigating through the fields. Again, we strongly recommend mapping your terms to PSI MI 2.5 whenever possible.

ProHits Installation and Setup

The ProHits server requires a Linux system that runs PHP, Apache and MYSQL. We have developed this software on Fedora Core v9. Use of other versions of Linux may cause problems. Make sure you have full administrative access to the machine via Root access. Persons installing this package must have a good knowledge of both windows and Linux operating systems.

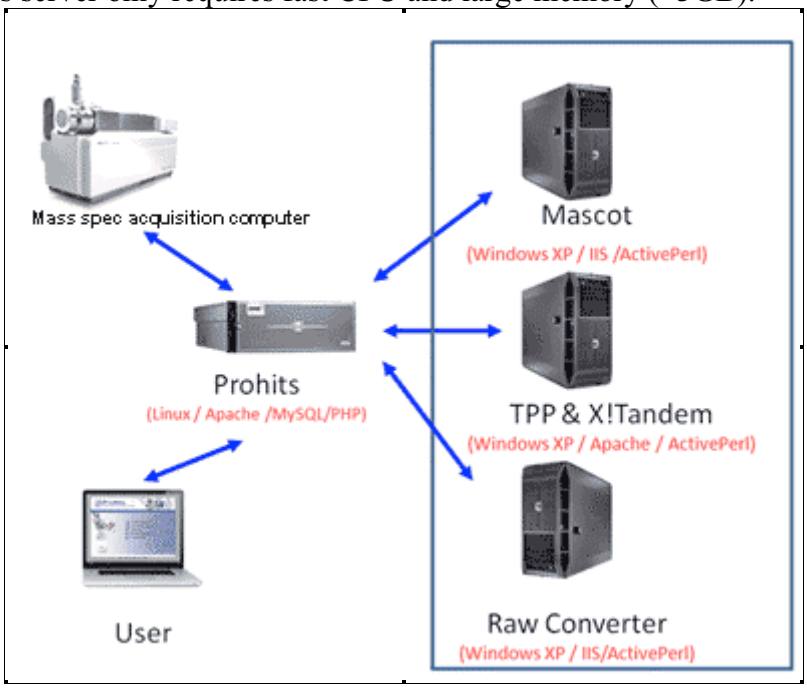
We make no claims that ProHits is stable or secured against hackers. We strongly suggest that ProHits be kept and used only behind a secured firewall. The following instructions apply to Fedora Core 9 (PHP5.2.5), Mascot 2.2 /2.3, TPP v4.3 and X!Tandem (2008.02.01).

Part 1: Prepare ProHits Installation

1. Computer requirement

The full version of ProHits (with MS Data Management) requires four web servers for full operation. If you use the ProHits Lite version (which only handles uploaded search results), only the ProHits server is required. All web server ports should be 80. Use of other ports may cause problems.

- ProHits server (Linux).
  - This server needs a large hard drive for raw file storage (>400GB) and large memory (>3GB).
- Mascot (Windows or Linux)
  - The Mascot server should have a large storage folder to keep search results (> 400GB).
- TPP & X!Tandem (Windows)
  - TPP and X!Tandem should be installed on the same server; this server should have a large storage folder to keep search results (>400GB).
- RawConverter (Windows)
  - This server only requires fast CPU and large memory (>3GB).



2. Setup the ProHits server

The ProHits server has been designed and tested only on an Intel-based Linux computer with particular emphasis on Fedora core v9. It requires the following packages that are available with all Linux distributions.

```
php
mysql
mysql-devel
mysql-server
php-mysql
httpd (Apache)
gd
php-gd
php-pear (after installing php-pear, use pear command to install
HTTP_Request)
sendmail
java-x.x.x-openjdk
java-x.x.x-openjdk-devel
```

To test whether you have these packages installed, log in as root and use the following command to see if you have installed packages from rpm:

- `yum list installed <PACKAGE NAME>`

For example, *yum list installed gd*
If you don't know package full name you can use grep command
For example, *yum list installed | grep openjdk*

If the package is installed, it should appear on your list. If it is missing, or you think it might not be the most recent version, type in:

- `yum install <PACKAGE NAME>`

- After you've installed **php-pear,** then use the following command to install HTTP_Request.

```
>pear list
>pear install Net_Socket
>pear install Net_URL
>pear install HTTP_Request
```

A. Set Network configuration

- Set a fixed IP address and DNS address for the Prohits server
  - Typically by clicking on **System --> Administration --> Network**

B. Make adjustments for the Trusted services in the iptables firewall

System > Administration > Security Level and Firewall > checkmark FTP, WWW(HTTP), DNS

- Click 'Other Ports' 'Add' port 3306 for Protocol 'tcp' (MYSQL)

C. **Edit the httpd.conf file** (/etc/httpd/conf/httpd.conf).

- Please set the web server to use **port 80**. Use of other ports may cause problems.
- Remove the option "Indexes" from document root directory setting (var/www/html), resulting in the following fragment:

```
<Directory "/var/www/html">
    Options FollowSymLinks ExecCGI Includes MultiViews
    AllowOverride None
    Order allow,deny
    Allow from all
</Directory>
```

- Find the line that starts with "DirectoryIndex"
  Add "index.php" and "index.html"
- Change the timeout if you need to process large raw files. 200 seconds is long enough to handle a 100 MB raw file. You may need to adjust this to your own needs.
  Timeout 200

D. Modify the PHP configuration file (/etc/php.ini) – the settings should be adjusted as follows:

```
register_globals = Off
default_socket_timeout=360
memory_limit = 20M
```

//you may increase the memory limit for a large raw file

```
upload_max_filesize = 700M
post_max_size= 700M
```

//you may increase the value if you want the user to upload a large raw file or TPP results files.

```
session.auto_start = 0
session.use_cookies = 1
```

//make sure session directory is writable for apache user (the directory is defined in php.ini)

```
Session.save_path =  "/var/lib/php/session"
session.use_only_cookies = 0
display_errors = On
```

E. Ensure your php session directory has write permissions for the Apache user

```
>ll /var/lib/php
>cd /var/lib/php
>chgrp apache session
>chmod g+wrx session
```

F. Restart computer services.

In the services configuration, make sure these services "Enabled" and "Running" services: Go to:

System > Administration > Server Settings > Services >

```
network
mysqld
sendmail
httpd
```

3. Setup the Mascot server (skip this step for Lite version)

A. ProHits supports Mascot 2.2/2.3 running on Linux or Windows web servers.

B. Copy '**ProhitsMascotParser.pl**' from Prohits/install/Mascot/ to the Mascot /cgi/ folder
Please make sure the first line of the file has the correct path to perl.exe in the Mascot server
e.g.
Mascot in Windows server
#!c:/perl/bin/perl.exe
Mascot in Linux server
#!/usr/local/bin/perl

4. Setup the TPP & X!Tandem server (skip this step for Lite version)

A. ProHits works with X!Tandem and TPP in both Windows and Linux servers. However, note that X!Tandem and TPP should be on the same server. We suggest using a Windows server for easy installation..

Install ActivePerl and Apache using default installation.

Apache: (IIS web server may not be supported)
http://httpd.apache.org/download.cgi
select Win32 Binary without crypto (no mod_ssl) (MSI Installer)
ActivePerl: (Be sure to install 5.8)
http://www.activestate.com/activeperl/downloads/
select ActivePerl 5.8.9.827 Windows Installer(MSI)

B. Please follow the instruction in
Prohits/install/GPM/install_TPP_GPM.html to setup TPP and X!Tandem.

5. Set up Raw file converter server (skip this step for Lite version) (must be a windows computer)

A. ProHits needs a Microsoft Windows computer running the IIS web server and ActivePerl to convert raw files to mgf and mzXML files.

B. Please follow the instructions in:
Prohits /install/RawConverter/install_rawConverter.html to setup RawConverter.

Part 2: Install ProHits

1. Place ProHits source code on ProHits server Apache document root directory

A. Unzip downloaded ProHits source in the Apache document root directory (/var/www/html).
Unzip ProHits source code:
```
> tar xvfz Prohits_v1.x_x.tar.gz
```

B. Change permission:
(find out Apache User from httpd.conf. The User should be apache or www-data)
```
>chown -R apache:apache Prohits
>chmod -R 755 Prohits
```

2. Run installation wizard.

Before running the installation wizard, make sure that SELinux http, ftp and cifs have been opened. You can turn off SELinux (**SELINUX=disabled** in **/etc/selinux/config),** if the ProHits server is firewall protected. For more information please read:
http://docs.fedoraproject.org/en-US/Fedora/13/html/Security-Enhanced_Linux/

```
> /usr/sbin/getenforce
```

A. Open installation wizard page from your browser.
   http://prohits_server_address/Prohits/install/install.php

B. Run installation check list after finish the wizard.
   http://prohits_server_address/Prohits/admin_office/check.php