

Exploring Proteomics Data (Exercise 12)

1. Find all genes with mass spec evidence from the basal body in Giardia. (For this exercise use <http://giardiadb.org>)

- The proteomics search is available under the heading “Protein Expression” in the “Identify gene by” section.

- How many genes did you identify? Examine the filter table below the strategy section. Why is the number in the ortholog groups column less than the number in the Assemblage A column? Why do the other cells have a zero?

All Results	Ortholog Groups	G.Assemblage A2 isolate DH	G.Assemblage A isolate WB	G.Assemblage B isolate GS_B	G.Assemblage E isolate P15
265	250	0	265	0	0

Gene ID	Genomic Location	Product Description	Selected Samples that Meet Criteria	Sum of Unique Peptides (Within Samples)	Unique Peptides (Across Samples)	Total Number of Spectra
GL50803_136020	GLCHR05: 3,625,897 - 3,627,040 (-)	Beta tubulin	3	234	177	2266
GL50803_112079	GLCHR03: 1,813,742 - 1,815,106 (-)	Alpha-tubulin	3	213	173	1874
GL50803_17230	GLCHR03: 483,217 - 484,152 (-)	Gamma giardin	3	175	136	1184

- What kinds of genes are in your result list? (hint: one option is to use the word cloud tool - click on the little graphic icon next to the column called “Product description”).

The screenshot shows a table of gene results with columns: Gene ID, Genomic Location, Product Description, Selected Samples, Sum of Unique Peptides (Within Samples), Unique Peptides (Across Samples), and Total Number of Spectra. A word cloud is overlaid on the table, showing terms like 'alpha', 'dynein', 'putative', 'ribosomal', 'chain', 'heavy', 'kinase', 'giardin', 'putative', 'ribosomal', 'specific', 'surface', 'synthetase', 'trna', 'variant', 'vsp', 'shippo', 'nekin', 'phosphatase', 'phosphate', 'pole', 'pyruvate', 'pyruvate', 'shippo', 'specific', 'spindle', 'surface', 'synthetase', 'trna', 'variant', 'vsp'.

- How many genes with mass spec data from the basal proteome experiment also have mass spec evidence from the “mitosome enriched fraction experiment”?

The screenshot shows a software interface for managing mass spec strategies. It includes a 'My Strategies' section with 'New', 'Opened (1)', 'All (38)', 'Basket', 'Public Strategies (5)', and 'Help'. A 'Mass Spec' strategy with 265 genes is shown. A red arrow points to the 'Add Step' button. Another red arrow points to the 'Add Step 2: Mass Spec. Evidence' dialog box, which shows the selection of 'Mitosome enriched fraction - nanoLC/MALDI TOF/TOF' (Jedelsk et al.) experiment. The dialog also shows parameters for 'Minimum Number of Unique Peptide Sequences' and 'Minimum Number of Spectra', both set to 1. A 'Combine Genes in Step 1 with Genes in Step 2' section offers options like 'Intersect 2', 'Union 2', and 'Relative to 2, using genomic collocation'. A 'Run Step' button is at the bottom.

- The default parameters of the mass spec search is to identify any gene with at least 1 peptide identified. How will your results change if you revise the two steps in your search strategy to only return genes with at least 5 peptides identified?

Any step in a strategy can be revised

2. Find gene in *Cryptosporidium* that have mass spec evidence from any of the sporozoite proteomics experiments available in CryptoDB. (For this exercise use <http://cryptodb.org>)

- Explore the available proteomics data and select samples that make sense. You may need to click on the '+' sign to expand experiments to see the underlying samples.

Identify Genes based on Mass Spec. Evidence

Experiment/Samples

- Cryptosporidium*
 - Cryptosporidium parvum*
 - Enriched cytoskeletal and membrane fractions separated by 1D-SDS-PAGE (Madrid-Alliste et al.)
 - 1D Gel 14Aug2006
 - 1D Gel 16May2006
 - 1D Gel 24Jun2006
 - Linear Ion Trap (LTQ) analysis of Oocyst Walls (Ferrari)
 - Intact Oocysts
 - Oocyst walls
 - Sporozoites
 - Mitochondrial Fraction Proteomics (Putignani)
 - Sporozoite 2D gel LC-MS/MS Analysis (Sanderson et al.)
 - 1D Gel LC-MS/MS
 - 2D Gel LC-MS/MS
 - MudPit Insoluble fractions
 - MudPit Soluble fractions
 - Sporozoite LC-MS/MS peptides, insoluble excysted fraction (Snelling et al.)
 - Insoluble Excysted Fraction LC-MS/MS
 - Insoluble Non-excysted Fraction LC-MS/MS
 - Soluble Excysted and Non-excysted Fraction LC-MS/MS

Minimum Number of Unique Peptide Sequences

Minimum Number of Spectra

- How can you remove any gene with peptide evidence from non-sporozoite samples? (hint: add a step for mass spec data and think about how you will combine your results.)

Add Step 2 : Mass Spec. Evidence

Experiment/Samples [select all](#) | [clear all](#) | [expand all](#) | [collapse all](#) | [reset to default](#)

- [-] **Cryptosporidium**
 - [-] **Cryptosporidium parvum**
 - Enriched cytoskeletal and membrane fractions separated by 1D-SDS-PAGE (Madrid-Aliste et al.)
 - Linear Ion Trap (LTQ) analysis of Oocyst Walls (Ferrari)
 - Intact Oocysts
 - Oocyst walls
 - Sporozoites
 - Mitochondrial Fraction Proteomics (Putignani)
 - Mitochondrial Fraction
 - Sporozoite 2D gel LC-MS/MS Analysis (Sanderson et al.)
 - Sporozoite LC-MS/MS peptides, insoluble excysted fraction (Snelling et al.)
 - Insoluble Excysted Fraction LC-MS/MS
 - Insoluble Non-excysted Fraction LC-MS/MS
 - Soluble Excysted and Non-excysted Fraction LC-MS/MS

[select all](#) | [clear all](#) | [expand all](#) | [collapse all](#) | [reset to default](#)

Minimum Number of Unique Peptide Sequences

Minimum Number of Spectra

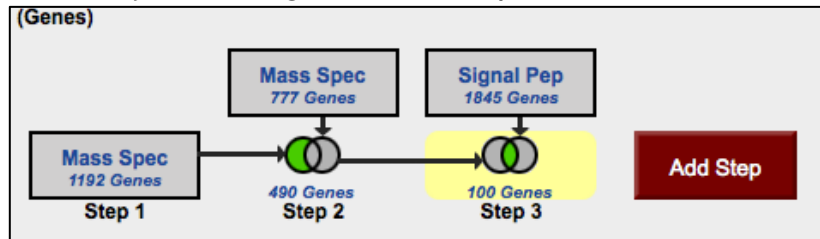
[Advanced Parameters](#)

Combine Genes in Step 1 with Genes in Step 2:

1 Intersect 2
 1 Union 2
 1 Minus 2
 2 Minus 1
 1 Relative to 2, using genomic colocation

[Run Step](#)

- How many of these genes are also predicted to be secreted?



- Note that so far you have been searching for *C. parvum* genes because we only have proteomics data from this species. However, what if you are interested in the orthologs of these genes in *C. muris*. How can you transform your *C. parvum* results to *C. muris* genes? (hint: add a step then select the “transform by orthology” option).

Add Step 4 : Transform by Orthology

Organism [select all](#) | [clear all](#) | [expand](#)

- Cryptosporidium hominis*
- Cryptosporidium muris*
- Cryptosporidium parvum*

[select all](#) | [clear all](#) | [expand](#)

Syntenic Orthologs Only?

[Advanced Parameters](#)

[Run Step](#)

(Genes)

```

graph LR
    S1[Mass Spec 1192 Genes Step 1] --> S2((490 Genes Step 2))
    MS1[Mass Spec 777 Genes] --> S2
    S2 --> S3((100 Genes Step 3))
    SP[Signal Pep 1845 Genes] --> S3
    S3 --> S4[Orthologs 285 Genes Step 4]
    AS[Add Step]
  
```

285 Genes from Step 4
Strategy: *Mass Spec(2)*

[Click on a number in this table to limit/filter your results](#)

All Results	Ortholog Groups	<i>Cryptosporidium</i>		
		<i>C.hominis</i>	<i>C.muris</i>	<i>C.parvum</i>
		TU502	RN66	Iowa II
285	100	101	75	109

- Why did the number of *C. parvum* genes increase?

3. Finding all genes with mass spec evidence in *T. cruzi*.

Note: for this exercise use <http://TriTrypDB.org>.

- How many genes in *T. cruzi* have mass spec evidence?
- How many genes from the results in a. have at least 10 unique peptide hits? (*hint*: try revising the step in 'a' and change the "minimum number of unique peptide sequences" option to 10.

STEP 1 : Mass Spec

Experiment/Samples : *Trypanosoma cruzi*, Proteome of *T. cruzi* C, Brenner epimastigotes (Bayona et al.), Proteome, Brenner, Proteome, esmeraldo-like, Proteome, nonesmeraldo-like, Life cycle proteome (Brenner) (Atwood et al.), amastigote, Brenner, amastigote, esmeraldo-like, amastigote, nonesmeraldo-like, epimastigote, Brenner, epimastigote, esmeraldo-like,

Revise Step 1 : Mass Spec. Evidence

Experiment/Samples ? select all | clear all | expand all | collapse all | reset to default

- Leishmania
- Trypanosoma
 - Trypanosoma brucei*
 - Trypanosoma cruzi*

select all | clear all | expand all | collapse all | reset to default

Minimum Number of Unique Peptide Sequences ?

Minimum Number of Spectra ?

Minimum Number of Unique Peptide Sequences : 1
Minimum Number of Spectra : 1

- Can you expand the list of results in 'b' to include possible orthologs/paralogs in *T. cruzi*?

Hint: you will have to use the ortholog transform option when adding a step and select only *T. cruzi*. Explore the columns in your result set.

3124 Genes from Step 2
Strategy: Mass Spec

Filter results by species (results removed by the filter will not be combined into the next step.)

All Results	Ortholog Groups	Leishmania						Trypanosoma brucei			Trypanosoma cruzi				Trypanosoma				
		braziliensis	donovani	infantum	major	mexicana	tarentolae	Distinct genes	TREU927	strain 427	gambiense	Distinct genes	esmeraldo	non-esmeraldo	unassigned	marinkellei	Sylvio	congolense	evansi
3124	376	0	0	0	0	0	0	0	0	0	3113	637	690	207	613	977	0	0	0

Gene Results Genome View

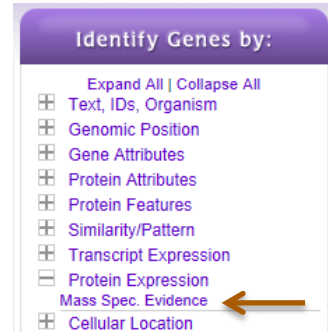
Gene ID	Organism	Genomic Location	Product Description	Input Ortholog(s)	Ortholog Group	Paralog count	Ortholog count
TCSYLVI0_000024	<i>T. cruzi</i> Sylvio X10/1	ADWP02000075: 3 - 665 (-)	retrotransposon hot spot (RHS) protein, putative	TcCLB.410923.20, TcCLB.459199.10, TcCLB.463155.20, TcCLB.503483.9, TcCLB.503607.4, TcCLB.503861.10, ...	OG5_126555	443	772
TCSYLVI0_000061	<i>T. cruzi</i> Sylvio X10/1	ADWP02000144: 81 - 671 (+)	retrotransposon hot spot (RHS) protein, putative	TcCLB.410923.20, TcCLB.459199.10, TcCLB.463155.20, TcCLB.503483.9, TcCLB.503607.4, TcCLB.503861.10, ...	OG5_126555	443	772
TCSYLVI0_000111	<i>T. cruzi</i> Sylvio X10/1	ADWP02000314: 1 - 663 (+)	retrotransposon hot spot (RHS) protein, putative	TcCLB.410923.20, TcCLB.459199.10, TcCLB.463155.20, TcCLB.503483.9, TcCLB.503607.4, TcCLB.503861.10, ...	OG5_126555	443	772
TCSYLVI0_000114	<i>T. cruzi</i> Sylvio X10/1	ADWP02000331: 2 - 661 (+)	retrotransposon hot spot (RHS) protein, putative	TcCLB.410923.20, TcCLB.459199.10, TcCLB.463155.20, TcCLB.503483.9, TcCLB.503607.4, TcCLB.503861.10, ...	OG5_126555	443	772
TCSYLVI0_000134	<i>T. cruzi</i> Sylvio X10/1	ADWP02000370: 140 - 1,198 (+)	retrotransposon hot spot (RHS) protein, putative	TcCLB.410923.20, TcCLB.459199.10, TcCLB.463155.20, TcCLB.503483.9, ...	OG5_126555	443	772

4. **Finding genes with mass spec evidence in *P. berghei* gametocytes.**
Note: For this exercise use <http://www.plasmodb.org>

a. **Find all *P. berghei* genes that have mass spec evidence in either or both male and female gametocytes.**

(*hint*: mass spec searches are in the “protein expression” expression section. Either or both is the Union of both results, not the intersection).

- How many genes did you get? How did you get to this number?
- Try running this search in two different ways:
 - i. Select both male and female gametocyte options and run the search.
 - ii. Select one of them first, run the search then add the other one using the add step button. How did you combine the two steps? Do you get the same results as in (i)?



b. **Find all genes that have mass spec evidence in both male and female gametocytes.**

(*hint*: use the strategy you developed in (ii) to get this answer, but change the union into an intersection).

c. **Find genes that have mass spec evidence only in male gametocytes and not in female ones.**

(*hint*: modify the set operation in b).

d. **Find genes that have mass spec evidence only in female gametocytes and not in male ones.**

(*hint*: modify the set operation in b).

e. **Which female gametocyte gene has the highest number of peptide sequences?**

(*hint*: look at the “number of peptide sequences” column in the list of results).

f. **What does the distribution of peptides in the gene from ‘e’ look like?**

(*hint*: go to the gene page and look at the “Protein features” section, or go to the genome browser from the gene page and turn on the right tracks).

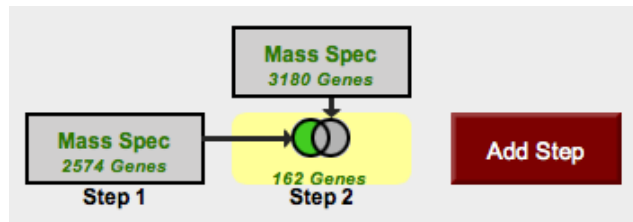
5. Finding genes with evidence of phosphorylation in intracellular *Toxoplasma* tachyzoites.

Note: For this exercise use <http://www.toxodb.org>

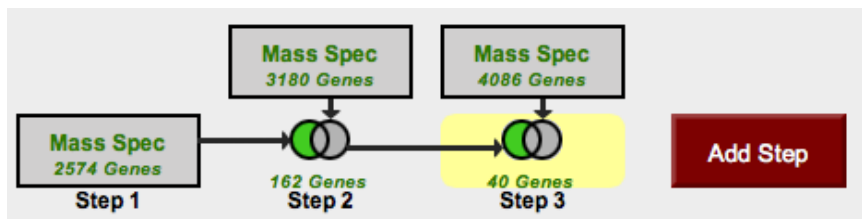
Hint: phosphorylated peptides can be identified by searching the appropriate experiments in the Mass Spec Evidence search page.

- a. Find all genes with evidence of phosphorylation in intracellular tachyzoites. Select the “Infected host cell, phosphopeptide-enriched (peptide discovery against TgME49)” sample under the experiment called “Tachyzoite phosphoproteome from purified parasite or infected host cell (RH) (Treeck et al.)”

- b. Remove all genes with phosphorylation evidence from purified extracellular tachyzoites.



- c. Remove all genes present in the phosphopeptide-depleted fractions (select both intracellular and extracellular).



d. Explore your results. What kinds of genes did you find? Are any of these results to be secreted? (Hint: add a step searching for genes with secretory

My Strategies: New Opened (1) All (2) Basket Examples Help

(Genes) Strategy: Mass Spec *
 Rename Duplicate Save As Share Delete

9 Genes from Step 4
 Strategy: Mass Spec Add 9 Genes to Basket | Download 9 Genes

Filter by organism or strain (results removed by the filter will not be combined into the next step.)

All Results	Ortholog Groups	Toxoplasma gondii				Neospora caninum		Eimeria tenella
		All	Non-redundant	GT1	ME49	VEG	RH	
9	9	9	9	0	0	9	0	0

Filter by strains (advanced) (results removed by the filter will not be combined into the next step.)

Gene Results Genome View

Advanced Paging Add Columns

Gene ID	Product Description
TGME49_288880	hypothetical protein
TGME49_257568	hypothetical protein
TGME49_229680	hypothetical protein
TGME49_231180	hypothetical protein
TGME49_269420	hypothetical protein
TGME49_200440	hypothetical protein
TGME49_216840	hypothetical protein
TGME49_308070	hypothetical protein
TGME49_219640	hypothetical protein

Advanced Paging

signal peptides).

e. Pick one or two of the hypothetical genes in your results and visit their gene pages. Can you infer anything about their function? (Hint: explore the protein and expression sections).

f. What about polymorphism data? Go back to your strategy and add columns for SNP data under the population section. Explore the gene page for the gene that has the most number of nonsynonymous SNPs.

Gene Results Genome View

Advanced Paging Add Columns

Gene ID	Product Description	Nonsynonymous SNPs All Strains	Synonymous SNPs All Strains	Total HTS SNPs All Strains	Total HTS Non-Synonymous SNPs	Total HTS Synonymous SNPs
TGME49_219640	hypothetical protein	33	60	383	81	302
TGME49_308070	hypothetical protein	20	34	188	42	146
TGME49_288880	hypothetical protein	17	27	221	52	169
TGME49_200440	hypothetical protein	14	13	72	36	36
TGME49_216840	hypothetical protein	13	42	189	71	118
TGME49_269420	hypothetical protein	9	27	45	31	14
TGME49_257568	hypothetical protein	5	16	30	20	10
TGME49_231180	hypothetical protein	3	15	54	24	30
TGME49_229680	hypothetical protein	0	6	33	2	31

Advanced Paging