

## Exploring Proteomics Data (Exercise 4)

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.

Identify Genes by:

- Expand All | Collapse All
- Text, IDs, Organism
- Genomic Position
- Gene Attributes
- Protein Attributes
- Protein Features
- Similarity/Pattern
- Transcript Expression
- Protein Expression
- Mass Spec. Evidence
- Cellular Location
- Putative Function
- Evolution

**Identify Genes based on Mass Spec. Evidence**

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

- Giardia
  - Giardia Assemblage A
    - Basal body Proteome (Lauwaet et al.)
      - Basal body enriched fraction (analysis 1)
      - Basal body enriched fraction (analysis 2)
      - Basal body enriched fraction (analysis 3)
    - DTA-Select Filter (Ratner)
    - Mitosome enriched fraction - nanoLC/MALDI TOF/TOF (Jedelsk et al.)

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

Minimum Number of Unique Peptide Sequences ?

Minimum Number of Spectra ?

[Advanced Parameters](#)

- 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?

My Strategies: [New](#) | [Opened \(1\)](#) | [All \(38\)](#) | [Basket](#) | [Public Strategies \(5\)](#) | [Help](#)

(Genes)
Strategy: Mass Spec \*

Mass Spec  
265 Genes

Add Step

[Rename](#)  
[Duplicate](#)  
[Save As](#)  
[Share](#)  
[Delete](#)

**265 Genes from Step 1**
[Add 265 Genes to Basket](#) | [Download 265 Genes](#)

Strategy: Mass Spec

Click on a number in this table to limit/filter your results

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

Filter Table

Assemblage A Genes  
 Not Deprecated |  Deprecated

265
0

Gene Results
Genome View

First 1 2 3 4 5 Next Last
Advanced Paging
Add Columns

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', 'kinase', and 'chain'. A red arrow points from the 'Product Description' column header to the word cloud.

Gene ID	Genomic Location	Product Description	Selected Samples	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_17236						
GL50803_4812						
GL50803_21942						
GL50803_86676						
GL50803_11231						
GL50803_16935						
GL50803_10524						
GL50803_14373						
GL50803_41512						
GL50803_10808						
GL50803_16648	3,785,915 - 3,788,233 (-)	hypothetical protein				
GL50803_16844	GLCHR04:	hypothetical protein	3	55	45	196

- 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 workflow tool interface. A red arrow points from the 'Add Step' button to the configuration window for 'Add Step 2 : Mass Spec. Evidence'. The configuration window shows the following settings:

- Experiment/Samples: Giardia Assemblage A
  - Basal body Proteome (Lauwaet et al.)
  - DTA-Select Filter (Ratner)
  - Mitosome enriched fraction - nanoLC/MALDI TOF/TOF (Jedelski et al.)
- Minimum Number of Unique Peptide Sequences: 1
- Minimum Number of Spectra: 1
- Combine Genes in Step 1 with Genes in Step 2:
  - 1 Intersect 2
  - 1 Union 2
  - 1 Relative to 2, using genomic colocation
  - 1 Minus 2
  - 2 Minus 1

- 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-Aliste 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

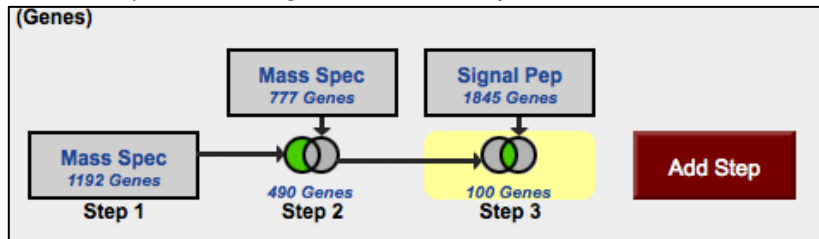
[Advanced Parameters](#)

**Combine Genes in Step 1 with Genes in Step 2:**

1 Intersect 2   
  1 Minus 2   
  1 Union 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)**

**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	Cryptosporidium		
		C.hominis	C.muris	C.parvum
		TU502	RN66	Iowa II
285	100	101	75	109

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