Functional Genomics I Proteomics

1. Find all Giardia genes with evidence of basal body expression based on mass spec/proteomics data.

For this exercise use <u>http://giardiadb.org</u>

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



- a. How many genes did you identify?
- **b.** What kinds of genes are in your result list?

Hint: analyze the Product Description column.

The **word cloud tool** counts the number of times a word appears in the column and then draws a word cloud in which the size of the word reflects how many times the word appears in the product description column. Click on the little graphic icon next to the column called "Product description".

	st 1 2 3 4 5 Nex	tt Last Advan	ced Paging				Add Colum
0	🗘 Gene ID	Genomic O	Product Obscription	Selected Samples Criteria	Sum of Unique Peptides (Within Samples)	Unique Peptides (Across Samples)	Total Number of Spectra
1	GL50803_136020	GLCHR05: 3,625,697 - 3,627,040 (-)	Beta tubulin	3	234	177	2266
2	GL50803_112079	GLCHR03: 1,813,742 - 1,815,106 (-)	Alpha-tubulin	3	213	173	1874
2	GL50803_17230	Word Cloud					
2	GL50803_4812	Word Cloud Data					
	GL50803_21942	Filter words by rank:	3 to 50				
	GI 50803 86676	Sort by: Rank A	-Z	Use slider or en	nter numbers to adjust filter		
	0200000_00010		Mou	se over a word to see its occur	ence in the column		
	GL50803_11231:	alaba	Мои	se over a word to see its occur	ence in the column		
	GL50803_11231. GL50803_16935	alpha.	Mountigen associated atp atpase b0ta bi	se over a word to see its occur nding calmodulin candidate carbi	ence in the column	coil coiled dehydrogena	9\$0 dynamin
	GL50803_11231: GL50803_16935 GL50803_10524	ory alpha . dynein oos elor	Mou religen associated atp atpase beta bi agation factor flavodoxin frag	se over a word to see its occur nding calmodulin candidate carbi gamma giardin gip	ence in the column explase chain coa heat heavy histone	coil coiled dehydrogena a hap isomerase king	ase _{dynamin} BSE like
	GL50803_11231; GL50803_16935 GL50803_10524 GL50803_14373	dynein ever eior	Moure referent associated alle alpasse beta bi agation factor favodesin frag	se over a word to see its occur ndrig calmodulin candidate carbs gamma giardin ge e oversvata rik	ence in the column explanse chain coa heat heavy histone	coil coiled dehydrogene a he isoverase kina mal sh	ase _{dynamin} 3SC like
	GL50803_11231: GL50803_16935 GL50803_10524 GL50803_14373 GL50803_41512	sovi alpha dynein «de elor nek phosphatase p specific spindle su	Mou refigen associated atp apase both bin gation factor favodavin teng phosphate pole putativ rface synthetase trma	ee over a word to see its occur rding calmodulin candidate carbi gamma giardin ge e pyruvate variant vsp	ence in the column explase chain coa heat heavy histore	coil coiled dehydrogena a hep isorressa kina mal sh	ase _{dynamin} BSE like ippo

Hint: analyze the Gene Ontology terms assigned to the genes in your result list. **The Result Analysis/Enrichment tool** applies the Fischer's Exact test to compare your gene result to the entire genome. Use the Gene Ontology Enrichment to find Biological Process ontology terms that are enriched in your gene result.

265 Genes from Step 1 Strategy: Mass Spec: ■ ▼ Cacco as sumber in this table to limit@ther your results ■ ↓ Results Gradue ↓ ↓ Results Gradue ↓ ↓ 265 250 0 ↓ ↓ 265 250 0 ↓ ↓ 3< ▼ Assemblage A Genes 0	265 Genes from Step 1 Strategy: Also Spec Strategy: Click as a number in this table to limit?liter your reference of the state of	suits Dardie 0.Assemblage 8 0.Assemblage 8 isolate GS_B isolate GS isolate P15 0 0 0	Spironucleus S saimonicida ATCC50377 0	Add 285 Genes to Basket Download 285 Gene
Indicaptempo (Jeperation 266 o o * Cene Results Centra View Analyze Rescalts Centra View First 1 2 3 4 5 Next Last Advanced Paging	Analyze Reserved Organization Analyze Reserved Organization Analyze Poor Reserved Organization Analyze Poor Reserved Organization Analyze Poor Reserved Organization	Analyza Results Stow. Metabolic Pattwey Er Metabolic Pattwey Er Metabolic Pattwey Er	richneat assis	kinase phosphatase exported membrane Word Earichment
Gene Ontology Enrichment Find Gene Ontology terms that are en	ched in your gene result. Read More			
	Organism® Giardia A Ontology® © Cellular Molecui ® Riotonic	ssemblage A isolate WB - Component ar Function al Process		
	GO Association Sources e Select all (P-Value Cutoff (a - 1.0) 0 005 Submit	Clear all predictions		

c. 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 St	rategies	: New	Opened (1)	All (38)	Baskel	t Public Strat	egies (5) Help		
(Genes) Mas 265	ss Spec	Add Step						Strategy	: Mass Spec * Rename Duplicate Save As Share Delete
S	tep 1								
265 G Strate	enes fri egy: Mas	om Step 1 is Spec imber in this tab	ie to limit/filter y	our results			Add 265	i Genes to Basket D	ownload 265 Gene
	Ortholiza	Giardia							
Results	Groups	Assemblage A2	G.Assemblage	G.Assem	blage B	G.Assemblage E	10.000	10.00	
	-	isolate DH	isolate WB	isolate GS_B	isolate GS	isolate P15	Filte	r Table	3
265	250	0	265	0	0	0		I I GD CC	
Not Dep 26 Gene First :	Results	Genome View	Advance	d Paging			Sum of		Add Columns
\$ ۵	Gene ID	Cenomi Locatio		Product Description	0 💼	Selected Samples that Meet Criteria	Deptides (Within Samples)	Unique Peptides (Across Samples)	Total Number @ of Spectra
GL	.50803_1360	020 GLCHR05: 3,625,697 - 3	3,627,040 (-) B	eta tubulin		3	234	177	2266
GL	.50803_1120	079 GLCHR03: 1,813,742 -	1,815,106 (-) A	lpha-tubulin		3	213	173	1874
GL	50803_172	30 GLCHR03: 483,217 - 48	4,152 (-)	iamma giardin		3	175	136	1184

d. How many genes with mass spec data from the basal proteome experiment also have mass spec evidence from the "Mitosome enriched proteome (WB) (Jedelsk et al.)"?

My Strategies: New Opened (*	1) All (38) Basket Public St	rategies (5) Help	
Genes) Mass Spec 26 Genes Add Step			Strategy: Mass Spec * Rename Duplicate Save As Share Delete
Step 1		Add Step	
Run a new S Transform by Add content Add existing Filter by assi	iearch for Genes y Orthology Genomic Segments s of Basket ORFs Strategy igned Weight	 Text, IDs, Organism Genomic Position Gene Attributes Protein Attributes Protein Attributes Similarity/Pattern Transcript Expression 	Mass Spec. Evidence
	Add Step	Bratida Barrania	(x)
Experiment/Samples <table-cell> Minimum Number of Unique Peptide Sequences 🔊</table-cell>	select all clear all expand all collapse all rese	t to default tet al.) Jedelsk et al.) to default	
Minimum Number of Spectra 😯	1		
	Advanced Parameters		
Combine Genes in Step 1 © © 0 0 1 1	with Genes in Step 2: 1 Intersect 2 ① 1 Minus 2 1 Union 2 ② 2 Minus 1 1 Relative to 2, using genomic colocatio	n	
	Run Step		

e. The default parameters of the mass spec search are set to identify any gene with at least 1 mapped 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?



- 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
- **a.** 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 Ge	enes based on Mass Spec. Evidence
Experiment/Samples () Minimum Number of Unique Peptide Sequences () Minimum Number of Spectra ()	select all (clear all expand all collapse all reset to default Cryptosporidium Cryptosporidium parvum Concyst Wall Proteome (lowall) (Ferrari) Concyst Wall Proteome (lowall) (Ferrari) Concyst Walls Sporozoites Proteome during Sporozoite Excystation (ISSC162) (Snelling et al.) Concyst walls Soluble Excysted Fraction LC-MSMS Soluble Excysted and Non-excysted Fraction LC-MSMS Soluble Excysted and Non-excysted Fraction LC-MSMS Soluble Excysted and Non-excysted Fraction LC-MSMS MudPit Insoluble fractions select all clear all expand all collapse all reset to default
	Advanced Parameters
	Get Answer

b. Remove any gene with peptide evidence from non-sporozoite samplesHint: add a step for mass spec data and think about how you will combine your results.

Add Step 2 : Mass Spec. E	vidence
Experiment/Samples 🤣	select all clear all expand all collapse all reset to default
5402 ST	🗄 🔳 Cryptosporidium
	🗄 🔳 Cryptosporidium parvum
	Enriched cytoskeletal and membrane fractions separated by 1D-SDS-PAGE (Madrid-Aliste et al.)
	E Ilinear Ion Trap (LTQ) analysis of Oocyst Walls (Ferrari)
	- MIntact Oocysts
	- Oocyst walls
	L. Sporozoites
	Mitochondrial Fraction Proteomics (Putignani)
	I. 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
	Soluble Exercised and Non-exercised Eraction LC MS/MS
	select all clear all expand all collapse all reset to default
Minimum Number of Unique Peptide Sequences (2)	1
Minimum Number of Spectra 2	1
	Advanced Parameters
Combine Genes in Step 1	with Genes in Step 2:
0	1 Intersect 2 💿 🔘 1 Minus 2
\odot \bigcirc	1 Union 2 🛛 🔘 2 Minus 1 🕌
	1 Relative to 2, using genomic colocation

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



- **d.** 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).
 - Did the number of *C. parvum* genes increase or decrease? Why?

Add Step 4 : Transform by Orthology	
Organism select all clear all expand	(Genes) Mass Spec 777 Genes Mass Spec 1845 Genes 1845 Genes 192 Genes 192 Genes 192 Genes 190 Genes 100 Genes
Advanced Parameter	6
Run Step	285 Genes from Step 4 Strategy: Mass Spec(2)
	□ ▼ Click on a number in this table to limit/filter your results
	Cryptosporidium
	All Ortholog C.hominis C.muris C.parvum
	TU502 RN66 Iowa II
	285 100 101 75 109

- 3. Finding all genes with mass spec evidence in *T. cruzi*. For this exercise use http://TriTrypDB.org
- a. How many genes in *T. cruzi* have expression evidence based on mass spec data?

Identify G	Identify Genes based on Mass Spec. Evidence								
Experiment/Samples 0	select al clear al expand al colapse al reset to defaut								
Minimum Number of Unique Peptide Sequences � Minimum Number of Spectra �	1								
	Advanced Parameters								
	Get Answer								

b. How many genes from the results in **'a'** have at least 10 uniquely mapped peptides? (*hint*: try revising the step in 'a' and change the "minimum number of unique peptide sequences" option to 10.

My Strategies: New ((Genes) Mass Spec 7874 Gene Step 1	Copened (1) Rename Vii Analyze Kevise Make Neelec STEP 1 Experiment/Samples : Trypanosoma c (Queiroz et a), esmeraldo-like, pimastigote, B b btm/miated epim Brener, control e	I Stralegy Insert Step Befg : Mass Spec :ruzi Epimastigote Cell Sus SHAVE epimastigote, non- rene, biodhyaled epimastig nastigote, non-esmeraldo-lig ipmastigote, esmeraldo-lig)	
	amastipole nonesmera epimastipol metacyclic nonesmera esmeraldo from insec GPI-anchor membrane Phosphopr Epimastipo bour of nutr nutritional s nutritional s 12 hour, Br esmeraldo honesmera esmeraldo	Revise Step 1 : Mass Sper Experiment/Samples • Minimum Number of Unique Peptide Sequences • Minimum Number of Spectra •	Revise Step c. Evidence select al coara al colapse al reset to default	×

c. Expand the list of results in **'b'** to include possible orthologs/paralogs in *T. cruzi*.

Hint: use the ortholog transform option when adding a step and select only *T. cruzi*. Explore the columns in your result set. Pay close attention to the organism filter table.

(Genes) Ma 97	(Genes) Strategy: Mass Spec Mass Spec Orthologs 4004 Genes Add Step Add															
4004 Strate	4004 Genes from Step 2 Add 4004 Genes to Basket Download 4004 Genes Strategy: Mass Spec(8) ■ ▼Click on a number in this table to limit@liter your results															
		Crithidia				Leishman	ia									Trypan
All	Ortholog	C.fasciculata	<i>L.brazilie</i> Gene	<i>nsis</i> (nr s:0)	L.donovani	L.infantun	L.major	L.mexica	na L	.tarentolae	T.brucei (nr Genes:		enes: 0)	T.congolense	è	T.c.
Result	Groups	strain Cf-Cl	MHOM/BR /75/M2903	MHOM/BR /75/M2904	BPK282A1	JPCM5	strain Friedlin	MHOM/0 /2001/U11	GT 103 F	Parrot-Tarll	Lister strain 427	rain 27 TREU927 gambiense DAL972		e IL3000	CL Brener Esmeraldo-like	(Non-E
4004	574	0	0	0	0	0	0	0		0	0	0	0	0	865	
•								1						1	•	۴
Gene	Results	Genome Vi	ew An	alyze Res	ults	2										
First	12345	i Next Last	A	dvanced P	aging										Add Colun	nns
1	Gene ID	\$ 01	ganism 🕹	♦ Gen Uoca	omic ition	Ø ‡Pi	roduct escriptio	n 🎱 📊 🗄	♦ Inp	put tholog(s)	0	Ortholog Group	🕲 🌲 Pa	ralog 🚳 📊 unt	Count	
Image: Control Contro Control Contecontrol Control Control Control Control Control Con						443	1275									
Т	TCSYLVIO_000061 T. cruzi Sylvio X10/1 ADWP02000144: 81 - 671 (+) retrotransposon hot spot (RHS) protein, putative TCLB.503480.1.0, TCLB.503407.4, TCCLB.503407.4, TCCLB.503407.4, TCCLB.503407.4, OG5_128555 443 1275															

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

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



Hint: When using a two-step search to find genes that have either or both characteristics, take the union of the two searches, not the intersection.



- 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? Select female gametocyte search step.

(*Hint*: look at the "Sum of Unique Peptides (Within Samples)" column in the list of results).

Му	Strategies:	Ne	w Op	ened (3)	All (3)	Basket	Public Stra	itegie	S (8)	Help				
(Gen	female 736 Genes	823 316 SI	lale Genes Genes tep 2	Add	Step						Stra	rtegy: Mass	Spec(3) Ri Duj Si	ename plicate ave As Share Delete
_								_						
736 Stra	Genes from tegy: Mass	n Step Spec(3	o 1 3) is table to lii	nit/filter y	our results				,	Add 736 Ger	nes to	Basket Do	wnload 7	'36 Gene
							Plasmod	ium						
AJ	Ortholog Pt	erghei	P.chabaudi	P.cynom	olgi P.falciparum	(nr Gene	s:0) P.gallinace	um P	knowlesi	Preichenowi	P.vivax.	P.yoelii	(nr Genes	3:0)
Resi	lits Groups	NKA	chabaudi	strain	B 3D7	ſ	F 8A		strain H_	Dennis	Sal-1	yoelii 17XNL	yoelii 17X	yoelii YM
73	6 726	736	0	0	0	() 0	+	0	0	0	0	0	0
Ge Fir	ne Results C st 1 2 3 4 5 No Gene ID	enome	view t	Analyze Advanc	Results	•	Selected Samples that Meet Criteria	3	Sum Uniq Pept (Wit Sam	of ue ides O hin ples)	♥ Un ♥ Pej (Ac Sal	ique ptides cross mples)	Add Tota Num of Spec	Columns I Iber @ ctra
	PBANKA_14630) berg 2,327	14: 7,865 - 2,335	,685 (-)	osmiophilic body (G377)	protein	1			48		48		48
	PBANKA_07119	bergi 420,1	07: 169 - 422,250	0 (-)	heat shock protei (HSP70)	n, putative	1			34		34		34
	PBANKA_08057	bergi 268,6	08: 90 - 271,84	7 (+)	heat shock protei putative (HSP90)	n 90,	1			33		33		33

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. Find genes with evidence of phosphorylation in intracellular *Toxoplasma* tachyzoites. For this exercise use http://www.toxodb.org

Phosphorylated peptides can be identified by searching the appropriate experiments in the <u>Mass Spec Evidence</u> search page.

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.)"

Identify Ge	Identify Genes based on Mass Spec. Evidence									
Experiment/Samples 🕈	select all [clear al] expand al] collapse all reset to default Eimeria Toxoplasma Toxoplasma Oocyst Partially Sporulated Proteome (VEG) (Possenti, et al.) Oocyst proteome (M4 Typell) (Wastling) Oocyst proteome - Fractionated (M4 type II) (Fritz et al.) Proteome During Infection in H. sapiens (Wastling) Tachyzoite Intra- and Extracellular Lysine-Acetylomes (RH) (Jeffers and Xue) Tachyzoite conoid proteome (RH) (Hu et al.) Tachyzoite membrane and cytosolic fractions (RH) (Nebl et al.) Tachyzoite phosphoproteome - Calcium dependent (RH) (Nebl et al.) Tachyzoite bhosphoproteome - Calcium dependent (RH) (Nebl et al.) Tachyzoite phosphoproteome - Calcium dependent (RH) (Nebl et al.) Tachyzoite bhosphopeptide-depleted (peptide discovery against TgME49) Infected host cell, phosphopeptide-depleted (peptide discovery against TgGT1) Purified tachyzoites phosphopeptide-enriched (peptide discovery against TgME49) Infected host cell, phosphopeptide-enriched (peptide discovery against TgME49) Purified tachyzoites phosphopeptide-enriched (peptide discovery against TgME49) Purified tachyzoites phosp									
Minimum Number of Unique Peptide Sequences 🕜	1									
Minimum Number of Spectra 📀	1									
	Advanced Parameters									
	Get Answer									

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? (Hint: use the Product description word column or perform a GO enrichment analysis of your results.)
- **e.** Are any of these results to be secreted? (*Hint*: add a step searching for genes with secretory signal peptides).

My S	strategies: New	Opened (1) All	(1) 💮 Basket	Public Strategies (4)	Help						
(Gene	s)					Strategy:	Mass Spec * 🛛 🔯				
M 2	Ass Spec 290 Genes 390 Genes Step 1 Step 2	Mass Spec 3880 Genes 105 Genes Step 3	Signal Pep 26271 Genes 22 Genes Step 4	Add Step			Rename Duplicate Save As Share Delete				
22 Genes from Step 4 Add 22 Genes to Basket Download 22 Genes Strategy: Mass Spec Image: Click on a number in this table to limit/filter your results Image: The provided interval of the pro											
Gene Results Genome View Analyze Results											
Firs	First 1 2 Next Last Advanced Paging Add Colum										
	韋 Gene ID	Product	Description 🎱 👔	L							
	TGME49_294940	hypothetical	hypothetical protein								
	TGME49_222870	hypothetical p	hypothetical protein								
	TGME49_320150	elongation fa	elongation factor Tu GTP binding domain-containing protein								
	TGME49_288370	hypothetical p	hypothetical protein								
	TGME49_288880	hypothetical	hypothetical protein								
	TGME49_259830	diacylglycero	diacylglycerol kinase catalytic domain-containing protein								
	TGME49_257640	hypothetical	hypothetical protein								
	TGME49_257595	hypothetical	hypothetical protein								
	TGME49_229680	hypothetical	hypothetical protein								
1	TGME49 231180	hypothetical	hypothetical protein								

- **f.** 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).
- **g.** What about polymorphism data? Go back to your strategy and add columns for SNP data found under the population biology section. Explore the gene page for the gene that has the most number of non-synonymous SNPs. (*Hint*: you can sort the columns by clicking on the up/down arrows next to the column names).

Fin	First 1 2 Next Last Advanced Paging Add Columns														
١	🗘 Gene ID	Product Oescription	SNPs All	NonSynonymous SNPs All Strains	SNPs All Strains	SNPs All Strains	SNPs with Stop Codons All Strains	NonSyn/Syn SNP Ratio All Strains	3						
	TGME49_271110	hypothetical protein	890	157	44	679	10	3.57							
	TGME49_257595	hypothetical protein	317	123	51	131	12	2.41							
	TGME49_219640	hypothetical protein	382	85	34	263	0	2.5							
	TGME49_288370	hypothetical protein	224	82	35	105	2	2.34							
	TGME49_216840	hypothetical protein	189	75	23	89	2	3.26							
	TGME49_257640	hypothetical protein	110	66	12	31	1	5.5							
٦	TGME49_320150	elongation factor Tu GTP binding domain-containing protein	378	65	22	286	5	2.95							
	TGME49_235960	hypothetical protein	155	58	14	77	6	4.14							
	TGME49_288880	hypothetical protein	220	56	17	147	0	3.29							
	TGME49_269750	CrcB family protein	95	54	20	18	3	2.7							
	TGME49_315700	hypothetical protein	338	54	14	265	5	3.86							
	TGME49_308070	hypothetical protein	188	43	22	123	0	1.95							
۵	TGME49_269420	hypothetical protein	45	37	8	0	0	4.63							
	TGME49_200440	hypothetical protein	72	35	11	24	2	3.18							
۵	TGME49_259830	diacylglycerol kinase catalytic domain-containing protein	176	32	3	139	2	10.67							
	TGME49_236220	PCI domain-containing protein	383	28	18	332	5	1.56							
	TGME49_231180	hypothetical protein	54	25	9	18	2	2.78							
	TGME49_294940	hypothetical protein	137	16	7	111	3	2.29							