Functional Genomics I

Exploring transcriptomics, proteomics, phenomics and metabolic pathways

1. Find T. gondii genes expressed in late enteroepithelial stages

Toxoplasma gondii is a zoonotic pathogen for which felids serve as definitive hosts. In cats, the parasite undergoes several rounds of asexual replication before entering the sexual cycle which gives rise to oocysts that are shed into the environment. These then sporulate and become infective to humans and livestock. To understand the genes involved in the parasite development in the felid host and identify potential intervention targets, we designed a transcriptomic approach to compare the cat intestinal stages with the well characterized tachyzoites that mediate acute infection and tissue cysts that are responsible for chronic infection. Cats were infected with *T. gondii* CZ clone H3 tissue cysts from mouse brain and the intestinal stages were sampled at day 3, 5 and 7 post infection. As an input sample, we also collected tissue cysts from mouse brain. In vitro cultivated tachyzoites were also harvested. Total RNA was extracted, enriched for mRNA and used for cDNA synthesis. RNA-Seq was then performed to describe the transcriptomic repertoire of each developmental stage. RNA-seq datasets from each time point post inoculation with bradyzoites in kittens were subjected to cluster analysis and assigned to five enteroepithelial developmental stages (EES) according to their profile.

Cat enteroepithelial stages:

- EES1 = very early enteroepithelial stages
- EES2 = early enteroepithelial stages
- EES3 = mixed enteroepithelial stages
- EES4 = late enteroepithelial stages
- EES5 = very late enteroepithelial stages
- Navigate to the RNAseq searches and identify the experiment of cat enterocyte stages. Configure the search to identify call *T. gondii* genes that are upregulated by at least 2fold in late and very late enteroepithelial stages (EES4 and EES5) compared to all other stages available from this experiment.



• What kinds of genes did this search identify? How can you determine if your results are enriched for a particular function? Try clicking on Analyze Results and explore the GO enrichment tool.



2. Finding genes based on high throughput mutagenesis and fitness analysis.

In EuPathDB we have a variety of studies where genome scale phenotypic analyses were carried out. In this exercise we'll use <u>ToxoDB.org</u> and look at fitness following CRISPR mutagenesis. You could also explore phenotyping studies in PlasmoDB or FungiDB if you prefer, the principles are the same.

 Navigate to the CRISPR phenotype search. Note that this search form is quite simple just requiring a range of fitness values. The defaults return all genes not limiting the search at all. This is only useful in as much as it tells you which genes were assayed which is nearly the entire genome. The tricky bit is deciding where to make the cutoffs. Again, the description on the search form is very helpful in this regard (as is the link to the paper ... remember these phenotypes were assayed under specific conditions so just because a particular gene doesn't

show phenotype а doesn't mean it wouldn't in other conditions (or infecting an actual host). The plot showing the phenotype score (fitness) is particularly useful. Red points along the plot are genes known to be essential under these conditions



while yellow are known to be expendable. This will help you determine where to set the values. The last essential gene has a fitness score just >= than -4 so setting the phenotype score <= -2 would provide a pretty stringent search but still return more than 1000 genes. Try it. Do you get the expected results based on the number of genes returned?



- Can you find additional evidence that these genes are essential? One way is to use the analysis tools to assess biological process and go function. Are the results what you would expect?

1343 130	9 0	c)	
Gene Results	Genome View	Analyz	e Result	s

- Try adding columns to show additional data or intersecting these results with other queries, perhaps expression queries, to further assess this list. NOTE: this experiment was done in GT1 while all *T. gondii* functional data in ToxoDB is mapped to ME49 so an ortholog transform to ME49 is required before adding any additional functional studies.
- To do this, click on add step and select the Transform to orthologs option and select *T. gondii* ME49 to transform to.

My Strategies: New Opened	0		Add Step	×
Hide search strategy panel Genes) CRISPR Tel: Genes Step 1 Add Step	Run a new Search for Transform by Orthology Transform by Orthology Transform to Diskipt Filter by assigned Weight Transform to Athways Transform to Compounds	Genes Genomic Segments SNPs ORFs	Test Gene ID(s) User Comments Amontation, cuantion and identifiers Amontal control is a comment of the image of th	
1542 Genes from Step 1 Revise Strategy: CRISPR(5)	Add Step 2 : Transfor	m by Orthology	Protein features and properties Protein targeting and localization Function prediction Pathways and interactions Proteomics	
☐ ♥ Click on a number in this table to limit	Organism		immunology	
Cyclospora Cystoisosp	1 selected, ou	t of 26	mammonoia weospora Sarcocysus	530
	me49	×		
	 Toxoplasma Toxoplasma gondii Toxoplasma gondii MI add these clear these select all clea 	E49 elect only these r all	CRISPR 1542 Genes Step 1 Step 2 CRISPR	

- How many of these genes are upregulated in *in vivo* chronic stages of *T. gondii*?
 - Click on add step
 - Select the RNAseq searches under the Transcriptomics category
 - Find the experiment with chronic stages and run a search based on differentially expressed genes (DE).

			Add S	tep			×
Add Step 3 : RNA	Seq Evidence						
Filter Data Sets: chroni		Legend:	DE Differential Ex	FC Fold Change	P Percentile	SA SenseAntisense	
 Organism 	Data Set					Shoose a search	
T. gondii ME49 (filtered from 20 total entries)	Transcriptome du	iring acute or	chronic infection in mo	ouse brain (Pittman et	al.)	DE FC P	

• Intersect genes that are 2-fold upregulated in chronic stages compared to acute stages.

-	
Experiment	
Acute and chronic T.gondii infection of mouse. unstranded	1
Reference Sample	
acute infection 10 days p.i. chronic infection 28 days p.i.	Tg In murine mad
Ocomparator Sample	314 Genes
 acute infection 10 days p.i. chronic infection 28 days p.i. 	CRISPR 1542 Genes Orthologs 1683 Genes 25 Genes Add Ster
2 Direction	Step 1 Step 2 Step 3
up-regulated	
<pre>o fold difference >=</pre>	
2	
adjusted P value less than or equal to	
0.1	
Combine Genes in Step 2 with Genes in Ste	ар 3:
	2 Intersect 3 2 Minus 3
	2 Union 3 3 Minus 2
	2 Relative to 3, using genomic colocation

- What do these results look like? Do you find any interesting genes?
- 3. Finding oocyst expressed genes in *T. gondii* based on microarray evidence.



a. Find genes that are expressed at 10 fold higher levels in one of the oocyst stages than in any other stage in the "Oocyst, tachyzoite, and bradyzoite developmental expression profiles (M4) (John Boothroyd)" microarray experiment.



- **b.** <u>Add a step</u> to limit this set of genes to only those for which all the non-oocyst stages are expressed below 50th percentile ... ie likely not expressed at those stages. (*Hint*: after you click on add step find the same experiment under microarray expression and chose the percentile search).
 - Select the 4 non-oocyst samples.
 - We want all to have less than 50th percentile so set minimum percentile to 0 and maximum percentile to 50.



- Since we want all of them to be in this range, choose ALL in the "Matches Any or All Selected Samples".
- To view the graphs in the final result table, turn on the columns called "TgM4 OoTachyBrady Marray - Expr Graph" and "TgM4 OoTachyBrady Marray - %ile Graph" (inside the "T. gondii ME49 Oocyst, tachyzoite, and bradyzoite developmental expression profiles (M4) (Fritz and Buchholz et al.)" Microarray).



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

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

4a. Find all genes with evidence of protein phosphorylation in intracellular tachyzoites. Navigate to the Post-Translational Modification search. 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.)"

4b. Remove all genes with phosphorylation evidence from purified tachyzoites and the

Search for Genes	Identify Genes based on Post-Translational Modificati
expand all collapse all Find a search	Type of Post-Translational Modification
Text Gene models Annotation, curation and identifiers Genomic Location Taxonomy	phosphorylation site Experiments and Samples
Orthology and synteny	Filter list helow
Phenotype Genetic variation Epigenomics Transcriptomics Sequence analysis Structure analysis Protein features and properties Protein targeting and localization Pathways and interactions Proteomics Mass Spec. Evidence Post-Translational Modification Quantitative Mass Spec. Evidence	Toxoplasma gondii Toxoplasma gondii Toxoplasma gondii GTI Toxoplasma gondii GTA Toxoplasma gondii GTA
Immunology	Number of modifications is Greater than or equal to O Number of Modifications
	1
	Get Answer

phosphopeptide depleted fractions.

Hint: Use the Mass Spec Evidence search to access the tachyzoite and depleted fractions. Subtract (1 minus 2) these results from your first search.

9	Add Step	×
un a new Search for ransform by Orthology dd contents of Basket dd existing Strategy litter by assigned Weight ransform to Pathways ransform to Compounds	Add Step Seness Text Gene models Annotation, curate nand Annotation, curate nand Post-Translational Modification DRFs Annotation, curate nand DRFs Gene models Add Step Curanitative Mass Spec. Evidence Gene models Gene models Curanitative Mass Spec. DRFs Add Step 2: Mass Spec. Evidence Image: Specific Cond Add Step 2: Mass Spec. Evidence Image: Specific Cond Add Step 2: Mass Spec. Evidence Image: Specific Cond Add Step 2: Mass Spec. Evidence Image: Specific Cond Specific Cond Image: Specific Cond <th>Close</th>	Close

4c. 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.*



4d. Are any of these genes likely to be secreted? Hint: add a step searching for genes with secretory signal peptides.

My Stra	tegles: New	Opened (3)	All (3)	Basket	Public S	trategies	(14)	lelp															
🗉 Hide se	∃ Hide search strategy panel																						
(Genes) Post-Tra 2332 Sta	Anslation Genesi ep 1 Ste	Spec Sense Storr Storr Storr Fa Ca p 2 Step	Pep benes o 3	Add Step																	Strategy: Post	-Translati	Rename Duplicate Save As Share Delete
18 Gene Strategy	s from Step 3	evise al Mod																					
H T Ch	Cyclospi	a table to smithiter y	our results				Eimeria						Hammondia	Neospo	ra Saro	ocyatia	_			1	oxopiasma		
All	Ortholog C.cayetane	naia C.auia	Elecerculina	E brunetti	E faicifo	rmia	E.maxima	E mitis	E necetrix	E praecox	81	tenella	Hhammondi	N.caninu	am S.neu	ona (0				T	pondii (18)		
Results	CHOUPS Strain CHN HER	401 Strain Wien I	Houghton	Houghton	Bayer Hab 1970	erkorn	Weybridge	Houghton	Houghton	Houghton	Ho	strain ughton	strain H.H.34	Liverpo	ol SN3	SO SN1	ARI F	ou	GA82-2007-GAL- DOM2	GT1 M	AS ME49 RH RU	B TgCatF	RC2 VAND VEG p89
10	18 0	0	0	0	0		0	0	0	0		0	0	0	0	0	0	0	0	0 (18 0 0	0	0 0 0
4																							•
Gene Ret	Genome View	Analyze Results																					
Rows per j	page 20 ¥																				Download Ad	d to Basket	Add Columns
-	© Gene ID	° Transcript ID	Product Description	on O	o 🛋	° # Trans	cripts 🛛	0 📠	Non-C SNPs Strain	Coding All 😧	0	• SNP Strain	iyn/Syn Ratio All 📢	• •	NonSy SNPs	nonymo All Strair	^{US} 0	•	SNPs with Stop Codons All Strains	9 0	Synonymous SNPs All Strains	0 0	Total SNPs All Strains
-	TGME49_288370	TGME49_288370- 126_1	hypothetical p	protein		1			83			2.34			75				0		32		190
-	TGME49_288880	TGME49_288880- t26_1	hypothetical	protein		1			158			3.29			56				0		17		231
-	TGME49_243290	TGME49_243290- 126_1	hypothetical (protein		1			216			1.08			43				0		40		299
-	TGME49_205625	TGME49_205625- t26_1	hypothetical (protein		1			207			1.62			55			0			34		295
-	TGME49_259830	TGME49_259830- 126_1	diacylglycero domain-conta	l kinase cal aining prote	talytic in	1			139			0.61			14						23		176
-	TGME49_257595	TGME49_257595- t26_1	hypothetical	protein		1			131			2.32			130				0		55		317
-	TGME49_229680	TGME49_229680-	hypothetical	protein		1			28			0			0						5		33

4e. 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.

4f. 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 highest number of non-synonymous SNPs. Hint: you can sort the columns by clicking on the up/down arrows next to the column names.

5. Find and explore the metabolic pathway for glycolysis.

Navigate to the search page for Identify Metabolic Pathways based on Pathway Name/ID.

- Metabolic pathway and compound searches are available in the "Identify Other Data Types" section on the home page. You can find metabolic pathways based on the pathway name, genes involved in the pathway, or compounds involved in the pathway. Search for the glycolysis pathway using the Pathway Name/ID option.
- This search is equipped with a type-ahead function for choosing the metabolic pathway name. Begin typing glycolysis and then choose the pathway name from the list that appears.



a. Examine the Glycolysis / Gluconeogenesis pathway.



- The search takes you straight to the record page for the Glycolysis / Gluconeogenesis (ec00010) metabolic pathway from KEGG. The overview section of the record page contains an interactive graphical representation of the pathway. The pathway map and the legend can be repositioned.
 - A. Initial pathway view is zoomed out.
 - B. Zoom in to see more details including EC numbers and metabolite structures.

- C. Click on a metabolite structure to get additional information.
- D. Click on the EC number to get more info about the enzyme including links to retrieve all genes in the database assigned to this EC number.



- E. The drop-down menu under the heading "Paint Enzymes" allows you paint the pathway based on experiments or based on phyletic pattern.
- F. Painting pathway by experiment provides a graphical representation of experimental results. Click on the graph to see more details.
- G. Painting pathway based on phyletic pattern provides a graphical representation of phyletic distribution. Clicking on the phyletic pattern graphic provides additional information.
- What do the rectangles with numbers like 2.7.1.11 represent zoom in closer to see EC numbers?
- What is the difference between the rectangular nodes that are orange and those that are not?
- Why are some enzymes grouped?
- Click on the 2.7.1.11 node to open a popup with information about this enzyme.



- How many genes in the database matched this EC number?
- Try the link 'Show 41 gene(s) which match this EC Number'. Where did you end up? What do the 41 genes in the result list represent? Is 6-phosphofructokinase unique to *Toxoplasma*? Notice the two columns called "EC numbers" and "EC numbers from OrthoMCL". What do these columns represent?

🗄 Hide search strategy panel																								
(Genes)																			Stra	tegy:	EC N	umber	(2) *	×
EC 41 S	EC Number of General Biep 1 Add Step										hame Jicate ve As Share Helete													
															-		-				-		_	
41 Gen Strateg	es from ny: EC M	Step 1 Rev lumber(2)	risa		_																			
		Cyclospora	Cystoisospora	r your result	•		Eimer	ia					Hammondia	Neospora	Sarc	ocvstis						Toxopk	is <i>m</i> a	
All	Ortholog	C.cayetanensis	C.suis	E.acervulina	E.brunetti	E.falciform	nis E.maxima	E.mitis	E.necatrix	E.pra	есах і	E.tenella	H.hammondi	N.caninum	S.ne	urona (0)					Τ	gondii	(28)
Hesults	Groups	strain CHN_HEN01	strain Wien I	Houghton	Houghton	Bayer Haberko 1970	m Weybridge	Houghton	Houghton	Houg	phton H	strain Houghton	strain H.H.34	Liverpool	SN3	SO SN1	ARI	FOU	GAB2-2007- GAL-DOM2	GT1	MAS	ME49	RH	RUB Tgi
41	2	0	1	1	1	0	1	2	1	0	D	1	2	3	0	0	3	3	3	3	3	1	0	2
Gene R	esults i	Genome View	Analyze Result	s														Dow	mioad Add t	to Baske	ot	Add C	olumr	15
-	° Ge	ne ID	Transcript ID	o <u>I</u> i o	Irganism 🌘	• •	© Genomic L	ocation (G	iene) 😧	•	≎ Pro De	oduct escription	00	EC n	umbe	ers 😧	0		E fr	C num rom rthoM	ibers CL	0	0	
-	CSU	_008020	CSUI_008020-t	36_1 C. su	iis strain W	ien I	MIGC0100436	8:4319,2	70(-)		phos conta	phofructo aining pro	okinase domai Itein	n- _{N/A}					2 P	.7.1.11 hospho	(6- ofruct	okinase	0	
-	EAH	00020360	EAH_00020360 t26_1	E. ac	ervulina Ho	sughton	HG670489:10	12520,07	79(-)		phos	phofructo	okinase, putati	ve 2.7.1	.11 (6	-phosp	hofru	ctokin	ase) fr 1	.7.1.90 uctose -phosp	(Diph -6-ph hotra	osphat osphat nsferas	в в •е)	
	FRH	0022950	EBH_0022950-	E br	unetti Hour	nton	HG711375-29	792 45 30	17(+)		nhos	ohofnuctr	kinase putati	ve 271	11 (6	-ohosr	hofru	ctokin	2 (1)	.7.1.90	(Diph	osphat	e	

- Is this enzyme missing from some organisms? Do you think this is possible? What step can you add to confirm this? (*hint: try an ortholog transform*)

E niue	searchis	trategy panel																					_	
(Genes	4)														Str	rateg	y: [C Numberta	2)			٦,	r	×
EC	Number 11 Genes Step 1	Ortholo 104 Gen Step 2	95 85 2	id Step															-			Ri Duj Si	name dicate tve As Share Delete	
									<u> </u>	-											_		_	_
104 G Strate	enes fror gy: EC M Click on a	m Step 2 Re Number(2) number in this t	wise able to limit/filte	er your results																				
		Cyclospora	Cystoisospora				Eimei	ria				Hammondia	Neospora	Sarco	cystis						Тохор	lasm	1	
All	Ortholog	C.cayetanensis	C.suis	E.acenulina	E.brunetti	E.falciformis	E.maxima	E.mitis	E.necatrix	E.praecox	E.tenella	H.hammondi	N.caninum	S.neu ((rona 3)					7	T.gondii	(6	3)	_
Result	s Groups	strain CHN_HEN01	strain Wien I	Houghton	Houghton	Bayer Haberkorn 1970	Weybridge	Houghton	Houghton	Houghton	strain Houghton	strain H.H.34	Liverpool	SN3	SO SN1	ARI	FOU	GAB2-2007- GAL-DOM2	GT1	MAS	ME49	RH	RUB	Tgt
104	2	2	4	2	2	2	1	6	2	1	2	3	3	3	3	7	9	6	4	10	3	0	6	
Gene	Results	Genome View	Analyze Resul	ta per pege: 100	•							-					Dow	micad Add tr	o Bask	ot	Add (Colum	ms	
-	≎ Gen	e ID	0 Transo	cript ID	4	Organism	0 0	Genomic	Location	(Gene) 🕜	0	Produ Descrit	ct iption	0	¢	Inpu	t Orti	nolog(s) 🖸						
-	CSUL	000558	CSUI_00	0558-136_1	0	C. suis strain	Wien I M	/IGC01000	216:54,49	169,152(+)	phospho	fructokinas	a pfkii	EA	AH_O	00203	60,EBH_002	2950,	EMH_	00198	70,EI	ин_0	0€
-	CSUL	003878	CSUI_00	3878-136_1		C. suis strain	Wien I M	/IGC01001	748:39,44	945,742(+)	6-phospi	ofructokina	ise	E	AH_0	00203	60,EBH_002	2950,	EMH_	00198	70,EI	MH_0	.0€
-	CSUI_	006039	CSUI_00	6039-136_1	0	C. suis strain	Wien I M	/IGC01003	014:1512	2,441(-)		6-phospi	ofructokina	ise	E	AH_0	00203	60,EBH_002	2950,	EMH_	00198	70,EI	NH_0	OE

 Use your Browser's back button to return to the Glycolysis pathway record page and open the Paint Experiment menu. Choose the experiment "T. gondii ME49 Feline enterocyte, tachyzoite, bradyzoite stage transcriptome (Hehl, Ramakrishnan et al.)". Be patient while the graphs appear in place of the EC numbers.

1.1 Metabolic pathways	-	
Cytoscape Drawing NOTE Click on nodes for more info. Nodes highlighted in orange are EC	Experiment Selector Paint	×
File < Layout ∅ Paint Enzymes ∅ Clear all By Experiment By General + By General + By General - Enzymes ∅ - Enzymes ∅	(entero x) ♥ Transcriptomics TRMSeq ○ T, gondi ME49 Peline enterocyte, tachyzoite, bradyzoite stage transcriptome (Heh), Ramakrishnan et al.) Paint	Cytoscape JB EVYYYAN D glu

 Do you see enzymes that appear to be differentially regulated? Note you can click on the graph to see a larger image.



- Use the Paint Genera option to paint the pathway with orthologs from across Apicomplexa and Chromerida. Explore the results.

