RESEARCH

A sample article title

Jane E Doe[??*??] and John RS Smith[??,??]

*Correspondence:
jane.e.doe@cambridge.co.uk
[??]Department of Zoology, Cambridge, Waterloo Road, London, UK
Full list of author information is available at the end of the article

Abstract

First part title: Text for this section.

Second part title: Text for this section.

Keywords: Parasite, apicomplexa, RNA-seq, transcriptome, life-cycle, interaction; article; author

Introduction

Text and results for this section, as per the individual journal's instructions for authors.

Results

Text for this section ...

GO term enrichments in heatmap clusters

The annotations referred to here are inferred from orthologs in other *Eimeria* spp. or in *T. gondii*

Preparation for invasion in oocysts

The mRNA profile in the oocyst stage is mainly determined by highly abundant genes in cluster 4. Overrepresented GO-terms in this cluster are enriched by ortholog genes to peptidases, microneme localized proteins reported to be involved in invasion, genes associated with adhesion in protozoans (and with clotting in higher eukaryotes), and genes that are annotated to be involved in amino acid biosynthesis. Aminopeptidase N ('related' annotation) is the reported ortholog for three genes with abundant mRNAs in oocysts. In humans, this enzyme has been reported to cleave peptides bound to major histocompatibility complex, MHC, II (UniProt reference if we want to keep this... but does any secretion happen from oocysts...? Or is this too far-fetched to be interesting?).

A Thrombospondin type 1 domain-containing protein ortholog is highly abundant in cluster 4 (high abundance in oocysts). Thrombospondin type 1 domains have been reported in *E. tenella* microneme localizing proteins, MIC, e.g. MIC4 (Tomley01) In *E. tenella* MIC4 mRNA was reported in sporozoites where it localizes to the apical end, and in late schizonts and late oocyst stages, when sporozoites are forming. (Tomley01). For the same gene, the *T. gondii* annotation is Sushi domain-containing protein, which is also the ortholog annotation of another gene in this cluster. In the related malaria parasite *P. falciparum* the apical sushi protein, ASP, (which has a sushi domain) localizes to micronemes in merozoites

but not other stages (OKeeffe05). Limulus clotting factor C, Coch-5b2 (Cochlin) and Lgl1, LCCL, (syn. F5/8 domain) domains are associated discoidin lectin domains and thereby with adhesion. (Pfam entries for 'LCCL domain' and 'Discoidin domain', May 2016). Taken together, this indicates that the thrombospondin and sushi-domain genes (EfaB_MINUS_4114.g412 and EfaB_PLUS_1425.g183) are involved in sporozoite invasion in *E. falciformis* and that the mRNAs are transcribed and available before excystation. A role in merozoite re-invasion in *E. falciformis* is not indicated by our data. The LCCL domain annotation and thrombospondins role in higher eukaryotes also indicates that adhesion or preparation for adhesion is important in oocysts. We suggest (speculate...?) that the thrombospondin annotated ortholog and the LCCL domain-containing protein (EfaB_MINUS_11233.g986) are involved in cell adhesion in *E. falciformis*.

Amino acid biosynthesis in oocysts

High abundance of aminotransferase mRNAs indicate amino acid biosynthesis or preparation for the same in oocysts (cluster 4). We identify D-3-phosphoglycerate dehydrogenase and alanine dehydrogenase orthologs, which are enzymes contributing to L-serine and L-alanine production, respectively. A putative *Eimeria* spp. cystathionine beta-synthase, CBS, in this cluster also indicates de novo cysteine production. Alkyl sulfatase mRNA is also abundant in oocysts. Generally, this enzyme enables an organism to exploit organic sulfur to produce and incorporate inorganic sulfur into the amino acids cysteine and methionine, when no inorganoc sulfur is available.

'Embryonic development' Nicalin 1, patched family protein (hedgehog)

Oocysts contain mRNA for fatty acid catabolism

MmgE/PrpD is overrepresented in oocysts. The enzyme is important for propionate catabolism in the 2-methylcitric acid cycle and has been shown to be used by the intestinal intracellular bacterium *Salmonella typhimurium* to generate pyruvate (Horswill99). Propionate is one of two most abundant small-chain fatty acids in the gut along with butyrate. Both fatty acids are largely produced as degradation products from food by commensal bacteria (Sun13). Sharing the intestines as a niche with *S. typhimurium* it is possible that also *E. falciformis* uses Mmg/PrpD to exploit available propionate for pyruvate production.

Oocyst highly abundant mRNAs are downregulated in sporozoites

Interestingly, the genes described above which are thought to be involved in amino acid biosynthesis and invasion are highly abundant in oocysts but are underrepresented in schizont stages (day 3 and day 5 samples) and even in sporozoites. An average abundance was detected on day 7 for these genes, indicating a role in either gametes or early oocyst formation. This pattern supports the suggestion that these specific mRNAs (cluster 4) for invasion and biosynthetic processes are prepared (and possibly expressed) in the oocyst stage but are no longer detectable in the cell at the timepont when the protein is assumed to be in use (sporozoites and merozoite stages). Therefore, correlating mRNA prevalence with biological function at the timepoint when mRNAs are detected must be done with care.

Down in sporozoites and oocysts –¿ cluster 3.....
Oocysts: profile for 3, 5, 6:
Sporozoites: 3 and 5
Day 7: 5 and 7
Specific genes.... Enolase 2, encoded by Eno2, is among the downregulated genes in oocysts and sporozoites. In *T. gondii* the paralog Eno1 is strongly associated with the cyst (bradyzoite) stage and Eno2 is associated with tachyzoite stages (kibe05). It is therefore expected that this mRNA is underrepresented in oocysts and our data also show that the same is true in sporozoites for *E. falciformis*. (TK: If important we could look specifically for Eno1 in cluster 4).

Down in sporozoites and oocysts –¿ cluster 3.....
Motility-related mRNAs indicate gamete development on day 7
Two clusters contain genes with mRNAs highly abundant on day 7 p.i; cluster 1 and 2. Dynein, kinesin and tubulin are annotations highly represented among orthologs of genes in both these clusters. The annotations indicate an important role for motility at this timepoint, probably reflecting development of microgametes. In addition, in cluster 2, there are two 'EF-hand domain containing proteins' annotations as well as caltractin, centrin-1, and troponin annotations. Caltractin and centrin-1 are associated with the centrosome and structure and function of microtubuli in mammals, and troponin is linked to muscle function (UniProt). Also potentially linked to motility is the occurence of growth arrest specific protein 8, Gas8, which in the mouse has been reported to be highly expressed in the testes and important for mouse sperm function (Yeh02).

Other genes among the 38 indicate carbon fixation (glycolysis/gluconeogenesis) or conversions of nucleoside phosphates. In addition, a Ras family protein, RNA polymerase II transcription initiation factor and Sec23 and Sec24 were among orthologs identified in *E. falciformis* cluster 2.

In cluster 1, carbon metabolism genes are represented by 6-phosphogluconate dehydrogenase and glycogen phosphorylase family protein 1. UDP-glucose 4-epimerase and amiloride-sensitive amine oxidase are reported as upregulated in gametocytes in *E. tenella* by RNA-seq (Walker15) and suggested by those authors to play a role in cyst wall synthesis.

Microneme proteins highly expressed on day 7 p.i.
Unintuitively for a protozoan organism, seven out of eight GO biological process terms in cluster 1 are associated with wound healing and blood coagulation. An explanation is offered by some of the orthologs to the three *E. falciformis* genes responsible for these terms. In protozoa, e.g., other *Eimeria* spp. and *Toxoplasma gondii* orthologs are annotated as 'Micronemal protein MIC4, related' (E. tenella) and more generally for several other protozoa, 'PAN domain containing proteins'. The PAN domain is found in the plasminogen/hepatocyte growth factor family and in coagulation factor XI family (REF), explaining why terms related to blood coagulation are enriched by these genes. Later publications on *T. gondii* (Marchant12) also associate PAN domains and proteins in apicomplexan parasites with micronemes and therefore invasion. In our case, this is peculiar, since the enrichment appears

on day 7 p.i.. A possible role at this timepoint is suggested by work on the fungi *Sclerotinia sclerotiorum* where Yu et al. reported an important role for PAN domain proteins in cell wall integrity (Yu12). This role for MIC proteins has to our knowledge not been investigated in apicomplexan parasites. The PAN domain domain has also been reported to be common in nematodes such as *Caenorhabditis elegans*, however the function is not understood. (Thordai99) The other two GO terms in the cluster of day seven upregulated genes are DNA replication and DNA replication initiation, which most likely reflects late stage schizogony or gamete formation. Six genes contribute to this enrichment and orthologs are either annotated as DNA replcation licencing factors, DNA polymerases or minichromosome maintenance proteins 2/3/5/7, Mcm2/3/5/7.

Gene and sample patterns by hierarchical clustering
Samples (columns) cluster into two major clusters where day 7 p.i. samples form one group distinct from other samples. In the second group, oocsts and sporozoites are distinct and sporozoites cluster most closely with day 3 and 5 p.i. samples. Day 3 and 5 p.i. samples also cluster into two groups, of which one contains all NMRI day 5 p.i. samples. Apart from this, the two day 3 and 5 p.i. sample clusters have no obvious patters.

For gene clusters (rows), the two groups with high mRNA abundance on day 7 p.i. (cluster 1 and 2) do not cluster most closely with each other, but with the cluster for high mRNA abundance in oocysts (cluster 1 association) and with the cluster for high mRNA abundance in sporozoites (cluster 2 association).

Discussion

In our analysis we demonstrate which biological processes are dominant in different life cycle stages of *E. falciformis* in the mouse. The RNAseq transcriptome provided here allows for detailed analysis of genes involved in those processes, providing candidates for life stage specific marker in *Eimeria* spp. research.

. . . . . .

Methods

. . .

Sub-heading for section

Text for this sub-heading . . .

Sub-sub heading for section

Text for this sub-sub-heading . . .

Sub-sub-sub heading for section Text for this sub-sub-sub-heading . . . text text text text text text text text text text text text text text text text text text

Competing interests
The authors declare that they have no competing interests.

Author's contributions
Text for this section . . .

Acknowledgements
Text for this section . . .

Figures

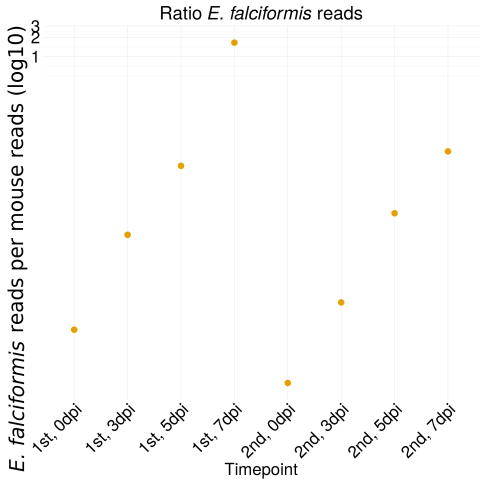

Figure 1 Parasite read sequences per mouse read sequences for each experimental condition. The more advanced the infection is, the higher the ratio of parasite reads is, reflecting parasite replication between day zero and day seven. .....compare 1st and 2nd ..... patterns between Rag and NMRI/C57BL/6? ..... Values are mean of replicates (n=2, if * n=1) on log10 scale. Each sample (replicate) consists of mRNA from three different mice.

Tables

Table 1 Sample table title. This is where the description of the table should go.

	B1	B2	B3
A1	0.1	0.2	0.3
A2	...	..	.
A3	..	.	.

Additional Files

Additional file 1 — Sample additional file title
Additional file descriptions text (including details of how to view the file, if it is in a non-standard format or the file extension). This might refer to a multi-page table or a figure.

Additional file 2 — Sample additional file title
Additional file descriptions text.

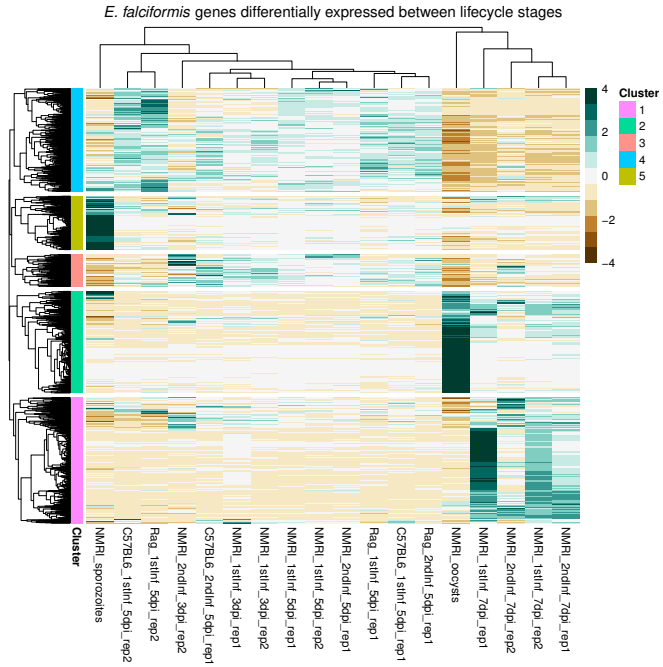

Figure 2 Parasite genes with different mRNA abundance between samples, sorted into five clusters by hierarchical clustering (method = complete, distance = euclidean). *E. falciformis* samples from seven days p.i. cluster together (NMRI mice only). These samples have distinct mRNA abundance patterns in all gene clusters, although more pronounced in clusters 1 (up) and 4 (down). Distinct groups of genes also define sporozoites (cluster 5, up) and oocysts (cluster 2, up). mRNA profiles on days three and five p.i. from all three mouse strains cluster together. These samples are distinct from oocysts, NMRI day 7 p.i., and sporozoites, however closest to the latter. On scale bar, 0 is mean mRNA abundance for each gene. Up and downregulation is standard deviations from mean.

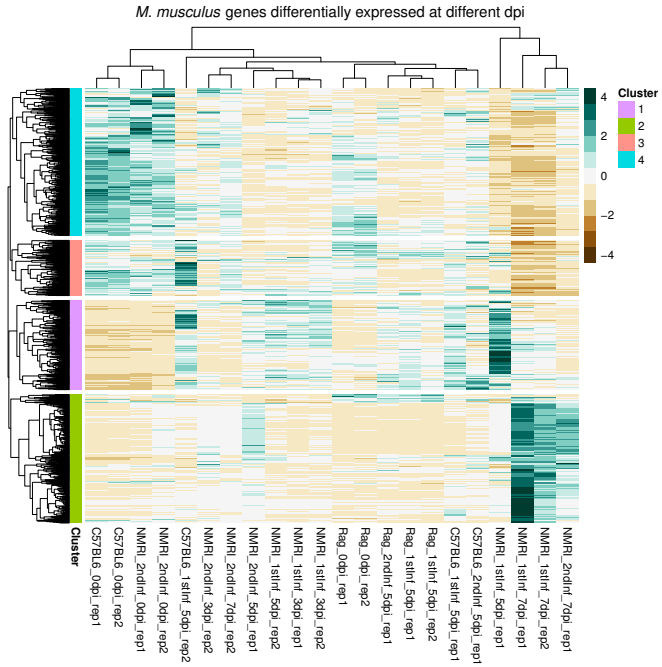

Figure 3 Mouse genes with different mRNA abundance between samples, sorted into four clusters by hierarchical clustering (method = complete, distance = euclidean). Three out of four samples from day 7 p.i. (NMRI only) cluster together. These samples are characterized by genes in cluster 2 (up), 3, and 4 (down). NMRI and C57BL/6 uninfected samples cluster together (left), defined by clusters 3, 4 (up), and 1 (down). Rag1-/- samples cluster together. These samples share a weak downregulation of most genes in cluster 2 as well as upregulation of a small group genes in the same cluster. Uninfected Rag1-/- samples are separated from infected ones with distinct profiles in clusters 1, 3, and 4. On scale bar, 0 is mean mRNA abundance for each gene. Up and downregulation is standard deviations from mean.