

iiiiii HEAD ===== llllllll ce7a348a5c7d0ee80f8fe6f44b4b32d973ca1d8f
iiiiii HEAD ===== llllllll ce7a348a5c7d0ee80f8fe6f44b4b32d973ca1d8f

RESEARCH

Dual host-parasite transcriptomes of apicomplexan *Eimeria falciformis* and its natural mouse host

Totta Kasemo, Simone Spork, Christoph Dieterich, Richard Lucius, Emanuel Heitlinger??

Abstract

First part title: Text for this section.

Second part title: Text for this section.

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

?? Correspondence:
?? Department of Biology,
Humboldt-Universitat zu Berlin,
Philippstr. 13, Haus 14, 10115
Berlin, Germany
?? 0
Full list of author information is
available at the end of the article

Introduction

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

Results

Text for this section ...

We found xx *E. falciformis* genes of the total predicted yy, which were expressed in at least one developmental stage measured. Of the total reported mouse genes aaa, hhh were expressed on at least one timepoint during *E. falciformis* infection and fff were expressed in controls.

1 Experimental overview

Replicate average of transcripts as order of magnitude for *E. falciformis*/mouse are indicated in Table.... Columns in table represent mouse strains used in infection experiments. Rows represent timepoints post infection plus oocyst and sporozoite samples. The upper part of the table shows data for first infection, and oocyst and sporozoite data. The lower part shows data for challenge infection. Averages were calculated after sample exclusions (see Methods). For exact values, see Table 1.

Table 1 Experimental overview with number of read for parasite/mouse indicated.

<i>Day, 1st infection</i>	NMRI	C57BL/6	Rag1-/-
0 (control)	$10^4 / 10^8$	$10^2 / 10^8$	$10^2 / 10^7$
3	$10^5 / 10^8$	NA	NA
5	$10^7 / 10^8$	$10^5 / 10^7$	$10^5 / 10^7$
7	$10^8 / 10^7$	NA	NA
Oocysts	$10^8 / \text{NA}$	NA	NA
Sporozoites	$10^7 / \text{NA}$	NA	NA
<i>Day, 2nd infection</i>			
0 (control)	$10^3 / 10^8$	NA	NA
3	$10^4 / 10^8$	NA	NA
5	$10^4 / 10^8$	$10^5 / 10^7$	$10^5 / 10^7$
7	$10^6 / 10^8$	NA	NA

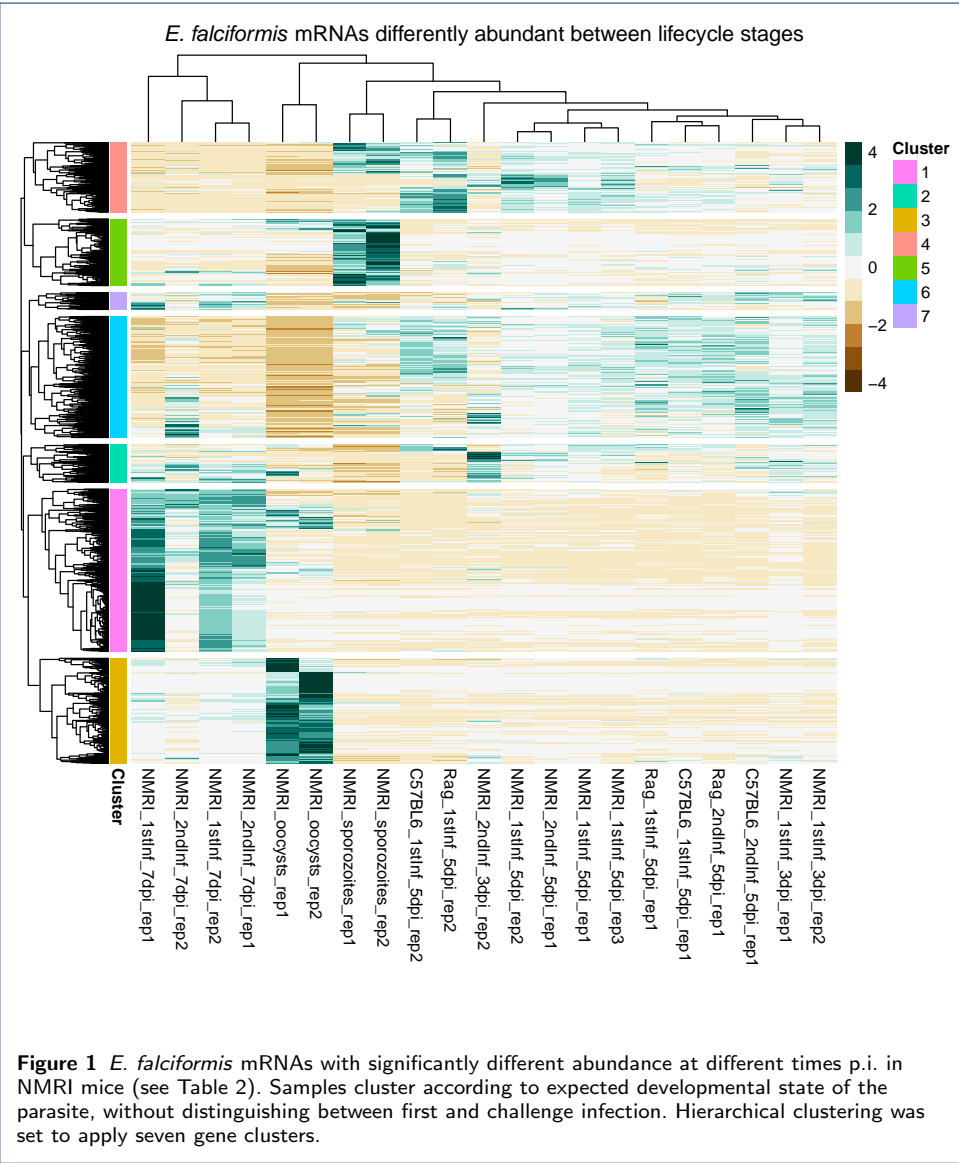
2 mRNA abundance differences between different experimental groups

Table 2 mRNA abundance differences between different experimental groups.

<i>Day post infection comparisons</i>	<i>Ef</i> genes different (FDR \leq 0.01)	Mouse genes different (FDR \leq 0.01/0.05)
NMRI 0 vs NMRI 3	NA	274
NMRI 0 vs NMRI 5	NA	1736
NMRI 0 vs NMRI 7	NA	2802
NMRI 3 vs NMRI 5	111	1
NMRI 3 vs NMRI 7	1385	1407
NMRI 5 vs NMRI 7	1895	873
C57BL/6 0 vs C57BL/6 5	NA	914
Rag1-/- 0 vs Rag1-/- 5	NA	45
<i>Day post infection, parasite relevant comparisons</i>		
Oocysts vs NMRI 3	3310	NA
Oocysts vs NMRI 5	3605	NA
Oocysts vs NMRI 7	3085	NA
Oocysts vs sporozoites	3421	NA
Sporozoites vs NMRI 3	1663	NA
Sporozoites vs NMRI 5	1605	NA
Sporozoites vs NMRI 7	2473	NA
<i>First and second infection comparisons</i>		
NMRI 3 1st vs NMRI 3 2nd	0	5
NMRI 5 1st vs NMRI 5 2nd	0	1
NMRI 7 1st vs NMRI 7 2nd	0	902
C57BL/6 1st vs C57BL/6 2nd (day 5)	0	mouse
Rag1-/- 1st vs Rag1-/- 2nd (day 5)	0	mouse
<i>Mouse strain comparisons</i>		
NMRI vs C57BL/6	22	NA*
NMRI vs Rag1-/-	0	NA*
C57BL/6 vs Rag1-/-	0	356

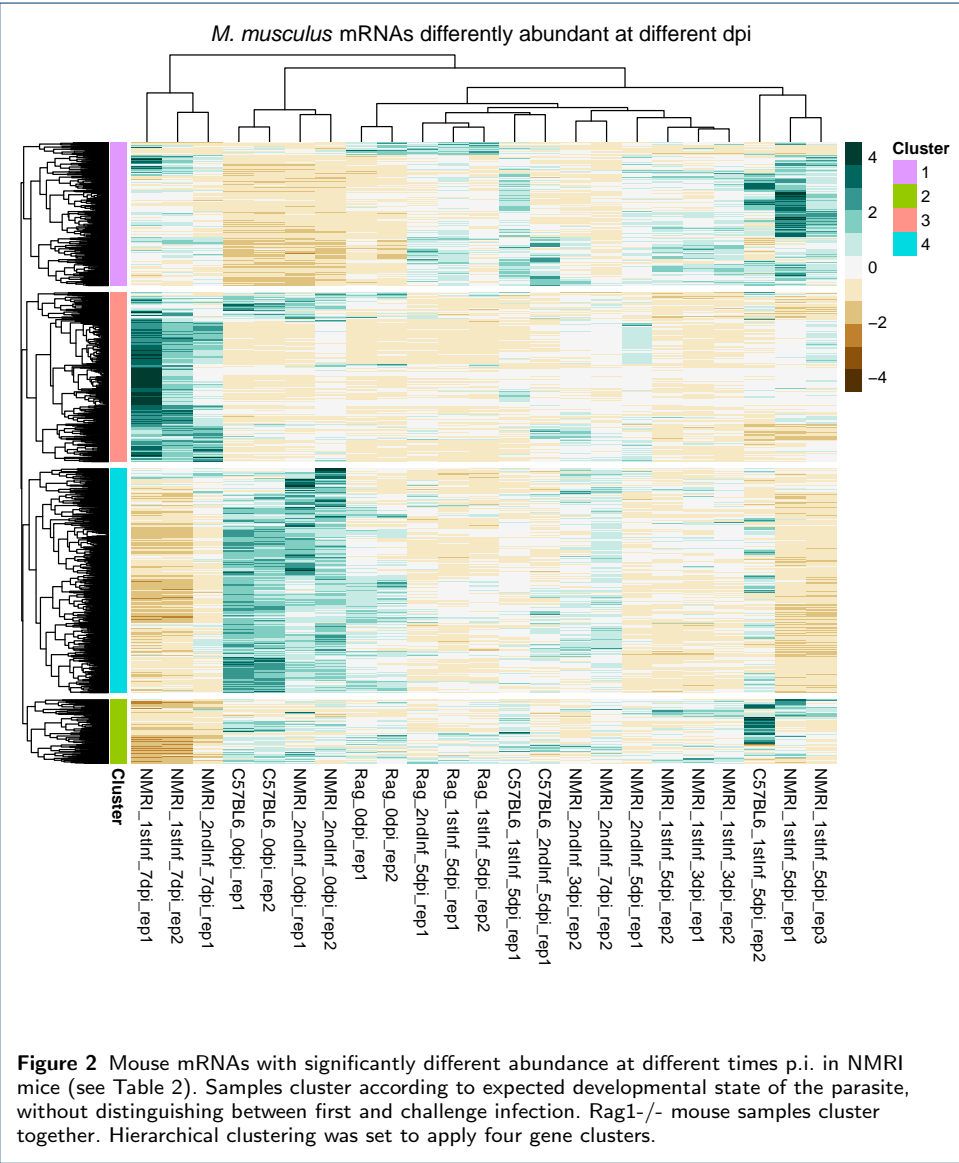
Table 2 lists pairwise sample comparisons and number of genes with differently abundant mRNAs per comparison. "NMRI" followed by number indicates day post infection (e.g. NMRI 3 = NMRI mouse on day 3 p.i.). Genes with Benjamini-Hochberg corrected p-values ≤ 0.01 as implemented in edgeR are included. NAs are missing samples or not applicable for the species. NA* is due to missing NMRI day 0 sample from first infection. For *E. falciparum* no mRNAs are significantly differently abundant between first and second infection, whereas in mouse there are some differences, especially on day 7 p.i. However, upon hierarchical clustering of these genes (union of 500 with lowest FDR), samples do not cluster according to first/second infection (Figure SX ?). In comparisons between mouse strains, 22 genes are detected to be different in the parasite data. The difference is however not between Rag1-/- mice as expected, but between NMRI and C57BL/6. For mouse....

2.1 Life cycle analysis... *E. falciformis*



We applied hierarchical clustering on *E. falciformis* mRNAs which were differently abundant at different timepoints p.i., including comparisons with oocysts and sporozoites (union of 500 with lowest FDR). Samples from day 7 p.i. form a separate cluster (NMRI mice only), not distinguishing between first and challenge infection. For these samples, the pattern is most pronounced in gene cluster 1 (up), and 4 and 6 (down). Within cluster 1, the second replicate from the challenge infection has a deviant profile (compare mouse data in Figure 2). Distinct groups of genes also define sporozoites (cluster 5 up) and oocysts (cluster 3 up, cluster 6 down). mRNA profiles on days 3 and 5 p.i. from all three mouse strains cluster together. These samples are distinct from oocysts, NMRI day 7 p.i., and sporozoites. Sporozoites cluster closer with days 3 and 5 than with oocysts or day 7.

2.2 Parasite development reflected in mouse data



When clustering mouse mRNAs which were different in time p.i. comparisons in Table 2, three out of four samples from day 7 p.i. (NMRI only) cluster together. These samples are characterized by gene cluster 3 (up) as well as 2 and 4 (down). The fourth day 7-sample (challenge infection, second replicate) clusters with day 3 and 5 samples. Upon visual inspection of this sample in gene cluster 4, it displays a profile similar to non-infected mice. The same sample is abnormal in the parasite profile (Figure 1). NMRI and C57BL/6 uninfected samples (0dpi) cluster together, defined by gene clusters 4 (up), and 1 (down). Rag1^{-/-} day 0 samples are however less pronounced in gene cluster 1. Day 3 and 5 samples cluster together, with Rag1^{-/-} forming a separate group and Rag1^{-/-} non-infected most distant in this sample cluster.

GO term enrichments in heatmap clusters

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

Data overview

In total, xx reads were sequenced. For each sample, mouse read numbers were 10^7 - 10^8 Quality filters applied resulted in xx reads being removed from further analysis

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 inorganic 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 timepoint 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 occurrence 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 replication licensing 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, oocysts 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 patterns.

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.

.....

3 Methods

3.1 Mice and infection procedure

Three strains of mice were used in our experiments: NMRI (Charles River Laboratories, Sulzfeld, Germany), C57BL/6 (), and Rag1-/- on C57BL/6 background (gift from Susanne Hartmann, FU?). Animal procedures were performed according to the German Animal Protection Laws as directed and approved by the overseeing authority Landesamt fuer Gesundheit und Soziales (Berlin, Germany). Animals were infected as described by Schmid et al., (Schmid12), but tapwater was used instead of PBS for administration of oocysts. Briefly, NMRI mice were infected two times, which will be referred to as first and second infection. For the first infection, 150 sporulated oocysts were administered in 100 L by oral gavage. During the first infection of 60 mice, all animals were weighed every day. On day zero, before infection, as well as on day three, five and seven days post infection, dpi, caeca from 3-4 sacrificed mice per time point were collected. Epithelial cells were isolated as described in Schmid et al.(schmid12). For challenge infection, mice recovered for four weeks before second infection. Recovery was monitored by weighing and visual inspection of fur. For the second infection, 1500 sporulated oocysts were applied by oral gavage. Three mice were used as non-second infection control, referred to as day 0, second infection.

3.2 Oocyst purification for infection and sequencing

Sporulated oocysts were purified by flotation from feces stored in potassium dichromate and administered orally in 100 uL tapwater. One *E. falciformis* isolate, *E. falciformis* Bayer Haberkorn 1970, was used for all infections and parasite samples. The strain is maintained through passage in NMRI mice in our facilities as described elsewhere (schmid12).

3.3 Sporozoite isolation

Sporozoites were isolated from sporocysts by excystment. For this, sporocysts were incubated at 37C in DMEM containing 0.04% tauroglycocholate (MP Biomedicals) and 0.25% trypsin (Applichem) for 30 min. Sporozoites were purified by the method of Schmatz et al (schmatz-).

3.4 RNA extraction

Total RNA was isolated from infected epithelial cells, sporozoites and sporulated oocysts using Trizol according to the manufacturer's protocol (Invitrogen). High quality *what is the meaning of 'high quality' here?* RNA was used to produce an mRNA library using the Illumina's TruSeq RNA Sample Preparation guide. *stolen from genome paper* Sporozoites were stored in 1 mL Trizol until RNA-isolation. Total RNA was isolated using the PureLink RNA Mini Kit (Invitrogen) and reverse transcribed into cDNA.

3.5 Sequencing, sequence quality assessment and alignment

cDNA samples were sequenced by either GAIIX or Illumina Hiseq 2000 as specified in SI xx (both unstranded). A fastq_quality_filter (FASTQ-toolkit, version 0.0.14, available at https://github.com/agordon/fastx_toolkit.git) was applied to Illumina Hiseq 2000 samples after replacing "N" bases by "." annotation. A phred score of 10 was applied. We further set $q = 60$. These settings require that nine out of ten bases or more are correct in at least 60% of the bases for each read.

3.5.1 Alignment and reference genomes

We used the published *Mus musculus* mm10 assembly (Genome Reference Consortium Mouse Build 38, GCA_000001635.2) as reference genome including annotations for mouse data. The *E. falciformis* genome (Heitlinger14) was downloaded from ToxoDB (Gajria07). For the alignment, the mouse and parasite genome files were merged into a dual reference genome, and files including mRNA sequences from both species were aligned against the dual reference genome using TopHat2 (version 2.0.14, Trapnell09) with -G specified, and a Bowtie2 (version 1.1.2, Langmead12) index of the dual genome. Single-end and pair-end sequence samples were aligned separately with library type 'fr-unstranded' specified for pair-end samples. Import into R was enabled by the R package Ballgown, which requires bam files to be processed by Tablemaker (Frazee15), in our case used with -qW -G specified. Tablemaker in turn makes use of Cufflinks (version 2.1.1, Trapnell10).

3.6 Differential mRNA abundance, data normalisation and sample exclusions

Complete scripts are available at [git@github.com:derele/Ef_RNAseq.git](https://github.com:derele/Ef_RNAseq.git). Count data was analysed in R, using the upperquartile normalisation method as implemented in edgeR (version 3.14.0). Briefly, genes with zero coverage in all samples (libraries) are removed and normalisation factors are calculated for the 75% quantile for each library. This normalisation assumes a negative binomial distribution of read densities. Two samples contradicted the assumption (parasite data) and both mouse and parasite data from these samples were excluded from further analysis: NMRI_1st.3dpi_rep1 and NMRI_2nd.5dpi_rep1 (SI ...). We used a the edgeR pipeline which estimates dispersion values for each gene (provided that a design matrix is defined). The method fits data for each gene to a negative binomial distribution using a generalized linear model (GLM) with the function glmFit() and glmLRT().

..... include outcome of MDS analysis here?

3.7 Differentially abundant mRNAs used for hierarchical clustering

A selection of differently abundant mRNAs are used for hierarchical clustering of *E. falciformis* life cycle relevant genes. In each comparison (see Table 3), the union of the 500 differentially abundant mRNAs with lowest FDR are selected. In the next step, the 500 mRNAs from each comparison (or less) are joined. For *E. falciformis* time p.i. comparison, this resulted in 4935 genes of which only the 1618 unique genes were selected. The 22 genes in the NMRI vs C57BL/6 comparison are not included in the *E. falciformis* life cycle analysis. The same comparisons for mouse yielded 8052 genes in total and 1313 unique ones. In heatmaps, all samples, i.e. also samples which themselves did not have any significantly different mRNAs according to our selection, were included in hierarchical clustering. Scale bar in heatmaps show 0 as mean mRNA abundance for each gene (row). Up (green) and down-regulation (brown) denote number of standard deviations from 0, i.e., row mean. Hierarchical clustering was performed using with Euclidean distances, using complete linkage ('complete', R package base).

Table 3 Genes used for hierarchical clustering of mRNAs differently abundant depending on time p.i..

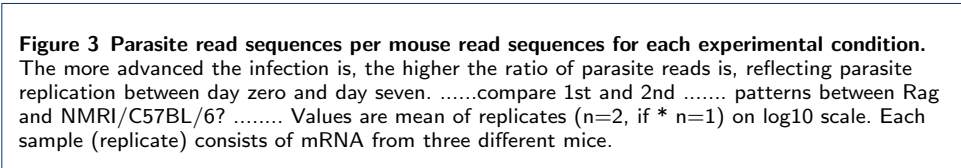
<i>Data description</i>	<i>E. falciformis</i> genes	Mouse genes
Sum of 1st infection NMRI sample differences (including oocysts and sporozoites if appl.)	4935	8052
Used in hierarchical clustering (heatmap)	1618	1313

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

Author's contributions
Text for this section ...

Acknowledgements
Text for this section ...

Figures



Tables

Table 4 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.