RESEARCH

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

Totta Kasemo[1], Simone Spork[1], Christoph Dieterich[2], Richard Lucius[1] and Emanuel Heitlinger[1,3*]

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:
emanuel.heitlinger@hu-berlin.de
[1] Institute of Biology, Humboldt-Universitat zu Berlin, Philippstr. 13, Haus 14, 10115 Berlin, Germany
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

A dual transcriptomics experiment

We performed mRNA sequencing of caecum tissue from mice infected with the apicomplexan parasite *E. falciformis*. Oocysts and sporozoites were included as "environmental" stages and processed "*in vitro*". To follow the life cycle of the parasite, we compare different time-points after infection.

We additionally used different mouse strains which display different immunocompetence in infection trials measured by oocyst output to assess the influence of host immunocompetence on parasite development (figure 1a,b). In these experiments we analysed mice lacking the Recombination activating gene 1, Rag1, and compared them to the parental C57BL/6 mouse line. Rag1 knock-out mice lack adaptive immune response due to the absence of mature B and T lymphocytes. All three mouse strains were used for infections of both naive mice and previously infected animals (onward referred to as "challenged").

Basic phenotyping showed differences in oocyst numbers between the immunocompetent (C57BL/6) and immunodeficient (Rag1-/-) host strain in infection of naive mice and in challenge infections. Immunocompetent NMRI mice were infected with a higher dosis of sporulated oocysts, and a drastic reduction of oocysts in faeces was seen in challenged mice compared to naive mice (Figure 1a). Oocyst numbers in faeces peaked on days 8-9 and all mice had cleared the infection by day 14. Development of *E. falciformis* intestinal stages is reflected by the percentage of parasite reads sequenced per time-point post infection (Figure 1b) and this was confirmed by quantitative reverse transcription PCR (RT qPCR) of parasitic 18S (Figure 1c).

We thus use dual RNA-seq to analyse the life cycle of *E. falciformis* under the influence of host immune responses at early stages of infection. We used an experimental design which allows to compare infections at 5 days post infection (dpi) for all experimental conditions (NMRI, C57BL/6 and Rag1-/- mouse strains in naive and challenge infection). In addition, additional time-points 3 dpi and 7 dpi were analysed from NMRI mice (Figure 1d).

Parasite and host dual transcriptomes can be assessed in parallel

In this RNA-seq experiment, each replicate sample was enriched for caecum epithelial tissue which was pooled from three mice. mRNA was extracted and sequencing libraries prepared. Two biological replicates were used for all but two conditions (with one and three replicates, respectively). These libraries were sequenced on several lanes of Illumina GAIIX (13 samples) and HiSeq machines (14 samples) and mapped to both mouse and parasite genomes simultaneously to avoid spurious assignments of reads in ultra conserved genomic regions. As samples and individual replicates were sequenced in batches to different depth and using different instrumentation (Table 1) we performed quality controls (additional files xyz). These confirm the absence of batch effects influencing analysis and quality of results.

Total numbers of sequenced reads as well as reads mapped to either the *E. falciformis* genome or the mouse genome are indicated in Table 1 for all replicates. The number of total read mappings for individual replicates ranged from 11,479,260 (sample NMRI_sporozoites_rep2) to 145,033,326 (NMRI_2ndInf_5dpi_rep1). At the latest time-point samples, 7 dpi, the overall mRNA output of the sampled caecum tissue is dominated by parasite material with proportional mRNA abundance of 77% (in NMRI_1stInf_7dpi_rep2) and 92% (in NMRI_1stInf_7dpi_rep1). (Figure 1b).

Exclusion of samples with uncertain infection status

A maximum of 86% (sample NMRI_1stInf_7dpi_rep1) and a minimum of 0.0326% (sample NMRI_2ndInf_3dpi_rep2) of mapped reads could be assigned to the *E. falciformis* genome in samples considered infected (Table 1). We excluded samples NMRI_2nd_3dpi_rep1 and NMRI_2nd_5dpi_rep2 due to low parasite contribution to the overall transcriptome. Technically, this exclusion made it possible to obtain read counts in agreement with a negative binomial distribution (see additional file x). It is also likely that the number of reads in the excluded samples would have been insufficient to fully normalise these datasets to those with the highest parasite contributions. From a biological point of view, both excluded are samples from challenge infection and we assume that the infection was cleared or reduced to a non-detectable level. One sample (NMRI_1stInf_0dpi_rep1) was excluded because the uninfected control showed unexpected mapping of reads to the *E. falciformis* genome. We consider the three excluded samples to display an uncertain state of infection.

The mouse transcriptome changes upon E. falciformis infection

Statistical testing for differential expression between infected and uninfected mice revealed changes in mRNA abundance becoming more pronounced (involving more

genes) at later time-points post infection (especially at 7 dpi; Table 2). For mouse, 325 mRNAs were considered differently abundant (DA; FDR1%) between controls and 3 dpi, 1,804 mRNAs between controls and 5 dpi and 2,711 mRNAs between controls and 7 dpi. This lead to a combined set of 3,453 genes responding to infection (Figure 2bi). DA mRNAs early in infection (3 dpi and 5 dpi) were not a pure subset of genes DA later in infection (7 dpi). Instead, the transcriptional profile of the mouse changes throughout the infection. Changes between controls and 7 dpi were in agreement with previously published microarry data. Fold-change data obtained from *E. falciformis* infected mice at 6 dpi on Agilent microarrays (Schmidt et al 2012) or 7 dpi (our RNAseq data) against uninfected controls show a strong correlation (Spearman's $\rho = 0.74$; $Figure 2a$). $Considering both biological (e.g. exact time-points) and technical differences between the two methods this comparison confirms the adequacy of using dual RNA-seq for assessing the host transcriptome.$

Epithelial responses depend on mouse immune status

To further validate the pattern of a distinct response in early infection we performed hierarchical clustering on the (union of) mouse genes DA between different time-points post infection (Figure 2c). By hierarchical clustering of mouse samples, three main sample clusters formed, "main clusters" (dendrogram of columns at top of Figure 2c). Immune deficient Rag1-/- mice cluster with control samples. There is no clear distinction between infected and non-infected Rag1-/- samples, which confirms the immune deficiency phenotype in these mice.

We identify a group of genes (cluster 4, Figure 2c) which change the transcriptional profile upon infection (all time-points) in immune competent mice only. In immune compromised Rag1-/- another group of genes (cluster 3, figure 2c) display the "infection profile" of immune competent mice also in control samples, indicating that these are mRNAs kept at a low abundance in a T and B cell dependent manner. Hence, these genes in some way depend on functional T and B cells also in controls (cluster 3) at the earliest time-point post infection (3 dpi, cluster 4). Gene Ontology, GO, terms enriched and regulated in cluster 3 are, e.g., "lipid modification" and "lipid metabolic process" and "protein intraciliary transport". Other enriched categories in cluster 3 are terms for regulation of different metabolic processes, for blood coagulation as well as terms containing "spinal cord", "axon" or "neuronal" regulation or development. (SI file x). We suggest that these are processes which are all intrinsically different in Rag1-/- mice compared to immune competent mice. The genes in cluster 4, which change only in immune competent mice, are highly abundant in controls including Rag1-/-. Infected immune competent mice have a lower abundance of these mRNAs, but in Rag1-/- there is no difference compared to controls. Therefore, these genes appear to be down-regulated upon infection by processes which depend on (mature) T and/or B cell activity, surprisingly also at 3 dpi. Enriched GO terms in this gene cluster include 13 terms for cytokines, particularly interleukin-1, IL-1, ("negative regulation of" and "secretion of"). Terms such as "negative regulation of viral (or inflammatory) response", "negative chemotaxis", "autophagy", "blood coagulation", "inositol phosphate-mediated signaling", and "positive regulation of calcineurin-NFAT" are also enriched in cluster 4. Inositol signaling can lead to release of calcium and calcineurin-dependent translocation of NFAT to the nucleus and activation of its target genes

in T cells, but also many other cell types (reviewed by macian05). GO enrichment of cluster 4 also highlights particular regulation of transforming growth factor $\beta, TGF\beta, epidermal growth factor, EGF, and tumor necrosis factor, TNF. TGF\beta is important for wound healing ini$
$dependent and capable to suppress inflammatory responses (noti10) but is also reported to regulate proliferation of epit$
$1 and IL-6 are among the enriched GO terms in cluster 4. The IL-1 receptor (type I) is similar to Toll-$
$like receptors and IL-1 induces innate immune responses in many cell types, and influence lymphocyte activity (dinar$
$6 has been shown to support repair and inhibit apoptosis after epithelial wounding (kuhn14) and it is known to be importa$

$Th17 .... Stange's work (discuss Richard). IL-6 induces the Janus kinase, JAK, and signal transducer and activator of$
$1 and IL-6 are main actors in the epithelial response to E. falciformis infection and that the response is T and B cell depen$
$points. The pronounced changes late in infection (7dpi) also reflect the expected onset of an adaptive immune response.$
$point. Among the same genes, natural killer cell regulation, JAK-STAT signalling, and IL-$
$1 and interleukin-2, IL-2 production are enriched biological processes. IL-2 is one target of NFAT signaling and as m$
$STAT signaling can be induced by IL-6. This indicates that the early responses in, e.g., NFAT and IL-$
$6 regulation induce distinct mRNA abundace differences later in infection (7dpi).$

Taken together, clusters 3 and 4 seem to capture the difference between Rag1-/- mice and immune competent mice in this parasitic infection. Some genes are different also between control animals (Rag1-/- and NMRI, cluster 3). These do not change upon infection in mice lacking mature T and B cells. Other genes (cluster 4) for which the transcriptional profile is similar in all controls, mRNA abundance also only changes upon infection only in animals with functional maturation of T and B cells. Using GO enrichment analysis, we show that regulation of IL-1, IL-6, TNF,TGFβ and EGF are regulated in both first and challenge infections and suggest wound healing as a common funct

1 RNA processing is enriched in challenge infected immune competent mice

Three challenge infected samples (3 dpi, 5 dpi and 7 dpi) from immune competent mice show a distinct profile in cluster 6. This cluster is highly enriched for RNA processing and splicing, as well as terms for histone and chromatin modification.

"UDP-N-acetylglucosamine-lysosomal-enzyme" (MF)

mRNA abundance .....

We performed statistical tests to evaluate significant differences (FDR1%) in mRNA abundance between different parasite life cycle stages, approximated by time post infection (Table 2). Between early time-points, 3 dpi and 5 dpi, 103 mRNAs were different, whereas between 3 dpi and 7 dpi 1399 mRNAs were DA, and between 5 dpi and 7 dpi 2084 mRNAs were DA (Figure 3a). This indicates that the major changes take place between 5 dpi and 7 dpi, and that variation is smaller between 3 dpi and 5 dpi. This motivated us to define 3 dpi and 5 dpi as "early infection" and 7 dpi as "late infection". Early and late infection samples were tested for DA compared to sporozoites and sporulated oocysts, resulting in 1697 and 3919 DA mRNAs, respectively. To evaluate this outcome further, we performed hierarchical clustering of (the union of) the DA mRNAs in the comparisons described above.

Major transcriptional differences in the E. falciformis life cycle are independent of mouse immune status

In the parasite transcriptome, we see no difference between infection in immune competent mice, or T and B cell deficient Rag1-/- mice, or between naive and challenge infected mice. This is surprising considering the measured differences in oocyst output in the same comparisons (Figure 1a), and the fact that these differences are visible in the mouse transcriptomes. On the parasite side, major patters instead seem to be determined by life cycle stages, independent of the host immune status. Distinct clusters of genes define early infection (3 dpi and 5 dpi) in which schizogony takes place, and separately, late infection (7 dpi) in which it is assumed that gametocytes are present. Extracellular (and in our experiments, in vitro) samples of sporozoites and oocysts also cluster separately and are defined by distinct gene clusters.

Early infection transcriptomes suggest parasite expansion

GO enrichment of the major sporozoite defining cluster with high mRNA abundance in this stage (cluster 4) proposes that different biosynthesis processes are important, as well as "maintenance of protein location in cell". Possibly, the latter is due to control of microneme protein localization as sporozoites prepare for invasion. Biosynthesis terms probably reflect preparation for asexual expansion and perhaps ATP production for motility in sporozoites, although no GO terms for motility are enriched in this cluster. In the following life cycle stages, measured 3 dpi and 5 dpi, in which several rounds of schizogony (asexual replication) take place, mRNAs in cluster 6 are abundant in all early samples (except one, discussed below). Among these "early infection" genes, several GO terms for biosynthetic acitivity are enriched, e.g., "ribosome biogenesis" and "cellular biosynthetic process", as well as terms for "gene expression" and RNA processes, including tRNAs and ncRNAs. "Cellular amino acid catabolic process" is also enriched. mRNAs in cluster 6 are low in abundance in both oocysts and late infection, whereas the sporozoite profile is less defined in this cluster. Therefore, this cluster is distinct for early infection with some features shared with invasive sporozoites. The enriched terms indicate replication and growth, highlighting the parasite's expansion in numbers on 3 dpi and 5 dpi, as supported both by previous knowledge about the life cycle and our increase in sequences from the parasite (Figure 1b).

Gametocyte presence likely determines transcriptome late in infection

Two gene clusters have a distinct profile with high mRNA abundance on 7 dpi (clusters 2 and 7). Both clusters display low mRNA abundance in other life cycle stages, especially in oocysts and sporozoites. Enriched GO terms such as "movement of cell or subcellular component" and "microtubule-based movement" along with terms suggesting ATP production ("ATP generation from ADP") indicate the presence of motile and energy demanding gametocytes in these samples. Other cluster 2 terms such as "chitin metabolic process" along with enriched metabolic and biosynthetic processes and "gene expression" in cluster 7 suggests start of or preparation for encystation at 7 dpi. In addition, cluster 2 is enriched for a number

of GO terms for "blood coagulation" and reflect the presence of Thrombospondin type I domains in the protein products of cluster 2 mRNAs. Thrombospondin type 1 domains have been reported in *E. tenella* microneme localizing proteins, MIC, e.g. MIC4 (Tomley01). 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. This suggests that *E. falciformis* prepares for invasion already in gametocytes or during oocyst formation, however this is speculative.

Oocysts...

Oocysts have strong high or low mRNA abundance profiles and few intermediate abundance mRNAS. Clusters 1 and 5 contain mRNAS with high abundance in oocysts and all other clusters have low abundance in this stage. GO enrichment in cluster 1 contains only one term (adj. p-value 0.11) for "DNA-templated transcription, initiation". Cluster 5 is enriched for terms related to stress responses, "DNA repair", "protein modification process" and for "cell differentiation". Stress responses and DNA repair can be a result of storage in potassium dichromate of mouse faeces with oocysts before purification. Initiation of transcription and cell differentiation could reflect slow preparation for invasion when oocysts are taken out of storage and purified for RNA extraction. Overall, the oocyst profile in five of 7 gene clusters is characterized by below average abundance of mRNAs, as can be expected in this life cycle stage suited for long-term survival outside of the host. Taken together, sample and gene clustering indicates that genes in clusters 6 are abundant only early in infection and are candidates for merozoite specific genes, clusters 2 and 7 might be useful to characterise gametocytes. Genes in cluster 4 can be evaluated as sporozoite specific genes, and clusters 1 and 5 as oocyst specific genes in *E. falciformis*.

Imperfect clustering might reflect true biological differences

One late challenge infection sample deviates from the "late infection pattern" and clusters with an early challenge infection sample. These two samples are, apart from controls and excluded samples (see above), the ones with the least number of detected parasite genes (1836 and 1580, respectively). This, along with the unclear transcriptional profile and the fact that they are both challenge infections in immune competent hosts, suggest that infections had been cleared before sampling. If this is the case, it raises the question of why it was cleared in these samples but not in other challenge infected immune competent mice. The answer can lie with technical problems in the experiment. Hierarchical clustering analysis in most cases does not cluster replicates together, which at a first glance suggests such problems. However, considering the strong overall patterns in the data and that these fit well with previous knowledge about this infection, it is worthwhile to consider the possibility that this imperfection reflects true biological variation between individual hosts and possible also parasites (oocysts used for infection). It is perceivable that the parasite accommodates to host variation with small differences in, e.g., host stress levels due to litter mates, draught, differences in light exposure or other factors which may vary also in a controlled animal facility. Parasite adjustment to such factors could explain the transcriptional profiles we see. We suggest that considering such

possibilities might be useful for understanding the basic biology of the parasite, and as an extension, successfully fight infections in less homogeneous hosts than laboratory mice.

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).

Evolutionary conservation of expression changes in E. falciformis

For comparison of changes in the lifecycle of *E. falciformis* with the economically important chicken parasite *E. tenella* and the model apicompomplexan *Toxoplasma gondii* we downloaded expresson datasets from ...

Oocysts, Sproozoites

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.

. . . . . .

2 Methods

2.1 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 where 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.

2.2 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).

2.3 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–).

2.4 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.

2.5 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.

2.6 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).

2.7 Differential mRNA abundance, data normalisation and sample exclusions

Count data was normalized using the R-package edgeR (version 3.14.0; cite) with the upperquartile normalisation method. 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 is suitable for read densities following a negative binomial distribution. Two samples contradicted this assumption (parasite data) for later modelling and both mouse and parasite data from these samples were excluded from further analysis: NMRI_1st_3dpi_rep1 and NMRI_2nd_5dpi_rep1 (SI ...). The method then fits a generalized linear model (GLM with a negative binomial link function) for each gene (glmFit) and then performs likelihood ratio tests for models w or w/o focal factor (glmLRT).

2.8 Selection of differentially abundant mRNAs nad 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 at most 500 genes differentially abundant with lowest FDR (¡0.05) 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 1618 unique genes selected as differently expressed . 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).

2.9 Differentially abundant mRNAs used for hierarchical clustering

All analyses were performed in R (cite R-core). Complete scripts are available at https://github.com/derele/Ef_RNAseq.git tagged as version 1.0.

Table 1 Genes used for hierarchical clustering of mRNAs differently abundant depending on time p.i..

Data description	*E. falciformis* 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 . . .

Author details
[1] Institute of Biology, Humboldt-Universitat zu Berlin, Philippstr. 13, Haus 14, 10115 Berlin, Germany. [2] Max Planck Institute for Biology of Ageing, Joseph-Stelzmann-Strasse 9B, 50931 Cologne, Germany. [3] Leibniz Institute of Zoo and Wildlife Research, Alfred-Kowalke-Str. 17, 10315 Berlin, Germany.

References

1. Koonin, E.V., Altschul, S.F., Bork, P.: Brca1 protein products: functional motifs. Nat Genet **13**, 266–267 (1996)
2. Kharitonov, S.A., Barnes, P.J.: Clinical Aspects of Exhaled Nitric Oxide. in press
3. Zvaifler, N.J., Burger, J.A., Marinova-Mutafchieva, L., Taylor, P., Maini, R.N.: Mesenchymal cells, stromal derived factor-1 and rheumatoid arthritis [abstract]. Arthritis Rheum **42**, 250 (1999)
4. Jones, X.: Zeolites and synthetic mechanisms. In: Smith, Y. (ed.) Proceedings of the First National Conference on Porous Sieves: 27-30 June 1996; Baltimore, pp. 16–27 (1996). Stoneham: Butterworth-Heinemann
5. Margulis, L.: Origin of Eukaryotic Cells. Yale University Press, New Haven (1970)
6. Orengo, C.A., Bray, J.E., Hubbard, T., LoConte, L., Sillitoe, I.: Analysis and assessment of ab initio three-dimensional prediction, secondary structure, and contacts prediction. Proteins **Suppl 3**, 149–170 (1999)
7. Schnepf, E.: From prey via endosymbiont to plastids: comparative studies in dinoflagellates. In: Lewin, R.A. (ed.) Origins of Plastids vol. 2, 2nd edn., pp. 53–76. Chapman and Hall, New York (1993)
8. Innovative Oncology
9. Smith, Y. (ed.): Proceedings of the First National Conference on Porous Sieves: 27-30 June 1996; Baltimore. Butterworth-Heinemann, Stoneham (1996)

10. Hunninghake, G.W., Gadek, J.E.: The alveloar macrophage. In: Harris, T.J.R. (ed.) Cultured Human Cells and Tissues, pp. 54–56. Academic Press, New York (1995). Stoner G (Series Editor): Methods and Perspectives in Cell Biology, vol 1
11. Advisory Committee on Genetic Modification: Annual Report. London (1999). Advisory Committee on Genetic Modification
12. Kohavi, R.: Wrappers for performance enhancement and obvious decision graphs. PhD thesis, Stanford University, Computer Science Department (1995)
13. The Mouse Tumor Biology Database. http://tumor.informatics.jax.org/cancer_links.html

Figures

Figure 1 Experimental design and outcome of a dual RNAseq experiment 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.

figure 1 label cotribution of mouse vs. parasite transcripts to the overall transcriptional output of the analysed tissue for all samples.

Figure 2 *E. falciformis* mRNAs with significantly different abundance at different times p.i. in NMRI mice (see Table 2). Samples cluster according to expected developmental state of the parasite, without distinguishing between first and challenge infection. Hierarchical clustering was set to apply seven gene clusters.

2.10 Parasite development reflected in mouse data

Figure 3 Mouse mRNAs with significantly different abundance at different times p.i. in NMRI mice (see Table 2). Samples cluster according to expected developmental state of the parasite, without distinguishing between first and challenge infection. Rag1-/- mouse samples cluster together. Hierarchical clustering was set to apply four gene clusters.

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 diplays 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.

Sample*	Sequencing method	batch	total reads	reads mapping Mouse	reads mapping *E. falciformis*	Percentage ** *E. falciformis*	detected *E. falciformis* genes
NMRI_2ndInf_0dpi_rep1	GAII	2	108,937,797	70,489,674	247	0.0004	1
Rag_1stInf_0dpi_rep1	hiseq	3	25,362,793	18,853,850	443	0.0023	2
C57BL6_1stInf_0dpi_rep1	hiseq	3	35,731,249	25,119,348	457	0.0018	2
C57BL6_1stInf_0dpi_rep2	hiseq	3	47,085,959	34,377,133	608	0.0018	2
Rag_1stInf_0dpi_rep2	hiseq	3	46,556,156	35,233,327	676	0.0019	2
NMRI_2ndInf_0dpi_rep2	hiseq	3	58,122,244	40,794,245	3,406	0.0083	51
NMRI_2ndInf_3dpi_rep1	hiseq	3	57,934,016	40,544,287	4,803	0.0118	95
NMRI_2ndInf_5dpi_rep2	hiseq	3	63,965,539	48,289,181	10,941	0.0227	407
NMRI_1stInf_0dpi_rep1	GAII	1	82,364,585	55,176,243	17,954	0.0325	701
NMRI_2ndInf_3dpi_rep2	hiseq	3	65,548,826	46,171,909	29,548	0.0640	1,580
NMRI_2ndInf_7dpi_rep2	hiseq	3	67,487,466	51,722,265	40,091	0.0775	1,836
Rag_1stInf_5dpi_rep1	hiseq	3	38,651,359	29,982,453	63,024	0.2098	2,548
Rag_1stInf_5dpi_rep2	hiseq	3	34,779,832	25,297,803	99,000	0.3898	2,828
C57BL6_1stInf_5dpi_rep1	hiseq	3	40,904,388	29,319,604	185,969	0.6303	4,173
Rag_2ndInf_5dpi_rep1	hiseq	3	50,049,848	37,093,621	192,856	0.5172	4,167
C57BL6_1stInf_5dpi_rep2	hiseq	3	29,511,368	18,062,349	215,696	1.1801	3,823
C57BL6_2ndInf_5dpi_rep1	hiseq	3	35,148,432	25,660,184	262,909	1.0142	4,563
NMRI_1stInf_3dpi_rep1	GAII	1	73,236,430	49,993,358	394,384	0.7827	5,220
NMRI_1stInf_3dpi_rep2	GAII	2	160,709,694	117,791,044	413,051	0.3494	4,862
NMRI_1stInf_5dpi_rep2	GAII	2	119,902,722	76,419,774	794,570	1.0290	5,333
NMRI_2ndInf_5dpi_rep1	GAII	2	230,773,955	143,186,486	1,846,840	1.2734	5,533
NMRI_2ndInf_7dpi_rep1	hiseq	3	70,366,762	41,467,146	8,634,201	17.2335	5,875
NMRI_1stInf_5dpi_rep1	GAII	2	76,702,168	47,037,087	8,669,701	15.5631	5,700
NMRI_sporozoites_rep2	GAII	0	19,551,681	8,656	11,470,604	99.9246	5,513
NMRI_1stInf_5dpi_rep3	GAII	0	191,099,180	83,735,624	27,839,458	24.9513	5,784
NMRI_1stInf_7dpi_rep1	GAII	1	66,505,514	3,310,666	39,400,884	92.2488	5,932
NMRI_sporozoites_rep1	GAII	1	67,325,397	4,334	43,774,401	99.9901	5,825
NMRI_oocysts_rep1	GAII	1	68,859,802	3,805	49,653,065	99.9923	5,695
NMRI_oocysts_rep2	GAII	0	151,090,783	18,524	71,019,860	99.9739	5,777
NMRI_1stInf_7dpi_rep2	GAII	1	139,749,046	21,699,324	73,539,445	77.2159	5,943

* sample names are given as a) mouse strain b) first or challenge infection c) days post infection (dpi) and d) replicate number seperated by undersocre .

** percentag mapping *E. falciformis* is given as percentage in total mapping reads

3 mRNA abundance differences between different experimental groups

Table 2 mRNA abundance differences between different experimental groups.

Day post infection comparisons	*Ef* genes different (FDR≤1%)	Mouse genes different (FDR≤1%/5%)
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
Day post infection, parasite relevant comparisons		
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
First and second infection comparisons		
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

Additional Files

Additional file 1 — Raw and normalized counts
Raw counts of reads mappins to the *E. falciformis* and mouse genome for individual samples in our study. Normalized counts for seperately for the host and parasite mappings (three compressed csv files).

Additional file 2 — Results of statistical tests (edgeR)
Focal contrast, fold-changes, likelihood ratio in/excluding this difference in models, p-values , and false discovery rates (adjusted p-values) are given for all tested contrasts (one compressed csv file).

Additional file 3 — Additional methods and resuts
Document containing additonal figures and summary tables (pdf).

Additional file 4 — Results of enrichment analyses (topGO)
Tables listing all tested gene sets and resulting significatn GO terms.