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

Apicomplexan parasites such as Plasmodium spp., Toxoplasma gondii and Eimeria spp. cause disease in humans, livestock and wild animals. The genus Eimeria contains thousands of niche specific intracellular parasites, including several species which cause losses in poultry industries. Eimeria falciformis naturally infects the cecum of mice. Infecting one of the best studied and available animal models in biological research, E. falciformis constitutes a perfect model to investigate Eimeria parasites. However, much is still unknown about the parasite's basic biology and no in vitro culture has been established for the full life cycle. We have performed a dual RNA-seq transcriptome study of the full life cycle in the mouse and of in vitro cultured sporozoites and oocysts. Drastic differences are seen in both parasite and host at three time-points post infection. Comparisons between immunocompetent and immunocompromised mice show differences in oocyst output as well as transcriptional differences indicated by Gene Ontology enrichments. In mouse, TGF-beta, EGF, TNF and IL-1 and IL-6 are examples of genes reacting differently depending on mouse immune status. Parasite transcriptomes have distinct profiles early and late in infection, characterized by biosynthesis and motility, respectively. Sporozoites and oocysts can also be identified by their respective transcriptional profiles. Taken together, the analysis highlights general patterns in the parasite's life cycle and links them to biological processes. It also lays the ground for detailed analysis of specific parasitic stages and the genes relevant in them. The use of hosts with different immune competence highlights the role of adaptive and innate immunity and offers a source for in-depth analysis of these responses.

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

Why Eimeria falciformis? Genome available

Other apicomplexa, especially E. tenella, E. maxima, T. gondii and Plasmodium spp.

Why is evolution of these parasites relevant/important

Why mouse host (e.g. compared to chicken)

Ef life cycle and timing

Knowledge-gaps: basic biology not understood, no in vitro system, no stable tranfection reported, unknown why so niche specific and Eimeria genus so numerous

We approach some of these questions by using dual RNA-seq to produce simultaneous host and parasite transcriptomes throughout in vivo infection infection

Dual RNA-seq has been performed in bacteria... also in protozoans? Intracellular...

Improved normalization methods should make analysis of reproducing (increasing in numbers and biological material) possible... we show/evaluate this

Transcriptome studies are a source of great amounts of data. Enrichment tests (in general) and GO enrichment are useful tools to elucidate patterns in such large data... some words about GO concept

Results

A dual transcriptomics experiment

We performed mRNA sequencing of caecum epithelial 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 in order to assess the influence of host immunocompetence on parasite development (Figure 1a and 1b). Immune competent NMRI mice were infected and sampled at three time-points post infection, p.i. We also infected mice lacking the Recombination activating gene 1, Rag1, (Rag1-/-) and compared them to the parental C57BL/6 mouse line. Rag1-/- mice lack mature B and T lymphocytes, which is taken as a proxy for absent adaptive immunity. All three mouse strains were used for infections of both naive mice and previously infected (and recovered) animals (onward referred to as "challenged" or "challenge infection"). Basic phenotyping (Figure 1a) showed differences in oocyst numbers between the immunocompetent (C57BL/6) and immunodeficient (Rag1-/-) host strains in naive and challenge infected mice. 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. Oocyst numbers in faeces peaked on days 8-9 andall mice had cleared the infection by day 14. Development and replication 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 and late 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). 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 and 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). 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 25,362,739

(sample Rag_1stInf_0dpi_rep1) to 139,749,046 (NMRI_1stInf_7dpi_rep2). 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 92% (sample NMRI_1stInf_7dpi_rep1) and a minimum of 0.064% (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 samples 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 (containing more genes) at later time-points post infection (Table 2). For mouse, 325 mRNAs were considered differently abundant (DA; FDR ¡ 0.01) 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 unique genes responding to infection (Figure 2bi). DA mRNAs early in infection (3 dpi and 5 dpi) were not a strict subset of genes DA later in infection (7 dpi). Instead, the transcriptional profile of the mouse changes throughout the infection. Differences 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 ([**?**]) or 7 dpi (our RNA-seq data) against uninfected controls show a strong correlation (Spearman's ρ = 0.74; Figure 2a). Considering both biological differences in the experiments (e.g. exact time-points for samples), and technical differences between the two methods this comparison confirms the adequacy of using dual RNA-seq for assessing the host transcriptome.

Epithelial responses vary with mouse immune status

A distinct response in early infection is indicated by the number of DA mRNAs between different time-points. To validate the pattern we performed hierarchical clustering on the (union of) mouse genes DA between different time-points post infection (Figure 2c). Three main sample clusters formed (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 changes transcriptional profile upon infection 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 healthy animals in a T and B cell dependent manner. Hence, these genes surprisingly appear to depend on functional T and B cells also in controls (cluster 3) and at the earliest time-point post infection (3 dpi, cluster 4). Gene Ontology, GO, terms enriched in cluster 3 are, e.g., "lipid metabolic process" and "protein intraciliary transport". Other enriched terms in cluster 3 are regulation of other metabolic processes and "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-/- and appear to be down-regulated upon infection by processes which depend on (mature) T and/or B cell activity, also as early as 3 dpi.

Renewal of infected epithelium appears heavily regulated

Several terms enriched in cluster 4 are associated with wound healing and proliferation. 13 terms for cytokines as well as "negative regulation of viral (or inflammatory) response", "negative chemotaxis", "autophagy", "blood coagulation", "inositol phosphate-mediated signaling", and "positive regulation of calcineurin-NFAT" are enriched. mRNAS supporting these terms are less abundant in infected samples. Although speculative, several of these processes can be linked. 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 also highlights regulation of transforming growth factorβ, TGFβ, epidermal growth factor, EGF, and tumor necrosis factor, TNF. TGFβ is important for wound healing in intestinal epithelium (beck03), and EGF regulates proliferation of epithelial cells and inhibits apoptosis (suzuki10). TNF is dose-dependent and can suppress inflammatory responses (noti10) and is reported to regulate proliferation of epithelial cells (kaiser97). Additionally, IL-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 influences lymphocyte activity (dinarello09). IL-6 has been shown to support repair and inhibit apoptosis after epithelial wounding (kuhn14, probably through the Janus kinase, JAK, and signal transducer and activator of transcription STAT3 (pickert09). IL-6 is also known to be important for development of Th17 responses (ref in kuhn intro) which play an important role in responses to *E. falciformis* (stange–). This analysis suggests that TGFβ, TNF, EGF, IL-1 and IL-6 are main actors in the epithelial response to *E. falciformis* infection and that the response is T and B cell dependent, also at early time-points. Enrichment tests suggest that hosts invest resources in intestinal healing also at early time-points of *E. falciformis* infection and that functional T and B cells are needed for these responses. An alternative interpretation is that pathology is lower in Rag1-/- mice

and that these responses are therefore not triggered in them. Based on unpunlished data from colleagues (talk to Stange), severe pathology in infected Rag1-/- mice makes this scenario unlikely.

Late infection is strongly enriched for adaptive immunity processes

The pronounced changes late in infection (7 dpi) reflect the expected onset of an adaptive immune response, undelined by enriched GO terms in gene clusters 1 and 5. Terms such as "antigen binding" and "immunoglobulin receptor binding" (molecular function, MF), and "immune system process", "adaptive immune response" and also "innate immune response" (biological process, BP) are highly enriched in cluster 5 and confirm an activated immune system and adaptive immune responses at this time-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 mentioned above, JAK-STAT signaling can be induced by IL-6. It is likely that the enriched early responses in, e.g., NFAT and IL-6 regulation induce distinct mRNA abundace differences later in infection (7 dpi) and it is encouraging that these links are detected by the methods applied here. Taken together, clusters 3 and 4 highlight 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), whereas in cluster 4 the profile is shared among all controls, including Rag1-/-. In both cases, the profile changes in immune competent mice but not in T and B cell deficient Rag1-/- mice. 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 probably function of this regulation, especially at 7 dpi.

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 GO terms for RNA processing and splicing, as well as terms for histone and chromatin modification. . . .

Include or not? - daunorubicin metabolic process ¿ cancer drug which prevents DNA replication by stabilizing DNA - doxorubicin metabolic process ¿ cancer drug "UDP-N-acetylglucosamine-lysosomal-enzyme" (MF)

Transcriptional differences in the parasite life cycle are independent of mouse immune status

We performed statistical tests to evaluate significant differences (FDR<0.01) 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 and applied GO enrichment analysis of gene clusters (annotations are inferred from orthologs in other *Eimeria* spp. or *T. gondii*). 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 (asexual reproduction) takes place, and separately, late infection (7 dpi) in which it is assumed that gametocytes are present. Extracellular samples of sporozoites and oocysts also cluster separately and are defined by distinct gene clusters.

Early infection transcriptomes reflect 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 could reflect preparation for asexual expansion. n reporting the genome of *E. falciformis*, two fructose-bisphosphate aldolase, FBA, paralogs were inferred (heitlinger14). Based on previously reported data on T. gondii, the authors speculate that localization of FBA is controlled in sporozoites in order to direct energy supply during the supposedly costly invasion. Our enrichment of control of protein localization support the idea that control over energy supply and probably other functions is important in the invasive stage. In the following intracellular life cycle stages, measured at 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 in cluster 6, several GO terms (biological process) for biosynthetic acitivity are enriched, e.g., "ribosome biogenesis" and "cellular biosynthetic process", as well as terms for "gene expression" and RNA processing, including terms for tRNAs and ncRNAs. "Cellular amino acid catabolic process" is also enriched. In cluster 3, with a similar profile as cluster 6, no GO terms are enriched. Cluster 6 is distinct in early infection with a biosynthesis profile shared with invasive sporozoites. Replication and growth-related processes being enriched highlights 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).

Gametocytes likely determine 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 ongoing encystation at 7 dpi, fitting the timing of oocyst output which peaks at 8-9 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 are characterized by stress responses and differentiation

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 probably reflects (slow) preparation for invasion when oocysts are taken out of storage and purified for RNA extraction. Overall, the oocyst profile in five out of seven gene clusters is characterized by below average abundance of mRNAs, as can be expected in this life cycle stage which endures long-term survival outside 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, whereas clusters 2 and 7 might be useful to characterise gametocytes. Genes in cluster 4 can be considered sporozoite specific, and clusters 1 and 5 as oocyst specific genes in *E. falciformis*.

Imperfect clustering might reflect true biological differences

In parasite data, 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 be technical problems in the experiment. However, other patterns suggest a different interpretation. Hierarchical clustering analysis in most cases does not cluster replicates together, which at a first glance suggests such problems. On the other hand, very distinct samples such as oocysts and sporozoites do cluster as replicates. Also, considering the early/late infection patterns in the parasite data and that these fit well with previous knowledge about this infection (asynchronous schizogony and gametocyte formation at 7 dpi), it is worthwhile to consider whether

replicate separation reflects true biological variation between hosts and possibly 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 ght 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, with overall patterns but replicate separation. We suggest that considering such possibilities is useful for interpreting results and understanding basic biology of the parasite, and, in extension better understand infections in less homogeneous hosts than laboratory mice.

Evolutionary conservation in life cycle-characteristic gene groups

To gain insight about the conservation status of life stage-specific gene groups in *E. falciformis*, we tested gene clusters in Figure 3c for enrichment of gene orthologs which are present in other species. Groups analyzed for orthologs are: i) one category with *E. falciformis* only ("Efalci"), ii) one category of 10 apicomplexan parasites ("Api") (see xxx), iii) one category of Api but excluding *Cryprosporidium hominis* ("Api minus C.h."), iv) one catogory with three *Eimeria* species: *E. falciformis, E. maxima* and *E. tenella* ("Eimeria"), and v) one conserved group containing a broad range of species with, e.g., *Saccharomyces cerevisiae* and *Arabidopsis thaliana* ("conserved"). Sporozoites are enriched for *E. falciformis* specific genes, and the group "conserved" is underrepresented, indicating a strong species specificity in the sporozoite stage. Gene clusters characteristic for early infection are enriched for "Api" (cluster 6), probably reflecting that similar processes and genes are involved in asexual reproduction of the selected apicomplexan species. In late infection (cluster 2), characteristic genes are underrepresented for conserved and "Api minus C.h", whereas the other characteristic cluster for late infection is enriched for "conserved" orthologs. This clustering and difference in enrichment of conserved versus apicomplexan (minus *C. hominis*) genes probably reflects activity of generic and phylum specific processes/genes in oocysts. One oocyst cluster (cluster 1) is underrepresented for "Eimeria", whereas the other oocyst genes (cluster 5) are enriched for "Api". . . .

In addition to analysing conservation by ortholog enrichment, we also performed Spearman correlation analysis between our RNA-seq transcriptomes and RNA-seq data from related parasites. Two datasets for the economically important chicken parasite *E. tenella* (walker15 and reid14) and one dataset of the model apicompomplexan parasite *Toxoplasma gondii* (hehl15) were included in the comparison. Interestingly, this analysis confirms the species specificity for the sporozoite transcriptome, by clustering *E. tenella* sporozoite samples together, but *E. falciformis* sporozoites with *E. falciformis* early infection samples. *E. falciformis* late infection samples correlate the most with *E. tenella* gametocytes, indicating similarity also between the species in this stage and supporting the presence of gametocytes in our samples. *E. tenella* merozoites from both independent studies are most similar to early *E. falciformis* samples, indicating similarity also during asexual reproduction which is shown by the conservation analysis as well (Figure?/Table?). *E. falciformis* oocysts cluster with unsporulated *E. tenella* oocysts, whereas *E. tenella* sporulated oocysts are most similar to *E. tenella* sporozoites.

In summary, we have performed a dual RNA-seq transcriptome of an apicomplexan parasite in its natural host. Our analysis of differentially abundant mRNAs at different time-points post infection highlights large groups of genes which characterize the different life stages of the parasite. The dual approach also allows insight into the host responses to this intracellular infection. Using GO enrichment tests and by analysing how species specific these life stage defining gene groups are, we show that the transcriptional profile of sporozoites is the most *E. falciformis* specific stage of the time-points we have analyzed. The sporozoite defining gene group is characterized by ATP production, regulation of protein localization and biosynthetic processes. The analysis further demonstrates that early infection, 3 and 5 dpi, is not homogenous between samples, and that the conformity seen in sporozoites (and oocysts) is regained late in infection. We speculate that parasites during asexual replication are more asynchronous than late stage merozoites or gametocytes on 7 dpi, possibly due to an increasing influence from the host adaptive immune response. The asynchrone early parasite asexual replication stages are enriched for growth related processes, and a large expansion in parasite numbers is reflected in the percentage of parasite genes, read numbers as well asin RT-qPCR data. Gametocyte formation on 7 dpi is supported by GO enrichments of terms for motility and ATP production, as well as 7 dpi samples correlating strongly with *E. tenella* gametocyte samples. Our analysis interestingly does not detect any influence of mouse immune status on the parasite transcriptome, even though a difference in oocyst output between Rag1-/- and its C57BL/6 control is clear. We sequenced enriched mRNA and it is possible that non-coding RNAs or other post transcriptional regulatory mechanisms in the parasite could explain differences in oocyst output. Another possibility is that immune competent mice mainly interfere with the parasite at a very late time-point before oocyst formation, which a different experimental design could elucidate.

Methods

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 μL 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 3 dpi, 5 dpi and 7 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.

Oocyst purification for infection and sequencing

Sporulated oocysts were purified by flotation from feces stored in potassium dichromate and administered orally in 100 μL 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).

Sporozoite isolation

Sporozoites were isolated from sporocysts by excystment. For this, sporocysts were incubated at 37 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–).

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. 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.

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" s 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.

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).

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).

Selection of differentially abundant mRNAs and 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. In heatmaps, all samples, i.e., also samples which did not have any significantly different mRNAs according to our selection, were included in hierarchical clustering. Scale bar in heatmaps shows 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, by the complete linkage method ('complete', R package base).

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.

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

Author's contributions
Animal and parasite experiments: Simone Spork, experimental design: Richard Lucius, Simone Spork, Emanuel Heitlinger, RNA sequecing: Christoph Dieterich, data anylsis: Emanuel Heitlinger, Totta Kasemo, text: Totta Kasemo, Emanuel Heilinger, Simone Spork (all?).

Acknowledgements
This project was funded by..... Totta Kasemo is funded by the German Research Foundation (DFG) program GRK4026: Parasite Infections: From Experimental Models to Natural Systems.

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
Figures

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.)	4	8052
Used in hierarchical clustering (heatmap)	1618	1313

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	x	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

1 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-/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.