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

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

1. Institute of Biology, Humboldt-Universitat zu Berlin

Philippstr. 13, Haus 14, 10115 Berlin, Germany

2. University Hospital Heidelberg - German Center for Cardiovascular Research (DZHK), Analysezentrum III, Im Neuenheimer Feld 669, 69120 Heidelberg, Germany

3. Leibniz Institute of Zoo and Wildlife Research, Alfred-Kowalke-Str. 17, 10315, Berlin, Germany

**ABSTRACT**

Apicomplexan parasites such as Plasmodium spp., Toxoplasma gondii and Eimeria spp. cause disease in humans, livestock and wild animals. The genus Eimeria comprises thousands of niche specific intracellular parasites, most prominent among them 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 can be used as a 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 enriched functional (Gene Ontology; GO) categories. 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

INTRODUCTION

Eimeria falciformis is an intracellular parasite in the phylum Apicomplexa. Among more than 4000 described species of Apicomplexa (Duszynski 2011) prominent pathogens of human are found such as Toxoplasma gondii, causative agent of toxoplasmosis, Plasmodium spp., causing malaria, and Cryptosporidium spp, which cause cryptosporidosis. Coccidiosis is a disease in livestock and wildlife caused by coccidian parasites which are dominated by > 1800 species of Eimeria (Duszynski 2011). Eimeria are best known for several species problematic for poultry industry (Chapman et al., 2013). A useful model for studying Eimeria spp. is E. falciformis, which naturally infects wild and laboratory mice, Mus musculus (Heitlinger et al. 2014).

Eimeria spp. infect the gut and are highly niche specific in the caecum and upper part of colon, where they mainly reside in the crypts (Haberkorn 1970, Schmid et al., 2012; Stange et al 2009). These monoxenous parasites go through asexual (schizogony) and sexual reproduction which results in hosts releasing high numbers of resistant oocysts. When a mouse ingests E. falciformis oocysts, one sporulated oocyst releases eight infective sporozoites inside the host, which can infect epithelial crypt cells of the caecum and colon. Within the epithelium, so called merozoite stages form in several rounds of schizogony. Parasite numbers increase drastically during schizogony, which is not completely synchronized, and the exact number of schizogony cycles is either not clear (Haberkorn 1970, Mesfin and Bellamy, 1978) or it varies naturally and can therefore not be determined. Haberkorn (1970) reported variation from one to four rounds of schizogony before gamete formation whereas Mesfin and Bellamy (1978) detected four schizont stages. (Any comment on unpublished data from your lab, Richard or Simone?) It therefore appears that the number and exact timing of schizont formation can vary in E. falciformis infections. Oocysts are first detected in faeces on day six to seven post infection, p.i., and output peaks on day eight to nine p.i. Oocysts are not detectable after day 13-15 p.i. (Haberkorn 1970 and this study). Oocysts form as merozoites differentiate into gametes, which fuse and form a zygote. Immature, unsporulated oocysts are shed into the environment where they mature into sporulated, infective oocysts (Mesfin and Bellamy, 1978).

Studies on host responses to this infection have shown that in E. falciformis infection in laboratory mice, IFNγ is upregulated (Schmid, 2014). It was however demonstrated that even though IFNγ and its receptor influences parasite development, the pathway is not required for clearance of a first infection, or for development of protective immunity in challenge infections. In the IFNγ deficient model which displays larger weight losses and intestinal pathology but also lower oocyst output, the phenotype was recovered by blocking IL-17A and IL-22 signaling. Only blocking IL-22 signaling recovered the oocyst shedding but had no effect on weight loss (Stange et al., 2012). These studies demonstrate the complex relationships between parasite and host, e.g., exemplifying that larger pathology in the host is not necessarily beneficial for the parasite, which was less successful in producing oocysts. The role and importance of adaptive immunity remains unclear.

Several studies address host parasite interactions in apicomplexan parasites (Periz et al., 2007; Reid and Berriman, 2012; Schmid, 2014; Stange et al., 2012) but often with a strong focus on either host or parasite. To our knowledge, few studies of apicomplexan parasites simultaneously and symmetrically asses both parasite and host. Reid and Berriman have analyzed two microarray datasets from independent studies of Plasmodium and mouse, and mosquito, respectively (Lovegrove et al., 2006, Xu et al., 2005). They correlate host and parasite transcriptional changes to identify interacting host-parasite gene pairs and confirm the results by also analyzing protein domain information from which they infer host-parasite interactions at the molecular level.

The genome sequences of seven Eimeria spp. infecting the chicken (Reid et al. 2014) and of E. falciformis (Heitlinger et al., 2014) lay the foundation for transcriptome studies based on high throughput sequencing of RNAs (RNAseq) in these parasites. Applications of such methods to parasites provide the opportunity to simultaneously assess RNA abundance (gene expression) in both host and parasite (dual RNAseq; Westermann et al. 2015, Fernandes et al. 2016, Li et. al 2016).

By applying such dual RNA-seq to produce such host and parasite transcriptomes from the same samples and tissue, we provide a first dataset which allows analysis of host and parasite mRNA profiles at several time-points post infection in an aplicomplexan parasite. Not only can we for the first time provide transcriptomes for the full parasite life cycle of E. falciformis. We analyze biological processes and apply gene correlation analysis (Reid and Berriman, 2012) between host and parasite and discuss candidate interaction-gene pairs. We included a mouse strain which lacks a gene for maturation of T and B cell receptors, *Recombination-activating gene 1*, *Rag1*. Both naive immune competent and T and B cell deficient mice were infected, and both groups were challenge infected after recovering from the first infection. This design allows a broad comparison of infections in mice with different immune competence, and analysis of how these differences affect the parasite transcriptomes and reproductive success. Our parasite RNA-seq transcriptome is analyzed together with published transcriptomes from the chicken parasite E. tenella and T. gondii, highlighting aspects of species specificity versus conservation between these species at different life cycle stages.

RESULTS & DISCUSSION

**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 for our experiment. To follow the life cycle of the parasite, we compare different time-points after infection. We additionally used different mouse strains which display different immune-competence in infection trials, measured by oocyst output, in order to assess the influence of host immune-competence 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 deficient in the *Rag1* gene (*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"). immunocompetent (C57BL/6) and immunodeficient (*Rag1-/-*) host strains in naive and challenge infected mice.

Basic phenotyping showed decreased in oocyst output (Figure 1 a) in these challenged host compared to naïve animals, as well as a reduction in parasite 18S transcripts determined by quantitative reverse transcription PCR (RT qPCR).

These data demonstrate reduced parasite numbers and reduced reproductive success in immune competent, challenge infected mice. In contrast, no difference was observed between first and challenge infection in immune-deficient *Rag1-/-* mice. Immune competent NMRI mice were infected with higher doses of sporulated oocysts (150 in first infection and 1500 in challenge infection, which was required to see a response. Data not shown), and a drastic reduction of oocysts in faeces was seen in both strains of challenged immune competent mice. Oocyst numbers in faeces peaked on days 8-9 and all mice had cleared the infection by day 14. We thereby note that E. falciformis infection is self-limiting also in T and B cell deficient mice with the same timing as in immune competent mice. Asexual replication and increase in parasite numbers of E. falciformis intestinal stages is reflected by the percentage of parasite reads sequenced per time-point post infection (Figure 1c) mirroring RT qPCR profile of 18S transcirpts (Figure 1b). We thus use dual RNA-seq to analyse the life cycle of E. falciformis under the influence of different host immune capacity 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) to assess further transcriptional changes in host and parasite associated with lifecycle progression of the parasite and immune response in the host.

*Parasite and host dual transcriptomes can be assessed in parallel*

We purified host and parasite RNA from infected tissues, and we here demonstrate that even early in infection there is sufficient parasite material for detection and quantification by RNA-seq. 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-points, 7 dpi, the overall mRNA output of the sampled caecum tissue is dominated by parasite material with proportional parasite mRNA abundance of 77% (in NMRI\_1stInf\_7dpi\_rep2) and 92% (in NMRI\_1stInf\_7dpi\_rep1) (Figure 1c) suggesting heavily infected tissues and large parasite numbers at this time-point and/or transcriptionally “silent” host cells.

*Exclusion of samples with uncertain infection status*

We report a proportional mapping of sequencing reads to the parasite genome of maximum 92% (sample NMRI\_1stInf\_7dpi\_rep1) and a minimum of 0.064% (sample NMRI\_2ndInf\_3dpi\_rep2) in samples considered infected (Table 1). We excluded samples NMRI\_2nd\_3dpi\_rep1 (0.012%) and NMRI\_2nd\_5dpi\_rep2 (0.023%) 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). Existing analysis methods of RNA-seq data (e.g. edgeR, DEseq) assume a negative binomial distribution of transcript counts for reliable normalization and differential gene expression analysis. It is also likely that the number of reads in the excluded samples would have been insufficient to fully normalize these datasets to those with the highest parasite contributions. From a biological point of view, both excluded samples are samples from challenge infection and it is likely that the infection had been cleared or reduced to a non-detectable level. One additional sample (NMRI\_1stInf\_0dpi\_rep1) was excluded because the uninfected control showed unexpected mapping of reads to the E. falciformis genome (0.033%). We consider the three excluded samples to display an uncertain state of infection.

**The mouse transcriptome undergoes broad changes upon *E. falciformis* infection**

We here show that upon infection with E. falciformis, which induces weight loss and intestinal pathology in mice, the host transcriptome undergoes drastic changes with more than 3000 genes changing their mRNA profile. Statistical testing for differential expression between infected and uninfected mice revealed changes in mRNA abundance becoming more pronounced (both in magnitude and number of genes affected) at later time-points post infection (Table 2 and Figure 2A). 325 mouse 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), which would be the case if the same genes are regulated throughout infection.

Instead, the transcriptional profile of the mouse changes more fundamentally with other genes regulated late compared to early infection. Our results are in agreement with previously published microarray data from mice infected with E. falciformis species. Differences between uninfected controls and samples 7 dpi were correlated with fold-change data obtained from E. falciformis infected mice at 6 dpi on Agilent microarrays (Schmid et al. 2014). Differential abundances between the data-sets show a strong correlation (Spearman's rho = 0.74; Figure 2a). Considering both biological differences in the experiments such as exact time-points for sampling, and technical differences between the two methods, this comparison confirms the adequacy of using dual RNA-seq for assessing the host transcriptome.

A distinct host response in late infection is indicated by over 1600 DA mRNAs on 7 dpi (Figure 2b). To analyze this further, we performed hierarchical clustering on the 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, including infected samples, cluster with unifected samples. The similarity between infected and non-infected *Rag1-/-* samples, confirms the immune deficiency phenotype (failure to react to infection) in these mice and suggest a strong influence of adaptive immune responses on overall transcriptional response.

***B and T-cell dependent and independent early responses to infection***

In immune compromised *Rag1-/-* hosts, a group of genes (cluster 3, Figure 2c) display a profile simmilar to immune competent hosts upon infection: RNA of these genes is generally reduced in abundance upon infection. Hence, these genes appear to not depend on functional T and B cells in their response to infection including a response at the earliest measured time-point (3 dpi). The Gene Ontology (GO) terms "lipid metabolic process" and "protein intraciliary transport" and "blood coagulation" are enriched in these genes (SI file x). Our data suggests that these processes are regulated even in *Rag1-/-* mice and hence might constitute an innate response to infection independent of B- and T-cells.

Another group of genes (cluster 4, Figure 2c) changes its transcriptional profile upon infection in immune competent mice only. The direction of these changes is, as for B- an T-cell independent changes before, from a high abundance in unifected to low a abundance in infected mice. Corresponding mRNAs thus appear to be downregulated upon infection in a T and B cell dependent manner i.e. by processes which depend on (mature) T and/or B cell activity, also as early as 3 dpi. Enriched GO biological processes include "signalling", "cell communication", “response to external stimuli", “wound healing” and “proliferation”. This shows that epithelial healing is induced. Enriched terms for “cytokines”, "negative regulation of viral (or inflammatory) response", "negative chemotaxis", "autophagy", "blood coagulation", "inositol phosphate-mediated signaling", and "positive regulation of calcineurin-NFAT" are addionally enriched.

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 differential abundance of transforming growth factor-β, 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 regulation are among the enriched GO terms annotated in genes with this reduced abundance upon infection in a B and T-cell dependent manner. 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 TGF$beta$, 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. This also implies that hosts invest resources in intestinal healing and possible regulatory functions (IL-1 down) at early time-points of E. falciformis infection.

These differences between immunocompetent and immunodeficient *Rag1-/-* mice suggest 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. However, severe pathology seen in infected *Rag1-/-* mice makes this scenario unlikely.

A distinct group of genes shows a similar response in the opposite direction, an upregulation upon infection limited to immunocomptent animals, not observed in Rag1-/- (cluster 1 in Figure 2c).

***Adaptive immune responses characterizes late infection***

Pronounced transcriptional changes late in infection (7 dpi; cluster 5 in Figure 2c) reflect the expected onset of an adaptive immune response. This is confirmed by enriched GO 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). These terms highlight an activated immune system and adaptive immune responses at 7 dpi. Among the same genes, natural killer cell regulation, JAK-STAT signaling, 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 abundance differences later in infection (7 dpi) and it is encouraging that these links are detected.

In addition, RNA processing distinguishes challenge infected mice from naïve infected mice.Three challenge infected samples (3 dpi, 5 dpi and 7 dpi) from immune competent mice show a distinct mRNA profile (cluster 6 in Figure 2c). These mRNAs are highly enriched for GO terms for RNA processing and splicing, as well as terms for histone and chromatin modification. These genes might point to differences in transcriptional activation and control of an adaptive immune response in challenged vs. naive mice.

Taken together, RNA abundance patterns of immunocompetent versus immunocompromised hosts suggest a process strongly influenced by adaptive immunity. The development of this immune reactions is the major biological processes which characterizes the infection. We additionally show regulation of IL-1, IL-6, TNF, TGFβ and EGF especially late (7dpi) in both first and challenge infections and suggest wound healing as a probable function of this regulation.

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

We present data which supports that E. falciformis is "transcriptionally blind" to the host immune status, since comparisons between differently immune-competent mice return no significantly different mRNAs. 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 DA mRNAs in the comparisons described above and applied GO enrichment analysis of gene clusters based on annotations from Heitlinger et al. 2014. This includes annotations with “conservation categories” derived phylogenetic patterns inferred for gene families in *E. falcifomis*.

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. Major patters in the parasite transcriptome 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, cluster 6) in which schizogony, asexual reproduction, takes place, and separately, late infection (7 dpi, clusters 2 and 7 ) in which we assume gametocytes to be present. Extracellular samples of sporozoites and oocysts also cluster separately and are defined by distinct gene clusters.

*Early infection transcriptomes reflect parasite expansion*

Sporozoites are defined by a group of genes (cluster 4 and Additional file 5) which are largely specific to E. falciformis, meaning that no orthologs are found as gene family members outside of this species. Interestingly, five out of 12 SAG genes (Heitlinger 2012) are among the mRNAs highly abundant in sporozoites, confirming their naming as “sporozoite suface antigens”. These surface bound proteins are thought to be involved in host cell attachment and possibly in inducing immune responses. The high enrichment in an easily accessible life cycle stage opens for analysis of the function of SAG genes in E. falciformis. Our GO enrichment data on sporozoites further suggests ATP production and biosynthesis processes as dominant features in sporozoites. Metabolic pathway analysis (ToxoDB) of the same genes reveals several pathways which suggest that ATP production takes place, including fatty acid degradation, oxidative phosphorylation and valine, leucine and isoleucine degradation (Additional file 6). The latter degradation of branched chain amino acids pathway is enriched for genes which have orthologs in T. gondii, where they are used to utilize amino acids for energy supply (Oppenheim 2014). In addition, the invasive stage is characterized by "maintenance of protein location in cell" and similar terms. Possibly, this is due to control of microneme or rhoptry protein localization as sporozoites prepare for invasion.

In stages following in the progression of the lifecycle are less distinct. Measured at 3 dpi and 5 dpi we expect several rounds of schizogony take place in an unsynchronized fashion. Abundand mRNAs cannot be destinguised between all early (3 dpi and 5 dpi) samples (cluster 6 in Figure 3c). Among these "early infection" genes, several GO terms (biological process) for biosynthetic acitivity are enriched, e.g., "ribosome biogenesis", "cellular biosynthetic process", "gene expression", “RNA processing”, including terms for “tRNA and ncRNA processing”. "Cellular amino acid catabolic process" is also enriched. Biosynthetic processes are enriched both in early infection and in invasive sporozoites, but driven by different mRNAs.

Analyzing these lifecycle stage defining gene groups, sporozoites specific genes showed a species specific profile (low number of orthologs outside E. falciformis), whereas early infection mRNAs are enriched for genes shared as orthologs by 10 apicomplexan species. This could reflect that similar processes and genes are involved in asexual reproduction among these apicomplexan species, but invasion may be a more species specific event. Among the early infection genes we found 4 out of 10 rhoptry kinases, ROPs, described for the E. falciformis genome. This is the largest number of ROPs in any group of genes in our analysis (EfaB\_PLUS\_24117.g1969, EfaB\_MINUS\_17096.g1521, EfaB\_MINUS\_42996.g2710 and EfaB\_PLUS\_7742.g778). Among the sporozoite defining genes are two rhoptry kinases (EfaB\_PLUS\_8664.g829 and EfaB\_PLUS\_15899.g1411).

Enrichment of replication and growth-related processes highlights the parasite's expansion in numbers on 3 dpi and 5 dpi. This is supported both by previous knowledge about the life cycle and by the increase in parasite derived sequences we measured (Figure 1b). The mRNAs supporting the processes described here all have a low abundance late in infection, on 7 dpi. This can be explained by a switch from the early asexual expansion towards sexual reproduction on the time-point one day before oocyst output peaks.

*Gametocytes likely determine transcriptome late in infection*

Two E. falciformis gene clusters show a distinct profile characterized by 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. Late stage defining mRNAs are enriched either for orthologues shared among core apicomplexan parasites (Apicomplexa excluding Cryptosporidium; cluster 2) or for genes conserved bejond the Apicomplexa (cluster 7, Additional file 5). 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 genes support ATP production in this life-cycle stage compared to sporozoites. For both gene groups that define late infection, different (peptide/nitrogen compound/cellular protein/macromolecule) "biosynthetic process" terms along with "chitin metabolic process" suggest that the parasite is producing building blocks for oocysts. This fits the timing of oocyst output which peaks at 8-9 dpi, one day after sampling for RNA-seq. In addition, the Apicomplexa specific genes expressed in oocysts are enriched for a number of distinct GO terms such as "blood coagulation". These reflect the presence of Thrombospondin type I domains in the protein products these mRNAs. Thrombospondin type 1 domains have been reported in E. tenella microneme localizing proteins, MIC, e.g. MIC4 (Tomley01). MIC4 mRNA was reported in E. tenella sporozoites where it localizes to the apical end, and in late schizonts and late oocyst stages, when sporozoites are forming. This has so far not been demonstrated in E. falciformis. Possibly, the mouse parasite's MICs play a role in gamete stages or oocyst formation, or E. falciformis prepares for invasion already in during oocyst formation, however this is highly speculative.

*Oocysts are characterized by stress responses and differentiation*

Overall, the oocyst profile with five out of seven gene clusters being characterized by below average abundance of mRNAs, can be expected in this inactive life cycle stage which endures long-term survival outside the host. However, two groups of mRNA are highly abundant (clusters 1 and 5). One of the gene sets (cluster 5) is enriched for apicomplexan-shared orthologues and contain GO terms for "DNA repair", "protein modification process" and "cell differentiation". GO enrichment in the other group of genes highlights only one term (adj. P-value 0.11): "DNA-templated transcription, initiation". Enrichment for these terms, along with initiation of transcription and cell differentiation are likely a result of DNA replication and DNA quality control which takes place during sporulation

*Evolutionary conservation in life cycle-characteristic gene groups*

A good understanding for similarities and differences between closely related parasites will be useful to draw parallels between the species, and importantly also to avoid incorrect assumptions about, e.g., gene function. One way to achieve such understanding is to identify which genes are species-specific, phylum specific or broadly conserved. In addition to the above described enrichment analysis for differently conserved orthologues we performed Spearman's 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. This analysis confirms the species specificity for the sporozoite transcriptome. Gene expression in E. falciformis sporozoite is more similar to E. falciformis early infection (3 dpi and 5dpi) than to orthologous genes in E. tenella. Conversely mRNA abundance in E. falciformis late infection samples correlate best with with orthologs in E. tenella gametocytes. This indicates a similarity between species at this stage of the lifecycle and supports the presence of gametocytes in our samples. The transcriptomes of E. tenella merozoites from both independent studies are most similar to early (3 dpi and 5 dpi) E. falciformis samples, indicating similarity also during asexual reproduction which is also shown by the conservation analysis (Figure?/Table?). Cross-species similarity is thus driven by genes upregulated during asexual expansion (cluster 6 in Figure 3c). These mRNAs determine early stages and are enriched for shared apicomplexa orthologs. E. falciformis oocysts cluster with unsporulated E. tenella oocysts, whereas E. tenella sporulated oocysts are most similar to E. tenella sporozoites, potentially reflecting differences in the in vitro preparation of these “environmental stages”.

We have identified groups of genes which are shared among apicomplexa and linked them to the asexual reproduction phase of E. falciformis. The analysis also highlights E. falciformis specific genes, abundant in sporozoites. Genes predominantly expressed in E. falciformis oocyst are underrepresented for E. tenella orthologs and can perhaps be used to understand differences between mammalian and avian Eimeria spp. oocysts.

**Imperfect clustering might reflect true biological differences**

Hierarchical clustering does not cluster replicates together especiall for the early stages of infection at 3 and 5 dpi. At a first glance this might suggest technical problems. On the other hand, for distinct samples such as oocysts and sporozoites replicates do cluster, and all 7 dpi samples form one cluster for parasite gene expression. Also, considering the early/late infection patterns in the parasite data and that these fit well with previous knowledge about asynchronous schizogony and gametocyte formation around 7 dpi, it is worthwhile to consider whether replicate separation reflects true biological variation. It is perceivable that the overall course and intensity of infection varies with small differences in, e.g., host stress levels due to litter mates, draught, differences in light exposure or other factors which do vary also in a controlled animal facility. Variation in the course of infection triggered by such factors could explain the transcriptional profiles we see, with distinct overall patterns but replicate separation in the asexual phase. We suggest that considering such possibilities is important for interpreting results and understanding basic biology of the parasite and, in extension better understand infections in less homogeneous hosts than laboratory mice.

Additionally, in the context of a dual RNAseq experiment, this variation might allow us to identify genes correlated in their expression between host and parasite beyond the major patterns in the data (e.g. adaptive immunity and gamete formation at later time points).

We thus applied the “ISIGEM method” of Reid and Berriman (2013) to enrich for genes showing a correlated expression pattern across host and parasite. Results indicate that interacting genes are enriched for XXX analysis still running XXX. We highligh that those genes are not enricht for of any of the clusters reported for distinct host and parsite response along the lifecycle. They instead constitute an independent gene-set. The identification of these genes was only possible making use of (partially unexplained) variation between samples. It is thus based on the assumption independent of the experssion profile of a particular sample, host and parasite were recorded.

**Conclusions**

In summary, we have performed the first dual RNA-seq transcriptome at representative timepoints of the entire life-cycle 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 adds insight into the host responses to this intracellular infection, e.g., demonstrating that the host transcriptome is different in a first versus challenge infection, whereas the parasite seems to be transcriptionally inert to the differences in mouse immune competence. We show that the transcriptional profile of sporozoites is characterized by and upregulation the mostly species specific genes. The sporozoite defining gene group is characterized by ATP production, regulation of protein localization and biosynthetic processes. The analysis further demonstrates that early infection is not synchronized between samples, neither in host nor parasite. For the parasite, the conformity which is seen in sporozoites and oocysts is however regained late in infection, where both parasite and host transcriptomes are synchronized between samples. We speculate that parasites during asexual replication are more asynchronous than developing gametes and oocysts on 7 dpi. The distinct profile on 7 dpi in both host and parasite suggests that synchronization is driven by interactive processes between the species. Whether gamete formation triggers the adaptive immune responses seen in the host, or previously activated immune responses lead to gamete and oocyst formation remains to be investigated.

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 $mu$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 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. (2012). 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 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 (Schmid 2012).

*Sporozoite isolation*

Sporozoites were isolated from sporocysts by in vitro excystation. For this, sporocysts were incubated at 37newcommand{degreeC 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). Purified 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 and quality assessment*

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 with very lax phred score of 10. These base settings require that nine out of ten bases or more are correct in at least 60% of the bases for each read alignment. We intentionally did not use a stringent trimming before mapping to genome assemblies.

*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 (Heitlinger et al., 2014) was downloaded from ToxoDB (Gajria et al., 2007). For the alignment, the mouse and parasite genome files were merged into a combined reference genome, and files including mRNA sequences from both species were aligned against this reference using TopHat2, version 2.0.14, (Trapnell et al., 2009) with –G specified, and a Bowtie2, version 1.1.2, (Langmead and Salzberg, 2012). Single-end and pair-end sequence samples were aligned separately with library type 'fr-unstranded' specified for pair-end samples. Bam were used as input for the function “featureCounts” from of the R package “Rsubread” (Liao et al., 2014).

*Differential mRNA abundance, data normalization and sample exclusions*

After import of data to R, mouse and parasite data was separated using obvious patterns in transcript IDs. For each species separately, count data was normalized using the R-package edgeR (version 3.14.0; cite) with the upperquartile normalization method. Briefly, genes with a coverage below 3000 (mouse ) and XXX (E.falciformis) summed over all samples (libraries) were removed and normalization factors were calculated for the 75% quantile for each library. This normalization 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 ...).

We then used EdgeR to fit generalized linear models (GLMs with a negative binomial link function) for each gene (glmFit) and to performs likelihood ratio tests for models with or w/o a focal factor (glmLRT).

*Selection of differentially abundant mRNAs and hierarchical clustering*

Mouse mRNAs responding to infection or differently abundant at different timepoints of infection (0 vs “any dpi” or “any dpi” vs “any dpi”; see Table 2) and E. falciformis genes showing differences between any life cycle stage (oocysts vs sporozoites, or any of those vs. “any dpi” or “any dpi” vs. “any dpi”) were selected and used for hierarchical clustering. Hierarchical clustering was performed using the complete linkage method based on Euclidean distances.

*Enrichment tests*

Gene Ontology (GO) enrichment analysis was performed using the R-package topGO with the “weight01” algorithm and Fisher's exact tests. We additionally perfomrmed a correction for multiple testing on the returned p-values (p.adjust using the BH-method). Similarly a Fisher's exact test and corrections for multiple testing were also used to test for an overrepresentation of transcripts with a signal sequence for entering the secretory pathway or with transmembran domains (as inferred using Signal P).

Evolutionary conservation of gene families was analyzed based from categories from Heitlinger et al. (2014) which are as follows: i) E. falciformis specific, ii) specific to the genus Eimeria, compiled by an analysis of E. falciformis, E. maxima and E. tenella, iii) Coccidia: Eimeria plus T. gondii and Neospora caninum, iv) Coccidia plus Babesia microti, Theileria annulata, Plasmodium falciparum and Plasmodium vivax v) the same apicomplexan parasites plus Cryprosporidium hominis, vi) universally conserved in the eukaryote super-kingdom inferred from an analysis of Saccharomyces cerevisiae and Arabidopsis thaliana. These categories were tested for overrrepresentation in gene-sets with particular patterns described in the text using Fisher's exact-tests and the resulting p-values were again corrected for multiple testing.

*Correlation analysis of apicomplexan transcriptomes*

Transcriptome datasets from Reid et al (2014), Walker et al. (2014) and Hehl et al (2012) were downloaded from ToxoDB (ref). Orthologs between E. falciformis, E. tenella and T. gondii were compiled as defined in Heitlinger et al. (2014) and only 1:1:1 otholog triplets were retained for analysis, as multi-paralog gene-families might contain member showing divergent evolution of gene-expression due to neo/sub functionalisation. Spearman correlation coefficients for expression over different samples in all studies and over different species represented by their orthologs were determined. Hierarchical clustering with complete linkage was used to cluster resulting correlations coefficients.

*Correlation analysis between host and parasite*

We used the ISIGEM method of Reid and Berriman (2013) to asses correlations of host and parasite transcripts. This method avoids spurious correlations for host-parasite gene pairs remaining e.g. unchanged over many conditions. It does so by estimating emmpircal p-values (ISIGEM-scores), which correspond to the proportion of cases in which a random shuffeling of samples produces better correlations than those observed in the given samples between host-parasite pairs. Similar to Reid and Berriman (2013) we used 5x10⁵ replicates of random shuffels and did not correct for multiple testing in our set of (79,764,454 tested correlations between 13,343 mouse and 5978 parasite genes).

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 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: all?

**ACKNOWLEDGEMENTS**

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

Tables

Figures(Legends only; figures in separate files)

Figure 1. Oocyst output and changes in intensity of E. falciformis infection in mouse. A) Oocyst counts in first and challenge infection are shown for different mouse strains. Immune competent mice (NMRI and C57BL/6) have reduced oocyst counts in challenge infection, whereas immune incompetent Rag1-/- mice do not. B) RT-qPCR data of E. falciformis 18S in NMRI mice displays an increase in parasite material over the course of infection. Less parasite material is detected in challenge infected mice. C) The percentage of parasite mRNA detected by RNA-seq increases during infection. More material is detected in naïve mice compared to challenge infected mice, as is also demonstrated in B. Sporozoites and oocysts are shown with 100% parasite material.

Figure 2: Differential abundance of mouse mRNAs and clutering of abundance profiles – A) Log2 fold-changes obtained from a comparison of RNAseq data at 7dpi vs. unifected controls are plotted against log2 fold changes obtained in microarray study (Schmid et al. 2014) at 6dpi vs. similar controls. B) I) Overlap between genes showing differential abundance (FDR < 0.01; EdgeR glm likelihood-ratio tests) between uninfected controls and different time-points after infection and ii) between different time points and all those genes reacting to infection is viszalized in venn diagramms (C) Hierarchical clustering of expression profiles for these mRNAs performed on Euklidean distances using complete linkage. Cluster cut-offs were set to identify gene-sets with profile interpretable in relation to the parasite lifecycle and between mice of different immune competence.

Figure 3 Differential abundance of E. falciformis mRNAs and clutering of abundance profiles – A) i) Overlap between genes showing differential abundance (FDR < 0.01; EdgeR glm likelihood-ratio tests) between intracellular stages at 3dpi, 5dpi and 7dpi. ii) Overlapp between earls vs. late chanes deduced from this and between “environmental stages” (oocysts and sporozoites) (C) Hierarchical clustering of expression profiles for these differentially abundand mRNAs performed on Euklidean distances using complete linkage. Cluster cut-offs were set to identify gene-sets with profile interpretable in relation to the parasite lifecycle.

Figure 4: Correlation of the E. falciformis transcriptome with other coccidian transcriptome- Correlations of our mRNA abundance data with orthologs from published RNA-seq transcriptomes from E. tenella (Reid et al. 2014, Walker et al. 2014) and T. gondii (Hehl et al 2012) are shown. Correlations coffeicients (Spearman's rho) were clustered using complete linkage. T. gondii and Eimeria spp. “late infection” samples cluster together. E. falciformis early infection samples cluster with E. tenella merozoites. E. falciformis sporozoites cluster with E. falciformis early infection, whereas oocysts cluster with E. tenella unsporulated oocysts.

Figure 5: Pairwise interaction of host and parasite genes deduced from correlated expressions – TO BE ADDED

**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

separately 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 results

Document containing additional figures and summary tables (pdf).

**Additional file 4**: Results of enrichment analyses (topGO)

**Additional file 5**: Results of enrichment analysis for categories of differently conserved gene orthologues

**Additional file 6**: Result of metabolic pathway enrichment of sporozoite defining genes

Tables listing all tested gene sets and resulting significant GO terms.

REFERENCES

Frazee, A.C., Pertea, G., Jaffe, A.E., Langmead, B., Salzberg, S.L., Leek, J.T., 2015. Ballgown bridges the gap between transcriptome assembly and expression analysis. Nat. Biotechnol. 33, 243–246. doi:10.1038/nbt.3172

Gajria, B., Bahl, A., Brestelli, J., Dommer, J., Fischer, S., Gao, X., Heiges, M., Iodice, J., Kissinger, J.C., Mackey, A.J., Pinney, D.F., Roos, D.S., Stoeckert, C.J., Wang, H., Brunk, B.P., 2007. ToxoDB: an integrated Toxoplasma gondii database resource. Nucleic Acids Res. 36, D553–D556. doi:10.1093/nar/gkm981

Heitlinger, E., Spork, S., Lucius, R., Dieterich, C., 2014. The genome of Eimeria falciformis - reduction and specialization in a single host apicomplexan parasite. BMC Genomics 15, 696. doi:10.1186/1471-2164-15-696

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. doi:10.1038/nmeth.1923

Liao, Y., Smyth, G.K., Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930. doi:10.1093/bioinformatics/btt656

Lovegrove, F.E., Peña-Castillo, L., Mohammad, N., Liles, W.C., Hughes, T.R., Kain, K.C., 2006. Simultaneous host and parasite expression profiling identifies tissue-specific transcriptional programs associated with susceptibility or resistance to experimental cerebral malaria. BMC Genomics 7, 295. doi:10.1186/1471-2164-7-295

Mesfin, G.M., Bellamy, J.E., 1978. The life cycle of Eimeria falciformis var. pragensis (Sporozoa: Coccidia) in the mouse, Mus musculus. J. Parasitol. 64, 696–705.

Periz, J., Gill, A.C., Hunt, L., Brown, P., Tomley, F.M., 2007. The Microneme Proteins EtMIC4 and EtMIC5 of Eimeria tenella Form a Novel, Ultra-high Molecular Mass Protein Complex That Binds Target Host Cells. J. Biol. Chem. 282, 16891–16898. doi:10.1074/jbc.M702407200

Reid, A.J., Berriman, M., 2012. Genes involved in host–parasite interactions can be revealed by their correlated expression. Nucleic Acids Res. gks1340. doi:10.1093/nar/gks1340

Schmid, M., 2014. Eimeria falciformis Infection of Mouse Cells Identifies Host Determinants of Parasite Development. Humboldt-Universität zu Berlin.

Schmid, M., Lehmann, M.J., Lucius, R., Gupta, N., 2012. Apicomplexan parasite, Eimeria falciformis, co-opts host tryptophan catabolism for life cycle progression in mouse. J. Biol. Chem. 287, 20197–20207. doi:10.1074/jbc.M112.351999

Stange, J., n.d. Studies on host-pathogen interactions at mucosal barrier surfaces using the murine intestinal parasite Eimeria falciformis.

Stange, J., Hepworth, M.R., Rausch, S., Zajic, L., Kühl, A.A., Uyttenhove, C., Renauld, J.-C., Hartmann, S., Lucius, R., 2012. IL-22 mediates host defense against an intestinal intracellular parasite in the absence of IFN-γ at the cost of Th17-driven immunopathology. J. Immunol. Baltim. Md 1950 188, 2410–8. doi:10.4049/jimmunol.1102062

Trapnell, C., Pachter, L., Salzberg, S.L., 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111. doi:10.1093/bioinformatics/btp120