

# Proteomic and transcriptomic analyses of early and late-chronic *Toxoplasma gondii* infection shows novel and stage specific transcripts

**Andrew Garfoot**

University of Wisconsin Madison

**Gary Wilson**

University of Wisconsin Madison

**Joshua Coon**

University of Wisconsin Madison

**Laura Knoll** (✉ [ljknoll@wisc.edu](mailto:ljknoll@wisc.edu))

University of Wisconsin Madison <https://orcid.org/0000-0002-9362-8088>

---

## Research article

**Keywords:** *Toxoplasma gondii*; cysts; bradyzoites; RNA-seq; proteomics

**Posted Date:** October 25th, 2019

**DOI:** <https://doi.org/10.21203/rs.2.12868/v3>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published on November 14th, 2019. See the published version at <https://doi.org/10.1186/s12864-019-6213-0>.

# Abstract

**Background :** The protozoan pathogen *Toxoplasma gondii* has the unique ability to develop a chronic infection in the brain of its host by transitioning from the fast growing tachyzoite morphology to latent bradyzoite morphology. A hallmark of the bradyzoite is the development of neuronal cysts that are resilient against host immune response and current therapeutics. The bradyzoite parasites within the cyst have a carbohydrate and protein-rich wall and a slow-replication cycle, allowing them to remain hidden from the host. The intracellular, encysted lifestyle of *T. gondii* has made them recalcitrant to molecular analysis *in vivo* .

**Results :** Here, we detail the results from transcriptional and proteomic analyses of bradyzoite-enriched fractions isolated from mouse brains infected with *T. gondii* over a time course of 21 to 150 days. The enrichment procedure afforded consistent identification of over 2,000 parasitic peptides from the mixed-organism sample, representing 366 *T. gondii* proteins at 28, 90, and 120 day timepoints. Deep sequencing of transcripts expressed during these three timepoints revealed that a subpopulation of genes that are transcriptionally expressed at a high level. Approximately one-third of these transcripts are more enriched during bradyzoite conditions compared to tachyzoites and approximately half are expressed at similar levels during each phase. The *T. gondii* transcript which increased the most over the course of chronic infection, *sporoAMA1*, shows stage specific isoform expression of the gene.

**Conclusions :** We have expanded the transcriptional profile of *in vivo* bradyzoites to 120 days post-infection and provided the first *in vivo* proteomic profile of *T. gondii* bradyzoites. The RNA sequencing depth of *in vivo* bradyzoite *T. gondii* was over 250-fold greater than previous reports and allowed us to identify low level transcripts and a novel bradyzoite-specific isoform of *sporoAMA1*.

## Background

The protozoan parasite *Toxoplasma gondii* is one of the most successful eukaryotic pathogens, infecting approximately a quarter of the world's population [1]. One of the drivers of its success as a pathogen is the ability to develop a chronic infection in the brain of any warm-blooded host. Within the brain, the parasite undergoes a transformation from the fast-growing tachyzoite form to the slow-growing bradyzoite form [2]. Bradyzoites remain shielded from the host immune system by changing surface protein expression and developing a cyst wall [2, 3]. *T. gondii* cysts cannot be cleared from the host and current therapeutics are ineffective against chronic infection. This persistence becomes particularly problematic if the host becomes immunocompromised. As the immune pressure wanes, bradyzoites transition back into tachyzoites and start to rapidly replicate, which can lead to encephalitis if left untreated [4]. Despite its clinical importance, the biology of bradyzoites and the molecular components of the cyst structure are not well understood.

*T. gondii* undergoes a major transformational change during the switch from a tachyzoite to a bradyzoite. The transition has been characterized by both microarray analysis of cysts 21 days post-infection (DPI) [5] and deep sequencing of cysts 28 DPI [6], which identified ~500-800 genes changing between the different stages *in vivo*. This data has provided a good snapshot of cysts during early chronic infection; however, bradyzoites have a dynamic growth pattern during chronic infection. Parasite growth measured

by ICM3 protein abundance showed a cyclic growth pattern of bradyzoites out to 8 weeks post-infection [7]. These results suggest that cysts are not static but are continually changing and responding to the environment.

The documented changes in cysts throughout chronic infection has led us to question whether the global transcriptional profile of bradyzoites changes over the course of chronic infection. For this study, we isolated *T. gondii* cysts from the infected brain tissue of mice 21, 28, 90, and 120 DPI using dextran centrifugation. This purification method resulted in enough parasite material to analyze both the global RNA and protein profile of *in vivo* bradyzoites. While the proteins present on the cyst wall have been recently analyzed [8], this current study is the first proteomic analysis of *T. gondii* bradyzoite parasites and includes the transcription profile of *T. gondii* bradyzoites throughout the time course of chronic infection.

## Results

*Enrichment of bradyzoite cysts decreases host protein interference.* *T. gondii* bradyzoite cysts were enriched from homogenized brain of host mice using dextran density centrifugation [9] for *in vivo* RNA-seq and proteomic analyses (Fig. 1A). Each replicate was a pool of several mouse brains (ranging from 3 mice to 13 mice per group, Fig. 1B). Bradyzoites were released from cysts by pepsin digestion. Approximately two-thirds of the parasites from each replicate were processed for protein analysis and one-third was processed for RNA (Fig. 1C). The highest parasite burden was at 28 and 90 DPI, and the lowest burden at the earliest (21 DPI) and latest (150 DPI) timepoints (Fig. 1B). As the chronic infection progresses, survival decreases [10], which limited the number of mice harvested at the latest timepoints. With few mice and lower parasites per brain, the 150 DPI timepoint resulted in the fewest total parasites collected (Fig. 1C).

Protein samples were analyzed by bottom-up LC-MS/MS to identify parasitic and host proteins. The total number of unique peptides ranged from 11,548 to 24,833 (Additional File 1). Dextran purification of the infected brains increased the number of peptides that mapped to the *T. gondii* genome for many of the samples. Without dextran purification, protein from infected brain tissue mapped 3.4% of the peptides to *T. gondii* (black bar, Fig. 2A), whereas many purified samples mapped more than 25% of the peptides to *T. gondii*. Notably, the 21 DPI samples were near the same *T. gondii* mapping percentage as unpurified tissue, possibly because the bradyzoite cysts were not mature enough to sediment under the dextran. The number of unique *T. gondii* peptides ranged from 462 at 21 DPI, to 8,163 at 28 DPI, and correlated with the number of bradyzoites in the sample (Fig. 2B). These peptides mapped to 1,683 unique *T. gondii* proteins (Additional File 2). We chose two samples in each of the middle timepoints (28, 90 and 120 DPI) with the highest *T. gondii* peptide count and enrichment for further analysis (Fig. 2A arrows). From these samples, we required peptides be identified in both replicates of each timepoint. This resulted in 6,528 peptides identified at 28 DPI, 3,617 at 90 DPI and 3,486 at 120 DPI, with 2,040 peptides common between all three timepoints (Fig. 2C). The identified peptides represent 870 proteins at 28 DPI, 504 proteins at 90

DPI, and 502 at 120 DPI for a total of 893 unique proteins with 366 common between all timepoints (Fig. 2D, Additional File 3).

*StringTie program predicts novel T. gondii bradyzoite transcripts.* RNA from the six samples with the highest percentage of *T. gondii* peptides (Fig. 2A arrows) were processed for Illumina sequencing. Samples were multiplexed and sequencing acquired over 265 million total reads, averaging 44 million reads per sample (Table 1). These samples were highly enriched for *T. gondii*, as 52-94% of the reads aligned to the *T. gondii* genome (21-41 million *T. gondii* reads per sample). This coverage is comparable to that of tissue culture tachyzoites, which ranged from 45-83% reads aligning to *T. gondii* (approximately 20 million total reads). This coverage is 265-fold greater than the number of reads acquired from infected whole brain tissue [6], which mapped only 0.1% of reads to *T. gondii*.

To obtain higher transcript resolution for the bradyzoite datasets, we used the StringTie program [11] to predict novel transcripts (Fig. 3). The alignment files of all datasets (Table 1) were used to update the original *T. gondii* annotation. This update changed the number of predicted genes from 8,920 to 8,805 and the number of predicted transcripts from 8,920 to 10,668. This updated annotation was used to remap the reads. Subsequently, the gene (Additional File 4) and transcript (Additional File 5) abundance was quantified using RSEM and DESeq2 [12, 13].

The principal component analysis (PCA) calculated from the normalized DESeq2 values shows distinct clustering of samples by their experimental groups (Fig. 4A). Tissue culture tachyzoites group on the far right of the X-axis, while the purified bradyzoites group on the left. Whole brain tissue lays in the middle, with acutely infected samples skewed towards tachyzoites and chronically infected samples skewed toward the bradyzoites. All timepoints show a similar number of highly expressed transcripts with a normalized expression value (transcripts per million; TPM) >50, as well as moderately expressed (TPM 11-50) transcripts. The overall number of low expression transcripts (TPM <10) are only similar between tachyzoites and purified bradyzoites (Fig. 4B). The lower read depth of *T. gondii* reads in the unpurified brain samples likely contributed to the increase in undetected transcripts and thus a shift away from the purified 28-day bradyzoite samples on the PCA plot.

*Most identified proteins had high gene expression.* We next analyzed normalized expression calculated from RSEM (TPM values). The 'gene abundance' output of RSEM is aggregated from all transcripts that map to each annotated gene. We focused on the 366 proteins identified by mass spectrometry that were common between all timepoints (Additional File 3). Of the 366 proteins, 266 were highly expressed in the purified and whole brain chronic infection samples with an average TPM >100 (Fig. 5). Of these, 89 genes were >2-fold higher during chronic infection relative to acute infection (Fig. 5, Group I, the average of all chronic TPM values relative to tachyzoite and acute whole brain sample). The genes in this subset include known bradyzoite markers such as BAG1, ENO1, and LDH2. Gene ontology (GO) enrichment analysis shows that DNA interaction and chromatin assembly GO terms are common in this group (Additional File 6), suggesting that bradyzoites require structural changes in its chromatin for bradyzoite gene expression. Also, the 200 highest expressed genes in the enriched bradyzoites had enriched GO

terms related to translation and peptide synthesis (Additional File 6). Approximately half of the genes are highly expressed during both tachyzoite and bradyzoite stages (Fig. 5, Group II, >100 average TPM value and <2-fold difference between stages). These highly expressed genes include many housekeeping genes such as genes for tubulin, actin, and GAPDH as well as several metabolism related GO terms enriched among this group (Additional File 6). Interestingly, 100 proteins that were identified in all 6 bradyzoite samples had low gene expression (Fig. 5, Group III, <100 TPM), with 25 of these genes are higher during acute infection (Fig. 5, Group IV). GO enrichment analysis of groups III or IV showed no GO terms significantly enriched.

*Transcriptomic analysis of purified bradyzoites results in a higher number of differentially expressed genes.* As described above, we used the StringTie program to predict novel transcripts (Fig. 3). Using this updated annotation, we remapped the sequencing reads from our previous acute and chronic infection whole brain datasets to analyze differential expression with DESeq2. Using the new annotation, we identified 643 transcripts changing in abundance between acute and chronic whole brain tissues using a q-value (false discovery rate) threshold <0.05 and 2-fold cut-off (Fig. 6A). This number compares to the 547 genes previously identified [6]. As mentioned, the dextran purification allowed greater depth of *T. gondii* sequencing. A direct comparison of *T. gondii* genes from whole brain tissue at 28 DPI with the purified bradyzoites from either the 28, 90 or 120 DPI resulted in approximately 2,000 differentially transcribed genes (Fig. 6A, tan colored boxes). Comparing all three purified brain bradyzoite timepoints to tissue culture tachyzoites showed approximately 4,000 of the transcripts changed during chronic infection (Fig. 6A, purple colored boxes).

*T. gondii transcription remains static after 3 months post-infection.* Comparing purified bradyzoites throughout chronic infection, 59 transcripts (from 50 genes) change at 90 DPI and 109 transcripts (from 93 genes) change by 120 DPI relative to 28 days (Fig. 6A, Additional Files 7 and 8). No transcripts change significantly between 90 and 120 DPI. Of the differentially expressed transcripts, 51 transcripts were differentially expressed at both 90 and 120 DPI: 15 increasing in abundance and 36 decreasing (Fig. 6B). Only 8 transcripts change specifically at 90 days, whereas 58 transcripts are specific for 120 days: 18 increasing and 40 decreasing in expression. 33 of the differentially expressed transcripts have an average TPM value <10 among all purified bradyzoite timepoints. This group of 33 included the transcript for SAG1 which had an average TPM value of 16 at 28 days, decreasing 3.5-fold and 4.8-fold at 90 and 120 DPI, respectively. This low level of bradyzoite expression is compared to the average TPM values of 8,000 and 13,000 for tachyzoites and acute whole brain tissue respectively. Enriched GO terms among the genes that are reduced in expression during late-chronic infection relate to protein phosphorylation and protein modification. Specifically at 120 DPI, GO terms related to polysaccharide processes were enriched, suggesting much of the developmental changes are complete by 28 DPI.

One of the transcripts which had one of the largest changes during chronic infection was for the gene *sporoAMA1* (transcript MSTRG.4390.2, ToxoDB ID: TGME49\_315730). *SporoAMA1* was identified as a sporozoite specific paralog of the AMA1 protein, which is an essential component of the moving junction during invasion of the host cell [14, 15]. While most other transcripts map to the entirety of the predicted

full gene, such as conventional AMA1 (TgME49\_255260, Fig. 7A), reads to sporoAMA1 mapped to only the 3' region of the gene and almost no reads mapped to the 5' region (Fig. 7B). Coverage for all bradyzoite timepoints starts in the predicted intron region between exons 6 and 7. Compared to 28 DPI, both 90 and 120 DPI had an approximately 7.0-fold increase in transcript abundance. RNA-seq coverage from *T. gondii* 7 DPI in the intestines of cats [16] shows a similar coverage profile of sporoAMA1, except coverage starts at the predicted intron between exons 4 and 5. Both of these isoforms for sporoAMA1 are shortened from that of sporulated oocysts [17], which include all 9 exons (Additional File 9, Fig. S1A). This coverage suggests that *T. gondii* has stage alternative isoforms of sporoAMA1 that are expressed only during late-chronic infection or the sexual stages in cats.

## Discussion

Although the host is largely asymptomatic during chronic *T. gondii* infection, bradyzoites are actively replicating within cysts [7]. Transcriptional activity of parasites during early chronic infection is high; however, the abundance of host material has hindered global analyses of *T. gondii* proteins and late-chronic transcripts. Our ability to collect tissue cysts and separate them from host material has allowed us the resolution to analyze low abundant *T. gondii* transcripts and to reproducibly identify many *T. gondii* peptides.

Dextran enrichment greatly increased the relative *T. gondii* peptide abundance. Unpurified infected brain found 3.4% of total unique peptides mapping to *T. gondii*, which increased to 30% for many of the enriched samples (Fig. 2A). Although some samples yielded fewer *T. gondii* peptides, comparing the most consistent samples between timepoints identified 2040 unique peptides in all samples, mapping to 366 different proteins (Fig. 2). While pepsin digestion removed host material and release bradyzoites from the cysts, many cyst wall proteins were likely lost due to the procedure. Despite the pepsin treatment, some known cyst wall markers, such as CST1 and GRA2, were seen [18, 19]. Also due to the enrichment procedure, we were unlikely to identify host proteins that were associated with the parasites or the cyst wall. It is interesting to note that two host transporter proteins associated with antigen processing (TAP1 and TAP2) were identified in every sample by MS/MS. Transcripts for these proteins are highly expressed in the brain during chronic infection [10], suggesting the TAP proteins may not be interacting directly with the cyst. In addition, 31 host proteins identified in our datasets were also identified from *in vitro* *T. gondii* cyst wall fractions (Additional File 2, right column) [8], suggesting there may be an association with the cyst wall among these proteins.

For many of the identified *T. gondii* proteins, the RNA expression levels were high during chronic infection (Fig. 5). These genes include the bradyzoite specific markers BAG1 and LDH2 [20, 21], as well as constitutively expressed genes such as Act1 and TubA1. Interestingly, 100 of the 366 proteins had low level gene expression during chronic infection. These results likely indicate either short-lived transcripts or long-lived proteins. Alternatively, as bradyzoites within tissue cysts are heterogeneously replicating *in vivo* [7], some proteins may be produced by a subset of metabolically active parasites or those undergoing reactivation into the tachyzoite state.

Cyst isolation has allowed us to identify transcripts from *in vivo* parasites at a depth similar to parasites *in vitro*. This depth allowed examination of alternative transcripts produced by bradyzoites from early to late chronic infection. Comparing timepoints with two biological replicates from each, over 100 transcripts were differentially expressed from early to late chronic infection. No transcripts change between 90 and 120 DPI, suggesting that, at the global level, the parasites are at a steady state by 90 DPI. This result is consistent with many host transcripts decreasing expression from 28 to 90 DPI and remaining similar from 90 to 180 DPI [10].

One of the transcripts which had the largest change from early to late infection was for *sporoAMA1*. The StringTie program identified two transcripts for *sporoAMA1* (MSTRG.4390.2 and TGME49\_315730-t26\_1). Both transcripts were significantly upregulated at 90 and 120 DPI from 28 DPI, and MSTRG.4390.2 had the largest increase at near 7-fold increased at both late-chronic timepoints compared to 28 DPI. *SporoAMA1* is the sporozoite specific paralog of the *AMA1* protein and a major component of the moving junction during entry into the host cell [14, 15]. The N-terminal region of *sporoAMA1* includes a binding domain to the *T. gondii* protein RON2 [14, 22] to help form the moving junction complex. The annotated gene for *sporoAMA1* includes 9 exons (transcript TGME49\_315730-t26\_1); however, the bradyzoite reads do not align until exon 7 (Fig. 7B). In comparison, RNA-seq data from *T. gondii* infected cat intestines showed that the reads to *sporoAMA1* do not align until after exon 4 (Fig. 7). Microarray evidence of this same region from sporulated oocysts [17], show reads aligning to all nine exons of *sporoAMA1* (Additional File 9). This result suggests that *sporoAMA1* has stage specific isoforms, with the expression of the shortest isoform starting in late chronic infection and longer isoforms expressed during different stages of the cat cycle.

## Future Directions

The functions of the shortened *sporoAMA1* isoforms are unknown. For each of these transcripts, an in frame start codon is present 150-300 base pairs after the reads start aligning to the genome, highlighting that these shortened transcripts may be translated. It is unlikely that either of these two short forms plays a role in invasion as the RON2 binding region is not present in either isoform (Additional File 9, Fig. S1B). In support of a non-invasion function for *sporoAMA1*, there are almost no transcripts *sporoRON2* (TgME49\_265120) transcripts expressed either in bradyzoite cysts or sexual stages in the cat intestine. The shortest isoform would produce an 8 kDa protein during late chronic infection that contains the cytoplasmic region and the conserved serines/threonines. The sequential phosphorylation of these residues is essential for *Plasmodium falciparum* to efficiently invade red blood cells [23]. Thus the 8 kDa bradyzoite isoform may be phosphorylated and playing a regulatory role. However, no *sporoAMA1* peptides were found in any of the six chronic infection cyst samples generated in this study. While this result could indicate that the short 8 kDa isoform is unstable or degraded in the sample preparation process, it could also indicate that the approximately 400 bp RNA molecule is a long non-coding RNA. Perhaps this RNA functions similar to the Herpes encoded Latency Associated Transcripts (LATs), which are essential to maintain viral latency and reactivation [24]. The generation of an antibody to this shared

cytoplasmic domain will help determine if the short isoforms are translated as well as their potential function.

Further future studies are also needed to determine the host interactions with the bradyzoite cyst during long-term chronic infection. Upwards of 90% of the RNA-seq reads obtained in this study aligned to the *T. gondii* genome, whereas the 75% of the peptides aligned to the mouse. This difference is likely due to the instability of the host mRNA and the stability of the host proteins during sample processing. The enrichment of certain host proteins may indicate their stable interaction with the cysts. We have not yet examined if any identified host proteins are directly associated with the parasites or the cyst wall. The membrane associated host transporters TAP1 and TAP2 were identified in each sample by MS/MS. This result may be a consequence of their high expression in the brain during infection [10], or it may indicate a direct interaction with the cyst. *T. gondii* interaction with TAP1 and TAP2 could block antigen processing to help hide the cyst during chronic infection or it might assist in nutrient acquisition across the cyst wall. Our dataset had 31 host proteins (Additional File 2, right column) that were also identified in *in vitro* *T. gondii* cyst wall fractions [8], suggesting that these host proteins may be an association with the cyst wall. Examination of these host proteins potentially interacting with cysts will be an important area of future research as it may indicate how cysts can reside in their warm-blooded host for so long.

## Conclusions

The isolation of *in vivo* bradyzoites has allowed us to look deeply at both the transcript and protein content of the parasites. This depth has allowed us to identify novel bradyzoite specific isoforms, one of which is sporoAMA1. Over the course of chronic infection, we have found that the global transcriptional profile remains constant after the first month of chronic infection. This result is similar to the protein profile, in which most of the proteins identified are found in each timepoint. This dataset is the first global protein profile of *in vivo* bradyzoites. Together with the RNA-seq data, they provide a better understanding of the transcriptional and translational activity of *T. gondii* throughout infection.

## Methods

*Mouse infection with T. gondii.* The ME49 strain of *T. gondii* was cultured as tachyzoites in human foreskin fibroblast (HFF) cells. 6-10 week-old CBA/J (JAX) were infected intraperitoneally with  $1 \times 10^4$  tachyzoites to establish chronic toxoplasmosis.

*Collection of tissue bradyzoites.* At their respective timepoint, the mice were euthanized by carbon dioxide inhalation in a manner approved by the institutional animal ethics committee, and the whole brains of infected CBA/J mice at 21, 28, 90, 120, and 150 days post-infection were harvested and homogenized in phosphate buffered saline (PBS) using mortar and pestle. Several brain tissues were homogenized together as a single replicate to obtain enough parasite material. Homogenized tissue was serially passed through increasing gauge needles (18, 20, 22 respectively 5x each) and filtered using 150  $\mu$ m filter. The volume of the homogenate was increased to 40 mL using PBS and centrifuged for 10 min

at 2,000 x g. Pellet was resuspended in 3 mL of 20% Dextran (150,000 average molecular weight; Alfa Aesar) per brain and separated into tubes so that 3 or fewer brains were in each. Samples were centrifuged for 10 min at 400 x g to pellet the cysts. Host material and dextran were removed, and the cyst pellet was resuspended with 5 µg/mL pepsin in a 1% NaCl solution at pH 2.1 to release the bradyzoites from the cyst (1 mL per brain). Cysts were digested for 5 min at 37°C, then neutralized with a 1x volume of 1% Na<sub>2</sub>CO<sub>3</sub>. Bradyzoites were centrifuged for 10 min at 250 x g and resuspended in PBS then counted using hemocytometer.

*Protein LC-MS/MS.* Bradyzoites were preserved in 6 M GuHCl prior to processing. Protein was isolated by boiling at 100°C for 5 min before precipitating in 90% methanol. Protein was pelleted by centrifugation at 10,000 x g for 5 min, the supernatant was discarded, and the pellet was resuspended in lysis buffer (8 M urea, 100 mM tris pH 8, 40 mM 2-chloroacetamide, 10 mM 2-carboxyethyl phosphine). Samples were incubated for 10 minutes at RT before diluting to < 2 M urea with 50 mM tris pH 8. Trypsin (Promega, Madison, WI) was added at a mass ratio of 1:100::trypsin:protein and the sample was incubated overnight at RT with gentle rocking. A second bolus of trypsin (1:200::trypsin:protein) was added the next morning and incubated for an additional hour before desalting with reversed phase cartridges (StrataX, Phenomenex, Torrance, CA). Peptide yield was calculated by bicinchoninic acid (BCA) assay (Thermo Peirce, Rockford, IL). 2 µg of each sample was analyzed by LC-MS/MS on a Q-IT-OT Fusion Lumos mass spectrometer (Thermo Fisher, San Jose, CA) equipped with a 75/360 µm inner/outer diameter fused silica capillary column with a laser pulled electrospray tip packed with BEH C18 (130 Å pore, 1.7 µm particle size, 35 cm, Waters Corp, Milford, MA). Eluting peptides entered into the mass spectrometer following positive mode electrospray ionization. MS1 survey scans were performed in the orbitrap (240K resolution, AGC target 1e6, 100 ms maximum injection time). MS2 analysis of HCD generated (25% NCE) product ions were performed in the ion trap (Rapid resolution, AGC target 4e4, 18 ms maximum injection time). Monoisotopic precursor selection and dynamic exclusion (60 sec) were enabled. Raw data were searched against a concatenated database of mouse and *T. gondii* proteins protein sequences through the MaxQuant computational platform (version 1.6.0.13) operating the Andromeda search algorithm [25]. Precursor and fragment mass tolerances were set to 50 ppm and 0.5 Da, respectively. Carbamidomethylation of cysteine was imposed as a fixed modification and oxidation of methionine was allowed as a variable modification. Protein inference was performed from peptide spectrum matches (PSMs) based on the rules of parsimony [26]. Search results were filtered to a 1% false discovery rate at both the PSM and protein levels.

*Generation of RNA and RNA-seq.* Trizol was added to bradyzoites for the preservation of RNA, and the RNA extracted using phenol/chloroform separation and isopropanol precipitation. Samples include brain enriched bradyzoites at 28 DPI (n=2), 90 DPI (n=2), and 120 DPI (n=2). 10 ng of total RNA was processed using the SMART-Seq v4 Ultra Low Input RNA Kit (Takara), using 7 rounds of cDNA amplification. 1 ng of cDNA was indexed using the Nextera XT Index Kit (Illumina). RNA libraries were multiplexed and run across 2 lanes for paired end, 125 base pair sequencing by Illumina HiSeq 2500 at the University of Wisconsin Biotechnology Center.

Tachyzoite RNA was processed from the ME49 parasites grown in human foreskin fibroblasts (HFF) cells. Just before lysis of the parasite vacuole, infected HFF cells were scraped in trizol and RNA extracted as stated above. RNA libraries were prepared using a TruSeq RNA Library Prep kit v2 set A (Illumina) and sequenced using paired-end sequencing as stated above.

*Transcriptome assembly.* The sequencing reads were processed to remove low quality reads (Trimmomatic, v0.39 [27]). The tachyzoite, purified bradyzoite, and whole brain acute and chronic [6] sequencing reads were aligned to the *T. gondii* ME49 genome (ToxoDB.org) using the Spliced Transcripts Alignment to a Reference program (STAR, v2.7.0f, [28]). The following STAR parameters were changed from the default: mismatch max=2, intron min=70, and intron max=100,000. The STAR output bam file from each sample was fed into the StringTie program (v1.3.6, [11]), along with the current *T. gondii* annotation file (ToxoDB.org) and using a minimum transcript coverage value of 150. The sequencing reads were remapped using the new gtf file output from StringTie. Transcript abundance was estimated using RNA-seq by Expectation-Maximization (RSEM v1.2.31, [12]). Fold change was calculated using the geometric mean values between samples for transcripts with at least 10 RSEM gene counts in any one sample. The rlog transformed DESeq2 values were used for the PCA plot and PCA calculated using “plotPCA” function of the DESeq2 package. GO term enrichment was performed using ToxoDB. Terms were limited to biological processes with a p-value <0.05 and false discovery rate <0.05.

## List Of Abbreviations

DPI: days post-infection

RNA-seq: High-throughput RNA sequencing

LC-MS/MS: liquid chromatography-tandem mass spectrometry

PCA: principal component analysis

TPM: transcripts per million

PBS: phosphate buffered saline

HFF: human foreskin fibroblast

STAR: Spliced Transcripts Alignment to a Reference program

GO: Gene ontology

## Declarations

Ethics approval: Animals were treated in compliance with the guidelines set by the Institutional Animal Care and Use Committee (IACUC) of the University of Wisconsin School of Medicine and Public Health

(protocol #M005217), which adheres to the regulations and guidelines set by the National Research Council. The University of Wisconsin is accredited by the International Association for Assessment and Accreditation of Laboratory Animal Care.

Consent for publication: NA

Availability of data and material: All raw RNA sequencing data and differential expression values have been deposited in NCBI's Gene Expression Omnibus (GEO) [29] and are accessible through GEO Series accession number GSE134099. RNA sequencing and protein sequencing data have been supplied for public availability to ToxoDB.org.

Competing interests: We confirm that none of the authors have any competing interests in accordance with BioMed Central's guidelines.

Funding: This work was funded by the National Institutes of Health (P41 GM108538, JJC; 1R01AI144016-01, LJK; T32AI55397, ALG; T32GM008349, GMW). Funding bodies had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors' contributions: All authors contributed to the design of the experiments. ALG and LJK produced and analyzed the RNA-seq data. GMW and JJC performed the mass spectrometry data collection and interpretation of proteomic analyses. All authors read and approved the final manuscript.

Acknowledgements: We would like to thank Patrick Cervantes for providing RNA-seq data from tachyzoite culture.

## References

1. Robert-Gangneux F, Dardé M-L. Epidemiology of and Diagnostic Strategies for Toxoplasmosis. *Clin Microbiol Rev.* 2012;25:264–96.
2. Pittman KJ, Knoll LJ. Long-Term Relationships: the Complicated Interplay between the Host and the Developmental Stages of *Toxoplasma gondii* during Acute and Chronic Infections. *Microbiol Mol Biol Rev.* 2015;79:387–401.
3. Boothroyd JC, Black M, Bonnefoy S, Hehl A, Knoll LJ, Manger ID, et al. Genetic and biochemical analysis of development in *Toxoplasma gondii*. *Philos Trans R Soc Lond B Biol Sci.* 1997;352:1347–54.
4. Luft BJ, Remington JS. Toxoplasmic Encephalitis in AIDS. *Clin Infect Dis.* 1992;15:211–22.
5. Buchholz KR, Fritz HM, Chen X, Durbin-Johnson B, Rocke DM, Ferguson DJ, et al. Identification of Tissue Cyst Wall Components by Transcriptome Analysis of In Vivo and In Vitro *Toxoplasma gondii* Bradyzoites. *Eukaryot Cell.* 2011;10:1637–47.
6. Pittman KJ, Aliota MT, Knoll LJ. Dual transcriptional profiling of mice and *Toxoplasma gondii* during acute and chronic infection. *BMC Genomics.* 2014;15:806.

7. Watts E, Zhao Y, Dhara A, Eller B, Patwardhan A, Sinai AP. Novel Approaches Reveal that *Toxoplasma gondii* Bradyzoites within Tissue Cysts Are Dynamic and Replicating Entities In Vivo. *mBio*. 2015;6. doi:10.1128/mBio.01155-15.
8. Tu V, Mayoral J, Sugi T, Tomita T, Han B, Ma YF, et al. Enrichment and Proteomic Characterization of the Cyst Wall from In Vitro *Toxoplasma gondii* Cysts. *mBio*. 2019;10:e00469-19.
9. Freyre A. Separation of *Toxoplasma* Cysts from Brain Tissue and Liberation of Viable Bradyzoites. *The Journal of Parasitology*. 1995;81:1008–10.
10. Garfoot AL, Cervantes PW, Knoll LJ. Transcriptional Analysis Shows a Robust Host Response to *Toxoplasma gondii* during Early and Late Chronic Infection in Both Male and Female Mice. *Infection and Immunity*. 2019;87:e00024-19.
11. Perteau M, Perteau GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*. 2015;33:290–5.
12. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12:323.
13. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 2014;15:550.
14. Poukchanski A, Fritz HM, Tonkin ML, Treeck M, Boulanger MJ, Boothroyd JC. *Toxoplasma gondii* Sporozoites Invade Host Cells Using Two Novel Paralogues of RON2 and AMA1. *PLOS ONE*. 2013;8:e70637.
15. Mital J, Meissner M, Soldati D, Ward GE. Conditional Expression of *Toxoplasma gondii* Apical Membrane Antigen-1 (TgAMA1) Demonstrates That TgAMA1 Plays a Critical Role in Host Cell Invasion. *MBoC*. 2005;16:4341–9.
16. Ramakrishnan C, Maier S, Walker RA, Rehrauer H, Joekel DE, Winiger RR, et al. An experimental genetically attenuated live vaccine to prevent transmission of *Toxoplasma gondii* by cats. *Sci Rep*. 2019;9. doi:10.1038/s41598-018-37671-8.
17. Fritz HM, Buchholz KR, Chen X, Durbin-Johnson B, Rocke DM, Conrad PA, et al. Transcriptomic Analysis of *Toxoplasma* Development Reveals Many Novel Functions and Structures Specific to Sporozoites and Oocysts. *PLoS One*. 2012;7. doi:10.1371/journal.pone.0029998.
18. Torpier G, Charif H, Darcy F, Liu JL, Darde ML, Capron A. *Toxoplasma gondii*: Differential Location of Antigens Secreted from Encysted Bradyzoites. *Experimental Parasitology*. 1993;77:13–22.
19. Tomita T, Bzik DJ, Ma YF, Fox BA, Markillie LM, Taylor RC, et al. The *Toxoplasma gondii* Cyst Wall Protein CST1 Is Critical for Cyst Wall Integrity and Promotes Bradyzoite Persistence. *PLOS Pathogens*. 2013;9:e1003823.
20. Bohne W, Gross U, Ferguson DJP, Heesemann J. Cloning and characterization of a bradyzoite-specifically expressed gene (hsp30/bag1) of *Toxoplasma gondii*, related to genes encoding small heat-shock proteins of plants. *Molecular Microbiology*. 1995;16:1221–30.
21. Yang S, Parmley SF. *Toxoplasma gondii* expresses two distinct lactate dehydrogenase homologous genes during its life cycle in intermediate hosts. *Gene*. 1997;184:1–12.

22. Crawford J, Tonkin ML, Grujic O, Boulanger MJ. Structural Characterization of Apical Membrane Antigen 1 (AMA1) from *Toxoplasma gondii*. *J Biol Chem*. 2010;285:15644–52.
23. Prinz B, Harvey KL, Wilcke L, Ruch U, Engelberg K, Biller L, et al. Hierarchical phosphorylation of apical membrane antigen 1 is required for efficient red blood cell invasion by malaria parasites. *Scientific Reports*. 2016;6:34479.
24. Nicoll MP, Proença JT, Efstathiou S. The molecular basis of herpes simplex virus latency. *FEMS Microbiol Rev*. 2012;36:684–705.
25. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Molecular & Cellular Proteomics*. 2014;13:2513–26.
26. Nesvizhskii AI, Aebersold R. Interpretation of Shotgun Proteomic Data: The Protein Inference Problem. *Molecular & Cellular Proteomics*. 2005;4:1419–40.
27. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
28. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21.
29. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002;30:207–10.

## Table

**Table 1. RNA-seq reads per sample.** The number of total reads sequenced from each sample and the number of which mapped to the *T. gondii* genome.

| Sample Type | DPI | Replicate ID | Input Reads | Toxo Reads | % Toxo | Reference  |
|-------------|-----|--------------|-------------|------------|--------|------------|
| Tachyzoites | 0   | 1            | 25837455    | 11650902   | 45.09  | This Study |
|             |     | 2            | 18310484    | 15269361   | 83.39  | This Study |
| Whole Brain | 10  | 1            | 102581171   | 116278     | 0.11   | [6]        |
|             |     | 2            | 125546828   | 97611      | 0.08   | [6]        |
|             |     | 3            | 81630704    | 51309      | 0.06   | [6]        |
| Whole Brain | 28  | 1            | 103765423   | 167181     | 0.16   | [6]        |
|             |     | 2            | 113094439   | 137975     | 0.12   | [6]        |
|             |     | 3            | 114177191   | 162423     | 0.14   | [6]        |
| Bradyzoites | 28  | A            | 40656512    | 21075587   | 51.84  | This Study |
|             |     | C            | 39129732    | 36606828   | 93.55  | This Study |
| Bradyzoites | 90  | A            | 37700412    | 35412305   | 93.93  | This Study |
|             |     | B            | 41174415    | 25965460   | 63.06  | This Study |
| Bradyzoites | 120 | B            | 46857160    | 40734629   | 86.93  | This Study |
|             |     | C            | 45074363    | 34443785   | 76.42  | This Study |

## Additional Files

Additional File 1. csv Table S1. Unique peptides identified from bradyzoites. Data shows either *T. gondii* or mouse peptides identified in each sample with their spectral count value. The pooled sample represents peptides identified when protein from all samples is mixed before processing for LC-MS/MS.

Additional File 2. csv Table S2. Unique proteins to which the identified peptides map to. Data shows proteins that were mapped to either *T. gondii* or mouse in each sample with their total spectral count value.

Additional File 3. csv Table S3. Proteins common between all bradyzoite samples. Data shows each protein with the protein spectral count for the six samples chosen for analysis.

Additional File 4. csv Table S4. Normalized sequencing read values for the *T. gondii* isoforms. TPM values calculated by RSEM for individual isoforms after aligning reads to the transcripts identified by StringTie. TZ, tissue culture tachyzoites; WholeBrain, data from whole tissue sequencing from either acute or chronically infected brain tissue; BZ, enriched in vivo bradyzoites.

Additional File 5. csv Table S5. Normalized sequencing read values for the sum of all *T. gondii* gene isoforms. TPM values calculated by RSEM for each gene after aligning reads to the transcripts identified by StringTie. TZ, tissue culture tachyzoites; WholeBrain, data from whole tissue sequencing from either acute or chronically infected brain tissue; BZ, enriched in vivo bradyzoites.

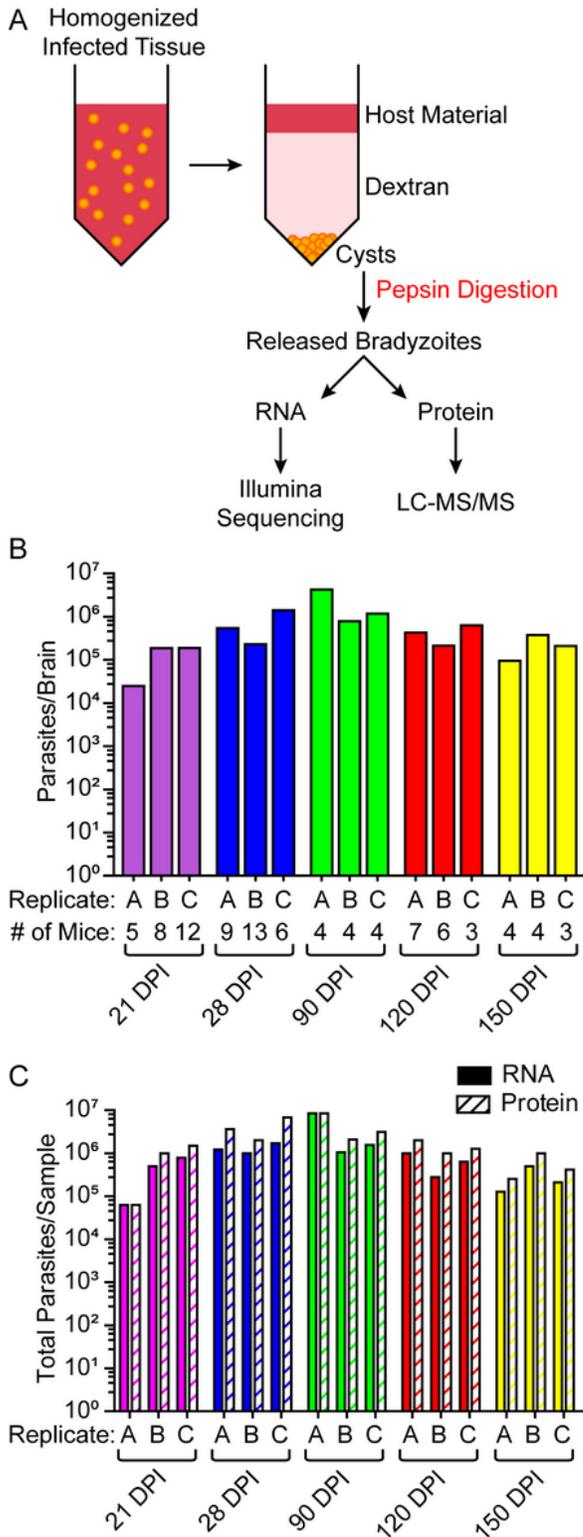
Additional File 6. csv Table S7. Enriched GO terms. GO terms significantly enriched among the highest expressed genes, chronically differentially expressed genes, and the groups of differently expressed genes among the proteins identified.

Additional File 7. csv Table S6. Log<sub>2</sub> values of the fold change for differentially expressed isoforms. Values were calculated with DESeq2. Values are only shown if there was >2-fold change between samples with q-value <0.05. Values below the threshold are represented by "-".

Additional File 8. csv Table S7. DE genes in chronic infection. Log<sub>2</sub> values of the fold change for differentially expressed genes. Values were calculated with DESeq2. Values are only shown if there was >2-fold change between samples with q-value <0.05. Values below the threshold are represented by "-".

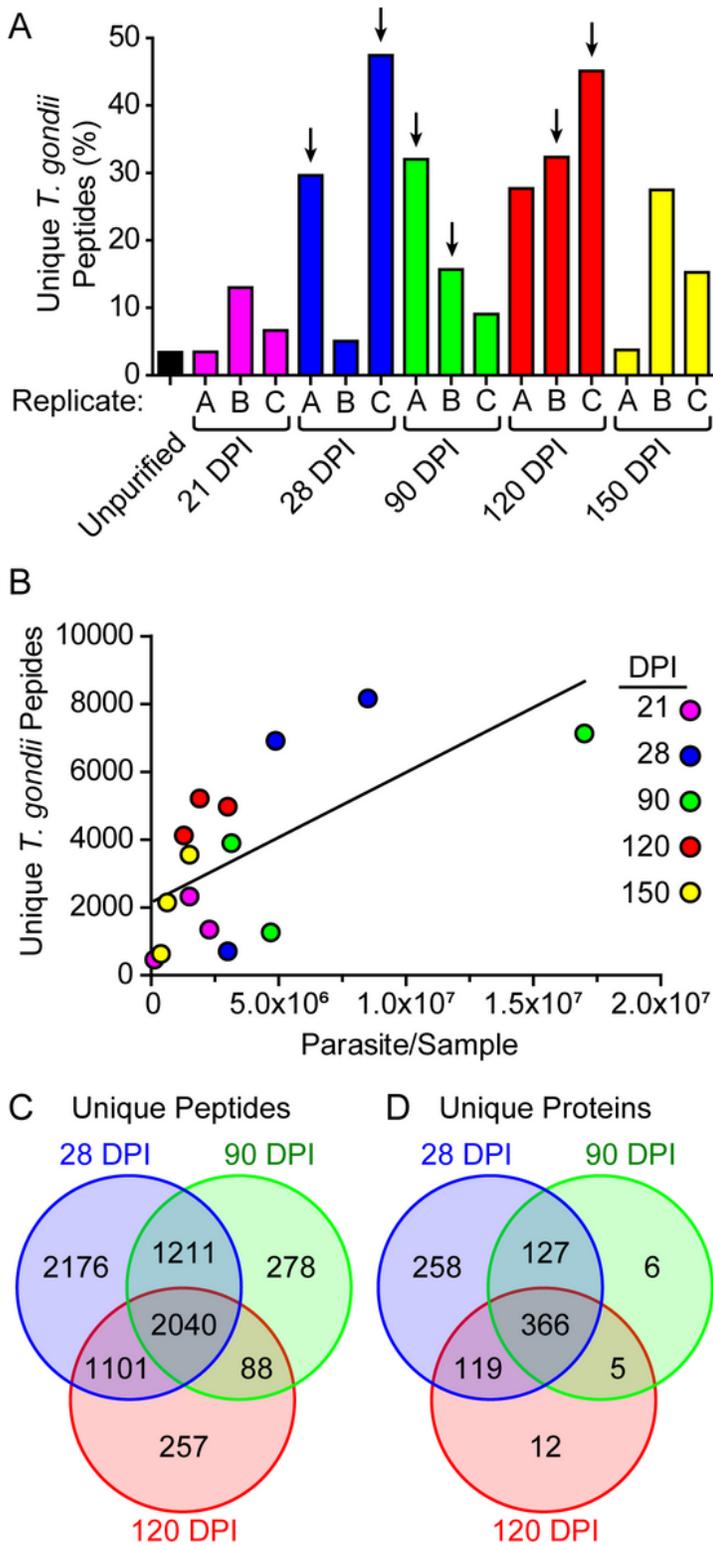
Additional File 9. tif Figure S1. Different isoforms of sporAMA1 are expressed during chronic infection, during the cat intestinal stage, and during oocyst sporulation. (A) All exons of sporAMA1 are expressed during oocyst sporulation. Sequencing coverage from a representative sample from each group viewed from the ToxoDB genome browser: purified bradyzoites at 120 DPI (red), cat stage (orange), and oocyst microarray data at 0, 4, and 10 days post-sporulation (Brown). The X-axis represents the genomic region for sporAMA1. The Y-axis represents the total read count at each nucleotide position. Predicted exons for the gene are represented by black rectangles under each panel. (B) Protein domains for the sporAMA1 isoforms. DI, Domain I highlighted in red; DII, Domain II in green; DIII, Domain III in yellow; TM, Transmembrane region in blue; CPD, cytoplasmic domain in orange. Red arrows represent phosphorylation sites and black arrow represents the RON2 binding site.

## Figures



**Figure 1**

*T. gondii* bradyzoites are enriched by dextran. (A) Schematic of the bradyzoite enrichment workflow to enrich for parasites for RNA and protein sequencing. (B) Number of parasites per brain of each replicate labeled A, B and C. Samples consisted of bradyzoites collected from a pool of homogenized brains and the number (#) of mice in each pool shown. (C) Number parasites used for either RNA (solid colored bars) or protein (striped bars) analysis.



**Figure 2**

Parasite number affects *T. gondii* peptide abundance. (A) Percentage of unique *T. gondii* peptides from the total number of unique peptides identified in each sample. Arrows represent samples used for RNAseq analyses. (B) Correlation of the number of parasites present in each sample compared against number of unique *T. gondii* peptides in that sample. Line represents the linear regression ( $r^2=0.42$ ).

Number of unique peptides (C) and proteins (D) identified in two samples at each timepoint (arrows from panel A). DPI=days post-infection.

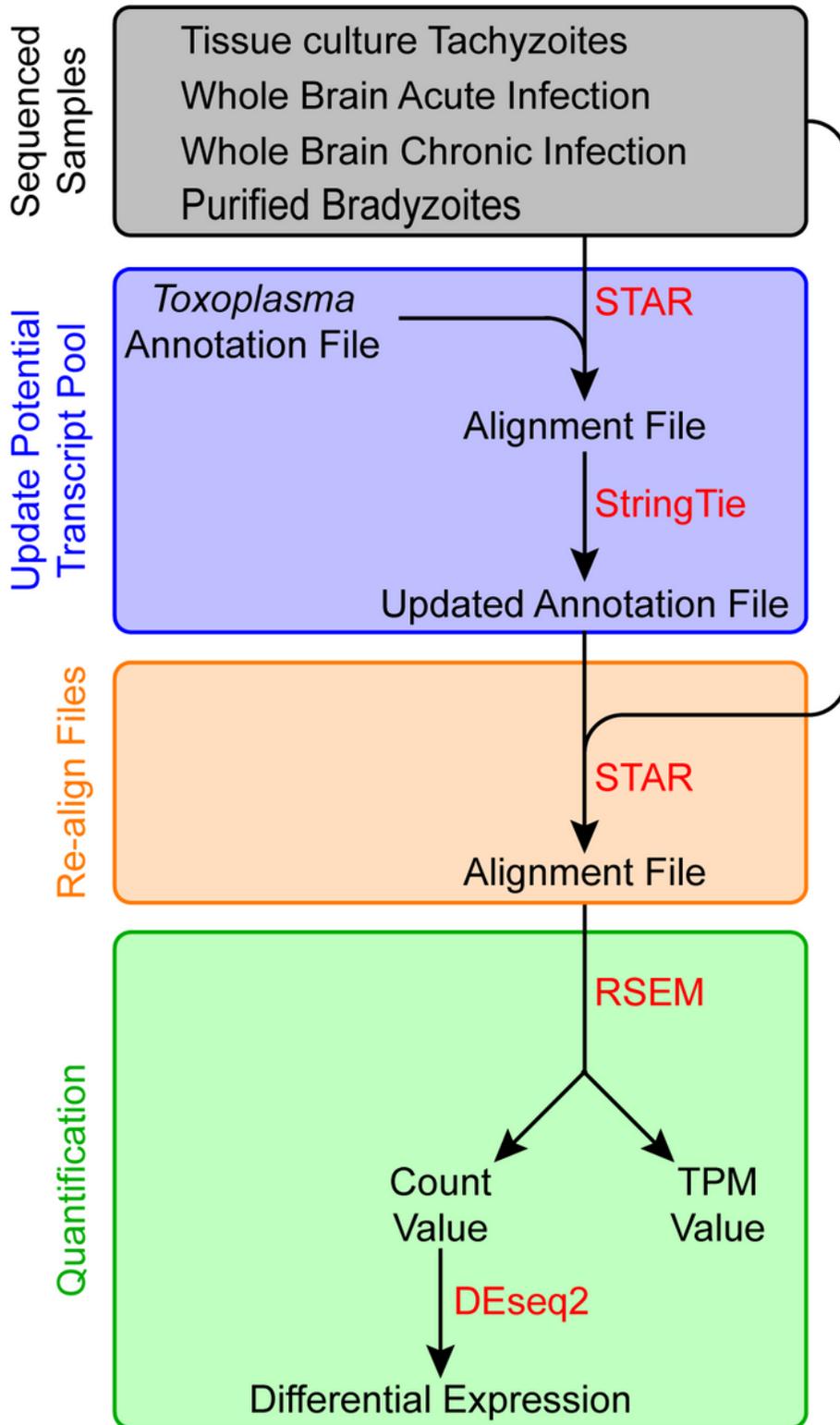
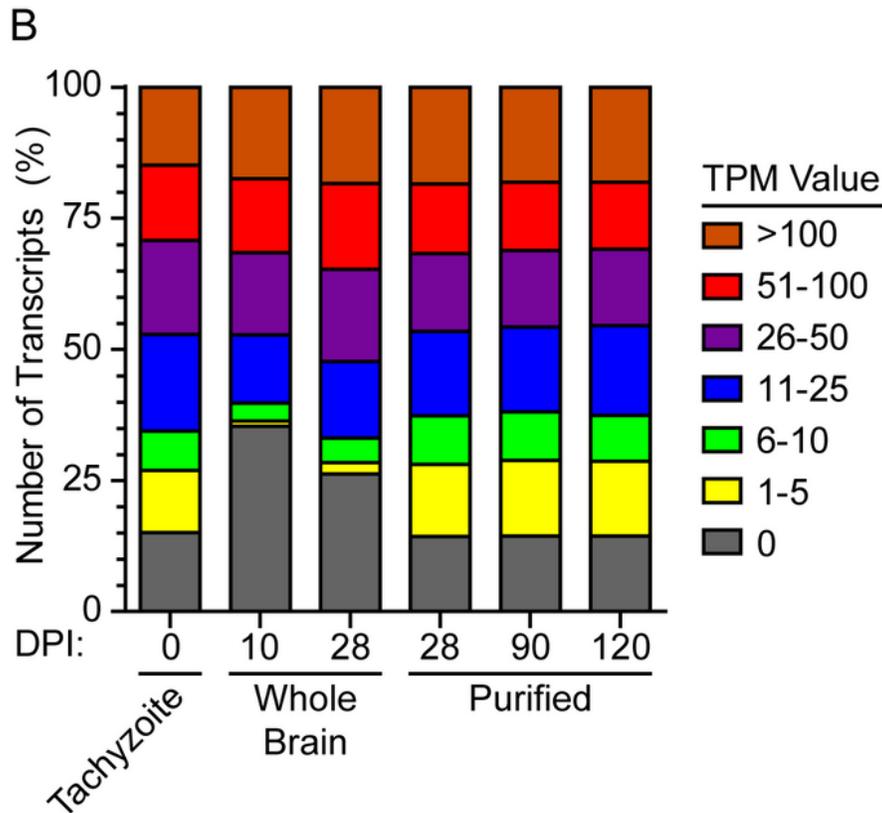
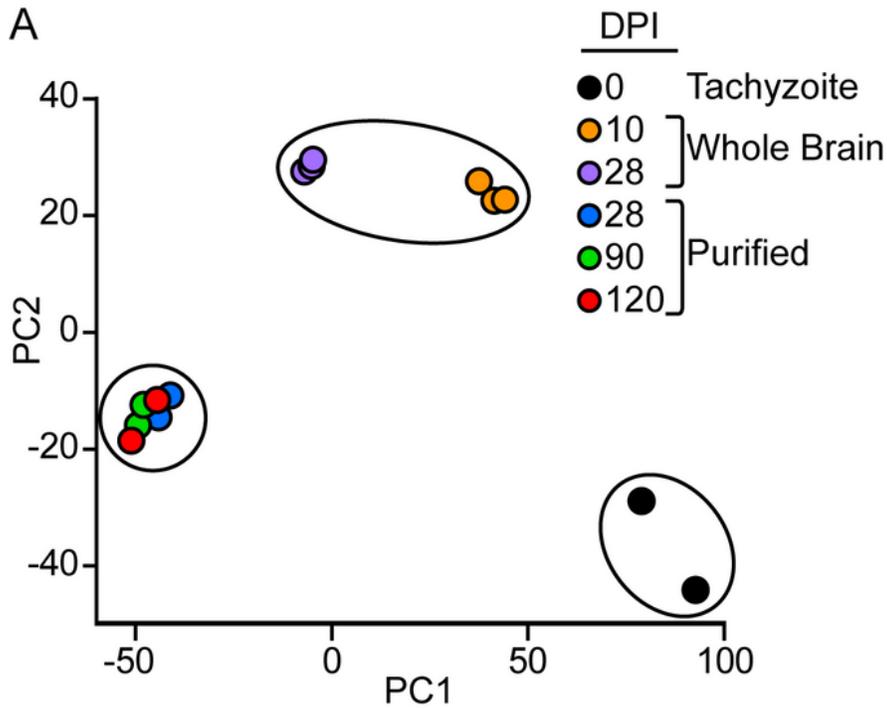


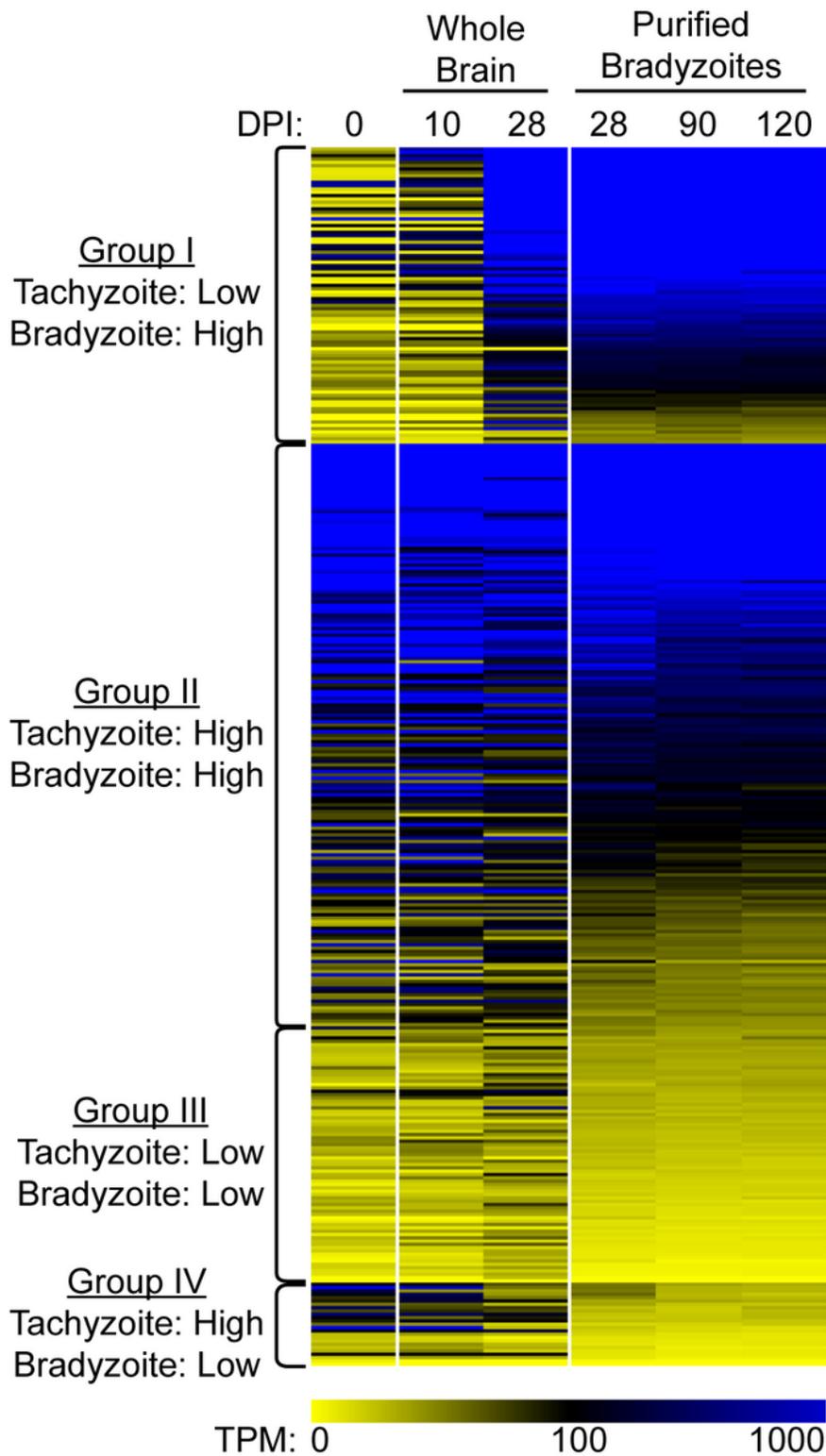
Figure 3

Schematic representing the RNAseq workflow. Samples include tissue culture tachyzoites (n=2), whole brain tissue from 10 DPI acute (n=3) and 28 DPI chronically (n=3) infected mice, and purified bradyzoites from 28 (n=2), 90 (n=2), and 120 (n=2) DPI.



**Figure 4**

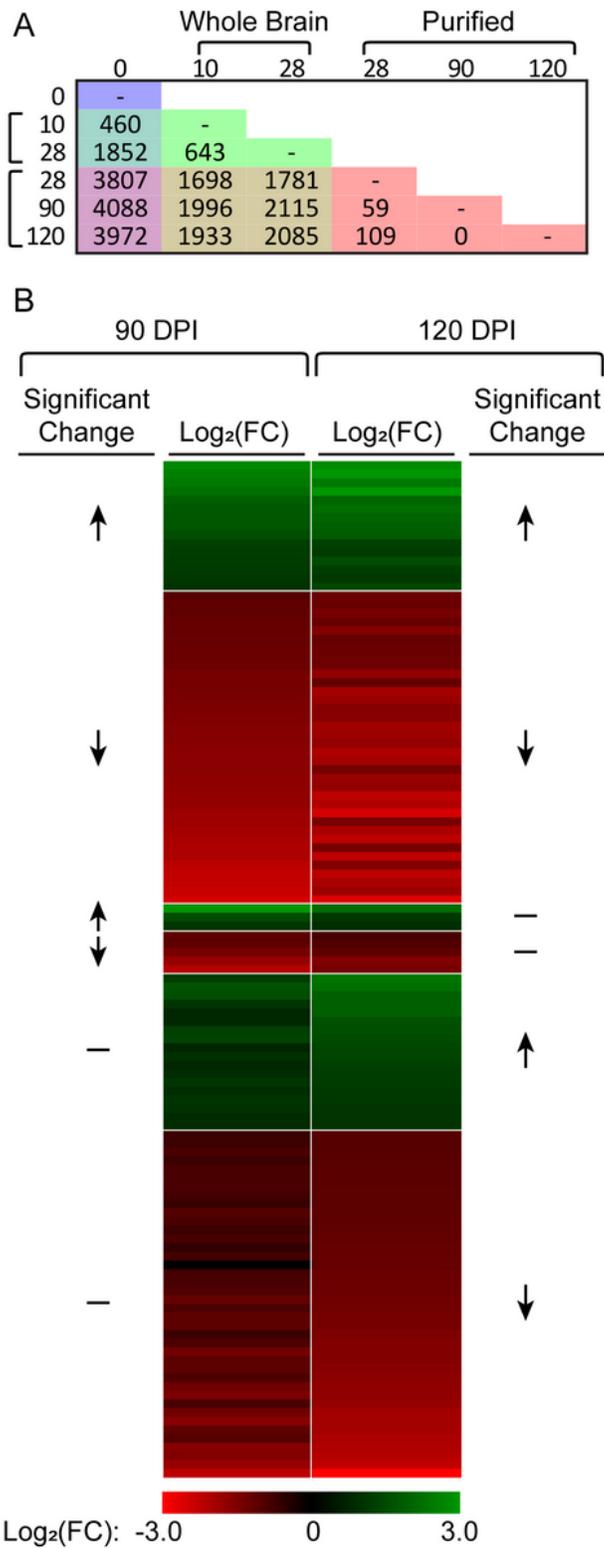
RNAseq samples group with experimental design and with infection state. (A) PCA plot from normalized values calculated by DESeq2. Symbol colors represent: tissue culture tachyzoites (black); whole brain acute infection (orange); whole brain chronic infection (purple); purified at 28 days (blue); purified at 90 days (green); purified at 120 days (red). (B) Percent of transcripts with TPM values within each range: 0 (grey), 1-5 (yellow), 6-10 (green), 11-25 (blue), 26-50 (purple), 51-100 (red), and >100 (brown).



**Figure 5**

Most in vivo bradyzoite proteins have high gene expression. TPM values of the genes expressing the proteins identified in all six analyzed samples. Each row represents the gene for the protein and columns represent the sample type: tissue culture tachyzoites (0); whole brain acute infection (10); whole brain chronic infection (28); purified at 28 days (28); purified at 90 days (90); purified at 120 days (120). The TPM value is the average between the replicates of each group (n=2-3). High expression is defined as an

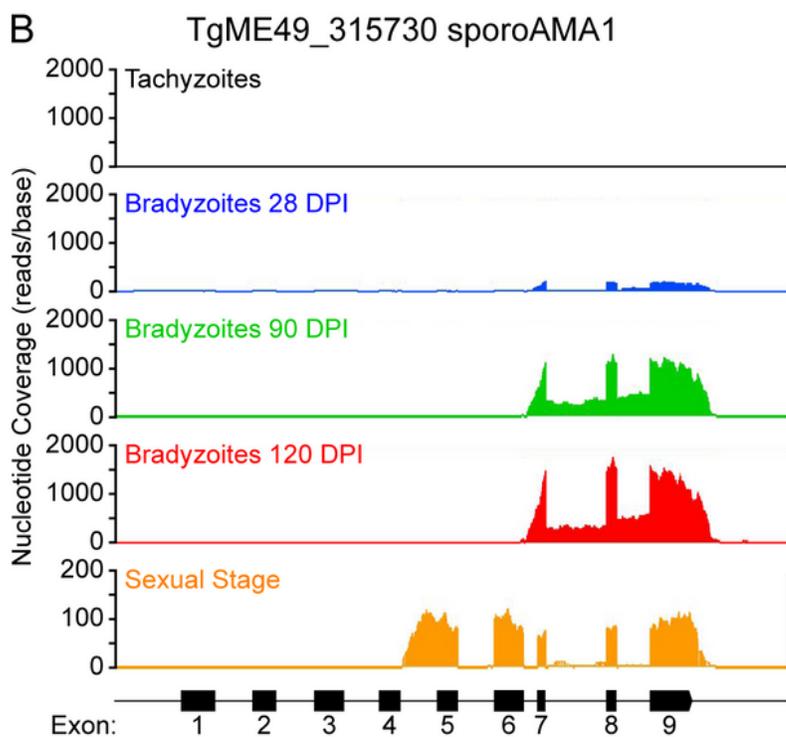
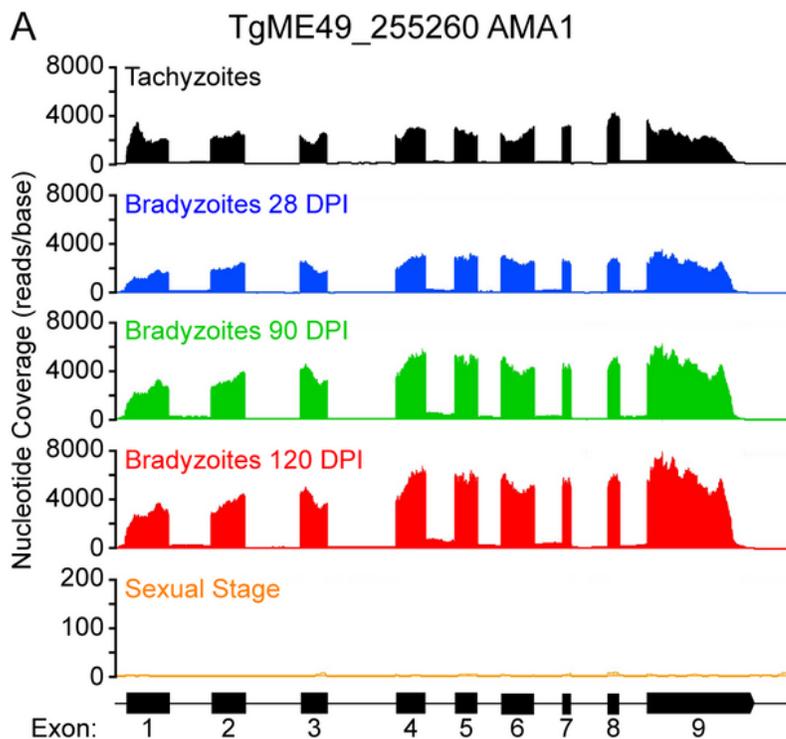
average TPM >100 between samples under tachyzoite (0 and 10 DPI) or bradyzoite (28-120 DPI) conditions.



**Figure 6**

Differentially expressed transcripts during late-chronic infection. (A) Number of differentially expressed transcripts (>2-fold difference and q-value <0.05) among all conditions. (B) Differentially expressed transcripts during late-chronic infection. Log<sub>2</sub> fold change values at 90 days post-infection (DPI) and 120

DPI relative to 28 DPI. Upward arrows indicate transcripts are significantly increasing at that timepoint, downward arrows indicate transcripts are significantly decreasing at that timepoint, and a dash indicates no significant change at that timepoint.



**Figure 7**

spoAMA1 has stage specific transcripts. Sequencing coverage from a representative sample from each group: Tachyzoite (black), purified bradyzoites at 28 DPI (blue), purified bradyzoites at 90 DPI (green),

purified bradyzoites at 120 DPI (red). Sequencing coverage for *T. gondii* during infection in cat intestines is shown in orange. X-axis represents the genomic region for AMA1 (A) and sporoAMA1 (B). Y-axis represents the total read count at each nucleotide position. Predicted exons for the gene are represented by black rectangles under each panel.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementalTable2ProteinsAll.csv](#)
- [SupplementalTable3.csv](#)
- [SupplementalTable6GOTerms.csv](#)
- [SupplementalFigure1sporoAMA1Oocysts.tif](#)
- [SupplementalTable4.csv](#)
- [SupplementalTable8DEGenes.csv](#)
- [SupplementalTable7DETranscripts.csv](#)
- [SupplementalTable1PeptidesAll.csv](#)
- [SupplementalTable5.csv](#)