A single-parasite transcriptional atlas of *Toxoplasma gondii* reveals novel control of antigen expression

Yuan Xue¹, Terence C. Theisen², Suchita Rastogi², Abel Ferrel², Stephen R. Quake¹,³,⁴,* John C. Boothroyd²,*

¹Department of Bioengineering, ³Department of Applied Physics, Stanford University, Stanford, CA, USA
²Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA
⁴Chan Zuckerberg Biohub, San Francisco, CA, USA

*Correspondence should be addressed to J.C.B. (jboothr@stanford.edu) and S.R.Q. (steve@quake-lab.org)

Abstract

*Toxoplasma gondii*, a protozoan parasite, undergoes a complex and poorly understood developmental process that is critical for establishing a chronic infection in its intermediate hosts. Here, we applied single-cell RNA-seq (scRNA-seq) on >5,400 *Toxoplasma* in both tachyzoite and bradyzoite stages using three widely studied strains to construct a comprehensive atlas of cell-cycle and asexual development, revealing hidden states and transcriptional factors associated with each developmental stage. Analysis of SAG1-related sequence (SRS) antigenic repertoire reveals a highly heterogeneous, sporadic expression pattern unexplained by measurement noise, cell cycle, or asexual development. Furthermore, we identified AP2IX-1 as a transcription factor that controls the switching from the ubiquitous SAG1 to rare surface antigens not previously observed in tachyzoites. In addition, comparative analysis between *Toxoplasma* and *Plasmodium* scRNA-seq results reveals concerted expression of gene sets, despite fundamental differences in cell division. Lastly, we built an interactive data-browser for visualization of our atlas resource.

Keywords: single-cell RNA sequencing, protozoa, cell cycle, *Toxoplasma*, development, AP2 transcription factor
31 Main text

32 Introduction

Toxoplasma gondii is an intracellular protozoan parasite that is thought to infect over a quarter of the world’s population\(^1\). Like some of its Apicomplexan cousins, Toxoplasma undergoes a complex developmental transition inside the host. In intermediate hosts, including humans and virtually all other non-feline, warm-blooded animals, Toxoplasma parasites remain haploid and transition from a replicative, virulent tachyzoite to an encysted, quasi-dormant bradyzoite. This asexual developmental transition is tightly coupled to the clinical progression of Toxoplasma infection. Although acute infection with tachyzoites produces few if any symptoms in healthy human children and adults, infected individuals, if left untreated, progress to a chronic stage wherein tachyzoites transition to bradyzoites that can persist for life in neurons and muscle cells. When infected individuals become immunocompromised, such as in chemotherapy, HIV infection, or organ transplantation\(^2,3\), bradyzoites can reactivate to become tachyzoites, causing severe neurological damage and even death. While no causal link has been established, a population-wide study has uncovered significant association of Toxoplasma infection with schizophrenia in chronically infected humans\(^4\). Chronic infection in mice has been observed to induce behavioral changes such as loss of aversion to cat urine, which is hypothesized to increase the transmission rate of Toxoplasma to its definitive feline host where sexual reproduction occurs\(^5\). As there are no therapeutic interventions to prevent or clear cysts in infected individuals, understanding how Toxoplasma transitions through its life stages remains of critical importance.

The development of in vitro methods to induce Toxoplasma differentiation have facilitated investigation of several aspects of chronic infection, including transition of tachyzoites to bradyzoites\(^6,7\). Bulk transcriptomic analyses of Toxoplasma gondii at distinct asexual stages reveal genetic modules that are expressed in each stage\(^8-15\), including AP2 transcription factors that are thought to play a role in differentiation\(^16,17\); however, transitioning parasites convert to the bradyzoite stage asynchronously and display a high degree of heterogeneity along the developmental pathway and in gene expression\(^18,19\). Furthermore, parasites within the same tissue cysts have been shown to display heterogeneity in the expression of bradyzoite marker proteins\(^20\). The transition of tachyzoites to the bradyzoite stage results in an overwhelming majority of mature bradyzoites in the G\(_1\) phase of the cell cycle that divide slowly, if at all\(^21,22\). Furthermore, tachyzoites exhibit slower growth kinetics immediately prior to the bradyzoite transition\(^21,23\). This suggests that parasites exit the cell cycle to differentiate into bradyzoites, a pattern consistent with developmental processes in several other eukaryotic organisms\(^24,25\). Dissecting these cell cycle aspects of stage conversion
requires a more detailed analysis than has been possible with bulk measurement of
tachyzoite or bradyzoite populations, or with the use of genetically modified parasites
coupled with chemical synchronization of cell cycle progression. This is because
the latter approaches require large quantities of synchronized parasites and can
potentially introduce artificial perturbations. Furthermore, bulk measurement fails to
distinguish parasite-to-parasite variation that is independent of cell cycle or known
developmental processes, potentially missing the phenotypic diversity intrinsic to a
population of cells.

Single-cell RNA sequencing (scRNA-seq) offers a powerful and unbiased
approach to reveal the underlying heterogeneity in an asynchronous population of cells.
Droplet and FACS-based approaches have already been applied towards multicellular
parasites such as Schistosoma to reveal developmental changes within different
hosts. Recently, scRNA-seq has revealed a surprising degree of heterogeneity in
another apicomplexan parasite, Plasmodium. Analyses derived from these single-
parasite measurements uncovered rare and critical transition events in parasite
development that were undetectable in bulk measurements. Combined with novel
analytical tools and increase in measurement throughput, scRNA-seq can help facilitate
the discovery of regulatory factors that mediate these developmental transitions in a
system-wide fashion.

Here, we performed scRNA-seq to reconstruct transcriptional dynamics of
asynchronous Toxoplasma parasites in the course of cell cycle and asexual
development in vitro. Our analysis reveals the existence of hidden cell states and rare
parasites that show highly unusual patterns of gene expression associated with specific
transcription factors. We also discover that individual parasites vary substantially in the
expression of surface antigen genes, suggesting the possibility of a novel form of
antigenic variation that may play a crucial role in host immunity evasion. Importantly, we
identified a single parasite from our scRNA-seq dataset that displayed an unusual
expression pattern of surface antigens, leading us to identify and validate the regulatory
role of a previously uncharacterized AP2 transcription factor typically associated with
parasites in sexual development. Lastly, we show that despite fundamental differences
in their modes of cell division, there are conserved transcriptional programs between the
asexual life cycles of Toxoplasma gondii and Plasmodium berghei. Our results
combined provide the first comprehensive single-cell atlas of Toxoplasma in the course
of asexual development and help reveal that the antigenic repertoire of this parasite is
much more heterogeneous than previously appreciated.
Results

Technical validation of single-parasite sorting and sequencing

There are more than a dozen approaches available for single-cell isolation and transcriptome amplification. Based on benchmark comparisons, Smart-seq2 generally has higher sensitivity than competing droplet-based approaches\textsuperscript{34,35}. We reasoned that sensitive measurement is crucial in our study, given that single \textit{Toxoplasma gondii} parasites are at least 50-fold smaller in volume than a typical mammalian cell, and thus the average parasite gene is likely expressed with much lower copy number per cell than a typical mammalian gene. For our initial studies, we used the common Type I lab strain of \textit{Toxoplasma}, RH, grown \textit{in vitro} in human foreskin fibroblasts (HFFs). Following such growth, individual tachyzoites were released by passage through a narrow-gauge needle and then purified by fluorescence activated cell sorting (FACS) into 384-well or 11696-well plates. We then synthesized, amplified, and barcoded cDNA using Smart-seq2 and Illumina Nextera protocols. We reduced the reagent cost in 384-well plates effectively by ten-fold compared to the 96-well format. The sequenced reads were bioinformatically deconvolved and grouped into individual parasites for analysis using modified bcl2fastq and custom python scripts (Materials and methods). A schematic to illustrate our experimental workflow is shown in Figure 1.

To ensure that our workflow efficiently captures single \textit{Toxoplasma} parasites, we mixed equal numbers of two transgenic lines of RH, one expressing GFP and the other expressing mCherry, and sorted individual parasites into a 384-well plate based on the presence of either green or red fluorescence. After Smart-seq2 amplification, we quantified the expression of GFP and mCherry mRNAs using quantitative polymerase chain reaction (qPCR). Across all 301 wells that we measured, we observed the presence of both GFP and mCherry mRNA in only one well, indicating that the rate of doublet events is below 1% (Figure 1 - Supplementary Figure 1a). To address the possibility that the reduced reagent volume in the 384-well format could potentially saturate the reaction chemistry and thus limit quantification range, we sorted varying numbers of RH and quantified with qPCR the mRNA of a gene encoding the abundantly expressed surface protein, SRS29B (SAG1) (Figure 1 - Supplementary Figure 1b).

The detected amount of SAG1 mRNA present in wells containing single, eight, or fifty RH fell into the expected distributions based on the number of parasites sorted, without signs of saturation, indicating that the assay is capable of quantitative measurement at the single \textit{Toxoplasma} level. We then proceeded to sort parasites into 384-well based on live/dead staining and sequence 729 RH (612 passed quality control) strain single \textit{Toxoplasma} parasites from asynchronous populations grown under tachyzoite conditions. We also sorted and sequenced 287 RH parasites (237 passed quality control) into 96-well plate for comparison, which we will discuss further in another
section. For Pru and ME49 strains, we collected parasites at several time points post alkaline treatment which induces differentiation from tachyzoites to bradyzoites to follow changes in their expression profiles during in vitro development (Materials and methods), yielding 2655 Pru (2198 passed quality control) and 1828 ME49 (1552 passed quality control) single parasites. RH reads were aligned to the GT1 strain genome, which is the most complete reference for Type I parasites, while Pru and ME49 were aligned to the ME49 strain Type II genome reference. Because many genes encoding Toxoplasma secretion factors and surface proteins are evolutionary products, we expected high sequence similarity amongst a substantial portion of the parasite genes. Thus, we modified our gene counting pipeline to account for duplicated genes by distributing reads across all regions with equal alignment score that passed thresholds (Materials and methods). The reason why we adopted a correction scheme for multiply-mapped reads is because the analysis of co-occurrence, or lack thereof, of pathogenic factors (e.g. surface antigens) hinges on sensitive detection of their expression. We thus faced a choice of increasing the false positive in gene alignment by correcting for multiply mapped reads or increasing its false negative by counting only uniquely aligned reads. We chose to account for multiply mapped genes, which would otherwise be obscured, as we hypothesized that surface antigen expression may vary between individual parasites, which we further discuss in another section. A comparison of counting methods does not reveal significant differences in the observed counts (Figure 1 - Supplementary Figure 1c). Further analysis reveals that our modified pipeline recovered the detection of more parasite genes than default parameters (Figure 1 - Supplementary Figure 1d).

To ensure that poorly amplified or sequenced parasites did not confound our downstream analysis, we filtered samples based on several quality metrics including percent reads mapping to ERCC spike-in sequences, number of genes detected, and sequencing depth (Materials and methods; Figure 1 - Supplementary Figure 2a). On average, each sequenced parasite contains 30-50% reads that mapped to Toxoplasma genes encoding proteins (top panel in Figure 1 - Supplementary Figure 2b). Most of the unmapped reads are from Toxoplasma's 28s ribosomal RNA. The relatively high rate of rRNA contamination was also observed in single-parasite RNA sequencing of Plasmodium. We suspect this occurred due to non-specific priming as protozoan cells have low RNA input. We normalized for sequencing depth across cells by dividing each read count by the read sum of each corresponding cell and then multiplied by the median of read sum within each dataset to yield “count per median” (CPM). After filtering ERCC spike-in and rRNA genes, we detected on average 862, 1290, and 970 genes per parasite with greater than or equal to 2 read counts (Materials and methods) in the RH, Pru, and ME49 datasets, respectively (bottom panel in Figure 1 - Supplementary Figure 2b). Characterization of our measurement sensitivity based on logistic regression modeling of ERCC spike-in standards (Materials and methods) reveals a 50% detection rate of 17, 17, and 26 molecules for RH, Pru, and ME49.
datasets, respectively (top panels in Figure 1 - Supplementary Figure 2c). The sensitivity of our 384-well Smart-seq2 measurement is comparable to the previously reported range for the 96-well format\textsuperscript{38}. As expected from our qPCR titration experiment, scRNA-seq measurement of gene expression is quantitative at single parasite resolution based on ERCC standards. We determined that the linear dynamic range of our scRNA-seq measurement spans over three orders of magnitude (bottom panels in Figure 1 - Supplementary Figure 2c). Taken together, we demonstrate a scalable and cost-effective approach to measure the transcriptomic changes of individual parasites with high sensitivity and accuracy.

\textbf{Cell cycle landscape of asynchronous \textit{Toxoplasma}}

Previous work posited a potential link between bradyzoite development and cell cycle, which poses a significant challenge to the bioinformatic analysis of either process\textsuperscript{21}. To characterize cell cycling changes without confounding contributions from developmental processes, we first analyzed an asynchronous population of Type I RH strain parasites grown under tachyzoite conditions; this extensively passaged lab strain is known to have little propensity to switch to bradyzoites under such conditions\textsuperscript{6} (Materials and methods). After filtering out genes whose expression levels did not vary significantly between individual parasites, we projected the data with principal component analysis (PCA) (Materials and methods). Interestingly, the first two principal components (PCs) reveal a circular trajectory that coincides with relative DNA content, determined using a cell permeable DNA content stain (top panel in Figure 2a). Unsupervised neighborhood clustering identified five distinct clusters of parasites based on their transcriptional profiles (middle panel in Figure 2a) (Materials and methods). We computed RNA velocity to infer transcriptional dynamics\textsuperscript{39,40} and the velocity vector field indicates a net “counter-clockwise” flow of transcriptional changes (bottom panel in Figure 2a) (Materials and methods). We assigned cell cycle phase to the clusters based primarily on change in DNA content (Figure 2 - Supplementary Figure 3a) but also considering previous bulk transcriptomic characterization\textsuperscript{28} (Figure 2 - Supplementary Figure 3b). Unsupervised clustering identified two distinct clusters in G\textsubscript{1} state, which we have designated as G\textsubscript{1}a and G\textsubscript{1}b. We found a list of differentially expressed genes between the two G\textsubscript{1} clusters. The G\textsubscript{1}a cluster is highly enriched for the expression of metabolic genes such as phenylalanine hydroxylase (\textit{TGGT1_411100}) and pyrroline-5-carboxylate reductase (\textit{TGGT1_236070}), as well as invasion-related secreted factors such as MIC2 (\textit{TGGT1_201780}), MIC3 (\textit{TGGT1_319560}), and MIC11 (\textit{TGGT1_204530}). On the other hand, G\textsubscript{1}b cluster is enriched for the expression of 3-ketoacyl reductase (\textit{TGGT1_217740}) and cytidine and deoxycytidylate deaminase (\textit{TGGT1_200430}), as well as numerous uncharacterized proteins (Supplementary File 2202). The relative abundance of G\textsubscript{1}a, G\textsubscript{1}b, S, M, and C states were determined to be 18\%, 28\%, 15\%, and 7\%, respectively. Without chemical synchronization, the
The correlation between the scRNA-seq data of asynchronous parasites and previously published bulk transcriptomic measurement suggests strong agreement in cluster assignment and cell cycle state identification (Figure 2 - Supplementary Figure 3b). This highlights a key advantage of scRNA-seq, as it enables identification of cell cycle status of a parasite without reliance on chemical induction, which may lead to unnatural cellular behavior.

To verify the cyclical nature of gene expression through the lytic cycle, we reconstructed a biological pseudotime of RH using Monocle 2 (Materials and methods). The results show a clear oscillatory expression pattern for the variably expressed genes along the pseudotime axis (Figure 2b). To further characterize cell cycle expression patterns, we clustered genes based on pseudotime interpolation and hierarchical clustering (Materials and methods). Some of the key organelles in tachyzoites are made at different times in the cell cycle. To confirm and refine this finding, we calculated the mean expression values for each set of organelle-specific genes based on their annotation in ToxoDB (Supplementary File 4). This showed the expected, strong oscillation with pseudotime (bottom panel in Figure 2 - Supplementary Figure 3c), which also strongly correlates with the oscillation of DNA and total mRNA content (top panels in Figure 2 - Supplementary Figure 3c). On the other hand, we also observed instances where a given gene’s expression was discordant to the dominant trend of its nominal organelle set (arrows in Figure 2 - Supplementary Figure 3d). For example, 63.5% of genes annotated as rhoptry (ROP) or rhoptry neck (RON) are assigned pseudotime cluster 3, while the remaining 36.5% rhoptry genes are assigned pseudotime clusters 1 or 2 (Figure 2 - Supplementary Figure 3e). Specifically, genes annotated as ROP33 and ROP34, based on their homology to genes encoding known rhoptry proteins, are assigned to cluster 2 instead of cluster 3 (left panel in Figure 2 - Supplementary Figure 3f). Recent reports have experimentally determined these two to be non-rhoptry-localizing proteins. This is consistent with our observation of discordance between their and known ROPs' expression profiles along the pseudotime. Through analysis of pseudotime clustering, we also identified genes not annotated as ROPs within the ROP-dominated cluster 3, such as TGGT1_218270 and TGGT1_230350, that have recently been shown to encode bona fide rhoptry and rhoptry neck proteins, now designated as ROP48 and RON11, respectively (left panel in Figure 2 - Supplementary Figure 3f). As another example, IMC2a peaks in expression level in G1, while the majority of inner-membrane complex (IMC) genes are expressed towards the M/C phase of the cell cycle (right panel in Figure 2 - Supplementary Figure 3f). A recent report has proposed reannotation of IMC2a as a dense granule (GRA) protein (GRA44) based on subcellular localization, which is consistent with our unsupervised group assignment of IMC2a as falling in cluster 1 where GRA genes dominate. A list of 8590 RH genes with their corresponding pseudotime clustering assignment is provided (Supplementary File 5). We observe high discordance of pseudotime expression for several genes in each
annotated organelle sets, suggesting that the current *Toxoplasma* annotation may need significant revision. Our scRNA-seq data provide an important resource to help identify mis-annotated genes and infer putative functions of uncharacterized proteins.

**Hidden heterogeneity in asexually developing *Toxoplasma***

*Toxoplasma* has one of the most complicated developmental programs of any single-celled organism; however, it is unknown how synchronized the transition is between developmental states. To address this, we assessed the inherent heterogeneity within asexually developing Pru, a type II strain that is capable of forming tissue cysts with characteristics that resemble early "bradyzoites" from *in vivo* source upon growth in *in vitro* alkaline conditions. We applied scRNA-seq to measure and analyze Pru parasites grown in HFFs as tachyzoites ("uninduced") and after inducing the switch to bradyzoites by growth in alkaline media for 3, 5, and 7 days. Projection of the first two PCs of uninduced Pru tachyzoites (Day 0) reveals the expected circular projection (*Figure 3 - Supplementary Figure 4a*), presumably reflecting cell cycle progression as seen for the RH tachyzoites, described above. To validate this, we developed a random forest classifier model based on our cell cycle assignment in RH (*Materials and methods*). Comparable to what we observed in RH, cell cycle prediction reveals that the uninduced population of Pru is composed of 28%, 41%, 21%, 7%, and 3% parasites in G₀, G₁, S, M, and C states, respectively. Consistent with previous observation, our data show most induced Pru parasites (Day 3 - 7) are in the G₁ state with a predominance of G₁b (*Figure 3 - Supplementary Figure 4b*).

To identify transcriptomic changes associated with the tachyzoite-bradyzoite transition, we next projected data from both induced and uninduced Pru parasites onto two dimensions using UMAP, a nonlinear dimensionality reduction method (*Materials and methods*). Unsupervised clustering revealed six distinct clusters of parasites, which we label P1-6 (*Figure 3a*). Cluster formations partially correlate with treatment time points and cell cycle states (*Figure 3b; Figure 3 - Supplementary Figure 4c*), suggesting that the asexual differentiation program overlaps with cell cycle regulation in *Toxoplasma*, as proposed previously. We stratified the datasets by days post alkaline induction (dpi) and observed elevated expression of previously described "early" bradyzoite marker genes in induced parasites, including *SRS44 (CST1)* and *BAG1*, with a concomitant reduction in expression of *SRS29B (SAG1)*, a tachyzoite-specific surface marker gene (*Figure 3 - Supplementary Figure 5*). The abundance of SAG1⁺ parasites in the induced population suggests two possible interpretations: (1) the depletion of SAG1 mRNA is relatively slow and we are measuring SAG1 transcripts made when the parasites were still tachyzoites, or (2) the asexual transition induced by alkaline treatment is highly asynchronous. Interestingly, RNA velocity analysis suggests that P3 may be a fate decision point as the trajectory trifurcates into either P4 (cell cycle), P1, or...
P2 as evident by the net transcriptional flow (compare Figure 3a to right panel in Figure 3b).

To determine the gene modules specific to a given cluster, we conducted differential gene expression for each cluster (Figure 3c and Supplementary File). P1 cluster cells are enriched for expression of bradyzoite-specific genes while P2-5 are enriched for that of tachyzoite-specific or cell cycle-associated genes (Figure 3c-d). In our scRNA-seq data, we also observe a small portion of BAG1+ bradyzoites (7.1%) annotated as either S, M, or C states, indicating that they are replicating (Figure 3 - Supplementary Figure 4d). Our data supports the notion that bradyzoites can undergo cell cycle progression. We observe a family of transcription factors known as Apetala 2 that are differentially expressed across different clusters, some of which are implicated in Toxoplasma development (Figure 3e). In particular, we identify AP2Ib-3131, AP2IX-1, AP2IX-6, and AP2VI-2 as over-expressed in P1, suggesting their potential roles in the regulation of developmental transition, while AP2IX-9, AP2X-8, AP2VIIa-6, 315AP2XI-1, AP2IX-3, AP2VIII-7, and AP2-domain protein TGME49_215895 (not yet assigned a formal AP2 number), are highly expressed in P6, hinting at their possible roles in defining this distinct cluster of parasites.

The most highly expressed genes in P6 include genes enriched in P2 as well as bradyzoite-specific genes found in P1 (Figure 3c). To identify genes that are specifically expressed in P6, we used Wilcoxon’s test (Figure 3 - Supplementary Figure 6a) (Materials and methods) between P6 and P2 or P1. Comparison of our data to previous bulk transcriptomic measurement in tachyzoites, tissue cyst, or isolates at the beginning or the end of sexual cycle showed no specific enrichment in known developmental stages (Figure 3 - Supplementary Figure 6b). Instead, we show that based on their expression, P6 forms a distinct sub-population of parasites which suggests that alkaline induced Toxoplasma may be more heterogeneous than previously thought. Thus, scRNA-seq resolves a transcriptomic landscape of asexual development and suggests the existence of an otherwise hidden state.

To determine the reproducibility of the phenomena we observed in the differentiating Pru strain parasites, we repeated the analysis with another widely used Type II strain, ME49, examining 1828 single ME49 parasites exposed to alkaline conditions to induce switching to bradyzoites. Data from the two experiments were computationally aligned using Scanorama to remove technical batch effects while retaining sample-specific differences. Unsupervised clustering revealed 5 distinct clusters in ME49 which share significant overlap in expression patterns with Pru (Figure 3 - Supplementary Figure 7a). Matrix correlation of batch-corrected expression across the two datasets, we next applied Partition-Based Graph Abstraction (PAGA) to present clusters of cells as nodes with connectivity based on similarity of the transcriptional profiles between clusters. A side-by-side comparison of expression of
tachyzoite, bradyzoite, and sexual stage specific genes reveals some key similarities and dissimilarities (Figure 3 - Supplementary Figure 7c). Clusters P1 and M1 are both enriched for the expression of bradyzoite marker genes, while clusters M4-5 and P4-5 are both predicted to be S/M/C phases of the cell cycle. Curiously, P6-specific genes (green panels in Figure 3 - Supplementary Figure 7c) are not enriched in any cluster in ME49, suggesting that P6 state is not conserved across the two type II strains. Such differences may not be surprising, however, as Pru and ME49 have entirely distinct passage histories.

scRNA-seq characterizes surface antigenic repertoire of Toxoplasma

A unique advantage of scRNA-seq over bulk RNA-seq is its ability to measure cell-to-cell variation that is independent of known biological processes. The Toxoplasma gondii genome encodes a family of over 120 SAG1-related sequence (SRS) proteins that fall into distinct subfamilies; most or all of these are presumed to be surface antigens based on their sequence similarities, including the presence of a predicted GPI-addition signal\(^54\). Whether SRSs constitute an antigenic repertoire that contribute to evasion of host adaptive immunity response is unclear; however, existing data on developmentally regulated expression of SRSs including SAG1 (SRS29B) and SRS16B lend support to that hypothesis\(^55,56\). In the experiments using 384-well plates, we noted that most SRS genes were detected in only a few sporadic cells (Figure 4a); however, biological variation was difficult to distinguish from measurement dropout (failure to capture a mRNA molecule) given the limited sensitivity of 384-well assay and the relatively low abundance of SRS transcripts. To increase our sensitivity, therefore, we performed scRNA-seq of extracellular Toxoplasma in 96-well plates which achieved ~40% sensitivity of detection for single molecules of ERCC spike-ins, compared to 14% sensitivity that we obtained in the 384-well experiments, at roughly equivalent sequencing depth (Figure 4a). To determine the extent of transcriptional variation in SRS genes independent of cell cycle or asexual development, we isolated RH parasites in G1 state using DNA content stain and FACS (Figure 4b). We noticed that the distributions of DNA content in single cells varied across the two experiments (compare Figure 4b to Figure 2a), suggesting that the distribution may depend on external factors such as tissue culture confluence, parasite load, or the amount of time passed since infection. We also measured on average 984 genes with read counts equal to or greater than two in 96-well plates, up from 862 genes in 384-well plates (Figure 4c). Analysis of the resulting 96-well plates data show that the vast majority of SRS genes are detected in only a small fraction of the population (Figure 4b), similar to what we observed in 384-well experiments. To control for measurement dropout as a cause of such variation, we first determine the non-zero median for those genes, i.e., the median expression level in...
381 cells where any transcript for that gene is detected. We then assess how often we fail to 382 detect an ERCC spike-ins that had a similar non-zero median expression level. If the 383 failure to detect a SRS was due to measurement dropout, then the fraction of cells 384 without detection for its transcript should be about the same as for the ERCC spike-in 385 with similar non-zero median expression. If, on the other hand, the failure to detect the 386 SRS transcript is due to underlying biological variation between cells, then we will find a 387 lower frequency of cells with detectable transcript for that gene than the ERCC spike-in 388 with similar expression. The results show that SRSs are indeed detected at a 389 substantially lower rate when compared to ERCC spike-ins (Figure 4c), indicating that 390 the variation of SRS detection cannot be explained by measurement noise. Compared 391 to most other genes, SRSs are expressed in a significantly smaller fraction of the 392 population and at a relatively lower abundance (Figure 4 – Supplementary Figure 8d-e). To determine whether variation of SRS expression is due to cell cycle or asexual development, we developed a bootstrapping approach to quantify the dependence of 395 expression on a topological network that represents either process (Materials and Methods). We show that apart from a few known SRSs, most SRSs do not co-vary with 397 cell cycle or asexual development, which suggests that the sporadic nature of SRS 398 expression may hint at a different biological role for their variation, beyond cell cycle and 399 asexual development, as discussed further below (Figure 4d). These results reveal new 400 insights into the antigenic repertoire of Toxoplasma, which appears far more heterogeneous, on a parasite-to-parasite basis, than previously known.

402 Transient expression of AP2IX-1 induces surface antigen switching

403 In examining our data, we were struck by a lone cell in the 384-well dataset of 404 RH that had no detectable level of SRS29B (SAG1), which was otherwise ubiquitously 405 and abundantly expressed in all tachyzoites (Figure 5a). We wondered if this cell was 406 damaged or unhealthy but its gene count and read depth were similar to its cohort 407 (Figure 5b). Further analysis of this SAG1- cell’s transcriptome revealed that it lacks 408 expression of the most abundant tachyzoite-specific and bradyzoite-specific genes; 409 instead, its gene expression most closely resembles the sexual stages from cat intestinal isolates, including abundant expression of a cat-stage SRS, SRS22C (Figure 411 5c). Within the set of genes uniquely expressed in the SAG1- cell are AP2IX-1 and 412 AP2III-4 which suggest a possible role for one or both of these transcription factors in 413 regulating the transcript differences observed in this outlier. To test this hypothesis, we 414 transiently expressed AP2IX-1 under the control of a strong promoter in RH parasites 415 (Figure 5d). Consistent with our hypothesis, immunofluorescence assay (IFA) revealed 416 a reduction of SAG1 surface protein expression within ~18-20 hours after transfection 417 (Figure 5e-f), while quantitative RT-PCR showed significantly higher mRNA expression 418 for several of the genes that were upregulated in the SAG1- cell: namely, 207965, 419 222305, 205210, and SRS22C, the latter being over 1000-fold higher in the transfected
population than control (Figure 5g). Note that the SAG1 transcript levels in the transfected population were not substantially lower, as expected because there remains a large number of untransfected cells in the population which still express high levels of this gene. As SRS22C is predicted to be a surface antigen like the rest of SRS, this suggests that AP2IX-1 induction can control the switching of surface antigens. We also attempted to express AP2III-4 that is upregulated in the SAG1- cell but were unable to obtain an epitope-tagged version of the gene (which is over 9 kbp, including the promoter). Overall, these results demonstrate the ability to infer transcriptional regulation from a single parasite cell that has an unusual co-expression pattern, revealing AP2IX-1 as a novel transcriptional factor that can alter antigen expression in Toxoplasma.

Comparative analysis of Plasmodium and Toxoplasma reveals shared expression programs

Plasmodium and Toxoplasma are both unusual for their prevalence in humans and their complex developmental transition, despite the fact that humans are non-definitive hosts for both pathogens. Toxoplasma has been used as an experimental model for other apicomplexans including Plasmodium, yet the replication modes amongst apicomplexans can be very different. For example, asexual replication of Toxoplasma involves endodyogeny, which is similar to canonical binary fission, in order to replicate and divide. Asexual division of Plasmodium, on the other hand, involves schizogony, a process that entails multiple rounds of DNA replication and nuclear division followed by a mass cytokinesis. Despite their fundamental differences in cell cycle progression, we wondered if our Toxoplasma dataset can be combined with the Malaria Atlas to provide insights into apicomplexan biology. Toward this end, we first identified 1830 one-to-one orthologous genes between Plasmodium berghei and Toxoplasma gondii Pru (Figure 6 - Supplementary Figure 9a). As expected, the vast majority of orthologous genes are not surface adhesion factors or effector molecules in Toxoplasma (Figure 6 - Supplementary Figure 9b). After removing non-orthologous genes, we combined the two species’ datasets to produce an integrated UMAP projection with Scanorama. In this integrated projection space, mutually similar clusters of cells in the two datasets are brought close together, while organism-specific cell types are not (Figure 6a). Transitional dynamics of the parasitic development are preserved in the integrated projection, as separation of the original cluster assignment indicates (Figure 6b). We highlight comparative similarity across the two organisms by calculating the fraction of cells that share the same topological neighborhood in the integrated network of Plasmodium and Toxoplasma (Figure 6c), revealing striking similarity of expression pattern between the two organisms. We discover the concerted transcriptional expression of several conserved orthologous gene sets in the life cycles of these two parasites (Figure 6d). For example, the pre-erythrocytic stages of
*Plasmodium* (sporozoite and merozoites) most closely match the “G1 a” stage of *Toxoplasma*. Both parasites at these stages are not actively replicating DNA and express high levels of ribosomal and mitochondrial genes. Meanwhile, the ring stage of *Plasmodium* most closely matches the “G1 b” phase of *Toxoplasma* because they both express high levels of ribosomal genes with minimal expression of IMC-related and microtubule-related genes. Exo-erythrocytic form (EEF), microgametes (male), and trophozoite stages of *Plasmodium berghei* most closely match the “S” phase of *Toxoplasma* cell cycle. This is because they express DNA-replication factors and centrosome components, which are required for condensation and segregation of chromosomes. Intriguingly, the schizont stage of *Plasmodium* most closely resembles the “M” and “C” phases due to the expression of microtubule, centrosome, and IMC genes. A more detailed illustration of these gene sets expression is shown in Figure 6 – Supplementary Figure 9c. Our results indicate that while the cellular morphology and replication strategies may differ drastically between the two parasites, the cellular state, as defined by the transcriptomic profiles, can bear striking analogy and resemblance in the course of asexual replication.

**An interactive resource for visualizing single- *Toxoplasma* atlas**

Toward enabling the larger scientific community to take advantage of the scRNA-seq data we collected on the three *Toxoplasma* strains, we have constructed an interactive single-cell atlas for *Toxoplasma gondii* ([http://st-atlas.org](http://st-atlas.org)) using Bokeh and Javascript. Our atlas resource allows users to visualize the expression pattern of individual cells by providing a gene ID of interest and using the built-in graphical interface toolset. A factor plot displays expression levels within the parasite population based on cell cycle status, cluster identities, days post induction, or other categories. A sub-panel on the bottom left shows additional information for each input gene including translated product, mean, and standard deviation of the expression. Users can selectively highlight a subset of parasites by using a simple click-and-drag interface for further analysis. Example use cases are provided in Figure 6e-f. We have also made the raw data available through the atlas website. We hope this will help make our data and results readily accessible for others interested in exploring *Toxoplasma* parasitology.
We describe here single-cell RNA sequencing (scRNA-seq) for measurement of mRNA transcripts from individual extracellular *in vitro* *Toxoplasma gondii*, an obligate intracellular protozoan parasite. The results show that scRNA-seq can reveal intrinsic biological variation within an asynchronous population of parasites. Two types of biological variation could be seen in our asynchronous populations: cell cycle progression and asexual differentiation. We found the existence of two distinct 1N transcriptional states in cycling parasites which we call G$_{1a}$ and G$_{1b}$, concurring with what was previously reported in bulk analyses of *Toxoplasma*\(^ {28}\). Interestingly, bradyzoites are found predominantly in G$_{1b}$ but not in G$_{1a}$, suggesting the possibility of a putative checkpoint between these two phases that may also play a role in regulating the developmental transition. Our data further shows a small fraction of bradyzoites to be cycling which supports the hypothesis that bradyzoites can in fact divide\(^ {22}\). Our results showed a very strong correlation between cell cycle and expression of genes encoding proteins in various subcellular organelles, as noted previously using synchronized bulk populations\(^ {28}\). The results here, however, show an even more dramatic and extreme dependence on cell cycle, allowing refinement of approaches that use such timing to predict a given protein’s ultimate organellar destination in the cell\(^ {43}\).

They also extend such analyses to the Type II strains, Pru and ME49, which have not previously been examined in this way.

In addition to the above, we observed some striking and unexpected heterogeneity within asexually developing parasites. We discovered a cluster of cells, labeled P6, in the differentiating Pru parasites that is distinct from the rest of the alkaline-induced population of cells. Constituting 21% of the alkaline-induced population, the P6 cluster is marked by a set of genes that were previously detected by bulk transcriptomics in bradyzoites of tissue cysts\(^ {51}\). Remarkably, while most of these genes have unknown functions, we identified an enriched gene with a predicted AP2 domain, which may contribute to the unique expression pattern observed in this group of parasites. We found that the P6 expression profile is intermediate to P2 tachyzoites and P1 bradyzoite clusters. Interestingly, the genes enriched in P6 overlap with a subset of canonical bradyzoite marker genes including *LDH2* and *SRS35A*, albeit expressed at a lower level than in P1 (Figure 3c). In addition, we observed a gradual increase in the proportion of P6 cells as induction proceeded from day 3 to day 7. Taken together, one possible explanation for the emergence of P6 cluster is a reverted conversion from bradyzoites to tachyzoites in which alkaline stress fails to maintain the bradyzoite state. Our data and previous reports are consistent with this interpretation\(^ {60}\). On the other hand, we cannot rule out the possibility that this cluster is developmentally “confused” by the presence of a general stressor such as alkaline. RNA velocity analysis in the Pru data does not reveal a strong transcriptional flow between P1 and P6. Rather, P6
appears to transcriptionally transition from P2 tachyzoites. Thus, the P1 bradyzoites and P6 parasites are either distinct and separate developmental trajectories, or the transition from P1 to P6 is a rapid and rare event. Regardless, our results reflect a surprising diversity in an asexually transitioning population of *Toxoplasma*. Future measurement of single parasites isolated from *in vivo* sources coupled with genetic manipulation of the parasite genome, will further clarify the underlying developmental states that we identified here.

To quantify the variation of SRSs, which are generally expressed at low copy number, we performed 96-well Smart-seq2, which greatly improved measurement sensitivity over the 384-well format, likely due to changes in the input mRNA concentration. For scRNA-seq of pathogens, which tend to have smaller size and lower mRNA content than mammalian cells, we think a careful selection of the measurement approach is necessary based on consideration of throughput and measurement sensitivity, between which there is often a tradeoff. Combined with a novel approach that we developed based on random permutation and K-nearest neighbor (KNN) averaging, we were able to quantify the association of gene expression variation to known biological processes, like cell cycle and development. We discovered that *Toxoplasma* exhibits unexplained, sporadic variation in the expression of most SRSs, which may have biological implications. For example, it could expand the mode of interactions with the host and be the result of strong selective pressure to maximize invasion efficiency and transmission in a variety of different host species of cell types. Maintaining a large phenotypic diversity can be beneficial in ensuring at least some members will be able to invade the cells it encounters and/or evade adaptive immune response, enabling propagation in whatever the host environment encountered.

Very surprisingly, our scRNA-seq analysis identified an atypical co-expression pattern in an *in vitro* RH “tachyzoite” that is indicative of sexual development, which has not been previously observed in these culture conditions. This suggests that at least the beginnings of sexual developmental can spontaneously occur even in the absence of the cat intestinal environment or other chemical cues. Combined with transient expression experiments, the data from this cell enabled us to show that AP2IX-1 is sufficient to drive a switching of surface antigen expression toward that resembling the sexual stages of the parasite. Assuming this change in mRNA abundance translates into a change in protein levels of TGGT1_222305, which is predicted to contain a transmembrane domain, and SRS22C, which is strongly indicated to be a surface antigen like the rest of SRS, our data indicate that AP2IX-1 contributes to remodeling of the surface antigen repertoire during differentiation. Furthermore, this suggests AP2IX-1 may play a causal role in controlling the sexual differentiation of *Toxoplasma*.

Considering that the family of AP2 transcription factors was originally found to regulate stress response and floral sexual differentiation in plants, our finding suggests the biological role of AP2 family in *Toxoplasma* is evolutionarily conserved. Switching of surface antigens in parasites may be particularly favorable to parasites under stressful conditions.
conditions, perhaps including the stress of an immune response, thereby enabling evasion of host immunity. Regardless, these results show that scRNA-seq can reveal rare parasite variants that are, presumably, a result of spontaneous epigenetic changes similar to what has been described in cancer cells or transcriptional noise which was previously characterized in Escherichia coli. An important difference from the situation with cancer cells, however, is that these individual variants may be non-viable and so impossible to obtain as a stable line; thus scRNA-seq may be uniquely able to provide a detailed understanding of their very interesting and informative gene expression. In this case, a previously uncharacterized AP2 factor was revealed and shown to be responsible for regulating at least some of the genes that were uniquely expressed in this variant, relative to the remainder of the tachyzoites in this population. Expanding the number of AP2 transcription factors (totaling over 68 of them) analyzed in this way could enable the deduction of the sets of “regulons” in this parasite. Thus, even though these rare variants may be non-viable “biological freaks”, they can be highly informative and would be completely undetectable in bulk measurement.

Recently, cross-species analysis of scRNA-seq datasets has attracted considerable interest. Our study provides the first comparative analysis of developmental processes between two apicomplexans, both of which cause prevalent and potentially devastating diseases. While Plasmodium and Toxoplasma undergo distinct modes of cell replication and asexual development, we identified cross-species clusters that share significant similarity in the expression of orthologous genes. We discovered that the timing of expression in gene sets involved in cell cycle are conserved in the erythrocytic cycle of Plasmodium berghei. Lastly, we have made the datasets of our study available by creating an interactive and easily accessible web-browser. We hope this encourages other individuals interested in single-cell parasitology to actively explore our dataset without having expertise in programming or bioinformatics. Building on the work described here, which lays a foundation for a detailed understanding of the parasite itself, we anticipate future, single-cell co-transcriptomic sequencing of both the host cell and the parasite as a potentially powerful approach to further deconstruct the complexity of parasite-host interactions.
### Materials and Methods

#### Key Resources Table

| Reagent type (species) or resource | Designation               | Source or reference | Identifiers                  | Additional information                                      |
|-----------------------------------|---------------------------|---------------------|------------------------------|-------------------------------------------------------------|
| Cell line (Toxoplasma gondii)     | ME49                      | PMID: 15664907      |                              |                                                             |
| Cell line (Toxoplasma gondii)     | Pru                       | PMID: 18347037      |                              |                                                             |
| Cell line (Toxoplasma gondii)     | RH mCherry / RH           | This work.          |                              |                                                             |
| Cell line (Toxoplasma gondii)     | RH Δhxpgrt                | PMID: 8662859       |                              |                                                             |
| Cell line (Toxoplasma gondii)     | RH GFP                    | Gift of Michael W Panas. |                            |                                                             |
| Recombinant DNA reagent           | pG2A-AP2IX1-V5            | This work.          |                              | Constitutive expression plasmid carrying AP2IX1 in tandem fusion to V5 tag. |
| Chemical compound, drug           | Propidium iodide (PI)     | ThermoFisher        | P3566                        |                                                             |
| Chemical compound, drug           | Sytox Green               | ThermoFisher        | S7020                        |                                                             |
| Chemical compound, drug           | Fixable blue dead cell stain kit | ThermoFisher | L34962                        |                                                             |
| Chemical compound, drug           | Vybrant DyeCycle Violet (DCV) | ThermoFisher | V35003                        |                                                             |
| Chemical compound, drug | VECTASHIELD Antifade Mounting Medium with DAPI | Vector Laboratories | H-1200-10 |
|------------------------|------------------------------------------------|--------------------|-----------|
| Chemical compound, drug | Nuclease-free water | IDT | 11-04-02-01 | For SmartSeq2 protocol |
| Chemical compound, drug | recombinant RNase inhibitor | Takara Clonetech | 2313A | For SmartSeq2 protocol |
| Chemical compound, drug | 10 mM dNTP | ThermoFisher | R0194 | For SmartSeq2 protocol |
| Chemical compound, drug | ERCC RNA Spike-in Mix | ThermoFisher | 4456740 | For SmartSeq2 protocol |
| Chemical compound, drug | 10% Triton X-100 | Sigma-Aldrich | 93443 | For SmartSeq2 protocol |
| Chemical compound, drug | Buffer EB (elution buffer) | QIAGEN | 19086 | For SmartSeq2 protocol |
| Chemical compound, drug | AMPure XP nucleic acid purification beads | Beckman Coulter | A63880 | For SmartSeq2 protocol |
| Chemical compound, drug | Nucleofector Solution | Lonza | P3 Primary Cell solution | For transient transfection of Toxoplasma gondii |
| Antibody | Rabbit anti-SAG1 polyclonal antibody | PMID: 15944311 | (1:500) |
| Antibody | Mouse anti-V5 tag monoclonal antibody | Invitrogen | R960-25 | (1:1000) |
| Antibody | Goat polyclonal Alexa 488 Fluor-conjugated secondary antibodies | Invitrogen | A28175 | (1:1000) |
| Commercial assay, kit | SsoAdvanced | Bio-rad | 1725271 | qPCR mastermix |
**Table:**

| Sequence-based reagent | All oligos used in this study | See "supplementary_file1_oligos.csv" |
|------------------------|--------------------------------|-------------------------------------|
| Software               | Analysis algorithm             | www.github.com/xuesoso/singleToxoplasmaSeq |
| Software               | Interactive browser            | st-atlas.org                        |

### Cell and Parasite Culture

All *Toxoplasma gondii* strains were maintained by serial passage in human foreskin fibroblasts (HFFs) cultured at 37 C in 5% CO₂ in complete Dulbecco’s Modified Eagle Medium (cDMEM) supplemented with 10% heat-inactivated fetal bovine serum (FBS), 2 mM L-glutamine, 100 U/ml penicillin, and 100 µg/ml streptomycin. *T. gondii* strains used in this study were RH, Pru-GFP, and ME49-GFP-luc.

### In vitro Bradyzoite Switch Protocol

Differentiation to bradyzoite was induced by growth under low-serum, alkaline conditions in ambient (low) CO₂ as previously described. Briefly, confluent monolayers of HFFs were infected with tachyzoites at a multiplicity of infection (MOI) of 0.025 in RPMI 1640 medium (Invitrogen) lacking sodium bicarbonate and with 1% FBS, 10 mg/ml HEPES, 100 U/ml penicillin, and 100 g/ml streptomycin at pH 8.2. The infected HFFs were cultured at 37°C without supplemented CO₂.

### Preparation of Parasites for Fluorescence Activated Cell Sorting (FACS)

HFF monolayers infected with parasites overnight were scraped, and the detached host cells were lysed by passing them through a 25-gauge needle three times. The released parasites were spun down at 800 rpm for 6 minutes to pellet out host cell debris, and the supernatant was spun down at 1500 rpm for 5 minutes to pellet the parasites. The parasites were then resuspended in 500 µL of FACS buffer (1x phosphate-buffered saline, PBS, supplemented with 2% FBS, 50 µg/ml DNAse I, and 5 mM MgCl₂·6H₂O), passed through both a 5 µm filter and a filter cap into FACS tubes, and stored on wet ice until it was time to sort. In samples stained for DNA content, the parasites were resuspended in 500 µL of FACS buffer plus 1.5 µL of
Vybrant DyeCycle Violet (from ThermoFisher, catalog number V35003) and incubated at 37°C and 5% CO₂ for 30 minutes. The parasites were also stained with either propidium iodide (PI), Sytox Green, or the live/dead fixable blue dead cell stain kit (catalog number L34962) prior to sorting in order to distinguish live cells from dead cells. To stain with PI, 10 µL of 0.5 mg/ml PI was added to every 500 µL of parasite suspension in FACS buffer, and the parasites were incubated covered on ice for at least 15 minutes. To stain with Sytox Green, 1 drop of Sytox Green per ml was added to the parasite suspension in FACS buffer, and the parasites were incubated at room temperature for at least 15 minutes. To stain with the live/dead fixable blue dead cell stain kit, 1.5 µL of the kit's viability dye was added to every 500 µL of parasites along with the secondary antibody, and parasites were washed and resuspended in FACS buffer as usual.

FACS of parasites

Eight mL of lysis buffer was prepared by mixing together: 5.888 mL of water, 164 µL recombinant RNase inhibitor (Takara Clonetech), 1.6 mL of 10 mM dNTP (ThermoFisher), 160 µL of 100 µM oligo-dT (iDT; see attached supplementary_file1_oligos.csv for oligos), 1:600,000 diluted ERCC spike-in (ThermoFisher), and 32 µL of 10% Triton X-100. All reagents are declared RNase free. Lysis plates were prepared by dispensing 0.4 µL of lysis buffer into each well of a 384 well hard-shell low profile PCR plate (Bio-rad) using liquid handler Mantis (Formulatrix). Single parasites were sorted using the Stanford FACS Facility's SONY SH800s sorter or BD Influx Special Order sorter into the 384-well plates loaded with lysis buffer. Single color and colorless controls were used for compensation and adjustment of channel voltages. The data were collected with FACSDiva software and analyzed with FlowJo software. RH parasites were index sorted with fluorescence signal of cell permeable DNA stain, DyeCycle Violet.

Single-Toxoplasma cDNA synthesis, library preparation, and sequencing

Smart-seq2 protocol was carried out as previously described using liquid handlers Mantis and Mosquito (TTP Labtech) with slight modifications. 384-well Smart-seq2 was performed with a 2 µL final reaction volume, while 96-well format was carried out in 25 µL final reaction volume as recommended by the original protocol. For 384-well Smart-seq2, we performed 19 rounds of cDNA pre-amplification after reverse transcription with oligo-dT primers. Each well is then diluted with 1 to 4 v:v in RNAse free elution buffer (QiAgen) to a total volume of 8 µL. For 96-well Smart-seq2, we performed 30 rounds of cDNA pre-amplification after reverse transcription. PCR is performed with "IS_PCR primers". Each well is then purified with Ampure XP beads at 670.8X volume ratio and resuspended in 20 µL RNAse free elution buffer. We measured the size distribution and concentration of each well in 96-well plate using Fragment Analyzer High-sensitivity NGS kit (Agilent). We normalized the concentration of cDNA of...
each well to a concentration of 0.4 ng/µL. Finally, for both 384-well and 96-well Smart-seq2 measurements, we conducted library preparation with in-house Tn5 tagmentation using custom cell barcode and submitted for 2 x 150 bp paired-end sequencing on NovaSeq 6000 at the Chan Zuckerberg Biohub Genomics core. All primer sequences are provided as a supplementary file.

675

676**AP2IX-1 transient expression**

677 The pGRA-AP2IX1-V5 plasmid for AP2IX1 transient expression was created using Gibson assembly (NEB) from the pGRA-V5 plasmid. RH parasites were transfected with pGRA-AP2IX1-V5 using the Amaxa 4D Nucleofector (Lonza). Tachyzoites were mechanically released in PBS, pelleted, and resuspended in 20 µL P3 Primary Cell Nucleofector Solution (Lonza) with 7 or 15 µg DNA for transfection. After transfection, parasites were allowed to infect HFFs in DMEM. After 18-20 hours of infection, parasites were prepared for Immunofluorescence Assay (IFA) or qRT-PCR. To compute statistical independence between the transfected and control samples, we applied Student's t-test by assuming unequal variance. One sigma (*) indicates one standard deviation in mean difference assuming null hypothesis.

687

688**Immunofluorescence Assay (IFA) quantification**

689 Monolayers of infected cell on glass coverslips were fixed with cold methanol for 12 minutes. Samples were washed with PBS and blocked using 3% bovine serum albumin (BSA) in PBS for at least 30 minutes. SAG1 was detected with rabbit anti-SAG1 polyclonal antibody and V5 was detected with mouse anti-V5 tag monoclonal antibody (Invitrogen). Primary antibodies were detected with goat polyclonal Alexa Fluor-conjugated secondary antibodies (Invitrogen). Primary and secondary antibodies were both diluted in 3% BSA in PBS. Coverslips were incubated with primary antibodies for 30 minutes, washed, and incubated with secondary antibodies for 30 minutes. Vectashield with DAPI stain (Vector Laboratories) was used to mount the coverslips on slides. Fluorescence was detected using wide-field epifluorescence microscopy and images were analyzed using ImageJ. All images shown for any given condition/staining 700 in any given comparison/dataset were obtained using identical parameters.

701

702**Quantitative polymerase chain reaction (qPCR)**

703 To quantify the purity of single parasite sort and to ensure the cDNA synthesis reaction was not saturated, GFP, mCherry, or SAG1 mRNA expression were measured using commercial qPCR mastermix, SsoAdvanced™ Universal SYBR Green mastermix (Bio-rad). Briefly, 0.1 µL of diluted cDNA was added in a total of 2.1 µL reaction volume per well on a 384 well plate with qPCR mastermix and 200 nM PCR primers. The reaction was incubated on a Bio-rad qPCR thermal cycler with the following programs: 5 minutes of 95°C, 45 cycles of 95°C for 5 seconds and 56°C for 1 minute, and imaging.

710 To quantify the transcriptional effects of AP2IX-1 transient expression in RH parasites,
infected cell monolayers were first lysed with Trizol (Invitrogen) and RNA was extracted using standard molecular biology technique. cDNA from each sample was generated with Smart-seq2 protocol using 20 µL total reaction volume and roughly 200 ng RNA input followed by 0.8X AMPure XP beads (Beckman Coulter) purification. For qPCR, 1 ng of cDNA was added in a total of 2.1 µL reaction volume per well on a 384 well plate with qPCR mastermix and 200 nM PCR primers as described above. The reaction was incubated on a Bio-rad qPCR thermal cycler with the following programs: 5 minutes of 95°C, 60 cycles of 95°C for 5 seconds and 61°C for 30 seconds, and imaging. Each gene was measured four times with samples collected from at least two separate wells. The transfection and qRT-PCR experiments were performed twice in separate experiments. Fold change of gene expression was calculated as shown previously using ACT1 expression as an internal control for samples. All primer sequences are provided in “supplementary_file1_oligos.csv”.

Sequencing alignment
BCL output files from sequencing were converted into gzip compressed FastQs via a modified bcl2fastq demultiplexer which is designed to handle the higher throughput per sequencing run. To generate genome references with spike-in sequences, we concatenated T. gondii ME49 NCBI genome assembly version 3011/2013 or T. gondii GT1 T NCBI Genome assembly version 7/19/2013 genome references with ERCC sequences (obtained from ThermoFisher website). The raw fastq files from sequencing are aligned to the concatenated genomes with STAR aligner (version 2.6.0c) using the following settings: “--readFilesCommand zcat --outFilterType BySJout --outFilterMutlimapNmax 20 --alignSJoverhangMin 8 --alignSJDBoverhangMin 351 --outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.04 --alignIntronMin 20 --alignIntronMax 1000000 --alignMatesGapMax 1000000 --outSAMstrandField intronMotif --outSAMtype BAM Unsorted --outSAMattributes NH HI AS NM MD --outFilterMatchNminOverLread 0.4 --outFilterScoreMinOverLread 0.4 --clip3pAdapterSeq CTGTCTCTTATACACATCT --outReadsUnmapped Fastx”. Transcripts were counted with a custom htseq-count script (version 0.10.0, https://github.com/simon-anders/htseq) using ME49 or RH GFF3 annotations (version 74236 on ToxoDB) concatenated with ERCC annotation. Instead of discarding reads that mapped to multiple locations, we modified htseq-count to add transcript counts divided by the number of genomic locations with equal alignment score, thus rescuing measurement of duplicated genes in the Toxoplasma genome. Parallel jobs of STAR alignment and htseq-count were requested automatically by Bag of Stars (https://github.com/iosonofabio/bag_of_stars) and computed on Stanford high-performance computing cluster Sherlock 2.0. Estimation of reads containing exonic and intronic regions is computed with Velocyto estimation on the BAM output files and requested automatically by Bag of Velocyto (https://github.com/xuesoso/bag_of_velocyto) on Sherlock 2.0. Gene count matrix is
obtained by summing up transcripts into genes using a custom python script. Scanpy velocyto package is then used to estimate transcriptional velocity on a given reduced dimension. Parameters used for generating the results are supplied as supplementary python scripts. Sample code to generate the analysis figures are provided in supplementary jupyter notebooks.

Data preprocessing

To filter out cells with poor amplification or sequencing reaction and doublet cells, we discarded cells based on gene counts (>0 reads), total reads sum, percent reads mapped to *Toxoplasma* genome, percent ERCC reads, and percent ribosomal RNA reads. We reported "% mapped" based on the meta-alignment output from STAR aligner. We checked for some of the unmapped reads on BLASTn and found the majority of them to map to Toxoplasma 28S ribosomal RNA. Next, we filtered “ribosomal RNA” genes from the gene count matrix. Gene count matrices are normalized as counts per median (CPM):

\[ X_{\text{norm}} = \frac{X}{\text{median}(\sum(X))} \]

where \( X \) is the gene count matrix, \( \sum(X) \) is the read sum for each cell, and \( \text{median}(\sum(X)) \) is the median of read sums. Normalized data are added with a pseudocount of 1 and log transformed (e.g. \( \log_2(X_{\text{norm}} + 1) \)). To determine the detection limit (e.g. 50% detection rate), we modeled the detection probability of ERCC standards with a logistic regression as a function of spike-in amount:

\[ \log_2(Y) = m \cdot \log_2(X_{\text{norm}} + 1) + b \]

where \( X_{\text{norm, ERCC}>0.5} \) is the observed CPM value for ERCC spike-ins above the detection limit, \( Y \) is the amount of ERCC spike-in, \( m \) is the regression coefficient, and \( b \) is the intercept. To reduce the influence of measurement noise, we fit the model only to ERCC spike-ins with mean expression above the detection limit.

Cell cycle analysis and annotation

To determine the transcriptional variation associated with cell cycle, we applied Self-Assembling Manifolds (SAM) to filter for highly dispersed gene sets (>0.35 SAM weights) in asynchronous RH population. Principal components analysis (PCA) is then applied to the filtered and normalized RH data, and the nearest neighbor graph (K=50) is computed using “correlation” as a similarity metric. We identified the putative “G1”
clusters with 1N based on DNA content stain. Parasites in “G1” cluster are further sub-clustered with Louvain Clustering, in which we identified “G1a” and “G1b” clusters with 94 distinct transcriptional profiles. Pearson correlation between single-cell and bulk 95 transcriptomic data is computed between bulk assignment and the scRNA-seq cluster 96 assignment through which each cluster is uniquely assigned with a cell cycle state. To 97 quantify genes that are differentially expressed across cell cycle clusters, we applied 98 Kruskal-Wallis test. Genes are considered differentially expressed if their p-values are 99 less than 0.05 and they are at least 2-fold over-expressed in a cluster compared to the 800 average expression level of other clusters. We computed differential expression across 801 all cell cycle clusters as well as between the “G1a” and “G1b” clusters; the results are 802 uploaded as supplementary files 2 and 3, respectively. To enable cell cycle assignment 803 transfer from RH to Pru and ME49 data, we implemented a random forest classification 804 model trained on RH data. Briefly, this is done by training a model with 1000 estimators 805 on L2-normalized RH expression data containing only cell cycle associated genes in a 806 60-40 split scheme. Then the model is applied to predict cell cycle labels of L2- 807 normalized Pru or ME49 data containing the homologous cell cycle associated genes. 808 The testing accuracy was over 95%.

810 Pseudotime construction and clustering
811 Pseudotime analysis is conducted with Monocle 2 package in R on preprocessed 812 dataset with highly dispersive genes as described previously. A cell in “G1a” is 813 designated as the root cell, and all other cells are placed after this cell in order of their 814 inferred pseudotime. To cluster genes based on their pseudotime expression pattern, 815 high frequency patterns are removed through a double spline smoothing operation. The 816 interpolated expression matrix is then normalized by maximum expression along 817 pseudotime such that the maximum value of gene expression along pseudotime is 818 bound by 1. We then applied agglomerative clustering on this interpolated and 819 normalized expression matrix using “correlation affinity” as similarity metric and 820 “average linkage” method to predict three distinct clusters of genes.

822 Projection dependence scoring
823 To quantify the dependence of expression variation on a two-dimensional projection, we 824 developed a novel approach based on k-nearest neighbor (KNN) averaging. First, a 825 KNN graph is computed by locating nearest neighborhood in a projection using 826 Euclidean distance. We then generated a null expression matrix by shuffling the gene 827 expression matrix along each cell column, such that its correlation with respect to the 828 coordinate on projection is completely lost. Next, we compute an updated gene 829 expression value by taking the average of expression values across the KNN. This is 830 equivalent to:
where $X_{KNN}$ is the updated KNN averaged expression, $M$ is the nearest-neighbor graph with $k$ being the number of nearest neighbor, and $X$ is the log-transformed CPM of observed or null expression matrices. We chose a $k$ of 5 for all our analysis as varying $k$ did not have a large effect on the results (data not shown). In our experiments, we have shown that the first two principal components (PCs) of PCA on RH correspond to the projection of cell cycle progression, and a two-dimensional UMAP projection of Pru corresponds to asexual development and cell cycle progression. We thus computed $X_{KNN}$ for both the original, observed expression matrix and the shuffled, null matrix on either projection to reflect dependence on cell cycle progression and/or asexual development. $X_{KNN}$ is further normalized to have identical sum as the original expression values. A Kolmogorov-Smirnoff two sample test is then computed between the normalized $X_{KNN}$ of the observed matrix and that of the shuffled matrix based on 100 random permutations. The projection-dependence score for each gene is then computed as:

$$S_g = \sqrt{-\log(\hat{p}_g)}$$

where $S_g$ is the projection-dependence score for gene $g$ and $\hat{p}_g$ is the average p-values of 100 tests. We present $S_g$ normalized by the maximum score within each respective data set.

Comparative Analysis of Plasmodium and Toxoplasma scRNA-seq

To integrate scRNA-seq data of *Plasmodium berghei* from Malaria Atlas with our *Toxoplasma* Pru dataset (measured both induced and induced population in 384-well), we first identified one-to-one orthologous genes obtained from PlasmoDB(https://plasmodb.org/) and ToxoDB(https://toxodb.org/toxo/). Next, we filtered each dataset with the ortholog genes. Using scanpy library, we filtered for the intersect of genes that are within the top 800 most dispersed genes in each dataset, resulting in a list of 403 genes. Finally, we used Scanorama with default parameter settings to integrate the two datasets. We calculated co-clustering similarity as follows. We first computed a Leiden clustering on the integrated graph and returned a cluster co-occurrence matrix to the original cluster assignment in *Plasmodium berghei* ("ShortenedLifeStage4") or *Toxoplasma* Pru ("cell_cycle"). Then, the dot product between the two matrices was calculated and normalized such that it has a maximum of one.
Data availability

Analysis scripts, preprocessing scripts, and instructions to obtain the processed data are provided on https://github.com/xuesoso/singleToxoplasmaSeq. A sample jupyter notebook that regenerates some of the analysis and figures is provided in the Git repository. Raw fastq and processed files are deposited on GEO and SRA repositories (GSE145080).
Acknowledgements

We thank Fabio Zanini, Felix Horns, and Geoff Stanley for illuminating discussion and advice to YX on experiments and analysis. We thank Saroja Korullu, Robert Jones, and Vickie Lin for assistance with library preparation and sample submission. We thank Meredith Weglarz and Lisa Nichols at the Stanford Beckman FACS facility for assistance with FACS. We thank Michael W Panas for providing Toxoplasma GFP strain used in this study. This study is supported by National Institute of Health (NIH) RO1 AI021423, AI129529, and Chan Zuckerberg Biohub. YX and TCT are supported by Stanford Interdisciplinary Graduate Bio-X Fellowships. SR is supported by NIH F30 AI124589-03. AF is supported by NIH 5T32AI007328-30 and a Gilliam Fellowship for Advanced Study from Howard Hughes Medical Institute.
**Figure 1.** Schematic of single-cell RNA-sequencing (scRNA-seq) based on a modified Smart-seq2 protocol for 384-well or 96-well plate. A table of strain types with the number of sequenced cells and cells that passed quality checking (QC) is provided.
**Figure 1 - Supplementary Figure 1.**

(a) qPCR measurement of mRNA expression in 301 transgenic Toxoplasma cells expressing GFP or mCherry mixed at 1:1 ratio.

(b) qPCR Ct values of abundant surface protein, SAG1, measured for 374 wells with zero, one, eight, or fifty sorted parasites at 16, 176, 176, and 6 replicates, respectively.

(c) Comparison between “uniquely aligned” (default htseq-count settings) and “multiply aligned” (count each feature with equal read alignment score divided by the number of aligned reads) in the detection rate (> 2 read count) in Type I strain, RH.

(d) A more detailed comparison of detection rate of several parasite-specific gene sets in Pru (uninduced + induced). Genes that are detected more frequently in “multiply aligned” setting are annotated with their gene IDs in the plot.
**Figure 1 - Supplementary Figure 2.**

(a) Comparison of gene counts (≥2 read counts) and total mapped read counts for RH, Pru, and ME49 from left to right, respectively. Text in top left corner indicates the number of parasites that passed cell filtering and were analyzed (blue open circles).

(b) Top panel: distributions of percentage of reads in the analyzed cells that aligned to *Toxoplasma* genome references. Bottom panel: distributions of gene counts (≥2 read counts) in analyzed cells. Uninduced Pru and ME49 were grown in the absence of alkaline (Day 0), whereas induced Pru and ME49 were grown in the presence of alkaline (Day 3 - 7).

(c) Top panel: Logistic regression modeling (green line) of detection limit (50% detection rate, black dotted line) of ERCC spike-ins. Text on top left of each sub-panel indicates the detection limit in absolute molecular counts. Bottom panel: Linear regression modeling (crimson line) of measurement accuracy fitted on ERCC spike-ins with abundance above the detection limit. Text on top left of each sub-panel indicates the coefficient of determination for the regression fit.
**Figure 2.** (a) Projection of the first two principal components in RH data set. Top panel: 612 RH cells are colored by fluorescence measurement (arbitrary unit) of a cell permeable DNA content stain, DyeCycle Violet. Center panel: cells are colored by cluster assignment and labeled by the inferred “cell cycle” state. Bottom panel: RNA velocity vector field is overlaid on top of the inferred state colors, with arrows pointing in the direction of net transcriptional change. (b) Heatmap of the 1465 most variable gene expression are displayed along the rows. Cells are ordered by pseudotime assignment in ascending order along the columns of the heatmap (from left to right). Top colorbar reflects the assignment of inferred state and bottom colorbar reflects the relative fluorescence of DNA content using the same color scheme as in (a).
Figure 2 - Supplementary Figure 3. (a) Top panel: Numbers of RH parasites in each inferred “cell cycle” state. Bottom panel: Density plots of DNA content distributions stratified by the inferred state. (b) Heatmap of average expression correlation between each inferred “cell cycle” state of RH and each time-point of bulk transcriptomic measurement based on chemically synchronized parasites. (c) Absolute mRNA abundance (top panel) and DNA content (center panel) ordered by “cell cycle” pseudotime with individual cells colored by their inferred states. A spline smoothing is applied to approximate a rolling average along the pseudotime (black solid line). Average expression of gene sets based on ToxoDB annotation of organellar destination of the protein product after double spline smoothing (bottom panel). (d) Heatmap of gene expression ordered by organelle sets (top colorbar) and pseudotime cluster (bottom colorbar). “Unannotated rhoptry” refers to genes not annotated in ToxoDB (v.36) as encoding a rhoptry protein but whose expression pattern is highly concordant with the dominant rhoptry pattern. (e) Pie charts of pseudotime cluster frequency for parasite organelle sets. (f) Expression of annotated rhoptry (left panel) and inner-membrane complex (IMC; right panel) genes along pseudotime with different colors indicating genes concordant (blue) and discordant (crimson and orange) to the major trend of their organelle sets. Discordant genes which are known to be misannotated are highlighted in orange.
Figure 3. (a) UMAP projection of 809 uninduced and 1389 induced Pru parasites with colors indicating Louvain cluster assignment. Top panel shows the number of parasites in each cluster. (b) UMAP projections of Pru parasites colored or labeled by days post induction (dpi), inferred cell cycle states, and RNA velocity from left to right. (c) Heatmap of differentially expressed genes (along columns) across Louvain clusters of cells ordered by hierarchical clustering (along rows). The top 5 most enriched genes from each cluster are presented. (d) UMAP projections of Pru colored by the neighbor-averaged expression (log$_2$ CPM) of bradyzoite (top panels, purple background) and tachyzoite (bottom panels, red background) marker genes. (e) Heatmap of average expression level of AP2 transcription factors for each Louvain cluster, normalized by the maximum cluster expression level within each AP2. Purple and green rectangles highlight AP2s enriched in clusters P1 and P6, respectively.
Figure 3 - Supplementary Figure 4. (a) PCA projection of Pru stratified by days post induction (dpi) and colored by predicted cell cycle state. (b) Frequency of predicted cell cycle states at different dpi time points. (c) Frequency of Louvain clusters (top panels) and predicted cell cycle states in each cluster (bottom panels). (d) Rolling average frequency of predicted cell cycle states (colored lines) ordered by expression level of the canonical bradyzoite marker, BAG1 (black line).
Figure 3 - Supplementary Figure 5. Expression level (log$_2$ CPM) of four “bradyzoite-specific” marker genes compared to that of “tachyzoite-specific” marker gene, SAG1, stratified by days post induction (dpi; columns).
Figure 3 - Supplementary Figure 6. (a) UMAP projections of Pru colored by the neighbor-averaged expression (log$_2$ CPM) of top 8 most differentially expressed and enriched genes in P6 cluster relative to P1 and P2, two most closely related clusters.

(b) Comparison of P6-specific genes in bulk measurement of tachyzoites, tissue cysts, or enteroepithelial stages (EES1-EES5) (left) and scRNA-seq of Pru Louvain clusters (right).
Figure 3 - Supplementary Figure 7. (a) UMAP projections of Pru and ME49 aligned by Scanorama. Cluster assignment was performed independently in each dataset. (b) Matrix correlation of cluster averaged expression between Pru and ME49. (c) Partition-based graph abstraction (PAGA) of aligned clusters with each being represented as a node connected by linkage with a connectivity threshold of 0.8. Node size reflects relative abundance of the cluster. Node colors reflect relative expression level of gene denoted in the bottom left of each panel, normalized to the maximum cluster expression of the corresponding data set (Pru or ME49). Three marker genes that are specific to bradyzoite (blue background), tachyzoite (red background), and Pru P6 cluster (green background) are shown.
Figure 4. (a) Read depth (top) and detection rate of ERCC spike-ins (bottom) reveal higher measurement sensitivity in 96 well plate over 384 well plate format at roughly equivalent sequencing depth. (b) Hierarchical clustering heatmap of binarized expression reveals sparse expression pattern of SRS. SRS gene expression is binarized by converting expression $>2$ read counts to 1 and $<2$ to 0. (c) Comparison of ERCCs with similar mean expression level reveals that low detection rate of SRS cannot be fully explained by measurement dropout. (d) We quantified the projection dependence score of SRS expression to cell cycle projection in 384-well RH or asexual development projection in 384-well Pru, reflecting the degree of co-variation of SRS expression with respect to these two biological processes. The detection rate in 96-well RH dataset is plotted along the y-axis. This analysis reveals that some of the SRS variation cannot be readily explained by either cell cycle or asexual development.
**Figure 4 - Supplementary Figure 8.**

(a) SRSs rank ordered by percent expression in each single-parasite dataset. Each open circle depicts a single SRS in the corresponding dataset.

(b) Fluorescence measurement of extracellular RH stained with DNA content stain. G1 sub-population was sorted into 96-well plate for scRNA-seq following Smart-seq 2 protocol.

(c) Gene count (≥2 read counts) and read depth of 96-well scRNA-seq for cells that failed (red) or passed (blue) quality checking. Inset bar plot highlights the number of cells before and after filtering, revealing that >82% of cells passed quality checking.

(d) Majority of SRS is detected at lower frequency than most other *Toxoplasma* genes and has low expression level.

(e)
Figure 5. (a) PCA projection of 384-well RH cells colored by the presence (blue) or absence (red) of the SRS29B (SAG1) gene. (b) Gene count and total reads mapped for SAG1+ (blue) and SAG1- (red) cells. (c) Comparison of house-keeping, tachyzoite-specific, bradyzoite-specific, and SAG1-enriched gene expression (log2 CPM) in scRNA-seq to bulk RNA-seq measurement of enteroepithelial stages (EES1-EES5), tachyzoites, and tissue cysts in RH. Genes that were pursued further by qRT-PCR are indicated in red; AP2IX-1 is colored orange. (d) Schematic illustration of transfection experiment. AP-2 IX-1 from genomic DNA of RH was cloned into a pGRA1 backbone vector and expressed constitutively in transfected parasites. Expression level of gene and protein was quantified with immunofluorescence assay (IFA) and qRT-PCR. (e) Representative IFA images showing anti-SAG1 (left), anti-V5-AP2IX-1 (center), and merged channels (right). Blue arrows point at parasites with detected expression of V5-AP2IX-1. (f) IFA quantification of anti-SAG1 signal shows dosage dependence of SAG1 protein expression on amount of AP2IX-1 plasmid transfected, with error bars depicting the standard deviation of quantification and asterisks representing statistical significance of two-tailed t-test for independence. (g) qPCR measurement shows that transfection of 15 µg AP2IX-1 plasmid, compared to no DNA control, induced the expression of AP2IX-1 and several putative surface antigens, including SRS22C, in two independent experiments with each measured in biological triplicates. AP2IX-1 transfection did not lead to increased expression of beta tubulin (bTUB) or SAG1. Error bars depict standard deviation of the measurement.
Figure 6. (a) Scanorama integration of *Plasmodium berghei* (red, left) and *Toxoplasma gondii Pru* (blue, right). (b) Cell cycle of *Toxoplasma* is well-aligned to the erythrocytic cycle of *Plasmodium berghei*, despite fundamental differences in cell cycle progression between these two apicomplexans. Each cell is colored by the original cluster assignment in the corresponding dataset. (c) Normalized cluster similarity between the original cluster assignment of *Plasmodium berghei* and *Toxoplasma Pru*. Cluster similarity is calculated by quantifying the fraction of cells that overlap in topological network in each cluster of the corresponding dataset. (d) Heatmap of concerted gene sets expression normalized to 1 within each cluster of cells in *Plasmodium* (left) and *Toxoplasma* (right). (e) Single-cell atlas of *Toxoplasma gondii* can be interactively visualized (http://st-atlas.org). Individual cells with distinct expression pattern can be highlighted using built-in graphical interface tools. (f) Groups of cells can be selectively displayed based on cell cycle state, developmental clusters, and other categories. Hovering over an individual cell with cursor reveals gene expression level, sample id, and group membership for further analysis. All our datasets are available for download on the atlas.
Figure 6 – Supplementary Figure 9. (a) Number of total *Toxoplasma* genes and one-to-one orthologs in *Plasmodium*. (b) Number of ortholog genes grouped by annotation in *Toxoplasma* genome. (c) UMAP projection of *Plasmodium berghei* (top row) or *Toxoplasma* Pru (bottom row) with cells colored by the average expression level of ortholog gene sets. The expression value is normalized to a maximum of 1 for each gene before taking the average such that the normalized value is not skewed towards genes with high dynamic range. Bottom panel of each subplot shows a violin plot of the average gene sets expression distribution stratified by cluster assignment. Gene sets covary in closely related regions of the projection space in both organisms.
Supplementary File 1. A list of sequence-based reagents (e.g. oligos) used in this study.

Supplementary File 2. A table of summary statistics and fold change of genes that are found to be differentially expressed between “G1 a” and “G1 b” parasites in RH. Genes are considered differentially expressed only if the adjusted p-values of Kruskal-Wallis test are less than 0.05 and that they are at least 2-fold over-expressed in either of the cell states.

Supplementary File 3. A table of summary statistics and fold change of genes that are found to be differentially expressed between across all cell cycle states in RH parasites. Genes are considered differentially expressed only if the adjusted p-values of Kruskal-Wallis test are less than 0.05 and that they are at least 2-fold over-expressed in any of the cell states.

Supplementary File 4. A table summarizing the hierarchical clustering results and encoded gene products. Clustering is performed on smoothened gene expression based on biological pseudotime in RH parasites.

Supplementary File 5. A table of summary statistics and fold change of genes that are found to be differentially expressed between across all clusters in Pru parasites. Genes are considered differentially expressed only if the adjusted p-values of Kruskal-Wallis test are less than 0.05.

Supplementary File 6. A table of summary statistics and fold change of genes that are found to be differentially expressed between across all clusters in ME49 parasites. Genes are considered differentially expressed only if the adjusted p-values of Kruskal-Wallis test are less than 0.05.
References

1. Pappas, G., Roussos, N. & Falagas, M. E. Toxoplasmosis snapshots: Global status of Toxoplasma gondii seroprevalence and implications for pregnancy and congenital toxoplasmosis. *Int. J. Parasitol.* (2009). doi:10.1016/j.ijpara.2009.04.003

2. Rabaud, C. *et al.* Extracerebral toxoplasmosis in patients infected with hiv a french national survey. *Med. (United States)* (1994). doi:10.1097/00005792-199411000-00004

3. Robert-Gangneux, F. *et al.* Molecular diagnosis of toxoplasmosis in immunocompromised patients: A 3-year multicenter retrospective study. *J. Clin. Microbiol.* (2015). doi:10.1128/JCM.03282-14

4. Sutterland, A. L. *et al.* Beyond the association. Toxoplasma gondii in schizophrenia, bipolar disorder, and addiction: Systematic review and meta-analysis. *Acta Psychiatr. Scand.* (2015). doi:10.1111/acps.12423

5. Vyas, A., Kim, S.-K., Giacomini, N., Boothroyd, J. C. & Sapolsky, R. M. Behavioral changes induced by Toxoplasma infection of rodents are highly specific to aversion of cat odors. *Proc. Natl. Acad. Sci.* (2007). doi:10.1073/pnas.0608310104

6. Soëté, M., Camus, D. & Dubremetz, J. F. Experimental induction of bradyzoite-specific antigen expression and cyst formation by the RH strain of Toxoplasma gondii in vitro. *Exp. Parasitol.* (1994). doi:10.1006/expr.1994.1039

7. Jeffers, V., Tampaki, Z., Kim, K. & Sullivan, W. J. A latent ability to persist: differentiation in Toxoplasma gondii. *Cellular and Molecular Life Sciences* (2018). doi:10.1007/s00018-018-2808-x

8. Buchholz, K. R. *et al.* Identification of Tissue Cyst Wall Components by Transcriptome Analysis of In Vivo and In Vitro Toxoplasma gondii Bradyzoites. *Eukaryot. Cell* (2011). doi:10.1128/ec.05182-11

9. Manger, I. D. *et al.* Expressed Sequence Tag Analysis of the Bradyzoite Stage of Toxoplasma gondii: Identification of Developmentally Regulated Genes. *Infect. Immun.* (1998).

10. Pittman, K. J., Aliota, M. T. & Knoll, L. J. Dual transcriptional profiling of mice and Toxoplasma gondii during acute and chronic infection. *BMC Genomics* (2014). doi:10.1186/1471-2164-15-806

11. Yip, K. Our implementation of the SCA method. 1–4 (2007).

12. Cleary, M. D., Singh, U., Blader, I. J., Brewer, J. L. & Boothroyd, J. C. Toxoplasma gondii Asexual Development: Identification of Developmentally Regulated Genes and Distinct Patterns of Gene Expression. *Eukaryot. Cell* (2002). doi:10.1128/ec.1.3.329-340.2002

13. Radke, J. R. *et al.* The transcriptome of Toxoplasma gondii. *BMC Biology* (2005). doi:10.1186/1741-7007-3-26

14. Chen, L. F. *et al.* Comparative studies of Toxoplasma gondii transcriptomes: Insights into stage conversion based on gene expression profiling and alternative splicing. *Parasites and Vectors* (2018). doi:10.1186/s13071-018-2983-5

15. Fouts, A. E. & Boothroyd, J. C. Infection with Toxoplasma gondii bradyzoites has a diminished impact on host transcript levels relative to tachyzoite infection. *Infect. Immun.* (2007). doi:10.1128/IAI.01228-06

16. Hong, D.-P., Radke, J. B. & White, M. W. Opposing Transcriptional Mechanisms Regulate Toxoplasma Development. *mSphere* (2017). doi:10.1128/msphere.00347-16

17. White, M. W., Radke, J. R. & Radke, J. B. Toxoplasma development - turn the switch on or off? *Cellular Microbiology* (2014). doi:10.1111/cmi.12267

18. Soete, M., Fortier, B., Camus, D. & Dubremetz, J. F. Toxoplasma gondii: Kinetics of bradyzoite-tachyzoite interconversion in vitro. *Exp. Parasitol.* (1993). doi:10.1006/expv.1993.1031
110719. Watts, E. et al. Novel Approaches Reveal that Toxoplasma gondii Bradyzoites within Tissue Cysts Are Dynamic and Replicating Entities In Vivo. _MBio_ (2015).
1109 doi:10.1128/mbio.01155-15
111020. Ferguson, D. J. P., Huskinson-Mark, J., Araujo, F. G. & Remington, J. S. A morphological study of chronic cerebral toxoplasmosis in mice: comparison of four different strains of Toxoplasma gondii. _Parasitol. Res._ (1994). doi:10.1007/BF00932696
111321. Radke, J. R., Guerini, M. N., Jerome, M. & White, M. W. A change in the premitotic period of the cell cycle is associated with bradyzoite differentiation in Toxoplasma gondii. _Mol. Biochem. Parasitol._ (2003).
111622. Sinai, A. P. et al. Reexamining Chronic Toxoplasma gondii Infection: Surprising Activity for a "Dormant" Parasite. _Current Clinical Microbiology Reports_ (2016). doi:10.1007/s40588-016-0045-3
111923. Jerome, M. E., Radke, J. R., Bohne, W., Roos, D. S. & White, M. W. Toxoplasma gondii bradyzoites form spontaneously during sporozoite- initiated development. _Infect. Immun._ (1998).
112224. Ali, F. et al. Cell cycle-regulated multi-site phosphorylation of Neurogenin 2 coordinates cell cycling with differentiation during neurogenesis. _Development_ (2011).
1124 doi:10.1242/dev.067900
112525. Kim, D. H. et al. The CRL4Cdt2 Ubiquitin Ligase Mediates the Proteolysis of Cyclin-Dependent Kinase Inhibitor Xic1 through a Direct Association with PCNA. _Mol. Cell. Biol._ (2010). doi:10.1128/mcb.01135-09
112826. Radke, J. R. & White, M. W. A cell cycle model for the tachyzoite of Toxoplasma gondii using the Herpes simplex virus thymidine kinase. _Mol. Biochem. Parasitol._ (1998).
1130 doi:10.1016/S0166-6851(98)00074-7
113127. Conde de Felipe, M. M., Lehmann, M. M., Jerome, M. E. & White, M. W. Inhibition of Toxoplasma gondii growth by pyrrolidine dithiocarbamate is cell cycle specific and leads to population synchronization. _Mol. Biochem. Parasitol._ (2008).
1134 doi:10.1016/j.molbiopara.2007.09.003
113528. Behnke, M. S. et al. Coordinated progression through two subtranscriptomes underlies the tachyzoite cycle of toxoplasma gondii. _PLoS One_ (2010).
1137 doi:10.1371/journal.pone.0012354
113829. Wang, B. et al. Stem cell heterogeneity drives the parasitic life cycle of Schistosoma mansoni. 1–23 (2018).
1139
114030. Reid, A. J. et al. Single-cell RNA-seq reveals hidden transcriptional variation in malaria parasites. _Elife_ 7, 1–29 (2018).
114231. Ngara, M. et al. Exploring parasite heterogeneity using single-cell RNA-seq reveals a gene signature among sexual stage Plasmodium falciparum parasites. _Exp. Cell Res._ (2018). doi:10.1016/j.yexcr.2018.08.003
114532. Poran, A. et al. Single-cell RNA sequencing reveals a signature of sexual commitment in malaria parasites. _Nature_ 551, 95–99 (2017).
114733. Howick, V. M. et al. The Malaria Cell Atlas: Single parasite transcriptomes across the complete Plasmodium life cycle. _Science (80-. )._ 365, eaaw2619 (2019).
114934. Svensson, V. et al. Power analysis of single-cell RNA-sequencing experiments. _Nat. Methods_ 14, 381–387 (2017).
115135. Ziegenhain, C. et al. Comparative Analysis of Single-Cell RNA Sequencing Methods. _Mol. Cell_ (2017). doi:10.1016/j.molcel.2017.01.023
115336. Reid, A. J. Large, rapidly evolving gene families are at the forefront of host-parasite interactions in Apicomplexa. _Parasitology_ (2014). doi:10.1017/S0031182014001528
115537. Lönnberg, T. et al. Single-cell RNA-seq and computational analysis using temporal mixture modeling resolves T H 1/T H fate bifurcation in malaria. _Sci. Immunol._ 2, eaal2192 (2017).
38. Ziegenhain, C. et al. Comparative Analysis of Single-Cell RNA Sequencing Methods. Mol. Cell 65, 631-643.e4 (2017).
39. La Manno, G. et al. RNA velocity of single cells. Nature (2018). doi:10.1038/s41586-018-0414-6
40. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. (2018). doi:10.1186/s13059-017-1382-0
41. Gajria, B. et al. ToxoDB: An integrated toxoplasma gondii database resource. Nucleic Acids Res. 36, 553–556 (2008).
42. Beraki, T. et al. Divergent kinase regulates membrane ultrastructure of the Toxoplasma parasitophorous vacuole. Proc. Natl. Acad. Sci. (2019). doi:10.1073/pnas.1816161116
43. Camejo, A. et al. Identification of three novel Toxoplasma gondii rhoptry proteins. Int. J. Parasitol. (2014). doi:10.1016/j.ijpara.2013.08.002
44. Jones, N. G., Wang, Q. & Sibley, L. D. Secreted protein kinases regulate cyst burden during chronic toxoplasmosis. Cell. Microbiol. (2017). doi:10.1111/cmi.12651
45. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. J. Open Source Softw. (2018). doi:10.21105/joss.00861
46. Ramakrishnan, C. et al. An experimental genetically attenuated live vaccine to prevent transmission of Toxoplasma gondii by cats. Sci. Rep. (2019). doi:10.1038/s41598-018-37671-8
47. Hie, B., Bryson, B. & Berger, B. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. Nat. Biotechnol. 37, 685–691 (2019).
48. Kim, S.-K. & Boothroyd, J. C. The surface of Toxoplasma tachyzoites is dominated by a family of glycosylphosphatidylinositol-anchored antigens related to SAG1. Infect. Immun. 66, 2237–2244 (1998).
49. Kim, S. K., Karasov, A. & Boothroyd, J. C. Bradyzoite-specific surface antigen SRS9 plays a role in maintaining Toxoplasma gondii persistence in the brain and in host control of parasite replication in the intestine. Infect. Immun. 75, 1626–1634 (2007).
50. Development in the Mosquito and Infection of the Mammalian Host. Annu. Rev. Microbiol. 46, 564–574 (2011).
51. Cowman, A. F., Healer, J., Marapana, D. & Marsh, K. Malaria: Biology and Disease. Cell 167, 610–624 (2016).
120959. Hie, B. L., Bryson, B. & Berger, B. Panoramic stitching of heterogeneous single-cell transcriptomic data. bioRxiv (2018). doi:10.1101/371179
121160. Weiss, L. M., Ma, Y. F., Takvorian, P. M., Tanowitz, H. B. & Wittner, M. Bradyzoite development in Toxoplasma gondii and the hsp70 stress response. Infect. Immun. (1998).
121361. Litzenburger, U. M. et al. Single-cell epigenomic variability reveals functional cancer heterogeneity. Genome Biol. 18, 1–12 (2017).
121562. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. Science (80-. ). 297, 1183–1186 (2002).
121763. Butler, A., Hoffman, P., Smiberg, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Physiol. Behav. 176, 139–148 (2017).
122064. Ding, H., Blair, A., Yang, Y. & Stuart, J. M. Biological process activity transformation of single cell gene expression for cross-species alignment. Nat. Commun. 10, 4899 (2019).
122265. WEISS, L. M. et al. A Cell Culture System for Study of the Development of Toxoplasma gondii Bradyzoites. J. Eukaryot. Microbiol. (1995). doi:10.1111/j.1550-7408.1995.tb01556.x
122566. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. (2014). doi:10.1038/nprot.2014.006
122767. Panas, M. W., Naor, A., Cygan, A. M. & Boothroyd, J. C. Toxoplasma controls host cyclin E expression through the use of a novel myr1-dependent effector protein, HCE1. mBio 10, (2019).
123068. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods 25, 402–408 (2001).
123269. Tarashansky, A. J., Xue, Y., Li, P., Quake, S. R. & Wang, B. Self-assembling manifolds in single-cell RNA sequencing data. Elife 8, 1–29 (2019).
123470. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep. 9, 1–12 (2019).