nature portfolio

Peer Review File

Single cell expression and chromatin accessibility of the Toxoplasma gondii lytic cycle identifies AP2XII-8 as an essential ribosome regulon driver

Corresponding Author: Dr Marc-Jan Gubbels

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The comprehensive study of the lytic cycle of Toxoplasma gondii, in particular the detailed insights into the tachyzoite phase, is both innovative and valuable to the field. The combination of studying changes in gene activity over time with advanced methods such as single-cell RNA sequencing (scRNA-seq) and ATAC-seq is impressive. It creates a solid foundation to better understand how chromatin packaging influences gene activity in this organism. In addition, the functional analysis of AP2XII-8, which reveals its essential role in the expression of a ribosomal regulon, is an important contribution to our understanding of how this transcription factor family influences gene expression in Toxoplasma. The discovery of the two DNA motifs within larger ribosomal protein motifs, suggesting a cooperative mechanism of gene regulation, is particularly intriguing. I am very impressed by the depth and clarity of the research presented. The manuscript is remarkably thorough and clear. My only reservation, however, is that the authors appear to have overlooked some of their data (see my comments below), and I believe this work would benefit from a more comprehensive discussion.

Major comments:

1) According to Sup. Table 4, AP2XII-8 binds near 13 AP2 TFs, including AP2VIIa-2, AP2XII-1, AP2XI-3 and others, and surprisingly was also found near its own gene, which was not emphasized by the authors. This suggests a possible feedback mechanism in the regulation of its own gene. It would be interesting if the authors could comment on this self-binding of AP2XII-8 or provide initial results. In addition, the sheer number of AP2s potentially regulated by AP2XII-8 is remarkable. Could the authors provide an integrated view of how these AP2s might collectively contribute to the final transcriptional outcome?

14 AP2s that are regulated by AP2XII-8: AP2VIIa-2, AP2XII-1, AP2XI-3, AP2V-2, AP2X-3, AP2XII-4, AP2VIIb-3, AP2X-5, AP2VIII-7, AP2IX-9, AP2X-7, AP2VI-3, AP2XI-5 and AP2XII-8.

2) Among the 970 genes identified as targets of AP2XII-8 in Supplementary Table 4, both MORC (CRC230) and HDAC3 are included, with indications suggesting an up-regulation in their expression. Remarkably, the HDAC3 gene is associated with a DNA motif 1 in its vicinity. This raises the interesting question of what influence the repressive HDAC3/MORC complex has on the transcriptional changes observed after depletion of AP2XII-8. Could the authors provide insights on how this complex might influence overall transcriptional outcomes?

3) The DEG analysis in combination with the ATAC-seq data shows an overlap of 95 genes with a representation bias that emphasizes genes encoding the RP regulon. While this analysis is robust and convincing, it overlooks the 518 genes that also vary in AP2XII-8 knockdown (KD) but are characterized by the absence of AP2XII-8 binding in their vicinity. In addition, 150 AP2XII-8 binding sites are transcriptionally inactive. The authors hypothesized that AP2XII-8 acts exclusively in cis on genes near its binding site (which could be true for RP genes), but the possibility that this factor could also act in trans by directing chromosome arms to transcription factories should not be ruled out. Another possible hypothesis, inspired by Antunes et al. (2024), is that AP2XII-8 could be homodimerized on these 95 genes and its depletion directly affects the transcriptional activities of the neighboring genes. At the other 150 binding sites, it could be associated with a different AP2, and transcriptional effects could only be observed when both are depleted. Could the authors comment on these hypotheses?

Minor comments:

• In Figure 1f, the authors describe three phenotypes: 1C0B (1 centrosome, 0 buds, indicating G1 phase), 2C0B (2 centrosomes, 0 buds, indicating S phase), and 2C2B (2 centrosomes, 2 buds, indicating M/C phase). They used α-Centrin to identify centrosomes, yet the image accompanying the right side of the graph only displays IMC and DNA. It would be more consistent with the quantification graph if the image also depicted α-Centrin to visually align with the described phenotypes.

• References 13 and 14, as the initial studies describing histone PTM and chromatin accessibility in bulk replicating tachyzoites, were constrained by either limited resolution or an inability to fully capture the complexity of the histone code. It would also be relevant to mention the work of Sindikubwabo et al. (eLife, 2017), Farhat et al. (Nature Microbiology, 2020), and Antunes et al. (Nature, 2024), which complement these two aforementioned references.

• Although references 12, 22, 33, and 34 are relevant, it is advisable for the authors to also include a citation of Antunes et al. (Nature, 2024) in their discussion on the capacity of AP2s to form either homo- or hetero-dimers. This is important because the study by Antunes et al. provides concrete evidence of two T. gondii AP2 proteins forming a heterodimer, demonstrated through both in vivo and in vitro experiments.

• The discussion in the paper does not comprehensively address the intricacies of the histone code and chromatin signaling, particularly in the context of genes that are tightly regulated during the cell cycle. Important concepts such as the role of bimodal chromatin in establishing a transcriptionally poised state and the potential influence of chromatin topology on transcription factories are not thoroughly explored. A more detailed consideration of these elements is crucial for a more comprehensive understanding of the underlying regulatory mechanisms.

Reviewer #2

(Remarks to the Author)
Introduction:

This manuscript investigates the intricate dynamics of transcriptional regulation in the asexual cell cycle of the apicomplexan parasite T. gondii. The study addresses a significant gap in understanding the relationship between chromatin accessibility and transcription factor binding, overcoming the challenges posed by T. gondii's rapid, asynchronous, and atypical replication cycle. The use of advanced single-cell transcriptomic and chromatin accessibility approaches enabled a detailed exploration of gene expression patterns and chromatin states, and the identification of functionally related gene sets relevant for transitioning from one cell cycle phase to another.
The authors find a strong correlation between chromatin opening and gene expression allowing for transcriptional bursts of four distinct RNA expression and chromatin accessibility clusters throughout the Toxoplasma cell cycle. Motif analysis of co-accessible promoter regions identified several known AP2 DNA binding sites but also motifs without any known interaction partners. Based on RNA velocity analysis, the authors identify AP2XII-8 as an important driver for the parasite's progression through the G1-S phase. The functional role of AP2XII-8 was explored using CUT&RUN, highlighting its involvement in the regulation of a ribosomal regulon through two DNA motifs.
Thus, this work has generated a comprehensive resource for the field as it not only underscores the role of AP2XII-8 in the G1-S phase transition but also sheds light on potential gene sets crucial for other processes throughout the Toxoplasma cell cycle, such as karyokinesis. However, the ambiguity of the knockdown of AP2XII-8, as shown in Figure 4d, raises questions about the interpretations derived from all short-term AP2XII-8 knockdown experiments conducted in this study.

Major comments:

Line 257 and line 838, Figure 4:
- The authors claim that AP2XII-8 can be depleted using IAA in a period as brief as 2 hours. However, the western blot presented in Figure 4d does not sufficiently support this claim, as the timepoint 0 contains numerous non-specific bands, complicating the clear identification of AP2XII-8. The presence of these non-specific bands is concerning and may suggest additional problems with the used strain. Moreover, an appropriate negative control should include vehicle-treated parasites in addition to parental control. Verifying the knockdown kinetics of AP2XII-8 is crucial for accurately interpret the various knockdown experiments conducted with this strain.

Line 261 and line 838, Figure 4:
- The observation in Figure 4f that the depletion of AP2XII-8 leads to an accumulation of parasites in the G1 stage does not conclusively establish AP2XII-8 as a critical factor for the G1 transcriptional burst. Notably, even after 24 hours of AP2XII-8 knockdown, several parasites are still in the S-phase or M/C phase, indicating that they can progress through the G1 stage without AP2XII-8. To show that AP2XII-8 is essential for the G1 transcriptional burst would require transcriptional profiling and velocity analysis (similar to Figure 4a) under AP2XII-8 knockdown conditions. Alternatively, conducting a more extended period of knockdown might also be informative.

Minor comments:

Line 66:
- AP2 actually refers to the Apetala2/ethylene response factor transcription factor family.

- The phrase “the key driving the transcriptional waves” may lead to potential misunderstandings of the cited study. Referring to AP2s as “the key” may imply that they are the sole drivers of transcriptional waves in T. gondii. Additionally, the phrase “the transcriptional waves” should be clarified as it is unclear what they are referring to in this context, especially since the cited study solely investigated Plasmodium falciparum blood-stage development.

Line 88:
- The use of “RNA explosion” in this context is potentially confusing as it might be interpreted as a term denoting destruction. A more appropriate and descriptive term, such as "transcriptional burst", would be a better description of the observation, and has been used in the literature to derscribe similar phenomena.

Line 147:
- It is unclear what “the expression level regulation is more granular” is supposed to describe in this context.

Line 156:
- The reference connecting the identified GO terms to the G1 phase of the cell cycle is missing.

Line 166 and line 821, Figure 3:
- The authors state that they “differentiated three different correlations between expression and chromatin accessibility”. However, the thresholds used for the three categories (loose, tight, and multi-functional) are not defined. Furthermore, it is unclear whether these gene sets are mutually exclusive since they represent different categories (some are cell cycle stages, others are protein localization). This should be specified in the figure legend or the text.

Line 169:
- The authors state that “when the whole curve shape is considered (CCS), several genes in each group appear below the 0.6 cut-off, confirming this ‘loose’ pattern”. However, in Figure 3b, the data show only one gene below the 0.6 cut-off in the loose S-phase set and two in the loose M-phase set. Contrarily, the majority of genes in both sets have a CCS score above 0.6. This observation appears to contradict the statement in the text, suggesting a need for correction or clarification in the manuscript.

Line 187:
- The reference supporting the statement that transcriptional waves are AP2-driven is missing.

Line 209 and line 821, Figure 3:
- The authors suggest that the data presented in Figure 3e demonstrates cascading due to overlapping functional modules. Yet, it appears that there is only a single instance of overlapping GO terms in subsequent scRNAseq clusters aligning with the author’s interpretation. The arrangement of GO terms appears to be arbitrary, which raises the question of whether evidence of cascading would still be apparent if the GO terms were organized differently.

Line 220 and 304:
- The discovered motif (T/C)GCATGC(G/A) has not only been previously reported in references 21 and 30 but also in Markus et al. (PMID: 33553008), which identified this motif associated with 44 % of TSS in Tachyzoites. This publication should be cited and discussed.

Line 259 and line 838, Figure 4:
- The tagged strain exhibits a noticeable growth defect in the -IAA plaque assay. This point warrants discussion and acknowledgment as a potential caveat for the subsequent analysis. In this context, it is puzzling that there is no observable difference between the parental strain and the AP2XII-8-mAID-5xTy strain under -IAA conditions in Figure 4f, a discrepancy that requires discussion.

Line 269 and line 996, Extended Figure 7:
- The term “the highest detection power” used in the manuscript is ambiguous. Detection power is generally understood as the capacity to identify a true effect. Therefore, this claim should be substantiated with an evaluation of peak quality, not merely based on the quantity of peaks identified. To assess the CUT&RUN protocol with the highest detection power, it would be beneficial to consider metrics such as signal-to-noise ratio or a comparable measure to evaluate the quality of the identified peaks.

Line 270 and 628:
- Despite the R code for CUT&RUN data analysis having been made available by the authors, key parameters that were used to transform the data should be stated in the methods section. These parameters, such as the specific distance criteria used for merging adjacent peaks, should be explicitly stated to enhance clarity.

Line 315:
- The detection of additional motifs is not sufficient to “investigate cooperative binding”. Determining the binding mode requires biochemical assays. Therefore, this sentence should be rephrased.

Line 639:
- The provided link (https://bammmotif.soedinglab.org/) is not reachable. Access: 01/16/2024 – 5:20 PM

Line 804, Figure 2:

- Panel a and c: The methodology for selecting "transition points" is somewhat ambiguous. It appears that the authors identified these points by looking for peaks and troughs in the first derivative of the graph, essentially points where the slope of the curve approaches zero. However, it is noticeable that not all of these points were selected. In the discussion, the authors mention that the “more fluid” points were excluded from being designated as transition points, but the threshold for this decision remains unclear.
- Panel e: It should be commented on why the accessibility cluster TA4 initiates after the onset of expression cluster TE4. This observation raises questions: Are these genes expressed from heterochromatin, or might this discrepancy be attributed to issues of sensitivity in the experimental methods used?

Line 821, Figure 3:
- Panel a and b: In the legend, it is specified that “box plots represent 75 percentile; dots indicate outliers”. Yet, numerous dots are positioned on the median line, contradicting their designation as outliers.
- Panel d: Similar to observations in Figure 2, there are instances where RNA levels peak even though chromatin accessibility is at its lowest. These occurrences warrant further discussion.
- Panel e: The identified DNA motifs are missing critical details necessary for evaluating their quality, including q-values and the count of genes within each cluster associated with the respective motif. Additionally, the methods section should clarify which dataset was utilized as a background for this analysis.

Line 838, Figure 4:
- Panel e: The labeling of the AP2XII-8-mAID-5xTy strain as AP2XII-8 cKD is confusing as AP2XII-8 is not expected to be knocked down in the absence of IAA. This labeling should be revised for greater clarity.

Line 859, Figure 5 and 996, Extended Figure 7:
- Figure 5a and Extended Figure 7d: The significance of the color gradient used in the figure is not specified. This should be clarified in the figure legend. Additionally, the x-axis is inaccurately labeled as “gene distance (bp)” when, in fact, the distance is depicted in kilobases (kb). This labeling error needs to be corrected.

Line 872, Figure 6:
- Panel a: It appears that AP2XII-8 KD results in a higher proportion of parasites in the S-phase. This appears contradictory to what is depicted in Figure 4f, where AP2XII-8 KD seems to increase the proportion of parasites in the G1 phase. How can such a discrepancy be explained?
- Panel d and e: Similar to Figure 3e, the identified DNA motifs are missing critical details necessary for evaluating their quality. How many up-regulated genes contain the motif and how many CUT&RUN peaks contain the motif? Additionally, the methods section should specify the gene set used as a background for the motif discovery process.
- Panel f: The title “ribosomal genes scRNA-seq” is confusing as the plot specifically refers to the AP2XII-8 down-regulated ribosomal genes as indicated in the figure legend. Therefore, the plot's title should be revised to accurately reflect this specificity.

Line 945, Extended Figure 2 and line 961, Extended Figure 3:
- In both figures, the current color scheme leads to multiple genes being represented by apparently identical colors. To help enhance the distinction between genes, it would be beneficial to modify and diversify the color palette used.

Line 996, Extended Figure 7:
- Panel e and f: The dataset labels are somewhat challenging to understand. Consistency in labeling, similar to what is seen in panel c, would be beneficial for clarity. For example, labels could be "AP2XII-8 BB2 vs. AP2XII-8 IgG1" and "AP2XII-8 BB2 vs. RHΔKU80 IgG1". The fact that only 25% of identified target genes overlap depending on the chosen negative control raises concerns and warrants discussion, especially in light of the growth defect observed in the AP2XII-8-mAID-5xTy strain under -IAA conditions.

Reviewer #3

(Remarks to the Author)
The manuscript by Lou and colleagues title “Single cell expression and chromatin accessibility of the Toxoplasma gondii lytic cycle identifies AP2XII-8 as an essential ribosome regulon driver” analyze the T. gondii transcriptome and chromatin accessibility at single cell level. The team clustered their data and used a pseudo-time trajectory analysis to identify four transcriptional bursts corresponding to the major cell cycle transition. The control of the dominant G1 burst was further analyzed by the functional analysis of AP2XII-8, a transcription factor thought to drive the identified burst. The team then used a combination of molecular, cellular and genome wide (CUT&RUN) experiments to identify the AP2XII’s target genes and concluded that AP2XII-8 is controlling the expression of the ribosome genes. To facilitate access to the datasets, an interactive web-app is provided where users can perform various analyses, including clustering and co-expression analysis.

Several studies have already analyzed the transcriptome and chromatome of several apicomplexan parasites using a combination of genome-wide tools including single-cell RNA and Transposase-Accessible Chromatin (ATAC) sequencing

in T. gondii. Results of these studies have demonstrated that chromatin opening facilitates mechanism underlying gene expression. The present study validates previous published results by analyzing simultaneously the expression and the chromatin accessibility of T. Gondii during its cell cycle progression. While the transcriptome and chromatin state data sets generated tend to have a low resolution, results presented are robust. The findings are however not entirely novel, nor surprising. I am nevertheless puzzled by the analysis and the interpretation of the data generated for the identification of AP2XII-8 target genes (the novelty of the manuscript) using CUT&RUN and scRNA-seq experiments. Please find below my major comments. Some of them will need to be fully addressed before this manuscript can be published.

Major comments:

1 - The Cleavage Under Targets and Release Using Nuclease (CUT&RUN) experiment has been developed in the epigenetic field to increase sensibility of the traditional ChiP-seq experiment. While ChiP-seq experiments require in general a few millions of cells (1 to 10 millions of cells in higher eukaryotes – 5 to 40 millions of cells in the T. gondii genome), the CUT&RUN protocol require less than 500,000 to be accurate. Exceeding 500,000 cells per sample (1 million for smaller eukaryotes) is not recommended. Generally, a good starting point to optimize this protocol is usually done with 50,000 cells. Some experiments were even successful with less than 200 cells. Using too many cells in such protocol will increase the background, reduce the yield, reduce the specificity and, more importantly, reduce the complexity of the library. The team optimized their protocol and determined that 200 millions of parasites had the highest detection power. This seems extremely high and is the exact opposite of what most people in the field have experienced. I am therefore wondering how the results generated using CUT&RUN were validated. I think it will be best to confirm the result using a classical ChIP-seq experiment with a maximum of 10-40 millions parasites.

2 - Data analysis of the CUT&RUN experiment is even more puzzling. The team identified over 2,338 binding events (peaks) that were assigned to a list of 970 targeted genes. This list was then narrowed down to 246 genes after background subtraction using signals detected in the negative controls. From this 246 genes only 95 exhibited a change in their expression profiles in a AP2XII-8 knock down experiment with 73 down regulated, 19 up regulated and 3 either down- or up-regulated in different phases. These results make no sense. Why less than 10% of the targeted genes are affected by AP2XII-8 knock down? Overall, the CUT&RUN results seem too noisy, and it is very likely that data generated (both in the positive and negative controls) are inaccurate due to the large number of parasites used in the experiment. Considering that the team needs 200 millions parasites for their CUT&RUN protocol to work, I will suggest a classical ChIP-seq experiment that if done well - with less than 20 million parasites - should generate accurate results with almost no peaks detected in the negative controls therefore increasing the specificity of the ChIP-seq signals in the samples analyzed.

3 - Also Considering that the genes coding for the ribosomes are the most highly expressed in the genome of all eukaryotes, they are usually the most easily-detectable by single cell sequencing experiment. it is very likely that AP2XII-8 mediate more than just those genes during the cell cycle transition. The resolution of the data presented remains low and I will be careful with the interpretation of the data.

4 - Can you specify in the text how many cells and genes were identified in each of the clusters. Looking at the figures presented, the number of genes detected in each cluster seems low. To accurately evaluate the resolution of the data, it will be important to add a few plots that exhibit the distributions of the unique molecular identifiers (UMIs) as well as the distribution of the genes detected in each cell and cluster for both scRNA-seq and scATAC-seq samples. Also, for the remaining genes that are not detected as DEG, what is the profile of their chromatin state? I don't see a representation of those genes in Figure 1.

Minor comments:

1-In line 110-111, please clarify the 20 time points mentioned in the manuscript. Are these 20 time points evenly distributed within pseudo time (i.e. 4 time points for each of the 5 cell cycle stages)? Perhaps it isn't pertinent to this point, but figure 1f appears to show 18 distinct pseudo-time points.

2 - In line 121-123 - Do you mean that of the 1,620 genes that are differentially expressed, 1,238 are unique to a specific cell cycle stage? This sentence is a bit confusing.

3 - In line 139-142 - Is there any relationship between the genes confined within each transition state for expression and accessibility, or are the 1,238 genes clustered independently within each transition state? In other words, are the same cyclic genes found within TE1 and TA1? In the text, the data are described as generally highly correlated, but it will be good to give an exact number of genes that correlate and those that don't.

4 - In line 277-279 - Please clarify how the comparisons were performed. I expressed some of my concerns already regarding the CUT&RUN experiment and the way the data were analyzed. I am still quite confused on how the data were normalized and analyzed. How it currently looks is that any genes found within each negative control gene list were first subtracted from the list of AP2XII-8 targeted genes, generating four separate lists of genes. Then the resulting lists of genes were compared, and the overlap identified as shown in t≤he Venn diagram. On the mean intensity plot (Extended Data Fig. 7e) it appears that the 246 genes have an intensity ranging from 0 to 25, with a large proportion of those genes being below 5-fold mean intensity. If the 246 genes only include those found in AP2XII-8 and none of the controls, thus the area only filled with blue (most of which is above 5-fold mean intensity) and not the red/blue overlap, then perhaps a different color is needed within the overlap. Furthermore, the usage of "intersection" in the intensity plot is somewhat confusing because

some could understand it to mean the intersection of AP2XII-8 genes and those in the controls, rather than those only found in AP2XII-8. This CUT&RUN experiment is terribly confusing and will need to be validated using a more standard approach.

Author Rebuttal letter:

Response to Reviewer comments

Reviewer #1 (Remarks to the Author):

The comprehensive study of the lytic cycle of Toxoplasma gondii, in particular the detailed insights into the tachyzoite phase, is both innovative and valuable to the field. The combination of studying changes in gene activity over time with advanced methods such as single-cell RNA sequencing (scRNA-seq) and ATAC-seq is impressive. It creates a solid foundation to better understand how chromatin packaging influences gene activity in this organism. In addition, the functional analysis of AP2XII-8, which reveals its essential role in the expression of a ribosomal regulon, is an important contribution to our understanding of how this transcription factor family influences gene expression in Toxoplasma. The discovery of the two DNA motifs within larger ribosomal protein motifs, suggesting a cooperative mechanism of gene regulation, is particularly intriguing. I am very impressed by the depth and clarity of the research presented. The manuscript is remarkably thorough and clear. My only reservation, however, is that the authors appear to have overlooked some of their data (see my comments below), and I believe this work would benefit from a more comprehensive discussion.

Major comments:

1) According to Sup. Table 4, AP2XII-8 binds near 13 AP2 TFs, including AP2VIIa-2, AP2XII-1, AP2XI-3 and others, and surprisingly was also found near its own gene, which was not emphasized by the authors. This suggests a possible feedback mechanism in the regulation of its own gene. It would be interesting if the authors could comment on this self-binding of AP2XII-8 or provide initial results. In addition, the sheer number of AP2s potentially regulated by AP2XII-8 is remarkable. Could the authors provide an integrated view of how these AP2s might collectively contribute to the final transcriptional outcome?

14 AP2s that are regulated by AP2XII-8: AP2VIIa-2, AP2XII-1, AP2XI-3, AP2V-2, AP2X-3, AP2XII-4, AP2VIIb-3, AP2X-5, AP2VIII-7, AP2IX-9, AP2X-7, AP2VI-3, AP2XI-5 and AP2XII-8.

Thank you for this feedback. Indeed, AP2XII-8 binds its own promoter, and upon KD of its protein, we do see an interesting change in the RNA expression of AP2XII-8 (see Figure below; new panel Fig 6d), indicating the presence of a feedback loop. There may be other factors in play in regulating AP2XII-8 and further studies are needed to decipher the exact mechanisms and map other factors. As such we did not emphasize this in the text. However, we have now added a sentence regarding this observation in the discussion.

Regarding the other ApiAP2 factors, in re-visiting the CUT&RUN data (please see comments from Reviewer 3), we realized that the control we originally named RH-neg, was in fact RH-IgG1, but with 500 million parasites and 0.5 Âµg IgG1 antibody. We removed this control from our analysis and updated all figures and tables accordingly.
As such, for other AP2s with AP2XII-8 CUT&RUN signals, while it is the case that a total of 14 AP2s were detected, depending on the negative control used, the signals were not statistically significant in all cases. After this stringent cutoff, only five AP2s (AP2XII-1, AP2XI-3, AP2XII-4, AP2XII-8, and AP2VIII-7) passed the filter. Moreover, for additional rigor, we only considered AP2s whose expression significantly changes upon AP2XII-8 depletion. Out of the five AP2s, only the expressions of AP2VIII-7 and AP2XII-8 are affected by the loss of AP2XII-8. AP2VIII-7 is clearly upregulated upon loss of AP2XII-8, making AP2XII-8 a potential repressor of AP2VIII-7. On the other hand, the loss of AP2XII-8 resulted in inversion of its mRNA expression profile, suggesting the presence of a feedback loop (see new Fig. 6d AP2XII-8 profile shift: dotted (+IAA) vs solid lines (-IAA)).

Taken together, this implies that some AP2 targets could be false positive detections, rendering the importance of good negative controls that were carried out in this manuscript.

Of the other AP2 factors modulated upon AP2XII-8 depletion for which there is no engagement signal by AP2XII-8 in the CUT&RUN data, we observe: 1) downregulated target AP2IX-10, which is a stress

factor (Primo et al mSystems, 2021; PMID: 34874774); 2) and upregulated AP2 targets IV-5, VIIa-3, and XII-9. All these upregulated AP2 factors peak in S-phase in unperturbed parasites (Supplementary Fig. 3). This is consistent with the scRNA-seq 3D UMAP, where the AP2XII-8-cKD profile displays a modified S-phase outcrop (Fig. 6a). However, on the cell biological level, these parasites arrest with a single centrosome (Fig. 4f) and do not progress into S-phase. From this, we can conclude that depletion of AP2XII-8 disrupts G1-S progression on the cell biological level, but that the transcriptional profile indicates progression into S-phase, and thus hints at factors in parallel of AP2XII-8 driving G1-progression at the transcriptional level. Indeed, we mapped such a putative factor with a severe fitness score, AP2X-7, which displays the same expression profile as AP2XII-8.

We have added a short section in the discussion to include comments on how these ApiAP2 factor data can be interpreted in terms of the cell cycle arrest, UMAP arrest and AP2 regulatory network.

2) Among the 970 genes identified as targets of AP2XII-8 in Supplementary Table 5, both MORC (CRC230) and HDAC3 are included, with indications suggesting an up-regulation in their expression. Remarkably, the HDAC3 gene is associated with a DNA motif 1 in its vicinity. This raises the interesting question of what influence the repressive HDAC3/MORC complex has on the transcriptional changes observed after depletion of AP2XII-8. Could the authors provide insights on how this complex might influence overall transcriptional outcomes?

Similar to AP2 factors mentioned above, in the original analysis MORC and HDAC3 were indeed assigned to at least one CUT&RUN peak using one of the negative backgrounds, but these two genes did not make the final cut using the stringent CUT&RUN signal selection criteria (CUT&RUN signal over all negatives). Our re-analysis of the CUT&RUN data using 3 controls shows that MORC is now engaged by AP2XII-8, whereas HDAC3 is still below the threshold. Both are indeed upregulated upon AP2XII-8 depletion (peak at G1b-S; see figure below; new Fig. 6i). Taken together, these data indicate that at least AP2XII-8 is a repressor of MORC, and possibly of HDAC3. Given the absence of stringent CUT&RUN signal we did not include HDAC3 in the interpretation, but have revised the text to report that MORC is under AP2XII-8 control.

3) The DEG analysis in combination with the ATAC-seq data shows an overlap of 95 genes with a representation bias that emphasizes genes encoding the RP regulon. While this analysis is robust and convincing, it overlooks the 518 genes that also vary in AP2XII-8 knockdown (KD) but are characterized by the absence of AP2XII-8 binding in their vicinity. In addition, 150 AP2XII-8 binding sites are transcriptionally inactive. The authors hypothesized that AP2XII-8 acts exclusively in cis on genes near its binding site (which could be true for RP genes), but the possibility that this factor could also act in trans by directing chromosome arms to transcription factories should not be ruled out. Another possible hypothesis, inspired by Antunes et al. (2024), is that AP2XII-8 could be homodimerized on these 95 genes and its depletion directly affects the transcriptional activities of the neighboring genes. At the other 150 binding sites, it could be associated with a different AP2, and transcriptional effects could only be observed when both are depleted. Could the authors comment on these hypotheses?

First, we would like to mention that after our CUT&RUN data re-analysis (illustrated above), now there are a total of 223 transcriptionally inactive sites by AP2XII-8. We agree with the reviewer that the 223 genes engaged by AP2XII-8 in their promoters for which no transcriptional regulation is observed must represent a different scenario. We have added the following comment in the discussion highlighting these scenarios: âThis could indicate that genes are co-regulated by additional TgAP2 factors, either in (an alternative) heterodimeric complex or as an additional, dominantly acting TgAP2 in the promoter of these genesâ.

Minor comments:

â¢ In Figure 1f, the authors describe three phenotypes: 1C0B (1 centrosome, 0 buds, indicating G1 phase), 2C0B (2 centrosomes, 0 buds, indicating S phase), and 2C2B (2 centrosomes, 2 buds, indicating M/C phase). They used Î±-Centrin to identify centrosomes, yet the image accompanying the right side of the graph only displays IMC and DNA. It would be more consistent with the quantification graph if the image also depicted Î±-Centrin to visually align with the described phenotypes.

We regret that the centrosomes in the images were not as clear as they should have been. We have replaced the 3 panels with images clearly displaying the centrin signal.

â¢ References 13 and 14, as the initial studies describing histone PTM and chromatin accessibility in bulk replicating tachyzoites, were constrained by either limited resolution or an inability to fully capture the complexity of the histone code. It would also be relevant to mention the work of Sindikubwabo et al. (eLife, 2017), Farhat et al. (Nature Microbiology, 2020), and Antunes et al. (Nature, 2024), which complement these two aforementioned references.

We thank the reviewer for pointing out these omissions; we have added them in the revision.

â¢ Although references 12, 22, 33, and 34 are relevant, it is advisable for the authors to also include a citation of Antunes et al. (Nature, 2024) in their discussion on the capacity of AP2s to form either homo- or hetero-dimers. This is important because the study by Antunes et al. provides concrete evidence of two T. gondii AP2 proteins forming a heterodimer, demonstrated through both in vivo and in vitro experiments.

Excellent point, and we have added the Antunes reference here.

â¢ The discussion in the paper does not comprehensively address the intricacies of the histone code and chromatin signaling, particularly in the context of genes that are tightly regulated during the cell cycle. Important concepts such as the role of bimodal chromatin in establishing a transcriptionally poised state and the potential influence of chromatin topology on transcription factories are not thoroughly explored. A more detailed consideration of these elements is crucial for a more comprehensive understanding of the underlying regulatory mechanisms.

The reviewer makes a valid point. Due to length limitation, we had not elaborated in this direction. In the revision we have added this broader context of the presented data in the same section. Please also see response to major comments #1 and #3, where we present additional data supporting this point.

Reviewer #2 (Remarks to the Author):

Introduction:

This manuscript investigates the intricate dynamics of transcriptional regulation in the asexual cell cycle of the apicomplexan parasite T. gondii. The study addresses a significant gap in understanding the relationship between chromatin accessibility and transcription factor binding, overcoming the challenges posed by T. gondiiâs rapid, asynchronous, and atypical replication cycle. The use of advanced single-cell transcriptomic and chromatin accessibility approaches enabled a detailed exploration of gene expression patterns and chromatin states, and the identification of functionally related gene sets relevant for transitioning from one cell cycle phase to another.
The authors find a strong correlation between chromatin opening and gene expression allowing for transcriptional bursts of four distinct RNA expression and chromatin accessibility clusters throughout the Toxoplasma cell cycle. Motif analysis of co-accessible promoter regions identified several known AP2 DNA binding sites but also motifs without any known interaction partners. Based on RNA velocity analysis, the authors identify AP2XII-8 as an important driver for the parasiteâs progression through the G1-S phase. The functional role of AP2XII-8 was explored using CUT&RUN, highlighting its involvement in the regulation of a ribosomal regulon through two DNA motifs.
Thus, this work has generated a comprehensive resource for the field as it not only underscores the role of AP2XII-8 in the G1-S phase transition but also sheds light on potential gene sets crucial for other processes throughout the Toxoplasma cell cycle, such as karyokinesis. However, the ambiguity of the knockdown of AP2XII-8, as shown in Figure 4d, raises questions about the interpretations derived from all short-term AP2XII-8 knockdown experiments conducted in this study.

Major comments:

Line 257 and line 838, Figure 4:
- The authors claim that AP2XII-8 can be depleted using IAA in a period as brief as 2 hours. However, the western blot presented in Figure 4d does not sufficiently support this claim, as the timepoint 0 contains numerous non-specific bands, complicating the clear identification of AP2XII-8. The presence of these non-specific bands is concerning and may suggest additional problems with the used strain. Moreover, an appropriate negative control should include vehicle-treated parasites in addition to parental control. Verifying the knockdown kinetics of AP2XII-8 is crucial for accurately interpret the various knockdown experiments conducted with this strain.

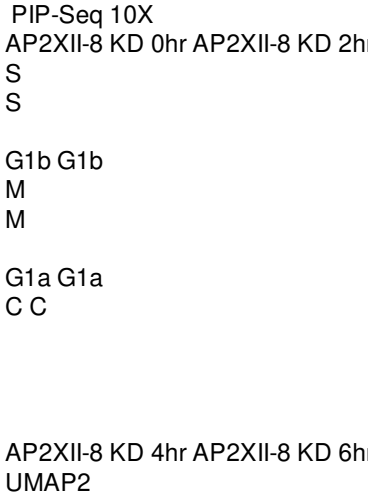
We agree with the reviewer that the multiple bands raise a concern. We did repeat this experiment several times, even varying the experimenter. We performed a western blot (WB) on a parallel strain wherein we tagged AP2XII-8 with a TurboID-5xV5-epitope tag, without a mAID tag, which resulted in a similar pattern (see below, left):

Moreover, we repeated the Ty WB experiment including the requested controls, and in addition included more extensive protease and kinase inhibitors while performing parasite harvest and lysis at 4Â°C to limit unspecific degradation (see above, right). Again, we obtained the same results. Furthermore, since we did not see a difference between untreated and vehicle control treated parasites, we kept the original panel in the same figure.

In addition, from all the published T. gondii AP2 WB results, we found that most figures were either cropped, or displayed similar banding ladder patterns, with some of them showing unspecific bands even with cropped representations, like AP2IX-5 by Khelifa et al., 2021 (PMID: 33414462), AP2XII-2 in RH strain by Srivastava et al., 2020 (PMID: 32938695), AP2XII-2 in ME49 strain by Srivastava et al., 2020 and Srivastava et al., 2023 (PMID: 32938695, PMID: 36786611), AP2XI-2 and AP2XII-1 by Antunes et al., 2024 (PMID: 38093015), AP2IX-7 and AP2X-8 by Wang et al., 2014 (PMID: 24391497), AP2XI-4 by Walker et al., 2014 (PMID: 23240624), AP2IX-4 by Huang et al., 2017 (PMID: 28317026). Taken together with our results, and similar banding ladder pattern observations with several other AP2 factors we recently tested (data not shown), we conclude that although we cannot provide a satisfying explanation at this point, the pattern persists and is consistent with other TgAP2 reports. It is possible that the ladder represents turn-over of the AP2 factor, but since we cannot synchronize the parasite cell cycle effectively, we cannot easily test this hypothesis. We would like to note that the nuclear localization is very clean (Fig. 4c) and that we do not see a cytoplasmic haze suggestive of lingering fragments.

Considering the level of knockdown after 2 hrs, we respectfully disagree with the reviewerâs assessment that knock-down is not sufficient as protein depletion only slightly increases at the 6 hr time point. Consequently, we initially opted for the 2-hour time point for scRNA-seq of the mutant, reasoning that earlier sampling would minimize potential secondary alterations in the transcriptional profile. In response to the subsequent comment, we performed additional scRNA-seq experiments (using Fluent-BioSciences PIP-seq rather than 10x Genomics as a more economical alternative) at 0, 2, 4 and 6 hrs (Figure below). The datasets presented in the manuscript were reprocessed along with the PIP-seq single-cell RNA sequencing data performed on the 4 time-points and visualized using UMAP projection. All the KD datasets (PIP-seq 2,4,6 hr and 10X 2hr) present a common group of cells that branch out at the end of G1b phase at the beginning of the S phase in similar proportions across all KD datasets, regardless of the treatment time. At 6 hrs, the bulge has slightly expanded, and some cells have now transitioned to show M-phase markers. We consider those secondary events, further justifying the early 2 hrs time point to capture the direct effects. Overall, this demonstrates that 2 hr is sufficient to capture the transcriptomic phenotype.

Remark: The PIP-seq technology is a relatively new platform for single-cell transcriptomics. The technology holds strong promise for capturing significantly higher cell numbers at a lower cost compared to competitors such as 10X. Unlike 10X, the full capacity and limitations of this technology have not been fully explored yet. We are currently investigating the properties of the PIP-seq data (e.g., signal-to-noise ratio, doublets, detection power, etc.) and have plans to systematically benchmark PIP-seq with other technologies. The data were used here to demonstrate the sufficiency of 2 hr KD only; however, the full exploration of the datasets is not within the scope of the current paper, and the raw data will be released in future studies.



UMAP1

Line 261 and line 838, Figure 4:
- The observation in Figure 4f that the depletion of AP2XII-8 leads to an accumulation of parasites in the G1 stage does not conclusively establish AP2XII-8 as a critical factor for the G1 transcriptional burst. Notably, even after 24 hours of AP2XII-8 knockdown, several parasites are still in the S-phase or M/C phase, indicating that they can progress through the G1 stage without AP2XII-8. To show that AP2XII-8 is essential for the G1 transcriptional burst would require transcriptional profiling and velocity analysis (similar to Figure 4a) under AP2XII-8 knockdown conditions. Alternatively, conducting a more extended period of knockdown might also be informative.

Indeed, there are still parasites in S- and M/C after 24 hrs, however their incidence is reduced by over 50% and there is a statistically significant accumulation of G1 stage parasites. This is a strong phenotype, and in line with other phenotypes, and even true checkpoints are almost universally leaky (e.g. Crk data from the White and Suvorova labs, as well as AP2IX-5 for instance by the Gissot lab, which drives daughter budding, but upon AP2IX-5 knock-down the percentage of parasites showing daughter buds only reduced from ~12 to ~5%). In fact, the field has been unable to exploit any of these factors to generate tightly synchronized cell cycle conditions, indicating all arrests are incomplete and/or leaky (or inherent to a very short 6-7 hr cell cycle).

In response to the second part of the comment, we added additional time points for the scRNA-seq analysis (please see response to previous point). Moreover, as suggested, we performed additional velocity analysis under the KD conditions (Figure below, which is added as new Fig. 6b). As shown, the burst signal is drastically dampened in the G1a transcriptional burst while there is added S noise and a heightened M phase burst (both due to the âbulgeâ), consistent with the 3D UMAP in Fig 6a. We would like to point out that we have not made a claim that AP2XII-8 is solely responsible for the burst in G1, but it certainly is part of it. We have added clarified this point in the âResultsâ text.
Minor comments:

Line 66:
- AP2 actually refers to the Apetala2/ethylene response factor transcription factor family.

Modified as suggested.

- The phrase âthe key driving the transcriptional wavesâ may lead to potential misunderstandings of the cited study. Referring to AP2s as âthe keyâ may imply that they are the sole drivers of transcriptional waves in T. gondii. Additionally, the phrase âthe transcriptional wavesâ should be clarified as it is unclear what they are referring to in this context, especially since the cited study solely investigated Plasmodium falciparum blood-stage development.

We regret that the reference here was indeed the incorrect one, for which we apologize: we have now replaced it with the correct Behnke et al 2010, and Xue et al 2020. In addition, we replaced the word âkeyâ with âpivotalâ.

Line 88:
- The use of âRNA explosionâ in this context is potentially confusing as it might be interpreted as a term denoting destruction. A more appropriate and descriptive term, such as "transcriptional burst", would be a better description of the observation, and has been used in the literature to derscribe similar phenomena.

Thank you for the comment. We adopted the suggestion.

Line 147:
- It is unclear what âthe expression level regulation is more granularâ is supposed to describe in this context.

We agree this word choice was unclear and have revised the text.

Line 156:
- The reference connecting the identified GO terms to the G1 phase of the cell cycle is missing.

There is no missing reference (if citation was meant?). We have simply listed the enriched GO terms (obtained from ToxoDB.org) in TE1. These comprise terms typically associated with (G1) growth, like biosynthesis of proteins and macromolecules, ribosomes, and metabolism.

Line 166 and line 821, Figure 3:

- The authors state that they âdifferentiated three different correlations between expression and chromatin accessibilityâ. However, the thresholds used for the three categories (loose, tight, and multi-functional) are not defined. Furthermore, it is unclear whether these gene sets are mutually exclusive since they represent different categories (some are cell cycle stages, others are protein localization). This should be specified in the figure legend or the text.

These criteria were indeed not clearly defined. We have added the following (underlined):
âLoose (RNA peaks clustered; ATAC multiple peaks or peaks hours offset from RNA peak), tight (peaks on both levels within 1.5 hr window) and multi-functional (both RNA and ATAC peaks for the group spread out over hours) (Fig. 3a, b).â

The reviewer is correct that gene sets are not mutually exclusive in different GO term gene sets. That is a natural feature of GO terms, as genes very often are present in multiple GO term gene sets and thus are not exclusive. Indeed, these can span the three different categories (Biological Processes, Molecular Functions, Cellular Components), and we explored all three categories.

Line 169:

- The authors state that âwhen the whole curve shape is considered (CCS), several genes in each group appear below the 0.6 cut-off, confirming this âlooseâ patternâ. However, in Figure 3b, the data show only one gene below the 0.6 cut-off in the loose S-phase set and two in the loose M-phase set. Contrarily, the majority of genes in both sets have a CCS score above 0.6. This observation appears to contradict the statement in the text, suggesting a need for correction or clarification in the manuscript.

That observation is correct, but we are considering these as a group sharing a similar process or function. Our choice of words also reflects this: âseveral genes in each group appear below the 0.6 cut-off, confirming this âlooseâ patternâ. To clarify this and avoid confusion, we added â....for the group.â at the end of this sentence.

Line 187:

- The reference supporting the statement that transcriptional waves are AP2-driven is missing.

We did add these in the introduction as stated above. To be complete, we included the Xue et al 2020 and Behnke et al 2010 references here as well.

Line 209 and line 821, Figure 3:

- The authors suggest that the data presented in Figure 3e demonstrates cascading due to overlapping functional modules. Yet, it appears that there is only a single instance of overlapping GO terms in subsequent scRNAseq clusters aligning with the authorâs interpretation. The arrangement of GO terms appears to be arbitrary, which raises the question of whether evidence of cascading would still be apparent if the GO terms were organized differently.

We would like to point out that there are 3 instances of overlapping GO terms: 1) âIMC-pellicleâ in TE2 and TE4; 2) âmembraneâ in TE2 and TE4; 3) âapical part of the cellâ in TE2, TE3, and TE4. All of these instances aligned well with our interpretations that S/M/C occur simultaneously in T. gondii. The way the GO terms were arranged was not arbitrary. They were first arranged based on the peaks in RNA transition cluster followed by the peaks of each ATAC sub-cluster order (from TE1CE1TA1 to TE4CE4TA4). Then only the sub-ATAC clusters with more than 8 genes were used in the GO term analysis. And then, the GO terms were arranged (left to right in Fig. 3e) based on the Benjamini significance values. This arrangement was not only logical but also systematic, and therefore, further supporting our interpretations that tachyzoite cell cycle was propelled by cascading transcriptional processes.

Line 220 and 304:

- The discovered motif (T/C)GCATGC(G/A) has not only been previously reported in references 21 and 30 but also in Markus et al. (PMID: 33553008), which identified this motif associated with 44 % of TSS in Tachyzoites. This publication should be cited and discussed.

Thanks for directing us to this as we had indeed missed the motifâs appearance in this paper. We have added this finding and the reference accordingly.

Line 259 and line 838, Figure 4:
- The tagged strain exhibits a noticeable growth defect in the -IAA plaque assay. This point warrants discussion and acknowledgment as a potential caveat for the subsequent analysis. In this context, it is puzzling that there is no observable difference between the parental strain and the AP2XII-8-mAID-5xTy strain under -IAA conditions in Figure 4f, a discrepancy that requires discussion.

Indeed, we consistently observed smaller plaques across three biological repetitions (not shown) in the AP2XII-8-mAID strain under permissive conditions. This is solely due to the introduction of Tir1 into the parent strain. This is known in the field but not well documented e.g. we generated Tir1 strains multiple times and in different genetic backgrounds with very reproducible, slightly slower growing outcomes. Typically, mAID tagging is performed into a parent line already expressing Tir1, and then plaque assays are compared to the Tir1 expressing parent line. Instead, we introduced mAID and Tir1 in a single integration event (see Suppl Fig 6a for strategy). To illustrate this effect we performed a comparative plaque assays (all in parallel) as shown below:

The two middle panels represent parasites expressing Tir1, panel 2 in a clean RHÎKu80 background, and panel three is the AP2XII-8-cKD line used in the paper (without IAA): these display similar plaque size, which is smaller than the RHÎKu80 parent line (far left panel). Moreover, the fourth panel (far right) shows an AP2XII-8 line tagged with TurboID, without simultaneous introduction of Tir1. These plaques are similar in size to the RHÎKu80 line in the first panel. Hence, we surmise that tagging AP2XII-8 does not affect parasite fitness, but introducing Tir1 does â that is what we see in Fig 4e.
A similar effect was seen by Wilde et al., 2023 (PMID: 36958824), where they created one-step auxin-inducible knockdown strains of TgOUTD3B, similar to the AP2XII-8-cKD strain that we generated. Under no IAA conditions, TgOUTD6B tagged strains produced approximately 25% smaller plaques. This corroborates our observation. We now refer to this in the Results section.

In summary, we are not alarmed about smaller plaques under permissive conditions compared to the parent line. That we did not detect a statistically relevant difference in Fig. 4f is likely due to an overall slowdown in cell cycle progression, not specific to a certain stage. Collectively, we think this alleviates the raised concern.

Line 269 and line 996, Extended Figure 7:
- The term âthe highest detection powerâ used in the manuscript is ambiguous. Detection power is generally understood as the capacity to identify a true effect. Therefore, this claim should be substantiated with an evaluation of peak quality, not merely based on the quantity of peaks identified. To assess the CUT&RUN protocol with the highest detection power, it would be beneficial to consider metrics such as signal-to-noise ratio or a comparable measure to evaluate the quality of the identified peaks.

We changed the term âthe highest detection powerâ to âoptimal detection capacityâ. MACS2 reports statistical thresholds (q-values) to control the overall rate of false discoveries. Only significant peaks (q-value < 0.05) were included in the subsequent analysis of the CUT&RUN data. In addition to q-values, MACS2 outputs signal values corresponding to the fold enrichment (FE) score for each called peak (note that when calling peaks using MACS2, we have positive samples and negative controls. Therefore, the signal value indicates the enrichment of the peak over the control).

In order to add more rigor to the CUT&RUN analysis and evaluate the quality of peaks, we used 6 CUT&RUN data with 6 different concentrations of antibody (0.5 ïg BB2, 1.0 ïg BB2, 2.0 ïg BB2, 0.625 ïg Diagenode, 1.25 ïg Diagenode and 2.5 ïg Diagenode) and analyzed each concentration and performed peak calling with respect to three negative control samples (RH-IgG1, AP2XII-8-IgG1 and RH-Ty). The box plot below (added as new Supplementary Fig. 7e) shows the Signal Values (Intensity) of all peaks per CUT&RUN data.

The average peak signal values detected in 2 ïg BB2 and 0.625 ïg Diag are the highest compared to the other data sets, with 2 ïg BB2 showing the tightest spread of all conditions. This analysis supports our choice of 2 ïg BB2 as the CUT&RUN data with the most optimal detection capacity.
Line 270 and 628:
- Despite the R code for CUT&RUN data analysis having been made available by the authors, key parameters that were used to transform the data should be stated in the methods section. These parameters, such as the specific distance criteria used for merging adjacent peaks, should

be explicitly stated to enhance clarity.

The parameters are specified in the method section in the line 640.
âmacs2 callpeak -p 0.05 -m 250 -t AP2XII-8.sorted.bam -c AP2XII-8-RH_S1.sorted.bam --nomodel -f BAM --extsize 120.â

To further clarify, the following is added to the manuscript:

âAll peaks detected by cellranger 10x pipeline (5444 total peaks) and gene annotation file were used to assign peaks to genes using bedtools. Peaks located entirely within the gene/coding region (exon-2 to exon-n) were filtered out from further analysis (427 total). The remaining peaks were assigned to the nearest downstream gene with distance cut-off < 3000 bp. Multiple peaks assigned to the same gene were merged. In total 5322 genes were uniquely assigned to a peak.â

Line 315:
- The detection of additional motifs is not sufficient to âinvestigate cooperative bindingâ. Determining the binding mode requires biochemical assays. Therefore, this sentence should be rephrased.

We removed the âcooperative bindingâ wording here.

Line 639:
- The provided link (https://bammmotif.soedinglab.org/) is not reachable. Access: 01/16/2024 â 5:20 PM

The previous link has been changed to https://bammmotif.mpibpc.mpg.de/ in the manuscript, which is where its availability changed into. For additional rigor, we have now included extra motif analysis using STREME command from the MEME-suit. Please also see response to minor comment regarding panel e below.

Line 804, Figure 2:
- Panel a and c: The methodology for selecting "transition points" is somewhat ambiguous. It appears that the authors identified these points by looking for peaks and troughs in the first derivative of the graph, essentially points where the slope of the curve approaches zero. However, it is noticeable that not all of these points were selected. In the discussion, the authors mention that the âmore fluidâ points were excluded from being designated as transition points, but the threshold for this decision remains unclear.

The reviewer is correct that not all the points of inflection were selected. The Figure below shows all detected points of inflection, however, as can be seen some of these points are very close to one another and analysis of genes peaking in these narrow regions did not yield any additional GO terms. We therefore merged the nearby points. We have updated the text to clarify the procedure.
RN

ATA

- Panel e: It should be commented on why the accessibility cluster TA4 initiates after the onset of expression cluster TE4. This observation raises questions: Are these genes expressed from heterochromatin, or might this discrepancy be attributed to issues of sensitivity in the experimental methods used?

The accessibility and expression clusters were identified independent of one another in the scRNA-seq or scATAC-seq data sets. Therefore, there is no direct connection between the genes in each transition. We interpret the transitions as different levels of regulation of gene expression, on either the chromatin level (epigenetic) or the transcription level (timing of availability and/or activation of the right set of transcription factors). As such, transitions on either level mark changes, which could be chromatin opening or transcription factor driven, but could equally be chromatin closing or transcription shut down. Directly correlating events at each level is therefore not relevant.

Line 821, Figure 3:
- Panel a and b: In the legend, it is specified that âbox plots represent 75 percentile; dots indicate outliersâ. Yet, numerous dots are positioned on the median line, contradicting their designation as outliers.

Thanks for catching this. All data plots are represented as dots indeed are not outliers. We changed the legend accordingly.
- Panel d: Similar to observations in Figure 2, there are instances where RNA levels peak even though chromatin accessibility is at its lowest. These occurrences warrant further discussion.

We interpret these as different levels of regulation. For example, chromatin might open, but transcription does not start till the right transcription factors are present. In addition, mRNA stability and turnover are other factors as we are not measuring active transcription but presence of transcripts. As such, in this specific situation it is possible that the chromatin is being set for G1, while the mRNA level peak is coming from genes in the previous chromatin switch. The list of genes in each cluster along with their expression and accessibility peak time is provided in Supplementary Table 3.

- Panel e: The identified DNA motifs are missing critical details necessary for evaluating their quality, including q-values and the count of genes within each cluster associated with the respective motif. Additionally, the methods section should clarify which dataset was utilized as a background for this analysis.

Thank you for the comment. We have now performed a major re-analysis of the motif search, using two software, The original BAMM motif finder as well as another widely tools STREME [PMID: 33760053] from the MEME-suit. BAMM reports both p-value and e-value, while we could only extract the p-values as reported by STREME. The negative background was not explicitly provided in the Motif analysis and default parameters were used. Motif search algorithms automatically construct the negative set, by shuffling the positive sequence while keeping the frequencies of k-mers, (e.g., 2-mers) fixed. Markov models are also used to generate negative sequences with estimated nucleotide and transition frequencies. The new analysis was restricted to classes with at least 50 positive sequences.

To increase the confidence in discovered motifs, we compared all the discovered motifs across the two software and quantified motif similarity by pairwise alignment and clustered the similar motifs using hierarchical clustering. For each cluster, a consensus motif was constructed from the PWM obtained by performing a multiple-sequence alignment on the motifs within each cluster (Please see methods for details). All discovered motifs along with their significances are presented in the new Supplementary Table 4. That application of these additional filters (consensus across two software and limiting the analysis to classes with at least 50 sequences), results in 21 overall motifs. Figures 3 and 6, as well as Supplementary Fig. 5 and 7, are updated according to these new analyses. All discovered motifs along with their significance and sequence numbers are presented in new Supplementary Table 4.

We have also reported the total number of genes within each represented sub-ATAC cluster in Fig. 3e. For a complete list of gene information within each sub-ATAC cluster, please refer to the updated Supplementary Table 3.

Motif search was performed under the ATAC peaks of genes within each ATAC sub-cluster (Fig. 3e). CUT&RUN peaks were searched to identify potential binding sites of AP2XII-8 TF.

Line 838, Figure 4:
- Panel e: The labeling of the AP2XII-8-mAID-5xTy strain as AP2XII-8 cKD is confusing as AP2XII-8 is not expected to be knocked down in the absence of IAA. This labeling should be revised for greater clarity.

The abbreviated use of âcKDâ or âiKDâ (inducible KD) for the parasite lines with a conditional allele is widespread in the field. The cKD fits in the context that the protein is âconditionallyâ knocked down only in the presence of IAA. We therefore are of the opinion that the labeling is in line with what is used in the field.

Line 859, Figure 5 and 996, Extended Figure 7:
- Figure 5a and Extended Figure 7d: The significance of the color gradient used in the figure is not specified. This should be clarified in the figure legend. Additionally, the x-axis is inaccurately labeled as âgene distance (bp)â when, in fact, the distance is depicted in kilobases (kb). This labeling error needs to be corrected.

We regret the oversight and replaced âbpâ with âkbâ in the figures, and explained in the legend that the gradient bars represents intensity.

Line 872, Figure 6:
- Panel a: It appears that AP2XII-8 KD results in a higher proportion of parasites in the S-phase. This appears contradictory to what is depicted in Figure 4f, where AP2XII-8 KD seems to increase the proportion of parasites in the G1 phase. How can such a discrepancy be explained?

Fig. 6a shows the transcriptional profiles of AP2XII-8 KD parasites at the RNA level, while Fig. 4f depicts the parasite cell cycle statuses after losing the proteins at the âcellularâ level. They were measuring two different parameters. The bulge portion away from the regular progression through G1b-S in Fig. 6a indicates that T. gondii is unable to progress properly through S-phase, but that its transcriptional profile still partly reflects S-phase. Please also see our rebuttal to Reviewer 1, comment 1, regarding the S-phase AP2 factors present in the bulge cells.

Overall, since RNA production precedes cellular activities, this indicates that T. gondii is physically stuck in a stage before the S-phase awaiting a cell biological cue, which is reflected as the G1 arrest in Fig. 4f.

- Panel d and e: Similar to Figure 3e, the identified DNA motifs are missing critical details necessary for evaluating their quality. How many up-regulated genes contain the motif and how many CUT&RUN peaks contain the motif? Additionally, the methods section should specify the gene set used as a background for the motif discovery process.

Information on the quality of detected motifs is included in the Supplementary Table 4. We did not explicitly specify background sequences (please see response to comment regarding motifs above). We have modified the text to include the information on gene sets used for motif discovery.

The list of DEGs and CUT&RUN targets containing DNA motifs can be found in Supplementary Table 5.

According to the motif analysis under the union of CUT&RUN peaks, out of 343 stringent CUT&RUN targets, 119 contain the AP2XII-8 associated motif (Fig. 6f; this is the motif found in cluster #6 in Extended Fig. 6a). Out of these 343 CUT&RUN targets, 120 genes are differentially expressed upon AP2XII-8 depletion:
â¢ 81 are down-regulated; 42 of them contain the motif
â¢ 36 are up-regulated; 5 of them contain the motif
â¢ 3 are modulated; 2 of them contain the motif
We have added these numbers in the Results section.
- Panel f: The title âribosomal genes scRNA-seqâ is confusing as the plot specifically refers to the AP2XII-8 down-regulated ribosomal genes as indicated in the figure legend. Therefore, the plot's title should be revised to accurately reflect this specificity.

We regret this confusion. These are actually the unperturbed expression and chromatin accessibility profiles, so headers are correct. We changed the legend to clarify this.

Line 945, Extended Figure 2 and line 961, Extended Figure 3:
- In both figures, the current color scheme leads to multiple genes being represented by apparently identical colors. To help enhance the distinction between genes, it would be beneficial to modify and diversify the color palette used.

Indeed, some colors are closely alike. We modified them to more contrasting palettes permitting easier differentiation.

Line 996, Extended Figure 7:
- Panel e and f: The dataset labels are somewhat challenging to understand. Consistency in labeling, similar to what is seen in panel c, would be beneficial for clarity. For example, labels could be "AP2XII-8 BB2 vs. AP2XII-8 IgG1" and "AP2XII-8 BB2 vs. RHâKU80 IgG1". The fact that only 25% of identified target genes overlap depending on the chosen negative control raises concerns and warrants discussion, especially in light of the growth defect observed in the AP2XII-8-mAID-5xTy strain under -IAA conditions.

First, we would like to point out that the growth defect should not be a concern in this study, as the protein can be successfully detected by IFA and western blot with the Ty epitope tag, which indicates that the protein could be pulled down during the CUT&RUN experiments. Second, in the past (T. gondii) protein binding sites studies involving similar approaches (ChIP-seq, CUT&Tag, CUT&RUN, etc.), negative control experiments are surprisingly sparse, suggesting that many binding sites may be

just false positives.

To this end, and to ensure that our findings were true positives, we opted for multiple negative controls, including AP2XII-8 IgG1, RHÎKu80 IgG1, and RHÎKu80 BB2.

In addition, in response to Reviewer 3, comment 1, we revisited our CUT&RUN (see below): our results for factors analyzed before, e.g. H3K4me3 and AP2IX-5, compare very well with published data. As such, we are convinced that our data here accurately reflect the AP2XII-8 engaged genes.

Reviewer #2 (Remarks on code availability):

As described in the main review, some parameters for usage should be added to the manuscript.

One data source (https://bammmotif.soedinglab.org/) is not reachable.

The previous link has been changed to https://bammmotif.mpibpc.mpg.de/ in the manuscript.
Reviewer #3 (Remarks to the Author):

The manuscript by Lou and colleagues title âSingle cell expression and chromatin accessibility of the Toxoplasma gondii lytic cycle identifies AP2XII-8 as an essential ribosome regulon driverâ analyze the T. gondii transcriptome and chromatin accessibility at single cell level. The team clustered their data and used a pseudo-time trajectory analysis to identify four transcriptional bursts corresponding to the major cell cycle transition. The control of the dominant G1 burst was further analyzed by the functional analysis of AP2XII-8, a transcription factor thought to drive the identified burst. The team then used a combination of molecular, cellular and genome wide (CUT&RUN) experiments to identify the AP2XIIâs target genes and concluded that AP2XII-8 is controlling the expression of the ribosome genes. To facilitate access to the datasets, an interactive web-app is provided where users can perform various analyses, including clustering and co-expression analysis.
Several studies have already analyzed the transcriptome and chromatome of several apicomplexan parasites using a combination of genome-wide tools including single-cell RNA and Transposase-Accessible Chromatin (ATAC) sequencing in T. gondii. Results of these studies have demonstrated that chromatin opening facilitates mechanism underlying gene expression. The present study validates previous published results by analyzing simultaneously the expression and the chromatin accessibility of T. Gondii during its cell cycle progression. While the transcriptome and chromatin state data sets generated tend to have a low resolution, results presented are robust. The findings are however not entirely novel, nor surprising. I am nevertheless puzzled by the analysis and the interpretation of the data generated for the identification of AP2XII-8 target genes (the novelty of the manuscript) using CUT&RUN and scRNA-seq experiments. Please find below my major comments. Some of them will need to be fully addressed before this manuscript can be published.

Major comments:

1 - The Cleavage Under Targets and Release Using Nuclease (CUT&RUN) experiment has been developed in the epigenetic field to increase sensibility of the traditional ChiP-seq experiment. While ChiP-seq experiments require in general a few millions of cells (1 to 10 millions of cells in higher eukaryotes â 5 to 40 millions of cells in the T. gondii genome), the CUT&RUN protocol require less than 500,000 to be accurate. Exceeding 500,000 cells per sample (1 million for smaller eukaryotes) is not recommended. Generally, a good starting point to optimize this protocol is usually done with 50,000 cells. Some experiments were even successful with less than 200 cells. Using too many cells in such protocol will increase the background, reduce the yield, reduce the specificity and, more importantly, reduce the complexity of the library. The team optimized their protocol and determined that 200 millions of parasites had the highest detection power. This seems extremely high and is the exact opposite of what most people in the field have experienced. I am therefore wondering how the results generated using CUT&RUN were validated. I think it will be best to confirm the result using a classical ChIP-seq experiment with a maximum of 10-40 millions parasites.

The reviewer is correct that taking into account genome size, translating the EpiCypher protocol from mammalian cells into T. gondii comes to the range of 10-40 million cells. To this end, we initially tried 50 million cells, and were successfully detecting H3K9Me3 using 50 million parasites, but only obtained a total of <1 ng of CUT&RUN enriched DNA for the AP2-XII-8 with Ty antibodies, which was way below the suggested 5 ng DNA input for library construction according to the used EpiCypher protocol. Since the kit is designed for widespread modification on histone tails, we reasoned that because AP2XII-8 is only expressed in the parasites in a narrow window of time, not even all input parasites express this protein (~47% of parasites are in C-G1a phase in a heterogenous population; see figure in response to comment 4), making it even more difficult to capture the signal if only a small

amount of parasites were used. Therefore, we next increased the cell input to 500 million to ensure the successful generation of the CUT&RUN library. This turned out to be too noisy (Supplementary Fig. 7d) upon which we tuned down the input cell number to the sweet spot of 200 million.

Since the time of submission, we have performed multiple additional CUT&RUN and CUT&Tag experiments on epigenetic factors, ApiAP2 factors and Myb factors, including a CUT&RUN & CUT&Tag experiment on H3K4me3 for which there is public ChIP-Seq data available as well as Ap2XI-5, for which there is publicly available ChIP-Chip data (PMID: 24025328). Accordingly, we are able to verify the quality and reproducibility of our data by comparing with public ChIP-seq and ChIP-Chip data. The AP2XI-5 ChIP-Chip data (PMID: 24025328), is a tiling microarray data obtained from Agilent. We then compared our CUT&RUN & CUT&Tag signal for the positive control sample H3K4me3 (above) as well as CUT&RUN & CUT&Tag for AP2XI-5 (below) with these publicly available data sets. The figures show IGV tracks of randomly selected sections.

Six IGV tracks for H3K4me3: ChIP-Seq (top 3); our H3K4me3 CUT&RUN (red); and our H3K4me3 CUT&TAG (blue).
Four IGV tracks for AP2XI-5: our CUT&RUN and CUT&Tag for AP2XI-5 (top 2 tracks) and ChIP-Chip for AP2XI-5 (PMID 24025328; bottom 2 tracks). CUT&Tag and CUT&RUN signals are much stronger and less noisy compared to the ChIP-Chip, but the dominant peaks agree across the genome.

We then calculated the Spearman correlation across these data sets across the genome (see data analysis below). The results show a high degree of agreement between the two data sets (Fig below). This is particularly remarkable in the case of AP2XI-5. Despite the fact that the old ChIP-chip technology was applied over a decade ago (PMID 24025328), the data sets overwhelmingly agree with our data, giving us high confidence in quality of our data sets and the validity of our results.

Correlation plots between CUT&RUN, CUT&TAG, and ChIP-Seq/ChIP-chip for H3K4me3 (left) and AP2XI-5 (right). There is a high-degree of agreement between these technologies, demonstrating the quality of our CUR&RUN data.

Taken together, these results show the reproducibility of CUT&RUN. We have added selected IGV tracks for H3K4me3 in the manuscript to illustrate that our CUT&RUN approach detects the same profiles as two independently published ChIP-seq experiments (new Supplementary Fig. 7f).
We note here that CUT&Tag using the recently described lysis buffer (PMID 38278808) in combination with the EpiCypher CUT&Tag kit displays very clean signals, although some minor peaks appear to drop below the threshold. We have not included this in the main manuscript though.

Data Processing Steps:
We processed the ChIP-Chip data as follows:
- The tiling array probe sequences were replicated based on the log (base 10) ratio of the intensity between âredâ and âgreenâ signals. All negative ratios were set to 0.
- The probe sequences list was then output as a Fasta file to reflect the intensities.
- This Fasta of probe sequences reflecting the intensity of log ratios obtained by ChIP-chip was aligned to the T. gondii genome V65 using bowtie2 (version 2.4.2)
- We then used multiBamSummary from deeptools (version 3.5.2) to obtain a count matrix representing read counts mapped to the genome at single base pair resolution for AP2XI-5. Only bases represented in AP2XI-5 ChIP-chip were used to cross-compare across the different technologies.
- The matrix was normalized by total read count and min-max normalization and a kernel density function was fitted across the matrix, with rows as samples and columns as genomic loci.
- Spearman correlation was run on the obtained normalized matrix and a correlation heatmap was plotted using the corrplot package in R with Euclidean distance clustering using hclust.
These steps ensure reproducibility and robust cross-technology comparisons in our genomic analyses.

2 - Data analysis of the CUT&RUN experiment is even more puzzling. The team identified over 2,338 binding events (peaks) that were assigned to a list of 970 targeted genes. This list was then narrowed down to 246 genes after background subtraction using signals detected in the negative controls. From these 246 genes only 95 exhibited a change in their expression profiles in a AP2XII-8 knock down experiment with 73 down regulated, 19 up regulated and 3 either down- or up-regulated in different phases. These results make no sense. Why less than 10% of the targeted genes are affected by AP2XII-8 knock down? Overall, the CUT&RUN results seem too noisy, and it is very likely that data generated (both in the positive and negative controls) are inaccurate due

to the large number of parasites used in the experiment. Considering that the team needs 200 millions parasites for their CUT&RUN protocol to work, I will suggest a classical ChIP-seq experiment that if done well - with less than 20 million parasites - should generate accurate results with almost no peaks detected in the negative controls therefore increasing the specificity of the ChIP-seq signals in the samples analyzed.

Please see our rebuttal above: our CUT&RUN protocol produces the same data as previously collected by ChIP-seq in other labs. In essence, it is not the protocol, but the biology. Moreover, while the CUT&RUN data identifies direct targets, not all targets are functional in the sense that expression of the target gene may only be modulated under the right conditions (e.g., signaling pathways, dimerization with other TFs, and presence with epigenetic factors). On the other hand, the KD data identifies genes whose expressions are potentially modulated upon the depletion of the TF and may include off-target effects. Therefore, looking at the intersection of these two data sets will reveal direct & functional targets. That being said, we agree that the biology is fascinating, and it is so interesting to see that only 10% of the target genes were also transcriptionally affected by AP2XII-8. To this observation, our interpretation is that for many other genes, AP2XII-8 is not the sole regulator, and the effects are only observable when the other gene is co-depleted. Such a phenomenon has been elaborately described in a recent study by Antunes et al. (Nature, 2024, PMID: 36711883), where the complete transition from tachyzoites to merozoites in T. gondii is only initiated when both AP2XII-1 and AP2XI-1 are depleted. More recently, for AP2XII-2 only 105 genes (66 up, 39 down; 1.5-fold) were reported under its transcriptional control, while 3,939 genes were bound (out of total of 5,527 peaks) (PMID: 36786611). As raised by reviewer 1, several scenarios could be at work here, and we added a section in the discussion pertaining to this conundrum.

In addition, please see our response to Reviewer 2 Major Comment 1 regarding additional scRNA-seq knock-down time points to assure we did not under- or over-estimate genes under AP2XII-8 transcriptional control.

3 - Also Considering that the genes coding for the ribosomes are the most highly expressed in the genome of all eukaryotes, they are usually the most easily-detectable by single cell sequencing experiment. it is very likely that AP2XII-8 mediate more than just those genes during the cell cycle transition. The resolution of the data presented remains low and I will be careful with the interpretation of the data.

The opposite scenario is actually what we see, as these highly expressed genes are downregulated upon AP2XII-8 depletion. So this observation does not arise from the high native expression level of ribosomal genes.

As discussed in rebuttal to Reviewerâs 1 and 2 above, we posit that AP2XII-8 likely cooperates with other TFs and/or epigenetic factors to co-regulate some other genes it engages. This, however, is beyond the scope of this study, and will be further investigated in our follow-up work.

4 - Can you specify in the text how many cells and genes were identified in each of the clusters. Looking at the figures presented, the number of genes detected in each cluster seems low. To accurately evaluate the resolution of the data, it will be important to add a few plots that exhibit the distributions of the unique molecular identifiers (UMIs) as well as the distribution of the genes detected in each cell and cluster for both scRNA-seq and scATAC-seq samples. Also, for the remaining genes that are not detected as DEG, what is the profile of their chromatin state? I donât see a representation of those genes in Figure 1.

T. gondii is a haploid organism with ~8,000 protein-coding genes, in contrast to humans that contain diploid genomes and have over 20,000 protein-coding genes. Also, there is no indication of gene number detected in each cluster. If the reviewer was commenting on the cell number, then we would like to argue that we had identified 12,735 cells in total for our WT scRNA-seq, exceeding our target cell recovery of 10,000. For the exact cell number distribution across each inferred cell cycle phase in both WT scRNA-seq and scATAC-seq, please see the figure below (now included as Supplementary Fig 1g-h).

The exact number of cells and reads is reported in the original submission in the first and second paragraphs of the result section. The number of differentially expressed genes in each of the cell cycle clusters is provided in Supplementary Table 2. All UMI and quality control measures are reported in Supplementary Table 1.

As for the Chromatin (or RNA expression) of non-cyclic genes, by default they are constitutive with minor fluctuations and hence were excluded from the heatmaps.

Minor comments:

1-In line 110-111, please clarify the 20 time points mentioned in the manuscript. Are these 20 time points evenly distributed within pseudo time (i.e. 4 time points for each of the 5 cell cycle stages)? Perhaps it isnât pertinent to this point, but figure 1f appears to show 18 distinct pseudo-time points.

Please note that single cell pseudo time does not account for variation in the duration of cell cycle phases, causing a âtime-warpingâ. To correct this, we relied on prior knowledge of cell cycle phase duration to rescale the pseudo-time to real time (Fig. 1f).

Please note that this analysis is distinct from generating time-series gene curves with splines. Once splines are fitted to the data, any interval can be selected to sample the curves, therefore providing high resolution time-series (pseudo) synchronized expression curves. The pseudo-time curve spans the range of the cell cycle [0,6 h]. The pseudo-time interval was partitioned into 18 equal 20-min bins to gather âsufficiently similarâ cells and increase the detection power (pseudo-bulk RNA-Seq). We corrected the mention of 20 time points to 19 time points in the text.

2 - In line 121-123 - Do you mean that of the 1,620 genes that are differentially expressed, 1,238 are unique to a specific cell cycle stage? This sentence is a bit confusing.

That was exactly what we meant to say. Of the 1,620 DEGs, 1,238 are specific to a single cell cycle stage and the remaining are differentially expressed in more than one stage.

3 - In line 139-142 - Is there any relationship between the genes confined within each transition state for expression and accessibility, or are the 1,238 genes clustered independently within each transition state? In other words, are the same cyclic genes found within TE1 and TA1? In the text, the data are described as generally highly correlated, but it will be good to give an exact number of genes that correlate and those that donât.

The 1,238 genes were clustered independently within each transition state, and genes in TE1 and TA1 have no direct relations, except for overlaps. The correlations that were described in Fig. 1h were based on the curve cross-correlations between the expressions and chromatin accessibilities of each of the 1,238 genes. This high correlation is irrelevant to the expression- or accessibility-derived transition clusters.

4 - In line 277-279 - Please clarify how the comparisons were performed. I expressed some of my concerns already regarding the CUT&RUN experiment and the way the data were analyzed. I am still quite confused on how the data were normalized and analyzed. How it currently looks is that any genes found within each negative control gene list were first subtracted from the list of AP2XII-8 targeted genes, generating four separate lists of genes. Then the resulting lists of genes were compared, and the overlap identified as shown in the Venn diagram. On the mean intensity plot (Extended Data Fig. 7e) it appears that the 246 genes have an intensity ranging from 0 to 25, with a large proportion of those genes being below 5-fold mean intensity. If the 246 genes only include those found in AP2XII-8 and none of the controls, thus the area only filled with blue (most of which is above 5-fold mean intensity) and not the red/blue overlap, then perhaps a different color is needed within the overlap. Furthermore, the usage of âintersectionâ in the intensity plot is somewhat confusing because some could understand it to mean the intersection of AP2XII-8 genes and those in the controls, rather than those only found in AP2XII-8. This CUT&RUN experiment is terribly confusing and will need to be validated using a more standard approach.

We regret the confusion that the figure and terminology has caused. Briefly, when calling peaks with MACS2, the positive CUT&RUN signal and the background negative CUT&RUN signal are both used by the software (MACS2), from which peaks FC enrichment and q-values are calculated. Each component of the Venn diagram represents genes detected in the (same) positive over a negative control and the genes at the intersection are those that always appear in the positive regardless of the negative control used.

We repeated this analysis for the same positive signal and, to properly control for the background, this time we used 3 negative samples, removing the 4th negative that corresponded to the control we originally named RH-neg, which was in fact RH-IgG1, but with 500 million parasites (please see response to Major comment 1 from Reviewer 1). Genes were identified in each, and the Venn Diagram in Supplementary Fig 7 shows the agreement between these genes. At the âintersectionâ; i.e., genes that appear in all 3 peak callings, we see 343 genes (36%). On the other hand, if we restrict the analysis to Ribosomal Protein genes, we see 59 at the intersection (78%).

The density figures on the other hand, show the distribution of signal intensity for all the âotherâ genes (red) vs. the 343 genes at the intersection of the Venn Diagram (similarly for Ribosomal; all the âotherâ Ribosomal vs, the 59). The signal intensities should be interpreted as follows. The intensity of the signal for genes called from the 3 data sets is significantly higher than the rest of the genes, indicating that the multiple negative controls were needed to eliminate false positive calls. Please note that the overlap between the intensity signals has no relation to the intersection in Venn Diagrams. The intensities are simply plotted on top of each other for direct comparison. We have updated the text and the caption and have changed the vocabulary for clarity.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have made a commendable revision of their manuscript in response to the reviewers' comments. The study is greatly improved by the inclusion of new data, and the textual revisions are adequately addressed. In my opinion, they have adequately addressed all concerns and presented a convincing study of the lytic cycle of Toxoplasma gondii. The detailed insights into the tachyzoite phase are both innovative and valuable and provide a solid foundation for understanding the effects of chromatin packaging on gene activity in this organism.

Reviewer #2

(Remarks to the Author)
This manuscript by the Zarringhalam and Gubbels labs investigates the intricate dynamics of transcriptional regulation in the asexual cell cycle of the apicomplexan parasite Toxoplasma gondii. The study addresses a significant gap in understanding the relationship between chromatin accessibility and transcription factor binding, while overcoming the challenges posed by T. gondii's rapid, asynchronous, and atypical replication cycle. The use of advanced single-cell transcriptomic and chromatin accessibility approaches enabled a detailed exploration of gene expression patterns and chromatin states, and the identification of functionally related gene sets relevant for transitioning from one cell cycle phase to another.
The authors find a strong correlation between chromatin opening and gene expression allowing for transcriptional bursts of four distinct RNA expression and chromatin accessibility clusters throughout the Toxoplasma cell cycle. Motif analysis of co-accessible promoter regions identified several known AP2 DNA binding sites but also motifs without any known interaction partners. Based on RNA velocity analysis, the authors identify AP2XII-8 as an important driver for the parasite’s progression through the G1-S phase. The functional role of AP2XII-8 was explored using CUT&RUN, highlighting its involvement in the regulation of a ribosomal regulon through two DNA motifs.
Thus, this work has generated a comprehensive resource for the field as it not only underscores the role of AP2XII-8 in the G1-S phase transition but also sheds light on potential gene sets crucial for other processes throughout the Toxoplasma cell cycle, such as karyokinesis. While the authors have been very responsive to the prior reviews, a few issues remain to be addressed. Noting appropriate caveats in the text, will be sufficient to address these remaining issues.

Major comments:

1. Line 257 and line 838, Figure 4: The authors claim that AP2XII-8 can be depleted using IAA in a period as brief as 2 hours. However, the western blot presented in Figure 4d shows many non-specific bands that complicate detection of AP2XII-8. The authors attempted to address this point and show no significant additional effects from prolonged KD periods (2h, 4h, 6h) via PIP-seq. However, the general concerns regarding the strain used remain.
1.1. The authors tagged AP2XII-8 with a TurboID-5xV5-epitope tag. The resulting strain again showed multiple non-specific bands, which raises concerns about the employed cloning strategy and may suggest random integration of the tag.

1.2. The authors added a vehicle-treated negative control and claimed no difference between both. Even though the banding pattern of the parental control and the vehicle-treated negative control appear very similar, they are not identical as claimed by the authors. In particular concerning the relative abundances of certain bands as for example between 28 and 39 kDa.

1.3 The authors claim that “from all the published T. gondii AP2 WB results, we found that most figures were either cropped or displayed similar banding ladder patterns, with some of them showing unspecific bands even with cropped representations”. Inter alia, the authors cite Khelifa et al., 2021 (PMID: 33414462) and Srivastava et al., 2020 (PMID: 32938695). Indeed, both WBs are cropped and display some unspecific bands, however, not to the same degree as the present study.

2. Line 261 and line 838, Figure 4: The observation in Figure 4f that the depletion of AP2XII-8 leads to an accumulation of parasites in the G1 stage does not conclusively establish AP2XII-8 as a critical factor for the G1 transcriptional burst. Notably, even after 24 hours of AP2XII-8 knockdown, several parasites are still in the S-phase or M/C phase, indicating that they can progress through the G1 stage without AP2XII-8. The authors addressed this point by toning down their language, stating that AP2XII-8 is an essential factor involved in the G1 transcriptional burst. In addition, the authors provided the requested velocity analysis under AP2XII-8 knockdown conditions, which shows a dampened G1 transcriptional burst,

although not as drastically as stated here.

Minor comments:

Line 156: The reference connecting the identified GO terms to the G1 phase of the cell cycle is missing. The authors state "There is no missing reference (if citation was meant?). We have simply listed the enriched GO terms (obtained from ToxoDB.org) in TE1. These comprise terms typically associated with (G1) growth, like biosynthesis of proteins and macromolecules, ribosomes, and metabolism." However, the citation connecting these GO terms with G1 is still missing. At a minimum the GO publications should be cited:
- Ashburner et al. Gene ontology: tool for the unification of biology. Nat Genet. 2000 May;25(1):25-9. DOI: 10.1038/75556
- The Gene Ontology Consortium. The Gene Ontology knowledgebase in 2023. Genetics. 2023 May 4;224(1):iyad031. DOI: 10.1093/genetics/iyad031

Line 209 and line 821, Figure 3:
- The authors suggest that the data presented in Figure 3e demonstrates cascading due to overlapping functional modules. Yet, it appears that there is only a single instance of overlapping GO terms in subsequent scRNAseq clusters aligning with the author's interpretation. The arrangement of GO terms appears to be arbitrary, which raises the question of whether evidence of cascading would still be apparent if the GO terms were organized differently. In response to this issue, the authors state that "the GO terms were arranged (left to right in Fig. 3e) based on the Benjamini significance values. This arrangement was not only logical but also systematic, and therefore, further supporting our interpretations that tachyzoite cell cycle was propelled by cascading transcriptional processes." Even though it remains unclear how GO terms with multiple Benjamini significance values were treated, this statement is an important clarification and should be added to the figure legend.

Reviewer #3

(Remarks to the Author)
The authors have addressed my comments. I am still puzzled about the need to optimize a CUT&RUN experiment that required over 200 million cells when you can use a standard ChIP-seq experiment in T. gondii with only 20-40 million cells.

Author Rebuttal letter:

Response to Reviewer comments
We would like to thank the reviewers and the editor for assessing our work and providing useful feedback. We have completely revised the manuscript based on the feedback and addressed all the concerns. Reviewer comments are listed below followed by our responses.

Point-by-point rebuttal

Reviewer #1 (Remarks to the Author):

The authors have made a commendable revision of their manuscript in response to the reviewers' comments. The study is greatly improved by the inclusion of new data, and the textual revisions are adequately addressed. In my opinion, they have adequately addressed all concerns and presented a convincing study of the lytic cycle of Toxoplasma gondii. The detailed insights into the tachyzoite phase are both innovative and valuable and provide a solid foundation for understanding the effects of chromatin packaging on gene activity in this organism.

We sincerely thank the reviewer for their positive feedback and thoughtful comments. We appreciate the recognition of our efforts to revise the manuscript and include new data. We are delighted that the revisions have strengthened the study and provided valuable insights into the lytic cycle of Toxoplasma gondii. Thank you for your thorough review and support.

Reviewer #2 (Remarks to the Author):

Introduction:

This manuscript by the Zarringhalam and Gubbels labs investigates the intricate dynamics of transcriptional regulation in the asexual cell cycle of the apicomplexan parasite Toxoplasma gondii. The study addresses a significant gap in understanding the relationship between

chromatin accessibility and transcription factor binding, while overcoming the challenges posed by T. gondii's rapid, asynchronous, and atypical replication cycle. The use of advanced single-cell transcriptomic and chromatin accessibility approaches enabled a detailed exploration of gene expression patterns and chromatin states, and the identification of functionally related gene sets relevant for transitioning from one cell cycle phase to another.
The authors find a strong correlation between chromatin opening and gene expression allowing for transcriptional bursts of four distinct RNA expression and chromatin accessibility clusters throughout the Toxoplasma cell cycle. Motif analysis of co-accessible promoter regions identified several known AP2 DNA binding sites but also motifs without any known interaction partners. Based on RNA velocity analysis, the authors identify AP2XII-8 as an important driver for the parasiteâs progression through the G1-S phase. The functional role of AP2XII-8 was explored using CUT&RUN, highlighting its involvement in the regulation of a ribosomal regulon through two DNA motifs.
Thus, this work has generated a comprehensive resource for the field as it not only underscores the role of AP2XII-8 in the G1-S phase transition but also sheds light on potential gene sets crucial for other processes throughout the Toxoplasma cell cycle, such as karyokinesis. While the authors have been very responsive to the prior reviews, a few issues remain to be addressed. Noting appropriate caveats in the text, will be sufficient to address these remaining issues.

Major comments:

1. Line 257 and line 838, Figure 4: The authors claim that AP2XII-8 can be depleted using IAA in a period as brief as 2 hours. However, the western blot presented in Figure 4d shows many non-specific bands that complicate detection of AP2XII-8. The authors attempted to address this point and show no significant additional effects from prolonged KD periods (2h, 4h, 6h) via PIP-seq. However, the general concerns regarding the strain used remain.

1.1. The authors tagged AP2XII-8 with a TurboID-5xV5-epitope tag. The resulting strain again showed multiple non-specific bands, which raises concerns about the employed cloning strategy and may suggest random integration of the tag.

Although we understand the reviewer's concerns, the bands are specific as there is no signal in the untagged controls. The possibility of random integration we can technically not exclude, but is inconsistent with the general experience in the field using the Ku80 knock-out background as parent, including our own endogenous locus tagging experiments (e.g. PMID 30279285, 31470470). Please keep in mind that asynchronously replicating parasites are harvested for WB and that AP2XII-8 turns over during the cell cycle, making presence of turnover products likely. There are no current methods to tightly synchronize the T. gondii cell cycle to test this hypothesis directly. Since our data are in line with published WB data on other AP2 factors (see #1.3 below), we think an interpretation of unstable AP2 or turnover is the most plausible, and the result does not originate in random integration of the tag.

1.2. The authors added a vehicle-treated negative control and claimed no difference between both. Even though the banding pattern of the parental control and the vehicle-treated negative control appear very similar, they are not identical as claimed by the authors. In particular concerning the relative abundances of certain bands as for example between 28 and 39 kDa.

Although there are indeed minor variations in intensities of a few protein bands, the general pattern is very similar. The source of this minor variation is most likely batch-to-batch variation: note that the figure in the paper also displays minor differences relative to the figure in the rebuttal. As such we do not believe the minor differences represent significant differences, despite our slightly stronger language used in the first rebuttal.

1.3 The authors claim that âfrom all the published T. gondii AP2 WB results, we found that most figures were either cropped or displayed similar banding ladder patterns, with some of them showing unspecific bands even with cropped representationsâ. Inter alia, the authors cite Khelifa et al., 2021 (PMID: 33414462) and Srivastava et al., 2020 (PMID: 32938695). Indeed, both WBs are cropped and display some unspecific bands, however, not to the same degree as the present study.

We would like to point out that the AP2 WB results cited â Khelifa et al., 2021 (PMID: 33414462) and Srivastava et al., 2020 (PMID: 32938695) â were cropped (at least partially), which does make it challenging to directly compare them to our uncropped results as we do not know what is not shown (why the blots were cropped?).

2. Line 261 and line 838, Figure 4: The observation in Figure 4f that the depletion of AP2XII-8

leads to an accumulation of parasites in the G1 stage does not conclusively establish AP2XII-8 as a critical factor for the G1 transcriptional burst. Notably, even after 24 hours of AP2XII-8 knockdown, several parasites are still in the S-phase or M/C phase, indicating that they can progress through the G1 stage without AP2XII-8. The authors addressed this point by toning down their language, stating that AP2XII-8 is an essential factor involved in the G1 transcriptional burst. In addition, the authors provided the requested velocity analysis under AP2XII-8 knockdown conditions, which shows a dampened G1 transcriptional burst, although not as drastically as stated here.

We appreciate the reviewer's thoughtful comments and recognition of our efforts to address this concern. Thank you for your valuable feedback.

Minor comments:

Line 156: The reference connecting the identified GO terms to the G1 phase of the cell cycle is missing. The authors state âThere is no missing reference (if citation was meant?). We have simply listed the enriched GO terms (obtained from ToxoDB.org) in TE1. These comprise terms typically associated with (G1) growth, like biosynthesis of proteins and macromolecules, ribosomes, and metabolism.â However, the citation connecting these GO terms with G1 is still missing. At a minimum the GO publications should be cited:
- Ashburner et al. Gene ontology: tool for the unification of biology. Nat Genet. 2000 May;25(1):25-9. DOI: 10.1038/75556
- The Gene Ontology Consortium. The Gene Ontology knowledgebase in 2023. Genetics. 2023 May 4;224(1):iyad031. DOI: 10.1093/genetics/iyad031

We apologize for the previous misinterpretation of this comment. We appreciate the reviewer's insights on this matter and have updated our citations accordingly.

Line 209 and line 821, Figure 3:
- The authors suggest that the data presented in Figure 3e demonstrates cascading due to overlapping functional modules. Yet, it appears that there is only a single instance of overlapping GO terms in subsequent scRNAseq clusters aligning with the authorâs interpretation. The arrangement of GO terms appears to be arbitrary, which raises the question of whether evidence of cascading would still be apparent if the GO terms were organized differently. In response to this issue, the authors state that âthe GO terms were arranged (left to right in Fig. 3e) based on the Benjamini significance values. This arrangement was not only logical but also systematic, and therefore, further supporting our interpretations that tachyzoite cell cycle was propelled by cascading transcriptional processes.â Even though it remains unclear how GO terms with multiple Benjamini significance values were treated, this statement is an important clarification and should be added to the figure legend.
We appreciate the reviewer's insights and feedback regarding Figure 3e. In response to the concerns raised, we have clarified in our statement that the GO terms were arranged from left to right in Figure 3e based on their Benjamini significance values. This systematic and logical arrangement supports our interpretation that the tachyzoite cell cycle is driven by cascading transcriptional processes. We have now included this clarification in the Figure 3 legend. Thank you for bringing this to our attention and for your valuable feedback.

Reviewer #3 (Remarks to the Author):

The authors have addressed my comments. I am still puzzled about the need to optimize a CUT&RUN experiment that required over 200 million cells when you can use a standard ChIP-seq experiment in T. gondii with only 20-40 million cells.

We thank the reviewer for their valuable feedback and for acknowledging that our comments have been addressed.