Supplementary Information:

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

Jingjing Lou[1]*, Yasaman Rezvani[2,3]*, Argenis Arriojas[3], Yihan Wu[1], Nachiket Shankar[3], David Degras[2], Caroline D. Keroack[4#], Manoj T. Duraisingh[4], Kourosh Zarringhalam[2,3§], Marc-Jan Gubbels[1§]

[1] Department of Biology, Boston College, Chestnut Hill, Massachusetts, USA

[2] Department of Mathematics, University of Mass. Boston, Boston, Massachusetts, USA

[3] Center for Personalized Cancer Therapy, University of Mass. Boston, Boston, Massachusetts, USA

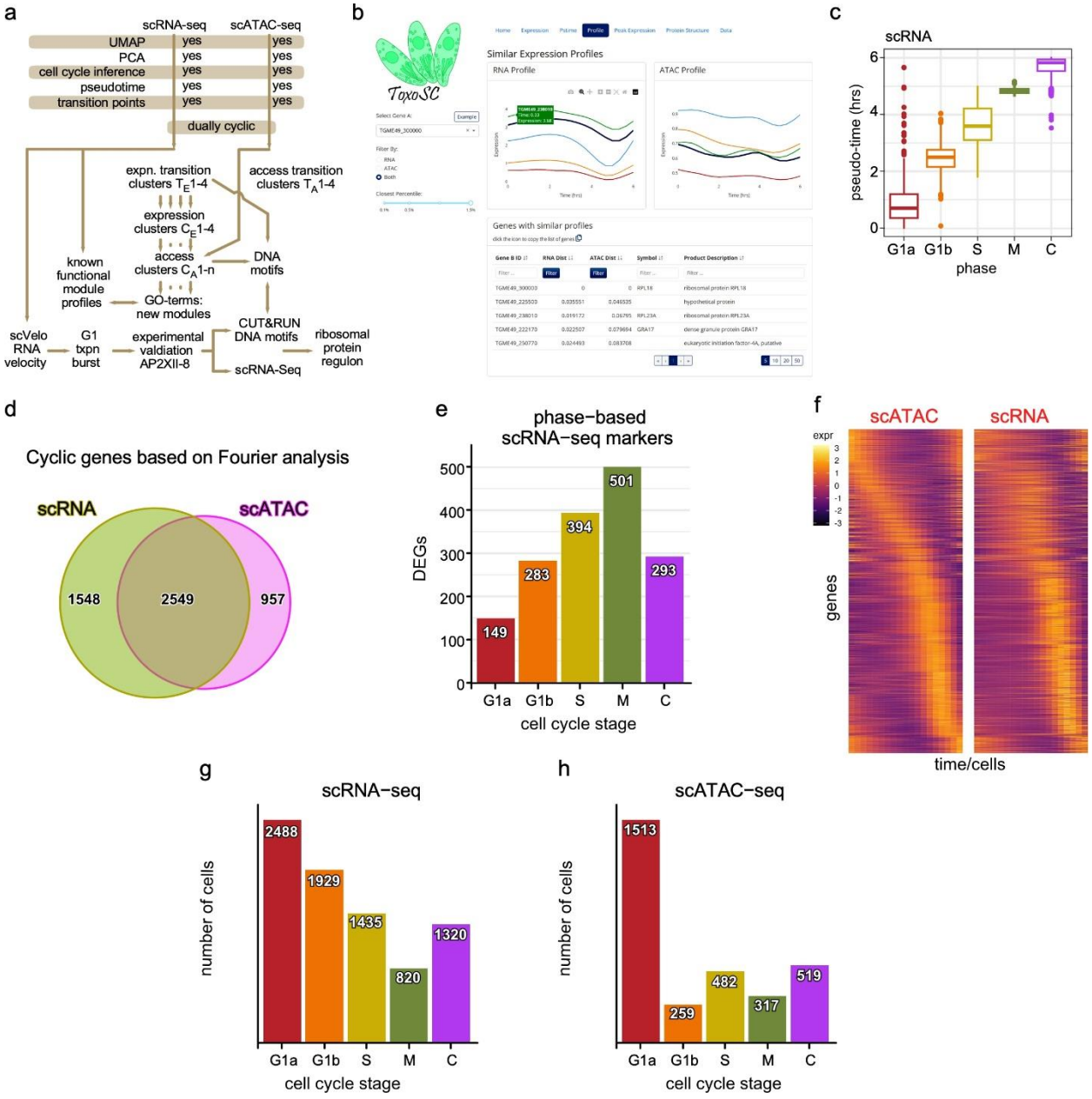
[4] Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health, Harvard University, Boston, MA, USA

*These authors contributed equally

[#]present address: Department of Molecular Microbiology and Immunology, Brown University, Providence, RI, USA

[§]These authors jointly supervised this work: kourosh.zarringhalam@umb.edu and gubbelsj@bc.edu

Supplementary Figures:



Supplementary Fig. 1. Study design, web-app and uncovering differentially expressed cyclic genes.

a. Schematic summary of the experimental and computational approaches taken in this study.

b. Web-app displaying query for genes by similarity in expression and/or chromatin accessibility pattern.

c. Box-plot representing the 75 percentile of distributions of cells at respective pseudo-time phases. Whisker indicates min/max and dots represent outliers. Horizontal bars indicate the median values.

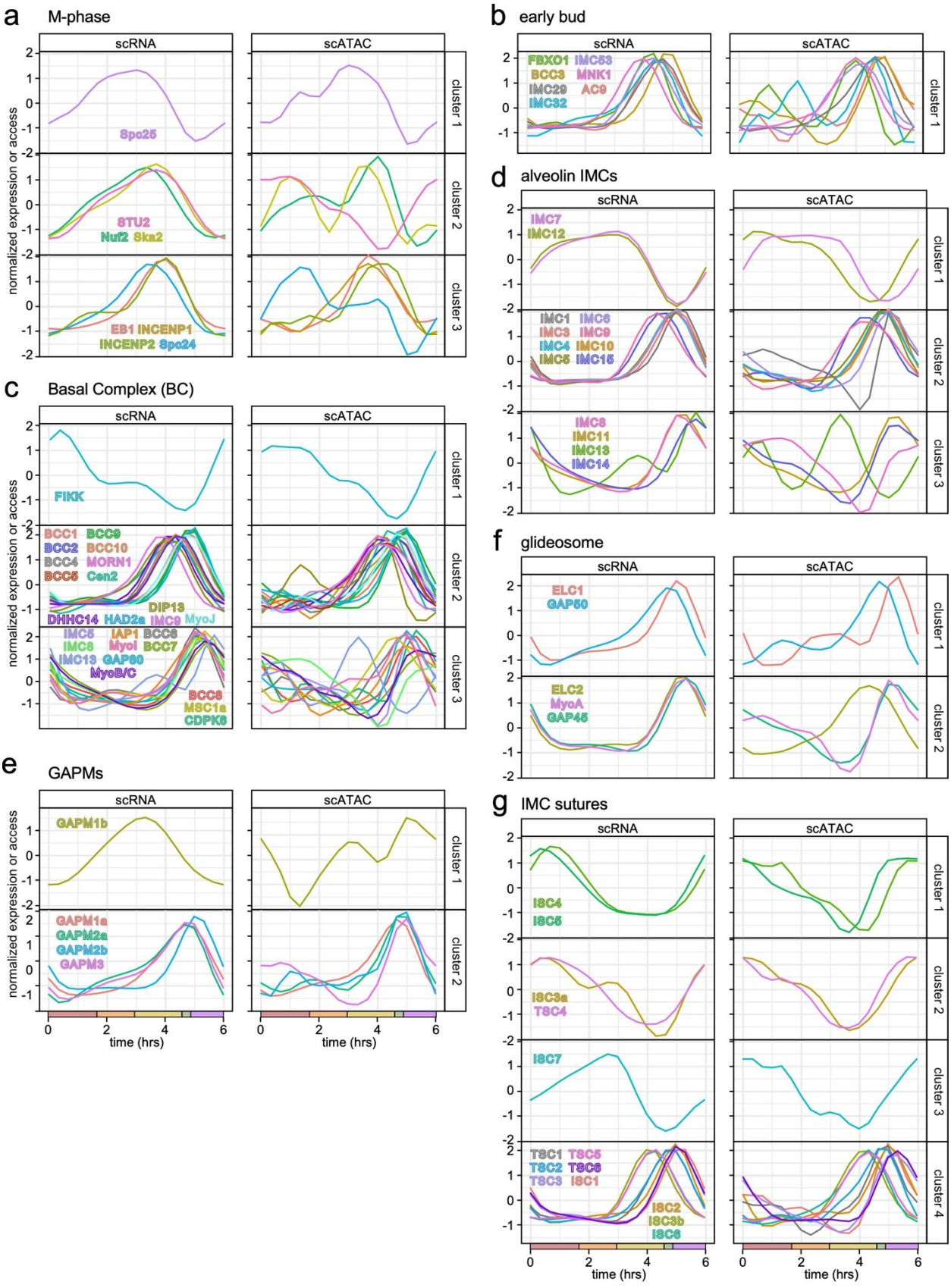
d. Fourier-based analysis identified 4,097 and 3,506 cyclic genes according to their expression and accessibility profiles, respectively. Ven diagram display 2,549 genes which were cyclic in both expression and accessibility curves.

e. DEGs within each inferred cell cycle phase (FC > 1.5, adj-p-value < 0.05).

f. Pseudo-time heatmaps of single cell transcriptome (right) and chromatin accessibility (left) for the 1,238 genes displaying cyclic expression and chromatin accessibility profiles. Top to bottom organized by timing of peak detection in scATAC-seq data.

g. Distribution of cells within each cell cycle phase after cell type annotation of scRNA-seq data.

h. Distribution of cells within each cell cycle phase after cell type annotation of scATAC-seq data.


a
M-phase
scRNA
scATAC
normalized expression or access
cluster 1
cluster 2
cluster 3
Spc25
STU2
Nuf2 Ska2
EB1 INCENP1
INCENP2 Spc24
b
early bud
FBXO1 IMC53
BCC3 MNK1
IMC29 AC9
IMC32
d
alveolin IMCs
IMC7
IMC12
IMC1 IMC6
IMC3 IMC9
IMC4 IMC10
IMC5 IMC15
IMC8
IMC11
IMC13
IMC14
c
Basal Complex (BC)
FIKK
BCC1 BCC9
BCC2 BCC10
BCC4 MORN1
BCC5 Cen2
DIP13
DHHC14 HAD2a IMC9 MyoJ
IMC5 IAP1 BCC6
IMC8 MyoI BCC7
IMC13 GAP60
MyoB/C
BCC8
MSC1a
CDPK6
f
glideosome
ELC1
GAP50
ELC2
MyoA
GAP45
e
GAPMs
GAPM1b
GAPM1a
GAPM2a
GAPM2b
GAPM3
time (hrs)
g
IMC sutures
ISC4
ISC5
ISC3a
TSC4
ISC7
TSC1 TSC5
TSC2 TSC6
TSC3 ISC1
ISC2
ISC3b
ISC6
cluster 4

Supplementary Fig. 2. Representative chromatin accessibility and expression patterns of functional groups. Profiles are automatically clustered based on their expression patterns (left) and shown together with their corresponding chromatin accessibility profiles plotted on the right.

a. M-phase genes

b. Early bud genes

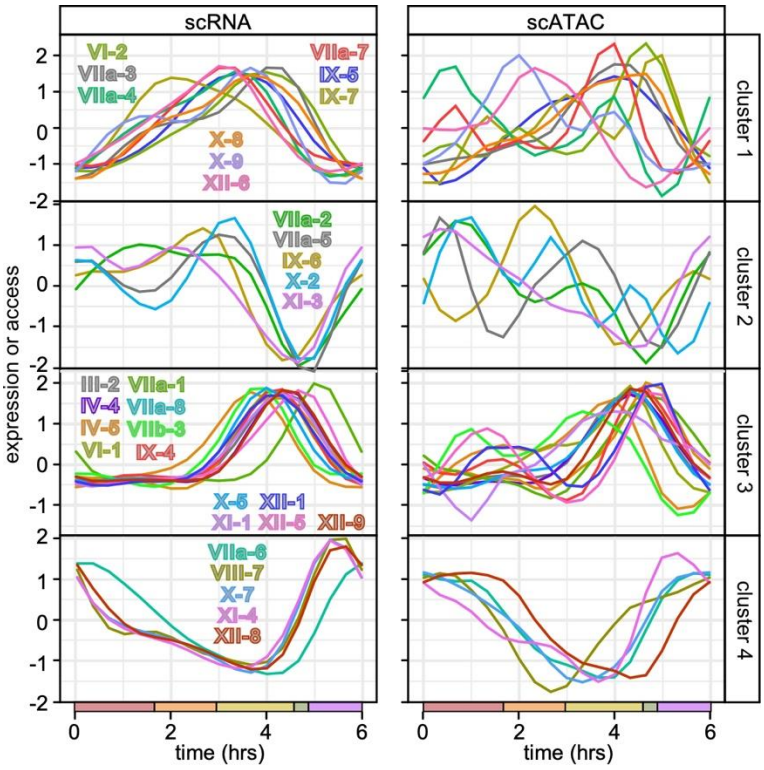
c. Basal complex (BC) genes. BC composition and function change through the cell division cycle[22], which is reflected in the expression patterns.

d. Alveolin domain containing IMC proteins. Alveolin gene expression patterns match their experimentally validated sequential assembly into the IMC[62].

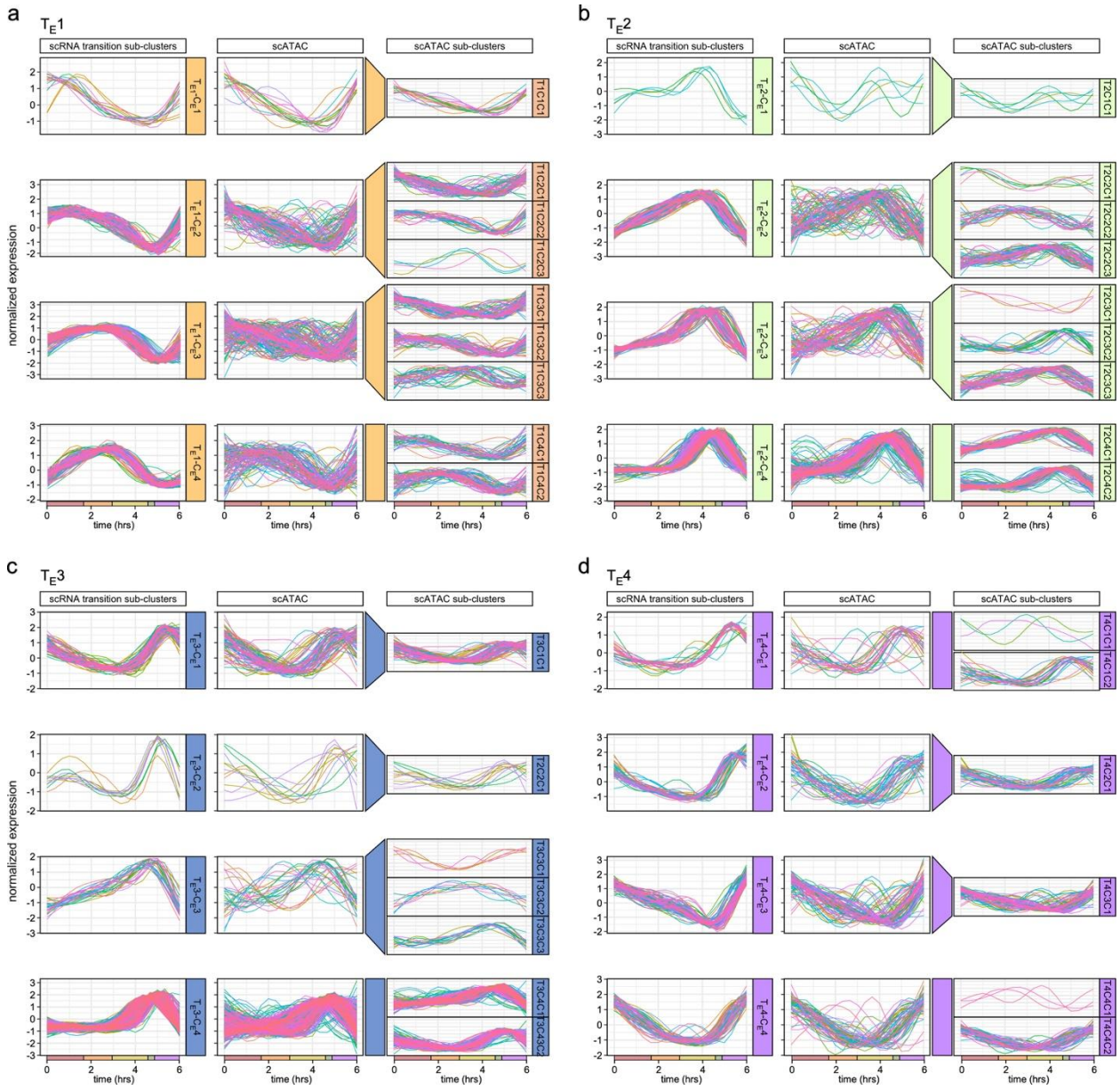
e. Gliding-associated membrane proteins (GAPM) genes connect the IMC to the subpellicular microtubules[63].

f. Glideosome associated genes.

g. IMC suture genes. The IMC sutures are found in between the alveolar vesicles making up the cortical membrane skeleton[24,25].



Supplementary Fig. 3. Chromatin accessibility and expression patterns of the 32 cyclic AP2 TFs. Profiles are automatically clustered (clusters 1-4) based on their expression patterns (left), and shown together with their corresponding chromatin accessibility profiles plotted on the right.



Supplementary Fig. 4. Clustering analyses of genes in transition expression clusters T_E1-4 based on their expression and chromatin accessibility profiles. DEGs in each expression-derived transition cluster **(Fig. 2f)**, T_E1-4 **(a-d)**, were separated into four sub-clusters (C_E1-4) based on the shapes of their scRNA-seq time-resolved expression profiles, shown on the left. The corresponding scATAC-plots for the genes in each scRNA transition sub-cluster are shown in the middle, which were subsequently automatically sub-clustered on the far right.

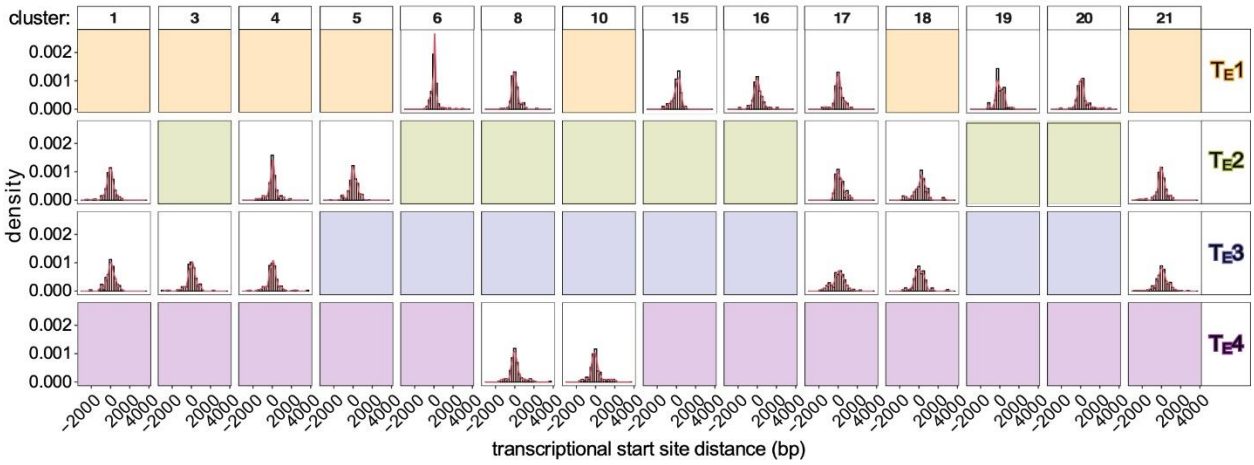
a

cluster #	consensus motif sequence	transcript. cluster
6	AP2XII-8	T_E1
8		
15		
16	AP2XII-8	
17		
19		
20		

cluster #	consensus motif sequence	transcript. cluster
1		T_E2
4	AP2XI-5	
5		
17		
18		
21	AP2XI-5	

cluster #	consensus motif sequence	transcript. cluster
1		T_E3
3		
4	AP2XI-5	
17		
18		
21	AP2XI-5	
8		T_E4
10	SAG1-promoters	

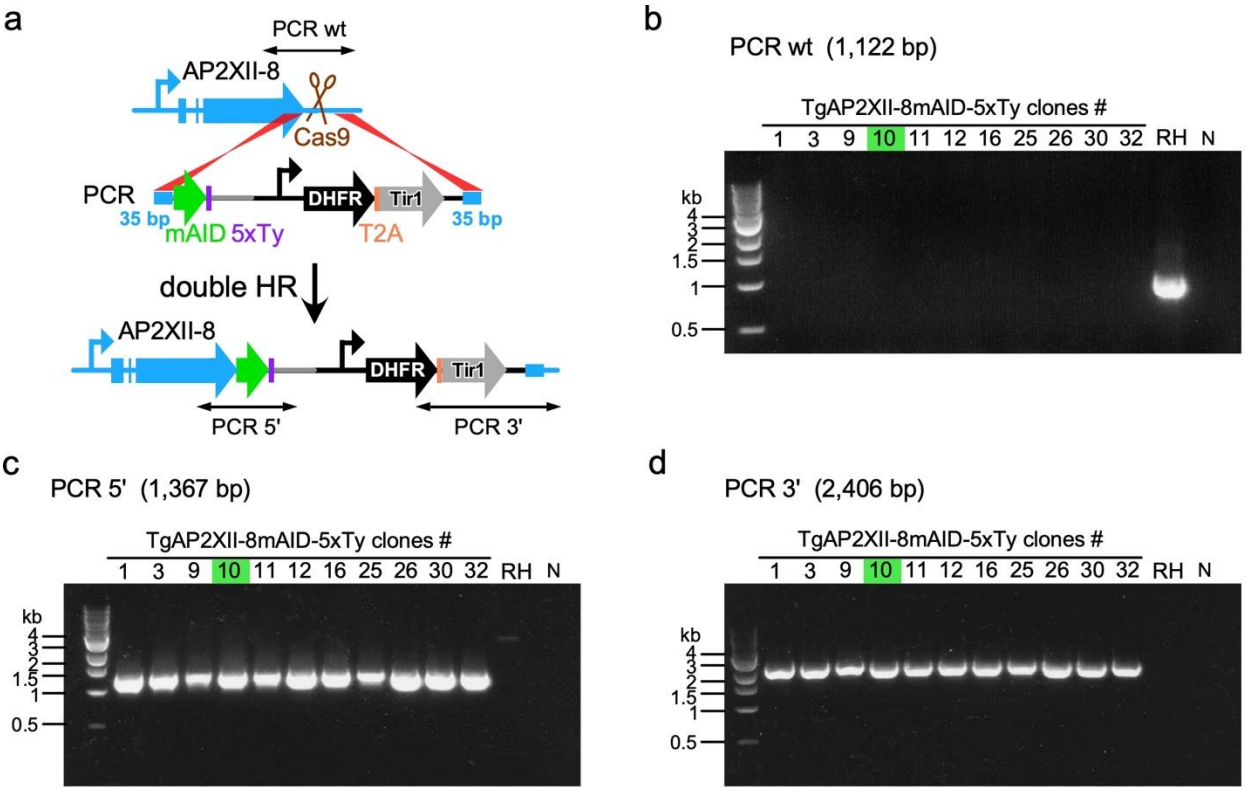
b



Supplementary Fig. 5. Motif mapping for co-expressed/co-accessible genes in expression transition clusters T_E1-4.

a. Fourteen motifs were identified for the DEGs in each expression-derived transition cluster **(Fig. 2f)** under the ATAC-peaks. These included several previously mapped motifs recognized by specific AP2 TFs as indicated.

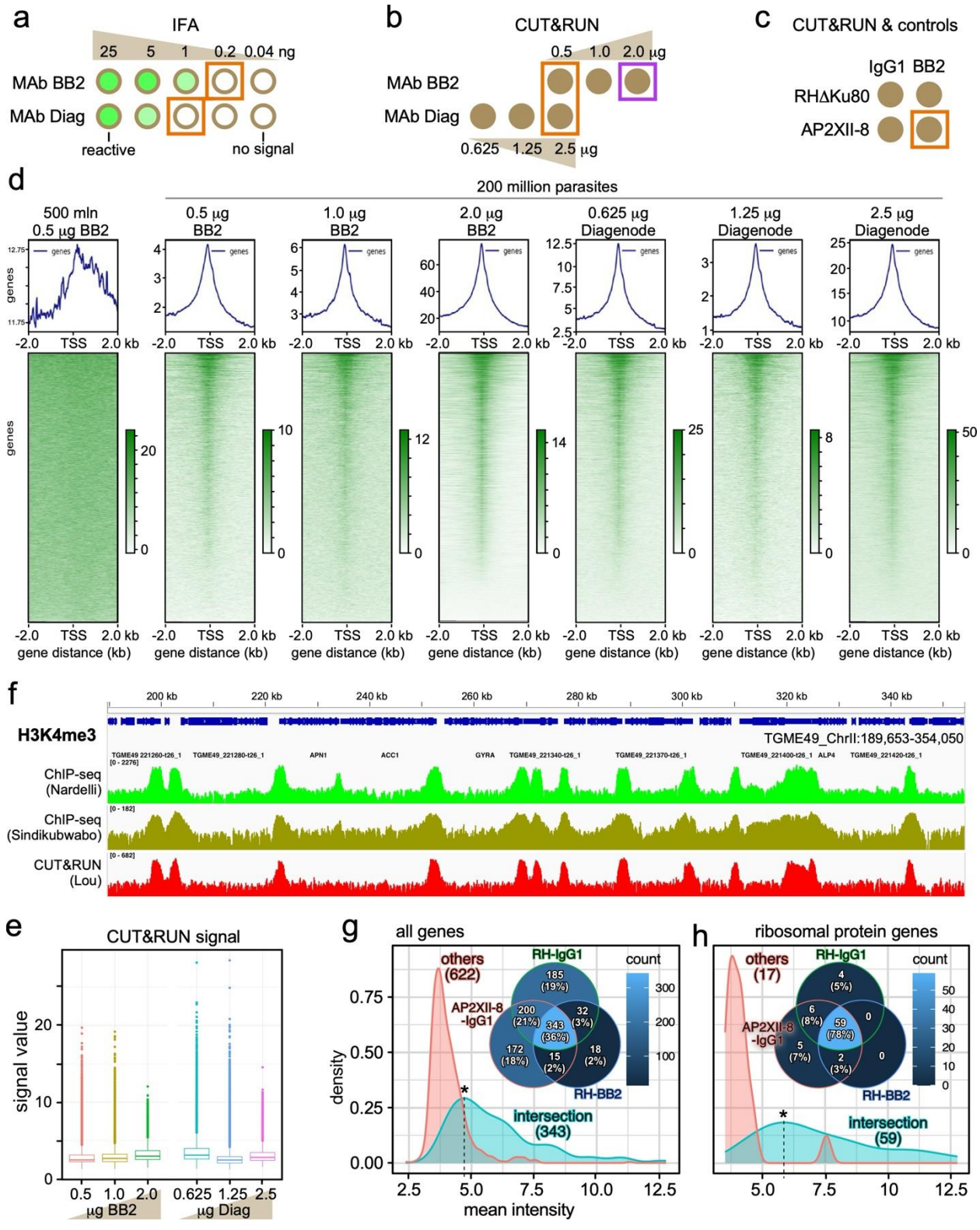
b. Each motif was mapped on the genes where it was found; the transcription start site (TSS) is set at 0 and plots show up to 4000 bp up- or down-stream as relevant for each factor.



Supplementary Fig. 6. AP2XII-8-mAID-5xTy/Tir1 generation and PCR validation.

a. Schematic representation of the one step strategy that simultaneously tags AP2XII-8 at the C-terminus with mAID-5xTy and introduces the DHFR selectable marker fused to the Tir1 open reading frame through a self-cleaving Thosea asigna virus 2A (T2A) peptide. Double homologous recombination (HR) is facilitated by 35 bp homologous flanks and a CRISPR/Cas9 generated dsDNA break.

b-d. Diagnostic PCR reactions as indicated on single parasite clones (numbered), RHΔ*Ku80* (RH) parent strain, and a no DNA negative (N) control. Expected lengths as indicated. Primer pair localizations are marked in panel **a** using the following primer pairs: PCR wt: primers P4 and P5; PCR 5': primers P4 and P7; PCR 3': primers P5 and P6 (see **Supplementary Data 8** for primer sequences). Clone #10 (green) was used in all further experiments.



Supplementary Fig. 7. CUT&RUN optimization, controls, and analysis of AP2XII-8 target genes and RPs.

a. Two different Ty-epitope specific monoclonal antibodies (MAb) were first compared for affinity by an IFA antibody dilution series on AP2XII-8-mAID-5xTy expressing parasites. The orange boxed dilutions defined the first value below the detection limit. BB2 is the standard antiserum used to detect Ty in IFA and western blots, whereas the 'Diag' antibody is marketed for Ty based ChIP experiments, which is in experimental principle comparable to CUT&RUN.

b. The two MAbs were used in two different dilution series for CUT&RUN above and beyond the value recommended by the manufacturer. The orange boxed region is functionally equivalent as derived from the dilution series in panel **a**. The purple boxed value was selected for subsequent experiments.

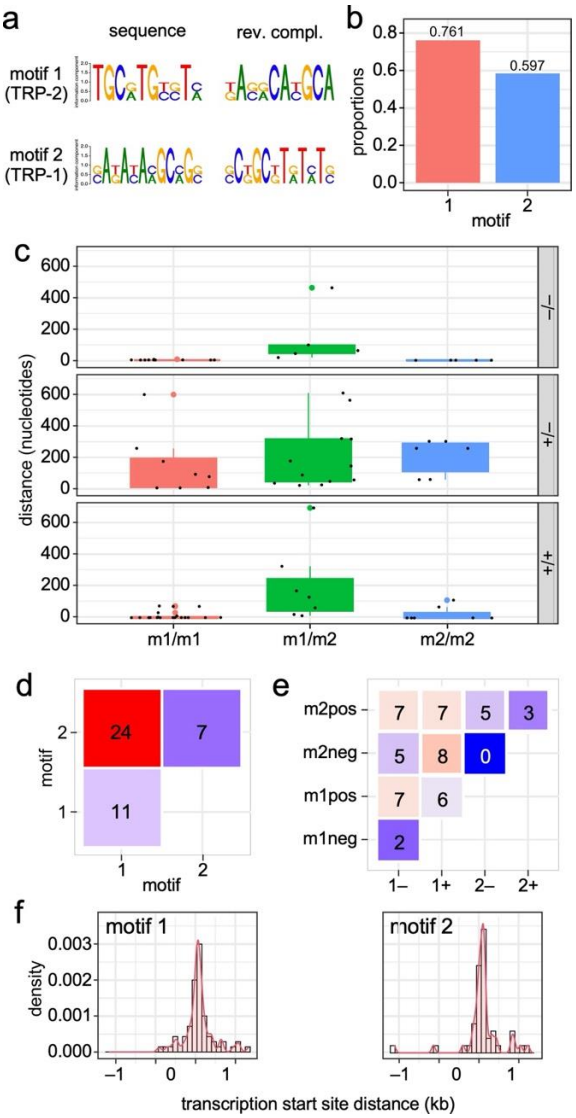
c. Experimental design of the definitive CUT&RUN experiments, including the negative controls. 2.0 μg BB2 MAb or IgG1 isotype control was used on 200 million parent (RHΔKu80) or AP2XII-cKD parasites harvested from asynchronously replicating cultures.

d. Heatmap of the CUT&RUN reads assembled around the TSS for the various controls and experimental samples under the conditions as indicated. Gradient bars represent signal intensity.

e. Signal value plots for CUT&RUN. Box-plot representing the 75 percentile of distributions of signal values per CUT&RUN data. Whisker indicates min/max and dots represent outliers. Horizontal bars indicate the median values.

f. Representative integrative genomics viewer (IGV) tracks of H3K4me3 of three independent data sets: ChIP-seq by Nardelli *et al*[36]; ChIP-seq by Sindikubwabo *et al* [14]; CUT&RUN by Lou *et al* (this study).

g, h. Venn diagrams of the unique genes of AP2XII-8 vs the indicated controls mapped to each CUT&RUN sample comparison as indicated. Panel **g** comprises all AP2XII-8 CUT&RUN target genes (965), while panel **h** comprises only the ribosomal genes (76). The density profiles display the mean intensity profiles of the loosely controlled genes (only genes found in all controls excluded) vs the most tightly (middle intersection in the Venn diagram; any gene detected in any of the negative controls excluded) selected CUT&RUN genes. Compare peaks marked with asterisks for the right-shifted peak of extended ribosomal gene mean density.



Supplementary Fig. 8. Motif identifications and co-occurrence analysis under the CUT&RUN peaks of all RPs.

a. Motifs identified under the CUT&RUN peaks of all RPs with signal over any of the negative sets. Motif-1 and Motif-2 are previously reported as *T. gondii* TRP-2 and TRP-1, respectively[37].

b. Proportion of RPs having the indicated motif.

c. Distance of successive motifs identified under the same CUT&RUN peaks. Facets indicate the strand on which the successive motifs were identified. Box plots represent 75 percentiles; whisker shows the standard deviation; the dots represent the actual data points.

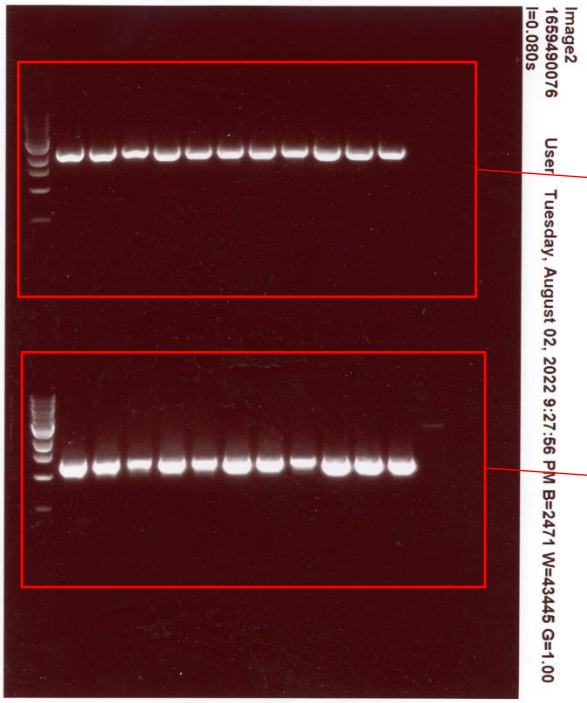
d. Heatmap shows motif co-occurrence.

e. Heatmap shows motif co-occurrence as in **d**, but considering the strand of the motif.

f. Motif distance to the TSS of the ribosomal genes.

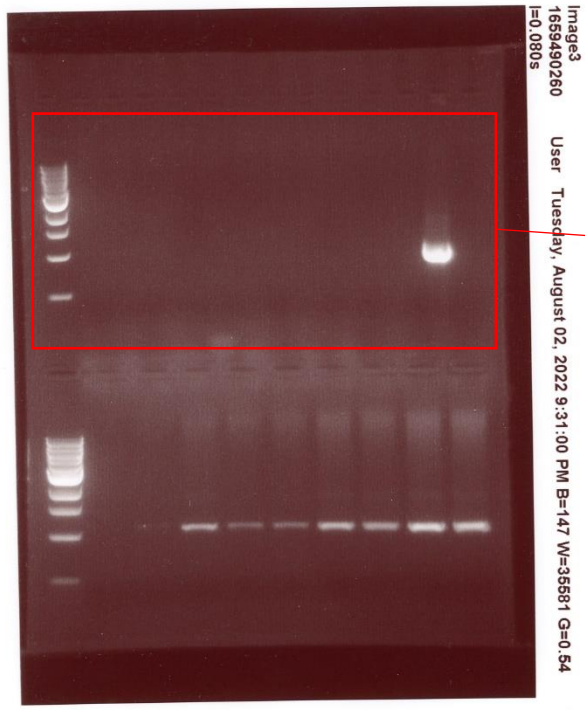
Source Data

Supplementary Fig. 6



Supplementary Fig. 6d

Supplementary Fig. 6c



Supplementary Fig. 6b

Additional References

62 Anderson-White, B. R. *et al.* A family of intermediate filament-like proteins is sequentially assembled into the cytoskeleton of Toxoplasma gondii. *Cell Microbiol* **13**, 18-31 (2011).

63 Harding, C. R. *et al.* Alveolar proteins stabilize cortical microtubules in Toxoplasma gondii. *Nat Commun* **10**, 401 (2019). https://doi.org/10.1038/s41467-019-08318-7