1 # Single cell transcriptomic analysis of bloodstream form *Trypanosoma brucei*
2 # reconstructs cell cycle progression and differentiation via quorum sensing
3
4 Emma M. Briggs[1,2*], Richard McCulloch[2], Keith R. Matthews[1], Thomas D. Otto[2]
5
6 **Author information**
7 These authors contributed equally: Keith R. Matthews and Thomas D. Otto
8
9 Corresponding author
10 *Correspondence to Emma M. Briggs (emma.briggs@ed.ac.uk)
11
12 Affiliations:
13 1. Institute for Immunology and Infection Research, School of Biological Sciences, University
14 of Edinburgh, Edinburgh, UK
15 2. Wellcome Centre for Integrative Parasitology, Institute of Infection, Immunity and
16 Inflammation, University of Glasgow, Glasgow, UK
17
18 **Abstract**
19
20 The life cycles of African trypanosomes are dependent on several differentiation steps,
21 where parasites transition between replicative and non-replicative forms specialised for
22 infectivity and survival in mammal and tsetse fly hosts. Here, we use single cell
23 transcriptomics (scRNA-seq) to dissect the asynchronous differentiation of replicative
24 slender to transmissible stumpy bloodstream form *Trypanosoma brucei*. Using oligopeptide-
25 induced differentiation, we accurately modelled stumpy development *in vitro* and captured
26 the transcriptomes of 9,344 slender and stumpy stage parasites, as well as parasites
27 transitioning between these extremes. Using this framework, we detail the relative order of
28 biological events during development, profile dynamic gene expression patterns and
29 identify putative novel regulators. Using marker genes to deduce the cell cycle phase of
30 each parasite, we additionally map the cell cycle of proliferating parasites and position
31 stumpy cell cycle exit at early G1, with subsequent progression to a distinct G0 state. We
32 also explored the role of one gene, ZC3H20, with transient elevated expression at the key
33 slender to stumpy transition point. By scRNA-seq analysis of ZC3H20 null parasites exposed
34 to oligopeptides and mapping the resulting transcriptome to our atlas of differentiation, we
35 identified the point of action for this key regulator. Using a developmental transition
36 relevant for both virulence in the mammalian host and disease transmission, our data
37 provide a paradigm for the temporal mapping of differentiation events and regulators in the
38 trypanosome life cycle.
39
40 **Introduction**
41
42 African trypanosome parasites cause both human[1] and animal[2] trypanosomiases and are
43 transmitted between hosts across sub-Saharan Africa by tsetse flies. During its life cycle
44 *Trypanosoma brucei* undergoes several developmental transitions, comprising changes in
45 nutrient-specific metabolism, morphology, organelle organisation and structure, and stage-
46 specific surface protein expression[3], facilitating parasite survival and transmission. In the
47 mammalian host, long slender bloodstream forms replicate extracellularly, increasing in

48 numbers to trigger differentiation into short stumpy bloodstream form parasites via a
49 quorum sensing (QS) process[4,5], with ill-defined intermediate forms between these
50 morphological extremes[6,7]. Stumpy forms remain arrested in the cell cycle[8] until ingested by
51 a feeding tsetse fly, where they are pre-adapted to survive in the midgut[9,10]. Here, stumpy
52 forms undergo a further differentiation event and re-enter the cell cycle as tsetse-midgut
53 procyclic forms[9,11].

55 Slender and stumpy forms differ at both the transcript[12–17] and protein level[18,19], as do
56 stumpy and procyclic parasites[15–17,19]. Reflecting their metabolism, slender forms show high
57 levels of transcripts encoding glycosomal components (specialist organelles housing
58 glycolytic enzymes)[9], whereas stumpy parasites upregulate transcripts related to a maturing
59 mitochondrion as they prepare for the tsetse midgut. This allows for metabolism of
60 pyruvate, as well as proline and threonine, to generate ATP in low glucose conditions[9,13–
61 15,20]. Consistent with exit from the cell cycle, stumpy parasites down-regulate histone, DNA
62 replication/repair, translation and cytoskeleton-related transcripts [15]. Additionally, PAD
63 (Proteins Associated with Differentiation) transcripts are upregulated in stumpy forms and
64 are required for further development into procyclics[21]. Transcripts encoding EP and GPEET
65 repeat procyclin surface proteins expressed in tsetse midgut forms are also elevated in
66 stumpy forms, whereas variant surface glycoprotein (VSGs) transcripts, required for immune
67 evasion by the parasite in the mammal, are reduced. Transcript analysis of *T. brucei*
68 parasites isolated during parasitaemia *in vivo* suggested some of these changes occur in
69 early differentiating parasites, before morphologically detectable stumpy forms dominate at
70 the peak of parasitemia[22].

72 QS based development between slender and stumpy forms has been recently characterised,
73 identifying several factors involved in detecting the differentiation stimulus[23], signal
74 propagation[24,25] and implementation of cellular changes[24,26–28]. Yet, understanding the
75 detailed developmental progression toward stumpy cells has been hampered by the
76 asynchrony of this differentiation step, as has the relationship of differentiation regulatory
77 genes to the various biological events involved. However, single-cell RNA sequencing
78 (scRNA-seq) offers the opportunity to study individual cells in a heterogenous population, to
79 identify rare cell types and decipher complex and transient developmental processes[29–31].
80 Recently, scRNA-seq has been used to study antigenic variation in *T. brucei*[32], as well as to
81 describe the diversity of parasites in the tsetse fly salivary gland[33]. The latter study revealed
82 early and late stages of metacyclic development, previously indistinguishable by population-
83 based RNA-seq[34], highlighting the differing expression of surface proteins within the
84 developing population[33].

86 Here, we applied scRNA-seq to analyse 9,344 differentiating parasites progressing from
87 bloodstream slender, through intermediate, to stumpy parasites *in vitro* using oligopeptide-
88 rich bovine brain heart infusion (BHI) broth[23], deriving a temporal map of this transition at
89 the transcript level based on individual cells. Detailed analysis of the associated expression
90 patterns revealed the absence of a discrete "intermediate" transcriptome, mapped the
91 relative timing of biological events during differentiation, including exit from the cell cycle
92 specifically prior to late G1, and identified novel genes regulated during the transition.
93 Moreover, scRNA-seq analysis of a null mutant for one important regulator elevated at the
94 slender to stumpy transition, ZC3H20[26,27], precisely mapped where development fails in its

95 absence in molecular terms. In combination, this provides a paradigm for the temporal
96 mapping of developmental events and regulators during the parasite's dynamic
97 differentiation programme in its mammalian host.
98
99 **Results**
100
101 **scRNA-seq identifies transcriptionally distinct long slender and short stumpy form *T.***
102 ***brucei***
103
104 To model stumpy development *in vitro*, pleomorphic *T. brucei* EATRO 1125 AnTa1.1 90:13
105 slender parasites were treated with oligopeptide-rich BHI broth, able to induce *T. brucei*
106 bloodstream form differentiation in a titratable manner[23]. In the presence of 10% BHI,
107 parasites underwent growth arrest (Fig. S1a), increased expression of the stumpy marker
108 protein PAD1[35] (Fig. S1b), and increased the percentage of parasites containing one copy of
109 the nucleus and one copy of the kinetoplast network (1N1K), indicating cell cycle
110 accumulation in G1/G0 and differentiation into stumpy forms[8] (Fig. S1c). After 72 hr, 72.5%
111 of cells expressed PAD1 (Fig. S1b) and 89.3% were in the 1N1K cell cycle configuration (Fig
112 S1c). To capture the transcriptomes of slender, intermediate and stumpy *T. brucei*, we
113 combined parasites after 0, 24, 48 or 72 hr of 10% BHI treatment in equal numbers. 15,000
114 cells of this heterogenous pool were then subjected to scRNA-seq using the Chromium
115 Single Cell 3' workflow (10X Genomics) and Illumina sequencing[36]. Two independent
116 biological replicates (WT 1 and WT 2) were generated and, after filtering to remove
117 transcriptomes of poor quality or likely doublets, 9,344 cells remained (5,793 and 3,551,
118 respectively; Fig. S2 and Supplementary data 1). The transcripts of 8,757 genes were
119 captured in at least 5 cells in both replicate experiments (10 cells total), with medians of
120 1,051 and 1,439 genes detected per cell, respectively. Cells from the two replicate
121 experiments were integrated and UMAP (Uniform Manifold Approximation and
122 Projection[37]) was used to visualise the relationship between individual *T. brucei*
123 transcriptomes in low dimensional space, where variation between transcriptomes dictates
124 the space between cells (Fig. 1a, b, d). Cells from WT 1 and WT 2 experiments overlapped
125 (Fig. 1a), indicating the capture of reproducible cell types in each replicate. Clustering
126 analysis identified four distinct groups containing transcriptionally similar cells (Fig. 1b;
127 Supplementary data 2), each appearing in comparable proportions per replicate (Fig. 1c).
128
129 Slender- and stumpy-like cells were clearly identifiable by the expression of marker genes:
130 slender-associated glycolytic genes GAPDH and PYK1[38]; and the stumpy markers, PAD2[21]
131 and EP1 procyclin[13] (Fig. 1d). Differential expression analysis of transcripts between the
132 slender A, slender B, stumpy A and stumpy B clusters identified 519 marker genes (adjusted
133 p-value < 0.05, logFC > 0.25); relative expression of the top unique markers is plotted in Fig.
134 1e. Markers of slender A and slender B overlapped extensively (Supplementary data 2); 178
135 slender marker genes were upregulated in both slender A and slender B relative to stumpy
136 clusters, including glycolytic genes (for example, hexokinase, phosphoglycerate kinase,
137 glucose-6-phosphate isomerase), genes involved in cytoskeleton organisation (e.g. beta
138 tubulin, cytoplasmic dynein 2 heavy chains 1 and 2), cell cycle regulating genes (e.g. cyclin
139 10, cyclin-like F-box protein 2 and T-complex protein 1 subunits gamma and delta) and RNA-
140 binding protein 10 (RBP10), which is a positive regulator of bloodstream form
141 transcripts[39,40]. Markers unique to slender A or slender B (183 and 95, respectively) were

142 generally related to the cell cycle phase of these cells and are discussed in detail below. The
143 top differentially expressed markers of slender A were FAZ2 (flagellum attachment zone
144 protein 2), CPC2 (chromosomal passenger complex 2) and histone H2B (Fig 1e and f).
145 Glycosomal membrane protein gim5B, a putative S-adenosylhomocysteine hydrolase
146 (AdoHcyase) and uridine phosphorylase (UPP) were the top markers of slender B (Fig. 1e
147 and f). 55 genes were upregulated in stumpy A cells relative to the other clusters, including
148 purine nucleoside transporter NT10 (known to be associated with stumpy forms[41,42]),
149 succinyl-CoA synthetase alpha subunit (SCS-alpha), and succinyl-CoA ligase [GDP-forming]
150 beta-chain (SUCLG2) (Fig. 1e and f). Just 9 genes significantly distinguished stumpy B cells,
151 including four encoded by the mitochondrial genome: cytochrome oxidase subunits I-III
152 (COI, COII, COIII) and NADH dehydrogenase subunit 1 (ND1) (Fig. 1e,f).

154 Gene ontology (GO) term enrichment analysis revealed the association of each cluster's
155 marker genes with distinct biological processes (Fig. 1g). Several terms relating to cell cycle
156 processes were enriched in slender A marker genes: organelle and cilium organisation,
157 chromosome segregation, and cell division. Cell cycle regulation genes, including Cytokinesis
158 Initiation Iactor 1 (CIF1)[43] and 14-3-3 protein 1 (14-3-3-I)[44], glycosylphosphatidylinositol-
159 specific phospholipase C (GPI-PLC), involved with GPI anchor release[45], and a positive
160 regulator of differentiation, ZC3H20[26,27,46], were all upregulated in slender A and slender B
161 cells compared to stumpy cells. Slender B was also associated with the genes implicated in
162 quorum sensing, including protein phosphatase 2C (PP2C) and Trichohyalin[24], as well as
163 genes putatively involved in cell communication (Ras-related protein Rab5A[47] and thimet
164 oligopeptidase). Stumpy A marker genes were associated with the TCA cycle (mitochondrial
165 malate dehydrogenase, two 2-oxoglutarate dehydrogenase E1 component encoding genes,
166 and succinyl-CoA ligase [GDP-forming] beta-chain[48]), oxidation-reduction process
167 (dihydrolipoyl dehydrogenase[4,49–51] and glutamate dehydrogenase[52]), rRNA metabolism
168 (splicing factor TSR1[53], nucleolar RNA-binding protein NOPP44/46-1[54–56] and Lupus LA
169 protein homolog[57]), and cell differentiation (zinc finger protein 2; ZFP2[58]). GO term analysis
170 of stumpy B marker genes was limited due to their small number but included post-
171 transcriptional regulators of gene expression, due to the presence of ZC3H11, which is also
172 involved in heat shock response[59], and PAD2, which is involved in detecting the
173 differentiation to procyclic forms stimulus via citrate transport[21].

175 Taken together, the above clustering analysis revealed distinct slender and stumpy clusters,
176 with significant variation within each population. Interestingly, a distinct cluster
177 representative of a discrete "intermediate" stage transcriptome between slender and
178 stumpy forms was not evident.

180 **Trajectory analysis of long slender to short stumpy differentiation**

182 As clustering analysis highlighted the transition from slender and stumpy cells involved
183 overlapping gene expression and GO term association, we conducted trajectory inference
184 and pseudotime analysis to study gene expression changes during stumpy development in
185 detail. Individual cells were replotted as a PHATE (Potential of Heat-diffusion for Affinity-
186 based Transition Embedding) map (Fig. 2a-c), which captures the local and global structure
187 of high-dimensional data to preserve the continual progression of developmental
188 processes[60]. Here, slender A and slender B clusters remained clearly separate, whereas

189   stumpy A and stumpy B showed more extensive overlap (Fig. 2a). A linear trajectory starting
190   from slender A cells was identified (Fig. 2b). Slender and stumpy marker gene (GAPDH,
191   PYK1, PAD2 and EP1) expression across the trajectory confirmed capture of the transition
192   from slender to stumpy forms (Fig. 2c). 2001 genes were identified as differentially
193   expressed as a function of pseudotime (p value < 0.05, fold change > 2) and were grouped
194   into 9 modules (A-I) of co-expressed genes that showed similar patterns of expression
195   across differentiation (Fig. 2d; Supplementary data 2). Of these, 1,335 genes were
196   previously found to be significantly (adjusted p-value < 0.05) differentially expressed
197   between slender and stumpy enriched populations of *T. brucei* isolated from low and peak
198   parasitaemia *in vivo*, respectively[12] , confirming the physiological relevance of *in vitro*,
199   oligopeptide induced, differentiation (Fig. 2e). Proportionally fewer genes in modules A
200   (transiently down regulated) and F (transiently upregulated) had been identified in bulk
201   RNA-seq data as differentially expressed (33.3% and 49%, respectively) compared to the
202   remaining modules (59.1 - 78.9%), highlighting the ability of single cell analyses to reveal
203   transient events in an asynchronous developmental trajectory.
204
205   GO term enrichment for biological processes associated with each gene module revealed
206   the relative order of biological events during slender to stumpy development (Fig. 2f).
207   Processes upregulated at the start of the trajectory included chromosome segregation and
208   regulation of cytokinesis (module B), and these were then followed by cell division (modules
209   C and D), indicating the progression of the later stages of the cell cycle. Module D also
210   included genes linked to stimulus response, including PKA-R (previously identified as a
211   potential signal transducer during stumpy development in response to hydrolysable-
212   cAMP[24]), cAMP-specific phosphodiesterases 1 and 2 (PDEB1 and PDEB2), the known
213   differentiation regulator RBP7B[24,25], DHFR-TS, an enzyme required for thymine synthesis[61],
214   and mitotic cyclin CYC8[62]. Module E genes, which are broadly expressed across the slender
215   parasites and peaked in expression slightly later in pseudotime, include GPI-PLC, which
216   releases the VSG coat via hydrolysis of the GPI-anchor[45], and genes associated with
217   cytoskeleton organisation (beta tubulin, actin A, cytoskeleton associated proteins CAP51V
218   and CAP5.5V[63,64]). Module F consisted of transiently upregulated genes, including the
219   known stumpy and procyclic developmental regulator ZC3H20[26,27,46], and MCP1, a pyruvate
220   transporter present in the mitochondrial membrane[65]. Module G-I genes peaked in the later
221   stages of development and include further genes identified in a reverse-genetic screen for
222   stumpy development factors: the chromatin regulator ISWI, a phosphoglycerate mutase
223   protein, KRIPP14, APPBP1, and hypothetical proteins Tb927.11.300 and Tb927.11.1640[24].
224   Module G included genes encoding components of the TCA cycle: mitochondrial chaperone
225   BCS1, and four ATP synthase genes (ATPF1A, ATPB and mitochondrial ATP synthase delta
226   chain[48]). Protein coding genes associated with rRNA metabolism (n = 31) and ribosome
227   biogenesis (n = 44) were also upregulated in the later stages of development, including 20
228   large ribosomal subunit components and 2 rRNA methyltransferases. These might be
229   related to the translational preparedness exhibited by quiescent stumpy forms for
230   development to procyclic forms and the resumption of translation[22]. Six genes linked to cell
231   cycle arrest were identified in module H, all encoding copies of retrotransposon hotspot
232   protein RHS4[66]. kDNA-encoded genes RPS12, ND1, COI-III, and NDH4[67], were all increased in
233   the later stages of development (modules G-I). Cytochrome B (Cyb), which was not sorted
234   into a gene module, peaked at the very end of development (Fig. 2g). Beyond these
235   annotated genes, 635 hypothetical genes were identified as differentially expressed during

236 slender to stumpy differentiation, including predicted gene regulators[68] Tb927.8.7820
237 (transiently upregulated), Tb927.8.4190 (transiently down regulated), and Tb927.11.11680,
238 which peaks in stumpy cells (Fig. 2g).
239
240 Pseudotime analysis was able to identify novel genes differentially expressed during
241 bloodstream form differentiation, as well as each gene's detailed expression pattern. The
242 relative timing of events, from proliferation (chromosome segregation, cytokinesis), cell
243 cycle exit, cell remodelling, through to a maturing mitochondrion and expression of
244 procyclin surface protein transcripts, can be inferred from these expression patterns.
245 Additionally, the expression peaks of known and putative developmental regulators were
246 identified relative to this progression.
247
248 **Transcript abundance during the bloodstream slender cell cycle**
249
250 As replicating slender bloodstream form cells were captured in these experiments, we next
251 asked if the scRNA-seq data could reveal greater detail than currently available on gene
252 expression changes during the cell cycle. We first assigned each cell to a cell cycle phase
253 using marker genes previously identified by bulk RNA-seq analysis[69], (Fig. 3a, S3). Slender A
254 and slender B cells were clearly grouped closer to cells of the same phase, with the parasites
255 most distal to the stumpy A and stumpy B cells labelled as late G1, followed by S and G2/M
256 phase cells. Slender B cells most proximal to stumpy A contained all four cell cycle phases,
257 although early G1 cells were enriched here. Stumpy A and stumpy B cells were marked as a
258 variety of all the cell cycle phases, indicating the phase of stumpy parasites cannot be clearly
259 identified with these markers and stumpy parasites do not clearly reside with the G1 phase
260 of actively replicating parasites.
261
262 Slender A and slender B clusters were next isolated and low dimensional projection of the
263 data was repeated, resulting in these cells organising into a circular profile (Fig. 3b). Cells
264 were clearly arranged by their cell cycle status within this ring, with late G1, S and G2/M
265 phase cells distinguishable. Although early G1 cells were enriched between G2/M and late
266 G1 cells, as expected, they overlapped with both neighbouring phases, indicating this stage
267 was less well defined by the scoring. To assess gene expression changes during the cell
268 cycle, we fitted a cyclical trajectory to this plot and assigned pseudotime values (Fig. 3c).
269 Testing each gene for expression patterns associated with pseudotime identified 1,897 that
270 were significantly (adj. p-value < 0.05) differentially expressed (Supplementary data 3). 854
271 of these changed in abundance by more than 2-fold (Fig. 3d). GO term enrichment revealed
272 expected GO terms associated with the cell cycle; 33 genes associated with the term "cell
273 cycle process" are highlighted in Fig. 3d. Amongst these 33 genes were several where
274 protein levels or distribution have been shown to match the scRNA-seq predicted cell cycle
275 timing of expression, including ORC1B[70], AUK1[71], PCNA[72] and KKIP5[73]. Many further genes
276 displayed cell cycle regulated expression that has not yet been explored (Supplementary
277 data 3). For instance, two cyclin genes were identified, each with a distinct expression
278 profile; CYC4 transcripts were more abundant before the appearance of late G1 and S phase
279 cells and CYC6 peaked post-late G1. CYC8 peaked specifically during late G1, as previously
280 documented by bulk RNA-seq analysis of cell cycle sorted populations[69] (Fig. 3e).
281

282 In addition to genes driving the cell cycle, we identified three genes previously shown to be
283 involved in stumpy development with differential expression patterns in slender cells (Fig.
284 3f). RBP7B[24] expression increased in late G1 cells and persisted though to G2/M. PPC2, in
285 contrast, showed the opposite pattern of expression, deceasing in late G1/S phase
286 parasites. ZC3H20[26,27,46] is highly abundant at the transcript level, yet dropped in expression
287 in late G1/S phase *T. brucei*.

288
289 **ZC3H20 null parasites fail to differentiate in response to BHI**

290
291 The above analysis highlighted differential expression patterns of several stumpy
292 development regulators, including ZC3H20, which peaks in expression at the slender B to
293 stumpy transition in pseudotime (Fig. 2g). As ZC3H20 has been previously shown to be
294 required for differentiation *in vivo* and *in vitro* at high density[26,27], we repeated scRNA-seq
295 analysis with a ZC3H20 null *T. brucei* line[27] to investigate where parasites fail in their
296 development to stumpy forms with respect to transcriptome changes and, potentially, to
297 identify direct or indirect mRNA targets of ZC3H20 itself. We first tested the effect of 10%
298 BHI broth on ZC3H20 null *T. brucei* parasites (ZC3H20 KO)[27] (Fig. 4a-c). Although the growth
299 rate was reduced in the presence of 10% BHI broth, ZC3H20 KO parasites continued to
300 replicate after WT cells arrested (Fig 4a) and, after 72 hr of culture with 10% BHI, ZC3H20 KO
301 parasites failed to express the PAD1 protein (Fig. 4b). Additionally, we tested the ability of
302 WT and ZC3H20 KO parasites exposed to BHI for 72 hr to differentiate into procyclic cells.
303 Consistent with their inability to generate stumpy forms, after 3 hr of cis-aconitate
304 treatment and incubation at 27 °C, none of ZC3H20 KO parasites expressed EP1 procyclin, in
305 contrast with 84.34% of WT parasites, confirming ZC3H20 KO *T. brucei* fail to differentiate
306 into functional stumpy cells when exposed to BHI broth (Fig. 4c).

307
308 ZC3H20 KO parasites were next cultured in 10% BHI for 0, 24, 48 or 72 hr and subjected to
309 scRNA-seq, as described for WT samples. After quality control filtering, 2,294 cells (median
310 1,051 genes per cell) remained and were integrated with the WT cells before dimensional
311 reduction was performed and results were plotted as UMAPs (Fig. 4d, g, h). Clustering the
312 ZC3H20 KO and WT integrated cells resulted in six distinct clusters: stumpy A and stumpy B,
313 and four slender clusters, slender A.1, slender A.2, slender B.1, slender B.2 (Fig. 4d, e). These
314 were identified as slender- and stumpy-like cells by the localised expression of marker genes
315 GAPDH, PYK1, PAD2 and EP1 (Fig. 4g). Whereas 77.3% of WT cells were found in clusters
316 stumpy A or stumpy B, only 0.3% of ZC3H20 KO cells were in either, consistent with near
317 complete ablation of stumpy formation in the mutant parasites (Fig. 4f). The majority of WT
318 slender parasites grouped as members of the slender A.2 and slender B.1 clusters (9.1% and
319 11.2% of all parasites, respectively), whereas ZC3H20 KO cells were divided between the
320 four slender clusters. Notably, the slender B.2 cohort was comprised almost entirely of
321 ZC3H20 KO parasites (comprising 40.1% of total ZC3H20 KO vs 0.4% of total WT cells).
322 Marker gene analysis between clusters (Fig. S3, Supplementary data 4), identified 94 marker
323 genes upregulated in slender B.2 cells. Of these just 18 genes were uniquely significantly
324 upregulated in cluster slender B.2, the top 5 genes being expression site-associated genes 2
325 and 5 (ESAG2 and ESAG5), one non-coding RNA gene (Tb2.NT.8), a serine peptidase (Clan
326 SC), Family S10 protein CBP1, and DNA topoisomerase II beta (TOP2B) (fig. 4h). Cell cycle
327 status contributed considerably to clustering analysis of WT and ZC3H20 KO integrated cells,
328 as slender cells clearly group by cell cycle phase (Fig. 4i).

329

Trajectory comparison between WT and ZC3H20 KO cells reveals functional separation of downregulation and upregulation of transcripts during differentiation

To compare the transcriptomic changes in ZC3H20 KO and WT *T. brucei* after BHI treatment in detail, we inferred a trajectory from the WT and ZC3H20 KO integrated parasites (Fig. 5a). Doing so identified a branched trajectory: early in pseudotime, WT and ZC3H20 KO parasites were transcriptionally similar and arranged on the same lineage; later, there was a clear branch in their comparative development, ending for WT in stumpy cells and in slender B.2 for ZC3H20 KO cells (Fig. 5a). To understand this change, we first assessed the expression of differentiation-associated genes identified previously in WT parasites (Fig 2d), mapping them across the truncated trajectory branch of the ZC3H20 KO cells (Fig 5b, c). 587 genes of the 2001 identified as differentially expressed during stumpy development in WT cells, significantly changed in expression in ZC3H20 KO parasites across the truncated trajectory (Fig. 5c), although the majority (94.2%) were less highly associated with the ZC3H20 KO trajectory relative to WT (Fig 5d; Supplementary data 4). 75.8% of these genes were part of co-expression modules B-E, which decreased in expression during stumpy development in WT parasites (Fig. 5c). These included genes involved with glycolysis, such as ATP-dependent 6-phosphofructokinase (PFK) and hexokinase 1 (HK1), and the mitotic cell cycle, including cdc2-related kinase 3 (CRK3) and aurora B kinase (AUK1) (fig. 5e). Genes differentially expressed in the ZC3H20 KO trajectory and belonging to expression modules F-H included heat shock 70 kDa protein mitochondrial precursor subunits B and C, and three components of the TCA cycle (succinyl-CoA ligase, mitochondrial malate dehydrogenase and 2-oxoglutarate dehydrogenase E1 component; 2-OGDH E1), but only 2-OGDH E1 increased to a similar level as seen in WT cells (Fig 5e). Hence, ZC3H20 KO cells down regulated transcripts associated with slender cells when exposed to the BHI differentiation stimulus, matching the response of WT cells. However, ZC3H20 KO parasites failed to upregulate transcripts later in development that are required for stumpy formation, and this point of dysregulation coincided with the peak of ZC3H20 expression during normal WT differentiation (Fig. 2g).

To identify regulators of early stumpy development, we looked for genes which changed significantly in abundance from the start of the trajectory to a point just downstream of the ZC3H20 branch (yellow dots, Fig. 5a). 234 genes changed in transcript abundance between these points and were associated with trajectory progression (Fig. 5f, Supplementary data 4). 117 of these genes were associated with both WT and ZC3H20 KO trajectories (p-value <0.05) and include cell cycle associated genes (for example CPC2, Kinetoplastid-specific Protein Phosphatase 1, and structural maintenance of chromosome 4), as expected. Additionally, differentiation associated genes RDK2[28] and PAD2[35] were associated with both trajectories but showed different patterns of expression (Fig 5g). 83 genes were differentially expressed only early in the trajectory of WT parasites. These genes include eight relating to ribosome biogenesis (including Midasin, PUF RNA binding protein 10, a putative RNA 3'-terminal phosphate cyclase, and the U3 snoRNA-associated protein UTP11). Other genes include pyruvate dehydrogenase E1 beta subunit, chromatin modifier NLP, TFIIH basal transcription factor complex helicase subunit, Exocyst complex component EXO99, Ras-related proteins RAB7 and RAB2B, and genes encoding hypothetical genes implicated in posttranscriptional regulation of gene expression (Tb927.6.2650,

376   Tb927.11.830 and Tb927.11.7590). 35 genes with early altered expression were associated
377   with the truncated ZC3H20 KO development only. These include G2/M regulator WEE1,
378   kinetoplastid-specific dual specificity phosphatase, Tb927.9.9990, and putative regulator of
379   post-transcriptional gene expression, Tb927.8.3780.
380
381   In summary, comparing the differentiation of WT and differentiation incompetent ZC3H20
382   KO cells through scRNA-seq has allowed the identification of a) the direct and indirect
383   targets of ZC3H20 altered specifically during differentiation, b) the failure point of ZC3H20
384   KO cells during the temporal profile of differentiation, and c) putative 'immediate early'
385   regulators of differentiation.
386
387   **Discussion**
388
389   Although extensively studied, *T. brucei* differentiation from slender to stumpy bloodstream
390   forms has remained difficult to dissect in detail due to the asynchronous nature of this life
391   cycle transition. Here, we used oligopeptide induction of differentiation[23] in combination
392   with scRNA-seq to deconvolve this process at the transcript level. This approach revealed
393   several details of this process, including: the lack of a discrete intermediate transcriptome;
394   the precise timing of cell cycle exit, immediately prior to late G1; the transient expression of
395   several genes not identified by bulk-analysis; and the expression timing of known and
396   putative differentiation factors during the developmental processes. Using scRNA-seq to
397   study ZC3H20 KO parasites, we were also able to validate the essentiality of ZC3H20 for
398   differentiation and position its action specifically at the major slender to stumpy transition
399   point where the transcripts of this gene peak in abundance. Additionally, we were able to
400   provide detailed gene expression patterns of both known and novel cell cycle regulated
401   genes during the slender cell cycle.
402
403   Clustering WT *T. brucei* into groups of transcriptionally similar cells clearly identified two
404   primary groups of slender- and stumpy parasites, each of which could be further classified
405   into two sub-slender and sub-stumpy clusters (Fig 1). The transcript differences between
406   clusters were mainly due to the cell cycle phase of slender cells and the stage of progression
407   towards stumpy development (Fig 1 and 2) with gene expression changes highlighting a
408   progressive transition to stumpy forms with relatively few genes transiently changing in
409   abundance during development (n = 96, FC >2). Thus, although discrimination of parasites
410   between the extremes of the slender and slender morphotypes is possible microscopically,
411   scRNA-seq analysis does not provide evidence for an intermediate form defined in
412   molecular terms. Rather, short stumpy cells appear to emerge directly from the G1 phase of
413   replicative slender cells (see below).
414
415   Previous bulk transcriptomics identified marker genes of the *T. brucei* cell cycle phases
416   (early and late G1, S phase and G2/mitosis)[69], allowing us to define the most likely cell cycle
417   position of each cell (Fig 3). Phase identification of the cycling parasites showed that late G1
418   stage cells are positioned at the start of the differentiation trajectory, consistent with a cell
419   cycle receptive window[74], followed by S and G2/M phase cells. Early G1 cells were enriched
420   closer to stumpy clusters, though less clearly grouped, and stumpy clusters showed no clear
421   cell cycle phase (Fig. 3, 4i, and 4j). Hence, the major switch in transcriptome from slender to
422   stumpy occurs during G1 and, specifically, before cells enter late G1. Cells at other stages

423 may be committed to differentiation but not yet arrested, as predicted by modelling[75]. The
424 inability to assign stumpy cells to any one proliferative cell cycle phase indicates these cells
425 persist in a distinct G0, as opposed to simply pausing in G1. The signalled progression of *T.*
426 *brucei* into G0 as stumpy forms provides a valuable, evolutionary divergent and tractable
427 model for studying the conservation of quiescence signalling pathways, which are critical in
428 many eukaryotic developmental processes[76–80].

430 We were additionally able to investigate the changes in transcript abundance during the
431 proliferative slender cell cycle. Pseudotime analysis allowed us to profile the dynamic
432 patterns of 1,897 genes found to be differentially expressed (Fig. 3), 332 of which had been
433 previously identified in bulk RNA-seq analysis of synchronised procyclic *T. brucei*[69]. For
434 example, we find CYC4, CYC6 and CYC8 peak at distinct points: CYC4 is a CYC2-like protein
435 predicted to act in G1[81], where the transcript levels increase in our analysis; CYC6[62,82,83] is
436 known to regulate nuclear division and peaks post S-phase; and predicted, but untested,
437 mitotic cyclin CYC8 peaks very precisely in late G1/S phase cells. Having not undergone
438 either selectional or chemical synchronisation procedures[69,84,85], this scRNA-seq derived cell
439 cycle atlas provides a relatively unperturbed picture of cell cycle regulated events in greater
440 detail than previously available, and suggests candidates for functional analysis. Distinctions
441 between developmentally competent (pleomorphic) slender forms and adapted
442 monomorphic forms used in previous studies may also be identified.

444 Trajectory inference and differential expression analysis of the slender to stumpy transition
445 revealed the relative order of events during differentiation of this asynchronous population.
446 Initially, there was higher abundance of transcripts linked to proliferation (Fig. 1 and 2), with
447 cells completing the later stages of the cell cycle during the early stages of the
448 differentiation trajectory, consistent with phase scoring analysis (Fig. 2 and 3). Thereafter,
449 metabolic changes and activation of the mitochondrion occurred as expected[12,38]. Finally
450 expression of several kDNA encoded genes (cytochrome oxidase subunits I-III and
451 cytochrome B) and procyclin surface protein encoding genes, EP1, EP2 and GPEET was
452 observed, reflecting preparation for differentiation to procyclic forms[12,15,17]. These changes
453 correlated well with bulk mRNA analysis of *in vivo* parasites[12], validating the use of BHI as
454 an *in vitro* model of stumpy development. Transient expression patterns of several genes,
455 not discernible in bulk RNA-seq and proteomic studies, were also observed. These included
456 several hypothetical genes, which may prove to be negative or positive regulators of
457 differentiation. For instance, Tb927.8.7820 peaked precisely at the slender to stumpy
458 transition point, is able to decrease mRNA stability[68] and is negatively targeted by RBP10[86],
459 which in our analysis decreased during differentiation consistent with its reported function
460 in maintaining the bloodstream cell state[39,40,86]. Other transiently regulated transcripts were
461 associated with cell cycle or mitochondrial control, including CYC6, required for nuclear
462 division[62,82,83], CIF1, a master regulator of cytokinesis[43] and MtHSP70B and MtHSP70C,
463 which locate to the mitochondrial outer membrane[87] and may facilitating protein folding
464 and targeting in the developing mitochondrion, matching the function of the human
465 homolog, HSPA9[88].

467 Transiently upregulated genes also included the known differentiation regulator
468 ZC3H20[26,27,46], confirming that we were able to identify developmental regulators via their
469 gene expression patterns. We postulated that scRNA-seq analysis may enable us to map

470   cells with diverse differentiation phenotypes onto our trajectory of WT differentiation, to
471   assess the point at which genetically perturbed parasites fail to develop. We therefore
472   exposed ZC3H20 KO parasites[27] to oligopeptides and confirmed that they remained
473   proliferative (Fig. 4i, j) and failed to develop to stumpy forms (Fig. 4 d-h). Trajectory
474   inference revealed that ZC3H20 KO cells downregulate transcripts also downregulated in
475   oligopeptide stimulated WT parasites, including several glycolysis factors, cell cycle
476   regulating genes, and posttranscriptional regulators of gene expression (Fig. 5). This
477   downregulation may contribute to the reduced growth of ZC3H20 KO parasites when
478   exposed to BHI. Interestingly, however, there was a clear distinction between this
479   downregulation of slender transcripts and the increase of stumpy-associated transcripts,
480   which ZC3H20 KO parasites failed to upregulate to WT levels, including 12/19 ZC3H20
481   regulated mRNAs[26]. This suggests that ZC3H20 KO parasites perceive the differentiation
482   signal and undergo early steps of differentiation, but do not commit to cell cycle exit and
483   further development to stumpy forms. Comparing the divergence of WT and ZC3H20 KO
484   cells in the trajectory of differentiation identified putative 'immediate early' regulators of
485   commitment (Fig. F), including 28 hypothetical genes, four of which are known to regulate
486   mRNA stability[68]. Further experimental work will be required to test the involvement of
487   these genes in differentiation.

489   In summary, our data demonstrate that transcript level changes in parasites could be used
490   to compile maps of both the cell cycle and the asynchronous slender to stumpy
491   differentiation process. These can be mined to identify regulatory genes of individual events
492   that make up each process. We further characterised mutant parasites by the same
493   approach, positioning the site of action of one regulator (ZC3H20) in the developmental
494   time course. If iterated for different genes, this method can be exploited to derive
495   hierarchies of gene action during differentiation in this and other life cycle stages, species
496   and development processes.

498   **Materials and methods**

500   ***Trypanosoma brucei* cell lines and culture**
501   Trypanosoma brucei EATRO 1125 AnTat1.1 90:13 parasites[89] were used as pleomorphic
502   wild-type (WT) in all experiments. The ZC3H20 KO null parasites were previously generated
503   in the same cell line transfected with plasmid pJ1399 (gifted by Dr. Jack Sunter), containing
504   T7 polymerase and CRISPR/cas9, by replacement of both alleles of Tb927.7.2660 with
505   blasticidin S deaminase[27]. All parasites were grown free from selective drugs in HMI-9
506   medium[90] (Life technologies), supplemented with 10% foetal calf serum at 37°C, 5% CO2.
507   For induction of differentiation, parasites were maintained below ~7x10^5 cells per ml for up
508   to 5 days prior to addition of brain heart infusion (BHI) broth (Sigma Aldrich).

510   **Single cell RNA-sequencing**

512   For each scRNA-seq sample, four staggered cultures were set up over four days all
513   maintained below ~8x10^5 cells per ml during the experiment by dilution. One culture was
514   maintained free from BHI, and the remaining had 10% BHI added 24, 48 or 72 hr prior to
515   sample preparation. Equal numbers of parasite from each culture were then combined to
516   generate one pooled sample. 1.5 ml of the pooled culture was centrifuged, and the pelleted

517 cells washed twice with ice-cold 1 ml 1X PBS supplemented with 1% D-glucose (PSG) and
518 0.04% Bovine Serum Albumin (BSA). Cells were then resuspended in ~ 500 ul PSG + 0.04%
519 BSA, filtered with 40 μm Flowmi™ Tip Strainer (Merck) and adjusted to 1,000 cells/μl. In all
520 steps, cells were centrifuged at 400 x g for 10 minutes. 15,000 cells (15 μl) from the mixed
521 sample were loaded into the Chromium Controller (10x Genomics) to capture individual
522 cells with unique barcoded beads. Libraries were prepared using the Chromium Single Cell 3'
523 GEM, Library & Gel Bead Kit v3 (10x Genomics). Sequencing was performed with the
524 NextSeq™ 500 platform (Illumia) to a depth of ~50,000 reads per cell. Library preparation
525 and sequencing was performed by Glasgow Polyomics. For the first WT replicate
526 experiment, *T. brucei* parasites were mixed 1:1 with *Leishmania mexicana* prepared by the
527 same method (data unpublished), so the heterogenous doublet rate of 8.04% could be
528 calculated.
529
530 **Read mapping and transcript counting**
531
532 The reference genome was complied with Cell Ranger, to combined the TREU927 *T. brucei*
533 nuclear reference genome[91] and *T. brucei* EATRO 1125 maxicircle kDNA sequence[67]. 3' UTR
534 annotations were extended to increase the proportion of reads correctly assigned to
535 annotated transcripts. 2500 bp immediately downstream of the stop code was assigned as
536 the 3'UTR of each protein coding gene, unless the existing 3'UTR was longer than 2500 UTR
537 in which case the full length was preserved. If the new 3'UTR was overlapped with other
538 genome features (coding and non-coding) the UTR was truncated to remove the overlap.
539 Reads were mapped and unique reads aligned to each annotated gene were counts and
540 assigned to a cell barcode with the Cell Ranger count function (Supplementary data 1). Cell
541 Ranger v3.0.2 (http://software.10xgenomics.com/single-cell/overview/welcome) was used
542 with all default settings.
543
544 **Data processing and integration**
545
546 Count data for individual samples (WT 1, WT 2 and ZC3H20 KO) was processed separately
547 prior to integration using the Seurat v3[92] and Scran v1.14.5[93] packages with R v3.6.1. The
548 percentage of transcripts encoded on the maxi circle kDNA was calculated per cell, as cells
549 with excess proportion of mitochondrial transcripts are likely to be poor quality[94]. The
550 percentage of transcripts per cell encoding ribosomal RNA was also calculated, as high levels
551 of rRNA indicate poor capture of polyadenylated transcripts. Low quality cells were removed
552 by filtering for low total RNA (<1,000), low unique transcripts (< 250), high proportion of
553 kDNA (> 2%) and high proportion of rRNA (>8%). Likely doublets were removed by filtering
554 for high total RNA (> 4,000) and high total unique transcripts counts (> 2,500). After filtering
555 rRNA transcripts were removed from each cell's transcriptome. For sample metrics, see
556 Supplementary data 1.
557
558 Each filtered sample was log normalised individually using the quick cluster method from
559 Scran[95]. To increase the robustness of variable genes selected for principle component (PC)
560 analysis, we used two selection methods[96]; Scran, which uses log normalised transcript
561 counts, and Seurat[92], which uses raw transcript counts. We identified 3,000 genes with each
562 method, selected those identified by both and removed VSG encoding genes[97] to avoid

563   clustering based on VSG expression. This left 2,029, 1,643 and 2,132 for WT 1, WT 2 and
564   ZC3H20 KO samples, respectively (Supplementary data 1).

565

566   For integration of WT replicate samples, the Seurat v3 package was used[92]. Common
567   variable features and integration anchors were identified, data for all genes integrated and
568   scaled before the PCs were calculated using the common variable features. The first 8 PC
569   dimensions each contributed > 0.1% of additional variance and were used to select anchors
570   and integrate data. The effect of total RNA per cell was regressed when scaling data. The
571   ZC3H20 KO cells were subsequently integrated with the previously integrated WT data using
572   the steps described above, however STACAS v1.01.1 was used to identify integration
573   anchors as the package is specialist for samples which don't fully overlap[98].

574

575   **Cluster analysis and mark gene identification**

576

577   For clustering and marker gene analysis the Seurat v3 package was used[92]. Cells were
578   plotted as dimensionality reduced UMAPs[37] and nearest neighbours were identified using 8
579   dimensions. A range of clustering resolutions were trialled, with 0.4 resulting in the highest
580   resolution clustering with significant mark genes identified for every cluster. Marker genes
581   were identified for each cluster using MAST[99]. Only genes expressed in > 25% of the cells in
582   the cluster, with a logFC of > 0.25 and adjusted p-value < 0.05 were considered marker
583   genes. Gene ontology (GO) terms concerning biological processes were identified via the
584   TriTrypDB[100] website (p <0.05) and redundant terms removed with REVIGO[101]
585   (allowed similarity = 0.5) and manually.

586

587   **Trajectory inference and pseudotime analysis**

588

589   For trajectory inference cells were plotted using PHATE maps[60] (using the same common
590   viable genes as for PCA and 8 dimensions) and trajectories were identified using slingshot[102],
591   with the slender A cluster defined as the starting point. For cell cycle analysis, a circular
592   trajectory was fitted as a principle curve[103]. To identify genes with expression patterns
593   associated with progression of the trajectory, generalise additive models were fit using the
594   tradeSeq package v1.3.18[104] with default parameters. The number of knots was tested to
595   find 6 knots provide sufficient detail for the highest number of genes without overfitting.
596   Differential expression analysis was performed with the tradeSeq associationTest function
597   using default parameters, and significant genes (p-value <0.05, FC >2) were clustered using
598   tradeSeq clusterExpressionPattern over 100 points on the trajectory. Gene clusters were
599   merged into co-expressed modules using default setting except the merging cut-off was set
600   to 0.95 to refine the number of modules from 58 to 9. For comparison with bulk RNA-seq
601   analysis, the fold-change of stumpy vs slender expression for all genes was taken from data
602   published by Silvester *et al.*[12]. All genes not found to be significant (p-value > 0.05) in bulk
603   analysis were given a fold-change value of 0 for comparison with scRNA-seq data.

604

605   **Immunofluorescence**

606

607   Parasites were fixed in 1% paraformaldehyde for 10 minutes at room temperature (RT).
608   Parasites were washed in 1X PBS and adhered to slide spread with Poly-L-lysine before
609   being permeabilised with 0.1% Igepal in 1X PBS for 3 minutes. Cells were then blocked with

610   2% BSA in 1X PBS for 45 minutes at RT, stained with primary antibody (anti-PAD1[35] 1:1,000,
611   EP1 procyclin [Cedar labs] 1:300) diluted in 0.2% BSA for 1 hr at RT. Three washes with 1X
612   PBS were performed before incubating with secondary Alexa Fluor 488 (ThermoFisher
613   Scientific) in 0.2% BSA for 1 hr at RT. Cells were washed a further three times before
614   mounting with Fluoromount G with DAPI (Cambridge Bioscience, Southern Biotech).
615   Imaging was performed with an Axioscope 2 fluorescence microscope (Zeiss) and a Zeiss
616   Plan Apochromat 63x/1.40 oil objective.
617
618   **Data availability**
619
620   Data can be sources via Supplementary Data Tables and the European Nucleotide Archive
621   with accession number PRJEB41744. Wild-type scRNA-seq data can be explored using the
622   interactive cell atlas (http://cellatlas.mvls.gla.ac.uk/TbruceiBSF/).
623
624   **Code availability**
625   Code used to perform analysis described can be accessed at GitHub
626   (https://github.com/emma23ed/Tbrucei_scRNA-seq.git).
627
628   **Extended data**
629
630   Supplementary Data 1. scRNA-seq sample metrics and variable genes.
631   Supplementary Data 2. Analysis of wild-type differentiating *T. brucei*
632   Supplementary Data 3. Cell cycle analysis of slender form *T. brucei*
633   Supplementary Data 4. Analysis ZC3H20 KO *T. brucei*
634
635
636
637

References

1. Simarro, P. P., Cecchi, G., Paone, M., Franco, J. R., Diarra, A., Ruiz, J. A., Fèvre, E. M., Courtin, F., Mattioli, R. C. & Jannin, J. G. The Atlas of human African trypanosomiasis: a contribution to global mapping of neglected tropical diseases. *Int. J. Health Geogr.* **9**, 57 (2010).

2. Giordani, F., Morrison, L. J., Rowan, T. G., DE Koning, H. P. & Barrett, M. P. The animal trypanosomiases and their chemotherapy: a review. *Parasitology* **143**, 1862–1889 (2016).

3. Fenn, K. & Matthews, K. R. The cell biology of Trypanosoma brucei differentiation. *Curr. Opin. Microbiol.* **10**, 539–546 (2007).

4. Reuner, B., Vassella, E., Yutzy, B. & Boshart, M. Cell density triggers slender to stumpy differentiation of Trypanosoma brucei bloodstream forms in culture. *Mol. Biochem. Parasitol.* **90**, 269–280 (1997).

5. Vassella, E., Reuner, B., Yutzy, B. & Boshart, M. Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. *J. Cell Sci.* **110 (Pt 2**, 2661–71 (1997).

6. Robertson, M. Notes on the polymorphism of Trypanosoma gambiense in the blood and its relation to the exogenous cycle in Glossina palpalis. *Proc. R. Soc. London. Ser. B, Contain. Pap. a Biol. Character* **85**, 527–539 (1912).

7. The morphology of the trypanosome causing disease in man in Nyasaland. *Proc. R. Soc. London. Ser. B, Contain. Pap. a Biol. Character* **85**, 423–433 (1912).

8. Shapiro, S. Z., Naessens, J., Liesegang, B., Moloo, S. K. & Magondu, J. Analysis by flow cytometry of DNA synthesis during the life cycle of African trypanosomes. **41**, 313–23 (1984).

9. Rico, E., Rojas, F., Mony, B. M., Szoor, B., MacGregor, P. & Matthews, K. R. Bloodstream form pre-adaptation to the tsetse fly in Trypanosoma brucei. *Front. Cell. Infect. Microbiol.* **3**, 78 (2013).

10. Silvester, E., McWilliam, K. R. & Matthews, K. R. The cytological events and molecular control of life cycle development of Trypanosoma brucei in the mammalian bloodstream. *Pathogens* **6**, (2017).

11. Turner, C. M., Aslam, N. & Dye, C. Replication, differentiation, growth and the virulence of Trypanosoma brucei infections. *Parasitology* **111 (Pt 3**, 289–300 (1995).

12. Silvester, E., Ivens, A. & Matthews, K. R. A gene expression comparison of Trypanosoma brucei and Trypanosoma congolense in the bloodstream of the mammalian host reveals species-specific adaptations to density-dependent development. *PLoS Negl. Trop. Dis.* **12**, e0006863 (2018).

13. Jensen, B. C., Sivam, D., Kifer, C. T., Myler, P. J. & Parsons, M. Widespread variation in transcript abundance within and across developmental stages of Trypanosoma brucei. **10**, 482 (2009).

14. Nilsson, D., Gunasekera, K., Mani, J., Osteras, M., Farinelli, L., Baerlocher, L., Roditi, I. & Ochsenreiter, T. Spliced Leader Trapping Reveals Widespread Alternative Splicing Patterns in the Highly Dynamic Transcriptome of Trypanosoma brucei. *PLoS Pathog.* **6**, e1001037 (2010).

15. Kabani, S., Fenn, K., Ross, A., Ivens, A., Smith, T. K., Ghazal, P. & Matthews, K. Genome-wide expression profiling of in vivo-derived bloodstream parasite stages and dynamic analysis of mRNA alterations during synchronous differentiation in

685   Trypanosoma brucei. *BMC Genomics* **10**, 427 (2009).

686 16.   Naguleswaran, A., Doiron, N. & Roditi, I. RNA-Seq analysis validates the use of
687   culture-derived Trypanosoma brucei and provides new markers for mammalian and
688   insect life-cycle stages. *BMC Genomics* **19**, 227 (2018).

689 17.   Queiroz, R., Benz, C., Fellenberg, K., Hoheisel, J. D. & Clayton, C. Transcriptome
690   analysis of differentiating trypanosomes reveals the existence of multiple post-
691   transcriptional regulons. *BMC Genomics* **10**, 495 (2009).

692 18.   Dejung, M., Subota, I., Bucerius, F., Dindar, G., Freiwald, A., Engstler, M., Boshart, M.,
693   Butter, F. & Janzen, C. J. Quantitative Proteomics Uncovers Novel Factors Involved in
694   Developmental Differentiation of Trypanosoma brucei. *PLoS Pathog.* **12**, e1005439
695   (2016).

696 19.   Gunasekera, K., Wüthrich, D., Braga-Lagache, S., Heller, M. & Ochsenreiter, T.
697   Proteome remodelling during development from blood to insect-form Trypanosoma
698   brucei quantified by SILAC and mass spectrometry. *BMC Genomics* **13**, 556 (2012).

699 20.   Lamour, N., Rivière, L., Coustou, V., Coombs, G. H., Barrett, M. P. & Bringaud, F.
700   Proline metabolism in procyclic Trypanosoma brucei is down-regulated in the
701   presence of glucose. *J. Biol. Chem.* **280**, 11902–11910 (2005).

702 21.   Dean, S., Marchetti, R., Kirk, K. & Matthews, K. R. A surface transporter family
703   conveys the trypanosome differentiation signal. **459**, 213–7 (2009).

704 22.   Capewell, P., Monk, S., Ivens, A., Macgregor, P., Fenn, K., Walrad, P., Bringaud, F.,
705   Smith, T. K. & Matthews, K. R. Regulation of Trypanosoma brucei Total and Polysomal
706   mRNA during Development within Its Mammalian Host. **8**, e67069 (2013).

707 23.   Rojas, F., Silvester, E., Young, J., Smith, T. K., Thompson, J., Matthews
708   Correspondence, K. R., Milne, R., Tettey, M., Houston, D. R., Walkinshaw, M. D., Pé
709   Rez-Pi, I., Auer, M., Denton, H. & Matthews, K. R. Oligopeptide Signaling through
710   TbGPR89 Drives Trypanosome Quorum Sensing Article Oligopeptide Signaling through
711   TbGPR89 Drives Trypanosome Quorum Sensing. *Cell* **176**, 306-317.e16 (2019).

712 24.   Mony, B. M., MacGregor, P., Ivens, A., Rojas, F., Cowton, A., Young, J., Horn, D. &
713   Matthews, K. Genome-wide dissection of the quorum sensing signalling pathway in
714   Trypanosoma brucei. *Nature* **505**, 681–685 (2014).

715 25.   McDonald, L., Cayla, M., Ivens, A., Mony, B. M., MacGregor, P., Silvester, E.,
716   McWilliam, K. & Matthews, K. R. Non-linear hierarchy of the quorum sensing
717   signalling pathway in bloodstream form African trypanosomes. *PLOS Pathog.* **14**,
718   e1007145 (2018).

719 26.   Liu, B., Kamanyi Marucha, K. & Clayton, C. The zinc finger proteins ZC3H20 and
720   ZC3H21 stabilise mRNAs encoding membrane proteins and mitochondrial proteins in
721   insect-form *Trypanosoma brucei*. *Mol. Microbiol.* **113**, 430–451 (2020).

722 27.   Cayla, M., McDonald, L., Macgregor, P. & Matthews, K. R. An atypical DYRK kinase
723   connects quorum-sensing with posttranscriptional gene regulation in Trypanosoma
724   brucei. *Elife* **9**, 1–52 (2020).

725 28.   Jones, N. G., Thomas, E. B., Brown, E., Dickens, N. J., Hammarton, T. C. & Mottram, J.
726   C. Regulators of Trypanosoma brucei cell cycle progression and differentiation
727   identified using a kinome-wide RNAi screen. *PLoS Pathog.* **10**, e1003886 (2014).

728 29.   Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The
729   Technology and Biology of Single-Cell RNA Sequencing. *Molecular Cell* **58**, 610–620
730   (2015).

731 30.   Potter, S. S. Single-cell RNA sequencing for the study of development, physiology and

732   disease. *Nature Reviews Nephrology* (2018). doi:10.1038/s41581-018-0021-7

733   31.   Hedlund, E. & Deng, Q. Single-cell RNA sequencing: Technical advancements and
734   biological applications. *Molecular Aspects of Medicine* (2018).
735   doi:10.1016/j.mam.2017.07.003

736   32.   Müller, L. S. M., Cosentino, R. O., Förstner, K. U., Guizetti, J., Wedel, C., Kaplan, N.,
737   Janzen, C. J., Arampatzi, P., Vogel, J., Steinbiss, S., Otto, T. D., Saliba, A.-E., Sebra, R. P.
738   & Siegel, T. N. Genome organization and DNA accessibility control antigenic variation
739   in trypanosomes. *Nature* (2018). doi:10.1038/s41586-018-0619-8

740   33.   Vigneron, A., O'Neill, M. B., Weiss, B. L., Savage, A. F., Campbell, O. C., Kamhawi, S.,
741   Valenzuela, J. G. & Aksoy, S. Single-cell RNA sequencing of Trypanosoma brucei from
742   tsetse salivary glands unveils metacyclogenesis and identifies potential transmission
743   blocking antigens. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 2613–2621 (2020).

744   34.   Savage, A. F., Kolev, N. G., Franklin, J. B., Vigneron, A., Aksoy, S. & Tschudi, C.
745   Transcriptome Profiling of Trypanosoma brucei Development in the Tsetse Fly Vector
746   Glossina morsitans. *PLoS One* **11**, e0168877 (2016).

747   35.   Dean, S., Marchetti, R., Kirk, K. & Matthews, K. R. A surface transporter family
748   conveys the trypanosome differentiation signal. *Nature* **459**, 213–217 (2009).

749   36.   Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S.
750   B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros,
751   L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P.
752   W., Hindson, C. M., *et al.* Massively parallel digital transcriptional profiling of single
753   cells. *Nat. Commun.* **8**, 1–12 (2017).

754   37.   McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and
755   Projection for Dimension Reduction. (2018).

756   38.   Vertommen, D., Van, R. J., Szikora, J.-P., Rider, M. H., Michels, P. A. M. & Opperdoes,
757   F. R. Differential expression of glycosomal and mitochondrial proteins in the two
758   major life-cycle stages of Trypanosoma brucei. **158**, 189–201 (2008).

759   39.   De Pablos, L. M., Kelly, S., Nascimento, J. D. F., Sunter, J. & Carrington, M.
760   Characterization of RBP9 and RBP10, two developmentally regulated RNA-binding
761   proteins in Trypanosoma brucei. *Open Biol.* **7**, (2017).

762   40.   Wurst, M., Seliger, B., Jha, B. A., Klein, C., Queiroz, R. & Clayton, C. Expression of the
763   RNA recognition motif protein RBP10 promotes a bloodstream-form transcript
764   pattern in Trypanosoma brucei. *Mol. Microbiol.* **83**, 1048–1063 (2012).

765   41.   Sanchez, M. A., Drutman, S., Van Ampting, M., Matthews, K. & Landfear, S. M. A novel
766   purine nucleoside transporter whose expression is up-regulated in the short stumpy
767   form of the Trypanosoma brucei life cycle. *Mol. Biochem. Parasitol.* (2004).
768   doi:10.1016/j.molbiopara.2004.04.009

769   42.   Spoerri, I., Chadwick, R., Renggli, C. K., Matthews, K., Roditi, I. & Burkard, G. Role of
770   the stage-regulated nucleoside transporter TbNT10 in differentiation and adenosine
771   uptake in Trypanosoma brucei. *Mol. Biochem. Parasitol.* (2007).
772   doi:10.1016/j.molbiopara.2007.04.006

773   43.   Zhou, Q., An, T., Pham, K. T. M., Hu, H. & Li, Z. The CIF1 protein is a master
774   orchestrator of trypanosome cytokinesis that recruits several cytokinesis regulators
775   to the cytokinesis initiation site. *J. Biol. Chem.* **293**, 16177–16192 (2018).

776   44.   Inoue, M., Nakamura, Y., Yasuda, K., Yasaka, N., Hara, T., Schnaufer, A., Stuart, K. &
777   Fukuma, T. The 14-3-3 proteins of Trypanosoma brucei function in motility,
778   cytokinesis, and cell cycle. *J. Biol. Chem.* **280**, 14085–14096 (2005).

779   45.   Carrington, M., Carnall, N., Crow, M. S., Gaud, A., Redpath, M. B., Wasunna, C. L. &
780   Webb, H. The properties and function of the glycosylphosphatidylinositol-
781   phospholipase C in Trypanosoma brucei. *Mol. Biochem. Parasitol.* **91**, 153–164 (1998).

782   46.   Ling, A. S., Trotter, J. R. & Hendriks, E. F. A zinc finger protein, TbZC3H20, stabilizes
783   two developmentally regulated mRNAs in trypanosomes. *J. Biol. Chem.* **286**, 20152–
784   20162 (2011).

785   47.   Field, H., Farjah, M., Pal, A., Gull, K. & Field, M. C. Complexity of trypanosomatid
786   endocytosis pathways revealed by Rab4 and Rab5 isoforms in Trypanosoma brucei. *J.*
787   *Biol. Chem.* **273**, 32102–32110 (1999).

788   48.   Van Hellemond, J. J., Opperdoes, F. R. & Tielens, A. G. M. The extraordinary
789   mitochondrion and unusual citric acid cycle in Trypanosoma brucei. *Biochem. Soc.*
790   *Trans.* **33**, 967–971 (2005).

791   49.   ELSE, A. J., HOUGH, D. W. & DANSON, M. J. Cloning, sequencing, and expression of
792   Trypanosoma brucei dihydrolipoamide dehydrogenase. *Eur. J. Biochem.* **212**, 423–429
793   (1993).

794   50.   Vickerman, K. Polymorphism and mitochondrial activity in sleeping sickness
795   trypanosomes. *Nature* (1965). doi:10.1038/208762a0

796   51.   Tyler, K. M., Matthews, K. R. & Gull, K. The bloodstream differentiation-division of
797   Trypanosoma brucei studied using mitochondrial markers. *Proc. R. Soc. B Biol. Sci.*
798   (1997). doi:10.1098/rspb.1997.0205

799   52.   Estévez, A. M., Kierszenbaum, F., Wirtz, E., Bringaud, F., Grunstein, J. & Simpson, L.
800   Knockout of the glutamate dehydrogenase gene in bloodstream Trypanosoma brucei
801   in culture has no effect on editing of mitochondrial mRNAs. *Mol. Biochem. Parasitol.*
802   **100**, 5–17 (1999).

803   53.   Gupta, S. K., Chikne, V., Eliaz, D., Tkacz, I. D., Naboishchikov, I., Carmi, S., Ben-Asher,
804   H. W. & Michaeli, S. Two splicing factors carrying serine-arginine motifs, TSR1 and
805   TSR1IP, regulate splicing, mRNA stability, and rRNA processing in Trypanosoma
806   brucei. *RNA Biol.* **11**, 715–731 (2014).

807   54.   Jensen, B. C., Brekken, D. L., Randall, A. C., Kifer, C. T. & Parsons, M. Species specificity
808   in ribosome biogenesis: A nonconserved phosphoprotein is required for formation of
809   the large ribosomal subunit in Trypanosoma brucei. *Eukaryot. Cell* **4**, 30–35 (2005).

810   55.   Chou, S., Jensen, B. C., Parsons, M., Alber, T. & Grundner, C. The Trypanosoma brucei
811   life cycle switch TbPTP1 is structurally conserved and dephosphorylates the nucleolar
812   protein NOPP44/46. **285**, 22075–81 (2010).

813   56.   Parsons, M., Ledbetter, J. A., Schieven, G. L., Nel, A. E. & Kanner, S. B. Developmental
814   regulation of pp44/46, tyrosine-phosphorylated proteins associated with
815   tyrosine/serine kinase activity in Trypanosoma brucei. *Mol. Biochem. Parasitol.*
816   (1994). doi:10.1016/0166-6851(94)90009-4

817   57.   Shan, F., Mei, S., Zhang, J., Zhang, X., Xu, C., Liao, S. & Tu, X. A telomerase subunit
818   homolog La protein from *Trypanosoma brucei* plays an essential role in ribosomal
819   biogenesis. *FEBS J.* **286**, 3129–3147 (2019).

820   58.   Hendriks, E. F., Robinson, D. R., Hinkins, M. & Matthews, K. R. A novel CCCH protein
821   which modulates differentiation of Trypanosoma brucei to its procyclic form. **20**,
822   6700–11 (2001).

823   59.   Droll, D., Minia, I., Fadda, A., Singh, A., Stewart, M., Queiroz, R. & Clayton, C. Post-
824   Transcriptional Regulation of the Trypanosome Heat Shock Response by a Zinc Finger
825   Protein. *PLoS Pathog.* **9**, e1003286 (2013).

60. Moon, K. R., van Dijk, D., Wang, Z., Gigante, S., Burkhardt, D. B., Chen, W. S., Yim, K., Elzen, A. van den, Hirn, M. J., Coifman, R. R., Ivanova, N. B., Wolf, G. & Krishnaswamy, S. Visualizing structure and transitions in high-dimensional biological data. *Nat. Biotechnol.* **37**, 1482–1492 (2019).

61. Sienkiewicz, N., Jarosławski, S., Wyllie, S. & Fairlamb, A. H. Chemical and genetic validation of dihydrofolate reductase-thymidylate synthase as a drug target in African trypanosomes. *Mol. Microbiol.* **69**, 520–533 (2008).

62. Li, Z. & Wang, C. C. A PHO80-like cyclin and a B-type cyclin control the cell cycle of the procyclic form of Trypanosoma brucei. *J. Biol. Chem.* **278**, 20652–20658 (2003).

63. Portman, N. & Gull, K. Identification of paralogous life-cycle stage specific cytoskeletal proteins in the parasite Trypanosoma brucei. *PLoS One* **9**, (2014).

64. Olego-Fernandez, S., Vaughan, S., Shaw, M. K., Gull, K. & Ginger, M. L. Cell Morphogenesis of Trypanosoma brucei Requires the Paralogous, Differentially Expressed Calpain-related Proteins CAP5.5 and CAP5.5V. *Protist* (2009). doi:10.1016/j.protis.2009.05.003

65. Colasante, C., Peña Diaz, P., Clayton, C. & Voncken, F. Mitochondrial carrier family inventory of Trypanosoma brucei brucei: Identification, expression and subcellular localisation. *Mol. Biochem. Parasitol.* **167**, 104–117 (2009).

66. Florini, F., Naguleswaran, A., Gharib, W. H., Bringaud, F. & Roditi, I. Unexpected diversity in eukaryotic transcription revealed by the retrotransposon hotspot family of Trypanosoma brucei. *Nucleic Acids Res.* **47**, 1725–1739 (2019).

67. Cooper, S., Wadsworth, E. S., Ochsenreiter, T., Ivens, A., Savill, N. J. & Schnaufer, A. Assembly and annotation of the mitochondrial minicircle genome of a differentiation-competent strain of Trypanosoma brucei. *Nucleic Acids Res.* **47**, 11304–11325 (2019).

68. Erben, E. D., Fadda, A., Lueong, S., Hoheisel, J. D. & Clayton, C. A Genome-Wide Tethering Screen Reveals Novel Potential Post-Transcriptional Regulators in Trypanosoma brucei. *PLoS Pathog.* **10**, e1004178 (2014).

69. Archer, S. K., Inchaustegui, D., Queiroz, R. & Clayton, C. The Cell Cycle Regulated Transcriptome of Trypanosoma brucei. *PLoS One* **6**, e18425 (2011).

70. Marques, C. A., Tiengwe, C., Lemgruber, L., Damasceno, J. D., Scott, A., Paape, D., Marcello, L. & McCulloch, R. Diverged composition and regulation of the Trypanosoma brucei origin recognition complex that mediates DNA replication initiation. *Nucleic Acids Res.* **44**, 4763–84 (2016).

71. Li, Z. & Wang, C. C. Changing roles of aurora-B kinase in two life cycle stages of Trypanosoma brucei. *Eukaryot. Cell* (2006). doi:10.1128/EC.00129-06

72. Kaufmann, D., Gassen, A., Maiser, A., Leonhardt, H. & Janzen, C. J. Regulation and spatial organization of PCNA in Trypanosoma brucei. *Biochem. Biophys. Res. Commun.* **419**, 698–702 (2012).

73. D'Archivio, S. & Wickstead, B. Trypanosome outer kinetochore proteins suggest conservation of chromosome segregation machinery across eukaryotes. *J. Cell Biol.* **216**, 379–391 (2017).

74. Matthews, K. R. & Gull, K. Evidence for an interplay between cell cycle progression and the initiation of differentiation between life cycle forms of African trypanosomes. *J. Cell Biol.* (1994). doi:10.1083/jcb.125.5.1147

75. MacGregor, P., Savill, N. J., Hall, D. & Matthews, K. R. Transmission stages dominate trypanosome within-host dynamics during chronic infections. *Cell Host Microbe* (2011). doi:10.1016/j.chom.2011.03.013

76. Orford, K. W. & Scadden, D. T. Deconstructing stem cell self-renewal: Genetic insights into cell-cycle regulation. *Nature Reviews Genetics* (2008). doi:10.1038/nrg2269

77. Heyman, J., Kumpf, R. P. & De Veylder, L. A quiescent path to plant longevity. *Trends in Cell Biology* (2014). doi:10.1016/j.tcb.2014.03.004

78. Gray, J. V., Petsko, G. A., Johnston, G. C., Ringe, D., Singer, R. A. & Werner-Washburne, M. "Sleeping Beauty": Quiescence in Saccharomyces cerevisiae. *Microbiol. Mol. Biol. Rev.* (2004). doi:10.1128/mmbr.68.2.187-206.2004

79. Sagot, I. & Laporte, D. The cell biology of quiescent yeast – a diversity of individual scenarios. *Journal of Cell Science* (2019). doi:10.1242/jcs.213025

80. de Virgilio, C. The essence of yeast quiescence. *FEMS Microbiology Reviews* (2012). doi:10.1111/j.1574-6976.2011.00287.x

81. Hammarton, T. C. Cell cycle regulation in Trypanosoma brucei. *Mol. Biochem. Parasitol.* **153**, 1–8 (2007).

82. Hayashi, H. & Akiyoshi, B. Degradation of cyclin B is critical for nuclear division in Trypanosoma brucei. *Biol. Open* (2018). doi:10.1242/bio.031609

83. Hammarton, T. C., Clark, J., Douglas, F., Boshart, M. & Mottram, J. C. Stage-specific differences in cell cycle control in Trypanosoma brucei revealed by RNA interference of a mitotic cyclin. *J. Biol. Chem.* (2003). doi:10.1074/jbc.M300813200

84. Forsythe, G. R., McCulloch, R. & Hammarton, T. C. Hydroxyurea-induced synchronisation of bloodstream stage Trypanosoma brucei. *Mol. Biochem. Parasitol.* (2009). doi:10.1016/j.molbiopara.2008.12.008

85. Benz, C., Dondelinger, F., McKean, P. G. & Urbaniak, M. D. Cell cycle synchronisation of Trypanosoma brucei by centrifugal counter-flow elutriation reveals the timing of nuclear and kinetoplast DNA replication. *Sci. Rep.* **7**, 17599 (2017).

86. Mugo, E. & Clayton, C. Expression of the RNA-binding protein RBP10 promotes the bloodstream-form differentiation state in Trypanosoma brucei. *PLoS Pathog.* (2017). doi:10.1371/journal.ppat.1006560

87. Niemann, M., Wiese, S., Mani, J., Chanfon, A., Jackson, C., Meisinger, C., Warscheid, B. & Schneider, A. Mitochondrial outer membrane proteome of trypanosoma brucei reveals novel factors required to maintain mitochondrial morphology. *Mol. Cell. Proteomics* **12**, 515–528 (2013).

88. Deocaris, C. C., Kaul, S. C. & Wadhwa, R. On the brotherhood of the mitochondrial chaperones mortalin and heat shock protein 60. in *Cell Stress and Chaperones* **11**, 116–128 (Cell Stress Chaperones, 2006).

89. Engstler, M. & Boshart, M. Cold shock and regulation of surface protein trafficking convey sensitization to inducers of stage differentiation in Trypanosoma brucei. *Genes Dev.* **18**, 2798–2811 (2004).

90. Hirumi, H. & Hirumi, K. Continuous cultivation of Trypanosoma brucei blood stream forms in a medium containing a low concentration of serum protein without feeder cell layers. *J. Parasitol.* **75**, 985–9 (1989).

91. Berriman, M. The Genome of the African Trypanosome Trypanosoma brucei. *Science (80-.).* **309**, 416–422 (2005).

92. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P. & Satija, R. Comprehensive Integration of Single-Cell Data. *Cell* (2019). doi:10.1016/J.CELL.2019.05.031

93. Lun, A. T. L., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data [version 1; referees: 5 approved with

920   reservations]. *F1000Research* **5**, (2016).

921   94.   Ilicic, T., Kim, J. K., Kolodziejczyk, A. A., Bagger, F. O., McCarthy, D. J., Marioni, J. C. &
922   Teichmann, S. A. Classification of low quality cells from single-cell RNA-seq data.
923   *Genome Biol.* **17**, 29 (2016).

924   95.   Lun, A. T. L., Bach, K. & Marioni, J. C. Pooling across cells to normalize single-cell RNA
925   sequencing data with many zero counts. *Genome Biol.* **17**, 75 (2016).

926   96.   Yip, S. H., Sham, P. C. & Wang, J. Evaluation of tools for highly variable gene discovery
927   from single-cell RNA-seq data. *Brief. Bioinform.* **20**, 1583–1589 (2018).

928   97.   Cross, G. A. M., Kim, H.-S. & Wickstead, B. Capturing the variant surface glycoprotein
929   repertoire (the VSGnome) of Trypanosoma brucei Lister 427. *Mol. Biochem. Parasitol.*
930   **195**, 59–73 (2014).

931   98.   Andreatta, M. & Carmona, S. J. STACAS: Sub-Type Anchor Correction for Alignment in
932   Seurat to integrate single-cell RNA-seq data. *Bioinformatics* (2020).
933   doi:10.1093/bioinformatics/btaa755

934   99.   Finak, G., McDavid, A., … M. Y.-G. & 2015, undefined. MAST: a flexible statistical
935   framework for assessing transcriptional changes and characterizing heterogeneity in
936   single-cell RNA sequencing data. *genomebiology.biomedcentral.com*

937   100. Aslett, M., Aurrecoechea, C., Berriman, M., Brestelli, J., Brunk, B. P., Carrington, M.,
938   Depledge, D. P., Fischer, S., Gajria, B., Gao, X., Gardner, M. J., Gingle, A., Grant, G.,
939   Harb, O. S., Heiges, M., Hertz-Fowler, C., Houston, R., Innamorato, F., Iodice, J., *et al.*
940   TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids*
941   *Res.* **38**, D457–D462 (2010).

942   101. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. Revigo summarizes and visualizes long
943   lists of gene ontology terms. *PLoS One* (2011). doi:10.1371/journal.pone.0021800

944   102. Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., Purdom, E. & Dudoit, S.
945   Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC*
946   *Genomics* **19**, 477 (2018).

947   103. Hastie, T. & Stuetzle, W. Principal curves. *J. Am. Stat. Assoc.* (1989).
948   doi:10.1080/01621459.1989.10478797

949   104. Van den Berge, K., Roux de Bézieux, H., Street, K., Saelens, W., Cannoodt, R., Saeys,
950   Y., Dudoit, S. & Clement, L. Trajectory-based differential expression analysis for
951   single-cell sequencing data. *Nat. Commun.* (2020). doi:10.1038/s41467-020-14766-3

952

953

954

Acknowledgements

956

964

Author Contributions
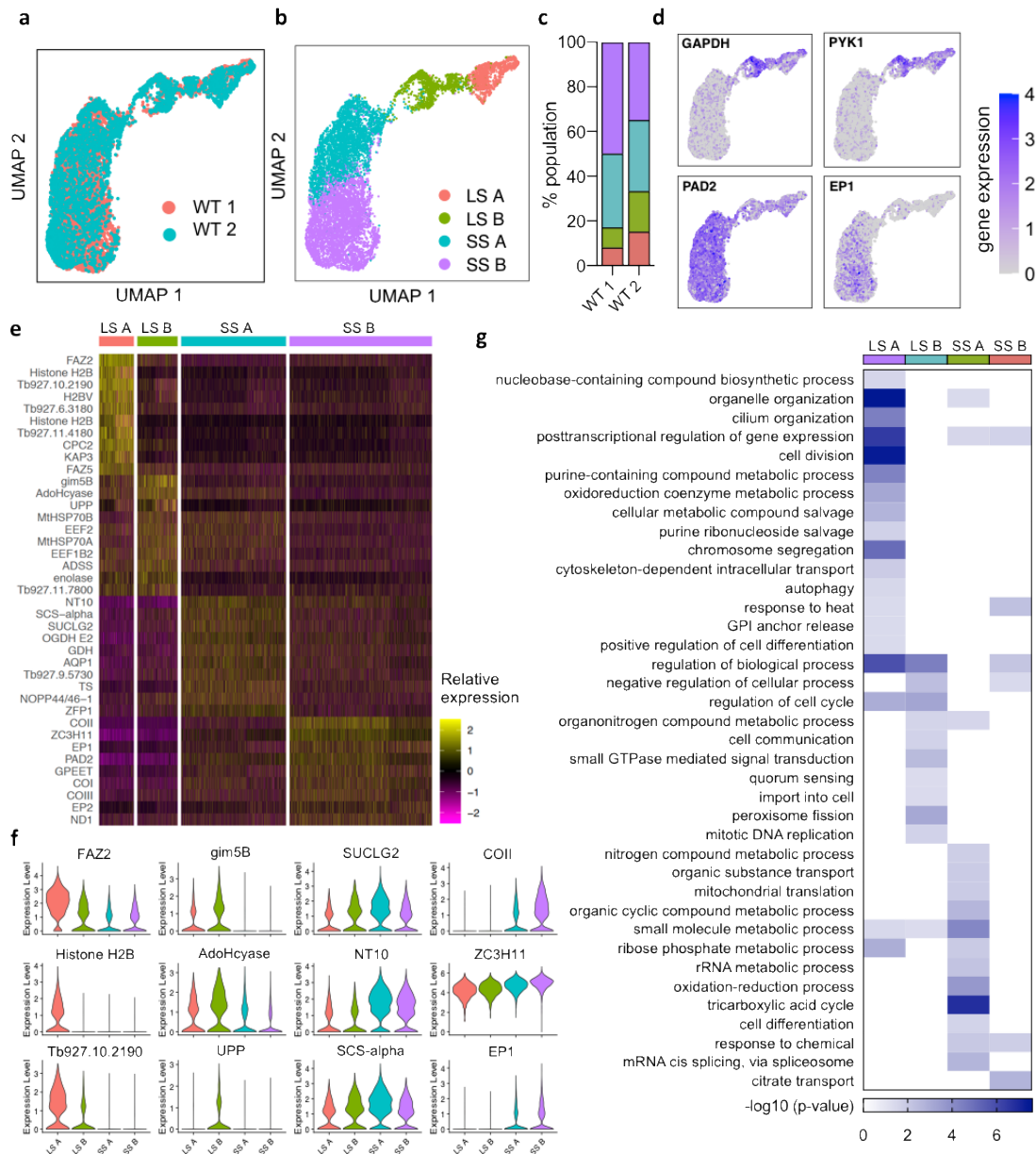
966

Methodology: E.M.B., R.M., T.D.O, and K.R.M. Data collection: E.M.B. Bioinformatic data analysis: E.M.B and T.D.O. Single cell atlas was created by T.D.O. All authors participated in discussions related to this work. All authors wrote, reviewed and approved the manuscript.

970

971

972

Figure 1. Sequencing of individual *T. brucei* transcriptomes during bloodstream differentiation *in vitro*. a). Low dimensional plot (UMAP) of each cell after filtering. Each point is the transcriptome of one cell positioned according to similarity with neighboring transcriptomes, coloured by replicate experiment. **b)** UMAP of WT parasites from both replicates, coloured by cluster: Long slender (LS) A, LS B, short stumpy (SS) A and SS B. **c)** Percentage of parasites in each cluster for each replicate experiment. **d)** UMAP of integrated WT parasites coloured by transcript counts for two SL marker genes (GAPDH; Tb927.6.4280 and PYK1; Tb927.10.14140) and stumpy marker genes (PAD2; Tb927.7.5940 and EP1; Tb927.10.10260). Scale shows raw transcript count per cell. **e)** Heatmap showing relative expression of the top 10 unique maker genes of each cluster identified in **c**. Each row is one gene coloured by relative expression. Where no gene name or symbol was available, the gene ID is shown. Each column is one cell grouped according to cluster. **f)** Violin plots showing the expression of each of the top 3 unique marker genes per cell, divided by cluster. X-axis shows the raw transcript count per cell. **g)** Gene ontology (GO) enrichment for biological process linked with maker genes for each cluster. Scale shows the -log(adjusted p-value) for each term enrichment per cluster.
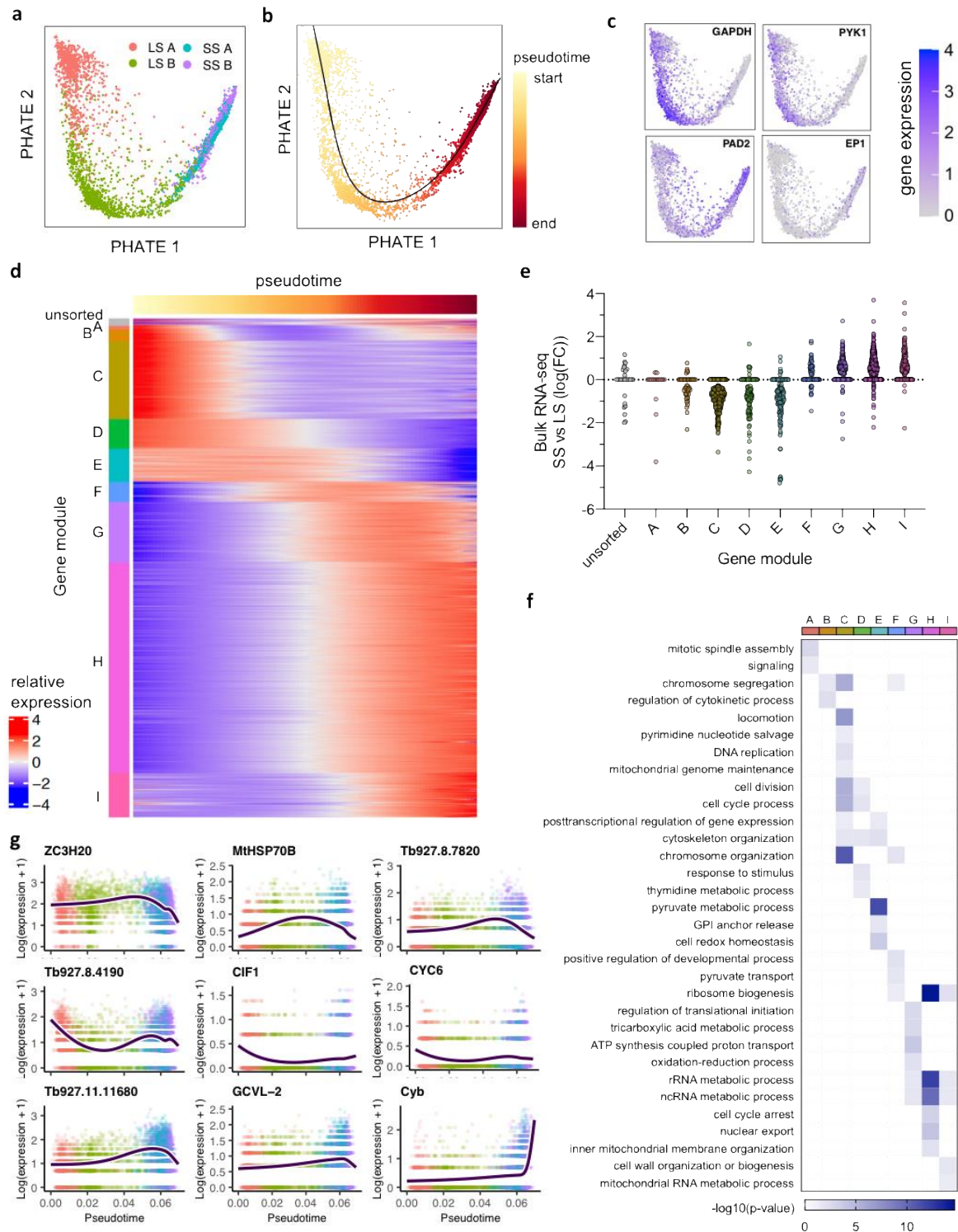
Figure 2. Pseudotime analysis reveals dynamic gene expression in slender to stumpy differentiation. PHATE plots of individual parasite transcriptomes coloured by (**a**) cluster identity, (**b**) pseudotime values and (**c**) raw marker gene transcript count as in **1e**. **d**) Heatmap plotting relative expression of genes significantly (>0.05 adjusted p value, FC > 2) associated with the trajectory (2001 genes). Top track shows pseudotime. Genes are clustered by expression pattern over pseudotime into 9 modules of co-expressed gene, indicated to the left. Unsorted genes are indicated (grey). **e**) Fold change of differentiation associated genes in bulk in vivo-derived RNA-seq data, comparing stumpy (peak parasitemia)

1000   and slender (low parasitemia). Each point is one gene, grouped and coloured according to
1001   the co-expressed module identified in d. **f**) Biological progress gene ontology (GO) term
1002   analysis of differentiation associated genes grouped by co-expressed module. **g**) Gene
1003   expression (log(transcript count +1)) across pseudotime from slender to stumpy
1004   differentiation of 9 genes identified as transiently upregulated (ZC3H20, MtHSP70B,
1005   Tb927.8.7820), transiently down regulated (Tb927.8.4190, CIF1, CYC6) or unsorted
1006   (Tb927.11.11680, GCVL-2, Cyb). Each point is one cell coloured by cluster as in **a**. Dark blue
1007   line is smoothed average expression across pseudotime.
1008

1009
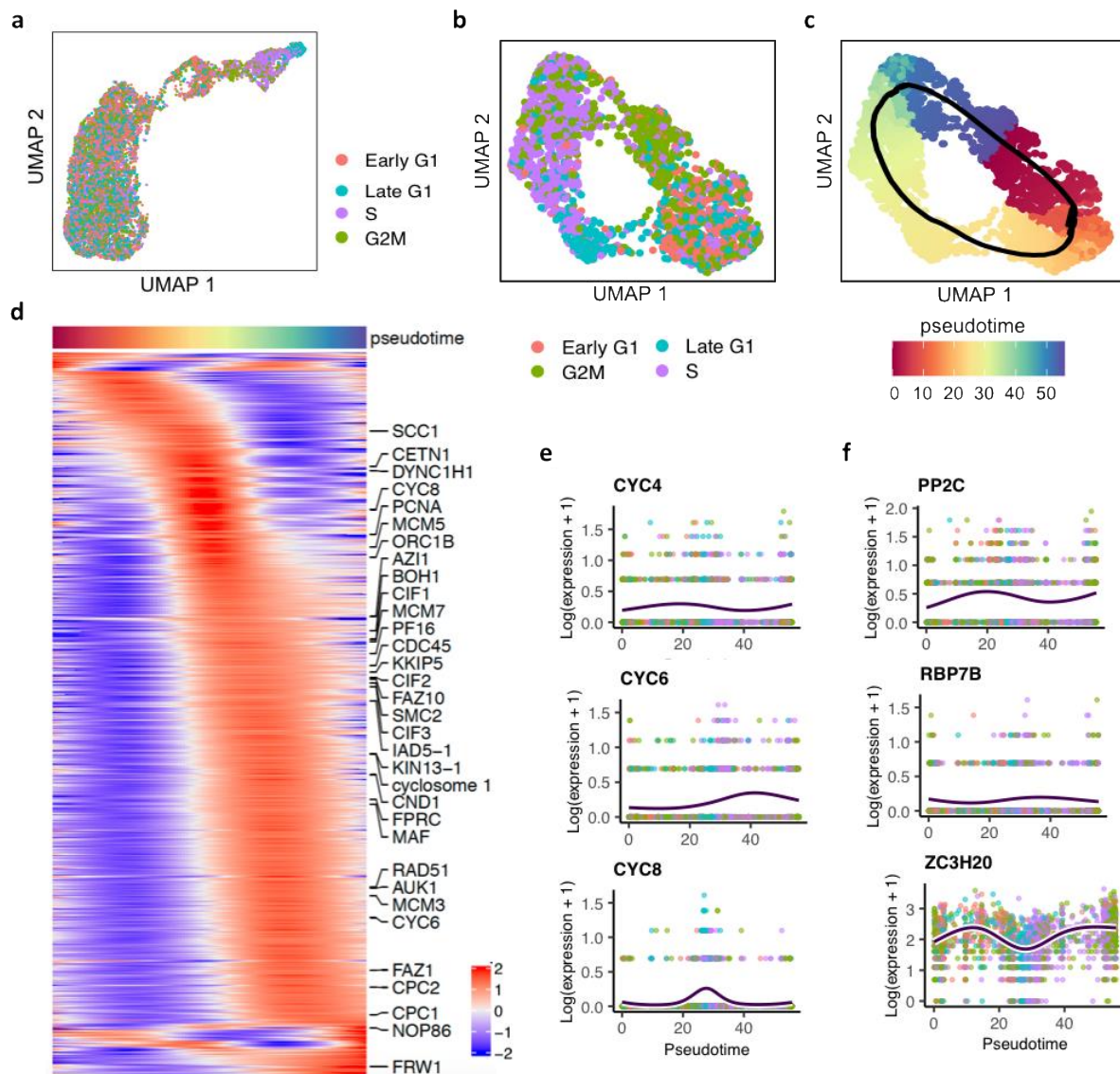
Figure 3. Identification of genes differentially expressed during the slender form cell cycle.
a) UMAP of WT cells coloured by assigned cell cycle phase. **b)** UMAP of re-plotted slender A and slender B clusters of cells, using genes variable within the slender population. **c)** UMAP of slender cells coloured by assigned pseudotime value. The black line indicates the inferred circular trajectory. **d)** Heatmap of relative expression of genes significantly differentially expressed (p-value <0.05), with FC >2 over the cell cycle. Genes associated with the GO term "cell cycle process" are labelled. **e)** Expression (y-axis; log(expression +1)) of 3 cyclin genes (CYC4, CYC6 and CYC8) and cell cycle pseudotime (x-axis). Each point is one cell coloured by cell cycle phase and dark blue line shows average expression over pseudotime. **f)** Expression of three genes previously shown to regulate stumpy formation. Labels as in **e**.
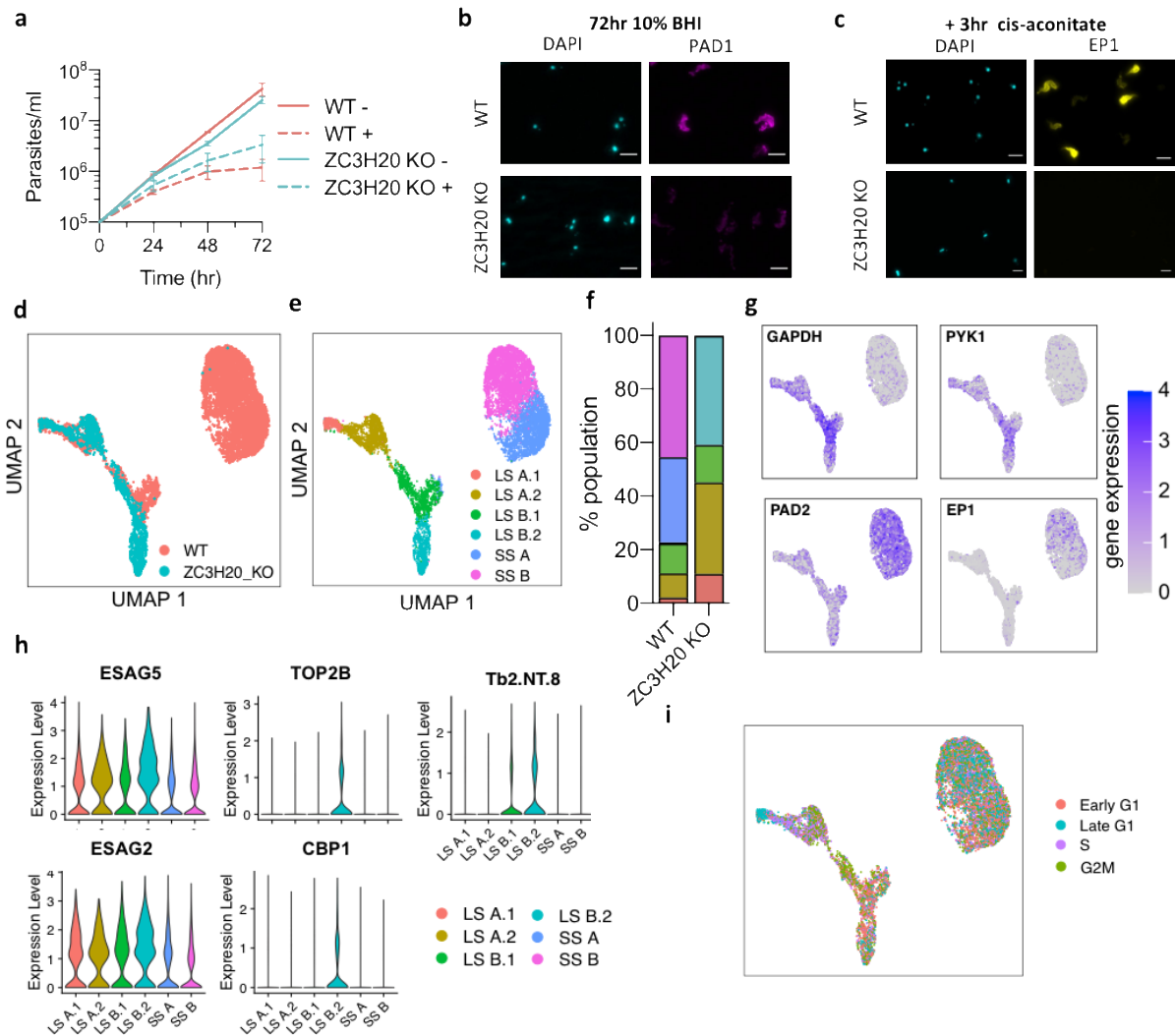
Figure 4. scRNA-seq analysis of differentiation incompetent ZC3H20 KO *T. brucei* parasites. Cumulative growth of WT (red) and ZC3H20 KO (blue) *T. brucei* in culture with (dashed line) and without (solid line) 10% BHI broth. Error bars, SD of three independent replicates. **b)** Staining of WT and ZC3H20 KO parasites with anti-PAD1 antibody after 72 hr incubation with 10% BHI broth. Scale, 5 μm. **c)**. EP1 staining of WT and ZC3H20 KO after 3 hr treatment with cis-aconitate to induced differentiation of 72hr BHI + samples into procyclic forms. Scale, 5 μm. UMAP plots of integrated WT cells (red) and ZC3H20 KO cells (blue) coloured by cell type (**d**) and by cluster identification (**e**). **f)** Proportion of cells in each cluster identified in integrated WT and ZC3H20 KO cells. **g)** UMAP of WT and ZC3H20 KO parasites coloured by transcript count of marker genes as in **1e**. **h)** Violin plots of top slender B.2 marker genes. **i)** UMAP of integrated WT and ZC3H20 KO cells coloured by cell cycle phase.
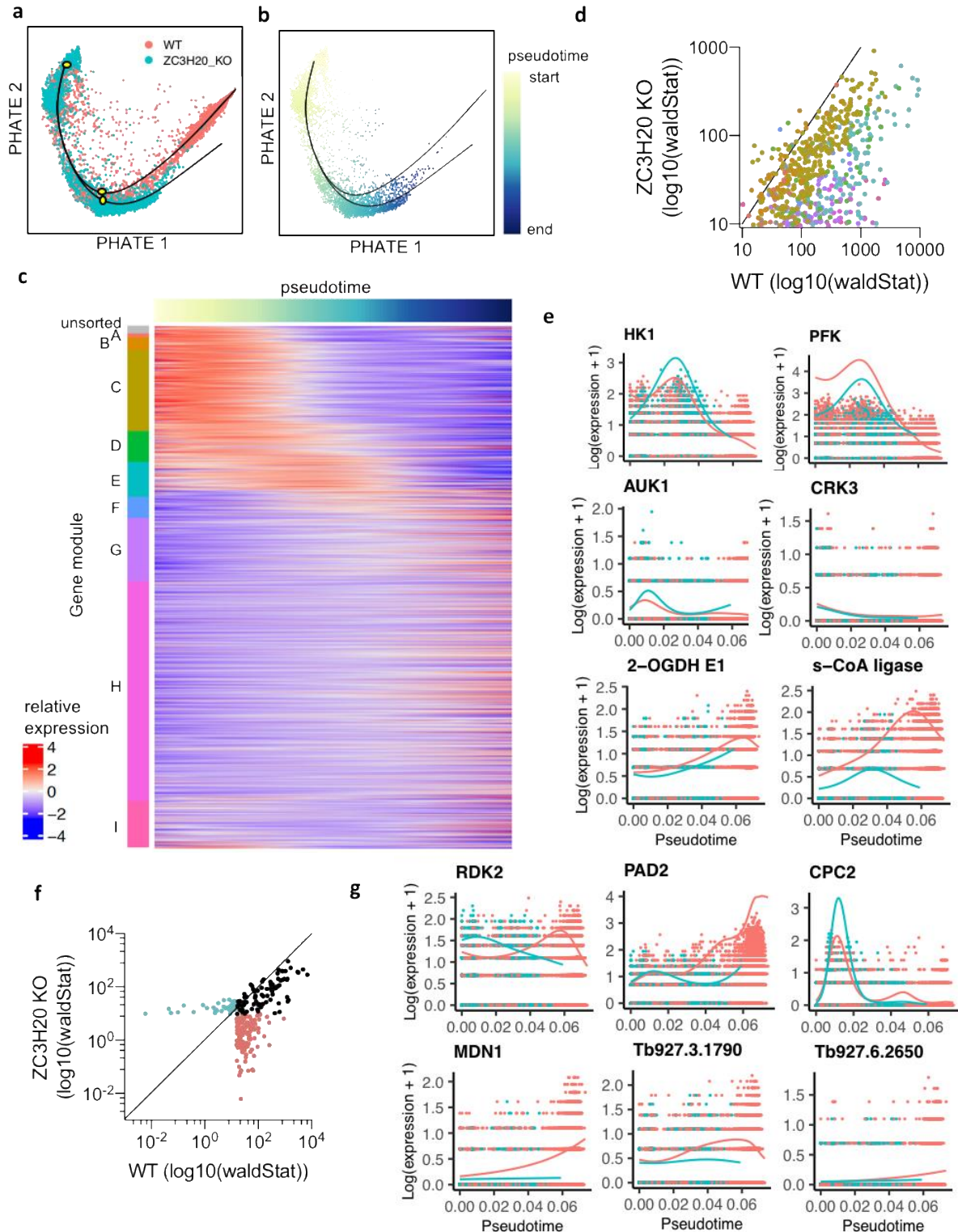
Figure 5. Comparison of differentially expressed genes during differentiation of WT cells and differentiation incompetent ZC3H20 KO cells. a) PHATE map of WT (red) and ZC3H20 KO (blue) parasites. Black line indicates branched trajectories. Yellow dots indicate points of analysis for early differentially expressed genes. **b)** PHATE map of ZC3H20 cells only, coloured by pseudotime values assigned for the second lineage of the branch trajectory, black line. **c)** Scatter plot of differentiation-associated genes also found to be differentially expressed in the ZC3H20 KO trajectory. Axes show the association score (log10(wald stat))

1043   for each gene with the WT differentiation trajectory (x-axis) and ZC3H20 KO trajectory (y-
1044   axis). Each gene is coloured by is co-expression module identified in **2d**. Black line indicates
1045   x=y. **d)** Heatmap of all differentiation associated genes (n = 2001) identified in **2d**. Relative
1046   expression across the ZC3H20 KO trajectory is plotted of each gene, grouped by co-
1047   expression module. **e)** Expression of example genes across the WT differentiation trajectory
1048   (red) and ZC3H20 KO trajectory (blue). X-axis; pseudotime, y-axis; log(expression +1). **f)**
1049   Scatter plot of genes identified as early differentially expressed across the branched
1050   trajectory (between white points in **a**). Axes show the association score (log10(wald stat)) for
1051   each gene with the WT differentiation trajectory (x-axis) and ZC3H20 KO trajectory (y-axis).
1052   Genes are coloured by their significant association with the WT (red), ZC3H20 KO (blue), or
1053   both trajectories (black). **g)** Expression patterns of example of early differentially expressed
1054   genes as in **e**).
1055

1056
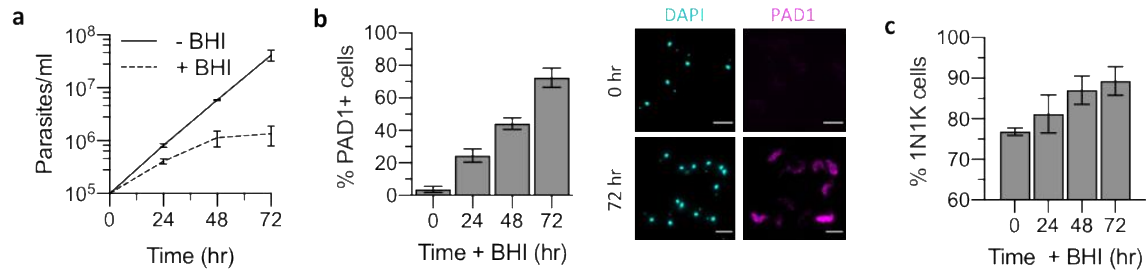
1057

1058 **Fig S1. Brain heat infusion (BHI) broth induces slender to stumpy differentiation *in vitro.***
1059 **a**) Cumulative growth of pleomorphic *T. brucei* in culture with (dashed line) and without
1060 (solid line) 10% BHI broth. Error bars, SD of three independent replicates. **b**) Left:
1061 Percentage of cells expressing stumpy marker protein PAD1 after culturing with 10% BHI.
1062 Error bars, SD of triplicate samples. Right: Staining of parasites with anti-PAD1 antibody,
1063 after 72 hr incubation with 10% BHI broth. Scale, 5 μm. **c**) Percentage of 1N1K *T. brucei* after
1064 culturing with 10% BHI broth. Error bars, SD of three independent experiments.

1065
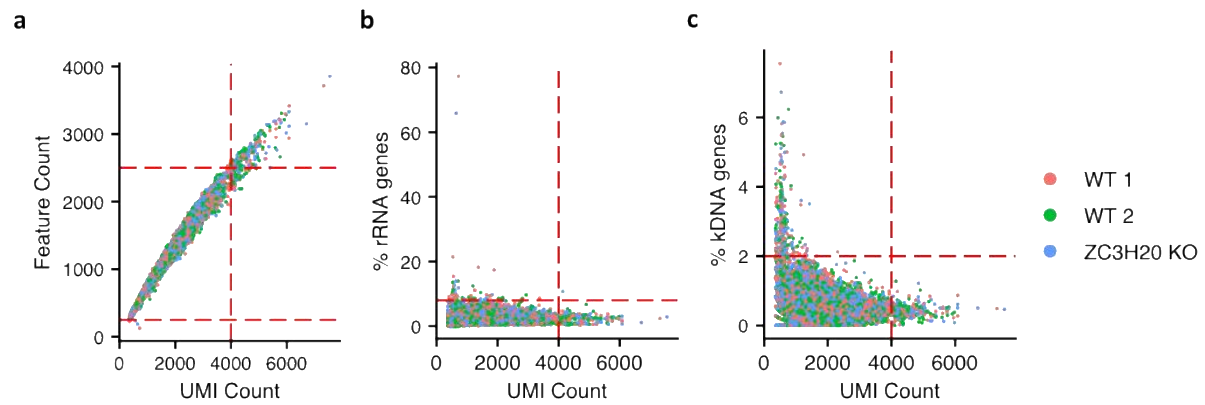
1066

1067 **Fig S2. Quality control and filtering of single transcriptomes.** Scatter plots for WT replicate
1068 1(red) and 2 (green), ZC3H20 KO (blue) experiments; each data point is one transcriptome.
1069 Plots show the relationship between the number of unique molecular identifiers detected
1070 per cell and (**a**) number of features (genes), (**b**) percentage of features encoding ribosomal
1071 RNA (rRNA) and (**c**) percentage of features encoding on the kDNA maxi circle genome. Red
1072 dashed lines indicate thresholds used to filter cells per experiment; Features count > 250, <
1073 2500, UMI count < 4000, % rRNA genes < 8, % kDNA genes < 2.
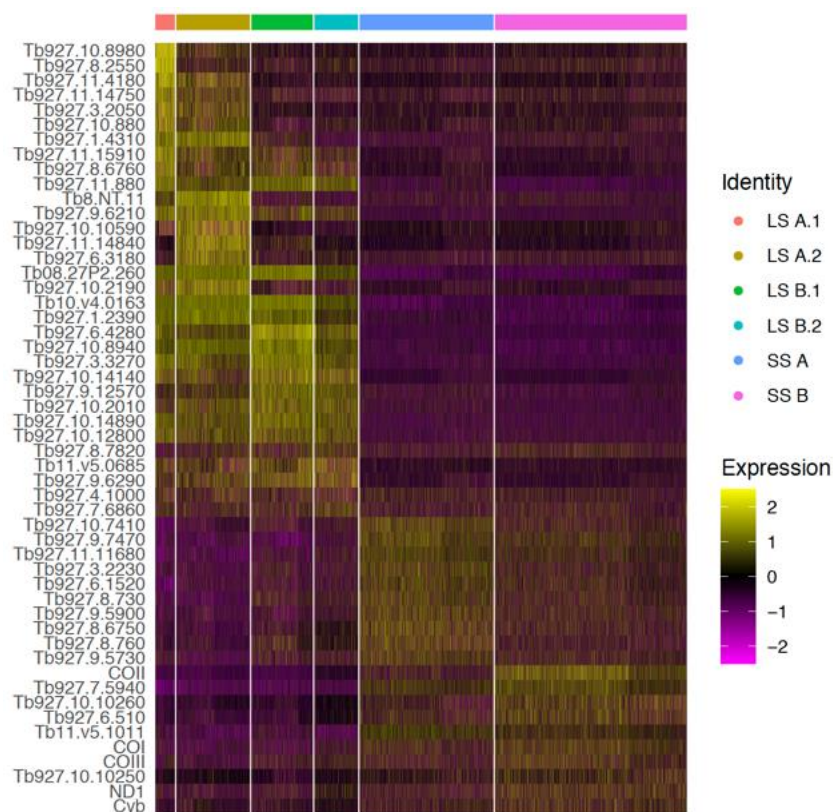
1074
1075 **Fig S3. WT and ZC3H20 KO cluster maker genes.** Heatmap showing relative expression of
1076 the top 10 unique maker genes of each cluster identified in **4d**. Each row is one gene
1077 coloured by relative expression. The gene ID is given for each marker. Each column is one
1078 cell grouped according to cluster identity.
1079