Single-cell transcriptomics reveals expression profiles of *Trypanosoma brucei* sexual stages

Virginia M. Howick[1,2,3*], Lori Peacock [4,5], Chris Kay [4], Clare Collett [4$], Wendy Gibson [4#], Mara K.N. Lawniczak [3#]

[1] Institute of Biodiversity, Animal Health, and Comparative Medicine, University of Glasgow, Glasgow G12 8QQ, UK
[2] Wellcome Centre for Integrative Parasitology, University of Glasgow, Glasgow G12 8QQ, UK
[3] Parasites and Microbes Programme, Wellcome Sanger Institute, Hinxton CB10 1SA, UK
[4] School of Biological Sciences, University of Bristol, Bristol BS8 1UG, UK
[5] Bristol Veterinary School, University of Bristol, Langford BS40 5DU, UK
[$] Present address: UK Health Security Agency, Porton Down, Salisbury SP4 0JG, UK

[#] These authors contributed equally to this work
[*] Correspondence: VMH, Virginia.Howick@glasgow.ac.uk

Running head: Single-cell transcriptomics of trypanosomes

Abstract

Early diverging lineages such as trypanosomes can provide clues to the evolution of sexual reproduction in eukaryotes. In *Trypanosoma brucei*, the pathogen that causes Human African Trypanosomiasis, sexual reproduction occurs in the salivary glands of the insect host, but analysis of the molecular signatures that define these sexual forms is complicated because they mingle with more numerous, mitotically-dividing developmental stages. We used single-cell RNA-sequencing (scRNAseq) to profile 388 individual trypanosomes from midgut, proventriculus, and salivary glands of infected tsetse flies allowing us to identify tissue-specific cell types. Further investigation of salivary gland parasite transcriptomes revealed fine-scale changes in gene expression over a developmental progression from putative sexual forms through metacyclics expressing variant surface glycoprotein genes. The cluster of cells potentially containing sexual forms was characterized by high level transcription of the gamete fusion protein HAP2, together with an array of surface proteins and several genes of unknown function. We linked these expression patterns to distinct morphological forms using immunofluorescence assays and reporter gene expression to demonstrate that the kinetoplastid-conserved gene Tb927.10.12080 is exclusively expressed at high levels by meiotic intermediates and gametes. We speculate that this protein, currently of unknown function, plays a role in gamete formation and/or fusion.

Keywords: African trypanosomes; single cell RNA sequencing; transcriptomics; tsetse fly; *Glossina*

Introduction

The African tsetse-transmitted trypanosomes are single-celled parasites that cause human and animal diseases, which are a heavy burden for many countries in sub-Saharan Africa. These trypanosomes survive in both the tsetse and mammalian host by taking on distinct morphological forms that suit the diverse metabolic and immune environments they encounter (Matthews, 2005). When blood infected with *Trypanosoma brucei* is imbibed by the tsetse fly (genus *Glossina*), trypanosome blood stream forms (BSF) rapidly change their transcriptional profile, including switching off Variant Surface Glycoprotein (VSG) transcription and upregulating expression of other surface proteins such as procyclins (Bütikofer et al., 2002; Vassella et al., 2001) They also switch their metabolism from dependence on glucose processed via glycolysis in the glycosome to exploitation of amino acids such as proline via the mitochondrial TCA cycle (Bringaud et al., 2006). Trypanosomes then multiply as procyclics in the fly midgut before migrating anteriorly, first colonising the proventriculus or cardia, the valve between the foregut and anterior midgut, and then the paired salivary glands (**Figure 1A**) (Sharma et al., 2008; Van Den Abbeele et al., 1999). Here trypanosomes attach and proliferate as epimastigotes characterised by BARP surface proteins (Urwyler et al., 2007), before final differentiation into infective metacyclics that are inoculated into a new host via the saliva.

Additionally, the salivary glands are the location of the non-obligatory sexual cycle of *T. brucei*, which involves meiosis and the production of haploid gametes (Peacock et al., 2014, 2011). As trypanosomes are early diverging eukaryotes, their sexual processes are of particular interest because they provide insights into the evolution of sexual reproduction and meiosis. Although the morphologies of the meiotic division stages and gametes have been described (Peacock et al., 2021, 2014, 2011), little is known about the transcriptional dynamics that characterise the sexual stages because these cells are a minority of the heterogenous cell population in the salivary glands. Sexual stages are found during the early phase of establishment of salivary gland infection, with numbers peaking about three weeks after fly infection (Peacock et al., 2014, 2011). The sexual cycle appears to be a sideshow in the normal mitotic developmental program, as it occurs in clonal trypanosome lines and does not need to be triggered by external factors such as the presence of another strain.

Single-cell RNAseq opens the door to study heterogenous populations of single-celled parasites by delineating expression patterns of individual cells allowing us to understand continuous developmental processes, cell-type specific patterns of co-expression and bet-hedging strategies (Howick et al., 2019; Nötzel and Kafsack, 2021; Poran et al., 2017; Real et al., 2021; Reid et al., 2018; Sà et al., 2020; Witmer et al., 2021). Recent studies have profiled *T. brucei* populations using single-cell droplet-based approaches from BSF culture to profile the development of stumpy forms (Briggs et al., 2021) as well as *in vivo* salivary gland parasites in order to identify a potential vaccine candidate among mature metacyclics (Vigneron et al., 2020) and the dynamics of *VSG* expression in the developing metacyclic parasites (Hutchinson et al., 2021). These studies complement the previous bulk transcriptomic studies in *T. brucei* that identified the major changes in transcriptional patterns over time and metacyclic development using either whole infected salivary glands or *in vitro*-derived metacyclics from the RBP 6-inducible system (Christiano et al., 2017; Doleželová et al., 2020; Kolev et al., 2012; Naguleswaran et al., 2018; Ramey-Butler et al., 2015; Savage et al., 2016, 2012; Telleria et al., 2014). Although these studies have been essential to our

88 understanding of the dynamics of gene expression in *T. brucei*, we still lack an understanding
89 of the molecular processing that characterize meiosis and sexual development in
90 kinetoplastids.
91
92 Here we have exploited scRNAseq to investigate transcriptomes of the sexual stages of *T.*
93 *brucei* that occur transiently in the heterogeneous trypanosome population in the fly salivary
94 glands. We used a modified Smart-seq2 protocol (Howick et al., 2019; Picelli et al., 2014; Reid
95 et al., 2018) to profile *T. brucei* cells from different tsetse tissues (midgut, proventriculus,
96 salivary glands) at different time points during development. We tied the observed
97 transcriptomic profiles to specific developmental stages, validated by immunofluorescence,
98 and identified cell-type specific markers, which revealed the dynamics of surface protein
99 expression as well as a new candidate gene that may be involved in sexual development.

100 Results and Discussion

101 **Generation of high-quality transcriptomes from *in vitro* procyclic forms**

102 To confirm that the modified Smart-seq2 protocol produces high-quality data for *T. brucei*, we
103 initially profiled 46 single-cell transcriptomes from *in vitro* procyclic forms. We found a mean
104 of $2.6x10^6$ reads mapped to the *T. brucei* 927 genome with the mean detection of 1756 genes
105 per cell **(Figure S1A, B).** Additionally, we observed high expression of genes that encode
106 known procyclic surface antigens including GPEET and EP1-3 (Bütikofer et al., 1997; Vassella
107 et al., 2001) **(Figure S1C).** These data support the utility of our protocol to profile single-cell
108 expression profiles in kinetoplastids.

109 **Transcriptomes of fly developmental stages**

110 Having confirmed that the modified Smart-seq2 protocol would produce high-quality data from
111 *T. brucei in vitro* procyclic forms**,** we profiled parasite transcriptomes isolated from diverse
112 tsetse tissues at different timepoints, and from two *T. brucei* strains as outlined in **Figure 1A**.
113 After quality control **(Figure S2)**, we obtained a total of 388 single-cell parasite transcriptomes:
114 78 from the midgut, 34 from the proventriculus and 276 from the salivary glands **(Figure 1A).**
115 Parasites from the midgut and proventriculus were isolated from flies infected with *T. brucei*
116 strain1738 dissected day 21 post infection (pi); 53 parasites from the SG were also derived
117 from these flies. A further 62 1738 parasites from the salivary glands were collected on day
118 40 pi; these cells were derived from free (spill-out from tissue) or attached (enzymatically
119 disassociated) cell populations to capture populations dominated by metacyclics or attached
120 epimastigotes and premetacyclics, respectively **(Figure 1A)**. A further experiment dissected
121 on day 24 pi aimed to analyse salivary gland parasites from an experimental cross between
122 strains 1738 and J10; 86 cells were from single infections of 1738 (38) or J10 (48), and 75
123 were from the experimental cross. We additionally used two cell preservation methods in the
124 collection of these data to allow for more flexibility in processing time. Although small
125 differences were observed in the number of genes detected between preservation methods,
126 this was confounded by timepoint, preventing us from fully understanding the impact of the
127 preservation techniques alone (**Figure S3**).
128
129 To understand transcriptional variation at the single-cell level across tissues, strains, and time,
130 we performed dimensionality reduction with all 388 cells using UMAP **(Figure 1B).** We

observed that cells grouped by their tissue of origin, with clusters representing midgut and proventriculus trypanosomes and two groups of salivary gland parasites that we hypothesized could represent different cell-types or stages **(Figure 1B).** This idea was supported by the distribution of time points across the two salivary gland groups with 21- and 24-day pi cells distributed throughout the two groups, while the 40-day pi cells (squares) occupied the centre and right-hand area; the left-hand area therefore represents early salivary gland developmental stages, such as sexual stages, which are frequent at 21 – 24 days pi, but are relatively scarce at 40 days pi compared to epimastigotes and metacyclics (Peacock et al, 2014).

We next used consensus clustering to partition the cells into six clusters based on the top average silhouette score in SC3 (Kiselev et al., 2017) **(Figure 1C)**. The midgut and proventriculus parasites each formed a single cluster (clusters C1 and C2 respectively), whereas the salivary gland cells divided into four clusters (C3-C6). We identified 238 marker genes across the six clusters allowing us to assign potential cell types (AUROC > 0.75, $p < 0.01$**, Table S1, Figure 1D)**. C1 and C5 had very few significant marker genes and C6 had none (hence not included in **Figure 1D**). This is likely due to different overall levels of transcriptional activity across the cell types as we observed fewer genes per cell for these clusters **(Table S1)**. For further characterization, we additionally identified the top 200 genes expressed in each cluster **(Table S2).** Based on both the cluster markers and top genes of each cluster, we were able to assign putative cell-types. C1 and C2 showed expression patterns consistent with midgut procyclics and proventricular forms, respectively. C1 showed high expression of several genes reported by previous studies (**Table S2**) (Savage et al 2012, 2016; Nuguleswaran et al 2021). We identified a single marker gene for this cluster, *FHc* (Tb927.3.4500), a fumarate hydratase, which catalyses conversion of fumarate to malate in the TCA cycle **(Figure 1)**, and other genes for TCA enzymes and oxidative phosphorylation were also highly expressed **(Table S2)**. C1 also expressed several genes encoding for surface proteins at high levels (*EP1*: Tb927.10.10260, *EP3-2:* Tb927.6.520, and *EP2:* Tb927.10.10250), as well as three Proteins Associated with Differentiation (*PAD1:* Tb927.7.5930, *PAD2:* Tb927.7.5940, *PAD7:* Tb927.7.5990) **(Table S2)**, which are implicated as sensors of environmental stimuli and trigger differentiation (Dean et al., 2009). C2 had several marker genes associated with transport (e.g. amino acid transporter *AATP11* and purine nucleotide transporter *NT10*, **Figure 1**), both also highly expressed in C1 cells **(Table S2)**. Procyclin *EP2* was identified as a marker gene for this cluster, though both *EP1* and *EP3* (Tb927.6.480) were also highly expressed **(Table S2)**.

Cluster C3 comprised day 21/24 pi early salivary gland developmental stages (**Figure 1B, C**) including potential sexual forms. Notably, we observed high expression of the gamete fusion protein HAP2 (Tb927.10.10770), which is known to be expressed in meiotic intermediates and gametes **(Table S1)** (Peacock et al., 2021). An analogous cluster was also identified by Hutchinson et al., 2021, which showed high expression of *HAP2* and *HOP1* (Tb927.10.5490), a meiosis-specific protein. Other notable marker genes for this cluster were two leucine-rich repeat (LRR) protein genes (Tb927.9.14570, Tb927.7.7180), with a third also highly expressed **(Table S2)**; as LRRs are protein recognition motifs, this could be significant in gamete interactions. Procyclin genes *EP1, EP3* and *GPEET* (Tb927.6.510) **(Table S2)** were highly expressed, together with *BARP* genes, which is the characteristic surface protein of epimastigotes (Urwyler et al., 2007).

179 Several *BARP* genes were the prominent marker transcripts in cluster C4 **(Figure 1D)**,
180 identifying this cluster as salivary gland epimastigotes, consistent with previous studies
181 (Vigneron et al, 2020; Hutchinson et al 2021). The majority of the later time point (day 40 pi)
182 cells were found in clusters C5 and C6 (**Figure 1B, C**), showing that these clusters represent
183 the later salivary gland developmental stages including mature metacyclics and/or their
184 immediate precursors, pre- and nascent metacyclics. However, the identities of these clusters
185 remain unclear. C5 had only three significant marker genes, two of which encode zinc finger
186 proteins and one hypothetical protein, and C6 none (**Figure 1D, Table S2**). The zinc finger
187 protein genes, *ZC3H11* (Tb927.5.810) and *ZC3H45* (Tb927.11.8470) were also identified as
188 biomarkers of pre-metacyclics (Meta 1) in Vigneron et al 2020, and ZC3H11 was also identified
189 as a highly expressed gene in purified, culture-derived mature metacyclics (Christiano et al
190 2017). Several other highly expressed genes of mature metacyclics (Christiano et al 2017)
191 were also identified in our data for either C5 or C6 (e.g. *ZFP2*, *HSP 110*; **Table S2**). It is
192 noteworthy that cluster C5 is predominantly J10, while cluster C6 is predominantly 1738
193 (**Figure 1D**), and hence these clusters may also represent strain-specific rather than stage-
194 specific expression differences. Additionally, mature and nascent metacyclics both have VSG
195 on the surface (Vickerman, 1985), and it is reasonable to suppose that pre-metacyclics already
196 transcribe *VSG*s. In support of this, both previous scRNAseq studies found high levels of
197 expression of *VSG*s in cells identified as pre-metacyclics (Meta 1, Vigneron et al, 2020; cluster
198 5, Hutchinson et al, 2021). Here, *VSG*s were not identified as abundant transcripts in either
199 C5 or C6, likely because of poor match between 1738/J10 *VSG* transcripts and the Tb927
200 reference genome, as *VSG* repertoires are strain-specific. We next endeavoured to identify
201 the metacyclic *VSG* repertoire of 1738 and J10, in order to confirm the cell-types of C5 and
202 C6.

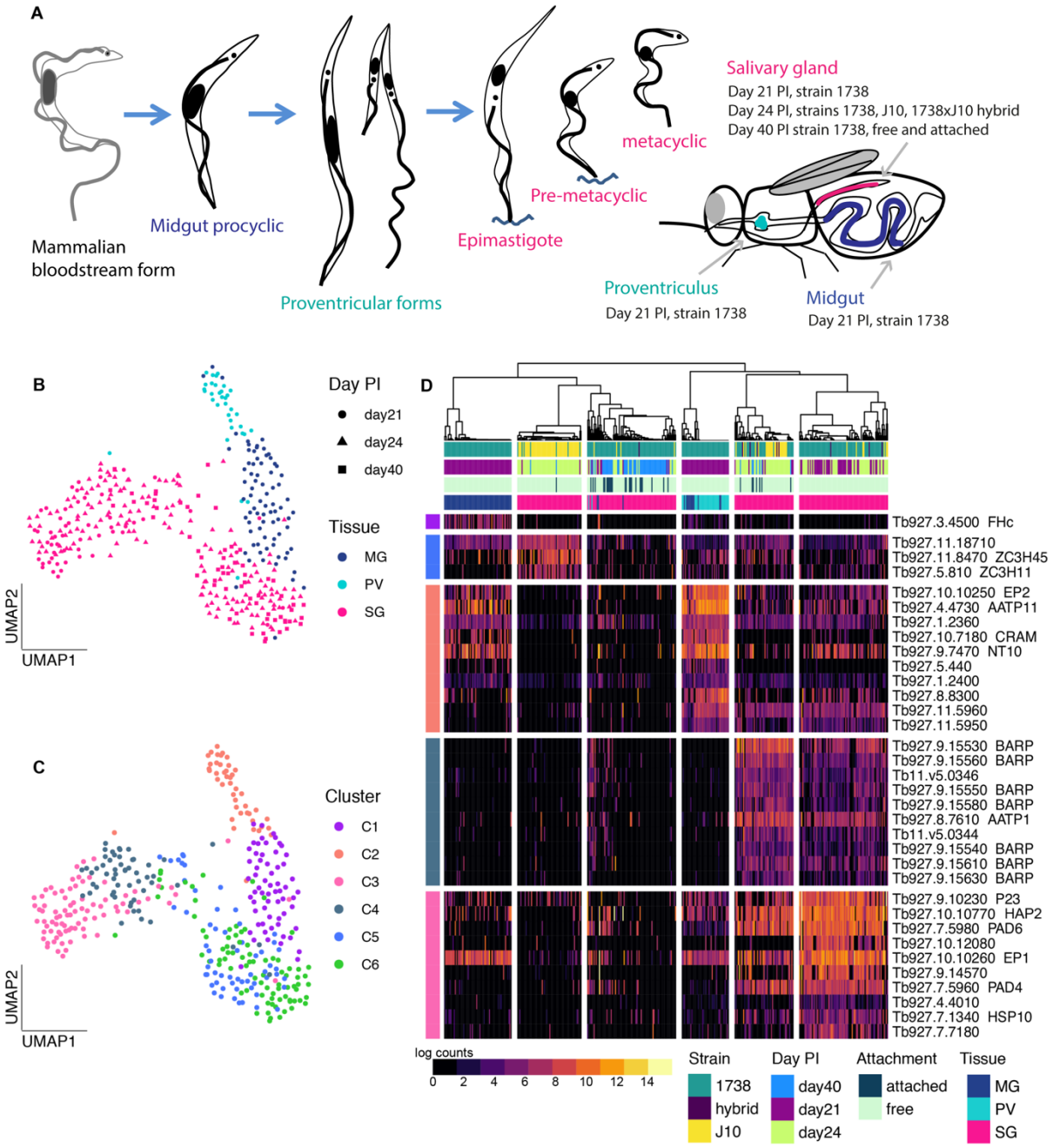
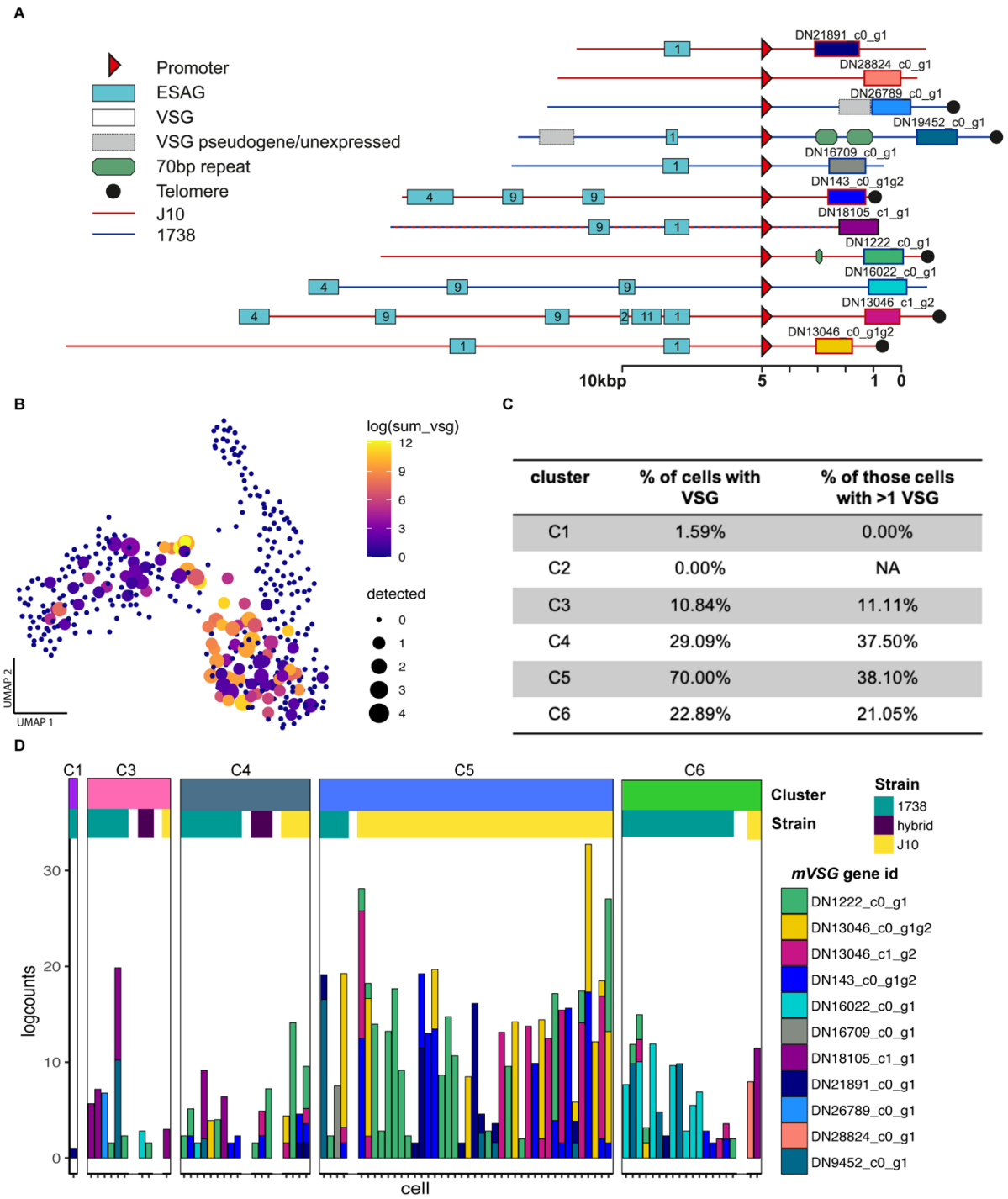


Figure 1. scRNA-seq analysis of trypanosome developmental stages in tsetse.
(A) A schematic of the trypanosome life cycle and collections of the single parasite transcriptomes from midgut (MG; blue), proventriculus (PV; turquoise) and salivary glands (SG; pink) from different time points and strains. Trypanosomes show two conformations: trypomastigote with kinetoplast (small black dot) posterior to nucleus *e.g.* bloodstream form, procyclic, metacyclic; epimastigote with kinetoplast anterior to nucleus. **(B)** A UMAP of the 388 cells that passed QC across collections, coloured by tissue of origin. **(C)** The UMAP coloured by cluster assignment. **(D)** A heatmap of the top significant marker genes from each of the five clusters that had marker genes (AUROC >0.75 & adjusted *p*-value < 0.01).

Expression of mVSG transcripts

Mature metacyclics can be unequivocally identified by their lack of expression of the cell surface proteins GPEET, EP and BARP, replaced by expression of a single *mVSG* gene in each cell (Christiano et al., 2017), but pre-metacyclics also transcribe *mVSG*s at high level in (Hutchinson et al., 2021). To explore *mVSG* expression in salivary gland-derived cells here,

218 we first needed to identify the *mVSG* repertoires of strains 1738 and J10, which differ from
219 those of previously published strains. We built a *de novo* transcriptome assembly based on all
220 reads from the 388 tsetse-derived cells and identified putative *mVSG*s by comparison of the
221 resulting ORFs to the total *VSG* repertoire of each strain, previously identified using an HMM
222 on full-genome Illumina sequence data (Sistrom et al., 2014). These assembled *mVSG*
223 transcripts were then mapped to 1738 or J10 contigs to place them in a genomic context.
224 Using this method, we identified 11 *mVSG*s that were expressed in our dataset all of which
225 contained an upstream *mVSG* promoter based on the consensus sequence (Cross et al.,
226 2014; Ginger et al., 2002). Additionally, downstream telomeric repeats were present in six of
227 these contigs (**Figure 2A**). Although several *ESAG*s were found on the contigs, these were
228 up- not downstream of the promotor and therefore not part of the *mVSG* expression site. The
229 presence of these characteristic features of *mVSG* gave us further confidence that the
230 identified transcripts were originating from *bona fide mVSG*. Interestingly, J10 and 1738
231 shared one *mVSG* (DN18105), suggesting some level of conservation across strains. This is
232 the first identification of the *mVSG* repertoire in these strains.
233
234 Individual parasite transcriptomes were then mapped to the assembly to generate counts for
235 each putative *mVSG*. *MVSG* transcripts were expressed by most cells in cluster C5, followed
236 in order by C4, C6 and C3, with negligible expression in C1 and C2 (**Figure 2B, C**). Overall
237 levels of expression were highest in cluster C5 (**Figure 2B, C, D**), with 38% (16/42 cells)
238 expressing more than one *mVSG* (**Figure 2B, C, D**). Multiple *mVSG*s were also expressed by
239 38% (6/16) and 21% (4/19) cells in clusters C4 and C6 respectively, and a single cell in C3
240 (**Figure 2C, 2D**). Hutchinson et al., 2021 confirmed expression of two *mVSGs* in pre-
241 metacyclics using single molecule mRNA-FISH and put forward a model where multiple
242 *mVSG*s are transcribed at low levels initially, with a single *mVSG* dominating expression in
243 the mature metacyclic forms. Based on this model, C4, C5 and C6 all contain a high proportion
244 of pre-metacyclics, as well as some mature metacyclics. The *mVSG*s expressed varied over
245 development and between strains, with DN1222 being the dominant transcript in C5, which
246 were primarily J10 cells, and DN16022 being the dominant transcript in C6, which were
247 primarily 1738 cells (**Figure 2D**). Observed expressed *mVSG*s are largely consistent with the
248 *VSG*s present in their strain (when there is sufficient read depth), but low read count and partial
249 coverage result in ambiguous assignment due to sequence conservation between *VSG*s.



cluster	% of cells with VSG	% of those cells with >1 VSG
C1	1.59%	0.00%
C2	0.00%	NA
C3	10.84%	11.11%
C4	29.09%	37.50%
C5	70.00%	38.10%
C6	22.89%	21.05%

Figure 2. mVSG expression in fly-derived trypanosomes

(A) The genomic context of 11 *mVSG*s identified in strains 1738 and J10. The sequences can be found in supplementary file 1. **(B)** The transcript abundance of *mVSG* across 388 fly-derived trypanosome cells on the UMAP coloured by the logged sum of the *mVSG* counts in each cell and sized by the number of different *mVSG* detected in that cell. **(C)** The breakdown of *mVSG* expression per cluster. C5 had the highest proportion of cells expressing *mVSG* and the greatest proportion of those cells expressing multiple *mVSG*. **(D)** A barchart of all cells expressing *mVSG* (>1 read) organised by cluster and strain. Strain-specific expression of *mVSG* was seen at high levels in C5, which is primarily composed of J10.

260 **Pseudotime analysis of salivary gland development**

261 To understand fine-scale changes in expression patterns during development of salivary gland
262 parasites, we focused on the 161 salivary gland cells of strain 1738 collected over three time
263 points (day 21, 24, 40 pi). The UMAP projection of these cells showed a general
264 correspondence with the clusters identified in **Figure 1C**, with the left-hand group of cells
265 representing clusters C3 and C4 (day 21/24 pi, early and late epimastigotes) and the right-
266 hand group predominantly representing cluster C6 which included most of the day 40 pi
267 transcriptomes (**Figure 3A, B**). The small branch connecting these two groups contained
268 many cells collected from the dissociated salivary gland tissue, which likely represent
269 attached epimastigotes and premetacyclics. Although we cannot rule out that these parasites
270 were trapped unattached cells, their enrichment at this bottleneck in the UMAP indicates their
271 importance in the developmental transition to metacyclics (**Figure 3C**).
272
273 We next used Slingshot to order these cells in pseudotime, revealing a trajectory running from
274 left to right from gametes and early epimastigotes to metacyclics (**Figure 3D**). We discovered
275 692 genes that were differentially expressed over this trajectory (**Figure 3E, Table S3**) and
276 used hierarchical clustering to identify modules of co-expressed genes. Modules expressed
277 early in development were enriched for genes involved in negative regulation of mitotic cell
278 cycle and ATP metabolism (modules 2 and 13, **Figure 3E, Table S3**), while middle-
279 late modules were enriched for genes involved in translation and the ribosome, perhaps
280 necessitated by the changes in surface proteins and metabolism associated with
281 differentiation from epimastigotes to metacyclics (**Figure 3E, Table S3**).

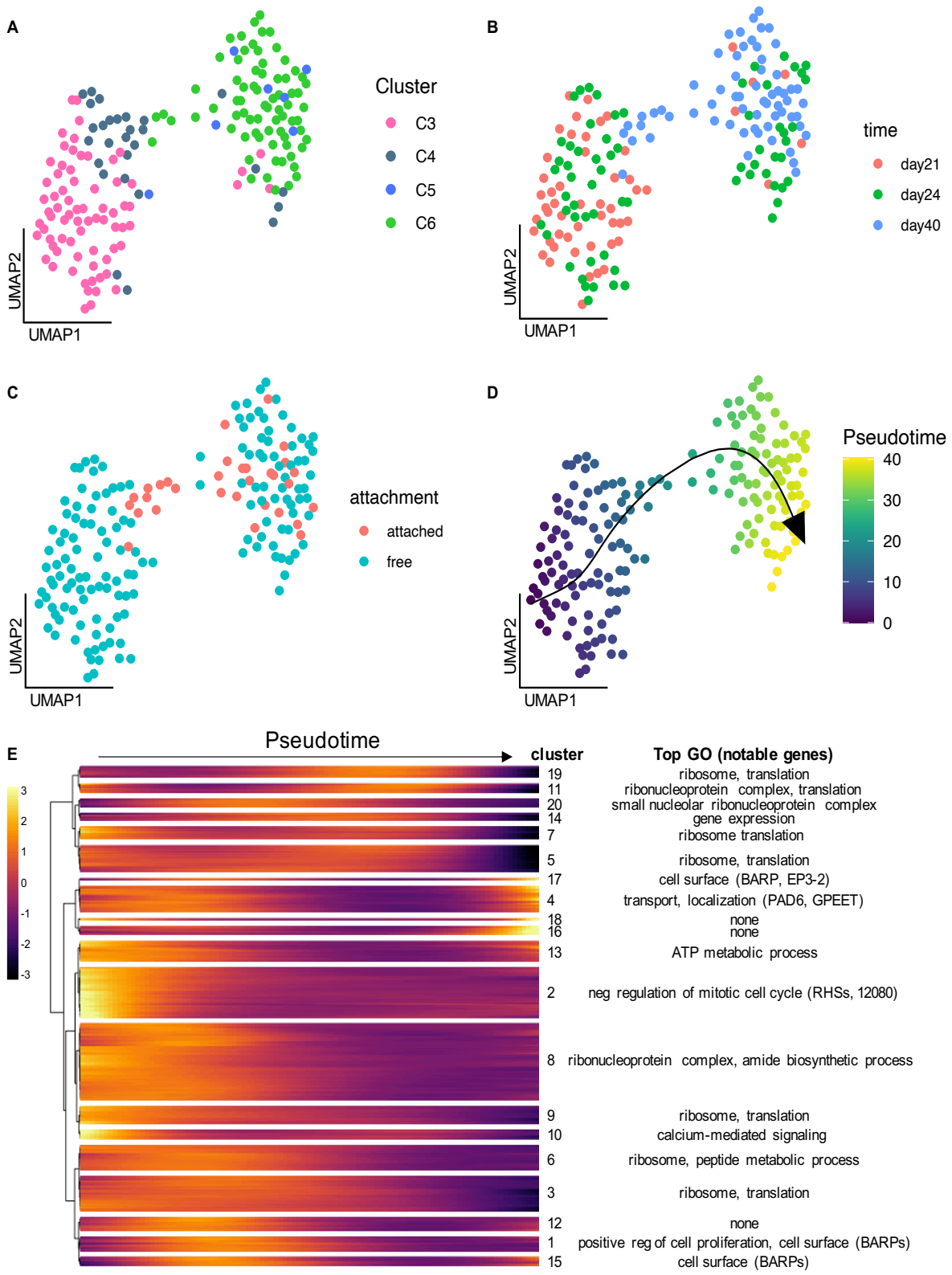


Figure 3. Pseudotime trajectory analysis of developing salivary gland parasites

1738 strain parasites collected from the salivary gland at day 21, 24 and 40 pi were used to map fine-scale changes in gene expression over development. **(A-D)** A UMAP of the 161 1738 salivary glands parasites coloured by global cluster assignment from Figure 1 **(A),** day PI **(B),** attachment treatment **(C)** and pseudotime assignment **(D). (E)** A heatmap of 20 clusters of genes differentially expressed over the pseudotime trajectory from **(D).**

Identification of hybrid progeny and strain-specific expression

To investigate potential cell-types and cell-type specific responses that could be involved in sexual reproduction at day 24 post-infection, we collected J10 and 1738 parasites from both single-strain infected and co-infected tsetse. In the co-infected treatment, we sorted cells from both strains based on fluorescence (1738 GFP+, J10 RFP+), and sorted a small number of RFP+/GFP+ potential hybrid parasites (16 sorted, 14 passed QC). To confirm strain assignment, we used Souporcell to cluster different genotypes based on SNPs in the RNA-seq reads (Heaton et al., 2020). The two genotype clusters identified were each composed of one of the strains based on fluorescent identification with FACS (1738 GFP+ = cluster 0; J10 RFP+ = 1), and the potential hybrid progeny were identified as inter-genotypic doublets (clusters 0/1 and 1/0), with alleles present from both strains (**Figure 4A)**. This confirmed that six of the RFP+/GFP+ cells were genuine hybrids, while a further three hybrid cells were identified from the RFP-/GFP+ or RFP+/GFP- groups; as the GFP and RFP genes are present on only one homologue, four hybrid genotypes with respect to fluorescent protein genes are expected (Gibson et al., 2008). Looking at a UMAP of all day 24 pi cells, we observed some separation of strain in both the early and late epimastigote clusters (C3 and C4) and clusters C5 and C6 (**Figure 4B**). However, we did not see clear clustering based on the infection treatment (co- vs single-infection), suggesting that there is no strong transcriptomic response to presence of another parasite strain (**Figure 4C**). In order to understand if the observed strain-specific clustering was a result of different cell-type composition or differential expression between strains within a cell-type, we integrated the data across strains using Seurat v3 (Butler et al., 2018). Using this method, we were able to co-cluster the early epimastigote cells and identify 11 genes differentially expressed between the two strains (**Figure S4, Table S4**). However, the later stage cells seen in J10 had no representation of 1738, suggesting this cell-type is unique to this strain at day 24 pi, which could be observed if the strains have different developmental rates (**Figure S4**).

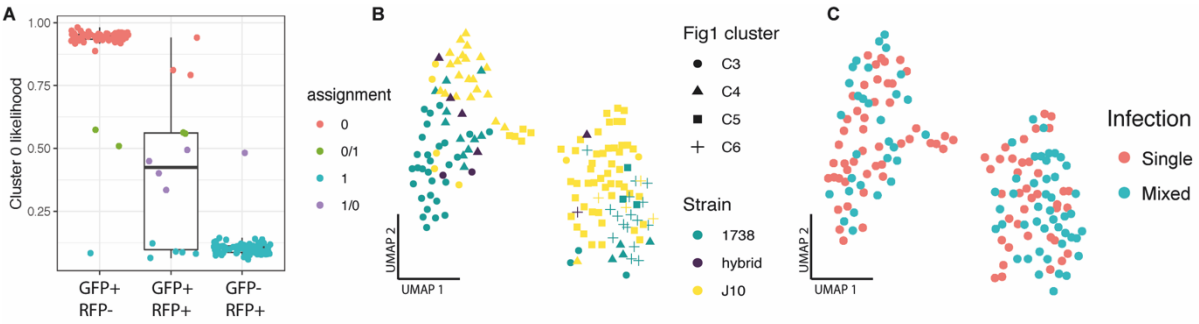


Figure 4. Classification of hybrid progeny

Souporcell was used to assign genotypes based on SNPs found between the two strains. The two genotype assignments (0,1) were each primarily composed of one of the strains based on fluorescent identification with FACS (1738 GFP+/RFP- = cluster 0; J10 GFP-/RFP+ = 1), and the potential hybrid progeny were classified as inter-genotypic doublets (clusters 0/1 and 1/0). The likelihood ratio of cluster 0 assignment is shown for each of the three sorted populations **(A).** The UMAP of day 24 mixed- and single-infection experiments coloured by strain assignment and shaped by Figure 1 cell cluster assignment **(B)** and infection treatment **(C).**

325 **Transcript levels of procyclin and candidate novel sexual stage genes correlate with**
326 **protein expression *in vivo*.**

327 A primary aim of this study was to identify the sexual stages of *T. brucei* and our results support
328 the hypothesis that cluster C3 (**Figure 1**, early epimastigotes) represents meiotic
329 intermediates and gametes, which are abundant around day 21 pi (Peacock et al., 2021, 2014,
330 2011). Looking at expression of genes encoding proteins known to be essential for sexual
331 reproduction, we found high levels of expression of *HAP2* and also *GEX1* in cluster C3, with
332 some signal from the meiosis-specific genes *DMC1* and *HOP1* (**Figure 5A**). Surprisingly,
333 these cells also expressed the procyclin gene *GPEET*, which is considered to be a marker of
334 early procyclics in the tsetse midgut, replaced by EP procyclins in late procyclics (Bütikofer et
335 al., 1997; Vassella et al., 2000). *GPEET*, *EP1*, *HAP2* and *GEX1* all have the highest
336 expression in cluster C3 (**Figure 5B**). We used immunofluorescence to tie these observations
337 to specific morphological forms and to validate the presence of GPEET on the surface of
338 salivary gland parasites (**Figure 5C, Table S5**). We found that GPEET, together with EP and
339 BARP were present in >90% of the meiotic dividers and gametes (identified by morphological
340 features), and as expected absent in metacyclics (**Figure 5C**). Epimastigotes showed a
341 similar pattern to the sexual forms but lower total proportions. Additionally, we looked at
342 proventricular parasites and found expression of EP and GPEET but no BARPs, further
343 confirming our gene expression data is matched at the protein level for these markers.

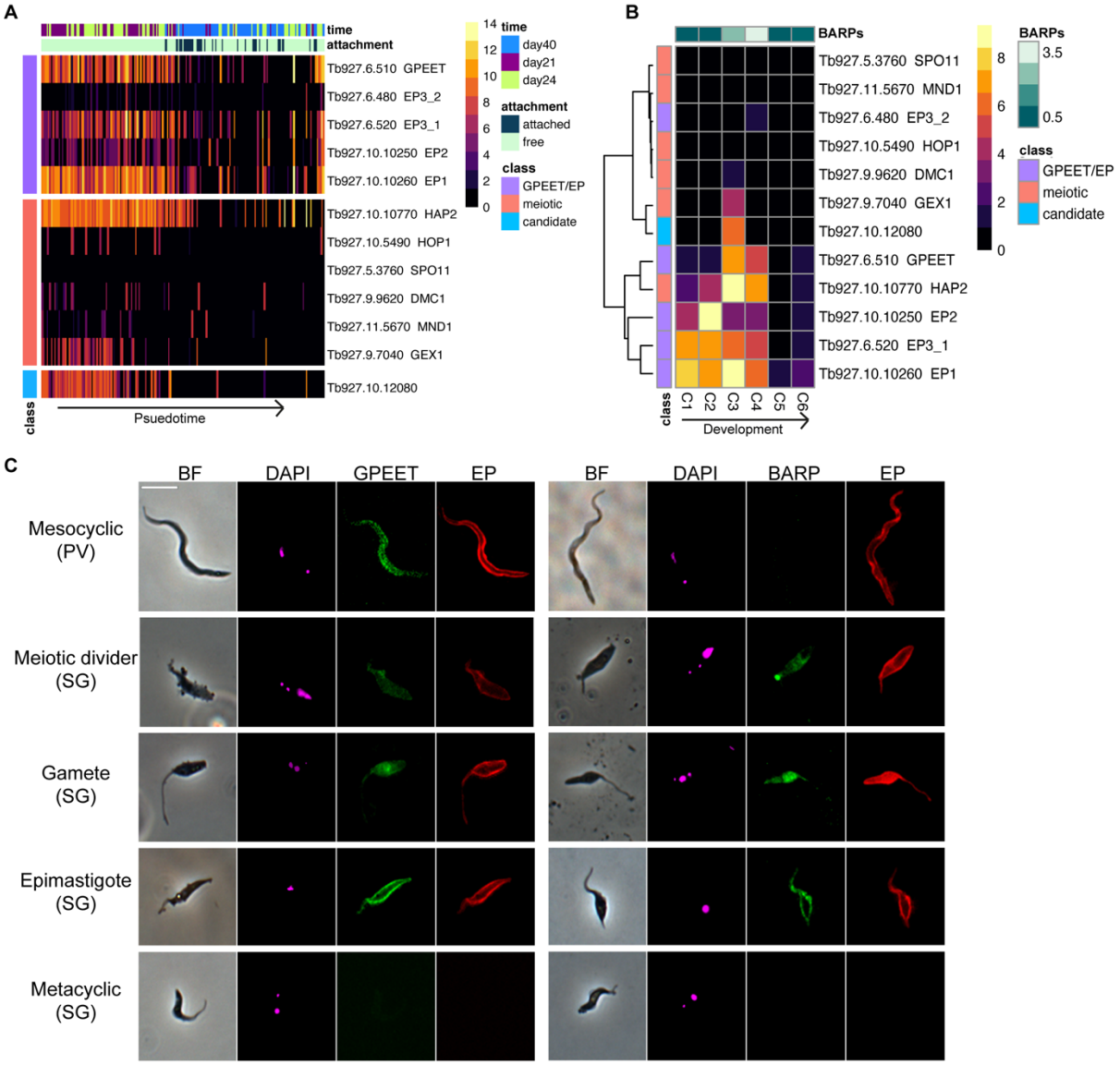


Figure 5. Expression of surface antigens and genes involved in sexual reproduction throughout development in the tsetse fly.

We observed co-expression of procyclic surface antigen genes and *HAP2* in early parasite development in the salivary glands **(A)** and this general pattern of expression was also seen in proventricular forms (C2) as well as putative epimastigotes (C4) that also had high expression of *BARP*s **(B).** Immunofluorescence assays confirmed that these surface proteins corresponded to their transcriptional profiles and were present on the epimastigote and sexual stages **(C).**

Cluster C3 was additionally characterised by strong and unique expression of Tb927.10.12080 (**Figure 1D, Figure 5A,B, Figure 6A**), and we hypothesized that this gene may play a role in sexual development. It encodes a hypothetical protein devoid of recognisable functional domains that is well-conserved in other trypanosomes including *T. congolense, T. vivax, T. cruzi* and *T. grayi,* and the C-terminal domain in more distantly related members of the trypanosomatid family such as *Leishmania* spp. Localisation data from Tryptag.org is equivocal, showing punctate cytoplasmic fluorescence for the N-terminal tagged protein and a mitochondrial location in a proportion of cells for the C-terminal tagged protein. To investigate expression of this protein during the *T. brucei* life cycle, we used the 3' UTR to regulate expression of GFP driven by the procyclin promotor (**Figure S5**). At 20-22 days post infection, there was very little detectable expression in midgut procyclics or proventricular

364 cells, but strong expression in salivary gland trypanosomes (**Figure 6B**). Overall, of the
365 parasites from the salivary gland scored, 39% (182/464) showed expression (**Table S6)**.
366 These included meiotic intermediates and both 1K1N and 2K1N gametes but not metacyclics
367 or unattached epimastigotes (**Figure 6C,D**). Cell types involved in the early stages of meiosis,
368 such as meiotic dividers and 3N cells with one diploid and two haploid nuclei, had lower
369 percentages of cells expressing (10% and 50% respectively) than did those involved in the
370 later stages of meiosis and in gametes (77-80%; **Figure 6C** & **Table S6**). At 37-38 days post
371 infection, the percentage of fluorescent trypanosomes dropped (103/682; 15%) and these cells
372 were misshapen with no recognisable gametes or sexual intermediates. Further work is
373 needed to understand the functional role of Tb927.10.12080 in meiosis and sexual
374 development; however, its unique pattern of transcript and protein expression indicate it could
375 play a vital role in processes that allow for genetic exchange in *T. brucei* and perhaps more
376 broadly in kinetoplastids.

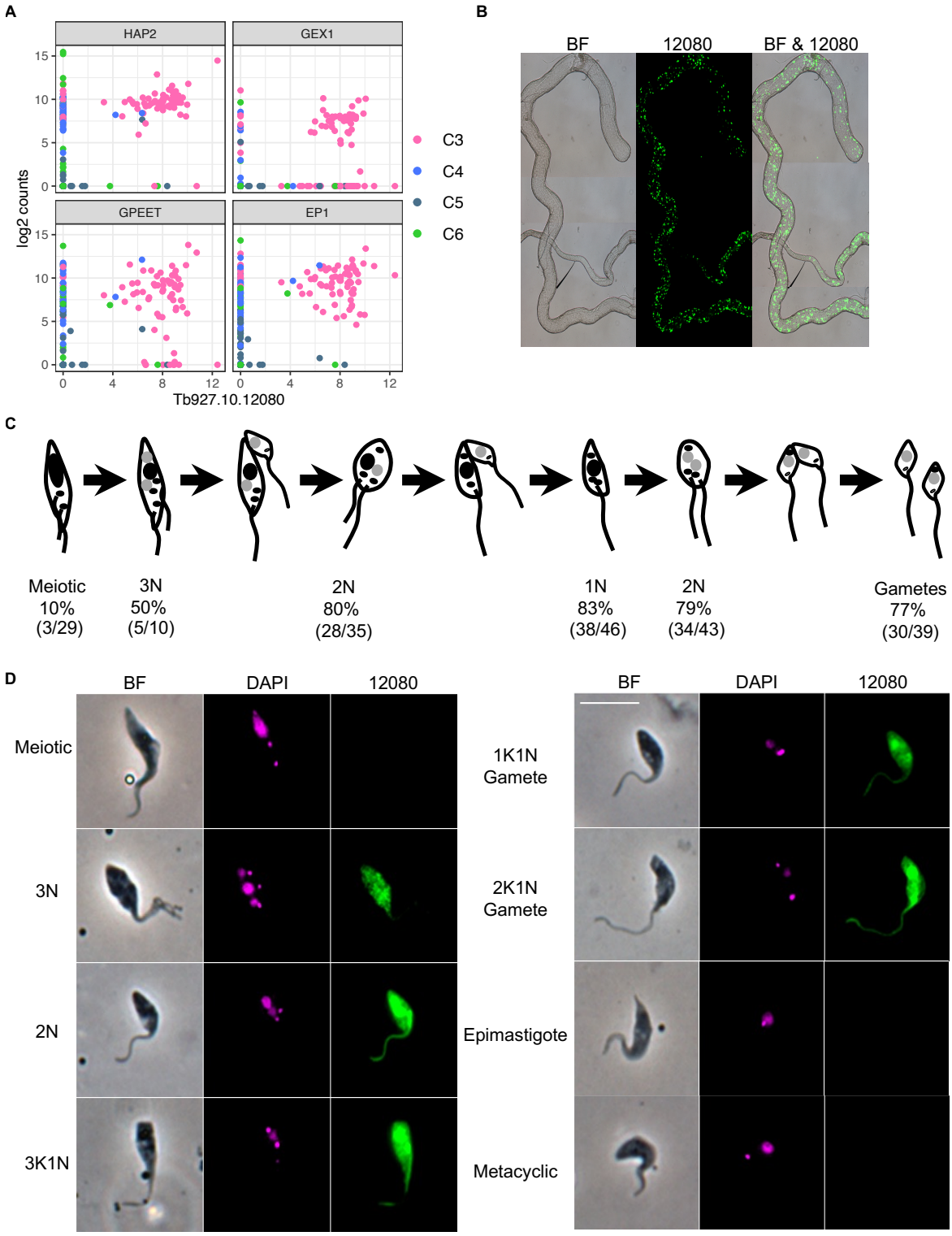


Figure 6. Expression of Tb927.10.12080 coincides with sexual forms.

(A) Co-expression of Tb927.10.12080 with genes encoding HAP2 and GEX1, proteins associated with gamete and nuclear fusion in eukaryotes, and the surface antigen genes GPEET and EP1; co-

381 expression is seen in a subset of cells from C3 (**Figure 1D**). **(B)** Salivary gland dissected 21 days pi
382 that is infected with *T. brucei* 1738 expressing GFP::Tb927.10.12080-3'UTR transcribed from the
383 procyclin promotor. **(C)** Diagram showing major cell types observed during meiosis in *T. brucei* (adapted
384 from Peacock et al 2021); nuclei are shown in black (4C or 2C DNA contents) or grey (1C, haploid) and
385 kinetoplasts are shown as smaller black dots. Values beneath are the numbers and percentages of
386 cells recorded for each cell type; both 1K1N and 2K1N gametes are included in the gamete total. Full
387 data are presented in **Table S6**. **(D)** Trypanosomes from tsetse fly salivary gland spill-out 16-21 days
388 pi with *T. brucei* 1738 expressing GFP::Tb927.10.12080-3'UTR transcribed from the procyclin
389 promotor. Left to right: phase contrast, DAPI, GFP::Tb927.10.12080-3'UTR. The scale bar represents
390 10 µm.
391

392 # Conclusion

393 Here we applied single-cell RNA sequencing to explore the heterogeneous trypanosome
394 populations in the tsetse fly. These data provide a resource for the parasitology community,
395 which we have made available via an interactive website (http://cellatlas.mvls.gla.ac.uk/).
396 Additionally, this data set allowed us to elucidate the transcriptional profiles of key life cycle
397 stages in the salivary glands including the sexual stages. From this mixture of cell types, we
398 were able to identify a cluster of cells that shared a particular transcriptomic profile
399 characterized by high expression of the gene encoding the gamete fusion protein HAP2,
400 together with several unstudied genes. One of these was a kinetoplastid-conserved gene
401 Tb927.10.12080, which was exclusively expressed at high levels by meiotic intermediates and
402 gametes. We speculate that this protein, currently of unknown function, plays a role in gamete
403 formation and/or fusion.
404

405 # Materials and Methods

406 ## Data collection

407 ### Trypanosome culture and tsetse infection

408 The following tsetse-transmissible strains of *Trypanosoma brucei brucei* were used: *T. b.*
409 *brucei* J10 (MCRO/ZM/73/J10) and 1738 (MOVS/KE/70/EATRO 1738); each was genetically
410 modified to express a fluorescent protein gene (J10 RFP, 1738 GFP). Mating between these
411 strains has been demonstrated previously (Gibson et al., 2008) and 1738 reliably produces
412 large numbers of gametes around day 21 post-infection (Peacock et al., 2014). Procyclic form
413 (PF) trypanosomes were grown in Cunningham's medium (CM) (Cunningham, 1977)
414 supplemented with 15 % v/v heat-inactivated foetal calf serum, 5 µg/ml hemin and 10 µg/ml
415 gentamycin at 27°C. Tsetse flies (*Glossina pallidipes*) were infected with PF trypanosomes,
416 maintained and dissected as described previously (Peacock et al., 2011).
417

418 ### Parasite isolation from tsetse tissues for scRNA-seq

419 Free swimming parasites were obtained from *G. pallidipes* by separately pooling tissues into
420 CM (5 midguts in 500 µl CM; 5 proventriculi in 50 µl CM; 20 sets of salivary glands in 50 µl
421 CM). Tissues were incubated at RT for 10 minutes prior to filtration through a 100 µm filter.
422 Cells were washed once with 1 ml CM prior to preservation or sorting. At day 40 pi, the

423 parasites attached to the salivary glands were isolated by disassociation of the tissue after the
424 10-minute incubation period. Forceps were used to transfer the tissue to an enzymatic solution
425 consisting of 200 µl Collagenase IV (1 mg/ml) and 25 µl of Elastase (4 mg/ml). The sample
426 was then incubated at 30°C for 40 min with shaking at 300 rpm. During the incubation, the
427 tissue was disrupted by pipetting up and down 40 times every 15 minutes at first with a p1000
428 pipette set to 150 µl and then with a p200 set to 100 µl once the tissue started to break up.

429 ### Cell preservation

430 A subset of cells was preserved prior to cell sorting to allow for greater flexibility in the time
431 between collections and FACS. The day 40 pi salivary glands parasites (attached and free)
432 cells were fixed by adding 200 µl of dithio-bis(succinimidyl propionate) (DSP; Lomant's
433 reagent) dropwise to the cell pellet as described in (Attar et al., 2018). DSP fixed samples
434 were incubated at room temperature for 30 minutes prior to adding 4 µl 1 M Tris-HCl pH 7.5.
435 Samples were then stored at 4°C for up to 24 hours. Prior to sorting, DTT was added to a final
436 concentration of 50 mM. The day 24 pi salivary gland parasites (cross vs single infection) were
437 preserved by resuspending the cell pellet in 200 µl Hypothermosol-FRS (BioLifeSolutions)
438 (Wang et al., 2018). Samples were then stored at 4°C for 5 hours prior to sorting.

439 ### Cell sorting, library preparation and sequencing

440 All parasite cells were sorted within 24 hours of collection on an Influx cell sorter (BD
441 Biosciences) with a 200 µm nozzle or a Sony SH800 with 100 µm chip. Parasites were sorted
442 based on RFP and/or GFP fluorescence into nuclease-free 96 well plates containing lysis
443 buffer as described previously (Reid et al., 2018). Sorted plates were spun at 1000 g for 10
444 seconds and immediately placed on dry ice. Reverse transcription, PCR, and library
445 preparation were performed as described in (Reid et al., 2018). Cells were multiplexed to 384
446 and sequenced on a single lane of Illumina HiSeq2500 v4 with 75 bp paired-end reads.

447 ### Immunofluorescence

448 Salivary glands, proventriculi and midguts from infected flies (20-22 days pi) were pooled
449 separately into CM and incubated at room temperature (RT) for 10 minutes (to allow
450 trypanosomes to swim out of tissue) and then filtered through a 100 µm filter with PBS.
451 Trypanosomes were pelleted by centrifugation and resuspended in 100 µl PBS. Cells were
452 fixed overnight at 4°C by adding 100 µl 6% paraformaldehyde, 0.1% glutaraldehyde in PBS,
453 and then washed twice with PBS before resuspension in 50 µl PBS. Cell suspensions were
454 pipetted onto 2 x 10 mm coverslips, allowed to settle for 20 mins in a humid chamber, and
455 then liquid was removed and replaced by 2% BSA in PBS. After 30 mins liquid was removed
456 and cells incubated with 2% BSA in PBS containing diluted antibody for 30 mins at RT. Rabbit
457 anti-GPEET (1:1000) and rabbit anti-BARP (1:1000) were a kind gift from Isobel Roditi,
458 University of Bern, Switzerland; mouse anti-EP mAB (1:100) was from Cedarlane. Cells were
459 washed three times with PBS and incubated with 2% BSA in PBS containing anti-rabbit FITC
460 (1:1000) and anti-mouse TRITC (1:1000) for 30 mins at RT. Cells were washed three times
461 with PBS, briefly air dried, stained with DAPI in VECTASHIELD mounting medium (Vector
462 Laboratories) and viewed using a DMRB microscope (Leica) equipped with a Retiga Exi
463 camera (QImaging) and Volocity software (PerkinElmer). The whole area of the coverslip was
464 scanned systematically from top to bottom, capturing FITC, TRITC, DAPI and phase contrast
465 images of each trypanosome. Digital images were analysed using ImageJ
466 (http://rsb.info.nih.gov/ij).

467 ### Tb927.10.12080 gene expression

468 The 3' UTR of Tb927.10.12080 was amplified from genomic DNA of 1738 using the primers
469 5'- GATCCTCGAGTAGTGGCGAGTGTTTACAACAGTGTC and 5'-
470 GATCGGGCCCCTTGTGCGGATCCAAACAA, and inserted immediately downstream of a
471 GFP gene driven by the procyclin promotor from plasmid backbone pHD449, which is
472 designed for insertion into the tubulin locus (**Figure S5**) (Biebinger et al., 1997; Bingle et al.,
473 2001). The plasmid construct was used for stable transfection of procyclic 1738 and following
474 hygromycin selection, clones were tsetse fly transmitted as described by Peacock et al 2011,
475 2014. Flies were dissected 10 to 40 days pi and organs viewed by fluorescence microscopy
476 and imaged live or fixed in 2% paraformaldehyde and stained with DAPI in VECTASHIELD
477 mounting medium (Vector Laboratories).

478 ## scRNAseq data analysis

479 ### Mapping and generation of expression matrices

480 Nextera adaptor sequences were trimmed from fastq files using trim_galore (*-q 20 -a*
481 *CTGTCTCTTATACACATCT --paired --stringency 3 --length 50 -e 0.1)* (v 0.4.3) (Martin, 2011).
482 Trimmed reads were mapped using HISAT2 (*hisat2 --max-intronlen 5000 -p 12*) (v 2.1.0) (Kim
483 et al., 2015) to the *T. b. brucei* 927 genome. The GFF was converted to GTF using the UCSC
484 genome browser tool (Kuhn et al., 2013). Reads were then summed against genes using
485 HTseq (*htseq-count -f bam -r pos -s no -t CDS*) (v 0.7.1) (Anders et al., 2015).

486 ### Assembly of VSG transcripts

487 Because there is a lack of conservation of *VSG*s across *T. brucei* strains, we built a *de novo*
488 transcriptome assembly to identify the *mVSG* transcripts expressed in 1738 and J10. First, we
489 merged the BAM files across the 388 cells and converted to FASTQ using bedtools (v. 2.29.2)
490 (Quinlan and Hall, 2010). Using Trinity (v. 2.1.1) (Grabherr et al., 2011) to assemble the
491 transcripts from this merged file, we detected 53521 'genes' with a mean contig length of 800
492 bp. We then mapped each cell to this assembly using RSEM (v. 1.3.3) to generate a counts
493 matrix and used Transdecoder (v. 5.5.0) to detect open reading frames (Haas et al., 2013).
494 BLASTp (v. 2.9.0) was used to match to putative *VSG*s that had been curated independently
495 from whole genome data as described below. Transcripts with >90% identity were used for
496 further analysis.

497 Genomes for the parent strains J10 and 1738 were assembled from 76 bp paired read Illumina
498 data using SPADES under default parameters. Predicted *mVSG*s were identified to genomic
499 loci, using BLAST against the assembled contigs with a percent identity across the entire
500 transcript >95%, alignments of a raw score of greater than 1000 were further investigated.
501 Additional open reading frames were identified by BLAST alignment of the curated 927
502 annotated CDS set. Nhmmer was used to identify putative *mVSG* promoters from the
503 alignments (Cross et al., 2014; Ginger et al., 2002).

504 ### Quality control and normalization

505 Quality control and visualisation was performed in Scater (v. 1.12.2) (McCarthy et al., 2017).
506 Cell quality was assessed based on the distribution of genes detected per single-cell
507 transcriptome. Cells with fewer than 250 genes or more than 3000 genes detected were

508 removed, as well as cells that had fewer than 5000 total reads. Out of 613 single cells
509 sequenced, 419 passed quality control and were used for downstream analyses. Raw count
510 data was normalized using a deconvolution size factor in Scran (v. 1.16.0) (L. Lun et al., 2016)
511 to account for differences in overall level of expression between cell-types.

512 Cell clustering, projection, and marker genes

513 In order to unbiasedly group transcriptomes based on similar expression profiles, 388 cells
514 collected from the tsetse were clustered using K-means clustering in SC3 (v. 1.12.0) (Kiselev
515 et al., 2017). Dimensionality reduction was performed in Scater (v. 1.12.2) (McCarthy et al.,
516 2017) using Principal Component Analysis (PCA) with the 500 most variable genes based on
517 log2 normalised expression values. Marker genes were identified for each cluster using SC3
518 (AUROC >0.75 & adjusted *p*-value < 0.01).

519 Identification of hybrid parasites and data integration

520 To select different parasite genotypes (J10, 1738, or J10x1738 cross) in the mixed infection
521 treatment, we first FACS sorted based on GFP+ (1738), RFP+ (J10), or GFP+/RFP+ (hybrid)
522 expression. We then used souporcell (*-k 2 -p 2*) (v2.0) (Heaton et al., 2020) to confirm
523 genotype assignment based on SNP profiles from the scRNA-seq.
524
525 To identify genes that were differentially expressed between the two strains, we used Seurat
526 (v3.1.5) to integrated the data by identifying anchors with the FindIntegrationAnchors() and
527 IntegrateData() functions. The data was then clustered using FindNeighbors() with the top 30
528 principal components and FindClusters() with a resolution of 0.5. Differential expression was
529 performed using the FindMarkers() function (adjusted *p*-value < 0.001).

530 Pseudotime and differential expression

531 To assess developmental progression in 1738 salivary glands parasites from day 21-, 24-,
532 and 40-days pi, Slingshot (v. 1.8.0) was used to infer a pseudotime developmental trajectory
533 (Street et al., 2018). Genes differentially expressed over this trajectory were identified using
534 the associationTest() function in TradeSeq (v. 1.4.0) (Van den Berge et al., 2020).
535

536 **Data availability**

537 Raw sequence data are available on the European Nucleotide Archive under accession
538 ERP132258. The expression matrix and associated code will be available on Github at
539 https://github.com/vhowick/tryps_single_cell. The data are explorable via the Glasgow Cell
540 Atlas website at http://cellatlas.mvls.gla.ac.uk/.
541

542 **Acknowledgements**

543 We thank the UK Biotechnology and Biological Sciences Research Council (BB/R016437/1
544 and BB/R010188/1) and Wellcome Trust (http://www.wellcome.ac.uk/) for funding. The
545 Wellcome Sanger Institute is funded by the Wellcome Trust (Grant 206194/Z/17/Z), which
546 supports M.K.N.L. V.M.H. is supported by a Sir Henry Dale Fellowship jointly funded by the
547 Wellcome Trust and the Royal Society (Grant 220185/Z/20/Z). We gratefully acknowledge the
548 generous supply of tsetse pupae from the International Atomic Energy Agency, Vienna. We
549 thank the Sanger Institute flow cytometry core facility for technical support and guidance. We
550 thank Thomas Otto and Jesse Ropp for development and deployment of the interactive
551 website.

552
553 **Author Contributions**
554 VMH, LP, CK, and CC designed and performed all experiments. VMH, LP, CK, and WG
555 analysed the data. WG and MKNL sourced funding, provided scientific direction, and
556 supported experimental design. VMH, LP, CK and WG wrote the manuscript with contributions
557 from all authors.
558
559 **References**
560

561 Anders S, Pyl PT, Huber W. 2015. HTSeq--a Python framework to work with high-throughput
562 sequencing data. *Bioinformatics* **31**:166–169. doi:10.1093/bioinformatics/btu638
563 Attar M, Sharma E, Li S, Bryer C, Cubitt L, Broxholme J, Lockstone H, Kinchen J, Simmons
564 A, Piazza P, Buck D, Livak KJ, Bowden R. 2018. A practical solution for preserving
565 single cells for RNA sequencing. *Sci Rep* **8**:2151. doi:10.1038/s41598-018-20372-7
566 Biebinger S, Wirtz LE, Lorenz P, Clayton C. 1997. Vectors for inducible expression of toxic
567 gene products in bloodstream and procyclic Trypanosoma brucei. *Mol Biochem*
568 *Parasitol* **85**:99–112.
569 Bingle LE, Eastlake JL, Bailey M, Gibson WC. 2001. A novel GFP approach for the analysis
570 of genetic exchange in trypanosomes allowing the in situ detection of mating events.
571 *Microbiology* **147**:3231–3240. doi:10.1099/00221287-147-12-3231
572 Briggs EM, Rojas F, McCulloch R, Matthews KR, Otto TD. 2021. Single-cell transcriptomic
573 analysis of bloodstream Trypanosoma brucei reconstructs cell cycle progression and
574 developmental quorum sensing. *Nat Commun* **12**:5268. doi:10.1038/s41467-021-
575 25607-2
576 Bringaud F, Rivière L, Coustou V. 2006. Energy metabolism of trypanosomatids: adaptation
577 to available carbon sources. *Mol Biochem Parasitol* **149**:1–9.
578 doi:10.1016/j.molbiopara.2006.03.017
579 Bütikofer P, Ruepp S, Boschung M, Roditi I. 1997. “GPEET” procyclin is the major surface
580 protein of procyclic culture forms of Trypanosoma brucei brucei strain 427. *Biochem J*
581 **326 (Pt 2)**:415–423. doi:10.1042/bj3260415
582 Bütikofer P, Vassella E, Mehlert A, Ferguson MAJ, Roditi I. 2002. Characterisation and cellular
583 localisation of a GPEET procyclin precursor in Trypanosoma brucei insect forms. *Mol*
584 *Biochem Parasitol* **119**:87–95. doi:10.1016/s0166-6851(01)00398-x
585 Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. 2018. Integrating single-cell
586 transcriptomic data across different conditions, technologies, and species. *Nat*
587 *Biotechnol* **36**:411–420. doi:10.1038/nbt.4096
588 Christiano R, Kolev NG, Shi H, Ullu E, Walther TC, Tschudi C. 2017. The proteome and
589 transcriptome of the infectious metacyclic form of Trypanosoma brucei define
590 quiescent cells primed for mammalian invasion. *Mol Microbiol* **106**:74–92.
591 doi:10.1111/mmi.13754
592 Cross GAM, Kim H-S, Wickstead B. 2014. Capturing the variant surface glycoprotein
593 repertoire (the VSGnome) of Trypanosoma brucei Lister 427. *Mol Biochem Parasitol*
594 **195**:59–73. doi:10.1016/j.molbiopara.2014.06.004
595 Cunningham I. 1977. New culture medium for maintenance of tsetse tissues and growth of
596 trypanosomatids. *J Protozool* **24**:325–329. doi:10.1111/j.1550-7408.1977.tb00987.x
597 Dean S, Marchetti R, Kirk K, Matthews KR. 2009. A surface transporter family conveys the
598 trypanosome differentiation signal. *Nature* **459**:213–217. doi:10.1038/nature07997

599 Doleželová E, Kunzová M, Dejung M, Levin M, Panicucci B, Regnault C, Janzen CJ, Barrett
600 MP, Butter F, Zíková A. 2020. Cell-based and multi-omics profiling reveals dynamic
601 metabolic repurposing of mitochondria to drive developmental progression of
602 Trypanosoma brucei. *PLoS Biol* **18**:e3000741. doi:10.1371/journal.pbio.3000741
603 Gibson W, Peacock L, Ferris V, Williams K, Bailey M. 2008. The use of yellow fluorescent
604 hybrids to indicate mating in Trypanosoma brucei. *Parasit Vectors* **1**:4.
605 doi:10.1186/1756-3305-1-4
606 Ginger ML, Blundell PA, Lewis AM, Browitt A, Günzl A, Barry JD. 2002. Ex vivo and in vitro
607 identification of a consensus promoter for VSG genes expressed by metacyclic-stage
608 trypanosomes in the tsetse fly. *Eukaryot Cell* **1**:1000–1009. doi:10.1128/EC.1.6.1000-
609 1009.2002
610 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,
611 Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di
612 Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-
613 length transcriptome assembly from RNA-Seq data without a reference genome. *Nat*
614 *Biotechnol* **29**:644–652. doi:10.1038/nbt.1883
615 Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles
616 D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N,
617 Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A.
618 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity
619 platform for reference generation and analysis. *Nat Protoc* **8**:1494–1512.
620 doi:10.1038/nprot.2013.084
621 Heaton H, Talman AM, Knights A, Imaz M, Gaffney DJ, Durbin R, Hemberg M, Lawniczak
622 MKN. 2020. Souporcell: robust clustering of single-cell RNA-seq data by genotype
623 without reference genotypes. *Nat Methods*. doi:10.1038/s41592-020-0820-1
624 Howick VM, Russell AJC, Andrews T, Heaton H, Reid AJ, Natarajan K, Butungi H, Metcalf T,
625 Verzier LH, Rayner JC, Berriman M, Herren JK, Billker O, Hemberg M, Talman AM,
626 Lawniczak MKN. 2019. The Malaria Cell Atlas: Single parasite transcriptomes across
627 the complete Plasmodium life cycle. *Science* **365**. doi:10.1126/science.aaw2619
628 Hutchinson S, Foulon S, Crouzols A, Menafra R, Rotureau B, Griffiths AD, Bastin P. 2021.
629 The establishment of variant surface glycoprotein monoallelic expression revealed by
630 single-cell RNA-seq of Trypanosoma brucei in the tsetse fly salivary glands. *Cold*
631 *Spring Harbor Laboratory*. doi:10.1101/2021.03.01.433049
632 Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory
633 requirements. *Nat Methods* **12**:357–360. doi:10.1038/nmeth.3317
634 Kiselev VY, Kirschner K, Schaub MT, Andrews T, Yiu A, Chandra T, Natarajan KN, Reik W,
635 Barahona M, Green AR, Hemberg M. 2017. SC3: consensus clustering of single-cell
636 RNA-seq data. *Nat Methods* **14**:483–486. doi:10.1038/nmeth.4236
637 Kolev NG, Ramey-Butler K, Cross GAM, Ullu E, Tschudi C. 2012. Developmental progression
638 to infectivity in Trypanosoma brucei triggered by an RNA-binding protein. *Science*
639 **338**:1352–1353. doi:10.1126/science.1229641
640 Kuhn RM, Haussler D, Kent WJ. 2013. The UCSC genome browser and associated tools.
641 *Brief Bioinform* **14**:144–161. doi:10.1093/bib/bbs038
642 L. Lun AT, Bach K, Marioni JC. 2016. Pooling across cells to normalize single-cell RNA
643 sequencing data with many zero counts. *Genome Biol* **17**:75. doi:10.1186/s13059-016-
644 0947-7
645 Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing
646 reads. *EMBnet.journal* **17**:10–12.

647 Matthews KR. 2005. The developmental cell biology of Trypanosoma brucei. *J Cell Sci*
648 **118**:283–290. doi:10.1242/jcs.01649
649 McCarthy DJ, Campbell KR, Lun ATL, Wills QF. 2017. Scater: pre-processing, quality control,
650 normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics*
651 **33**:1179–1186. doi:10.1093/bioinformatics/btw777
652 Naguleswaran A, Doiron N, Roditi I. 2018. RNA-Seq analysis validates the use of culture-
653 derived Trypanosoma brucei and provides new markers for mammalian and insect life-
654 cycle stages. *BMC Genomics* **19**:227. doi:10.1186/s12864-018-4600-6
655 Nötzel C, Kafsack BFC. 2021. There and back again: malaria parasite single-cell
656 transcriptomics comes full circle. *Trends Parasitol*. doi:10.1016/j.pt.2021.07.011
657 Peacock L, Bailey M, Carrington M, Gibson W. 2014. Meiosis and haploid gametes in the
658 pathogen Trypanosoma brucei. *Curr Biol* **24**:181–186. doi:10.1016/j.cub.2013.11.044
659 Peacock L, Ferris V, Sharma R, Sunter J, Bailey M, Carrington M, Gibson W. 2011.
660 Identification of the meiotic life cycle stage of Trypanosoma brucei in the tsetse fly.
661 *Proc Natl Acad Sci U S A* **108**:3671–3676. doi:10.1073/pnas.1019423108
662 Peacock L, Kay C, Farren C, Bailey M, Carrington M, Gibson W. 2021. Sequential production
663 of gametes during meiosis in trypanosomes. *Commun Biol* **4**:555. doi:10.1038/s42003-
664 021-02058-5
665 Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R. 2014. Full-length
666 RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**:171–181.
667 doi:10.1038/nprot.2014.006
668 Poran A, Nötzel C, Aly O, Mencia-Trinchant N, Harris CT, Guzman ML, Hassane DC,
669 Elemento O, Kafsack BFC. 2017. Single-cell RNA sequencing reveals a signature of
670 sexual commitment in malaria parasites. *Nature* **551**:95–99. doi:10.1038/nature24280
671 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic
672 features. *Bioinformatics* **26**:841–842. doi:10.1093/bioinformatics/btq033
673 Ramey-Butler K, Ullu E, Kolev NG, Tschudi C. 2015. Synchronous expression of individual
674 metacyclic variant surface glycoprotein genes in Trypanosoma brucei. *Mol Biochem*
675 *Parasitol* **200**:1–4. doi:10.1016/j.molbiopara.2015.04.001
676 Real E, Howick VM, Dahalan FA, Witmer K, Cudini J, Andradi-Brown C, Blight J, Davidson
677 MS, Dogga SK, Reid AJ, Baum J, Lawniczak MKN. 2021. A single-cell atlas of
678 Plasmodium falciparum transmission through the mosquito. *Nat Commun* **12**:3196.
679 doi:10.1038/s41467-021-23434-z
680 Reid AJ, Talman AM, Bennett HM, Gomes AR, Sanders MJ, Illingworth CJR, Billker O,
681 Berriman M, Lawniczak MKN. 2018. Single-cell RNA-seq reveals hidden
682 transcriptional variation in malaria parasites. *Elife* **7**:e33105. doi:10.7554/eLife.33105
683 Sà JM, Cannon MV, Caleon RL, Wellems TE, Serre D. 2020. Single-cell transcription analysis
684 of Plasmodium vivax blood-stage parasites identifies stage- and species-specific
685 profiles of expression. *PLoS Biol* **18**:e3000711. doi:10.1371/journal.pbio.3000711
686 Savage AF, Cerqueira GC, Regmi S, Wu Y, El Sayed NM, Aksoy S. 2012. Transcript
687 expression analysis of putative Trypanosoma brucei GPI-anchored surface proteins
688 during development in the tsetse and mammalian hosts. *PLoS Negl Trop Dis* **6**:e1708.
689 doi:10.1371/journal.pntd.0001708
690 Savage AF, Kolev NG, Franklin JB, Vigneron A, Aksoy S, Tschudi C. 2016. Transcriptome
691 Profiling of Trypanosoma brucei Development in the Tsetse Fly Vector Glossina
692 morsitans. *PLoS One* **11**:e0168877. doi:10.1371/journal.pone.0168877

693 Sharma R, Peacock L, Gluenz E, Gull K, Gibson W, Carrington M. 2008. Asymmetric cell
694 division as a route to reduction in cell length and change in cell morphology in
695 trypanosomes. *Protist* **159**:137–151. doi:10.1016/j.protis.2007.07.004
696 Sistrom M, Evans B, Bjornson R, Gibson W, Balmer O, Mäser P, Aksoy S, Caccone A. 2014.
697 Comparative genomics reveals multiple genetic backgrounds of human pathogenicity
698 in the Trypanosoma brucei complex. *Genome Biol Evol* **6**:2811–2819.
699 doi:10.1093/gbe/evu222
700 Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, Purdom E, Dudoit S. 2018. Slingshot:
701 cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*
702 **19**:477. doi:10.1186/s12864-018-4772-0
703 Telleria EL, Benoit JB, Zhao X, Savage AF, Regmi S, Alves e Silva TL, O'Neill M, Aksoy S.
704 2014. Insights into the trypanosome-host interactions revealed through transcriptomic
705 analysis of parasitized tsetse fly salivary glands. *PLoS Negl Trop Dis* **8**:e2649.
706 doi:10.1371/journal.pntd.0002649
707 Urwyler S, Studer E, Renggli CK, Roditi I. 2007. A family of stage-specific alanine-rich proteins
708 on the surface of epimastigote forms of Trypanosoma brucei. *Mol Microbiol* **63**:218–
709 228. doi:10.1111/j.1365-2958.2006.05492.x
710 Van Den Abbeele J, Claes Y, van Bockstaele D, Le Ray D, Coosemans M. 1999.
711 Trypanosoma brucei spp. development in the tsetse fly: characterization of the post-
712 mesocyclic stages in the foregut and proboscis. *Parasitology* **118 (Pt 5)**:469–478.
713 doi:10.1017/s0031182099004217
714 Van den Berge K, Roux de Bézieux H, Street K, Saelens W, Cannoodt R, Saeys Y, Dudoit S,
715 Clement L. 2020. Trajectory-based differential expression analysis for single-cell
716 sequencing data. *Nat Commun* **11**:1201. doi:10.1038/s41467-020-14766-3
717 Vassella E, Acosta-Serrano A, Studer E, Lee SH, Englund PT, Roditi I. 2001. Multiple procyclin
718 isoforms are expressed differentially during the development of insect forms of
719 Trypanosoma brucei. *J Mol Biol* **312**:597–607. doi:10.1006/jmbi.2001.5004
720 Vassella E, Den Abbeele JV, Bütikofer P, Renggli CK, Furger A, Brun R, Roditi I. 2000. A
721 major surface glycoprotein of trypanosoma brucei is expressed transiently during
722 development and can be regulated post-transcriptionally by glycerol or hypoxia. *Genes*
723 *Dev* **14**:615–626.
724 Vigneron A, O'Neill MB, Weiss BL, Savage AF, Campbell OC, Kamhawi S, Valenzuela JG,
725 Aksoy S. 2020. Single-cell RNA sequencing of Trypanosoma brucei from tsetse
726 salivary glands unveils metacyclogenesis and identifies potential transmission blocking
727 antigens. *Proc Natl Acad Sci U S A*. doi:10.1073/pnas.1914423117
728 Wang W, Penland L, Gokce O, Croote D, Quake SR. 2018. High fidelity hypothermic
729 preservation of primary tissues in organ transplant preservative for single cell
730 transcriptome analysis. *BMC Genomics* **19**:140. doi:10.1186/s12864-018-4512-5
731 Witmer K, Dahalan FA, Metcalf T, Talman AM, Howick VM, Lawniczak MKN. 2021. Using
732 scRNA-seq to Identify Transcriptional Variation in the Malaria Parasite Ookinete Stage. *Front*
733 *Cell Infect Microbiol* **11**:68. doi:10.3389/fcimb.2021.604129
734
735
736

Supplementary Information

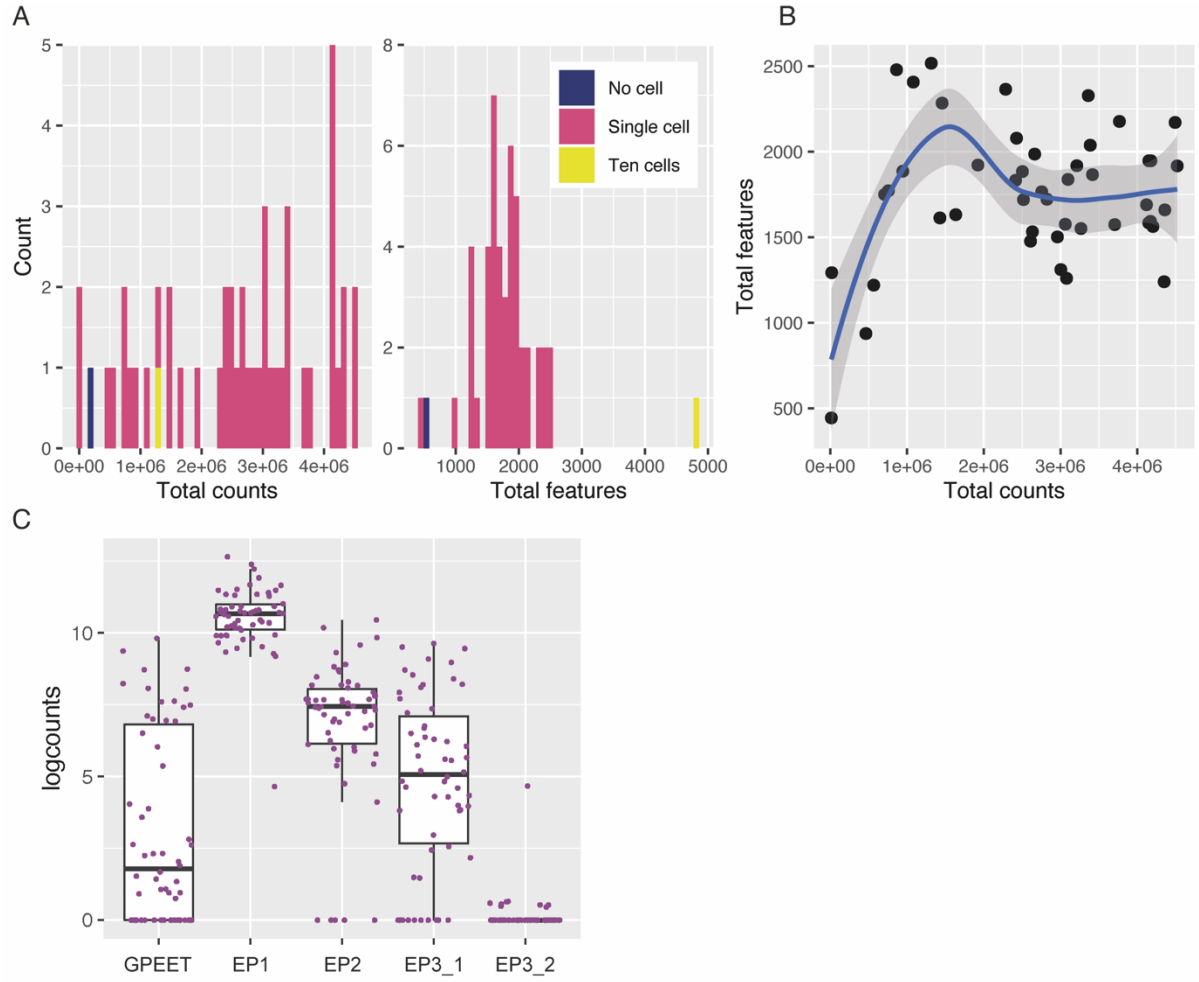


Fig S1. Quality assessment and expression of marker genes in procyclic culture singe-cell transcriptomes

Forty-eight transcriptomes were generated using Smart-seq2 from parasites in a procyclic culture including a no cell and ten cell control. **(A)** The distribution of the total counts (left) and total features (right) detected in these 48 transcriptomes. **(B)** The total features plotted against total counts for the 46 single-cell transcriptomes shows a plateau as features and counts increase, suggesting that sequencing was saturated for these cells. We detected a mean of 2.6×10^6 reads and 1756 features per single-cell transcriptome. **(C)** Expression of procyclic surface antigen genes *GPEET*(Tb927.6.510), *EP1* (Tb927.10.10260), *EP2* (Tb927.10.10250), *EP3_1* (Tb927.6.520), *EP3_2* (Tb927.6.480).

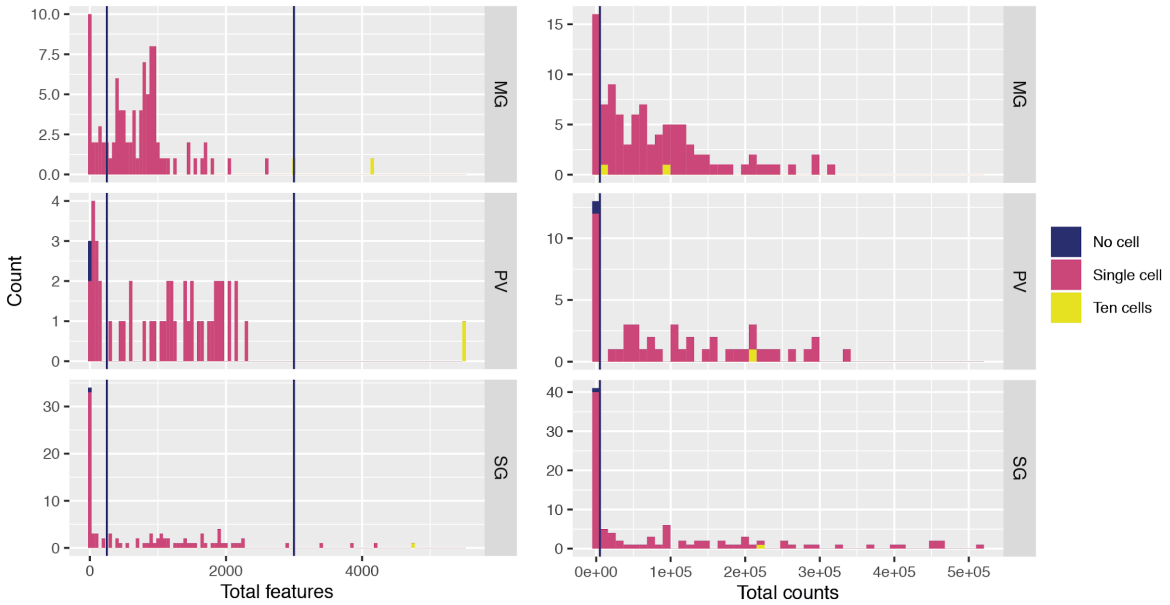


Fig S2. Quality control of Experiment 2 insect stage parasites

The distribution of features (left) and counts (right) in each cell across the three insect tissues: midgut (MG), proventriculus (PV), and salivary glands (SG). Cells with less than 250 or more than 3000 genes per cell were removed. Additionally, cells with less than 5000 reads were removed. Cutoffs are represented by the dark blue vertical lines in each histogram. Of 234 single cells sequenced, 157 passed quality control. After quality control the median number of features per cell was 950. The median number of reads per cell was 99,993.

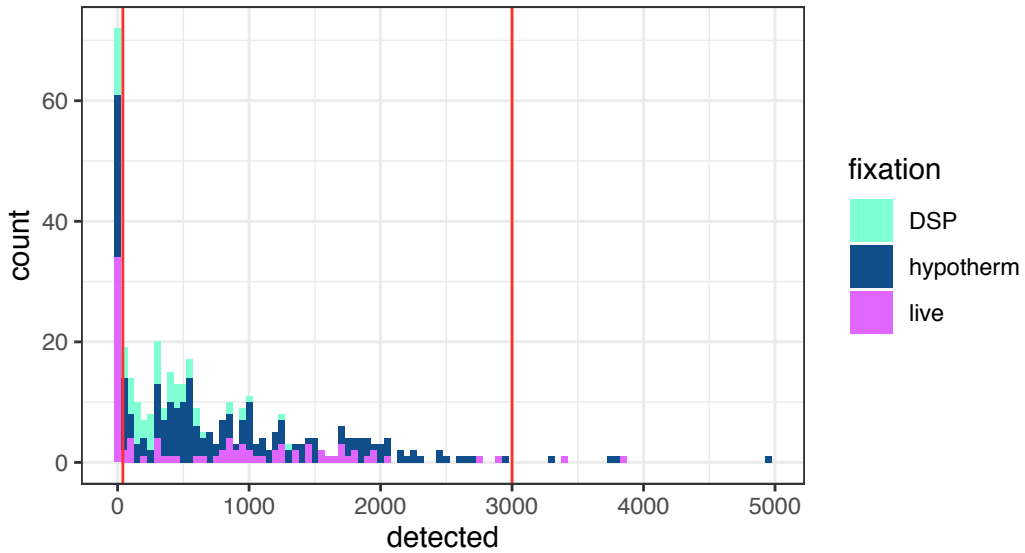


Figure S3. Assessment of preservation methods of salivary gland parasites

The distribution of features detected in salivary gland cells across the two preservation treatments (DSP and hypothermosol) compared to live parasites. Although there were slight differences in detection between the different treatments, caution must be taken in interpreting these differences as the fixation methods are confounded with the different timepoints collected (DSP: day 40; hypothermosol: day 24; live: day 21).

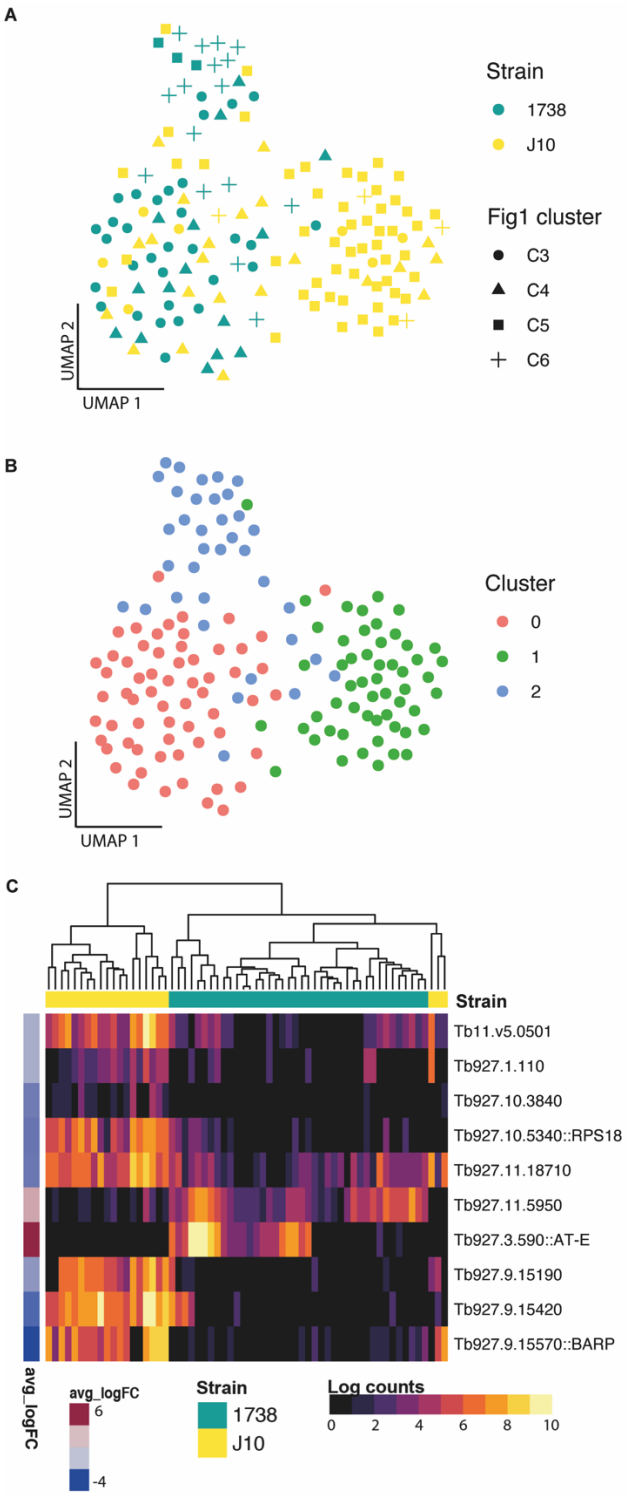


Figure S4. Strain-specific gene expression (A). A UMAP of the day 24 1738 and J10 SG parasites integrated by strain. Points are coloured by strain and shaped by Figure 1 cluster. **(B).** The integrated UMAP coloured by new cluster from the integration analysis. Cluster 0 has a representation of both strains, whereas cluster 1 and 2 are composed primarily of 1738 or J10, respectively. **(C)** Differential expression was performed between strains within cluster 0. The ten genes were differentially expressed between the two strains are displayed on a heatmap.

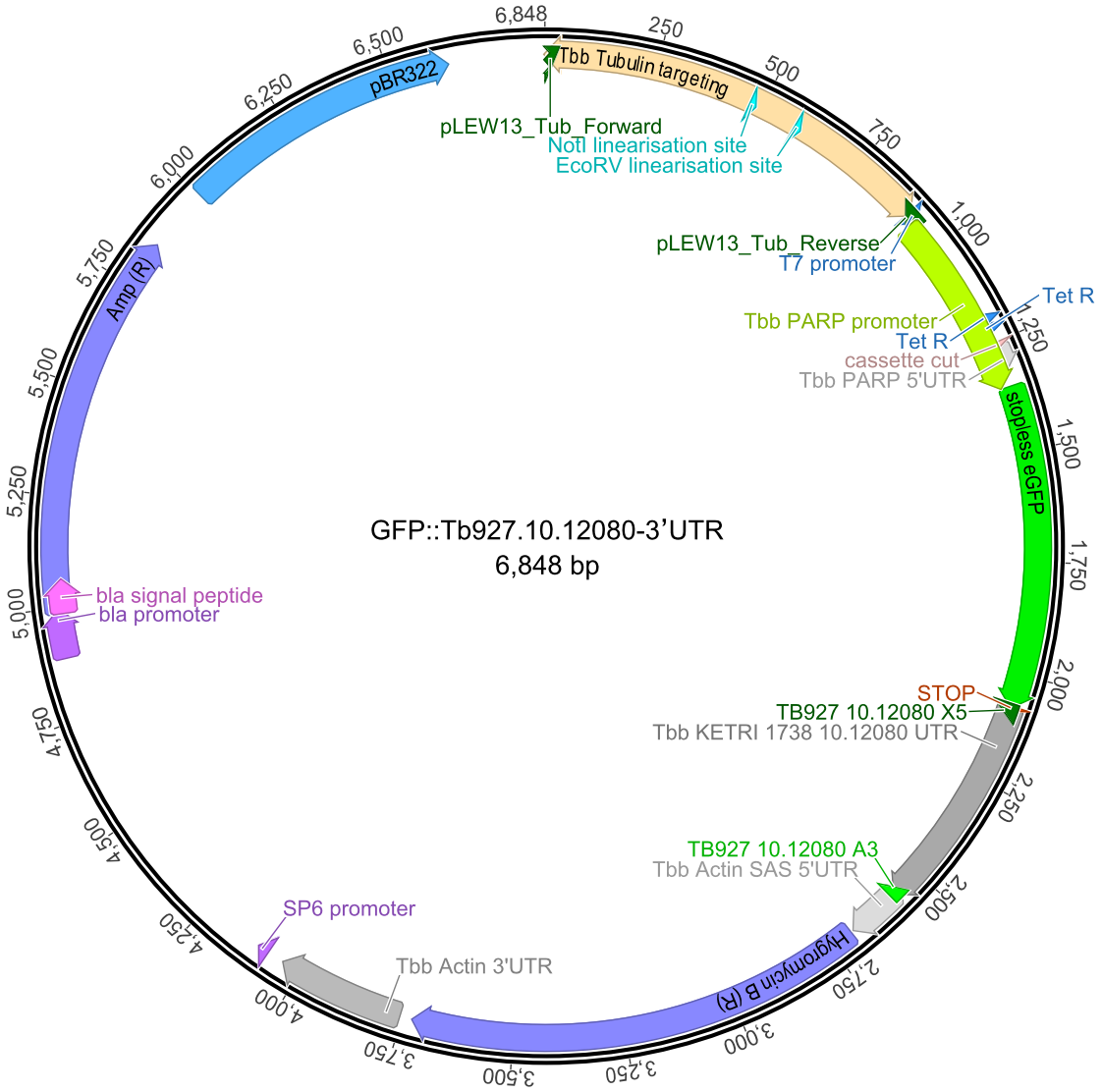


Figure S5. Plasmid map for GFP::Tb927.10.12080-3'UTR construct
Life cycle selective expression of Tb927.10.12080 was investigated through a reporter construct where the expression of GFP was controlled by ~500 bp of UTR downstream of the gene. For this study a stable transformant line was generated in 1738 using the 3' UTR from its endogenous gene and integrated into the tubulin locus.

Supplemental data tables

(A) Table S1 Marker genes from Figure 1
(B) Table S2 Top 200 genes expressed in each cluster
(C) Table S3 DE between strains
(D) Table S4 DE over pseudotime
(E) Table S5 IFA data
(F) Table S6 12080 data

Table S5. Surface protein expression of trypanosomes from salivary glands from tsetse dissected 19-21 days post infected feed, using immunofluorescence.

Cell type	GPEET- GFP				EP-RFP			Total	Notes	
	No	V.faint	Faint	Yes	No	Faint	Yes	cells		
1K1N gamete		2	5	8	0	0	15	15	all GPEET&EP	**Total gametes:**
2K1N gamete	2	3	3	11			19	19	82% GPEET&EP	**90% GPEET&EP, 10% only EP**
2K1N meiotic		1		1			2	2	all GPEET&EP	
3K1N				3			3	3		
4K1N				1			1	1		**Total intermediates:**
1K2N		2		2			4	4		**1/22 (5%) only EP;**
2K2N		1		3			4	4		**21/22 (95%) GPEET&EP**
3K2N	1	1		5			7	7		
4K2N				1			1	1		
1K3N				1			1	1		
3K3N				1			1	1		
1K1N epi	2		2	2	1		5	6	1/6(17%)none;1/6(17%)EP;4/6(66%)GPEET&EP	
1K1N short epi			1			1	0	1	GPEET&EP	
1K1N trypo	3	1	1	1			6	6	3/6 (50%) only EP; 3/6 (50%) GPEET&EP	
2K2N div trypo	1				1		0	1	no GPEET or EP	
1K1N meta	6				6		0	6	no GPEET or EP	
Total	15	11	12	40	8	1	69	78		
Cell type	**BARP- GFP**				**EP-RFP**			**Total**	**Notes**	
	No	**V.faint**	**Faint**	**Yes**	**No**	**V.faint**	**Yes**	**cells**		
1K1N gamete	4		1	10			15	15	73% BARP&EP	**Total gametes:**
2K1N gamete	4	3	2	9			18	18	78% BARP&EP.	**76% BARP&EP, 24% only EP**
2K1N meiotic	1		2	2			5	5	80% BARP&EP	
3K1N	1	2					3	3		
1K2N	2		1	3			6	6		**Total intermediates:**
2K2N	1	2	4	5			12	12		**7/36 (19%) only EP;**
3K2N	2			1			3	3		**29/36 (81%) BARP&EP**
4K2N		1					1	1		
5K2N			1	1			2	2		
2K3N				2			2	2		
3K3N		1	2	2			5	5		
4K3N	1			1			2	2		
1K1N epi		1	2	6			9	9	BARP&EP	
2K1N div epi				3			3	3	BARP&EP	
2K2N div epi	2	1		3			6	6	2/6 (33%) only EP; 4/6 (67%) BARP&EP	
1K1N trypo	2	2	1	2			7	7	2/7 (29%) only EP; 5/7 (71%) BARP&EP	
1K1N pre-meta			1	1	1	1	0	2	1/2 (50%) only BARP; 1/2 (50%) BARP&EP	
1K1N meta	8				8		0	8	no BARP or EP	
Total	28	13	17	51	9	1	99	109		

799 **Table S6.** Cell types recovered from tsetse salivary gland exudate 16-21 days post infection
800 with *T. brucei* 1738 expressing GFP::Tb927.10.12080-3'UTR scored for GFP fluorescence.
801

Cell Type	**Fluorescent**	**Non-Fluorescent**	**Total Cells**
Epimastigote/Dividing epimastigote	0 (0%)	86 (100%)	86
Trypomastigote/Dividing trypomastigote	0 (0%)	24 (100%)	24
Asymmetric Divider	0 (0%)	19 (100%)	19
Pre-metacyclic/Dividing pre-metacyclic	0 (0%)	29 (100%)	29
Metacyclic	0	3 (100%)	3
Trypomastigote-Epimastigote divider	0	4 (100%)	4
Meiotic Divider	3 (10%)	26 (90%)	29
Meiotic intermediate 3N	5 (50%)	5 (50%)	10
Meiotic intermediate 2N	28 (80%)	7 (20%)	35
Meiotic intermediate 3K1N or 4K1N	38 (83%)	8 (17%)	46
Meiotic intermediate - final division to two gametes	34 (79%)	9 (21%)	43
1K1N Gamete	11 (65%)	6 (35%)	17
2K1N Gamete	19 (86%)	3 (14%)	22
Unidentifiable	44 (45%)	53 (55%)	97
TOTAL	**182 (39%)**	**282 (61%)**	**464**

802
803