

***Trypanosoma congolense* Variant Surface Glycoprotein gene expression occurs in the absence of monoallelic transcription control**

Jane C Munday^{1*}, Marija Krasilnikova^{1*}, Catarina A Marques^{1*}, Stephen D Larcombe², Pieter Monsieurs³, Guy Oldrieve², Kevin O Kidambasi², Emma M Briggs⁴, Craig Lapsley¹, Andrew P Jackson⁵, Jan Van Den Abbeele³, Liam J Morrison⁶, Keith R Matthews² and Richard McCulloch^{1**}

Supplementary figure legends

Supplementary Figure 1. Diversity of VSG transcripts expressed over time in culture by wild type *T. congolense* or null mutants of RAD51 or BRCA2. Bulk RNA-seq data from WT, RAD51^{-/-} and BRCA2^{-/-} cells mapped to the *T. congolense* transcriptome, showing the relative proportion of all detected VSG transcripts per sample, with each colour representing a distinct VSG gene. P0, P11 and P22 refer to the passage number of the sample, in duplicate or triplicate. Data show an independent growth experiment, from independent starting clones, to those presented in Figure 1.

Supplementary Figure 2. VSG transcript diversity in wild type *T. congolense* evaluated VSGseq2. The *de novo* VSG transcript assembly and quantification pipeline vsgseq2 was used to analyse bulk RNA-seq data from WT *T. congolense* serial passage samples; the left-hand panels of the figure show the resulting assembled VSGs relative to their expression, with each colour/shade representing a distinct VSG gene. On the right-hand side, the analysis of the same data is shown by Salmon, using the *T. congolense* IL3000 transcriptome of the haplotype-resolved genome, for comparison (the same data are shown in Fig. 1 and Fig. S1). Colours of the transcripts detected by vsgseq2 that pass the following blastn search thresholds are matched to the *T. congolense* IL3000 transcriptome colours: 100% sequence identity, ≥95% query coverage. Those vsgseq2-detected and quantified VSG transcripts that did not pass the above thresholds are plotted in shades of grey.

Supplementary Figure 3. Expressed VSG diversity in *T. congolense* persists regardless of defining criteria or inclusion TPM thresholds. Bulk RNA-seq data (as analysed by salmon) of one replicate each of WT clone 1 P0, P11 and P22 samples, re-analysed using varying minimum TPM values (A) or VSG inclusion criteria (B). Numbers of VSGs detected is shown below each stacked bar.

Supplementary Figure 4. Published IL3000 bulk RNA-seq data shows similar VSG diversity. Previously published (ENA accession number PRJEB41578)³⁵ *T. congolense* IL3000 bulk RNA-seq data is shown mapped to

the *T. congolense* transcriptome, with sequencing run accession numbers for the samples shown below the plot. **A.** Panel shows the relative proportion of all detected VSGs per sample, with each colour representing a distinct VSG gene. **B.** Expression of all annotated genes and pseudogenes, with VSGs in red and non-VSGs in grey.

Supplementary Figure 5. Enrichment of VSG from two clones of wild type *T. congolense*. An SDS-PAGE gel of protein samples used for characterisation of VSGs expressed in two *T. congolense* WT clonal populations (C10, G03). The soluble VSG fractions (sVSG) were examined by mass spectrometry.

Supplementary Figure 6. Quality controls for single cell transcriptomics of *T. brucei* and *T. congolense* *in vitro* samples. **A.** and **B.** Plots showing the cutoffs (dashed lines) used to exclude cells with either low or high number of transcripts, number of genes, percentage of kDNA (mitochondrial) transcripts, or rRNA genes (top panels). Middle panels show the sample pre- and post-QC. Bottom panels show the cutoffs used to exclude cells that had a low percentage of VSG transcripts (of total transcripts) as dashed lines. The last four plots show the cells pre- and post-QC coloured by the percentage of VSG per cell as shown in the legend's gradient. (A) shows the data for *T. brucei*, while (B) shows the data for *T. congolense*.

Supplementary Figure 7. Single cell transcriptomics clusters. Top panels show the Unifold manifold approximation and projection (UMAP) plots for *T. congolense* (left) and *T. brucei* (right), where cells are coloured by cluster as shown in Fig. 4A. Middle panels show dot plots of the top 5 genes identified for each cluster, while the bottom panels show the same information but as heatmaps; both indicate the clusters are relatively well-defined.

Supplementary Figure 8. VSG expression levels in *T. brucei* and *T. congolense*. **A.** Median expression level vs maximum expression level for every gene in *T. brucei* (left) and *T. congolense* (right); VSGs are coloured either blue (*T. brucei*) or purple (*T. congolense*). In *T. brucei* the gene with highest mean and maximum expression is VSG Tbr-000111900/AnTat1.1; in *T. congolense* the gene with highest mean and maximum expression is Tco-IL3000-001881300, an rRNA gene that was annotated as hypothetical in the genome; VSG Tco-IL3000-000997800 is also highlighted. **B.** Violin plots showing the expression levels of the top 50 VSGs (ordered by the number of cells with expression level >4) for *T. brucei* (top) and *T. congolense* (bottom).

Supplementary Figure 9. Percentage of cells expressing different numbers of VSGs and the number of cells expressing specific combinations of VSGs. **–A.** Stacked bar graph showing the percentage of cells expressing different numbers of VSG in the *in vitro* samples. **B.** Data relating to the *T. congolense in vitro* sample, where the number of cells expressing a specific combination of VSGs

(indicated by the dot and connecting line graph) is shown; asterisks highlight the combinations shown in Fig.5D.

Supplementary Figure 10. Expression of minichromosome-resident *T. congolense* VSGs in the scRNA-seq data. Locations of all VSGs is shown on the *T. congolense* minichromosomes (<100 kb in size); those coloured dark green indicate expressed VSGs (expression >2), while those in light green are not detectably expressed. The location of telomeric repeats (TTAGGG) is indicated in grey.

Supplementary Figure 11. Localisation of VSGs expressed in the scRNA-seq data on the larger chromosomes of *T. congolense*. A. Locations of all VSGs is shown on the larger *T. congolense* chromosomes (the >800 kb megabase chromosomes, and two intermediate chromosomes, Tc_intermediate_02); those coloured green indicate expressed VSGs (expression >2), while those in black are not detectably expressed. B. –Distribution of genes around expressed VSGs with predicted functions based on Pfam domains.

Supplementary Figure 12. VSG expression differs in culture and in mice. A. *T. congolense* IL3000 *in vitro* (top, purple) and *in vivo* (bottom, salmon) data shown as: left panel, percentage of total VSG transcripts in the sample (left violin plot); middle panel, dot plot showing the expression level (y-axis) of a VSG in a given cell and how many reads (UMIs, x-axis) does that expression level correspond to (note: a cell is represented multiple times if expressing more than one VSG); right panel, sensitivity curve showing the number of cells called as expressing more than one VSG when one VSG is considered expressed if one, two or more UMIs are detected in the cell. B. Same as in (A) but for *T. brucei* samples. C. Percentage of cells expressing more than one VSG for all the samples when the expression threshold employed to call a VSG as expressed is set to 1, 1.5 or 2.

Supplementary Figure 13. VSG co-expression sensitivity curves post SoupX. Sensitivity curves are shown for four samples: *T. congolense* IL3000 *in vitro* (purple) and *in vivo* (salmon; only the day 6 sample was used), and *T. brucei* *in vitro* (blue) and *in vivo* (yellow; only the d7mS1 sample was used), after using SoupX to clean ambient RNA contamination using rho values of 0.05, 0.1 or 0.15 (corresponding to 5%, 10% and 15% contamination, respectively).

Name	Sequence (5' - 3')	Index	Well
JM125 nanopore-anchored oligodT	ACTTGCCTGCTGCTCTATCTTCNVTTTTTTTTTTTTTTTTTTTT		
5endSL_8mer_001	TTTCTGTTGGTGCTGATATTGCTATCTAATACAGTTTCTGTACTATATTG	TATCTAAT	A1
5endSL_8mer_002	TTTCTGTTGGTGCTGATATTGCTCGAAACAACAGTTTCTGTACTATATTG	TCGAAACA	A2
5endSL_8mer_003	TTTCTGTTGGTGCTGATATTGCGGACGAGCACAGTTTCTGTACTATATTG	GGACGAGC	A3
5endSL_8mer_004	TTTCTGTTGGTGCTGATATTGCAAACCCGTGACAGTTTCTGTACTATATTG	AAACCCGTG	A4
5endSL_8mer_005	TTTCTGTTGGTGCTGATATTGCTGTGGGTAACAGTTTCTGTACTATATTG	TGTGGGTA	A5
5endSL_8mer_006	TTTCTGTTGGTGCTGATATTGCAACGGACAACAGTTTCTGTACTATATTG	AACGGACA	A6
5endSL_8mer_007	TTTCTGTTGGTGCTGATATTGCCCTGTGCAACAGTTTCTGTACTATATTG	CCTGTGCA	A7
5endSL_8mer_008	TTTCTGTTGGTGCTGATATTGCCCGACTTCACAGTTTCTGTACTATATTG	CCGACTTC	A8
5endSL_8mer_009	TTTCTGTTGGTGCTGATATTGCATATGCTAACAGTTTCTGTACTATATTG	ATATGCTA	A9
5endSL_8mer_010	TTTCTGTTGGTGCTGATATTGCTAGTAATGACAGTTTCTGTACTATATTG	TAGTAATG	A10
5endSL_8mer_011	TTTCTGTTGGTGCTGATATTGCTAATCAGGACAGTTTCTGTACTATATTG	TAATCAGG	A11
5endSL_8mer_012	TTTCTGTTGGTGCTGATATTGCCGGCACGGACAGTTTCTGTACTATATTG	CGGCACGG	A12
5endSL_8mer_013	TTTCTGTTGGTGCTGATATTGCCCTAAGTAACAGTTTCTGTACTATATTG	CCTAAGTA	B1
5endSL_8mer_014	TTTCTGTTGGTGCTGATATTGCAGCGACTAACAGTTTCTGTACTATATTG	AGCGACTA	B2
5endSL_8mer_015	TTTCTGTTGGTGCTGATATTGCCAGTATACACAGTTTCTGTACTATATTG	CAGTATAC	B3
5endSL_8mer_016	TTTCTGTTGGTGCTGATATTGCGGACGGCTACAGTTTCTGTACTATATTG	GGACGGCT	B4
5endSL_8mer_017	TTTCTGTTGGTGCTGATATTGCTCTTATCGACAGTTTCTGTACTATATTG	TCTTATCG	B5
5endSL_8mer_018	TTTCTGTTGGTGCTGATATTGCCGGGCCCTTACAGTTTCTGTACTATATTG	CGGGCCTT	B6
5endSL_8mer_019	TTTCTGTTGGTGCTGATATTGCAATTACACACAGTTTCTGTACTATATTG	AATTACAC	B7
5endSL_8mer_020	TTTCTGTTGGTGCTGATATTGCTTATCGTTACAGTTTCTGTACTATATTG	TTATCGTT	B8
5endSL_8mer_021	TTTCTGTTGGTGCTGATATTGCCTAGAGCGACAGTTTCTGTACTATATTG	CTAGAGCG	B9
5endSL_8mer_022	TTTCTGTTGGTGCTGATATTGCGTACTTCAACAGTTTCTGTACTATATTG	GTAATTCA	B10
5endSL_8mer_023	TTTCTGTTGGTGCTGATATTGCGTCCGCAAACAGTTTCTGTACTATATTG	GTCCGCAA	B11
5endSL_8mer_024	TTTCTGTTGGTGCTGATATTGCGGCTCTCGACAGTTTCTGTACTATATTG	GGCTCTCG	B12
5endSL_8mer_025	TTTCTGTTGGTGCTGATATTGCGTTAAAAATACAGTTTCTGTACTATATTG	GTAAAAAT	C1
5endSL_8mer_026	TTTCTGTTGGTGCTGATATTGCAGGTAATGACAGTTTCTGTACTATATTG	AGGTAATG	C2
5endSL_8mer_027	TTTCTGTTGGTGCTGATATTGCACAGCTCGACAGTTTCTGTACTATATTG	ACAGCTCG	C3
5endSL_8mer_028	TTTCTGTTGGTGCTGATATTGCACACTGCGACAGTTTCTGTACTATATTG	ACACTGCG	C4
5endSL_8mer_029	TTTCTGTTGGTGCTGATATTGCTCAGCCCGACAGTTTCTGTACTATATTG	TCAGCCCG	C5
5endSL_8mer_030	TTTCTGTTGGTGCTGATATTGCTCCCCCAGACAGTTTCTGTACTATATTG	TCCCCCAG	C6
5endSL_8mer_031	TTTCTGTTGGTGCTGATATTGCGCCGTTCTACAGTTTCTGTACTATATTG	GCCGTTCT	C7
5endSL_8mer_032	TTTCTGTTGGTGCTGATATTGCACTTTGAAACAGTTTCTGTACTATATTG	ACTTTGAA	C8
5endSL_8mer_033	TTTCTGTTGGTGCTGATATTGCTCTTGACCACAGTTTCTGTACTATATTG	TCTTGACC	C9
5endSL_8mer_034	TTTCTGTTGGTGCTGATATTGCCATCTTTTACAGTTTCTGTACTATATTG	CATCTTTT	C10
5endSL_8mer_035	TTTCTGTTGGTGCTGATATTGCCCTTACCAAACAGTTTCTGTACTATATTG	CTTACCAA	C11
5endSL_8mer_036	TTTCTGTTGGTGCTGATATTGCAAAGTACTACAGTTTCTGTACTATATTG	AAAGTACT	C12
5endSL_8mer_037	TTTCTGTTGGTGCTGATATTGCTCTTCCCGACAGTTTCTGTACTATATTG	TCTTCCCG	D1
5endSL_8mer_038	TTTCTGTTGGTGCTGATATTGCTAGGCCCCACAGTTTCTGTACTATATTG	TAGGCCCC	D2
5endSL_8mer_039	TTTCTGTTGGTGCTGATATTGCGCACGCCGACAGTTTCTGTACTATATTG	GCACGCCG	D3
5endSL_8mer_040	TTTCTGTTGGTGCTGATATTGCGAGTAGCGACAGTTTCTGTACTATATTG	GAGTAGCG	D4

5endSL_8mer_041	TTTCTGTTGGTGCTGATATTGCCGAGGGGAACAGTTTCTGTACTATATTG	CGAGGGGA	D5
5endSL_8mer_042	TTTCTGTTGGTGCTGATATTGCACGTCGTAACAGTTTCTGTACTATATTG	ACGTCGTA	D6
5endSL_8mer_043	TTTCTGTTGGTGCTGATATTGCCGTTTCTGACAGTTTCTGTACTATATTG	CGTTTCTG	D7
5endSL_8mer_044	TTTCTGTTGGTGCTGATATTGCAATCAAACGACAGTTTCTGTACTATATTG	ATCAAACG	D8
5endSL_8mer_045	TTTCTGTTGGTGCTGATATTGCCCTCTACGGACAGTTTCTGTACTATATTG	CTCTACGG	D9
5endSL_8mer_046	TTTCTGTTGGTGCTGATATTGCCAGTGGGAACAGTTTCTGTACTATATTG	CAGTGGGA	D10
5endSL_8mer_047	TTTCTGTTGGTGCTGATATTGCTCTACTTTACAGTTTCTGTACTATATTG	TCTACTTT	D11
5endSL_8mer_048	TTTCTGTTGGTGCTGATATTGCCCTTCGCTACAGTTTCTGTACTATATTG	CCTTCGCT	D12
5endSL_8mer_049	TTTCTGTTGGTGCTGATATTGCAATGGTTAACAGTTTCTGTACTATATTG	AATGGTTA	E1
5endSL_8mer_050	TTTCTGTTGGTGCTGATATTGCCGTGCTTCCACAGTTTCTGTACTATATTG	CTGCTTCC	E2
5endSL_8mer_051	TTTCTGTTGGTGCTGATATTGCAAAACGCAACAGTTTCTGTACTATATTG	AAAACGCA	E3
5endSL_8mer_052	TTTCTGTTGGTGCTGATATTGCCCTCAATTACAGTTTCTGTACTATATTG	CCTCAATT	E4
5endSL_8mer_053	TTTCTGTTGGTGCTGATATTGCACCAATACACAGTTTCTGTACTATATTG	ACCAATAC	E5
5endSL_8mer_054	TTTCTGTTGGTGCTGATATTGCACCGAGCGACAGTTTCTGTACTATATTG	ACCGAGCG	E6
5endSL_8mer_055	TTTCTGTTGGTGCTGATATTGCAAGGACAAACAGTTTCTGTACTATATTG	AAGGACAA	E7
5endSL_8mer_056	TTTCTGTTGGTGCTGATATTGCAGGGCTATACAGTTTCTGTACTATATTG	AGGGCTAT	E8
5endSL_8mer_057	TTTCTGTTGGTGCTGATATTGCACATTATCACAGTTTCTGTACTATATTG	ACATTATC	E9
5endSL_8mer_058	TTTCTGTTGGTGCTGATATTGCCACAAAGCACAGTTTCTGTACTATATTG	CACAAAGC	E10
5endSL_8mer_059	TTTCTGTTGGTGCTGATATTGCCACGTCGCACAGTTTCTGTACTATATTG	CACGTCGC	E11
5endSL_8mer_060	TTTCTGTTGGTGCTGATATTGCTCCGTGACACAGTTTCTGTACTATATTG	TCCGTGAC	E12
5endSL_8mer_061	TTTCTGTTGGTGCTGATATTGCGGAAGTCTACAGTTTCTGTACTATATTG	GGAAGTCT	F1
5endSL_8mer_062	TTTCTGTTGGTGCTGATATTGCCGAATTTAACAGTTTCTGTACTATATTG	CGAATTTA	F2
5endSL_8mer_063	TTTCTGTTGGTGCTGATATTGCTATGTTGCACAGTTTCTGTACTATATTG	TATGTTGC	F3
5endSL_8mer_064	TTTCTGTTGGTGCTGATATTGCTTTCTTTTACAGTTTCTGTACTATATTG	TTTCTTTT	F4
5endSL_8mer_065	TTTCTGTTGGTGCTGATATTGCACAACCTCTACAGTTTCTGTACTATATTG	ACAACCTCT	F5
5endSL_8mer_066	TTTCTGTTGGTGCTGATATTGCATATAGAAACAGTTTCTGTACTATATTG	ATATAGAA	F6
5endSL_8mer_067	TTTCTGTTGGTGCTGATATTGCGATGCTGTACAGTTTCTGTACTATATTG	GATGCTGT	F7
5endSL_8mer_068	TTTCTGTTGGTGCTGATATTGCGAGTATGCACAGTTTCTGTACTATATTG	GAGTATGC	F8
5endSL_8mer_069	TTTCTGTTGGTGCTGATATTGCAATGTCACACAGTTTCTGTACTATATTG	AATGTCAC	F9
5endSL_8mer_070	TTTCTGTTGGTGCTGATATTGCTGCTCTGCACAGTTTCTGTACTATATTG	TGCTCTGC	F10
5endSL_8mer_071	TTTCTGTTGGTGCTGATATTGCCGTGGGGGTACAGTTTCTGTACTATATTG	CTGGGGGT	F11
5endSL_8mer_072	TTTCTGTTGGTGCTGATATTGCCCTTTTAAACAGTTTCTGTACTATATTG	CCTTTTAA	F12
5endSL_8mer_073	TTTCTGTTGGTGCTGATATTGCTGGTGACAGTTTCTGTACTATATTG	TGGTGCAG	G1
5endSL_8mer_074	TTTCTGTTGGTGCTGATATTGCAGTTTGTAAACAGTTTCTGTACTATATTG	AGTTTGTG	G2
5endSL_8mer_075	TTTCTGTTGGTGCTGATATTGCGCTGTCTGACAGTTTCTGTACTATATTG	GCTGTCTG	G3
5endSL_8mer_076	TTTCTGTTGGTGCTGATATTGCAATCTAGTACAGTTTCTGTACTATATTG	AATCTAGT	G4
5endSL_8mer_077	TTTCTGTTGGTGCTGATATTGCGCGATGCAACAGTTTCTGTACTATATTG	GCGATGCA	G5
5endSL_8mer_078	TTTCTGTTGGTGCTGATATTGCCCGCGCAGACAGTTTCTGTACTATATTG	CCGCGCAG	G6
5endSL_8mer_079	TTTCTGTTGGTGCTGATATTGCACATTTGAACAGTTTCTGTACTATATTG	ACATTTGA	G7
5endSL_8mer_080	TTTCTGTTGGTGCTGATATTGCCGGTGCTGACAGTTTCTGTACTATATTG	CGGTGCTG	G8
5endSL_8mer_081	TTTCTGTTGGTGCTGATATTGCTGACTGTACAGTTTCTGTACTATATTG	TGACTGTC	G9
5endSL_8mer_082	TTTCTGTTGGTGCTGATATTGCCATCTTTGACAGTTTCTGTACTATATTG	CTATCTTG	G10
5endSL_8mer_083	TTTCTGTTGGTGCTGATATTGCAGCCATAAACAGTTTCTGTACTATATTG	AGCCATAA	G11
5endSL_8mer_084	TTTCTGTTGGTGCTGATATTGCCCATCGAAACAGTTTCTGTACTATATTG	CCATCGAA	G12

5endSL_8mer_085	TTTCTGTTGGTGCTGATATTGCC7CGTACTACAGTTTCTGTACTATATTG	CTCGTACT	H1
5endSL_8mer_086	TTTCTGTTGGTGCTGATATTGCA7CGTAGCACAGTTTCTGTACTATATTG	ATCGTAGC	H2
5endSL_8mer_087	TTTCTGTTGGTGCTGATATTGCGCGCC7TGACAGTTTCTGTACTATATTG	GCGCCTTG	H3
5endSL_8mer_088	TTTCTGTTGGTGCTGATATTGCAGCAGGAGACAGTTTCTGTACTATATTG	AGCAGGAG	H4
5endSL_8mer_089	TTTCTGTTGGTGCTGATATTGCACACACCGACAGTTTCTGTACTATATTG	ACACACCG	H5
5endSL_8mer_090	TTTCTGTTGGTGCTGATATTGCAGAAGAGGACAGTTTCTGTACTATATTG	AGAAGAGG	H6
5endSL_8mer_091	TTTCTGTTGGTGCTGATATTGCT7ATTAGAACAGTTTCTGTACTATATTG	TTATTAGA	H7
5endSL_8mer_092	TTTCTGTTGGTGCTGATATTGCGGTCTGAAACAGTTTCTGTACTATATTG	GGTCTGAA	H8
5endSL_8mer_093	TTTCTGTTGGTGCTGATATTGCAATCAACGACAGTTTCTGTACTATATTG	AATCAACG	H9
5endSL_8mer_094	TTTCTGTTGGTGCTGATATTGCA7CAACGAACAGTTTCTGTACTATATTG	ATCAACGA	H10
5endSL_8mer_095	TTTCTGTTGGTGCTGATATTGCTATGGAATACAGTTTCTGTACTATATTG	TATGGAAT	H11
5endSL_8mer_096	TTTCTGTTGGTGCTGATATTGCAAGCTAAGACAGTTTCTGTACTATATTG	AAGCTAAG	H12

97

98

Fig.S1

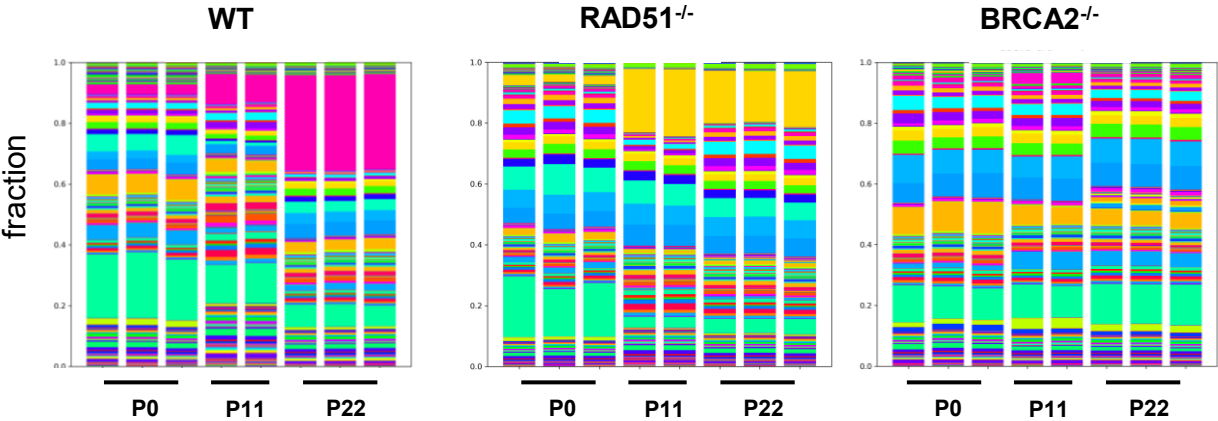


Fig.S2

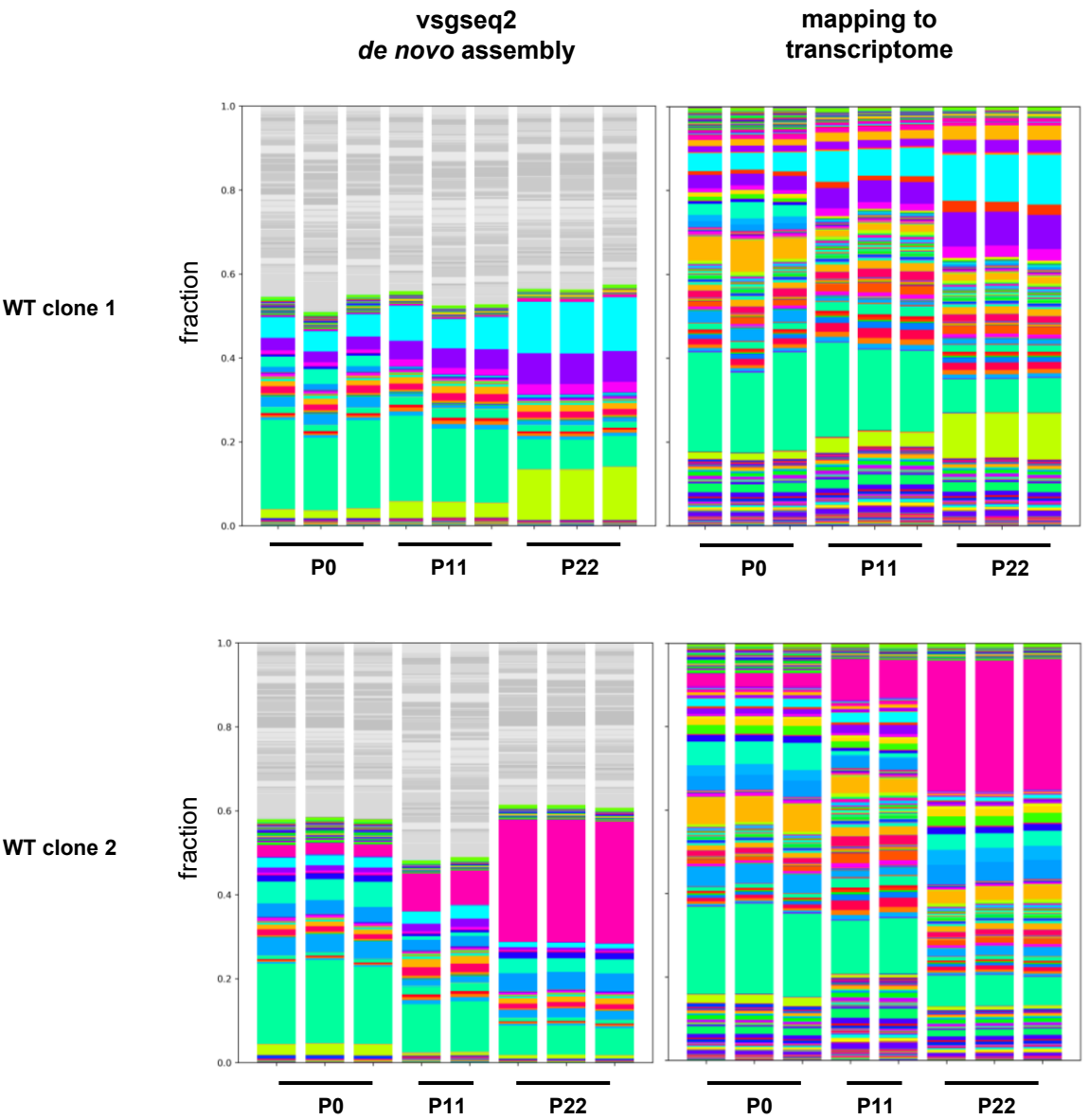
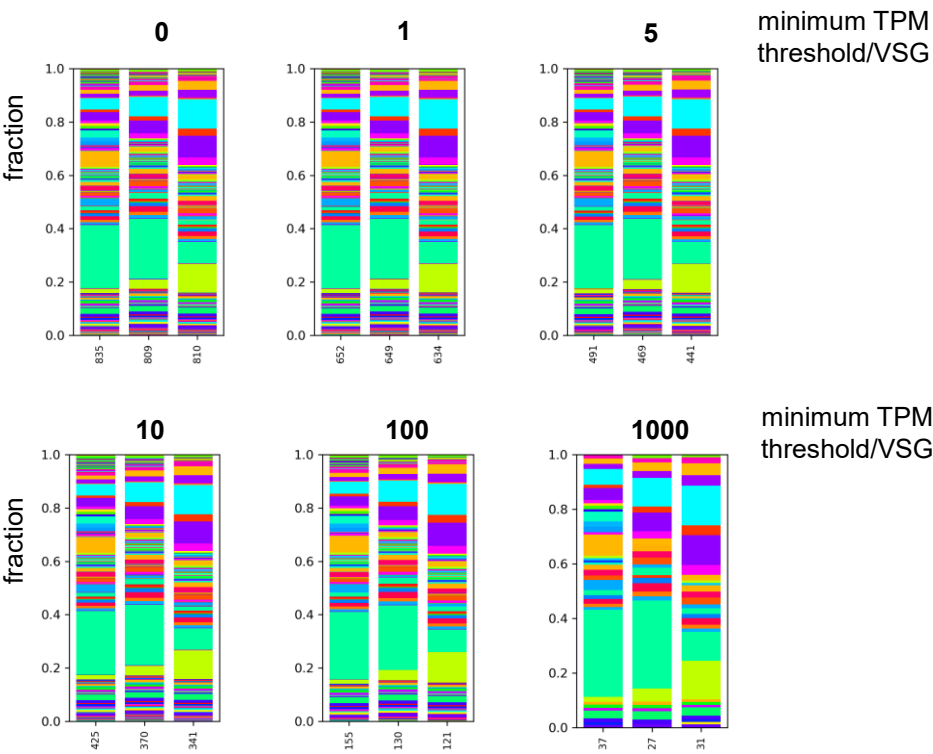


Fig.S3

A



B

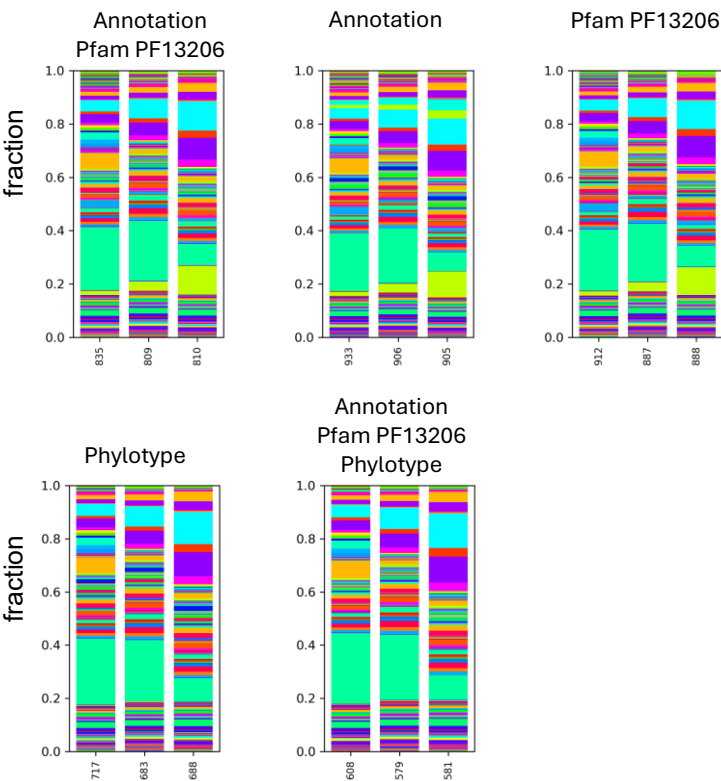
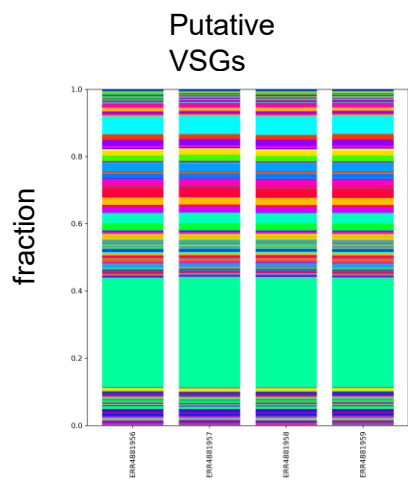


Fig.S4

A



B

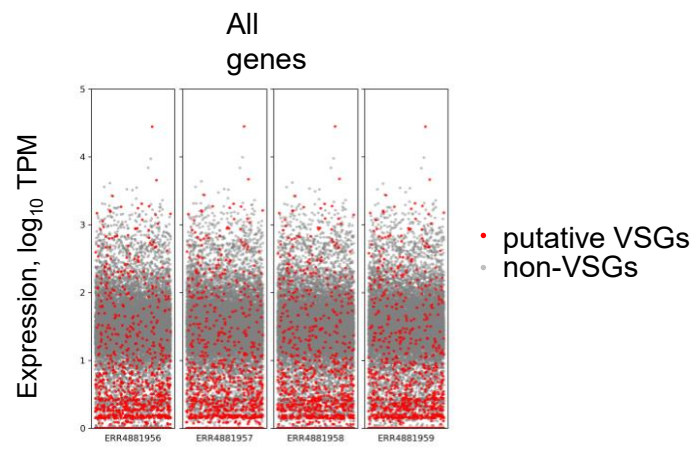


Fig.S5

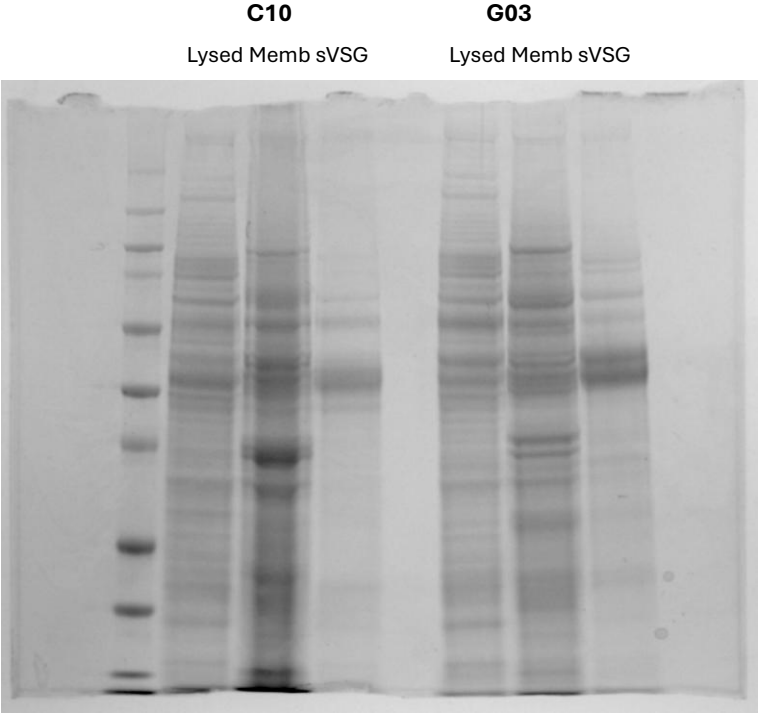


Fig.S6

A

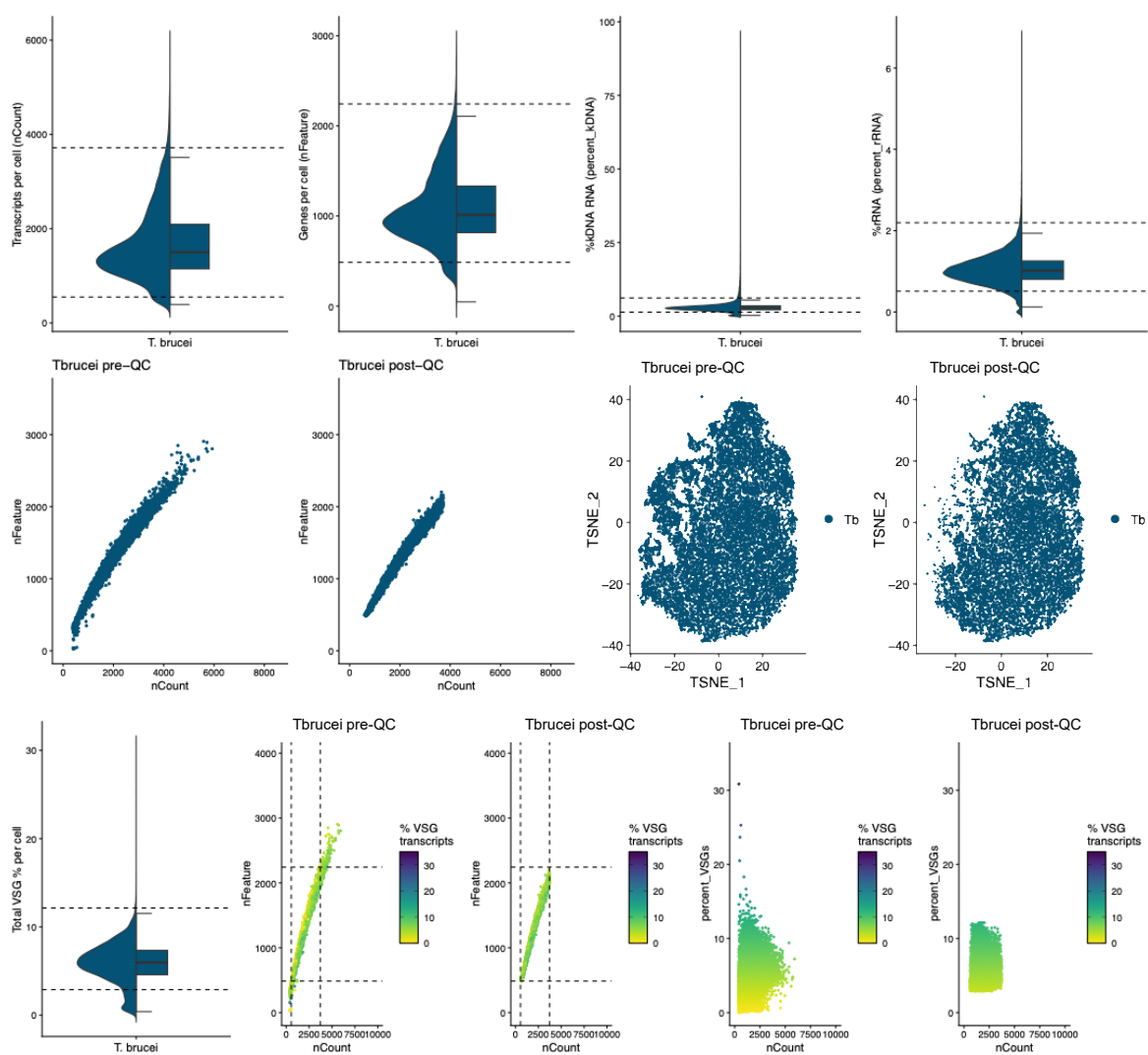


Fig.S6 (continue)

B

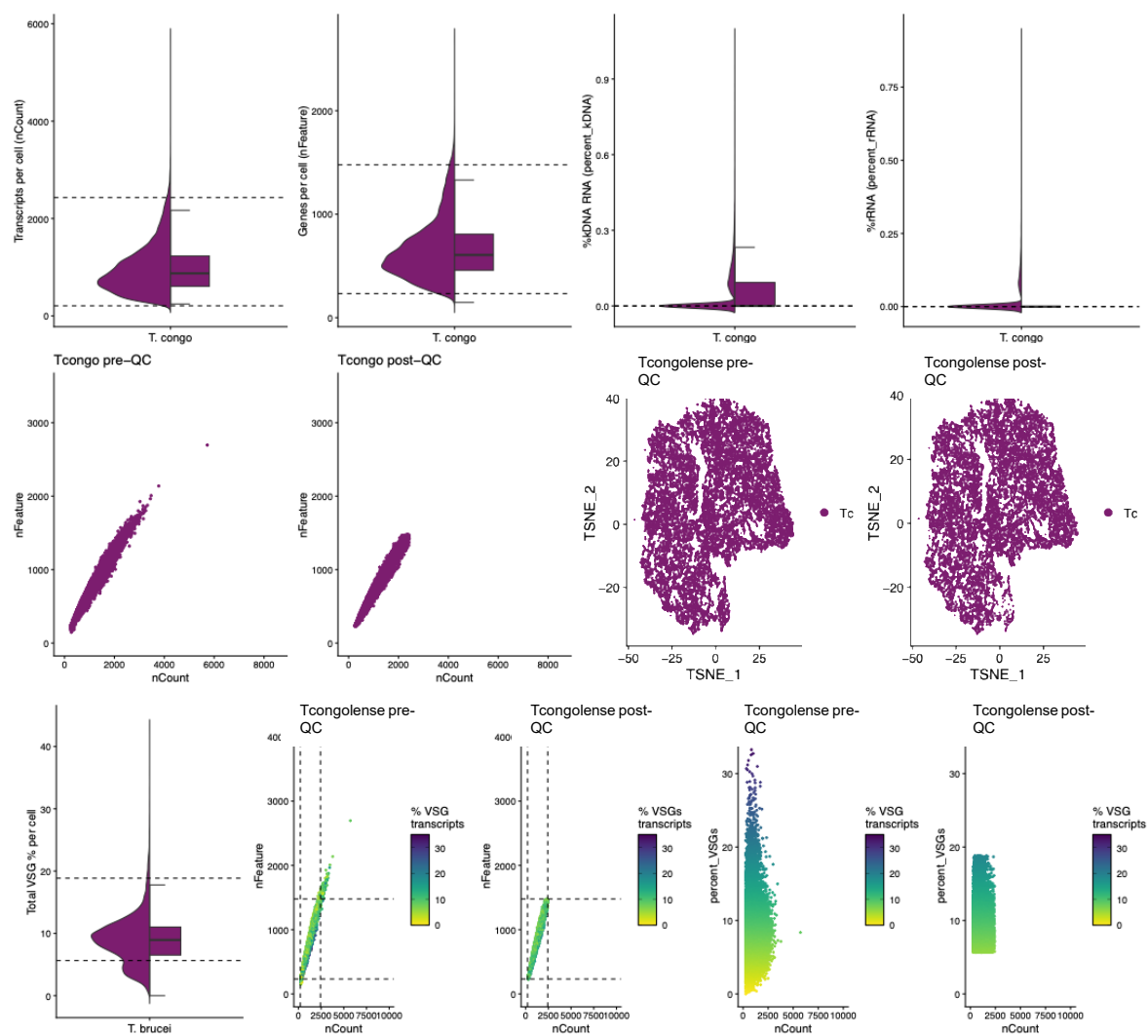


Fig.S7

T. brucei

T. congolense

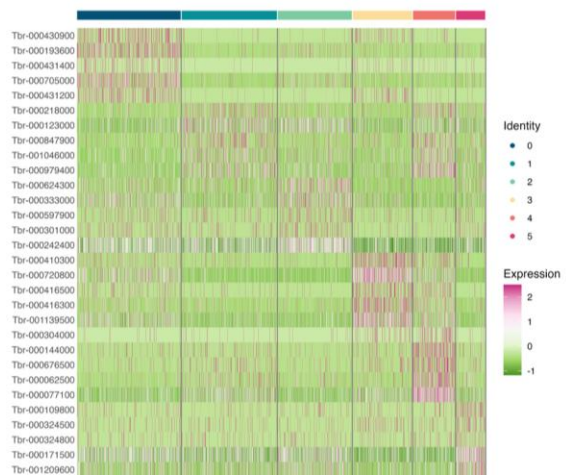
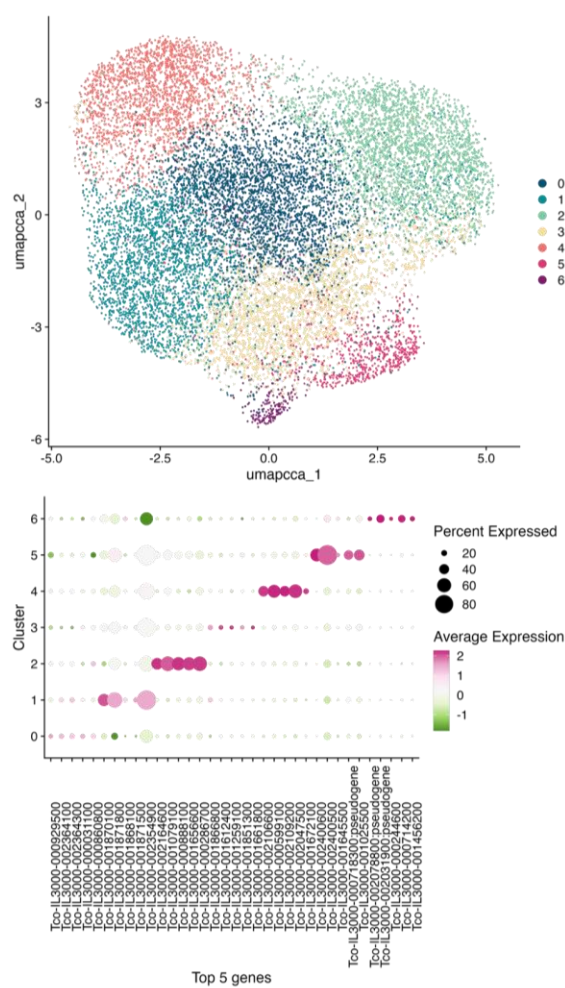
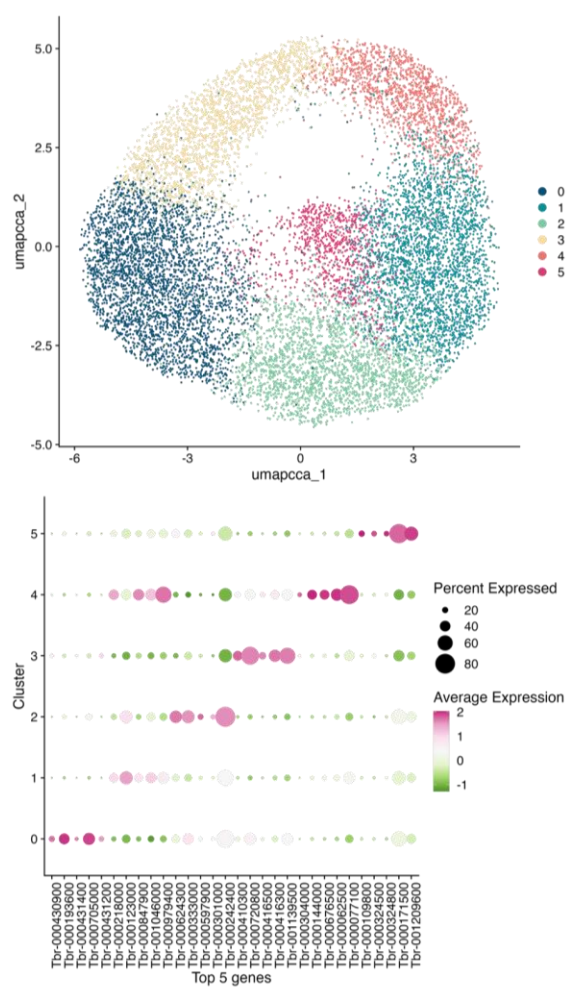
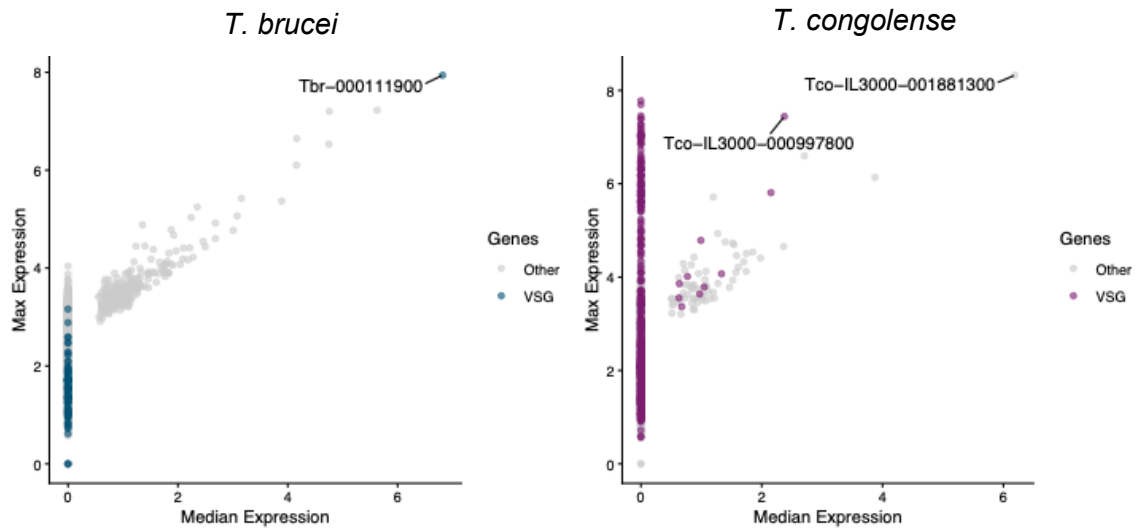


Fig.S8

A



B

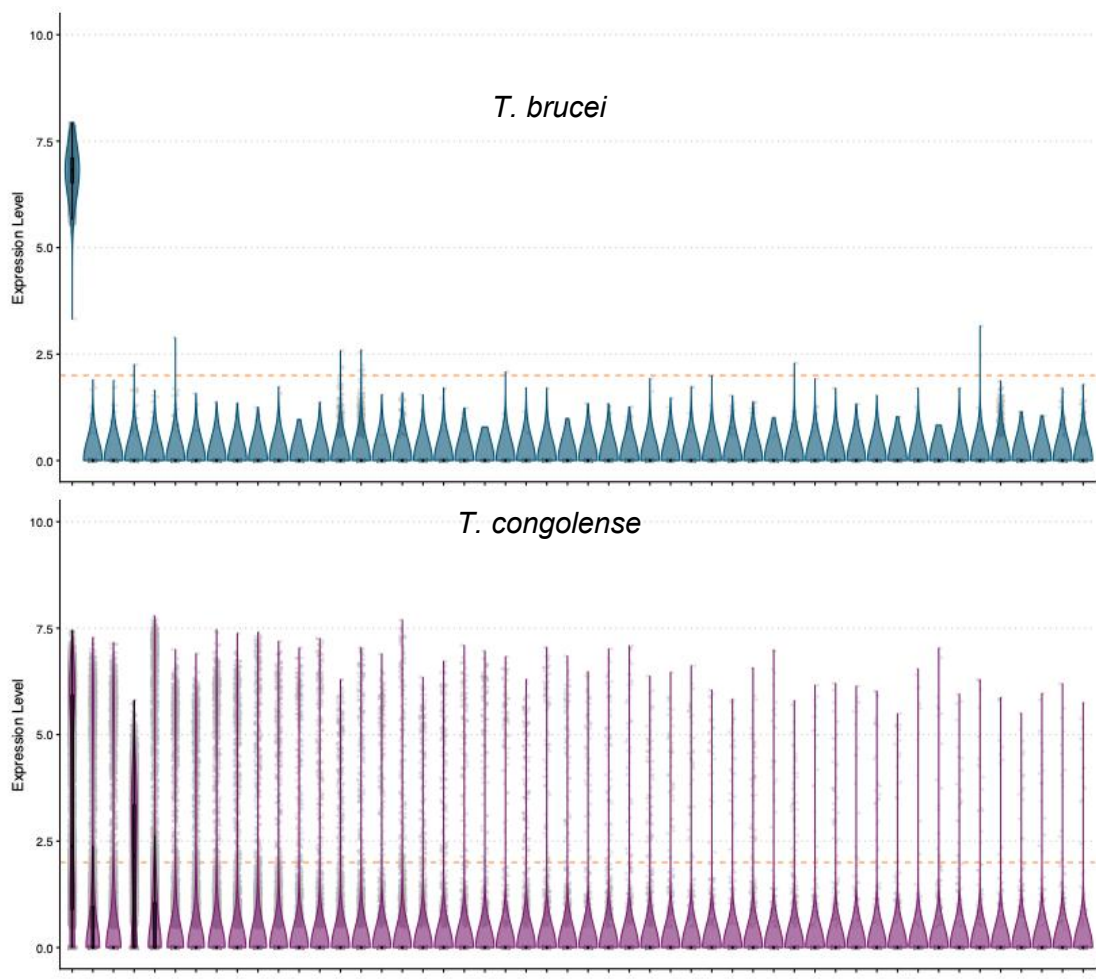


Fig.S9

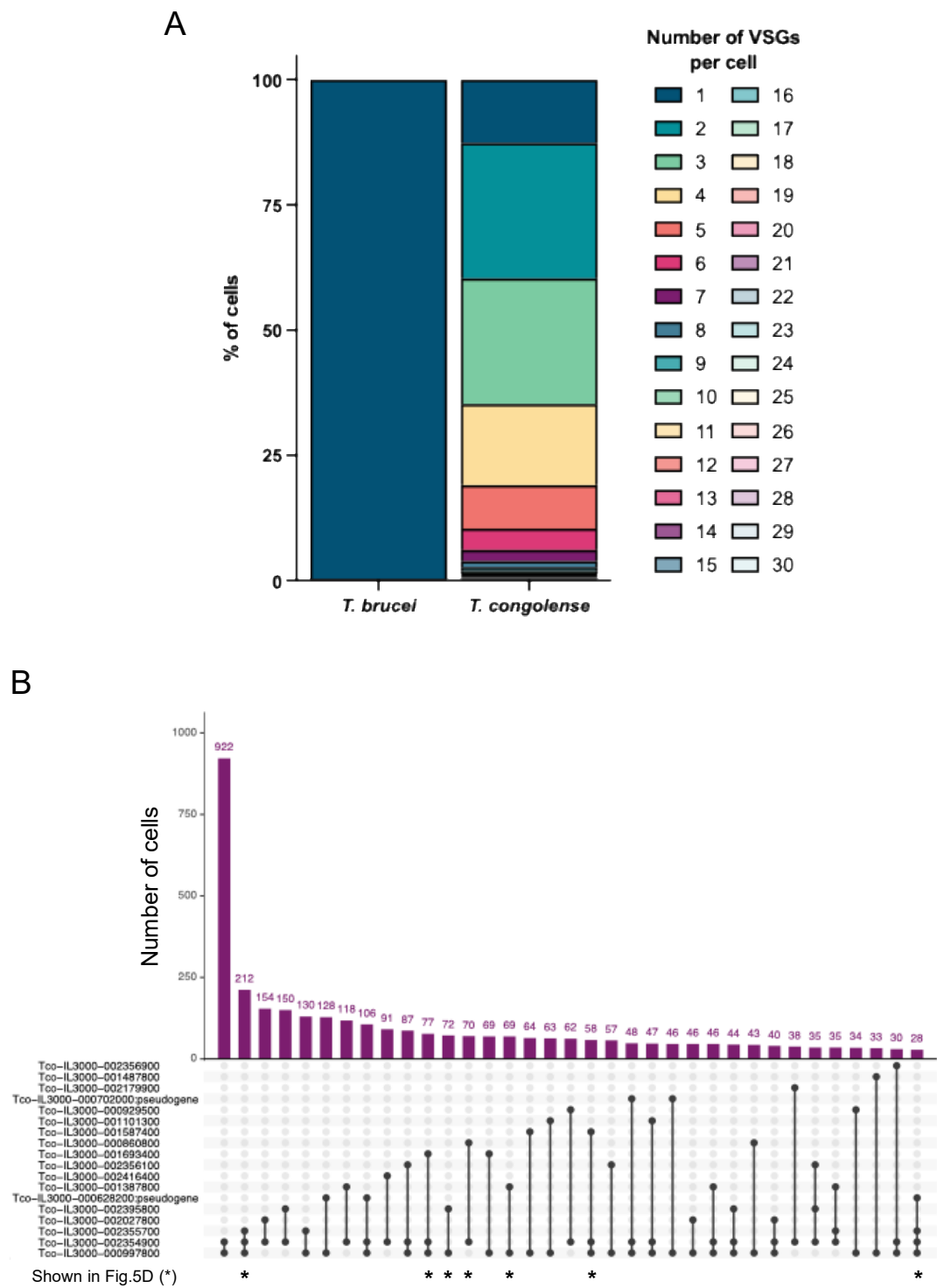
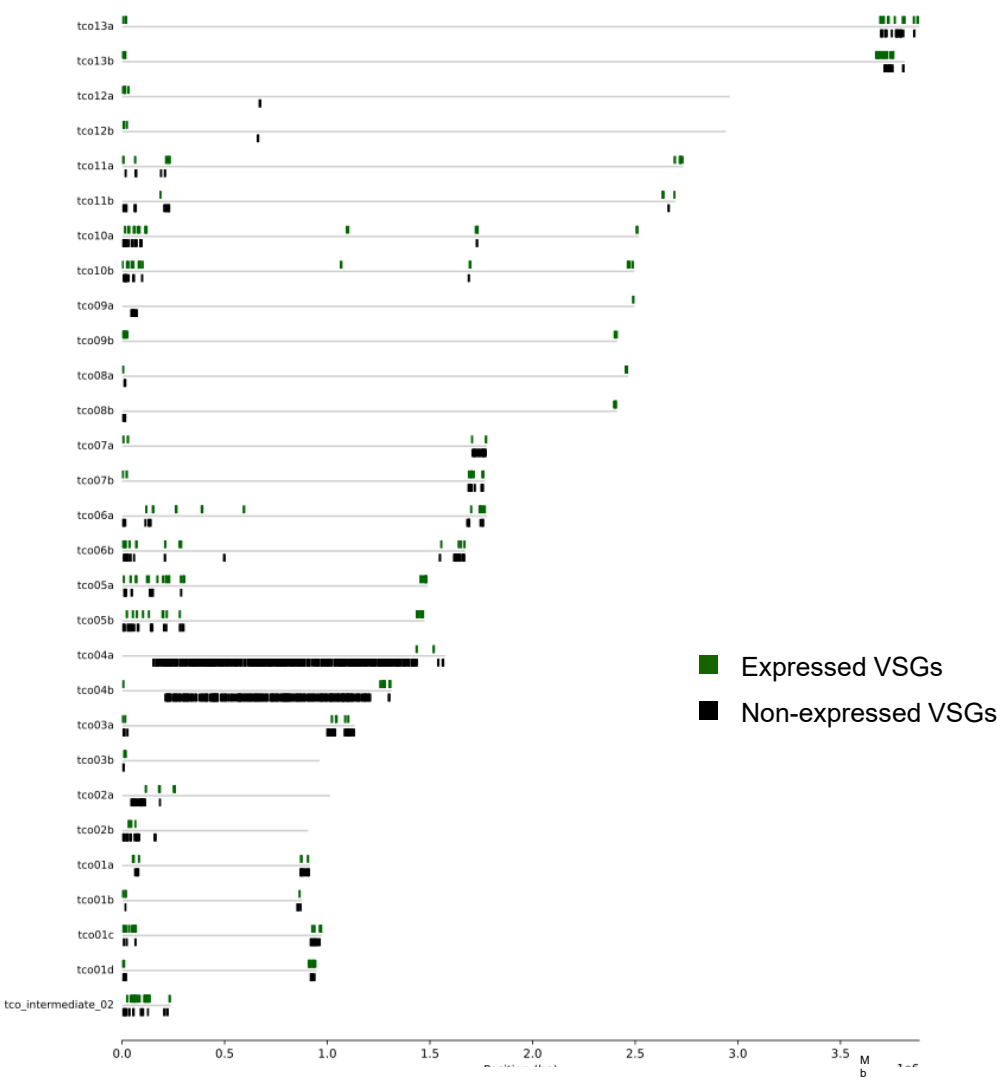


Fig.S10



Fig.S11

A



B

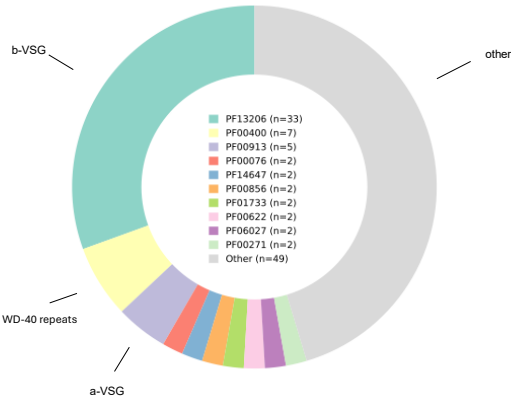
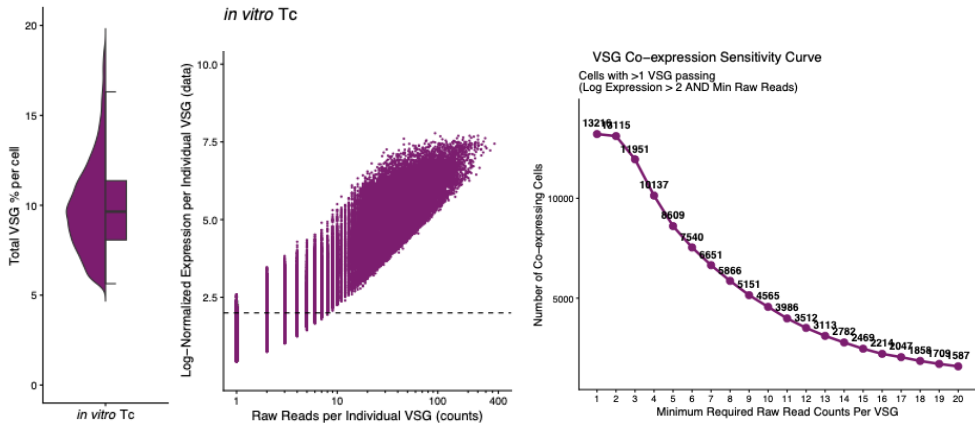


Fig.S12

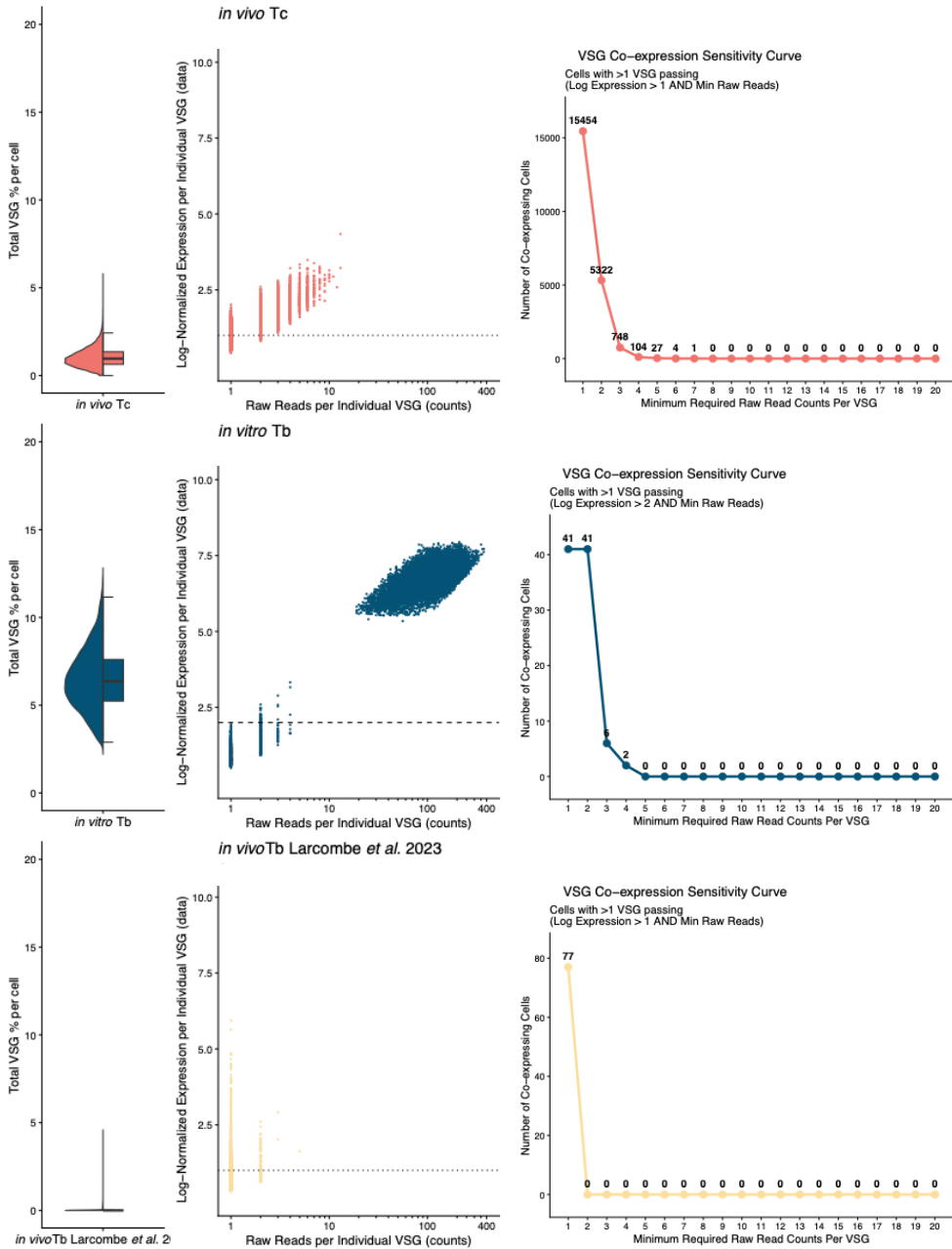
A

T. congolense



B

T. brucei



C

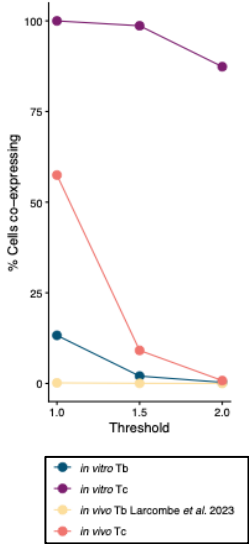
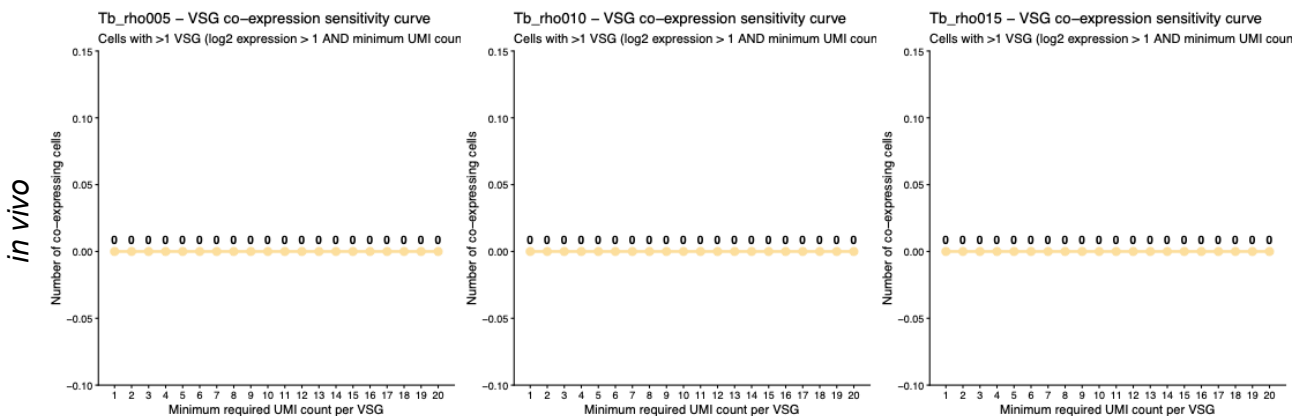
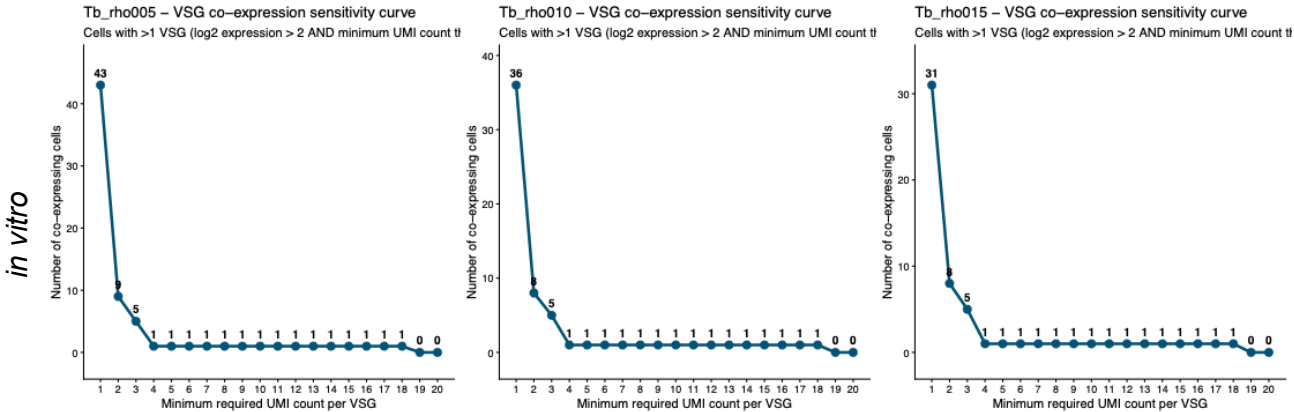
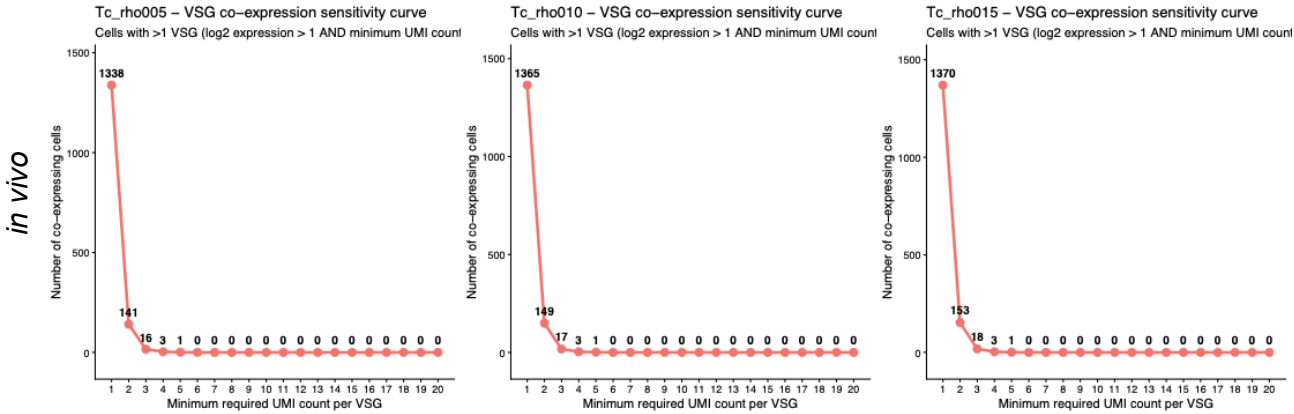
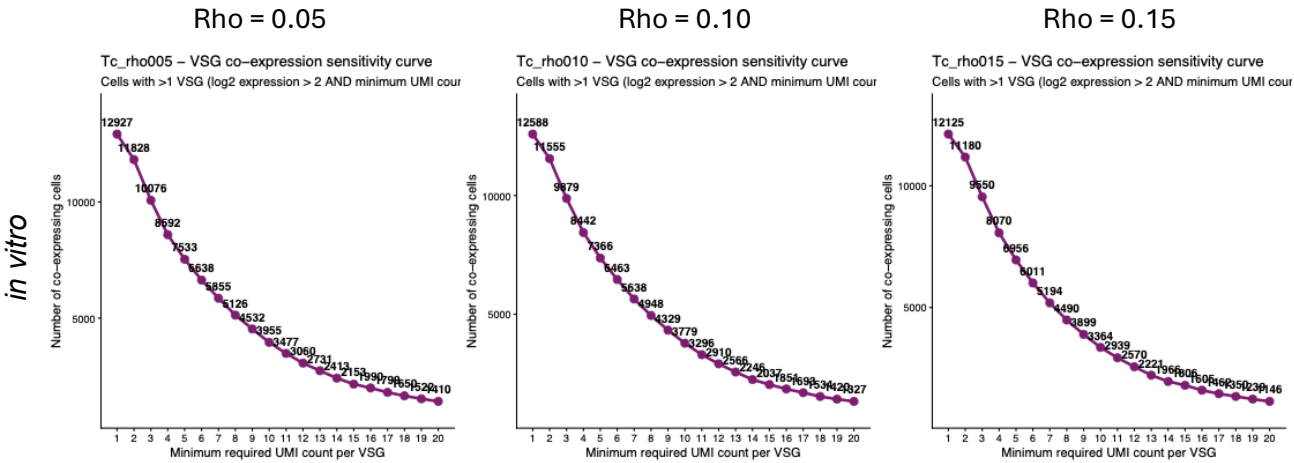


Fig.S13

T. congolense



T. brucei