1 # Genome-scale functional profiling of cell cycle controls in African trypanosomes

2
3
4

5 Catarina A. Marques[1,3,§], Michele Tinti[1,§], Andrew Cassidy[2] and David Horn[1,*]
6
7
8
9
10 [1] Wellcome Trust Centre for Anti-Infectives Research, School of Life Sciences, University of
11 Dundee, United Kingdom
12
13 [2] Tayside Centre for Genomic Analysis, Ninewells Hospital and School of Medicine,
14 University of Dundee, United Kingdom
15
16 [3] Current address: Wellcome Trust Centre for Integrative Parasitology, University of
17 Glasgow, United Kingdom
18
19
20
21 [§] joint first authors
22
23
24
25 * Correspondence: d.horn@dundee.ac.uk
26
27
28

29 Keywords: checkpoint, kinetoplastid, RIT-seq, sleeping sickness, *Leishmania, Trypanosoma*
30 *brucei, Trypanosoma cruzi.*
31

Abstract

Trypanosomatids, which include major pathogens of humans and livestock, are divergent eukaryotes for which cell cycle controls and the underlying mechanisms are not completely understood. Here, we describe a genome-wide RNA-interference library screen for cell cycle regulators in bloodstream form *Trypanosoma brucei*. We induced massive parallel knockdown and sorted the perturbed population into cell cycle stages using flow cytometry. RNAi-targets were deep-sequenced from each stage and cell cycle profiles were digitally reconstructed at a genomic scale. We identify hundreds of proteins that impact cell cycle progression; glycolytic enzymes required for G_1S progression, DNA replication factors, mitosis regulators, proteasome and kinetochore complex components required for G_2M progression, flagellar and cytoskeletal components required for cytokinesis, mRNA-binding factors, protein kinases and many previously uncharacterised proteins. The outputs facilitate functional annotation and drug-target prioritisation and provide comprehensive functional genomic evidence for the machineries, pathways and regulators that coordinate progression through the trypanosome cell cycle.

The data can be searched and browsed using an interactive, open access, online data visualization tool (https://tryp-cycle.onrender.com).

Introduction

51

52   The canonical eukaryotic cell cycle encompasses discrete phases: G_1 (gap 1), when

53   the cell prepares for DNA replication; S (synthesis) phase, when nuclear DNA replication

54   takes place; G_2 (gap 2), when the cell prepares for mitosis; and M (mitosis) when the

55   replicated DNA is segregated and the nucleus divides (1). Mitosis is immediately followed by

56   cytokinesis (cell division), generating two daughter cells (2). Anomalies occurring during cell

57   cycle progression can result in cell cycle arrest, to allow the cell to resolve the anomaly; in

58   cell death, if the anomaly cannot be resolved or, among other outcomes, carcinogenesis.

59   Therefore, progression through the cell cycle is typically under strict checkpoint control; the

60   G_1-S, intra S phase, G_2-M and spindle checkpoints control the onset of S phase, S phase

61   progression, the onset of M phase and M phase progression, respectively (3). These

62   processes have been extensively studied, particularly because cell cycle defects are

63   common triggers for carcinogenesis (4). However, our understanding of the evolution and

64   mechanisms of eukaryotic cell cycle progression control derives primarily from studies on the

65   opisthokonts (including animals and fungi), with relatively fewer studies on more divergent

66   eukaryotes, such as the trypanosomatids (1).

67   The trypanosomatids are flagellated protozoa and include parasites that cause a

68   range of neglected tropical diseases that have major impacts on human and animal health.

69   The African trypanosome, *Trypanosoma brucei*, is transmitted by tsetse flies and causes

70   both human and animal diseases, sleeping sickness and nagana, respectively, across sub-

71   Saharan Africa (5). *T. brucei* has emerged as a highly tractable experimental system, both

72   as a parasite and as a model organism (6). For example, the *T. brucei* flagellum (7) serves

73   as a model for studies on human ciliopathies (8-11). Divergent features, shared with other

74   pathogenic trypanosomatids, such as *Trypanosoma cruzi* and *Leishmania*, include glycolysis

75   compartmentalised within glycosomes (12), a single mitochondrion with a complex

76   mitochondrial DNA structure known as the kinetoplast (13) and polycistronic transcription of

77   almost every gene (14). Widespread, and constitutive, polycistronic transcription in

78   trypanosomatids places major emphasis on post-transcriptional controls by mRNA binding

79   proteins (RBPs) and post-translational controls, involving protein phosphorylation in

80   particular.

81   Studies that focus on cell cycle controls in *T. brucei* have revealed features that are

82   conserved with other well-studied eukaryotes, but also features that are divergent (13,15).

83   Most notably, the available evidence suggest that certain cell cycle checkpoints are absent.

84   For example, cytokinesis is not dependent upon either mitosis or nuclear DNA synthesis in

85   the insect stage of *T. brucei* (16). Moreover, functions previously thought to be fulfilled by

86   highly conserved proteins employ lineage-specific or highly divergent proteins in

87   trypanosomatids. The kinetochore complex, which directs chromosome segregation, is

3

88   trypanosomatid-specific, for example (17), while the origin recognition complex (ORC),

89   involved in DNA replication initiation, is highly divergent (18). In terms of high-throughput

90   studies, transcriptome (19) and proteome (20) monitoring during the *T. brucei* cell cycle

91   revealed hundreds of regulated mRNAs and proteins, while phosphoproteomic analysis

92   revealed dynamic phosphorylation of several RBPs (21). Nevertheless, divergence presents

93   a substantial challenge, many *T. brucei* genes have not yet been assigned a specific

94   function, and many cell cycle regulators likely remain to be identified.

95   High-throughput functional genetic screens can be used to simultaneously assess

96   every gene in a genome for a role in a particular process. We developed RNA interference

97   Target Sequencing (RIT-seq) for *T. brucei* (22) and previously generated genome-scale

98   fitness profiles, facilitating essentiality predictions and the prioritisation of potential drug

99   targets (23). Here we describe a genome-scale RIT-seq screen to identify cell cycle controls

100   and regulators in bloodstream form African trypanosomes. Following induction of

101   knockdown, the cells were sorted according to their DNA content using fluorescence-

102   activated cell sorting (FACS). The sorted populations were the G_1, S and G_2M cell cycle

103   stages as well as perturbed cell populations with either less DNA than typically found in G_1

104   or more DNA than typically found in G_2M. RIT-seq analysis was then carried out for each

105   sorted population and cell cycle profiles were digitally reconstructed for each gene using

106   sequencing read-counts. This genome-wide screen reveals the protein complexes, pathways

107   and signalling factors that coordinate progressive steps through the trypanosome cell cycle,

108   both improving our understanding of trypanosome cell biology and also further facilitating the

109   prioritisation of new potential drug targets.

110 **Results**

111 **A genome-wide conditional knockdown screen for cell cycle progression defects**

112   Bloodstream form *T. brucei* are readily grown in cell culture, with exponential

113 proliferation and a doubling time of approximately 6.5 h. The *T. brucei* genome is diploid

114 such that G_1 cells have a 2C genome content; C represents the haploid DNA content. Cells

115 progressing through nuclear S phase, and replicating their DNA, have a genome content

116 between 2C and 4C, while cells that have completed DNA replication (G_2M) have a 4C DNA

117 content (*Figure 1A*). Cytokinesis then produces two daughter cells with a 2C DNA content.

118 Some perturbations can yield defects which become apparent when cytokinesis produces

119 sub-2C cells or over-replicated, polyploid (>4C) cells (*Figure 1A*). These polyploid cells arise

120 due to endoreduplication, additional rounds of DNA replication without cytokinesis, either

121 with (24) or without (25,26) mitosis, yielding cells with multiple diploid nuclei or cells with

122 polyploid nuclei, respectively.

123   We devised a high-throughput RNA interference (RNAi) Target sequencing (RIT-seq)

124 screen to identify cell cycle controls and regulators at a genomic scale. Key features of RIT-

125 seq screening include: first, use of a high-complexity *T. brucei* RNAi library comprising, in

126 this case, approximately one million clones; second, massive parallel tetracycline-inducible

127 expression of cognate dsRNA; and third, deep sequencing, mapping and counting of

128 mapped reads derived from RNAi target fragments (22). Each clone in the library has one of

129 approximately 100,000 different RNAi target fragments (250-1500 bp) between head-to-head

130 inducible T7-phage promoters, with each cassette inserted at a chromosomal locus that

131 supports robust expression. Inducibly expressed long dsRNA is then processed to siRNA by

132 the native RNAi machinery (27). Complexity and depth of genome coverage in the library are

133 critical, in that similar phenotypes produced by multiple RNAi target fragments against a

134 single gene provide cross-validation. Improvements in reference genome annotation (28),

135 next generation sequencing technology and sequence data analysis tools (see Materials and

136 Methods) have also greatly facilitated quantitative phenotypic analysis using short-read

137 sequence data.

138   Briefly, we induced massive parallel knockdown in an asynchronous *T. brucei*

139 bloodstream form RNAi library for 24 h, fixed the cells, stained their DNA with propidium

140 iodide (PI) and then used high-speed fluorescence-activated cell sorting (FACS) to divide the

141 perturbed cell population into; sub-diploid (<2C), G_1 (2C), S (between 2C and 4C), G_2M (4C)

142 and over-replicated (>4C) pools (*Figure 1—figure supplement 1*). Fixation and staining

143 with the fluorescent DNA intercalating dye were pre-optimised for high-speed sorting (see

144 Materials and Methods). Approximately 10 million cells were collected for each of the G_1, S

145 and G_2M pools and samples from these pools were checked post-sorting to assess their

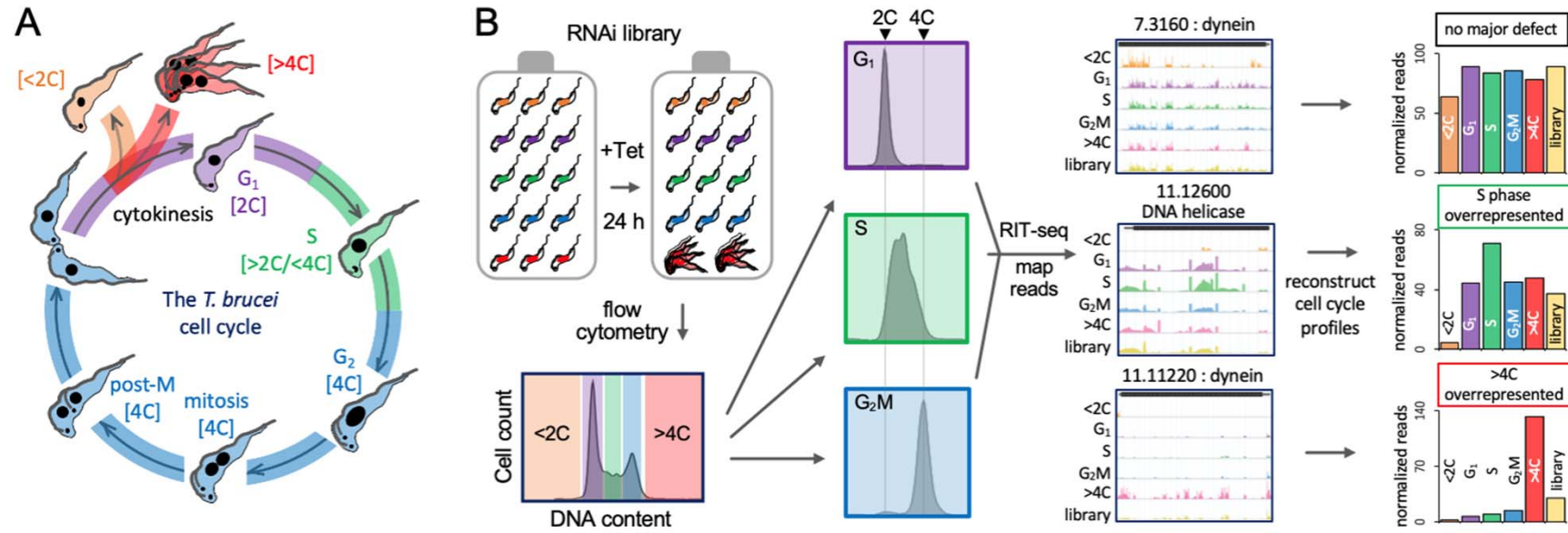146 purity (*Figure 1B, Figure 1—figure supplement 1*). For the perturbed and less abundant

Figure 1

Figure 1. A genome-wide conditional knockdown screen for cell cycle progression defects. (**A**) Schematic representation of the bloodstream form *T. brucei* cell cycle, also showing aberrant sub-2C and >4C phenotypes. (**B**) The schematic illustrates the RIT-seq screen; massive parallel induction of RNAi followed by flow cytometry and RIT-seq, allowing for reconstruction of cell cycle profiles, using mapped reads from each knockdown. Each read-mapping profile encompasses the gene of interest and associated untranslated regions present in the cognate mRNA. The library data represents the uninduced and unsorted population. GeneIDs, Tb927.7.3160 for example, are shown without the common 'Tb927.' component.

147   <2C and >4C pools, less than one million cells were collected; these pools were retained in

148   their entirety for RIT-seq analysis.

149   RIT-seq was carried out for both the uninduced and induced, unsorted library

150   controls, and for each of the five induced and sorted pools of cells as described in the

151   Materials and Methods section. Briefly, we extracted genomic DNA from each sample,

152   amplified DNA fragments containing each RNAi target fragment in PCR reactions (*Figure*

153   *1—figure supplement 1*) and used the amplified products to generate Illumina sequencing

154   libraries. Analysis of sequencing reads mapped to the reference genome yielded counts for

155   both total reads as well as reads containing the barcode (GTGAGGCCTCGCGA) that flanks

156   each RNAi target fragment; the presence of the barcode confirmed that reads were derived

157   from a specific RNAi target fragment and not from elsewhere in the genome. We derived

158   counts of reads mapped to each of >7,200 non-redundant gene sequences in the uninduced

159   and induced, unsorted library controls and in each of the five sorted samples. We selected

160   the 24 h timepoint, equivalent to approximately 3.5 population doubling times, to allow

161   sufficient time for the development of robust inducible phenotypes, but also to capture

162   perturbed populations before they were critically diminished. Indeed, reads for 23.4% of

163   genes dropped by >3-fold following 72 h of knockdown in a prior RIT-seq study (23), while a

164   comparison of the unsorted control samples indicated that reads for only 0.6% of genes

165   dropped by >3-fold following 24 h of knockdown here (see *Figure 1—figure supplement 2*).

166   Each sorted-sample library yielded between 22.6 and 37 million mapped read-pairs; <2C =

167   37 M, G_1 = 35.1 M, S = 30.3 M, G_2M = 22.6 M, >4C = 24.9 M; yielding >35,000 RNAi data-

168   points (*Supplementary File 1*).

169   The RIT-seq digital data for individual genes following knockdown provided a

170   measure of abundance in each pool and were, therefore, used to digitally reconstruct cell

171   cycle profiles for individual gene knockdowns (*Figure 1B*). We expected to observe

172   accumulation of particular knockdowns in specific cell cycle phase pools, thereby reflecting

173   specific defects. This was indeed the case, and some examples are shown to illustrate; no

174   major defect, S phase overrepresented or >4C overrepresented, following knockdown

175   (*Figure 1B*). These outputs suggest that loss of the cytoplasmic dynein heavy chain

176   (Tb927.7.3160) does not perturb cell cycle distribution; that a putative DNA helicase

177   (Tb927.11.12600) is required for the completion of S phase; and that knockdown of the

178   axonemal dynein heavy chain (Tb927.11.11220) results in endoreduplication in the absence

179   of cytokinesis. Dyneins are cytoskeletal motor proteins that move along microtubules, to

180   produce a flagellar beat, for example (29).

181

182   **Validation and identification of >1,000 candidates linked to cell cycle defects**

183   The *T. brucei* core genome comprises a non-redundant set of over 7,200 protein-

184   coding sequences, for which we were now able to digitally reconstruct cell cycle profiles

185   following knockdown; the full dataset is shown mapped to the eleven megabase

186   chromosomes in *Figure 2—figure supplement 1*. The data can also be searched and

187   browsed using an interactive, open access, online data visualization tool (see *Figure 2—*

188   *figure supplement 2*; https://tryp-cycle.onrender.com). First, we examined knockdowns

189   reporting an overrepresentation of >4C cells, indicating endoreduplication, and this yielded

190   201 genes for which reads in the >4C pool exceeded 1.5-fold the sum of reads in the G_1, S

191   phase and G_2M pools (*Figure 2A*, left-hand panel; *Supplementary File 1*). The >4C

192   phenotype was previously observed following α-tubulin knockdown in a landmark study that

193   first described RNAi in *T. brucei* (24) and, indeed, we observed pronounced

194   overrepresentation of >4C cells for both adjacent α-tubulin (Tb927.1.2340) and β-tubulin

195   (Tb927.1.2350) gene knockdowns (*Figure 2A*, middle and right-hand panel). We then

196   examined knockdowns reporting an overrepresentation of <2C cells, indicating a reduced

197   DNA content. This yielded 119 genes for which reads in the <2C pool exceeded 1.5-fold the

198   sum of reads in the G_1, S phase and G_2M pools (*Figure 2B*, left-hand panel;

199   *Supplementary File 1*). Haploid cells were previously observed following DOT1A

200   knockdown (30) and, consistent with the previous report, we observed pronounced

201   overrepresentation of <2C cells for the DOT1A (Tb927.8.1920) gene knockdown (*Figure 2A*,

202   middle and right-hand panel); we are not aware of other knockdowns reported to yield a

203   similar phenotype. Together, these results provide initial validation for the >4C and <2C

204   components of the screen.

205   Next, we turned our attention to knockdowns reporting an overrepresentation of G_1, S

206   phase or G_2M cells. The pools of knockdowns that registered >25% overrepresented read

207   counts in each of these categories are highlighted in *Figure 2C* (left-hand panels and

208   *Supplementary File 1*) and data for an example from each category are shown (*Figure 2C*;

209   right-hand panels); the glycolytic enzyme, aldolase (Tb927.10.5620), reported 104%

210   increase in G_1 cells (further details below); the proliferating cell nuclear antigen (PCNA;

211   Tb927.9.5190), a DNA sliding clamp that is a central component of the replication machinery

212   (31), reported 25% increase in S phase cells, consistent with prior analysis (32); and

213   PrimPol-like 2 (PPL2; Tb927.10.2520), a post-replication translesion polymerase, reported

214   65% increase in G_2M cells, also consistent with prior analysis (33). These results provided

215   initial validation for the G_1, S phase and G_2M components of the screen.

216   Overall, the five components of the screen yielded 1,158 genes that registered a cell

217   cycle defect, based on the thresholds applied above. This is 16.1% of the 7,204 genes

218   analysed, and the distribution of these genes among the five arms of the screen are shown
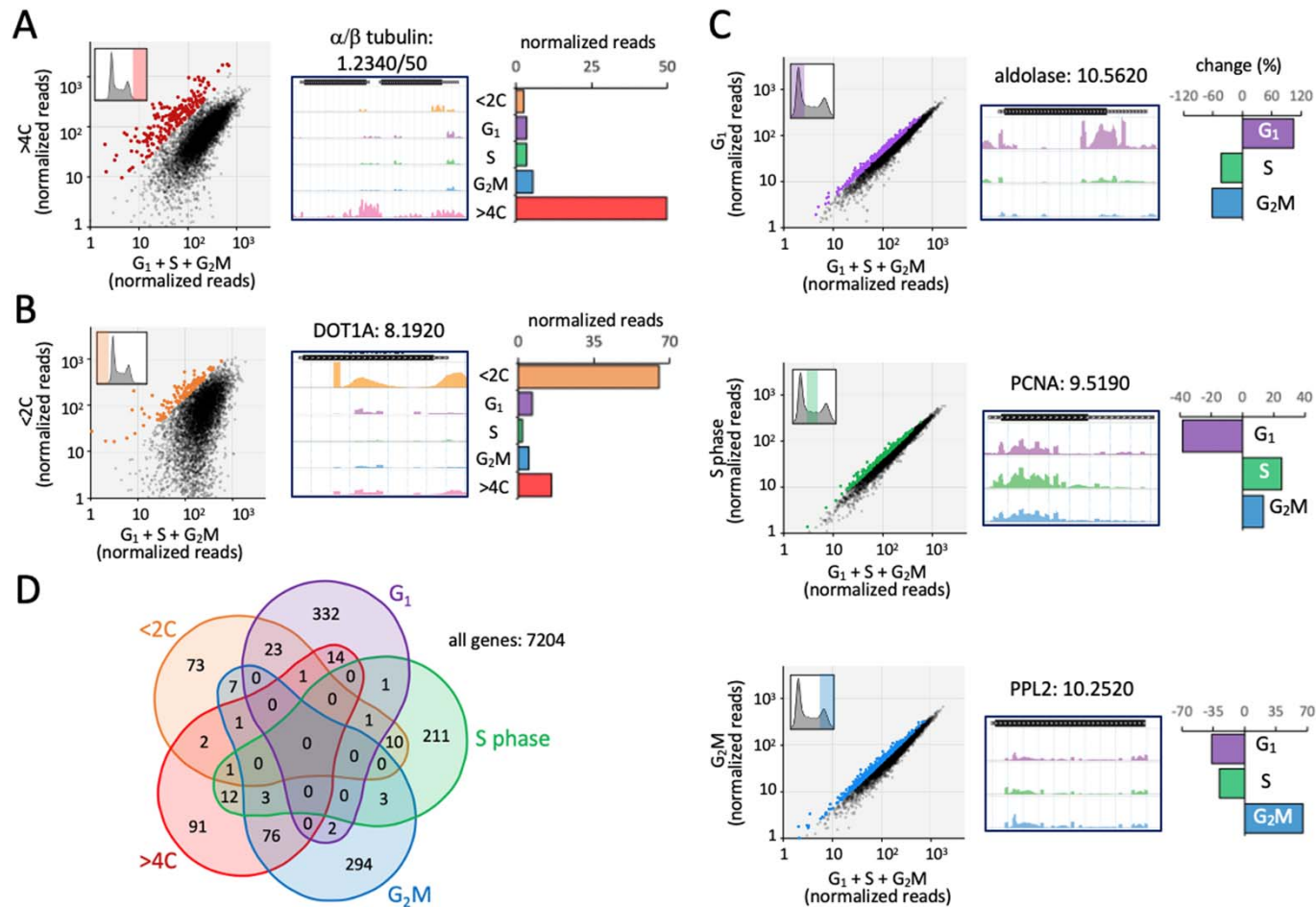
Figure 2. Validation and identification of >1000 candidates linked to cell cycle defects. (**A**) The plot on the left shows knockdowns overrepresented in the >4C experiment in red; those with >1.5-fold the sum of reads in the G_1, S phase and G_2M samples combined. The read-mapping profile and read-counts for α/β-tubulin are shown to the right. (**B**) The plot on the left shows knockdowns overrepresented in the sub-2C experiment in orange; those with >1.5-fold the sum of reads in the G_1, S phase and G_2M samples combined. The read-mapping profile and read-counts for DOT1A are shown to the right. (**C**) The plots on the left shows knockdowns overrepresented in the G_1, S phase and G_2M experiments in purple, green and blue, respectively; those that were >25% overrepresented in each category. Read-mapping profiles and relative read-counts for example hits are shown to the right. PCNA, proliferating cell nuclear antigen; PPL2, PrimPol-like 2. (**D**) The Venn diagram shows the distribution of knockdowns overrepresented in each arm of the screen.

219   in the Venn diagram in *Figure 2D*. Since we predicted that knockdowns associated with a

220   cell cycle defect were more likely to also register a growth defect, we compared these

221   datasets to prior RIT-seq fitness profiling data (23). All groups of genes that registered cell

222   cycle defects, except for the <2C set, were significantly enriched for genes that previously

223   registered a loss-of-fitness phenotype following knockdown in bloodstream form cells (χ^2

224   test; <2C, $p = 0.15$; G_1, $p = 0.015$; S, $p = 4.7^{-4}$; G_2M, $p = 3.5^{-24}$; >4C, $p = 4.4^{-199}$). This is

225   consistent with loss-of-fitness as a common outcome following a cell cycle progression

226   defect. Taken together, the analyses above provided validation for the RIT-seq based cell

227   cycle phenotyping approach and yielded >1,000 candidate genes that impact specific steps

228   during *T. brucei* cell cycle progression.

229

230   **Cytokinesis defects associated with endoreduplication**

231   In bloodstream form *T. brucei*, defective >4C cells can arise due to endoreduplication

232   without cytokinesis, either with (24) or without (25,26) mitosis. Cytokinesis-only defects were

233   previously observed following knockdown of α-tubulin (24) or flagellar proteins (7,34);

234   flagellar beat is thought to be required for cytokinesis in bloodstream form *T. brucei*.

235   Consistent with these observations, α-tubulin (see *Figure 2A*) and axonemal dynein heavy

236   chain (see *Figure 1B*) knockdowns were amongst 201 genes overrepresented in the >4C

237   pool in our screen. Gene Ontology (GO) annotations provide structured descriptions of gene

238   products in terms of functions, processes, and compartments. Analysis of overrepresented

239   GO annotations within the >4C-enriched cohort revealed 'dynein', 'intraflagellar transport'

240   (IFT), 'axoneme' and 'cytoskeleton' terms, and also 'chaperonin T-complex', 'cytokinesis'

241   and multiple other 'cell cycle' terms (*Figure 3A*). The violin plot in *Figure 3B* shows specific

242   enrichment of IFT and dynein knockdowns in the >4C pool, relative to other cohorts of

243   knockdowns. Exocyst components, primarily involved in exocytosis (35), were included as a

244   negative control cohort; indeed, none of the exocyst components register enrichment in the

245   >4C pool, nor in any other pool (see below). Enrichment of individual chaperonin T-complex

246   components, dyneins, and IFT factors in the >4C pool is illustrated in *Figure 3C*. The

247   chaperonin T-complex is involved in tubulin and actin folding (36) and, notably, actin

248   knockdown was also associated with >4C enrichment (*Figure 3—figure supplement 1*).

249   The heat-map in *Figure 3D* shows the data for all five sorted pools for the cohorts

250   described above and for additional cohorts of knockdowns enriched in the >4C pool; these

251   include radial spoke proteins, extra-axonemal paraflagellar rod (PFR) proteins, as well as

252   nucleoporins. The gallery in *Figure 3E* shows examples of RIT-seq read-mapping profiles

253   for twenty-four individual genes that register >4C enrichment following knockdown. In

254   addition to the categories above, these include the inner arm dynein 5-1 (37), FAZ proteins
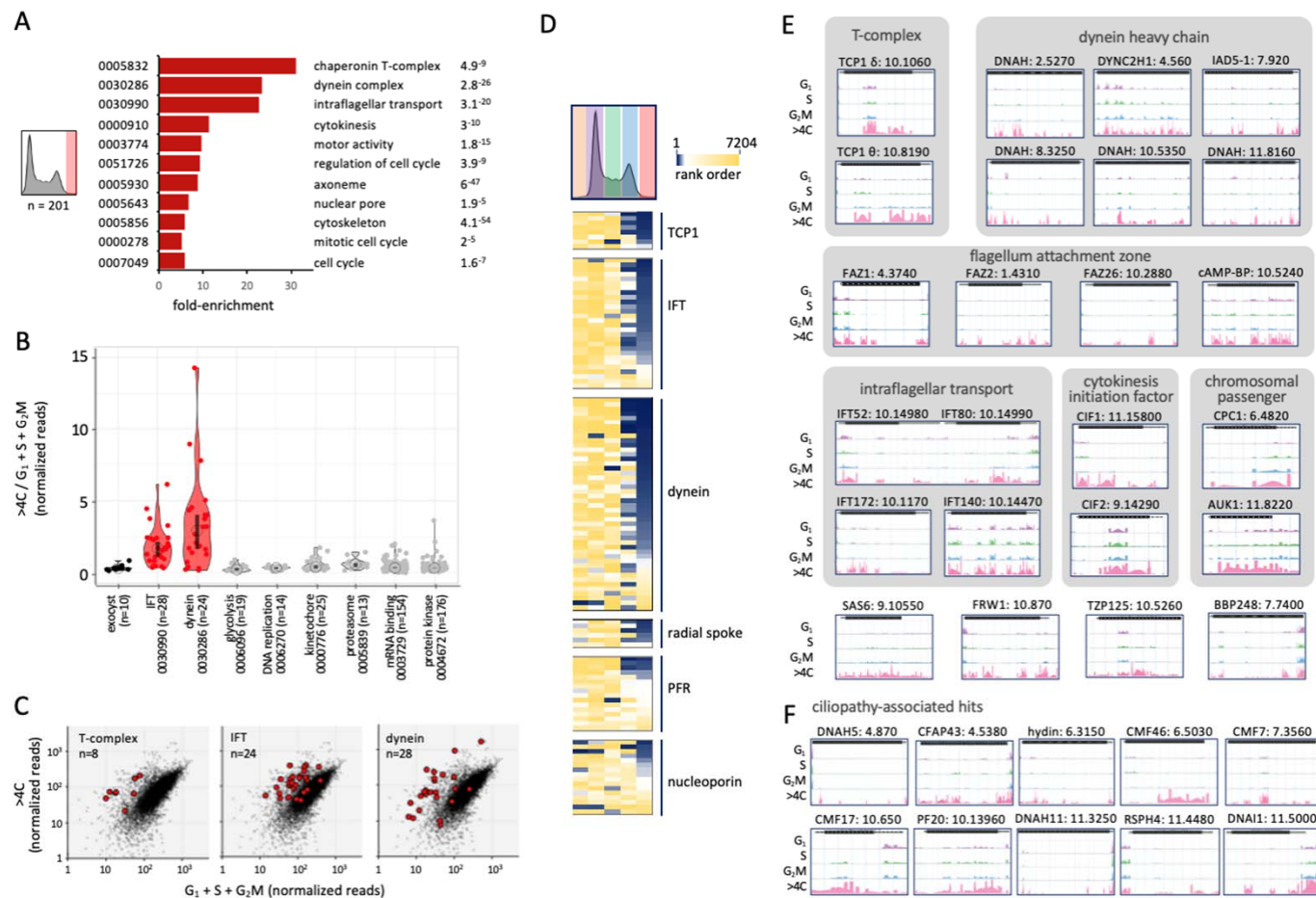
Figure 3

Figure 3. Cytokinesis defects associated with endoreduplication. (**A**) The bar-graph shows enriched Gene Ontology terms in the >4C overrepresented dataset. (**B**) The violin plot shows relative >4C read-counts for cohorts of genes and reflects data distribution. Open circles indicate median values and the vertical bars indicate 95% confidence intervals. Significantly overrepresented cohorts are indicated in red. (**C**) The plots show overrepresentation of T-complex, dynein and intraflagellar transport (IFT) factors in red in the >4C experiment. (**D**) The heatmaps show relative representation in all five sorted pools for the above and additional cohorts of knockdowns; blue, most overrepresented. (**E**) Example read-mapping profiles for hits overrepresented in the >4C pool. (**F**) Example read-mapping profiles for ciliopathy-associated hits overrepresented in the >4C pool. CMF, Component of Motile Flagella; CFAP, Cilia and Flagella Associated Protein.

255   which mediate attachment of the flagellum to the cell body (38); cytokinesis initiation factors

256   CIF1/TOEFAZ1 (Tb927.11.15800) and CIF2 (Tb927.9.14290) (39) and chromosomal

257   passenger complex components, including CPC1 (Tb927.6.4820) and the aurora B kinase,

258   AUK1 (Tb927.11.8220). AUK1 and CPC1 are spindle-associated and regulate mitosis and

259   cytokinesis (26,40). Notably, endoreduplication was reported previously following AUK1

260   knockdown in bloodstream form *T. brucei* (41) and this is the kinase with the most

261   pronounced overrepresentation in our >4C dataset. The next >4C overrepresented kinase is

262   the CMGC/RCK (Tb927.3.690), knockdown of which previously yielded a dramatic

263   cytokinesis defect (42).

264   Additional examples of genes registering >4C overrepresentation include the

265   centriole cartwheel protein SAS6 (Tb927.9.10550) (43), the cleavage furrow-localizing

266   protein FRW1 (Tb927.10.870) (44), the basal body – axoneme transition zone protein

267   TZP125 (Tb927.10.5260) (45) and the basal body protein BBP248 (Tb927.7.7400) (46). One

268   hundred additional examples are shown in *Figure 3—figure supplement 1*, including

269   intermediate and light chain dyneins; other flagellum-associated factors, radial spoke

270   proteins, components of motile flagella, flagellum attachment and transition zone proteins,

271   intraflagellar transport proteins, kinesins (47,48), nucleoporins (49), and many previously

272   uncharacterised hypothetical proteins. Some other notable examples include the

273   microtubule-severing katanin KAT80 (Tb927.9.9960) (50), the dynein regulatory factor

274   trypanin (Tb927.10.6350) (51), the AIR9 microtubule associated protein (Tb927.11.17000)

275   (52), CAP51V (Tb927.7.2650) (53) and importin, IMP1 (Tb927.9.13520) (54).

276   Orthologues of several *T. brucei* flagellar proteins have previously been linked to

277   debilitating human ciliopathies, such that the trypanosome flagellum is used as a model for

278   studies on these defects. Defects in intraflagellar dynein transport are associated with

279   respiratory infections, for example (9). Some examples of ciliopathy-associated orthologues

280   which register overrepresentation in the >4C pool are shown in *Figure 3F* and *Figure 3—*

281   *figure supplement 1*. These include proteins linked to primary ciliary dyskinesia (DNAI1,

282   Tb927.11.5000; DNAH5, Tb927.4.870; DNAH11, Tb927.11.3250; RSPH4, Tb927.11.4480)

283   (11); male infertility (CFAP43, Tb927.4.5380; CMF7/TbCFAP44, Tb927.7.3560) (10); and

284   cone-rod dystrophies, as well as other ocular defects (CMF17, Tb927.10.650; CMF39,

285   Tb927.4.5370; CMF46, Tb927.6.5030) (8).

286   From analysis of knockdowns overrepresented in the >4C pool, we conclude that

287   RIT-seq screening provided comprehensive genome-scale identification of cytokinesis

288   defects in bloodstream form *T. brucei*. Endoreduplication appears to be a common outcome

289   following a cytokinesis defect. Amongst hundreds of genes required for progression through

290   cytokinesis, flagellar proteins featured prominently, including the majority of dynein chains

291   and intraflagellar transport factors. Many of these factors are essential for viability and
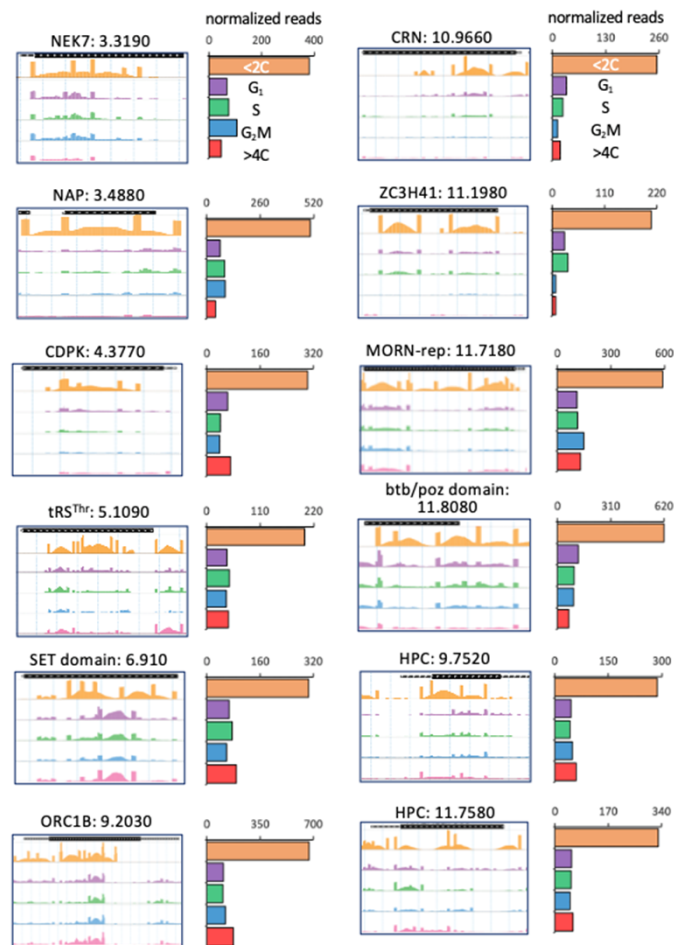
9

Figure 4

Figure 4. Defects producing sub-diploid cells. Read-mapping profiles and read-counts for example hits overrepresented in the <2C experiment. HPC, Hypothetical Protein, Conserved.

292   include potential druggable targets in trypanosomatids, as well as orthologues of proteins

293   associated with ciliopathies.

294

Defects producing sub-diploid cells

296   A DNA replication or mitosis defect followed by cytokinesis may result in generation

297   of cells that retain nuclear DNA with a sub-2C DNA content. We emphasise retention of

298   nuclear DNA here because *T. brucei* cells lacking nuclear DNA, referred to as zoids, have

299   been reported previously as a result of asymmetrical cell division. Zoids are typically

300   observed when DNA replication or mitosis are perturbed in insect stage cells (16,25,55). The

301   zoid phenotype is typically either absent or less abundant in the developmentally distinct

302   bloodstream form cells (56) that we analysed here. Nevertheless, any zoids present in the

303   <2C pool will not have been detected using RIT-seq, since detection relies upon the

304   presence of a nuclear RNAi target fragment.

305   One hundred and nineteen knockdowns were overrepresented in the <2C RIT-seq

306   screening dataset. A particularly prominent hit was the previously identified histone

307   methyltransferase, DOT1A (see above). DOT1A is responsible for dimethylation of histone

308   H3K76, and DOT1A knockdown results in premature mitosis without DNA replication,

309   generating cells with a haploid DNA content (30). The gallery in *Figure 4* shows examples of

310   RIT-seq read-mapping profiles for twelve additional genes that register <2C enrichment

311   following knockdown. These include the DNA replication origin recognition complex-

312   associated protein, ORC1B (Tb927.9.2030) (57), two putative protein kinases (NEK7,

313   Tb927.3.3190; CDPK, Tb927.4.3770), a putative mRNA-binding helicase (ZC3H41,

314   Tb927.11.1980), a nucleosome assembly protein (NAP, Tb927.3.4880) and threonyl-tRNA

315   synthetase (Tb927.5.1090). Notably, threonine metabolism impacts histone methylation in

316   mammalian cells (58), while a protein containing a putative methyllysine-binding SET

317   domain (Tb927.6.910) also registers enrichment in the <2C pool. These hits present new

318   candidate regulators of DNA replication, mitosis or meiosis (59), and further potential links to

319   post-translational protein methylation as a key player in coordinating these processes.

320

A profile of G_1, S phase and G_2M defects

322   We next analysed knockdowns overrepresented in the G_1, S phase or G_2M pools.

323   Several hundred knockdowns registered >25% overrepresented read counts in each of

324   these categories, as shown in a RadViz plot (*Figure 5A*). GO annotations within each cohort

325   revealed a number of enriched terms (*Figure 5B*). Overrepresented knockdowns were

326   associated with glycolysis and mRNA binding in the G_1 pool, with DNA replication in the S

327   phase pool and with a similar profile to that seen for the >4C set in the G_2M pool.
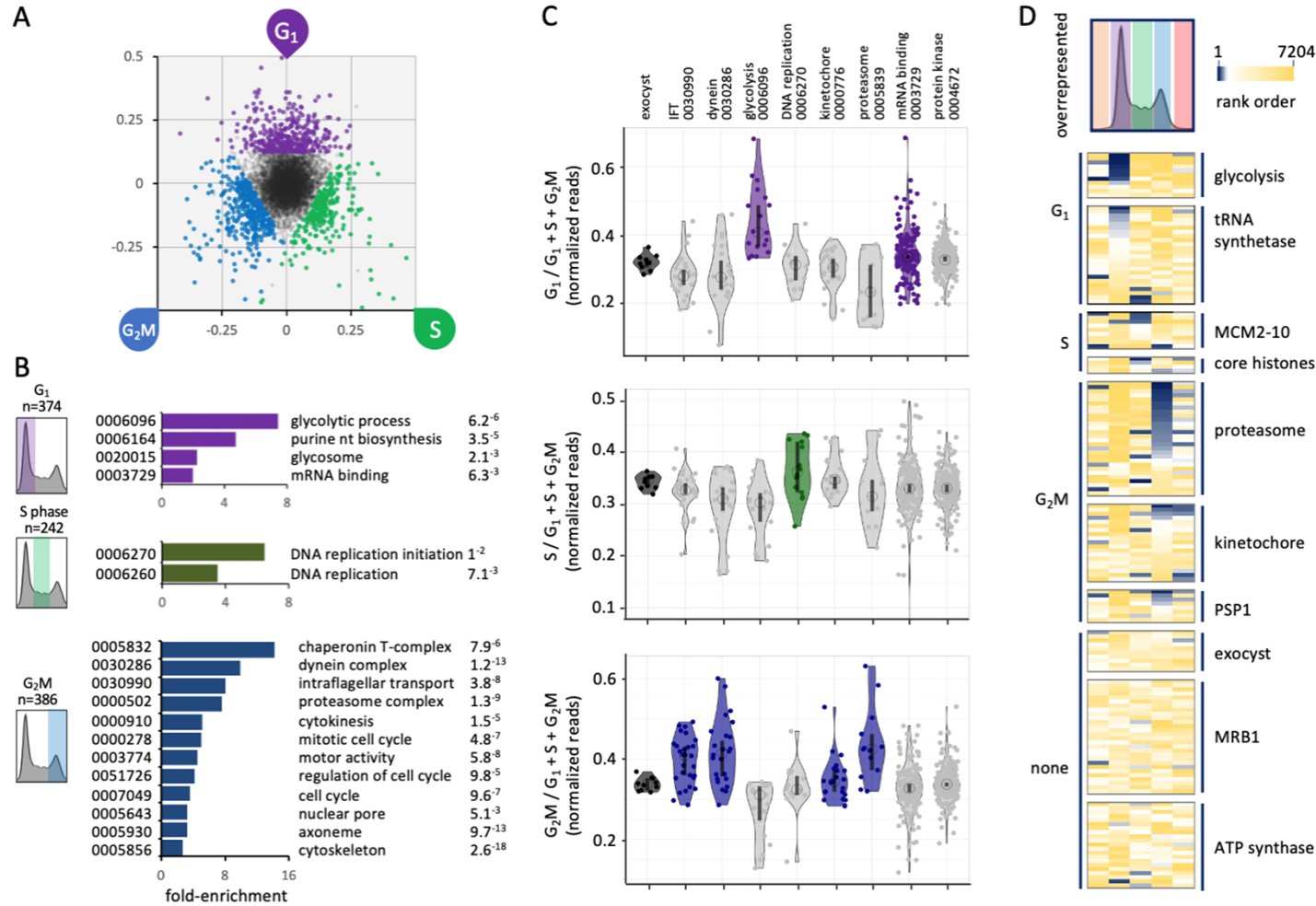
Figure 5. A profile of G_1, S phase and G_2/mitosis defects. (**A**) The RadViz plot shows knockdowns that registered >25% overrepresented read-counts in the G_1, S phase or G_2M categories. (**B**) The bar-graphs shows enriched Gene Ontology terms in the G_1, S phase or G_2M overrepresented datasets. (**C**) The violin plots show relative G_1, S phase or G_2M read-counts for cohorts of genes and reflects data distribution. Open circles indicate median values and the vertical bars indicate 95% confidence intervals. Overrepresented cohorts are indicated in purple, green and blue, respectively. (**D**) The heatmaps show relative representation in all five sorted pools for the above and additional cohorts of knockdowns; blue, most overrepresented.

328   The violin plot in **Figure 5C** shows specific enrichment of individual knockdowns for

329   glycolytic enzymes and a subset of mRNA binding protein in the G_1 pool, for DNA replication

330   factors in the S phase pool, and proteasome components and a subset of kinetochore

331   component in the G_2M pool (**Figure 5C**, lower panel). Overlap between knockdowns that

332   accumulate in both the G_2M and >4C pools likely reflects mitosis and cytokinesis defects

333   both before and after endoreduplication; perhaps 24 h of knockdown is insufficient for

334   endoreduplication in all perturbed cells or not all mitosis or cytokinesis-perturbed phenotypes

335   result in endoreduplication; compare >4C and G_2M data for IFT factors and dyneins in

336   **Figure 3B** and **Figure 5C**, for example. Once again, exocyst components provided a control

337   cohort with no components registering enrichment in the G_1, S phase or G_2M pools following

338   knockdown (**Figure 5C**).

339   The heat-map in **Figure 5D** shows the data for all five sorted pools for the cohorts

340   described above and for additional knockdowns enriched in the G_1 or S phase (tRNA

341   synthetases), S phase (core histones) or G_2M pools (PSP1, DNA polymerase suppressor 1),

342   or not enriched in any pool. These latter sets provide further controls that do not appear to

343   have specific impacts on cell cycle progression, including the mitochondria RNA editing

344   accessory complex MRB1 (60) and the mitochondrial ATP synthase complex V (61). Thus,

345   we identify a number of protein complexes, pathways and regulatory factors that are

346   specifically required for progressive steps through the trypanosome cell cycle.

347

348   **Pathways and protein complexes associated with G_1, S phase and G_2M defects**

349   We next explored some of the cohorts of hits described above in more detail.

350   Glycolytic enzymes are particularly prominent amongst knockdowns that accumulate in G_1

351   and we illustrate the RIT-seq profiling data for these enzymes in **Figure 6A**. Seven of eleven

352   glycolytic enzyme knockdowns register >25% overrepresentation in the G_1 pool; hexokinase

353   (Tb927.10.2010), phosphofructokinase (Tb927.3.3270), aldolase (see **Figure 2C**),

354   triosephosphate isomerase (Tb927.11.5520), glyceraldehyde 3-phosphate dehydrogenase

355   (Tb927.6.4280), phosphoglycerate kinase C (Tb927.1.700) and pyruvate kinase

356   (Tb927.10.14140). Glycolysis operates in peroxisome-like organelles known as glycosomes

357   in trypanosomes and is thought to be the single source of ATP in bloodstream form cells

358   (12). Glycolysis also provides metabolic intermediates that support nucleotide production.

359   Notably, mammalian cell proliferation is accompanied by activation of glycolysis, and the

360   Warburg effect relates to this phenomenon in oncology (62). Indeed, hexokinase regulates

361   the G_1/S checkpoint in tumour cells (63). The results are also consistent with the observation

362   that *T. brucei* accumulate in G_1 or G_0 under growth-limiting conditions (64) or during

363   differentiation to the non-dividing stumpy form (65), possibly reflecting a role for glucose

364   sensing in differentiation (66). Notably, glycolytic enzymes are downregulated 6.7+/-5.2-fold
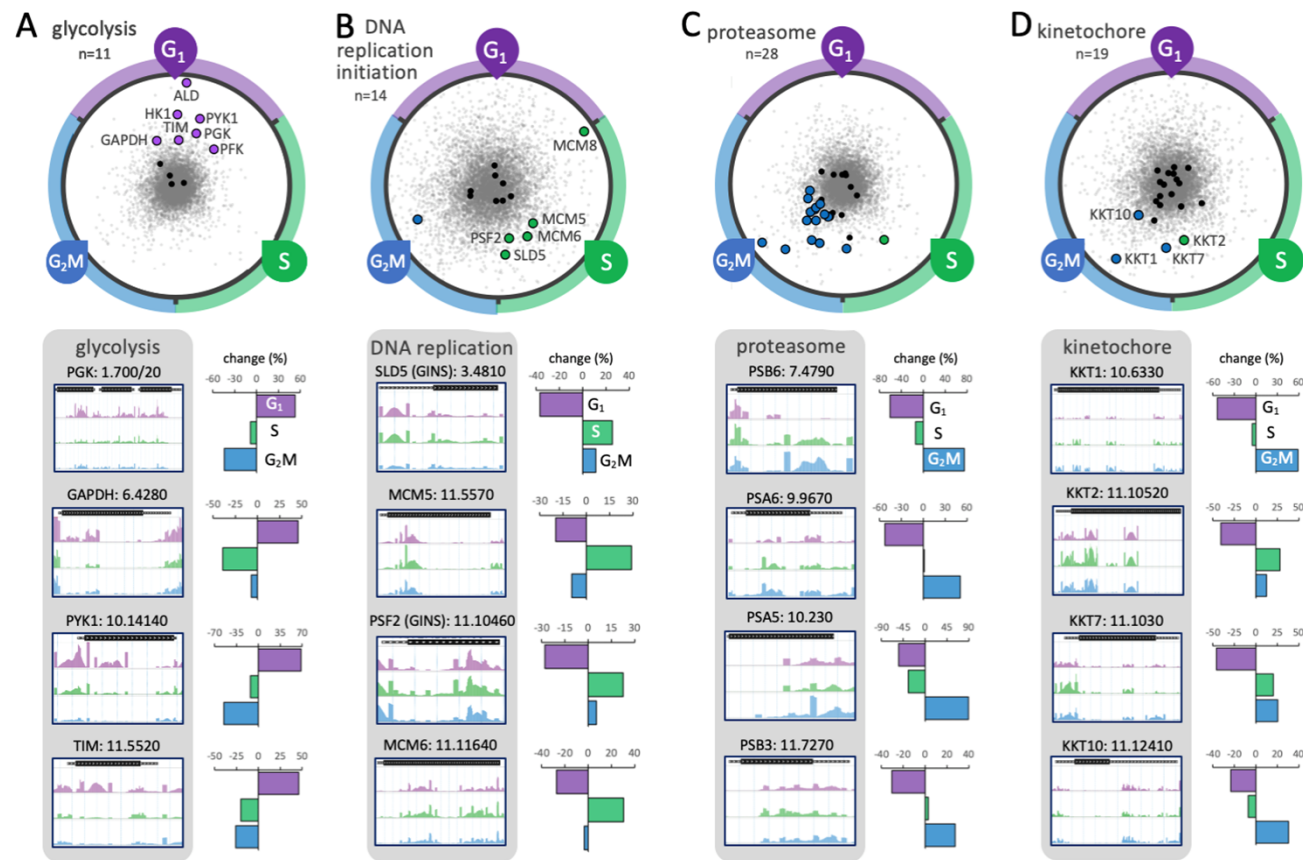
11

Figure 6

Figure 6. Protein complexes and pathways associated with G_1, S phase and G_2/mitosis defects. (**A**) The RadViz plot shows glycolytic enzyme knockdowns. Those that registered >25% overrepresented read-counts in the G_1 category are indicated, purple. Black data-points indicate other genes from each cohort. Grey data-points indicate all other genes. The read-mapping profiles and relative read-counts in the lower panel show example hits. (**B**) As in a but for DNA replication initiation factor knockdowns that registered >25% overrepresented read-counts in the S phase category, indicated in green. (**C**) As in a but for proteasome component knockdowns that registered >25% overrepresented read-counts in the G_2M category, indicated in blue. (**D**) As in a but for kinetochore component knockdowns that registered >25% overrepresented read-counts in the S phase or G_2M categories, indicated in green or blue, respectively.

365 in stumpy-form cells (67). We conclude that, as in other organisms (68), there is metabolic

366 control of the cell cycle and a nutrient sensitive restriction point in *T. brucei*, with glycolysis

367 playing a role in the G_1 to S phase transition.

368 DNA replication initiation factors are particularly prominent amongst knockdowns that

369 accumulate in S phase and we illustrate the RIT-seq profiling data for these factors in *Figure*

370 *6B*. Five knockdowns that register >25% overrepresentation in the S phase pool are

371 components of the eukaryotic replicative helicase, the CMG (Cdc45-MCM-GINS) complex.

372 At the core of this complex is the minichromosome maintenance complex (MCM2-7), a

373 helicase that unwinds the duplex DNA ahead of the moving replication fork (69).

374 Identification of this subset of components suggests that these particular subunits are

375 limiting for progression through S phase.

376 Proteasome components are particularly prominent amongst knockdowns that

377 accumulate in G_2M, and we illustrate the RIT-seq profiling data for this protein complex in

378 *Figure 6C*. Sixteen of 28 proteasome component knockdowns register >25%

379 overrepresentation in the G_2M pool. This output suggests that the *T. brucei* proteasome is

380 most likely responsible for degrading cell cycle regulators, such as poly-ubiquitinated cyclins,

381 which are known to control cell cycle checkpoints in other eukaryotes. Candidate target

382 cyclins in *T. brucei* include: cyclin 6 (CYC6, Tb927.11.16720), degradation of which is

383 required for mitosis (70); cyclin-like CFB2 (Tb927.1.4650), required for cytokinesis (71);

384 cyclin 2 (CYC2, Tb927.6.1460) or cyclin 3 (CYC3, Tb927.6.1460), which have short half-

385 lives and a candidate destruction box motif in the case of CYC3 (72).

386 Kinetochore components (17) are also amongst knockdowns that accumulate in G_2M

387 and we illustrate the RIT-seq profiling data for this protein complex in *Figure 6D*. Although

388 knockdown of KKT2 (Tb927.10.10520), a putative kinase, registered overrepresentation in

389 the S phase pool, KKT1 (Tb927.10.6330), KKT7 (Tb927.11.1030) and KKT10 (CLK1,

390 Tb927.11.12410) knockdowns registered >25% overrepresentation in the G_2M pool,

391 suggesting that these particular kinetochore components, which all display temporal patterns

392 of phosphorylation from S phase to G_2M (21), are limiting for progression through mitosis.

393 Notably, KKT10 is a kinase responsible for phosphorylation of KKT7, which is required for

394 the metaphase to anaphase transition (73); as well as for the phosphorylation of KKT1,

395 which is required for kinetochore assembly (74). These findings are consistent with the view

396 that kinetochore components control a non-canonical spindle checkpoint in trypanosomes

397 (73).

398

399 **RBPs, kinases and hypothetical proteins associated with G_1, S phase and G_2M defects**

400 Widespread polycistronic transcription in trypanosomatids places great emphasis on

401 post-transcriptional controls and, consistent with this, knockdowns overrepresented in the
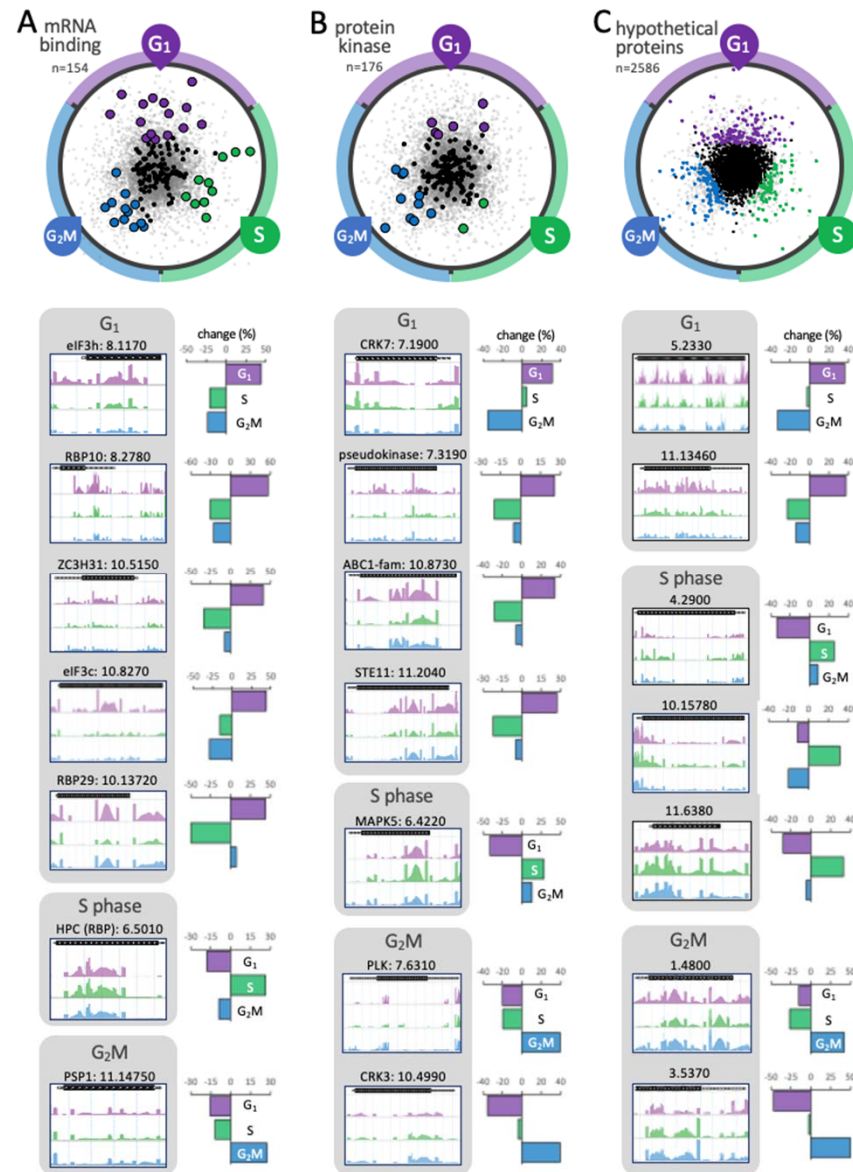
12

Figure 7

Figure 7. RBPs kinases and hypothetical proteins associated with gap and S phase defects. (**A**) The RadViz plot shows mRNA binding protein knockdowns (RBPs). Those that registered >25% overrepresented read-counts in the G_1, S phase or G_2M categories are indicated, in purple, green and blue, respectively. The read-mapping profiles and relative read-counts in the lower panels show example hits. (**B**) As in a but for protein kinase knockdowns. (**C**) As in a but for hypothetical (conserved) protein knockdowns.

402   G_1, S phase and G_2M pools revealed many putative mRNA binding proteins (RBPs) and

403   kinases. Indeed, RBPs are significantly enriched amongst knockdowns that registered G_1, S

404   phase or G_2M cell cycle defects (χ^2 test, $p = 7^{-5}$). We show the RIT-seq profiling data for

405   seven RBP knockdowns that register >25% overrepresentation in these pools (*Figure 7A*).

406   These include knockdowns for two components of the translation initiation factor, eIF3,

407   linked to accumulation in G_1 (Tb927.8.1170; Tb927.10.8270). Proposed regulatory functions

408   for eIF3 include reinitiation of translation on polycistronic mRNAs and as a substrate for

409   translation regulatory kinases (75). The RBP10 (Tb927.8.2780), RBP29 (Tb927.10.13720)

410   and ZC3H31 (Tb927.10.5150) knockdowns were also enriched in G_1. RBP10, in particular,

411   has been characterised in some detail and promotes the bloodstream form state (76). Thus,

412   the RIT-seq cell cycle screen implicated a number of specific RBPs in post-transcriptional

413   control of cell cycle progression, most likely through modulation of mRNA stability and/or

414   translation of cell cycle regulators.

415   We show data for several protein kinases above, linked to enriched >4C (*Figure 3D*),

416   sub-2C (*Figure 4*), S phase or G_2M (*Figure 6D*) phenotypes, and now show the RIT-seq

417   profiling data for seven additional protein kinase knockdowns that register >25%

418   overrepresentation in the G_1, S phase or G_2M pools (*Figure 7B*). These include knockdowns

419   for CRK7 (Tb927.7.1900), linked to accumulation in G_1; MAPK5 (Tb927.6.4220), linked to

420   accumulation in S phase and polo-like kinase (PLK, Tb927.7.6310) and cdc2-related kinase

421   3 (CRK3, Tb927.10.4990), linked to accumulation in G_2M. PLK was previously shown to

422   control cell morphology, furrow ingression and cytokinesis (77-79), while CRK3 was shown

423   to play a role in G_2M progression in bloodstream form *T. brucei* (42,80).

424   Finally, we analysed genes encoding proteins annotated as hypothetical (conserved).

425   Despite excellent progress in genome annotation, 35% of the non-redundant gene-set in *T.*

426   *brucei* retain this annotation, amounting to >2,500 genes. We show data for more than

427   twenty of these knockdowns above, linked to enriched >4C (*Figure 3—figure supplement*

428   *1*) or sub-2C (*Figure 4*) phenotypes, and we identify additional hypothetical (conserved)

429   protein knockdowns that register >25% overrepresentation in the G_1, S phase or G_2M pools

430   RIT-seq profiling data are shown for seven examples in *Figure 7C* and for twenty four

431   examples in *Figure 7 – figure supplement 1*. Amongst fifty-four other examples of

432   knockdowns shown in *Figure 7—figure supplement 1*, are alternative oxidase

433   (Tb927.10.7090) (81), linked to G_1 enrichment; tRNA synthetases linked to G_1 (Trp,

434   Tb927.3.5580; His, Tb927.6.2060) or S phase enrichment (Glu, Tb927.6.4590; Lys,

435   Tb927.8.1600); kinesins linked to S phase (Tb927.7.7120) or G_2M enrichment, including

436   both chromosomal passenger complex kinesins (26) (Tb927.11.2880, KIN-A and

437   Tb927.7.5040, KIN-B) and KIN-G (Tb927.6.1770); CYC6 (25,82), CFB2 (71), centrin 3

438   (Tb927.10.8710) (83) and, finally, both components of the histone chaperone FACT

439   (facilitates chromatin transcription) complex (84) Spt16 (Tb927.3.5620) and Pob3

440   (Tb927.10.14390), linked to G_2M enrichment. Notably, the FACT complex has been linked to

441   centromere function in human cells (85).

442

443   **Linking cell cycle regulated transcripts and proteins to cell cycle progression defects**

444   Some factors required for cell cycle progression are themselves cell cycle regulated.

445   To identify some of these factors, we compared our functional assignment data with

446   quantitative transcriptome (19), proteome (20) and phosphoproteome (21) cell cycle profiling

447   data. An initial survey of all 1,158 genes that registered a cell cycle defect here (see *Figure*

448   *2D*) revealed significant enrichment of cell cycle regulated mRNAs (overlap = 105 of 485, χ^2

449     $p = 8.5^{-4}$), as well as proteins displaying cell cycle regulated phosphorylation (overlap = 106

450   of 547, χ^2 $p = 0.035$). It is important to note here, however, and for the analyses below, that

451   these transcriptome and (phospho)proteome datasets were derived from insect stage *T.*

452   *brucei*, meaning that regulation may differ in some cases in bloodstream *T. brucei* cells used

453   for RIT-seq analysis here.

454   Specific transcripts required for cell cycle progression are likely upregulated prior to

455   peak demand for the encoded protein, and we found evidence to support this view. For

456   example, transcripts upregulated in late G_1 or in S phase were enriched amongst those

457   knockdowns linked to accumulation in the G_2M pool here (χ^2 $p = 3.3^{-3}$ and $p = 1.1^{-2}$

458   respectively); G_1 upregulated transcripts included, for example, both components of the

459   FACT complex (see *Figure 7—figure supplement 1*). In addition, S phase and G_2M

460   upregulated transcripts, including those encoding multiple flagellum-associated proteins,

461   were enriched amongst knockdowns linked to accumulation in the >4C pool (χ^2 $p = 4.6^{-18}$

462   and $p = 2.4^{-5}$ respectively).

463   Although the overlap between cell cycle regulated proteins (20) and those

464   knockdowns that registered a cell cycle defect here failed to achieve significance (overlap =

465   71 of 367, χ^2 $p = 0.09$), cell cycle regulated proteins do appear to be required for progression

466   through specific stages of the cell cycle. For example, multiple glycolytic enzymes

467   upregulated in G_1 were linked to accumulation in the G_1 pool following knockdown (χ^2 $p =$

468   7.9^{-11}). In addition, proteins highly upregulated in G_2 and M were linked to accumulation in

469   the G_2M (χ^2 $p = 1.8^{-8}$) or >4C pools (χ^2 $p = 8.9^{-9}$) following knockdown, including multiple

470   kinetochore and chromosomal passenger complex components, respectively.

471   In terms of specific cell cycle regulated genes/proteins, we focused on those that

472   previously registered a significant loss of fitness following knockdown (23) and now with a

473   RIT-seq based cell cycle progression functional assignment. Examples include putative

474   RBPs of the DNA polymerase suppressor 1 (PSP1) family, which display mRNA

475   upregulation in G_1, protein upregulation in S phase, cell cycle regulated phosphorylation and

476   accumulation in G_2M following knockdown (see *Figure 5D* and *Figure 7A*). The kinetochore

477   components, KKT1 and KKT7, and also CRK3, all display mRNA upregulation in S phase,

478   protein upregulation in G_2 and M, cell cycle regulated phosphorylation and accumulation in

479   G_2M following knockdown (see *Figure 6D* and *Figure 7B*); KKT10 and CYC6 report a

480   similar profile (see *Figure 6D*), except for the mRNA regulation component. The cytokinesis

481   initiation factors, CIF1 and CIF2, also display mRNA upregulation in S phase, protein

482   upregulation in G_2 and M and cell cycle regulated phosphorylation, but instead accumulation

483   in the >4C pool following knockdown (see *Figure 3E*). Finally, the chromosomal passenger

484   complex components, CPC1 and AUK1, as well as furrow localized FRW1, report mRNA

485   and protein upregulation in G_2M and accumulation in the >4C pool following knockdown (see

486   *Figure 3E*). Thus, several regulators linked to specific cell cycle progression defects by RIT-

487   seq profiling, are themselves regulated.

Discussion

488

489   Despite intense interest and study (13,15), many cell cycle regulators in

490   trypanosomatids remain to be identified and much remains to be learned about cell cycle

491   control and progression in these parasites. DNA staining followed by flow cytometry is a

492   widely used approach for quantifying cellular DNA content and analysing cell cycle

493   distribution across otherwise asynchronous populations. Here, we combined genome scale

494   loss-of-function genetic screening with flow cytometry in bloodstream form African

495   trypanosomes and identify hundreds of genes required for progression through specific

496   stages of the cell cycle.

497   Functional annotation of the trypanosomatid genomes will continue to benefit from

498   novel high-throughput functional analyses, and RNAi-mediated knockdown has proven to be

499   a powerful approach for *T. brucei*. RIT-seq profiling provides data for almost every gene and,

500   using this approach, we previously described genome-scale loss-of-fitness data (23).

501   Amongst 3117 knockdowns that scored a significant loss-of-fitness in bloodstream-form cells

502   in that screen (42% of all genes analysed) were genes encoding all 18 intraflagellar transport

503   complex subunits (χ^2 $p = 1^{-6}$), 12 of 13 dynein heavy-chains (χ^2 $p = 4^{-4}$), all 8 TCP-1

504   chaperone components (χ^2 $p = 1^{-3}$), 27 of 30 nucleoporins (χ^2 $p = 2^{-7}$), all eleven glycolytic

505   enzymes (χ^2 $p = 2^{-4}$) and 30 of 31 proteasome subunits (χ^2 $p = 2^{-9}$). This set also included 18

506   of 19 kinetochore proteins (χ^2 $p = 6^{-6}$), only later identified as components of this essential

507   complex (17). With this study, we now link many of these genes and many more to specific

508   cell cycle defects following RNAi knockdown. A large number of flagellar protein

509   knockdowns, in particular, yield cells with excess DNA, suggesting that DNA replication

510   typically continues following failure to complete cytokinesis. We identified a number of

511   pathways and protein complexes that impact cell cycle progression, such as glycolysis (G_1/S

512   transition) and the proteasome (likely G_2/M transition). We also identify many mRNA binding

513   proteins and protein kinases implicated in control of cell cycle progression. Notably, we link

514   multiple known potential and promising drug targets to cell cycle progression defects, such

515   as glycolytic enzymes (86), the proteasome (87), kinetochore kinases (74,88) and other

516   kinases (89).

517   Prior cell cycle studies have often focused on trypanosome orthologues of known

518   regulators from other eukaryotes. Since genome-scale profiling is unbiased, it presents the

519   opportunity to uncover divergent as well as novel factors and regulators that impact cell

520   cycle progression. Accordingly, we link many previously uncharacterised and hypothetical

521   proteins of unknown function to specific cell cycle progression defects. Thus, we uncover

522   mechanisms with an ancient origin in a common eukaryotic ancestor and others likely

523   reflecting trypanosomatid-specific biology. We also compared our functional data with cell

524   cycle regulated transcriptome and (phospho)proteome datasets.

525   The digital dataset provided in *Supplementary File 1* facilitates further interrogation

526   and further analysis of the genome-scale cell cycle RIT-seq data. We have also made the

527   data available via an interactive, open access, online data visualization tool, which allows

528   searching and browsing of the data (see *Figure 2—figure supplement 2*). Comparison with

529   existing and new datasets, including with high-throughput subcellular localisation data (90)

530   www.tryptag.org, should also facilitate future studies. Since high-throughput genetic screens

531   typically yield a proportion of false positive 'hits', we do urge some caution, however, in

532   particular where outputs are predominantly generated by a single RIT-seq fragment. On the

533   other hand, there are knockdowns in the current dataset that show specific cell cycle phase

534   enrichment yet fail to register a sufficient enrichment to surpass the thresholds applied

535   above. Considering both of these points, we hope that the digital dataset and the online

536   database will serve as valuable resources. Since other important trypanosomatid parasites,

537   including *Trypanosoma cruzi* and *Leishmania,* share a high degree of conservation and

538   synteny with *T. brucei* (91), the current datasets can also assist and inform studies on other

539   trypanosomatids.

540   In summary, we report RNAi induced cell cycle defects at a genomic scale and

541   identify the *T. brucei* genes that underlie these defects. The outputs confirm known roles in

542   cell cycle progression and provide functional annotation for many additional genes, including

543   many with no prior functional assignment and many that are trypanosomatid-specific. As

544   such, the data not only improve our understanding of cell cycle progression in these

545   important and divergent pathogens but should also accelerate further discovery. Taken

546   together, our findings further facilitate genome annotation, drug-target prioritisation and

547   provide comprehensive genetic evidence for the protein complexes, pathways and

548   regulatory factors that coordinate progression through the trypanosome cell cycle.

549 **Materials and Methods**

550

551 ***T. brucei* RNAi library growth and manipulation**

552 The bloodstream form *T. brucei* RNAi library (22) was thawed in HMI-11 containing 1 μg.ml^{-1}

553 of blasticidin and 0.2 μg.ml^{-1} of phleomycin and incubated at 37°C in 5% CO_2. After

554 approximately 48 h, six flasks, each containing 2 x10^7 cells in 150 ml of HMI-11 as above,

555 were prepared; 1 μg.ml^{-1} of tetracycline was added to five of them, while one served as the

556 non-induced control. The cells were grown under these conditions for 24 h and then

557 harvested by centrifugation for 10 min at 1000 g. Cells from each flask were then re-

558 suspended in 25 ml of 1x PBS (pH 7.0) supplemented with 5 mM EDTA and 1 % FBS

559 ("supplemented PBS"), centrifuged again for 10 min at 1000 g, and then re-suspended in 0.5

560 ml of supplemented PBS. To each cell suspension, 9.5 ml of 1 % formaldehyde in

561 supplemented PBS was added dropwise, with regular vortexing. The cells were incubated

562 for a further 10 min at room temperature and then washed twice in 10 ml of supplemented

563 PBS using centrifugation as above. The cells were finally re-suspended at 2.5 x 10^7 per ml in

564 supplemented PBS and were subsequently stored at 4°C, in the dark.

565

566 **Flow cytometry**

567 Fixed cells, 3 x 10^8 Tet-induced and 10^7 uninduced, were centrifuged for 10 min at 1000 g,

568 and re-suspended in 10 ml of supplemented PBS containing 0.01% Triton X-100 (Sigma

569 Aldrich). The cells were incubated for 30 min at room temperature, centrifuged for 10 min at

570 700 g and washed once in 10 ml of supplemented PBS. The cells were then re-suspended in

571 4 ml of supplemented PBS containing 10 μg.ml^{-1} of propidium iodide (Sigma Aldrich) and

572 100 μg.ml^{-1} of RNaseA (Sigma Aldrich), and incubated for 45 min at 37°C, in the dark; cells

573 were subsequently kept on ice and in the dark. Immediately prior to sorting, the Tet-induced

574 cells were filtered (Filcon Cup-type filter, 50 μm mesh, BD™ Medimachine) into 5 ml

575 polystyrene round-bottom tubes (BD Falcon). Cells were sorted using the BD Influx™

576 (Becton Dickinson) cell sorter, with BD FACSort™ software, at the Flow Cytometry and Cell

577 Sorting Facility in the School of Life Sciences, University of Dundee. The cells were sorted

578 into pools of <2C (~8 x 10^5 cells), 2C (G$_1$, 1 x 10^7 cells), 2-4C (S, 1 x 10^7 cells), 4C (G$_2$M, 1 x

579 10^7 cells) and >4C (~5 x 10^5 cells) based on their DNA content, and collected into 50 ml

580 Falcon tubes (BD Falcon); sorting time was approx. 4 h. The 2C, 2-4C and 4C sorted

581 samples were then run on a FACS LSR Fortessa flow cytometry analyser for a post-sorting

582 quality check. FlowJo v10 was used for data analysis and visualisation.

583

584 **RNA interference target amplification**

585

586 The five pools of Tet-induced, sorted cells as well as uninduced or induced, but unsorted

587 cells, were lysed overnight at 56°C in 50% (v/v) of Buffer AL (Qiagen) and 0.5 mg.ml^{-1} of

588 Proteinase K (Qiagen), to reverse formaldehyde crosslinking. Genomic DNA was then

589 extracted using the DNeasy Blood and Tissue DNA extraction kit (Qiagen), according to the

590 manufacturer's instructions, with the exception that each sample was eluted in 50 μl of

591 Buffer AE. The whole sample (range = 140-840 ng) was used for PCR, in a 100 μl reaction,

592 using OneTaq (NEB), and the Lib3F (CCTCGAGGGCCAGTGAG) and Lib3R

593 (ATCAAGCTTGGCCTGTGAG) primers and with the following programme: 94°C for 4 min,

594 followed by 27 cycles of 94°C for 30 sec, 55°C for 30 sec and 68°C for 2 min and 10 sec,

595 and a final extension of 68°C for 5 min. The PCR products were then purified using the

596 Qiaquick PCR extraction kit (Qiagen), as per the manufacturer's instructions, and eluted in

597 30 μl of nuclease-free water (Ambion); two columns per sample.

598

599 **RIT-seq library preparation and sequencing**

600 Purified PCR products were used for library preparation and sequencing at the Tayside

601 Centre for Genomic Analysis at the University of Dundee. The PCR products were

602 fragmented using a Covaris M220 sonicator (20% duty factor, 75W peak/displayed power,

603 60 seconds duration – 3 x 20 sec with intermittent spin down step, 18-20°C temperature;

604 resulting in 250-300 bp enriched fragments), and the libraries were prepared using the

605 Truseq Nano DNA Library Prep kit (Illumina). The samples were multiplexed, and sequenced

606 on an Illumina NextSeq 500 platform, on a 150 cycle Output Cartridge v2, paired-end. Each

607 library was run on 4 sequencing lanes. Base call, index deconvolution, trimming and QC

608 were performed in BaseSpace using bcl2fastq2 Conversion Software v2.17.

609

610 **RIT-seq data mapping and analysis**

611 The sequencing data analysis pipeline was adapted from (22). The FASTQ files with

612 forward and reverse paired end reads (4 technical replicates for each samples) were

613 concatenated and aligned to the reference genome v46 of *T. brucei* clone TREU927

614 downloaded from TriTrypDB (28) using Bowtie2 (92), with the 'very-sensitive-local' pre-set

615 alignment option. The alignments were converted to BAM format, reference sorted and

616 indexed with SAMtools (93). The quality of alignments was evaluated with Qualimap 2 (94)

617 using the bamqc and rnaseq options. The Qualimap 2 output files were aggregated with

618 MultiQC (95) and inspected. The alignments were deduplicated with the Picard tools

619 package using the MarkDuplicates function (http://broadinstitute.github.io/picard/); to

620 minimise the potential for overrepresentation of the shortest RIT-seq fragments. Alignments

19

621   with properly paired reads were extracted with SAMtool view using the -f 2 option and

622   parsed with a custom python script to extract the paired reads containing the barcode

623   sequence (GTGAGGCCTCGCGA) in forward or reverse complement orientation. The

624   genome coverage of the aligned reads was extracted from the bam files using bedtools (96)

625   with the -bg option to output bedGraph files. The bedGraph files were visualized with the

626   svist4get python package (97). Read counts for protein coding sequences and associated

627   untranslated regions (where annotated) were determined from the bam files using

628   featureCounts (98) and normalized to Transcripts Per Kilobase Million (TPM). Dimensionality

629   reduction of the G_1, S and G_2M TPM values was performed with the radviz algoritm

630   implemented in the pandas python package (99). The bash script containing the analysis

631   pipline, a conda environment specification file for its execution, the python script to extract

632   barcoded reads and a basic usage example are available at GitHub

633   (https://github.com/mtinti/ritseq_cellcycle). Data were subsequently analysed using a GO-

634   slim set and Gene Ontology tools available via tritrypdb.org and visualised using tools

635   available at huygens.science.uva.nl/PlotsOfData.

20

646   **Author contributions**

647   Catarina A. Marques, Conceptualization, Formal analysis, Investigation, Methodology,

648   Writing—review and editing; Michele Tinti, Data curation, visualisation and analysis,

649   Writing—review and editing; Andrew Cassidy, Investigation, Methodology; David Horn,

650   Conceptualization, Formal analysis, Supervision, Funding acquisition, Project administration,

651   Writing—original draft, Writing—review and editing.

652

653   **Author ORCIDs**

654   David Horn   http://orcid.org/0000-0001-5173-9284

655   Catarina A. Marques   https://orcid.org/0000-0003-1324-5448

656   Michele Tinti   https://orcid.org/0000-0002-0051-017X

657

658   **Competing interests:** The authors declare that they have no competing interests.

659

660   **Data and materials availability:** High-throughput sequencing data generated for

661   this study have been deposited in the Short Read Archive (SRA) at

662   https://www.ncbi.nlm.nih.gov/sra/PRJNA641153 under primary accession number

663   PRJNA641153.

21

References

664

665 1. Harashima, H., Dissmeyer, N. and Schnittger, A. (2013) Cell cycle control across the
666   eukaryotic kingdom. *Trends Cell Biol*, **23**, 345-356.
667 2. Pollard, T.D. and O'Shaughnessy, B. (2019) Molecular mechanism of mytokinesis.
668   *Annu Rev Biochem*, **88**, 661-689.
669 3. Poon, R.Y. (2016) Cell cycle control: A system of interlinking oscillators. *Methods*
670   *Mol Biol*, **1342**, 3-19.
671 4. Visconti, R., Della Monica, R. and Grieco, D. (2016) Cell cycle checkpoint in cancer:
672   a therapeutically targetable double-edged sword. *J Exp Clin Cancer Res*, **35**, 153.
673 5. Buscher, P., Cecchi, G., Jamonneau, V. and Priotto, G. (2017) Human African
674   trypanosomiasis. *Lancet*, **390**, 2397-2409.
675 6. Matthews, K.R. (2015) 25 years of African trypanosome research: From description
676   to molecular dissection and new drug discovery. *Mol Biochem Parasitol*, **200**, 30-40.
677 7. Broadhead, R., Dawe, H.R., Farr, H., Griffiths, S., Hart, S.R., Portman, N., Shaw,
678   M.K., Ginger, M.L., Gaskell, S.J., McKean, P.G. *et al.* (2006) Flagellar motility is
679   required for the viability of the bloodstream trypanosome. *Nature*, **440**, 224-227.
680 8. Baron, D.M., Kabututu, Z.P. and Hill, K.L. (2007) Stuck in reverse: loss of LC1 in
681   *Trypanosoma brucei* disrupts outer dynein arms and leads to reverse flagellar beat and
682   backward movement. *J Cell Sci*, **120**, 1513-1520.
683 9. Bonnefoy, S., Watson, C.M., Kernohan, K.D., Lemos, M., Hutchinson, S., Poulter,
684   J.A., Crinnion, L.A., Berry, I., Simmonds, J., Vasudevan, P. *et al.* (2018) Biallelic
685   mutations in LRRC56, encoding a protein associated with intraflagellar transport,
686   cause mucociliary clearance and laterality defects. *Am J Hum Genet*, **103**, 727-739.
687 10. Coutton, C., Vargas, A.S., Amiri-Yekta, A., Kherraf, Z.E., Ben Mustapha, S.F., Le
688   Tanno, P., Wambergue-Legrand, C., Karaouzene, T., Martinez, G., Crouzy, S. *et al.*
689   (2018) Mutations in CFAP43 and CFAP44 cause male infertility and flagellum
690   defects in *Trypanosoma* and human. *Nat Commun*, **9**, 686.
691 11. Vincensini, L., Blisnick, T. and Bastin, P. (2011) [The importance of model
692   organisms to study cilia and flagella biology]. *Biol Aujourdhui*, **205**, 5-28.
693 12. Allmann, S. and Bringaud, F. (2017) Glycosomes: A comprehensive view of their
694   metabolic roles in *T. brucei*. *Int J Biochem Cell Biol*, **85**, 85-90.
695 13. Wheeler, R.J., Gull, K. and Sunter, J.D. (2019) Coordination of the cell cycle in
696   trypanosomes. *Annu Rev Microbiol*, **73**, 133-154.
697 14. Clayton, C. (2019) Regulation of gene expression in trypanosomatids: living with
698   polycistronic transcription. *Open Biol*, **9**, 190072.
699 15. Zhou, Q., Hu, H. and Li, Z. (2014) New insights into the molecular mechanisms of
700   mitosis and cytokinesis in trypanosomes. *Int Rev Cell Mol Biol*, **308**, 127-166.
701 16. Ploubidou, A., Robinson, D.R., Docherty, R.C., Ogbadoyi, E.O. and Gull, K. (1999)
702   Evidence for novel cell cycle checkpoints in trypanosomes: kinetoplast segregation
703   and cytokinesis in the absence of mitosis. *J Cell Sci*, **112 (Pt 24)**, 4641-4650.
704 17. Akiyoshi, B. and Gull, K. (2014) Discovery of unconventional kinetochores in
705   kinetoplastids. *Cell*, **156**, 1247-1258.
706 18. Marques, C.A. and McCulloch, R. (2018) Conservation and variation in strategies for
707   DNA replication of kinetoplastid nuclear genomes. *Curr Genomics*, **19**, 98-109.
708 19. Archer, S.K., Inchaustegui, D., Queiroz, R. and Clayton, C. (2011) The cell cycle
709   regulated transcriptome of *Trypanosoma brucei*. *PLoS One*, **6**, e18425.
710 20. Crozier, T.W.M., Tinti, M., Wheeler, R.J., Ly, T., Ferguson, M.A.J. and Lamond, A.I.
711   (2018) Proteomic analysis of the cell cycle of procylic form *Trypanosoma brucei*. *Mol*
712   *Cell Proteomics*, **17**, 1184-1195.

713   21.   Benz, C. and Urbaniak, M.D. (2019) Organising the cell cycle in the absence of
714   transcriptional control: Dynamic phosphorylation co-ordinates the *Trypanosoma*
715   *brucei* cell cycle post-transcriptionally. *PLoS Pathog*, **15**, e1008129.
716   22.   Glover, L., Alsford, S., Baker, N., Turner, D.J., Sanchez-Flores, A., Hutchinson, S.,
717   Hertz-Fowler, C., Berriman, M. and Horn, D. (2015) Genome-scale RNAi screens for
718   high-throughput phenotyping in bloodstream-form African trypanosomes. *Nat Protoc*,
719   **10**, 106-133.
720   23.   Alsford, S., Turner, D.J., Obado, S.O., Sanchez-Flores, A., Glover, L., Berriman, M.,
721   Hertz-Fowler, C. and Horn, D. (2011) High-throughput phenotyping using parallel
722   sequencing of RNA interference targets in the African trypanosome. *Genome Res*, **21**,
723   915-924.
724   24.   Ngo, H., Tschudi, C., Gull, K. and Ullu, E. (1998) Double-stranded RNA induces
725   mRNA degradation in *Trypanosoma brucei*. *Proc Natl Acad Sci U S A*, **95**, 14687-
726   14692.
727   25.   Hammarton, T.C., Clark, J., Douglas, F., Boshart, M. and Mottram, J.C. (2003) Stage-
728   specific differences in cell cycle control in *Trypanosoma brucei* revealed by RNA
729   interference of a mitotic cyclin. *J Biol Chem*, **278**, 22877-22886.
730   26.   Li, Z., Umeyama, T. and Wang, C.C. (2008) The chromosomal passenger complex
731   and a mitotic kinesin interact with the tousled-like kinase in trypanosomes to regulate
732   mitosis and cytokinesis. *PLoS One*, **3**, e3814.
733   27.   Kolev, N.G., Tschudi, C. and Ullu, E. (2011) RNA interference in protozoan
734   parasites: achievements and challenges. *Eukaryot Cell*, **10**, 1156-1163.
735   28.   Aslett, M., Aurrecoechea, C., Berriman, M., Brestelli, J., Brunk, B.P., Carrington, M.,
736   Depledge, D.P., Fischer, S., Gajria, B., Gao, X. *et al.* (2010) TriTrypDB: a functional
737   genomic resource for the Trypanosomatidae. *Nucleic Acids Res*, **38**, D457-462.
738   29.   Springer, A.L., Bruhn, D.F., Kinzel, K.W., Rosenthal, N.F., Zukas, R. and Klingbeil,
739   M.M. (2011) Silencing of a putative inner arm dynein heavy chain results in flagellar
740   immotility in *Trypanosoma brucei*. *Mol Biochem Parasitol*, **175**, 68-75.
741   30.   Gassen, A., Brechtefeld, D., Schandry, N., Arteaga-Salas, J.M., Israel, L., Imhof, A.
742   and Janzen, C.J. (2012) DOT1A-dependent H3K76 methylation is required for
743   replication regulation in *Trypanosoma brucei*. *Nucleic Acids Res*, **40**, 10302-10311.
744   31.   Choe, K.N. and Moldovan, G.L. (2017) Forging ahead through darkness: PCNA, still
745   the principal conductor at the replication fork. *Mol Cell*, **65**, 380-392.
746   32.   Valenciano, A.L., Ramsey, A.C. and Mackey, Z.B. (2015) Deviating the level of
747   proliferating cell nuclear antigen in *Trypanosoma brucei* elicits distinct mechanisms
748   for inhibiting proliferation and cell cycle progression. *Cell Cycle*, **14**, 674-688.
749   33.   Rudd, S.G., Glover, L., Jozwiakowski, S.K., Horn, D. and Doherty, A.J. (2013) PPL2
750   translesion polymerase is essential for the completion of chromosomal DNA
751   replication in the African trypanosome. *Mol Cell*, **52**, 554-565.
752   34.   Ralston, K.S., Lerner, A.G., Diener, D.R. and Hill, K.L. (2006) Flagellar motility
753   contributes to cytokinesis in *Trypanosoma brucei* and is modulated by an
754   evolutionarily conserved dynein regulatory system. *Eukaryot Cell*, **5**, 696-711.
755   35.   Boehm, C.M., Obado, S., Gadelha, C., Kaupisch, A., Manna, P.T., Gould, G.W.,
756   Munson, M., Chait, B.T., Rout, M.P. and Field, M.C. (2017) The trypanosome
757   exocyst: A conserved structure revealing a new role in endocytosis. *PLoS Pathog*, **13**,
758   e1006063.
759   36.   Wang, D.Y., Kamuda, K., Montoya, G. and Mesa, P. (2020) The TRiC/CCT
760   chaperonin and its role in uncontrolled proliferation. *Adv Exp Med Biol*, **1243**, 21-40.

761   37.   Zhang, X., Hu, H., Lun, Z.R. and Li, Z. (2019) Functional analyses of an axonemal
762   inner-arm dynein complex in the bloodstream form of *Trypanosoma brucei* uncover
763   its essential role in cytokinesis initiation. *Mol Microbiol*, **112**, 1718-1730.
764   38.   Sunter, J.D., Varga, V., Dean, S. and Gull, K. (2015) A dynamic coordination of
765   flagellum and cytoplasmic cytoskeleton assembly specifies cell morphogenesis in
766   trypanosomes. *J Cell Sci*, **128**, 1580-1594.
767   39.   Hu, H., Majneri, P., Li, D., Kurasawa, Y., An, T., Dong, G. and Li, Z. (2017)
768   Functional analyses of the CIF1-CIF2 complex in trypanosomes identify the structural
769   motifs required for cytokinesis. *J Cell Sci*, **130**, 4108-4119.
770   40.   Li, Z., Lee, J.H., Chu, F., Burlingame, A.L., Gunzl, A. and Wang, C.C. (2008)
771   Identification of a novel chromosomal passenger complex and its unique localization
772   during cytokinesis in *Trypanosoma brucei*. *PLoS One*, **3**, e2354.
773   41.   Li, Z. and Wang, C.C. (2006) Changing roles of aurora-B kinase in two life cycle
774   stages of *Trypanosoma brucei*. *Eukaryot Cell*, **5**, 1026-1035.
775   42.   Jones, N.G., Thomas, E.B., Brown, E., Dickens, N.J., Hammarton, T.C. and Mottram,
776   J.C. (2014) Regulators of *Trypanosoma brucei* cell cycle progression and
777   differentiation identified using a kinome-wide RNAi screen. *PLoS Pathog*, **10**,
778   e1003886.
779   43.   Hu, H., Liu, Y., Zhou, Q., Siegel, S. and Li, Z. (2015) The centriole cartwheel protein
780   SAS-6 in *Trypanosoma brucei* is required for probasal body biogenesis and flagellum
781   assembly. *Eukaryot Cell*, **14**, 898-907.
782   44.   Zhang, X., An, T., Pham, K.T.M., Lun, Z.R. and Li, Z. (2019) Functional analyses of
783   cytokinesis regulators in bloodstream stage *Trypanosoma brucei* parasites identify
784   functions and regulations specific to the life cycle stage. *mSphere*, **4**.
785   45.   Dean, S., Moreira-Leite, F., Varga, V. and Gull, K. (2016) Cilium transition zone
786   proteome reveals compartmentalization and differential dynamics of ciliopathy
787   complexes. *Proc Natl Acad Sci U S A*, **113**, E5135-5143.
788   46.   Dang, H.Q., Zhou, Q., Rowlett, V.W., Hu, H., Lee, K.J., Margolin, W. and Li, Z.
789   (2017) Proximity interactions among basal body components in *Trypanosoma brucei*
790   identify novel regulators of basal body biogenesis and inheritance. *mBio*, **8**.
791   47.   An, T. and Li, Z. (2018) An orphan kinesin controls trypanosome morphology
792   transitions by targeting FLAM3 to the flagellum. *PLoS Pathog*, **14**, e1007101.
793   48.   Hu, L., Hu, H. and Li, Z. (2012) A kinetoplastid-specific kinesin is required for
794   cytokinesis and for maintenance of cell morphology in *Trypanosoma brucei*. *Mol
795   Microbiol*, **83**, 565-578.
796   49.   DeGrasse, J.A., DuBois, K.N., Devos, D., Siegel, T.N., Sali, A., Field, M.C., Rout,
797   M.P. and Chait, B.T. (2009) Evidence for a shared nuclear pore complex architecture
798   that is conserved from the last common eukaryotic ancestor. *Mol Cell Proteomics*, **8**,
799   2119-2130.
800   50.   Benz, C., Clucas, C., Mottram, J.C. and Hammarton, T.C. (2012) Cytokinesis in
801   bloodstream stage *Trypanosoma brucei* requires a family of katanins and spastin.
802   *PLoS One*, **7**, e30367.
803   51.   Ralston, K.S. and Hill, K.L. (2006) Trypanin, a component of the flagellar dynein
804   regulatory complex, is essential in bloodstream form African trypanosomes. *PLoS
805   Pathog*, **2**, e101.
806   52.   May, S.F., Peacock, L., Almeida Costa, C.I., Gibson, W.C., Tetley, L., Robinson,
807   D.R. and Hammarton, T.C. (2012) The *Trypanosoma brucei* AIR9-like protein is
808   cytoskeleton-associated and is required for nucleus positioning and accurate cleavage
809   furrow placement. *Mol Microbiol*, **84**, 77-92.

810 53. Portman, N. and Gull, K. (2014) Identification of paralogous life-cycle stage specific
811 cytoskeletal proteins in the parasite *Trypanosoma brucei*. *PLoS One*, **9**, e106777.
812 54. Dostalova, A., Kaser, S., Cristodero, M. and Schimanski, B. (2013) The nuclear
813 mRNA export receptor Mex67-Mtr2 of *Trypanosoma brucei* contains a unique and
814 essential zinc finger motif. *Mol Microbiol*, **88**, 728-739.
815 55. Marques, C.A., Tiengwe, C., Lemgruber, L., Damasceno, J.D., Scott, A., Paape, D.,
816 Marcello, L. and McCulloch, R. (2016) Diverged composition and regulation of the
817 *Trypanosoma brucei* origin recognition complex that mediates DNA replication
818 initiation. *Nucleic Acids Res*, **44**, 4763-4784.
819 56. Gluenz, E., Sharma, R., Carrington, M. and Gull, K. (2008) Functional
820 characterization of cohesin subunit SCC1 in *Trypanosoma brucei* and dissection of
821 mutant phenotypes in two life cycle stages. *Mol Microbiol*, **69**, 666-680.
822 57. Dang, H.Q. and Li, Z. (2011) The Cdc45.Mcm2-7.GINS protein complex in
823 trypanosomes regulates DNA replication and interacts with two Orc1-like proteins in
824 the origin recognition complex. *J Biol Chem*, **286**, 32424-32435.
825 58. Shyh-Chang, N., Locasale, J.W., Lyssiotis, C.A., Zheng, Y., Teo, R.Y.,
826 Ratanasirintrawoot, S., Zhang, J., Onder, T., Unternaehrer, J.J., Zhu, H. *et al.* (2013)
827 Influence of threonine metabolism on S-adenosylmethionine and histone methylation.
828 *Science*, **339**, 222-226.
829 59. Peacock, L., Bailey, M., Carrington, M. and Gibson, W. (2014) Meiosis and haploid
830 gametes in the pathogen *Trypanosoma brucei*. *Curr Biol*, **24**, 181-186.
831 60. Ammerman, M.L., Downey, K.M., Hashimi, H., Fisk, J.C., Tomasello, D.L.,
832 Faktorova, D., Kafkova, L., King, T., Lukes, J. and Read, L.K. (2012) Architecture of
833 the trypanosome RNA editing accessory complex, MRB1. *Nucleic Acids Res*, **40**,
834 5637-5650.
835 61. Zikova, A., Schnaufer, A., Dalley, R.A., Panigrahi, A.K. and Stuart, K.D. (2009) The
836 F0F1-ATP synthase complex contains novel subunits and is essential for procyclic
837 *Trypanosoma brucei*. *PLoS Pathog*, **5**, e1000436.
838 62. Tudzarova, S., Colombo, S.L., Stoeber, K., Carcamo, S., Williams, G.H. and
839 Moncada, S. (2011) Two ubiquitin ligases, APC/C-Cdh1 and SKP1-CUL1-F (SCF)-
840 beta-TrCP, sequentially regulate glycolysis during the cell cycle. *Proc Natl Acad Sci
841 U S A*, **108**, 5278-5283.
842 63. Hu, J.W., Sun, P., Zhang, D.X., Xiong, W.J. and Mi, J. (2014) Hexokinase 2 regulates
843 G1/S checkpoint through CDK2 in cancer-associated fibroblasts. *Cell Signal*, **26**,
844 2210-2216.
845 64. Morgan, G.A., Hamilton, E.A. and Black, S.J. (1996) The requirements for G1
846 checkpoint progression of *Trypanosoma brucei* S 427 clone 1. *Mol Biochem
847 Parasitol*, **78**, 195-207.
848 65. Silvester, E., McWilliam, K.R. and Matthews, K.R. (2017) The cytological events and
849 molecular control of life cycle development of *Trypanosoma brucei* in the
850 mammalian bloodstream. *Pathogens*, **6**.
851 66. Qiu, Y., Milanes, J.E., Jones, J.A., Noorai, R.E., Shankar, V. and Morris, J.C. (2018)
852 Glucose signaling is important for nutrient adaptation during differentiation of
853 pleomorphic African trypanosomes. *mSphere*, **3**.
854 67. Naguleswaran, A., Doiron, N. and Roditi, I. (2018) RNA-Seq analysis validates the
855 use of culture-derived *Trypanosoma brucei* and provides new markers for mammalian
856 and insect life-cycle stages. *BMC Genomics*, **19**, 227.
857 68. Kalucka, J., Missiaen, R., Georgiadou, M., Schoors, S., Lange, C., De Bock, K.,
858 Dewerchin, M. and Carmeliet, P. (2015) Metabolic control of the cell cycle. *Cell
859 Cycle*, **14**, 3379-3388.

69. Burgers, P.M.J. and Kunkel, T.A. (2017) Eukaryotic DNA replication fork. *Annu Rev Biochem*, **86**, 417-438.

70. Hayashi, H. and Akiyoshi, B. (2018) Degradation of cyclin B is critical for nuclear division in *Trypanosoma brucei*. *Biol Open*, **7**.

71. Benz, C. and Clayton, C.E. (2007) The F-box protein CFB2 is required for cytokinesis of bloodstream-form *Trypanosoma brucei*. *Mol Biochem Parasitol*, **156**, 217-224.

72. Van Hellemond, J.J. and Mottram, J.C. (2000) The CYC3 gene of *Trypanosoma brucei* encodes a cyclin with a short half-life. *Mol Biochem Parasitol*, **111**, 275-282.

73. Ishii, M. and Akiyoshi, B. (2020) Characterization of unconventional kinetochore kinases KKT10 and KKT19 in *Trypanosoma brucei*. *J Cell Sci*, **133**.

74. Saldivia, M., Fang, E., Ma, X., Myburgh, E., Carnielli, J.B.T., Bower-Lepts, C., Brown, E., Ritchie, R., Lakshminarayana, S.B., Chen, Y.-L. *et al.* (2020) Targeting the trypanosome kinetochore with CLK1 protein kinase inhibitors. *Nature Microbiology*.

75. Hinnebusch, A.G. (2006) eIF3: a versatile scaffold for translation initiation complexes. *Trends Biochem Sci*, **31**, 553-562.

76. Mugo, E. and Clayton, C. (2017) Expression of the RNA-binding protein RBP10 promotes the bloodstream-form differentiation state in *Trypanosoma brucei*. *PLoS Pathog*, **13**, e1006560.

77. Hammarton, T.C., Kramer, S., Tetley, L., Boshart, M. and Mottram, J.C. (2007) *Trypanosoma brucei* Polo-like kinase is essential for basal body duplication, kDNA segregation and cytokinesis. *Mol Microbiol*, **65**, 1229-1248.

78. Kumar, P. and Wang, C.C. (2006) Dissociation of cytokinesis initiation from mitotic control in a eukaryote. *Eukaryot Cell*, **5**, 92-102.

79. McAllaster, M.R., Ikeda, K.N., Lozano-Nunez, A., Anrather, D., Unterwurzacher, V., Gossenreiter, T., Perry, J.A., Crickley, R., Mercadante, C.J., Vaughan, S. *et al.* (2015) Proteomic identification of novel cytoskeletal proteins associated with TbPLK, an essential regulator of cell morphogenesis in *Trypanosoma brucei*. *Mol Biol Cell*, **26**, 3013-3029.

80. Tu, X. and Wang, C.C. (2004) The involvement of two cdc2-related kinases (CRKs) in *Trypanosoma brucei* cell cycle regulation and the distinctive stage-specific phenotypes caused by CRK3 depletion. *J Biol Chem*, **279**, 20519-20528.

81. Menzies, S.K., Tulloch, L.B., Florence, G.J. and Smith, T.K. (2018) The trypanosome alternative oxidase: a potential drug target? *Parasitology*, **145**, 175-183.

82. Li, Z. and Wang, C.C. (2003) 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.

83. Wei, Y., Hu, H., Lun, Z.R. and Li, Z. (2014) Centrin3 in trypanosomes maintains the stability of a flagellar inner-arm dynein for cell motility. *Nat Commun*, **5**, 4060.

84. Denninger, V. and Rudenko, G. (2014) FACT plays a major role in histone dynamics affecting *VSG* expression site control in *Trypanosoma brucei*. *Mol Microbiol*, **94**, 945-962.

85. Prendergast, L., Muller, S., Liu, Y., Huang, H., Dingli, F., Loew, D., Vassias, I., Patel, D.J., Sullivan, K.F. and Almouzni, G. (2016) The CENP-T/-W complex is a binding partner of the histone chaperone FACT. *Genes Dev*, **30**, 1313-1326.

86. Haanstra, J.R., Gerding, A., Dolga, A.M., Sorgdrager, F.J.H., Buist-Homan, M., du Toit, F., Faber, K.N., Holzhutter, H.G., Szoor, B., Matthews, K.R. *et al.* (2017) Targeting pathogen metabolism without collateral damage to the host. *Sci Rep*, **7**, 40406.

909   87.   Khare, S., Nagle, A.S., Biggart, A., Lai, Y.H., Liang, F., Davis, L.C., Barnes, S.W.,
910   Mathison, C.J., Myburgh, E., Gao, M.Y. *et al.* (2016) Proteasome inhibition for
911   treatment of leishmaniasis, Chagas disease and sleeping sickness. *Nature*, **537**, 229-
912   233.
913   88.   Torrie, L.S., Zuccotto, F., Robinson, D.A., Gray, D.W., Gilbert, I.H. and De Rycker,
914   M. (2019) Identification of inhibitors of an unconventional *Trypanosoma brucei*
915   kinetochore kinase. *PLoS One*, **14**, e0217828.
916   89.   Urbaniak, M.D., Mathieson, T., Bantscheff, M., Eberhard, D., Grimaldi, R., Miranda-
917   Saavedra, D., Wyatt, P., Ferguson, M.A., Frearson, J. and Drewes, G. (2012)
918   Chemical proteomic analysis reveals the drugability of the kinome of *Trypanosoma*
919   *brucei*. *ACS Chem Biol*, **7**, 1858-1865.
920   90.   Halliday, C., Billington, K., Wang, Z., Madden, R., Dean, S., Sunter, J.D. and
921   Wheeler, R.J. (2019) Cellular landmarks of *Trypanosoma brucei* and *Leishmania*
922   *mexicana*. *Mol Biochem Parasitol*, **230**, 24-36.
923   91.   El-Sayed, N.M., Myler, P.J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G.,
924   Caler, E., Renauld, H., Worthey, E.A., Hertz-Fowler, C. *et al.* (2005) Comparative
925   genomics of trypanosomatid parasitic protozoa. *Science*, **309**, 404-409.
926   92.   Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2.
927   *Nat Methods*, **9**, 357-359.
928   93.   Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G.,
929   Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009) The
930   Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.
931   94.   Okonechnikov, K., Conesa, A. and Garcia-Alcalde, F. (2016) Qualimap 2: advanced
932   multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, **32**,
933   292-294.
934   95.   Ewels, P., Magnusson, M., Lundin, S. and Kaller, M. (2016) MultiQC: summarize
935   analysis results for multiple tools and samples in a single report. *Bioinformatics*, **32**,
936   3047-3048.
937   96.   Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for
938   comparing genomic features. *Bioinformatics*, **26**, 841-842.
939   97.   Egorov, A.A., Sakharova, E.A., Anisimova, A.S., Dmitriev, S.E., Gladyshev, V.N.
940   and Kulakovskiy, I.V. (2019) svist4get: a simple visualization tool for genomic tracks
941   from sequencing experiments. *BMC Bioinformatics*, **20**, 113.
942   98.   Liao, Y., Smyth, G.K. and Shi, W. (2014) featureCounts: an efficient general purpose
943   program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923-
944   930.
945   99.   Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau,
946   D., Burovski, E., Peterson, P., Weckesser, W., Bright, J. *et al.* (2020) SciPy 1.0:
947   fundamental algorithms for scientific computing in Python. *Nat Methods*, **17**, 261-
948   272.
949
950

951   **Figure Legends**

952

953   **Figure 1.** A genome-wide conditional knockdown screen for cell cycle progression defects.

954   (**A**) Schematic representation of the bloodstream form *T. brucei* cell cycle, also showing

955   aberrant sub-2C and >4C phenotypes. (**B**) The schematic illustrates the RIT-seq screen;

956   massive parallel induction of RNAi followed by flow cytometry and RIT-seq,

957   allowing for reconstruction of cell cycle profiles, using mapped reads from each knockdown.

958   Each read-mapping profile encompasses the gene of interest and associated untranslated

959   regions present in the cognate mRNA. The library data represents the uninduced and

960   unsorted population. GeneIDs, Tb927.7.3160 for example, are shown without the common

961   'Tb927.' component.

962

963   **Figure 2.** Validation and identification of >1000 candidates linked to cell cycle defects. (**A**)

964   The plot on the left shows knockdowns overrepresented in the >4C experiment in red; those

965   with >1.5-fold the sum of reads in the G_1, S phase and G_2M samples combined. The read-

966   mapping profile and read-counts for α/β-tubulin are shown to the right. (**B**) The plot on the

967   left shows knockdowns overrepresented in the sub-2C experiment in orange; those with

968   >1.5-fold the sum of reads in the G_1, S phase and G_2M samples combined. The read-

969   mapping profile and read-counts for DOT1A are shown to the right. (**C**) The plots on the left

970   shows knockdowns overrepresented in the G_1, S phase and G_2M experiments in purple,

971   green and blue, respectively; those that were >25% overrepresented in each category.

972   Read-mapping profiles and relative read-counts for example hits are shown to the right.

973   PCNA, proliferating cell nuclear antigen; PPL2, PrimPol-like 2. (**D**) The Venn diagram shows

974   the distribution of knockdowns overrepresented in each arm of the screen.

975

976   **Figure 3.** Cytokinesis defects associated with endoreduplication. (**A**) The bar-graph shows

977   enriched Gene Ontology terms in the >4C overrepresented dataset. (**B**) The violin plot

978   shows relative >4C read-counts for cohorts of genes and reflects data distribution. Open

979   circles indicate median values and the vertical bars indicate 95% confidence intervals.

980   Significantly overrepresented cohorts are indicated in red. (**C**) The plots show

981   overrepresentation of T-complex, dynein and intraflagellar transport (IFT) factors in red in the

982   >4C experiment. (**D**) The heatmaps show relative representation in all five sorted pools for

983   the above and additional cohorts of knockdowns; blue, most overrepresented. (**E**) Example

984   read-mapping profiles for hits overrepresented in the >4C pool. (**F**) Example read-mapping

985   profiles for ciliopathy-associated hits overrepresented in the >4C pool. CMF, Component of

986   Motile Flagella; CFAP, Cilia and Flagella Associated Protein.

987

988   **Figure 4.** Defects producing sub-diploid cells. Read-mapping profiles and read-counts for

989   example hits overrepresented in the <2C experiment. HPC, Hypothetical Protein,

990   Conserved.

991

992   **Figure 5.** A profile of G_1, S phase and G_2/mitosis defects. (**A**) The RadViz plot shows

993   knockdowns that registered >25% overrepresented read-counts in the G_1, S phase or G_2M

994   categories. (**B**) The bar-graphs shows enriched Gene Ontology terms in the G_1, S phase or

995   G_2M overrepresented datasets. (**C**) The violin plots show relative G_1, S phase or G_2M read-

996   counts for cohorts of genes and reflects data distribution. Open circles indicate median

997   values and the vertical bars indicate 95% confidence intervals. Overrepresented cohorts are

998   indicated in purple, green and blue, respectively. (**D**) The heatmaps show relative

999   representation in all five sorted pools for the above and additional cohorts of knockdowns;

1000   blue, most overrepresented.

1001

1002   **Figure 6.** Protein complexes and pathways associated with G_1, S phase and G_2/mitosis

1003   defects. (**A**) The RadViz plot shows glycolytic enzyme knockdowns. Those that registered

1004   >25% overrepresented read-counts in the G_1 category are indicated, purple. Black data-

1005   points indicate other genes from each cohort. Grey data-points indicate all other genes. The

1006   read-mapping profiles and relative read-counts in the lower panel show example hits. (**B**) As

1007   in a but for DNA replication initiation factor knockdowns that registered >25%

1008   overrepresented read-counts in the S phase category, indicated in green. (**C**) As in a but for

1009   proteasome component knockdowns that registered >25% overrepresented read-counts in

1010   the G_2M category, indicated in blue. (**D**) As in a but for kinetochore component knockdowns

1011   that registered >25% overrepresented read-counts in the S phase or G_2M categories,

1012   indicated in green or blue, respectively.

1013

1014   **Figure 7.** RBPs kinases and hypothetical proteins associated with gap and S phase defects.

1015   (**A**) The RadViz plot shows mRNA binding protein knockdowns (RBPs). Those that

1016   registered >25% overrepresented read-counts in the G_1, S phase or G_2M categories are

1017   indicated, in purple, green and blue, respectively. The read-mapping profiles and relative

1018   read-counts in the lower panels show example hits. (**B**) As in a but for protein kinase

1019   knockdowns. (**C**) As in a but for hypothetical (conserved) protein knockdowns.

29

1020   **Supplementary Figure Legends**

1021

1022   *Figure 1—figure supplement 1.* Induced library sorting and RNAi target amplification. (**A**)

1023   BD Influx™ cell sorter BD FACSort™ software workspace data for the sorting session

1024   showing the gates and population frequencies. (**B**) Flow cytometry quality control data for

1025   the sorted samples. Three types of graphs are shown: SSC-Area x FSC-Area (cell

1026   morphology); PI-Area x PI-Width (cells stained with PI; the gate excludes cell aggregates);

1027   Modal x PI-Area (cells within the gate set in the PI-Area x PI-Width plot). Top row –

1028   unsorted; second row – G_1; third row – S; fourth row – G_2M; bottom row – overlay of the

1029   sorted and unsorted samples. 2C and 4C refer to unduplicated (diploid genome) and

1030   duplicated DNA content, respectively. (**C**) PCR amplification of each sorted sample with Lib3

1031   primers, prior to sequencing.

1032

1033   *Figure 1—figure supplement 2.* RIT-seq data comparing unsorted libraries. (**A**) The plot

1034   shows read-counts for 7,204 genes and for uninduced and 24 h induced samples. Reads for

1035   only 0.6% of genes dropped by >3-fold following 24 h of knockdown. (**B**) The violin plots

1036   show relative read-counts for cohorts of genes and reflect data distribution. Open circles

1037   indicate median values and the vertical bars indicate 95% confidence intervals. Read-counts

1038   remain relatively high after 24 h knockdown in the current study, when compared to read-

1039   counts after 72 h knockdown in a prior RIT-seq study.

1040

1041   *Figure 2—figure supplement 1.* RIT-seq data mapped to *T. brucei* chromosomes; for 7,204

1042   genes and five experiments = 36,020 data-points. Light and dark green indicate polycistronic

1043   transcription in the forward and reverse directions, respectively, on each chromosome. The

1044   heat maps (dark-blue to yellow) indicate rank enrichment for knockdowns in each cell cycle

1045   phase; most enriched are dark-blue. The coloured data-points below indicate those 'hits'

1046   enriched in each cell cycle phase and used for much of the analysis reported here; 1,158 or

1047   16.1% of genes. See the text for more details.

1048

1049   *Figure 2—figure supplement 2.* Online RIT-seq data visualization (https://tryp-

1050   cycle.onrender.com). In the radial visualization, experimental-points are hours on the clock-

1051   face (i.e. related to the angle of the polar coordinate system). The orthogonal axis (i.e. the

1052   distance) relates to the relative read-counts across the five experiments. The table on the

1053   right shows: Geneid, gene identification number; relative abundance of reads in each sorted

1054   sample; desc, gene description; class, the experiment where the gene shows maximum

1055   abundance; selected, a binary tag where '1' indicates genes in the radial visualization. Gene

30

1056 coverage images are displayed when hovering over dots on the radial visualization or over

1057 table rows.

1058   Data transformation to aid visualization: Normalised TPM values from

1059 *Supplementary file 1* were used, except the <2C and >4C values were divided by the sum

1060 of TPM values from all five sorted samples. we elevated the values to the power of 2.1 to

1061 maximize differences. We then normalized the values raw wise for each gene, by dividing

1062 values by the maximum. The transformed data was then fed to the radial visualization

1063 algorithm implemented in D3.js (https://github.com/d3/d3); code and web page were adapted

1064 from the repository at https://github.com/WYanChao/RadViz.

1065

1066 *Figure 3—figure supplement 1.* Cytokinesis defects associated with endoreduplication.

1067 One hundred example RIT-seq cell cycle profiles are shown for hits overrepresented in the

1068 >4C pool. Page 1 shows the heatmaps indicating relative representation in all five sorted

1069 pools; blue, most overrepresented. Subsequent pages show read-mapping profiles for each

1070 gene; see Figure 1B for further details.

1071

1072 *Figure 7—figure supplement 1.* Knockdowns associated with gap and S phase defects.

1073 Seventy-eight example RIT-seq cell cycle profiles are shown for hits overrepresented in the

1074 G_1, S phase and G_2M experiments. Page 1 shows the heatmaps indicating relative

1075 representation in all five sorted pools; blue, most overrepresented. Subsequent pages show

1076 read-mapping profiles for each gene; see Figure 1B for further details.

1077

1078 *Supplementary file 1.* RIT-seq digital data. The Excel file reports the total fragment counts

1079 for the uninduced (column C) and induced (column D) RNAi libraries, for the five sorted

1080 samples (columns E-I), normalised read-counts for the sorted samples (columns J-N) and

1081 normalised barcoded read-counts for the G1 (column O), S (column P) and G2M (column Q)

1082 samples. The final column (R) indicates Figure numbers for genes shown in the manuscript.

1083 TPM, Transcripts Per kilobase Million. Coloured values in column J-Q indicates enrichment

1084 in those samples; >1.5-fold the sum of normalised TPM in the G1+S+G2M samples in

1085 columns J and N, >41.66% (>25% above the mean) in columns K-M and also >40% (>20%

1086 above the mean) in columns O-Q. Genes considered in Figure 2 and used to generate Gene

1087 Ontology profiles surpass both the 25% total reads and 20% barcoded reads thresholds i.e.

1088 coloured in columns O-Q. Data are presented for 7,204 genes, which is 98% of the non-

1089 redundant gene set; all genes register >99 total reads across the five sorted samples.
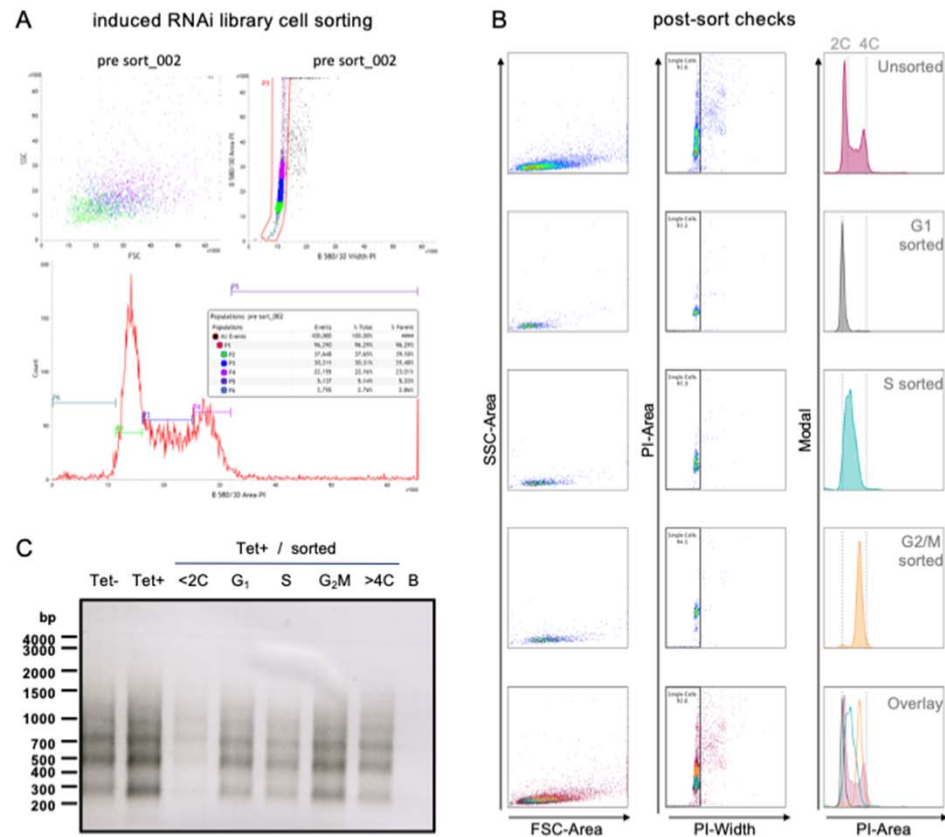
1090

Figure 1—figure supplement 1. Induced library sorting and RNAi target amplification. (**A**) BD Influx™ cell sorter BD FACSort™ software workspace data for the sorting session showing the gates and population frequencies. (**B**) Flow cytometry quality control data for the sorted samples. Three types of graphs are shown: SSC-Area x FSC-Area (cell morphology); PI-Area x PI-Width (cells stained with PI; the gate excludes cell aggregates); Modal x PI-Area (cells within the gate set in the PI-Area x PI-Width plot). Top row – unsorted; second row – G_1; third row – S; fourth row – G_2M; bottom row – overlay of the sorted and unsorted samples. 2C and 4C refer to unduplicated (diploid genome) and duplicated DNA content, respectively. (**C**) PCR amplification of each sorted sample with Lib3 primers, prior to sequencing; B, blank.
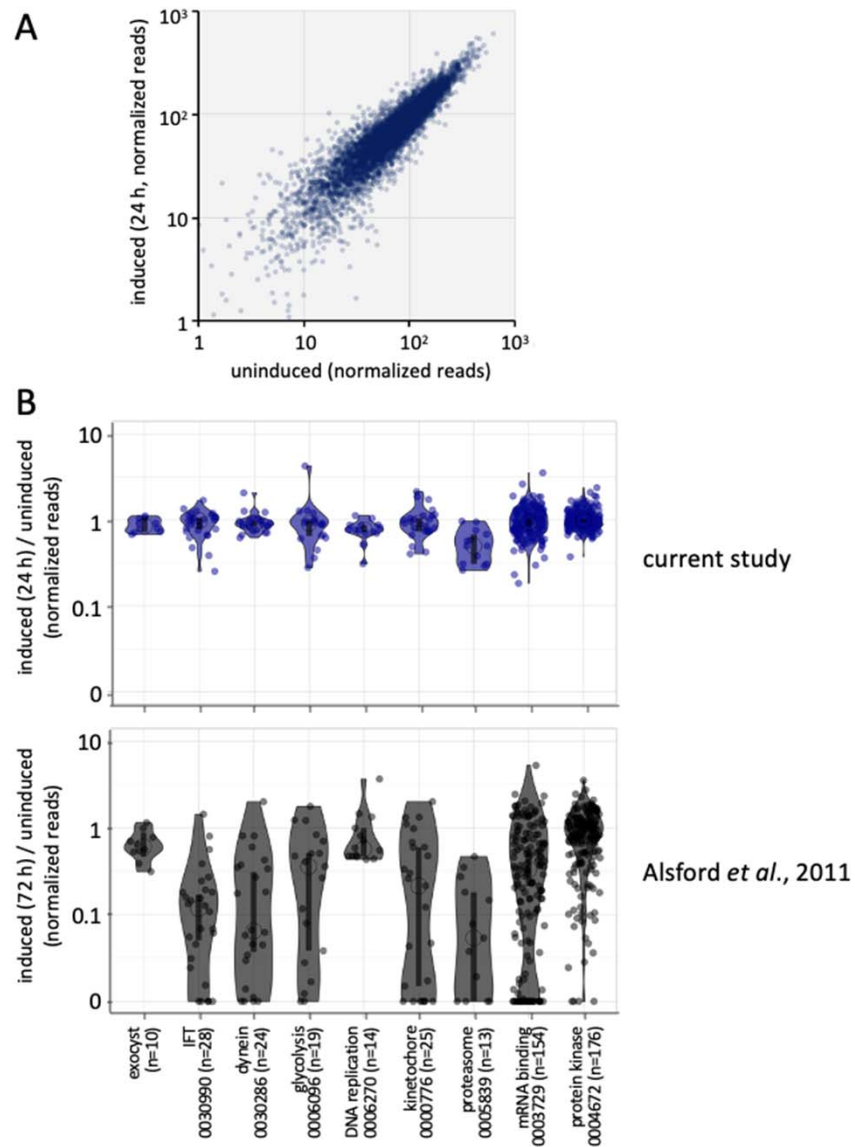
Figure 1—figure supplement 2. RIT-seq data comparing unsorted libraries. (**A**) The plot shows read-counts for 7,204 genes and for uninduced and 24 h induced samples. Reads for only 0.6% of genes dropped by >3-fold following 24 h of knockdown. (**B**) The violin plots show relative read-counts for cohorts of genes and reflect data distribution. Open circles indicate median values and the vertical bars indicate 95% confidence intervals. Read-counts remain relatively high after 24 h knockdown in the current study, when compared to read-counts after 72 h knockdown in a prior RIT-seq study.
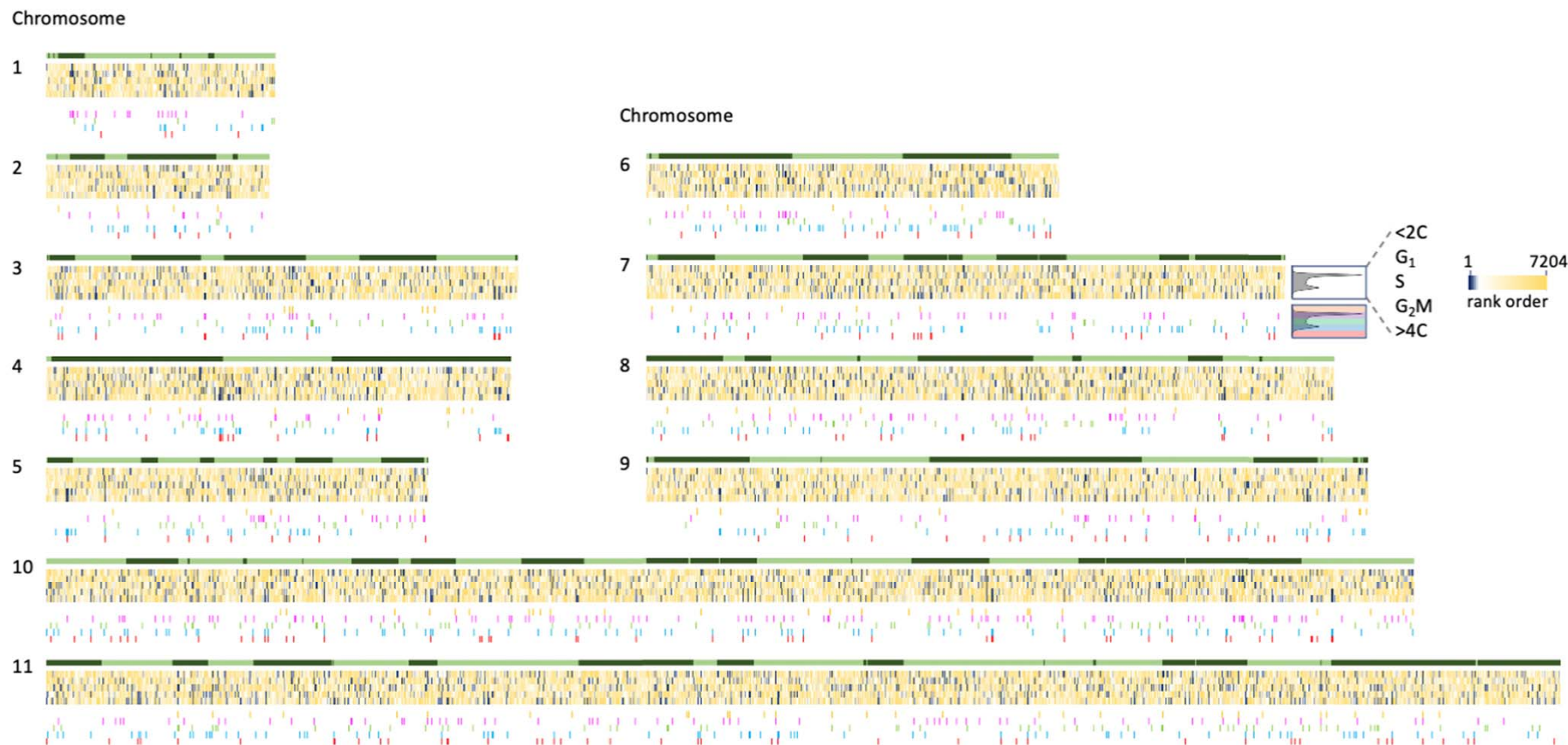
Figure 2—figure supplement 1. RIT-seq data mapped to *T. brucei* chromosomes; for 7,204 genes and five experiments = 36,020 data-points. Light and dark green indicate polycistronic transcription in the forward and reverse directions, respectively, on each chromosome. The heat maps (dark-blue to yellow) indicate rank enrichment for knockdowns in each cell cycle phase; most enriched are dark-blue. The coloured data-points below indicate those 'hits' enriched in each cell cycle phase and used for much of the analysis reported here; 1,158 or 16.1% of genes. See the text for more details.
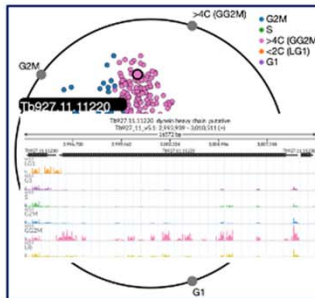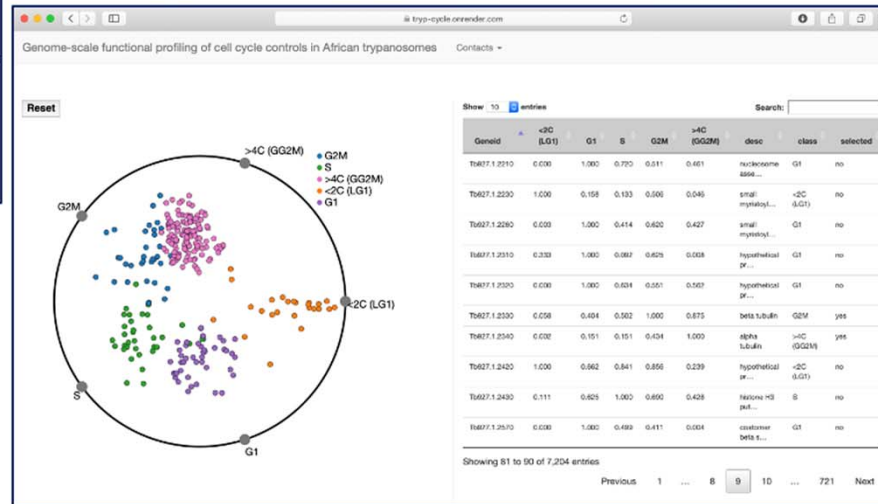
Figure 2—figure supplement 2. Online RIT-seq data visualization (https://tryp-cycle.onrender.com). In the radial visualization, experimental-points are hours on the clock-face (i.e. related to the angle of the polar coordinate system). The orthogonal axis (i.e. the distance) relates to the relative read-counts across the five experiments. The table on the right shows: Geneid, gene identification number; relative abundance of reads in each sorted sample; desc, gene description; class, the experiment where the gene shows maximum abundance; selected, a binary tag where 'yes' indicates genes in the radial visualization. Gene coverage images are displayed when hovering over dots on the radial visualization or over table rows.

Data transformation to aid visualization: Normalised TPM values from *Supplementary file 1* were used, except the <2C and >4C values were divided by the sum of TPM values from all five sorted samples. we elevated the values to the power of 2.1 to maximize differences. We then normalized the values raw wise for each gene, by dividing values by the maximum. The transformed data was then fed to the radial visualization algorithm implemented in D3.js (https://github.com/d3/d3); code and web page were adapted from the repository at https://github.com/WYanChao/RadViz.
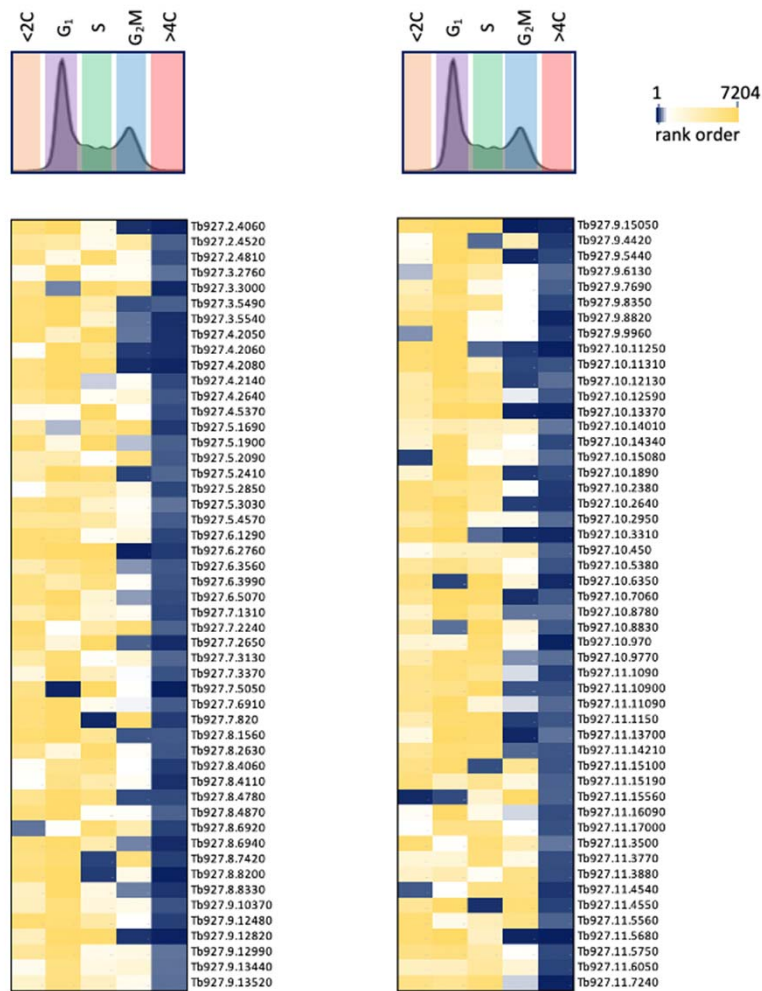
Figure 3—figure supplement 1. Cytokinesis defects associated with endoreduplication. One hundred example RIT-seq cell cycle profiles are shown for hits overrepresented in the >4C pool. Page 1 shows the heatmaps, indicating relative representation in all five sorted pools; blue, most overrepresented. Subsequent pages show read-mapping profiles for each gene; see Figure 1B for further details.

Figure 3 – figure supplement 1 – p2 of 5

flagellar attachment

FAZ8: 4.2060　　FLA3: 5.4570　　FLA2: 8.4060　　FLA5: 10.8830　　FLA3: 8.4110

<2C
G₁
S
G₂M
>4C
Lib

DUF4200

11.4550　　11.15190

actin (related)

actinP: 9.5440　　actinP: 11.3880

basal body

BLD10: 11.13700

nucleoporin

NUP96: 10.7060　　NUP48: 11.4540　　NUP140: 11.11090　　NUP53a: 11.15560

kinesin

KIN2A: 5.2090

CC2D: 4.2080　　CC40: 10.11250　　RP2: 10.14010　　AIR9: 11.17000

KIN-E: 5.2410

CAP51V: 7.2650　　CAP5.5V: 8.8330　　FLAM3: 8.4780　　calmodulin: 9.6130

KIN-C: 8.2630

FLAM5: 9.13440　　FAP59: 8.8200　　TerD domain: 8.7420　　TAX-1: 9.10370　　Tb5.20: 11.15100

Figure 3 – figure supplement 1 – p3 of 5

Figure 3 – figure supplement 1 – p4 of 5
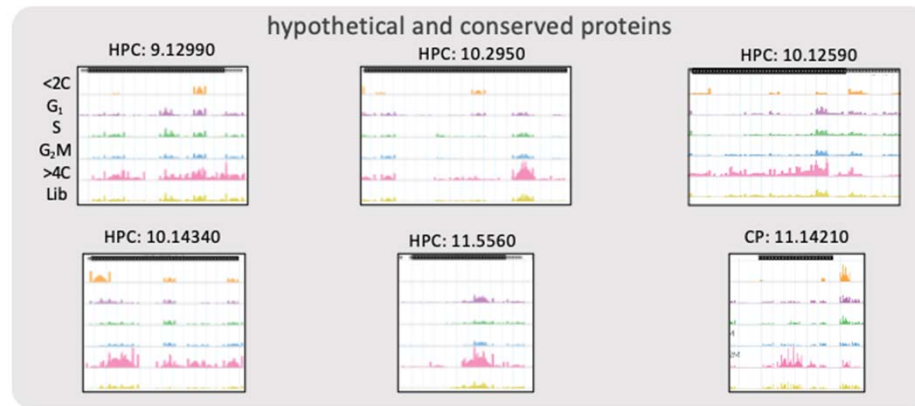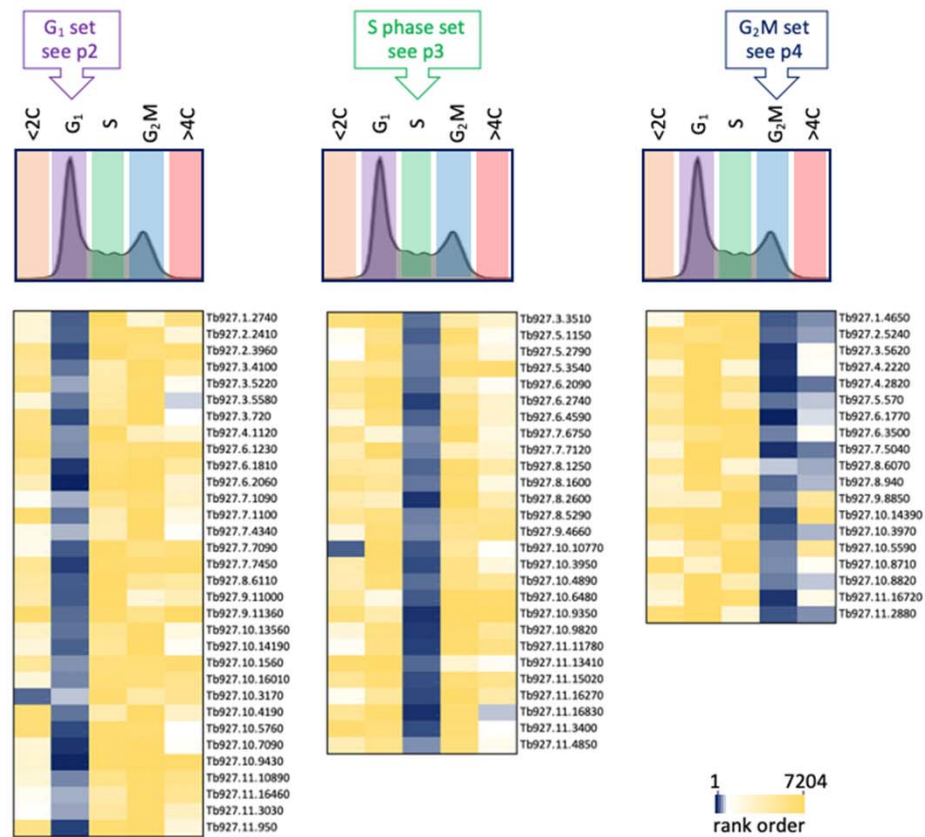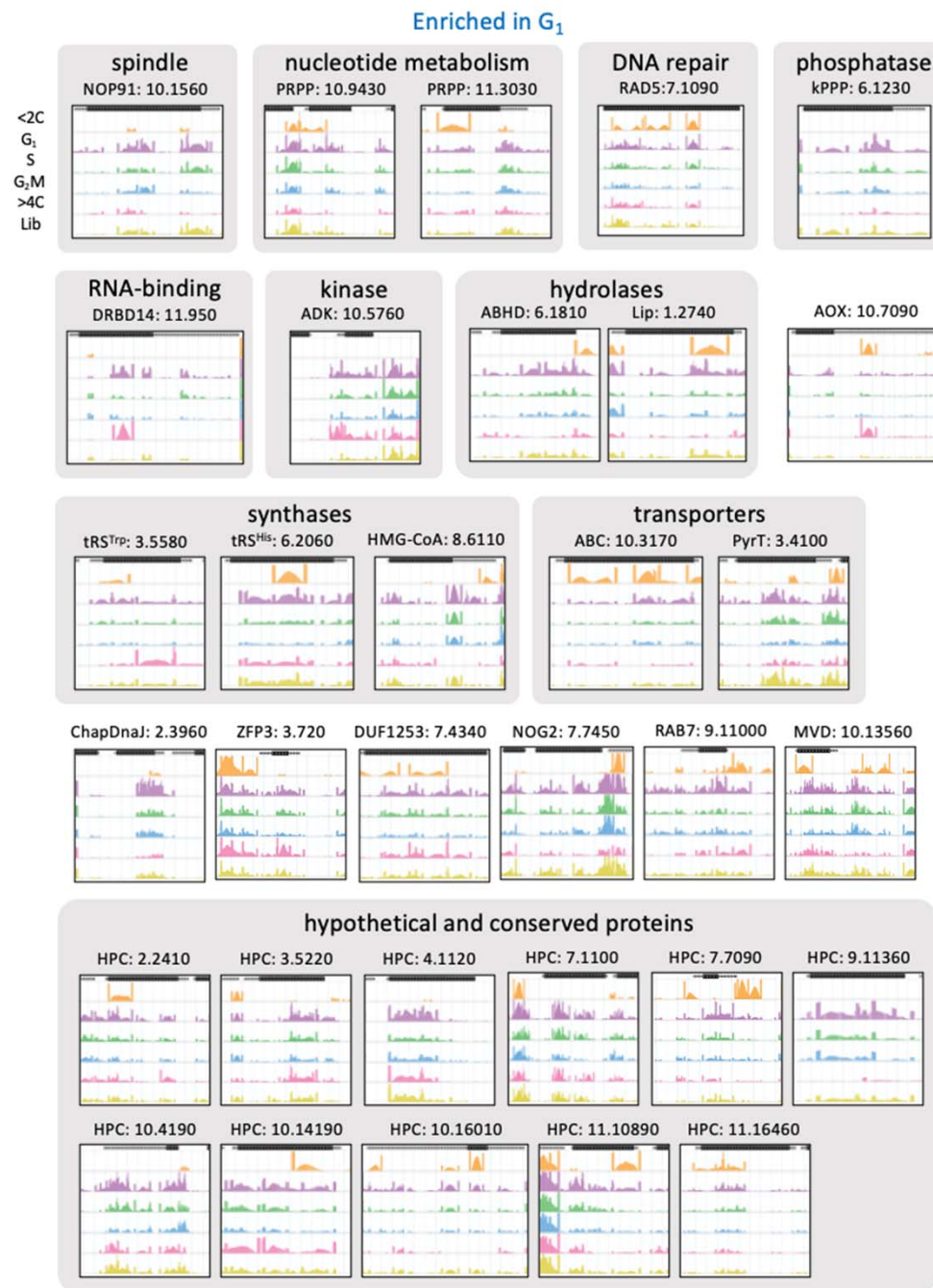
Figure 3 – figure supplement 1 – p5 of 5

Figure 7—figure supplement 1. Knockdowns associated with gap and S phase defects. Seventy-eight example RIT-seq cell cycle profiles are shown for hits overrepresented in the G_1, S phase and G_2M experiments. Page 1 shows the heatmaps indicating relative representation in all five sorted pools; blue, most overrepresented. Subsequent pages show read-mapping profiles for each gene; see Figure 1B for further details.

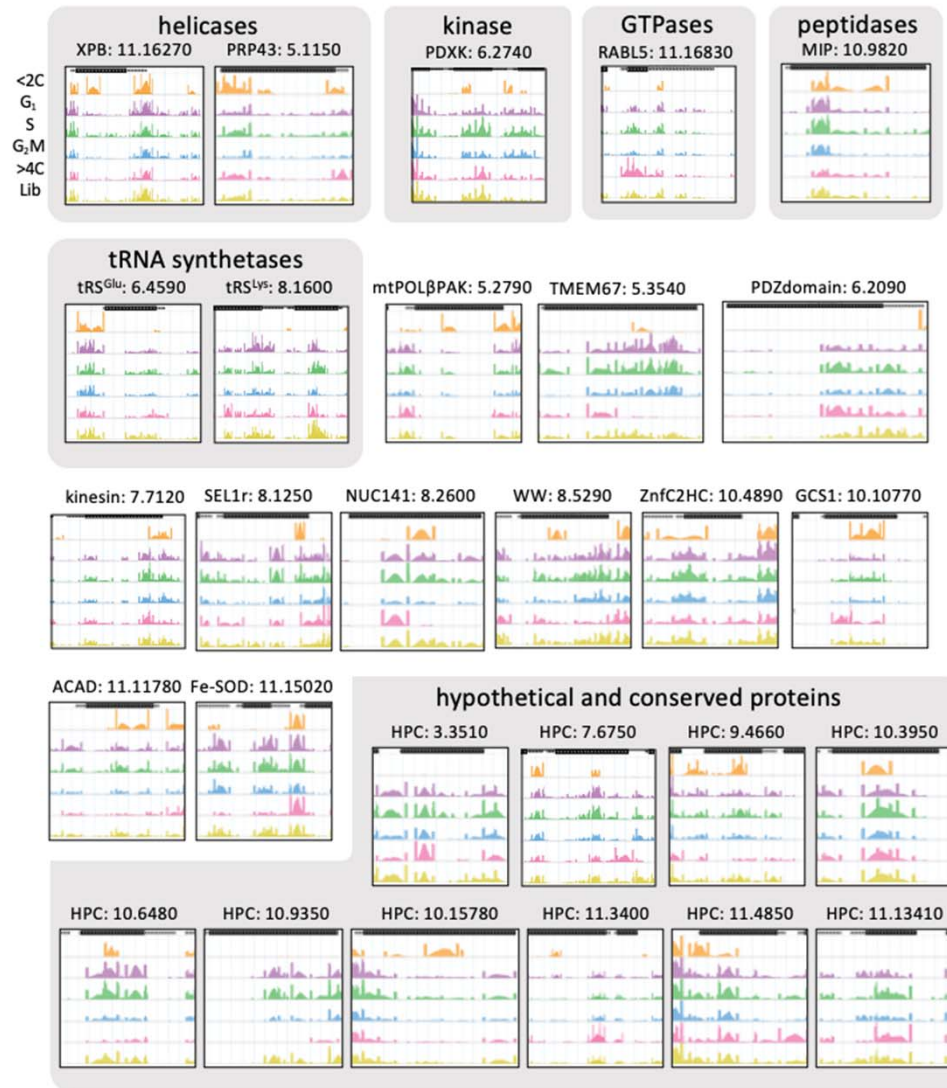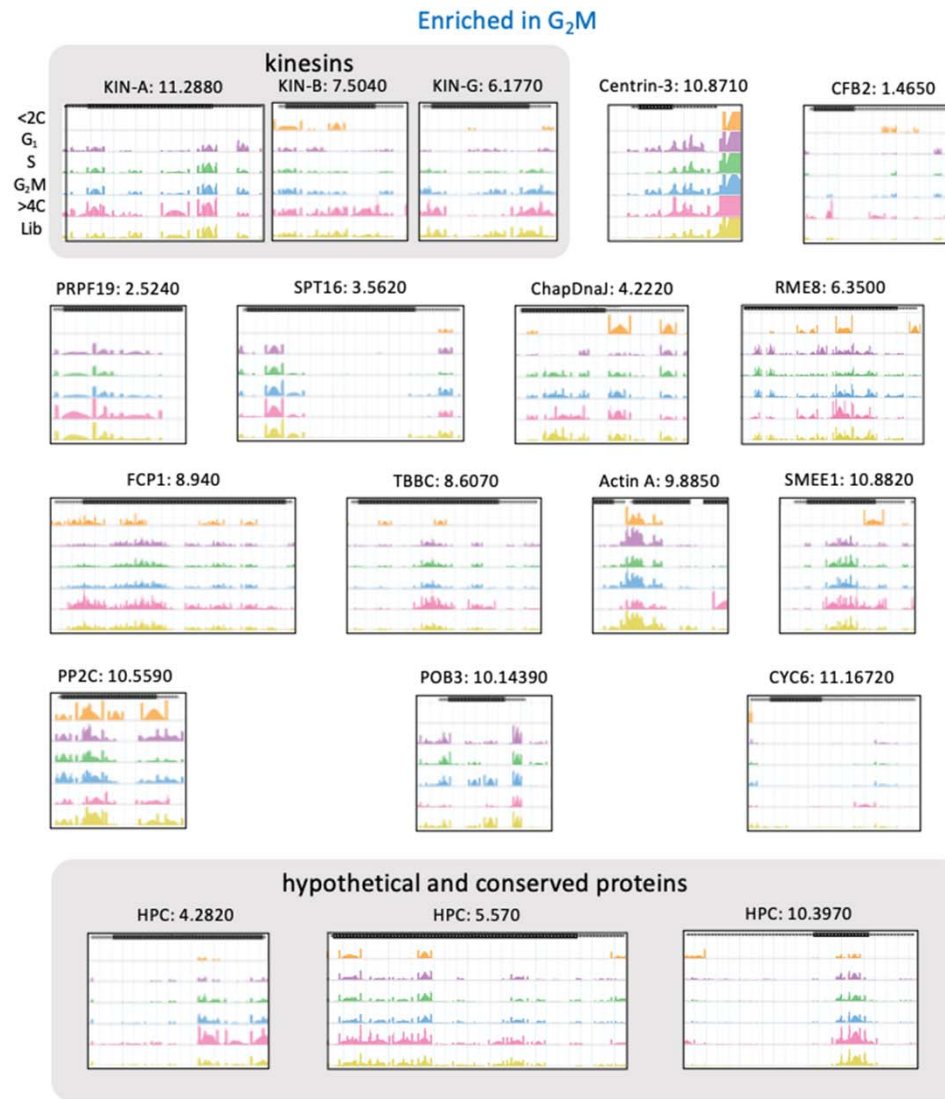*Figure 7 – figure supplement 1 – p2 of 4*

Enriched in S phase

Figure 7 – figure supplement 1 – p3 of 4

Figure 7 – figure supplement 1 – p4 of 4