

1            Transcriptome pyrosequencing of abnormal  
2 phenotypes in *Trypanosoma cruzi* epimastigotes after  
3 ectopic expression of a small zinc finger protein

4            Gaston Westergaard<sup>1,3</sup>, Marc Laverrière<sup>2</sup>, Santiago Revale<sup>3</sup>, Marina  
5 Reinert<sup>1</sup>, Javier G. De Gaudenzi<sup>2</sup>, Adriana Jäger<sup>2</sup>, Martin P. Vazquez<sup>1,3\*</sup>

6

7 1. Departamento de Fisiología y Biología Molecular y Celular, Facultad de  
8 Ciencias Exactas y Naturales, University of Buenos Aires, Argentina

9 2. Instituto de Investigaciones Biotecnológicas (IIB-INTECH), Universidad  
10 Nacional de San Martín, Buenos Aires, Argentina

11 3. Instituto de Agrobiotecnología Rosario (INDEAR), Rosario, Santa Fe,  
12 Argentina

13 \* Corresponding Author

14 martin.vazquez@indear.com

15

16 **Abstract**

17

18 The TcZFPs are a family of small zinc finger proteins harboring WW domains or  
19 Proline rich motifs. In *Trypanosoma brucei*, ZFPs are involved during stage  
20 specific differentiation. TcZFPs interact with each other using the WW domain  
21 (ZFP2 and ZFP3) and the proline rich motif (ZFP1). The tcZFP1b member is  
22 exclusive to *Trypanosoma cruzi* and it is only expressed in trypomastigote stage.  
23 We used a tetracycline inducible vector to express ectopically tcZFP1b in the  
24 epimastigote stage. Upon induction of tcZFP1b, the parasites stopped dividing  
25 completely after five days. Visual inspection showed abnormal distorted-  
26 morphology (monster) cells with multiple flagella and increased DNA contents.  
27 We were interested in investigate global transcription changes occurred during  
28 the generation of this abnormal phenotype. Thus, we performed RNA-seq  
29 transcriptome profiling with a 454 pyrosequencer to analyze the global changes  
30 after ectopic expression of tcZFP1b. The total mRNAs sequenced from induced  
31 and non-induced control epimastigotes showed, after filtering the data, a set of  
32 70 genes having equal or more than 3X fold change upregulation, while 35 genes  
33 showed equal or more than 3X fold downregulation. Interestingly, several trans-  
34 sialidase-like genes and pseudogenes were upregulated along with several  
35 genes in the categories of amino acid catabolism and carbohydrate metabolism.  
36 On the other hand, hypothetical proteins, fatty acid biosynthesis and  
37 mitochondrial functions dominated the group of downregulated genes. Our data  
38 showed that several mRNAs sharing related functions and pathways changed  
39 their levels in a concerted pattern resembling post-transcriptional regulons. We  
40 also found two different motifs in the 3'UTRs of the majority of mRNAs, one for  
41 upregulated and other for downregulated genes

42

## 43 **Introduction**

44

45 Trypanosomes are intriguing organisms in many aspects of their biology. In fact,  
46 they emerge as the paradigm to “the exception of the rule” in the eukaryotic  
47 lineage during the last decade. However, what was once considered “rare and  
48 exceptional” in these organisms was later shown to be more common than  
49 previously thought in the eukaryotic kingdom such as the mRNA trans- splicing or  
50 RNA editing processes [1].

51 Trypanosomes life cycles alternate between a mammalian host and an  
52 invertebrate vector. The adaptation to these two disparate environments requires  
53 a fine-tuning temporal control of gene expression and significant changes in the  
54 expression patterns of several genes. Notably, this regulation occurs almost  
55 entirely at the post-transcriptional level and typical polymerase II promoters are  
56 absent [1,2].

57 In accordance, genome organization and expression in trypanosomatids is  
58 unusual. Transcription of protein coding genes is not regulated at the level of  
59 transcription initiation by RNA polymerase II and their genes are organized into  
60 densely packed units with relatively short intergenic regions and mostly devoid of  
61 introns. In this way, gene expression is polycistronic and controlled mainly by  
62 post-transcriptional processes [1,3,4].

63 Polycistronic units contain unrelated genes that are co-transcriptionally  
64 processed to individual mRNAs by two coupled reactions controlled by the  
65 intergenic regions: 5′trans-splicing and 3′polyadenylation. Thus, the cis-acting  
66 sequences and trans-acting factors controlling the post-transcriptional gene  
67 expression in these parasites are extremely important [5].

68 Several trans-acting factors involved in these processes were identified, such as  
69 RRM (RNA Recognition Motif) containing proteins, PUF proteins and CCCH  
70 containing proteins, comprising a group generally known as RBPs (RNA Binding  
71 Proteins) [6,7,8].

72 RBPs have shown to regulate mRNA abundance in trypanosomes through a  
73 number of cis-acting sequences, most notably AU rich elements (AREs) for  
74 RRM and UGUR core elements for PUF proteins, both type of sequences  
75 present in the 3′end of their target transcripts [9,10].

76 Genome-wide analysis of these proteins in the Trityps (*Leishmania major*,  
77 *Trypanosoma brucei* and *Trypanosoma cruzi*) showed that they contain between  
78 75 (*T. brucei*) and 139 (*T. cruzi*) RRM-type proteins [6].

79 They also contain 10 different PUF proteins and between 48 and 54 CCCH-type  
80 proteins encoded in their genomes [10].

81 The majority of the CCCH-type proteins are unique to the Trityps and they share  
82 a core of 39 proteins in common with differences due to either loss or gain of a  
83 single gene-by-gene duplication.

84 An important set of CCCH-type proteins was previously studied in *T. brucei* and  
85 *T. cruzi*: the tbZFPs and tcZFPs [11,12].

86 The tbZFPs are a group of two small proteins, tbZFP1 (101 residues) and  
87 tbZFP2 (139 residues). The tbZFP2 also contains a WW domain characteristic of  
88 protein-protein interactions with proline-rich motifs. Genetic perturbation assays  
89 provided the evidence that the two tbZFPs can regulate differentiation and  
90 morphogenesis in *T. brucei*. Overexpression of tbZFP2 generated a posterior  
91 extension of the microtubule corset, a mechanism responsible for kinetoplast  
92 repositioning during differentiation [13] and RNAi mediated knockdown of tbZFP2  
93 severely compromised differentiation from bloodstream to procyclic forms. It was  
94 also shown that tbZFP1 is enriched through differentiation to procyclic forms.

95 Later on, a new tbZFP was described, tbZFP3 (CCCH and WW domains), which  
96 enhances development among life cycle stages in *T. brucei*. Ectopic expression  
97 of tbZFP3 in the insect stage of the parasite produced elongated forms (nozzle  
98 phenotype) typical of induced differentiation; while ectopic expression in the  
99 bloodstream stage enhanced differentiation by upregulating EP procyclin protein  
100 expression [14].

101 In *T. cruzi*, four tcZFPs were described, tcZFP1a, tcZFP1b, tcZFP2a and  
102 tcZFP2b. The tcZFP1 proteins present CCCH and proline-rich motifs, whereas  
103 tcZFP2 proteins present the CCCH motif and a WW domain. Interestingly,  
104 tcZFP2s engage in protein-protein interactions with tcZFP1s via the WW domain  
105 and the proline-rich motif respectively. Another interesting thing to note is that  
106 tcZFP1b is *T. cruzi* specific due to a partial gene duplication event that conserved  
107 the central core region of the tcZFP1a protein and diverged in the N- and C-  
108 terminal parts of the protein [11].

109 Different tcZFPs are expressed in different life-cycle stages, thus allowing for a  
110 modularization of the protein-protein interactions and it is proposed that this may  
111 allow the control of expression of distinct cohorts of genes in different life-cycle  
112 stages in a form of post-transcriptional operonic regulation [15].

113 In this work, we focus our attention in tcZFP1b, the *T. cruzi* specific protein.  
114 Since we previously showed that tcZFP1b mRNA was absent in epimastigotes  
115 (the insect stage) and was only detectable in trypomastigotes (the bloodstream  
116 stage), we devised that through the analysis of its ectopic expression in  
117 epimastigotes we could gain insight into the function of this protein.

118 Overexpression of tcZFP1b translates into a cell cycle arrest. An RNA-seq  
119 transcriptome profiling comparison of non-induced and induced cells showed  
120 several mRNAs of related functions changing in a concerted post-transcriptional  
121 pattern resembling post-transcriptional regulons.

122

## 123 **Results**

124

### 125 **Ectopic overexpression of tcZFP1b in epimastigotes (insect stage)**

126 Initial attempts to overexpress tcZFP1b in epimastigotes using constitutive  
127 overexpressing vectors such as pTREX [16] were unsuccessful for the selection  
128 of stable transgenic populations. In fact, parasites stop dividing two weeks after  
129 transfection and died systematically after various attempts. Visual inspection  
130 showed several monster cells (aberrant phenotype with multiple flagella and  
131 possible multiple nuclei or kinetoplasts) few days before parasites died. On the  
132 other hand, transfection with unrelated genes such as eGFP or other *T. cruzi*  
133 genes were successful [16].

134 Thus, we decided to clone tcZFP1b and the control gene eGFP in the inducible  
135 vector pTcINDEX [17] (Fig. 1A). After transfection of culture epimastigotes and  
136 clonal selection of strains, parasites were growing normally. After addition of  
137 tetracycline to induce gene expression, eGFP transgenic parasites grew normally  
138 for six days before reaching a plateau. In contrast, tcZFP1b transgenic parasites  
139 showed a marked decrease in cell counts by day three and eventually stop  
140 dividing by day four (Fig. 1A). Control tcZFP1b parasites without tetracycline  
141 addition grew normally until reaching a plateau by day six (Fig. 1A).

142 Visual analysis of induced expression and background leakage from the system  
143 was done by confocal microscopy for eGFP and by western blot for tcZFP1b.  
144 Results showed high levels of expression upon induction, while the system  
145 showed no background leakage as measured by day four (Fig. 1B)

146 **Monster cells in tcZFP1b transgenic cell line are arrested in cytokinesis**  
147 **and G2 phase of the cell cycle**

148 Presence of abnormal shaped (Monster) cells was observed in the culture of  
149 tetracycline induced tcZFP1b transgenic epimastigotes.

150 Confocal microscopy inspection showed that epimastigotes were arrested at an  
151 early step of cytokinesis. Parasites presented two flagella but no cytoplasmic  
152 division. In contrast, non-induced control parasites were often seen with two  
153 flagella in opposite orientation and proceeding with final steps of cell division  
154 (anti-tubulin staining, Fig. 2A). By using propidium iodide staining for DNA, it was  
155 clearly seen that monster cells also contained two nuclei and two kinetoplasts  
156 (2N2K content) (PI staining, Fig. 2B).

157 This was confirmed by FACS analysis (Fig. S1A). A significant number of cells  
158 that left G1 were arrested at G2/M tcZFP1b induced cells respect to non-induced  
159 control cells. Even a considerable number of epimastigotes presented 4N content  
160 compared to control. DAPI staining of DNA from the same samples confirmed  
161 multiple nuclei and kinetoplasts in cells with multiple flagella (Fig. S1B).

162 The monster cells phenotypes were almost identical to those obtained by treating  
163 epimastigotes with taxol (an anti-tumoral agent and stabilizer of microtubules)  
164 [18]. This fact was strongly indicative that overexpression of tcZFP1b led to a  
165 cytokinesis arrest similar to that of taxol treatment [18].

166 Analysis of the induced expression of tcZFP1b in the same samples was  
167 confirmed by confocal microscopy and indirect immunofluorescence using anti-  
168 tcZFP1b serum. Results showed that tcZFP1b was highly expressed upon  
169 induction and that the protein is distributed in a particulate form over the  
170 cytoplasm excluding nucleus and kinetoplast (Fig. 2C)

171 To gain insights into the induced cell cycle arrest, we performed Transmission  
172 Electron Microscopy (TEM) in the tcZFP1b induced samples. The micrographs  
173 showed, as expected, multiple nucleus and kinetoplast per cell and also showed  
174 evidence of multiple flagellum (Fig. 3). Interestingly, we detected cases with two  
175 kinetoplasts and four basal bodies (asterisks in Fig. 3B) suggesting basal body  
176 duplication without kinetoplast duplication, indicative of the start of another round  
177 of cell division without cytokinesis. It is also interesting to note the distribution of  
178 the chromatin within the nuclei. According to Elias et al. [19], the concentration of  
179 the chromatin in dense granules in the nuclear periphery attached to the  
180 envelope is indicative of cells being in the G2 phase of the cell cycle. A great  
181 majority of the cells showed this particular distribution of the chromatin (see Cr in  
182 Fig. 3).

183 **Changes in the epimastigote transcriptome profile upon overexpression of**  
184 **tcZFP1b as determined by 454 pyrosequencing**



185 Ectopic expression of tcZFP1b produced aberrant phenotypes and cytokinesis  
186 arrest. Since tcZFP1b is an RBP and trypanosomes control gene expression  
187 post-transcriptionally, we decided to investigate global changes in the expression  
188 profile using RNA-seq analysis.

189 Samples were taken in duplicate at 70 hours after tetracycline addition when cell  
190 replication arrest had begun (Induced) or no-addition (Control) (Fig. S2A).

191 The mRNA pyrosequencing produced a total of 233,310 reads for Control and  
192 206,703 reads for Induced. The uniquely mapped reads used for the analysis of  
193 quantitative expression were 95,792 and 107,111 for each duplicate in control,  
194 while 73,490 and 112,570 were mapped for Induced (Fig. S2B). A total of 30,407  
195 and 20,643 reads remained unmapped for Control and Induced, respectively.

196 Uniquely mapped reads were normalized by depth and gene length as indicated  
197 in Methods section. As a standard internal control, we measured the changes in  
198 gene expression of tcZFP1b after induction as well as the endogenous tcZFP1a.  
199 It was confirmed that tcZFP1b mRNA presented a 126X fold change upon  
200 induction while endogenous tcZFP1a expression was low and showed no  
201 changes (Fig. S2C). Importantly, it was also confirmed that leakage from  
202 pTcINDEX inducible vector was very low to non-existent.

203 To analyze coverage and bias in the 454 RNA-seq, we used three different  
204 genes as example: the internal control tcZFP1b (378 b), a downregulated gene  
205 ( $\alpha$ -tubulin, 1318 b), and an upregulated gene (proline racemase, 1065 b) (Fig.  
206 S3). Coverage was very good in all cases as expected for the long read  
207 sequences of the 454 technology, even in low-medium expressed genes (i.e.  
208 proline racemase in control) (Fig. S3). The coverage presented a little bias to the  
209 3' end, which was also expected for this methodology, although lower due to the  
210 mRNA fragmentation procedure instead of cDNA fragmentation (Fig. S3) [20].

211 To analyze global changes, we established a set of rules in order to look for the  
212 most prominent changes. First, we filtered the data set so that genes with at least  
213 five uniquely mapped reads in any condition were retained. This produced a data  
214 set of 2737 genes for analysis. Then, we filtered for those genes between 10 and  
215 100 unique mapped reads that presented a 3X fold or more change. We also  
216 filtered for those genes with unique mapped reads between 100 and 1000 that  
217 presented a 2X fold or more change. In this way, we established a data set of  
218 112 genes that match our criteria of regulation above threshold (Fig. 4). Under  
219 these conditions, a total of 73 genes were upregulated and 39 genes were  
220 downregulated.

221 **Upregulated genes in induced versus control epimastigotes**

222 Among the top 15 expressed genes in epimastigotes, the tyrosine  
223 aminotransferase (TAT), mucin TcSMUGS and hexose transporter genes were  
224 upregulated, while  $\alpha$ -tubulin and prostaglandin F<sub>2</sub> $\alpha$  synthase genes were  
225 downregulated (Fig. 5). Interestingly, TAT genes, which were the top fourth and  
226 eighth expressed genes in the control epimastigote, became the two most  
227 expressed in the induced epimastigote (3.6X fold change) (Fig. 5).

228 Among the 3X fold change upregulated genes, we could establish five different  
229 categories: infectivity and differentiation, carbohydrate metabolism, amino acid  
230 metabolism, ribosomal function and hypothetical proteins (Fig. 6). Interestingly,  
231 the dominant top 20 upregulated genes belonged to the categories of infectivity  
232 and differentiation and amino acid metabolism.

233 The infectivity and differentiation category is populated with a majority of trans-  
234 sialidase like family of genes, a family composed of hundreds of members,  
235 dispersed throughout all chromosomes and mostly expressed on the surface of  
236 trypomastigotes (bloodstream stage) [21]. Notably, several trans-sialidase like  
237 pseudogenes were also expressed and upregulated. Fold changes for trans-  
238 sialidase like members were from 35X to 3X (Fig. 6).

239 Another important upregulated gene was proline racemase (7.1X fold change),  
240 which was demonstrated to participate in differentiation from epimastigotes to  
241 trypomastigotes and to enhance infectivity of host cells (Fig. 6) [22,23].

242 The second important category was amino acid metabolism represented by the  
243 genes of aspartate aminotransferase (cytoplasmic and mitochondrial), 2-amino-  
244 3-ketobutyrate coenzyme A ligase and tyrosine aminotransferase. Their  
245 upregulation ranged from 13X to 3X fold change. The products of these genes  
246 participate in amino acid catabolism. Interestingly, they use pyridoxal phosphate  
247 as a cofactor and pyridoxal kinase was also one of the upregulated genes (11X  
248 fold change). Another interest correlation was the upregulation of an amino acid  
249 transporter gene (4.4X fold change).

250 The third important category was related to carbohydrate energy metabolism.  
251 The glycosomal phosphoenolpyruvate carboxykinase (5.3X fold change) and  
252 malate dehydrogenase (5.5X fold change) were known to contribute to ATP  
253 regeneration in the glycosome [24]. Moreover, D-2-hydroxy-acid dehydrogenase  
254 (3.8X fold change) was involved in the conversion of lactate to pyruvate and 2-  
255 hydroxy-3-oxopropionate reductase (4.4X fold change) was involved in the  
256 glyoxylate-dicarboxylate metabolism. The glycerate kinase gene was upregulated  
257 by 7X fold and was also related to glyoxylate-dicarboxylate metabolism (Fig. 6).



258 These processes were also linked to the glycine, serine, threonine metabolism  
259 described above to provide glyoxylate and hydroxypyruvate.

260 A minor group of genes related to ribosomal functions were upregulated such as  
261 elongation factor-1 gamma (6.6X fold), L14 (3.1X fold) and L44 (3X fold)  
262 ribosomal proteins. Additionally, 12 genes coding for hypothetical proteins were  
263 upregulated ranging from 7X to 3X fold change (Fig. 6)

264 It is important to note that all these upregulated genes are not located close to  
265 each other in the genome. Moreover, they are dispersed in different  
266 chromosomes in most cases.

### 267 **Downregulated genes in induced versus control epimastigotes**

268 Among the top 15 expressed genes, it is worth to mention that the top three were  
269 downregulated:  $\alpha$ -tubulin (2.4X fold) and the two genes for prostaglandin F2 $\alpha$   
270 synthase (2.6X and 2.5X fold) (Fig. 5). Interestingly, it was shown that  
271 prostaglandin F2 $\alpha$  (PGF2) was mainly produced in fast dividing forms of the  
272 parasite (i.e epimastigotes) and it was lower in non-dividing forms or during  
273 stationary phase in culture [25]. Downregulation of PGF2 synthase correlated  
274 perfectly with the fact that induced epimastigotes stopped dividing at the time of  
275 sampling (Fig. S2).

276 Among the group of 3X fold downregulated genes, we established three different  
277 categories: hypothetical proteins, fatty acids biosynthesis and mitochondrial  
278 functions (Fig. 7). The hypothetical proteins dominated the group of  
279 downregulated genes in induced epimastigotes. This fact was drastically different  
280 from the situation with the upregulated genes. This precludes a comprehensive  
281 analysis of the downregulated functions. However, it is worth mentioning that  
282 almost 60% of the hypothetical proteins were conserved among the  
283 trypanosomes, while the other 40% were *T. cruzi* specific (Fig. 7).

284 Downregulated mRNAs with annotated functions involved an important group  
285 related to fatty acid biosynthesis, such as fatty acid elongase (9X fold) and fatty  
286 acid desaturase (5X fold). A possible downregulation of mitochondrial functions  
287 could be also suggested due to downregulation of ATPase beta subunit (8X fold)  
288 and cytochrome C oxidase subunit IV (3.5X fold).

289 Other downregulated genes include amastin (12.7X fold), which was known to be  
290 downregulated in trypomastigotes [26], an amino acid permease (7.9X fold),  
291 histone H1 (4.4X fold), the UBP-2 RNA binding protein (mRNA metabolism, 3.1X  
292 fold) and the paraflagellar rod component par4 (3X fold)

293 To validate the quantification analysis of RNA-seq, we performed qPCR of  
294 selected genes. We chose three upregulated genes (aspartate aminotransferase  
295 cytoplasmic and mitochondrial, and piridoxal kinase), four downregulated genes  
296 ( $\alpha$ -tubulin, PGF2 synthase, ATPase beta subunit, and carboxipeptidase) and  
297 three genes with no expression changes (ribosomal protein L35a, gapdh, and  
298 Acyl carrier protein). The qPCR was performed in triplicate for each gene in  
299 control and induced samples. The results indicated a very good correlation with  
300 the RNA-seq data for all the genes tested (Fig. S4).

### 301 **Sequence motifs in upregulated and downregulated genes**

302 To evaluate the presence of common sequence motifs in the 3'UTRs among the  
303 data set of 112 genes with changed expression above threshold, we selected a  
304 fixed window of 300 nt downstream the stop codon.

305 The bioinformatics analysis was done as described in the Methods section. We  
306 found two different high-confidence motifs in the data set: up-h12 and down-h12  
307 (Fig. 8). The up-h12 motif presented a conserved core of UGuxxxGxGc (Fig. 8A),  
308 while down-h12 was a well-conserved AU-rich (ARE) element [6,9] with a  
309 conserved core UxUAU (Fig. 8B). Both motifs could be folded into stem-loop  
310 structures with the conserved cores exposed in the loops (Fig. 8).

311 The motif up-h12 was found with coverage of 72% in the data set of upregulated  
312 genes and 38% in the downregulated genes. The motif down-h12 was found with  
313 coverage of 89% in downregulated and 27% in the upregulated genes (for the  
314 complete list of genes containing up-h12 and down-h12 and their positions in the  
315 3'UTR see Table S2). Statistical significance was determined using a chi-square  
316 test by comparing the motifs against 115 groups composed by 50 randomly  
317 selected 3'UTR sequences. Results indicated that up-h12 could be found by  
318 chance in the data set with coverage of 45.49% ( $p < 0.01$ ), while down-h12 could  
319 be found at random with coverage of 45.37% ( $p < 0.005$ ). Thus, the presence of  
320 up-h12 and down-h12 were not random and has statistically significant  
321 correlations.

322 The identification of common sequence motifs in the 3'UTR of the up and down  
323 regulated mRNAs reinforces the idea of possible post-transcriptional regulons in  
324 *T. cruzi*.

325

## 326 **Discussion**

327

328 Post-transcriptional regulation is highly dependent on RBPs to achieve a fine-  
329 tuning control of mRNA levels. This is particularly important in a parasite adapted  
330 to two disparate environments in different hosts encountering abrupt changes  
331 that occur in a timeframe of seconds.

332 Several RBPs were described in trypanosomes and they were shown to be  
333 involved in post-transcriptional regulation [6]. Accordingly, the RRM, PUF and  
334 zinc finger domains and motifs are expanded in these parasite genomes [7].

335 Within the ZFP family, tcZFP1b is exclusive to *T. cruzi*. It presents a zinc finger  
336 CCCH motif and a proline rich motif. The tcZFP2 proteins interact with tcZFP1  
337 via the WW domain and the proline-rich motif, respectively [11]. Since they are  
338 expressed in different life-cycle stages, they may allow for a possible modular  
339 control of post-transcriptional regulation depending on the parasite stage [12,14].  
340 In fact, It is known that the three *T. brucei* ZFPs were involved in morphological  
341 changes and differentiation of the parasite [13,15].

342 Interestingly, tcZFP1b is expressed only in trypomastigotes (bloodstream stage)  
343 suggesting a stage-specific function.

344 Our results showed that ectopic expression in epimastigotes (insect stage) led to  
345 cell replication arrest with incomplete cytokinesis and subsequent cell death. The  
346 observed phenotype is almost identical to that obtained when epimastigotes were  
347 treated with taxol (a microtubule stabilizing drug). Treatment using as low as  
348 0.1uM taxol inhibited the growth curve in a very similar way as the tetracycline  
349 induction of tcZFP1b ectopic expression [18,27]. In accordance, transmission  
350 electron micrographs of thin sections of both monster phenotypes looked almost  
351 identical. They also showed chromatin in dense granules attached to the nuclear  
352 envelope suggesting an arrest in G2 phase of the cell cycle [19,28].

353 Comparison between the taxol treatment and ectopic expression of tcZFP1b is  
354 strongly indicative in favor of a similar mode of action. Thus, we could speculate  
355 that ectopic overexpression of tcZFP1b could be stabilizing the sub-pellicular  
356 microtubule corset and, in turn, blocking cytokinesis [27,29]. Although it is more  
357 difficult to speculate on a direct mode of action of tcZFP1b in this event, we  
358 argue in favor of an indirect action through one or more intermediates.

359 It is well known that cytokinesis depends on the dynamics of this sub-pellicular  
360 corset. In fact, the microtubule array is cross-linked together and is present  
361 throughout the full cell cycle with new microtubules being added and the array  
362 being inherited in a semi-conservative manner by the two daughter cells [27,29].  
363 Thus, it is clear that a stabilization of microtubules would end up blocking the  
364 cytokinesis process.

365 The tcZFP1b is an RBP that could regulate a set of targets that, in turn, could  
366 potentially regulate another set of targets. In an attempt to understand the whole  
367 picture of its mode of action we decided to look genome-wide instead of looking  
368 at the specific targets of its RNA binding motif.

369 An interesting fact observed in the genome-wide analysis was the 2.4X fold  
370 downregulation of  $\alpha$ -tubulin. Since new microtubules are constantly needed in  
371 order to progress to cytokinesis, this downregulation could be one of the hints  
372 pointing to the observed cytokinesis arrest. Non-dividing forms, such as  
373 trypomastigotes, would have more stabilized microtubules in the sub-pellicular  
374 corset than the fast dividing forms (epimastigotes and amastigotes).

375 Cause or consequence of epimastigotes stopping cell division several hours  
376 upon induction was the fact that PGF2 synthase mRNAs were downregulated by  
377 2.6 and 2.5X fold. It is known that production of PGF2 $\alpha$  decreased significantly in  
378 non-dividing forms [25].

379 One of the components of the flagellum, paraflagellar rod component par4, was  
380 also downregulated 3X fold. It is not clear the function of this component in the  
381 flagellum biology but it is tempting to speculate that its downregulation could be  
382 linked to the observed phenotype.

383 Another interesting observation of the genome-wide analysis is the fact that  
384 several genes of related functions were upregulated or downregulated in a  
385 concerted form. Most of these genes are located in different chromosomes and,  
386 thus, in different post-transcriptional units. However, their post-transcriptional  
387 levels appear to be concerted.

388 Notably, several genes involved in amino acid catabolism that use pyridoxal  
389 phosphate as cofactor were upregulated between 13X and 3X fold while the  
390 pyridoxal kinase mRNA was concomitantly up by 11X fold.

391 Several genes related to the glyoxylate and dicarboxylate metabolism were also  
392 upregulated. Interestingly, amino acid catabolism is linked to the former  
393 metabolism through the production of hydroxypyruvate and glyoxylate.

394 One important conclusion obtained from this work is that post-transcriptional  
395 regulons are evident in *T. cruzi*, as it also seems to emerge from other works in  
396 *T. brucei* [30,31]. The finding of common sequence motifs in the 3'UTRs  
397 reinforces the idea of the regulons model. In this work, we found two different  
398 high-confidence motifs, up-h12 in the upregulated and down-h12 in the  
399 downregulated genes. Both motifs presented conserved sequence cores  
400 exposed in loops. Interestingly, down-h12 resembles a classical AU-rich (ARE)  
401 element previously involved in the instability of mRNAs in *T. cruzi* [9]. The list of

402 down-h12 containing genes included fatty acid elongase, fatty acid desaturase,  
403 ATPase beta subunit, cytochrome C oxidase and amastin among others (Table  
404 S2)

405 Important to note is that most of the upregulated genes belonged to the family of  
406 trans-sialidase like members, a family expressed almost exclusively in  
407 trypomastigotes, the parasite form that interact with the mammalian host. The  
408 expression of trans-sialidase pseudogenes was also evident. Expression of  
409 pseudogenes was shown to have a role in gene expression via the RNA  
410 interference (RNAi) system in *T. brucei* [32]. However, since RNAi is lacking in *T.*  
411 *cruzi* [33,34], the upregulation of these pseudogenes remains puzzling.

412 A general view of the genome-wide analysis seemed to point to the activation of  
413 at least part of a program to differentiate to trypomastigotes, although the sole  
414 overexpression of tcZFP1b might not be sufficient to accomplish the task.  
415 Upregulation of proline racemase (7.1X fold) is remarkable in this context since it  
416 was shown that its ectopic expression in epimastigotes resulted in enhanced  
417 differentiation to trypomastigotes and enhanced infectivity [22,23].

418 In addition, the upregulation of several trans-sialidase like mRNAs, the possible  
419 stabilization of the microtubule sub-pellicular corset to enter a non-dividing form,  
420 the upregulation of a mechanism of energy production through the glycosome  
421 and amino acid catabolism, and the possible downregulation of mitochondrial  
422 functions are compelling evidence pointing towards that direction.

423 With this genome-wide analysis in hand, one of the challenging tasks for the near  
424 future would be to dissect the chain of events unleashed after overexpression of  
425 tcZFP1b in epimastigotes, beginning by looking for direct mRNA targets of its  
426 zinc-finger motif. Since the tcZFP2 proteins interact with tcZFP1b, it will be  
427 interesting also to look for direct mRNA targets of their zinc-fingers as well.

428

## 429 **Materials and Methods**

430

### 431 **Trypanosome cultures and tetracycline induction of pTcINDEX**

432 For inducible expression of tcZFP1b in the parasite, we first generated a cell  
433 line expressing T7 RNA polymerase and tetracycline repressor genes by  
434 transfecting epimastigotes with the plasmid pLew13 by electroporation as  
435 previously described [35]. Stable transfectants were selected and grown in  
436 brain-heart-tryptose (BHT) medium supplemented with 10% inactivated fetal  
437 calf serum (FCS) and 200 µg/ml G418 (Gibco). This cell line was then



438 transfected with pTcINDEX construct [17] carrying tcZFP1B gene or with the  
439 GFP gene, and transgenic parasites were obtained after selection with 200  
440  $\mu\text{g/ml}$  G418 and 200  $\mu\text{g/ml}$  hygromycin B (Calbiochem). Epimastigote cultures  
441 were grown to reach a cell density of  $5 \times 10^6$  parasites/ml and protein  
442 expression was induced by the addition of 5- $\mu\text{g/ml}$  of tetracycline for 60-72  
443 hours. Epimastigotes ( $1 \times 10^9$  cells) were harvested by centrifugation at 1,000  
444 x g for 5 min, washed twice in PBS and lysed on ice by incubation with  
445 Laemmli's sample buffer (for Western Blot) or fixed by 4% paraformaldehyde  
446 (for indirect immunofluorescence)

#### 447 **Microscopy Analysis**

448 Confocal microscopy, GFP detection and indirect immunofluorescence (IFI)  
449 analysis were done as previously described [16].

450 For Transmission Electron Microscopy (TEM),  $10^7$  epimastigotes were fixed in  
451 2.5% glutaraldehyde in 0.1 M phosphate buffer, pH 7.2, for 60 minutes, washed  
452 in the same buffer, post-fixed in 1%  $\text{OsO}_4$  and 0.8% potassium ferrocyanide in  
453 0.1 M sodium cacodylate buffer at room temperature for 40 minutes, washed in  
454 0.1 M phosphate buffer, dehydrated in acetone, and embedded in Polybed resin.  
455 Ultrathin sections were stained with uranyl acetate and lead citrate and observed  
456 using a TEM Philips EM 301 at CMA (Centro de Microscopias Avanzadas,  
457 University of Buenos Aires).

#### 458 **Fluorescence Activated Cell Sorting (FACS) analysis**

459 For FACS analysis, epimastigotes expressing inducible tcZFP1b were  
460 compared to a non-induced control. Samples were taken at 70 hours after  
461 induction when parasites stopped cell division. A total of  $10^7$  epimastigotes  
462 were washed twice with PBS and fixed in 500  $\mu\text{ml}$  of 70% (v/v) ice cold  
463 ethanol/PBS overnight at 4°C. The fixed cells were then resuspended in  
464 500  $\mu\text{ml}$  of PBS supplemented with 50  $\mu\text{g/ml}$  propidium iodide, 20  $\mu\text{g/ml}$   
465 RNase A and 2mM EDTA in PBS before incubation at 37°C for 30 min.  
466 FACS analysis was performed with a Becton Dickinson FACSCalibur using  
467 FL2-A (detecting fluorescence emission between 543 and 627 nm, propidium  
468 iodide), the forward scatter and the side scatter detectors. A total of 10,000-  
469 gated events were harvested from each sample. Data were interpreted using  
470 the WinMDI 2.9 software (Scripps Research Institute).

#### 471 **RNA extraction, processing and Pyrosequencing**

472 Total RNA from  $1 \times 10^9$  induced and non-induced control epimastigotes were  
473 extracted from biological duplicates using standard procedures [35].



474 Total RNA quantity and quality were assessed using the Agilent 2100  
475 Bioanalyzer (Agilent technologies).

476 Poly-A+ RNA was selected by oligo-dT chromatography in two rounds  
477 (Dynabeads mRNA DIRECT kit, Invitrogen), using the total RNA extracted from  
478 biological duplicates for each condition. The two rounds purification of polyA+  
479 allowed diminishing the rRNA contamination below 10%. A total amount of 200  
480 ng RNA was quantitated by Ribogreen and quality assessed on an RNA 6000  
481 Pico Chip on the Agilent 2100 Bioanalyzer. RNA was fragmented using a solution  
482 of ZnCl<sub>2</sub> according to 454 cDNA rapid library preparation method manual,  
483 generating fragments with a mean size of 500 bp. Finally, the cDNA was  
484 synthesized using random hexamers according to manufacturers instructions  
485 (Roche).

486 The cDNA quality was assessed using a high-sensitivity Chip on the Agilent 2100  
487 Bioanalyzer and subjected to 454 sequencing using standard protocols (Roche)  
488 at INDEAR sequencing facility (Rosario, Argentina). A half of the PicoTiter Plate  
489 (PTP) was divided in quarters, using one quarter for the biological duplicates of  
490 the non-induced condition with two MIDs (Multiplex Identifiers) and the other  
491 quarter for the duplicates of induced condition with two MIDs. Raw sequencing  
492 data produced 233,310 reads and 206,703 reads for each quarter respectively  
493 with median read length of 464 and 455 bases respectively.

#### 494 **Bioinformatics analysis of transcriptome data**

495 Raw data was filtered for artificial duplicate reads and mapped against the *T.*  
496 *cruzi* CL-Brener esmeraldo and non-esmeraldo haplotypes as references  
497 genomes. The mapping was done using the 454 GS Reference Mapper software  
498 (Roche). The uniquely mapped reads were taken into account for further  
499 processing. Mapping statistics could be found in Fig. S2C.

500 For normalization purposes, the unique read counts were normalized first by  
501 gene length. Since RNA fragmentation produced an average of 500 b, we  
502 introduced a factor for each gene as a ratio of gene length to fragment size (500  
503 b). If fragment size is greater than gene length then factor is 1. If fragment size is  
504 lower than gene length then a correction factor is introduced in the formula  
505 below:

$$factor(n) = \begin{cases} \frac{geneLength}{fragmentSize} & fragmentSize < geneLength_n \\ 1 & fragmentSize \geq geneLength_n \end{cases}$$

$$normalizedReads_A(n) = \frac{uniqueReadsCount(gene_n)_A \cdot \min(totalReads_A, totalReads_B)}{totalReads_A \cdot factor_n}$$

506

507 Once normalized reads for conditions A and B were calculated for the biological  
508 duplicates, a fold change was calculated as a ratio between induced and non-  
509 induced control normalized reads.

510 The results were parsed into a tabulated spreadsheet format with the GeneDB  
511 accession number, fold change, GeneDB description and GO annotation (if  
512 available) for each gene (Table S1).

### 513 **Computational analysis for sequence motifs search.**

514 For 3' UTR sequence definition, a length of 300 nt downstream to the CDS was  
515 used to obtain sequences resembling the 3'UTR, in agreement to previously  
516 reported data from trypanosomes [36,37] was downloaded using TcruziDB  
517 sequence retrieval tool. Homologue genes within each group with similar 3'UTR  
518 were filtered to avoid duplicated sequences. Consensus motifs were predicted  
519 from each dataset using CMfinder 0.2 [38]. Candidate motifs obtained were used  
520 to build the stochastic context-free grammar (SCFG) model (INFERNAL  
521 program). The SCFG for each candidate motif was used to search against the  
522 specific data set and the complementary database to obtain the number of hits  
523 for each motif (CMSEARCH program). The motif with the highest enrichment in  
524 the specific data set was considered to be the best candidate motif. The motif  
525 logo was constructed using WebLogo (<http://weblogo.berkeley.edu/>). Finally,  
526 RNAfold server [39] was used to plot the secondary structure of the  
527 representative RNA motifs. Differences between groups were examined for  
528 statistical significance using chi-square test. Comparison was made between the  
529 motif-containing group and random 3'UTR groups (115 lists composed by 50  
530 randomly selected sequences).

531

### 532 **Acknowledgments**

533 The authors wish to thank Dr. Guillermo Alonso, Dr. Claudio Pereira, Dr. Maria  
534 Teresa Tellez-Iñon, Dr. Martin Edreira and Dr. Ken Kobayashi for helpful  
535 discussions during the writing of this manuscript. The authors thank John Kelly  
536 and Martin Taylor (London School of Tropical Medicine, London, United  
537 Kingdom) for kindly providing the pTcINDEX expression vector and Dr. Juan  
538 José Cazzulo (IIB-INTECH, Buenos Aires, Argentina) for support. JGDG and  
539 MPV are members of the career of scientific investigator of CONICET, Argentina.  
540

### 541 **References**

542

543 1. Clayton C (2002) Life without transcriptional control? From fly to man and back

- 544 again. EMBO J 21: 1881–1888.
- 545 2. Clayton C, Shapira M (2007) Post-transcriptional regulation of gene  
546 expression in trypanosomes and leishmanias. Mol Biochem Parasitol 156:  
547 93–101.
- 548 3. Martínez-Calvillo S, Yan S, Nguyen D, Fox M, Stuart K, et al. (2003)  
549 Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both  
550 directions within a single region. Mol Cell 11: 1291–1299.
- 551 4. Worthey E, Martinez-Calvillo S, Schnaufer A, Aggarwal G, Cawthra J, et al.  
552 (2003) *Leishmania major* chromosome 3 contains two long convergent  
553 polycistronic gene clusters separated by a tRNA gene. Nucleic Acids Res 31:  
554 4201-4210.
- 555 5. Liang X, Haritan A, Uliel S, Michaeli S (2003) trans and cis splicing in  
556 trypanosomatids: mechanism, factors, and regulation. Eukaryot Cell 2: 830–  
557 840.
- 558 6. De Gaudenzi J, Frasch AC, Clayton C (2005) RNA-binding domain proteins in  
559 Kinetoplastids: a comparative analysis. Eukaryot Cell 4: 2106–2114.
- 560 7. Kramer S, Carrington M (2011) Trans-acting proteins regulating mRNA  
561 maturation, stability and translation in trypanosomatids. Trends Parasitol 27:  
562 23–30.
- 563 8. Kramer S, Kimblin NC, Carrington M (2010) Genome-wide in silico screen for  
564 CCCH-type zinc finger proteins of *Trypanosoma brucei*, *Trypanosoma cruzi*  
565 and *Leishmania major*. BMC Genomics 11: 283.
- 566 9. Cassola A, De Gaudenzi J, Frasch AC (2007) Recruitment of mRNAs to  
567 cytoplasmic ribonucleoprotein granules in trypanosomes. Mol Microbiol 65:  
568 655–670.
- 569 10. Caro F, Bercovich N, Atorrasagasti C, Levin MJ, Vázquez MP (2006)  
570 *Trypanosoma cruzi*: analysis of the complete PUF RNA-binding protein  
571 family. Exp Parasitol 113: 112–124.
- 572 11. Caro F, Bercovich N, Atorrasagasti C, Levin MJ, Vazquez MP (2005) Protein  
573 interactions within the TcZFP zinc finger family members of *Trypanosoma*  
574 *cruzi*: implications for their functions. Biochem Biophys Res Commun 333:  
575 1017–1025.

- 576 12. Hendriks EF, Robinson DR, Hinkins M, Matthews KR (2001) A novel CCCH  
577 protein which modulates differentiation of *Trypanosoma brucei* to its procyclic  
578 form. EMBO J 20: 6700–6711.
- 579 13. Hendriks E, Matthews K (2005) Disruption of the developmental programme  
580 of *Trypanosoma brucei* by genetic ablation of TbZFP1, a differentiation-  
581 enriched CCCH protein. Mol Microbiol 57: 706–716.
- 582 14. Paterou A, Walrad P, Craddy P, Fenn K, Matthews K (2006) Identification  
583 and stage-specific association with the translational apparatus of TbZFP3, a  
584 CCCH protein that promotes trypanosome life-cycle development. J Biol  
585 Chem 281: 39002–39013.
- 586 15. Ling AS, Trotter JR, Hendriks EF (2011) A Zinc Finger Protein, TbZC3H20,  
587 Stabilizes Two Developmentally Regulated mRNAs in Trypanosomes. J Biol  
588 Chem 286: 20152–20162.
- 589 16. Westergaard GG, Bercovich N, Reinert MD, Vázquez MP (2010) Analysis of  
590 a nuclear localization signal in the p14 splicing factor in *Trypanosoma cruzi*.  
591 Int J Parasitol, 40: 1029–1035.
- 592 17. Taylor M, Kelly J (2006) pTcINDEX: a stable tetracycline-regulated  
593 expression vector for *Trypanosoma cruzi*. BMC Biotechnol 6: 32–32.
- 594 18. Baum SG, Wittner M, Nadler JP, Horwitz SB, Dennis JE, et al. (1981) Taxol,  
595 a microtubule stabilizing agent, blocks the replication of *Trypanosoma cruzi*.  
596 Proc Natl Acad Sci USA 78: 4571–4575.
- 597 19. Elias MCQB, Faria M, Mortara RA, Motta MCM, de Souza W, et al. (2002)  
598 Chromosome localization changes in the *Trypanosoma cruzi* nucleus.  
599 Eukaryot Cell 1: 944–953.
- 600 20. Wang Z, Gerstein M, Snyder M. (2009) RNA-Seq: a revolutionary tool for  
601 transcriptomics. Nat Rev Genet 10: 57–63.
- 602 21. Frasch AC (2000) Functional diversity in the trans-sialidase and mucin  
603 families in *Trypanosoma cruzi*. Parasitol today 16: 282–286.
- 604 22. Coutinho L, Ferreira MA, Cosson A, Batista MM, Batista DDGJ, et al. (2009)  
605 Inhibition of *Trypanosoma cruzi* proline racemase affects host-parasite  
606 interactions and the outcome of in vitro infection. Mem. Inst. Oswaldo Cruz  
607 104: 1055–1062.

- 608 23. Chamond N, Goytia M, Coatnoan N, Barale JC, Cosson A, et al. (2005)  
609 *Trypanosoma cruzi* proline racemases are involved in parasite differentiation  
610 and infectivity. *Mol Microbiol* 58: 46–60.
- 611 24. Urbina JA, Osorno CE, Rojas A. (1990) Inhibition of phosphoenolpyruvate  
612 carboxykinase from *Trypanosoma* (Schizotrypanum) *cruzi* epimastigotes by  
613 3-mercaptopycolinic acid: in vitro and in vivo studies. *Arch. Biochem. Biophys*  
614 282: 91–99.
- 615 25. Kubata BK, Duszenko M, Kabututu Z, Rawer M, Szallies A, et al. (2000)  
616 Identification of a novel prostaglandin F(2alpha) synthase in *Trypanosoma*  
617 *brucei*. *J. Exp. Med* 192: 1327–1338..
- 618 26. Coughlin BC, Teixeira SM, Kirchhoff LV, Donelson JE (2000) Amastin mRNA  
619 abundance in *Trypanosoma cruzi* is controlled by a 3'-untranslated region  
620 position-dependent cis-element and an untranslated region-binding protein. *J*  
621 *Biol Chem* 275: 12051–12060..
- 622 27. Hammarton T, Monnerat S, Mottram J (2007) Cytokinesis in  
623 trypanosomatids. *Curr Opin Microbiol* 10: 520–527.
- 624 28. Elias M, Dacunha J, Defaria F, Mortara R, Freymuller E , et al. (2007)  
625 Morphological Events during the *Trypanosoma cruzi* Cell Cycle. *Protist* 158:  
626 147–157.
- 627 29. de Souza W (2009) Structural organization of *Trypanosoma cruzi*. *Mem. Inst.*  
628 *Oswaldo Cruz* 104 Suppl 1: 89–100.
- 629 30. Archer SK, Inchaustegui D, Queiroz R, Clayton C (2011) The Cell Cycle  
630 Regulated Transcriptome of *Trypanosoma brucei*. *PLoS ONE* 6: e18425
- 631 31. Archer SK, Luu VD, De Queiroz R., Brems S, Clayton C, et al. (2009)  
632 *Trypanosoma brucei* PUF9 Regulates mRNAs for Proteins Involved in  
633 Replicative Processes over the Cell Cycle. *PLoS Pathogens* 5: e1000565
- 634 32. Wen YZ, Zheng LL, Liao JY, Wang MH, Wei Y, et al. (2011) Pseudogene-  
635 derived small interference RNAs regulate gene expression in African  
636 *Trypanosoma brucei*. *Proc Natl Acad Sci USA* 108: 8345–8350.
- 637 33. Ullu E, Tschudi C, Chakraborty T (2004) RNA interference in protozoan  
638 parasites. *Cell. Microbiol* 6: 509–519.
- 639 34. DaRocha WD, Otsu K, Teixeira SMR, Donelson JE (2004) Tests of

- 640 cytoplasmic RNA interference (RNAi) and construction of a tetracycline-  
641 inducible T7 promoter system in *Trypanosoma cruzi*. Mol Biochem  
642 Parasitol,133: 175–186.
- 643 35. Ben-Dov CP, Levin MJ, Vázquez MP (2005) Analysis of the highly efficient  
644 pre-mRNA processing region HX1 of *Trypanosoma cruzi*. Mol Biochem  
645 Parasitol 140: 97–105.
- 646 36. Brandão A, Jiang T (2009). The composition of untranslated regions in  
647 *Trypanosoma cruzi* genes. Parasitol Int 58: 215–219.
- 648 37. Benz C, Nilsson D, Andersson B, Clayton C, Guilbride DL (2005). Messenger  
649 RNA processing sites in *Trypanosoma brucei*. Mol Biochem Parasitol 143:  
650 125–134.
- 651 38. Yao Z, Weinberg Z, Ruzzo WL (2006) CMfinder--a covariance model based  
652 RNA motif finding algorithm. Bioinformatics 22: 445–452.
- 653 39. Hofacker IL (2003) Vienna RNA secondary structure server. Nucleic Acids  
654 Res 31: 3429–3431.

655

## 656 **Figure Legends**

657

### 658 **Figure 1 – Inducible overexpression of tcZFP1 in epimastigotes.**

659 **A:** Upper panel, schematic representation of pTclINDEX, as appeared in [17],  
660 with the multiple cloning site used to introduce the eGFP and tcZFP1b genes.  
661 Both vectors were transfected into epimastigotes to generate independent cell  
662 lines. The eGFP was used as a control to test overexpression and leakage.  
663 Lower panel, growth curves for GFP and tcZFP1b cell lines induced with  
664 tetracycline and non-induced (control). **B:** Left panel, overexpression test of GFP  
665 using confocal microscopy. Results showed no leakage for control (non-induced)  
666 and strong induction after tetracycline addition (induced). Right panel, Western  
667 blot confirmation of tcZFP1b overexpression using a tcZFP1b specific mouse  
668 antiserum. Antibodies specificity was tested against a His-tcZFP1b recombinant  
669 protein produced in bacteria.

670

### 671 **Figure 2 – Monster cells appeared upon overexpression of tcZFP1b in 672 epimastigotes.**

673 **A:** Confocal microscopy images of tcZFP1b induced and control (non-induced)  
674 epimastigotes detected using anti-tubulin specific antibodies (FITC conjugated



675 anti-mouse antibody). Induced cells showed monster phenotypes with arrested  
676 cytokinesis while non-induced cells showed wild type phenotypes and parasites  
677 proceeding with typical cell division. **B:** Confocal microscopy images of  
678 epimastigotes stained with propidium iodide (PI) for detection of DNA. Induced  
679 and control parasites were mixed 50:50 in one slide for visualization and direct  
680 comparison of wild type and monster phenotypes. N, nucleus; K, kinetoplast  
681 DNA. **C:** Confocal microscopy images of induced epimastigotes detected using  
682 anti-tcZFP1b specific polyclonal serum (FITC conjugated anti-mouse antibody)  
683 and DAPI staining for DNA. The central panels show merged images of DIC  
684 (differential interface contrast), FITC detection and DAPI staining. Overexpressed  
685 tcZFP1b is distributed in the cytoplasm excluding nucleus and kinetoplast.

686

687 **Figure 3 – Monster epimastigotes arrested in G2-phase of the cell cycle**  
688 **upon overexpression of tcZFP1b**

689 Six images of monster phenotypes obtained using Transmission Electron  
690 Microscopy (TEM). N, nucleus; Nu, nucleolus; K, kinetoplast; FL, flagellum; Cr,  
691 chromatin. Asterisks in image **B** denotes basal body duplications. Dense  
692 granules in the nucleus indicate chromatin attached to the envelope and it is  
693 indicative of the G2-phase of the cell cycle.

694

695 **Figure 4 – Transcriptome profile of induced versus control (non-induced)**  
696 **epimastigotes.**

697 Normalized unique read counts from induced and control sequenced RNA  
698 samples of epimastigotes were plotted against each other in a log scale. Each  
699 blue dot represents a different gene. Genes with equal or more than five  
700 normalized read counts were plotted (2737 genes). The red line represents the  
701 threshold of upregulated genes and the green line represents the threshold for  
702 the downregulated ones in the induced sample. Two different thresholds were  
703 considered as significant: a) 3 or more fold change in the range from 5 to 100  
704 unique read counts; and b) 2 or more fold change in the range from 100 to 1000  
705 unique read counts. A total of 112 genes changed expressions above these  
706 thresholds.

707

708 **Figure 5 – Expression profile of top 15 expressed genes in control versus**  
709 **induced epimastigotes.**

710 The top 15 expressed genes in the control (non-induced) RNA sample were  
711 compared with their expression in the induced RNA sample and the normalized  
712 unique read counts were plotted. A star below the gene description name  
713 indicates a significant change in mRNA levels between the two samples. A blue  
714 star denotes a downregulation and a red star an upregulation in the induced

715 sample. Similar description names indicate similar genes from the two *T. cruzi*  
716 haplotypes reference genomes, esmeraldo and esmeraldo-like, of the CL-Brener  
717 strain.

718

719 **Figure 6 – Upregulated genes in epimastigotes upon overexpression of**  
720 **tcZFP1b**

721 The fold change (induced versus control) was plotted considering only those  
722 genes with changes equal or more than 3X. Upregulated genes were categorized  
723 using different colored dots as indicated. For proteins with predicted function, the  
724 description name is provided. For hypothetical proteins, the GeneDB number is  
725 provided

726

727 **Figure 7 – Downregulated genes in epimastigotes upon overexpression of**  
728 **tcZFP1b**

729 The fold change (induced versus control) was plotted considering only those  
730 genes with changes equal or more than 3X. Downregulated genes were  
731 categorized using different colored dots as indicated. For proteins with predicted  
732 function, the description name is provided. For hypothetical proteins, the  
733 GeneDB number is provided.

734

735 **Figure 8 – Common motifs in the 3'UTR of upregulated and downregulated**  
736 **genes in induced epimastigotes.**

737 **A:** Consensus motif found in upregulated genes. **B:** Consensus motif found in  
738 downregulated genes. Upper panel, sequence logo representations of consensus  
739 motifs. Middle panel, secondary structures for eight selected sequences  
740 representing the found motif, each one with their unique identifier of the GeneDB  
741 number below (<http://tritrypdb.org>). Lower panel, linear sequences for the  
742 selected 3'UTRs with their corresponding consensus. The selected 3'UTR for up-  
743 h12 are 504105.140, enolase; 504147.30, hypothetical protein, conserved;  
744 505807.180, 2-hydroxy-3-oxopropionate reductase; 506211.70, RNA-binding  
745 protein; 506597.40, trans-sialidase; 507185.40, trans-sialidase (pseudogene);  
746 08293.90, elongation factor 1-gamma (EF-1-gamma); 508441.20, glycosomal  
747 hosphoenolpyruvate carboxykinase. The selected 3'UTR for down-h12 are  
748 467287.30, ATPase beta subunit; 506529.360, cytochrome c oxidase subunit IV;  
749 508175.189, hypothetical protein, conserved; 509541.4, paraflagellar rod  
750 component par4, putative; 509747.80, hypothetical protein, conserved;  
751 510719.100, hypothetical protein, conserved; 511073.10, fatty acid desaturase;  
752 511439.40, hypothetical protein.

753

754 **Figure S1 – FACS analysis and DNA content of normal and monster cells**

755 **A:** FACS analysis. Upper panel, histogram analysis of control and induced cells.  
756 Lower panel, two-dimensional dot plot analysis of the corresponding histogram  
757 analysis shown above. The ploidies of the peaks and dots are shown.  
758 Epimastigotes were prepared 60hs after induction and fixed. **B:** Confocal  
759 microscope analysis of the FACS samples. Left panel shows normal cell division  
760 of non-induced sample, right panel shows arrested cytokinesis in induced  
761 sample.

762

### 763 **Figure S2 – 454 transcriptome sampling and metrics**

764 **A:** Epimastigotes growth curve for non-induced (control) and induced cell lines.  
765 An arrow indicates the time point of sampling for RNA-seq analysis. **B:**  
766 Reference mapping metrics for 454 reads obtained by pyrosequencing for  
767 Control and Induced samples. Uniquely mapped 1 and uniquely mapped 2 refers  
768 to the biological duplicates reads. Unmapped reads are reported as total for  
769 condition. **C:** Overexpression of tcZFP1b after tetracycline induction. The  
770 endogenous non-induced tcZFP1a is reported for comparison purposes.

771

### 772 **Figure S3 - 454 Transcriptome sequence coverage and depth.**

773 **A:** tcZFP1b coverage and SNPs detected. SNPs are indicated above and  
774 correspond to differences between *T. cruzi* CL-Brener strain (reference genome)  
775 and the *T. cruzi* I strain used in this study; aa, amino acids; nt, nucleotides; syn,  
776 synonyms. **B:** High expressed gene coverage example. **C:** Low expressed gene  
777 coverage example. Darker reads indicate forward direction. Lighter reads  
778 indicate reverse direction.

779

### 780 **Figure S4 – Validation of RNA-seq using qPCR**

781 Real-time PCR (qPCR) was used to validate RNA-seq results on selected genes.  
782 Measurements are the results of triple technical replicates for each biological  
783 duplicate. AATc, AATm, Aspartate aminotransferase cytoplasmic and  
784 mitochondrial respectively (Tc00.1047053503841.70, Tc00.1047053510945.70);  
785 Piryk, pyridoxal kinase (Tc00.1047053507925.40); CPEP, carboxypeptidase  
786 (Tc00.1047053504153.160); GAPDH, glyceraldehyde-3-P dehydrogenase;  
787 AcylCar, acyl carrier protein, mitochondrial precursor  
788 (Tc00.1047053511867.140); L35A, ribosomal protein L35A  
789 (Tc00.1047053506559.470); PGF2A, prostaglandine F2 $\alpha$  synthase  
790 (Tc00.1047053507617.9); ATUB;  $\alpha$ -tubulin (Tc00.1047053411235.9); ATPase,  
791 mitochondrial ATPase beta subunit (Tc00.1047053509233.180).

792

793 **Table S1 – Complete 454 transcriptome data for control and induced**  
794 **epimastigotes**

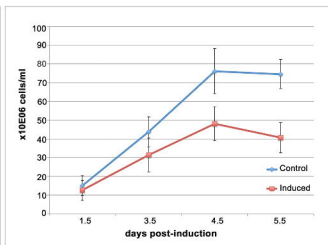
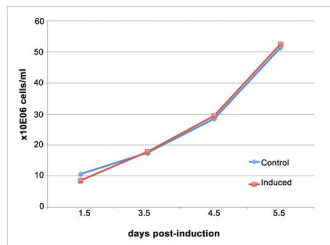
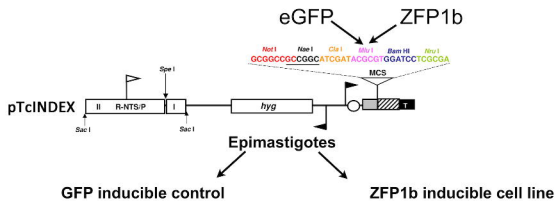
795 Results obtained from the 454 GS Reference Mapper software and normalized  
796 using a custom script as indicated in Materials and Methods were parsed to a  
797 tabulated spreadsheet format. The term f1 refers to the Control condition and f2  
798 to the induced condition. Gene Ontology (GO) annotation was included when  
799 available.

800

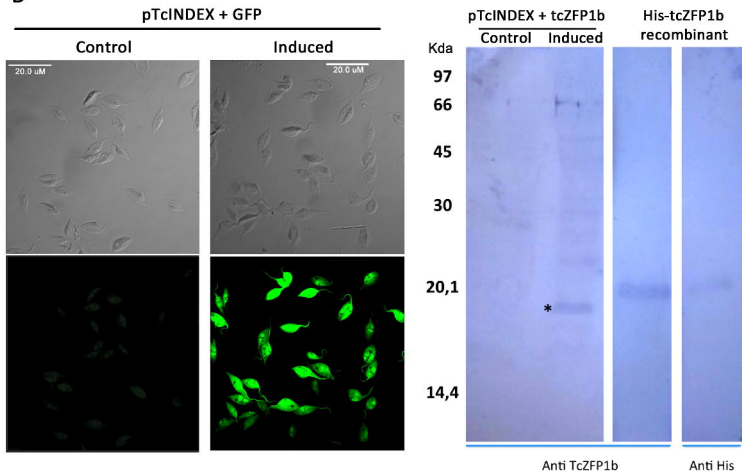
801 **Table S2 – Presence of up-h12 and down-h12 motifs in the upregulated and**  
802 **downregulated genes.**

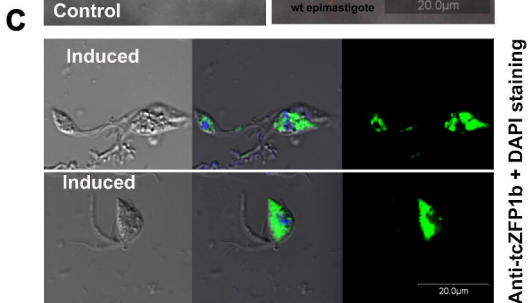
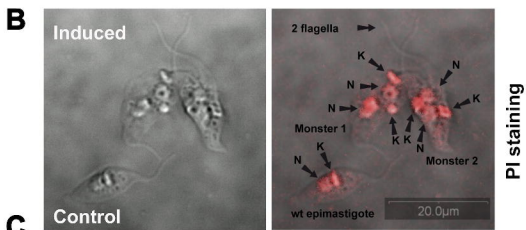
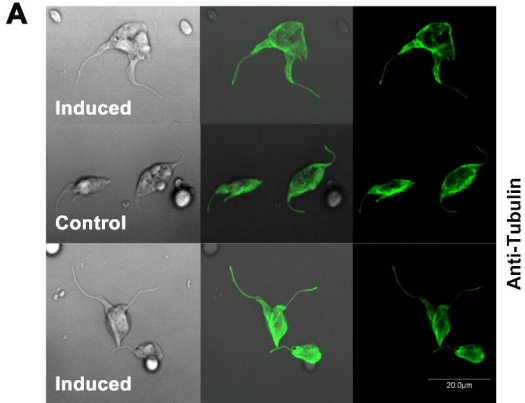
803 The gene ID, description, motif presence and position in their respective 3'UTRs  
804 are indicated. Positions indicate distances from stop codon.

A

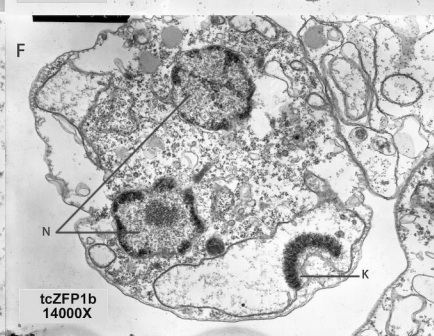
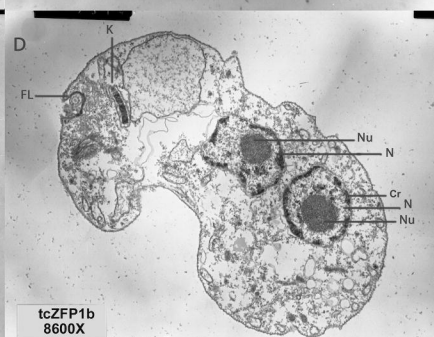
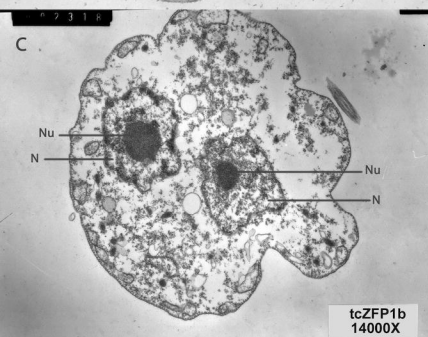
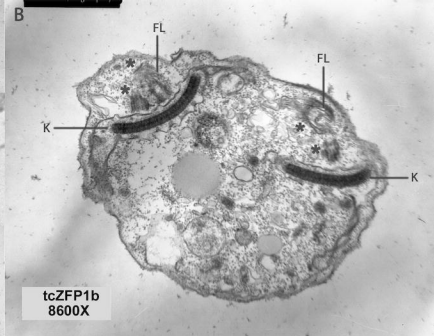
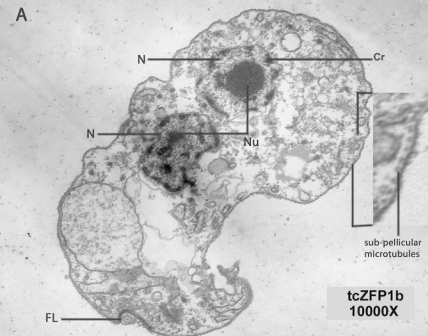


B

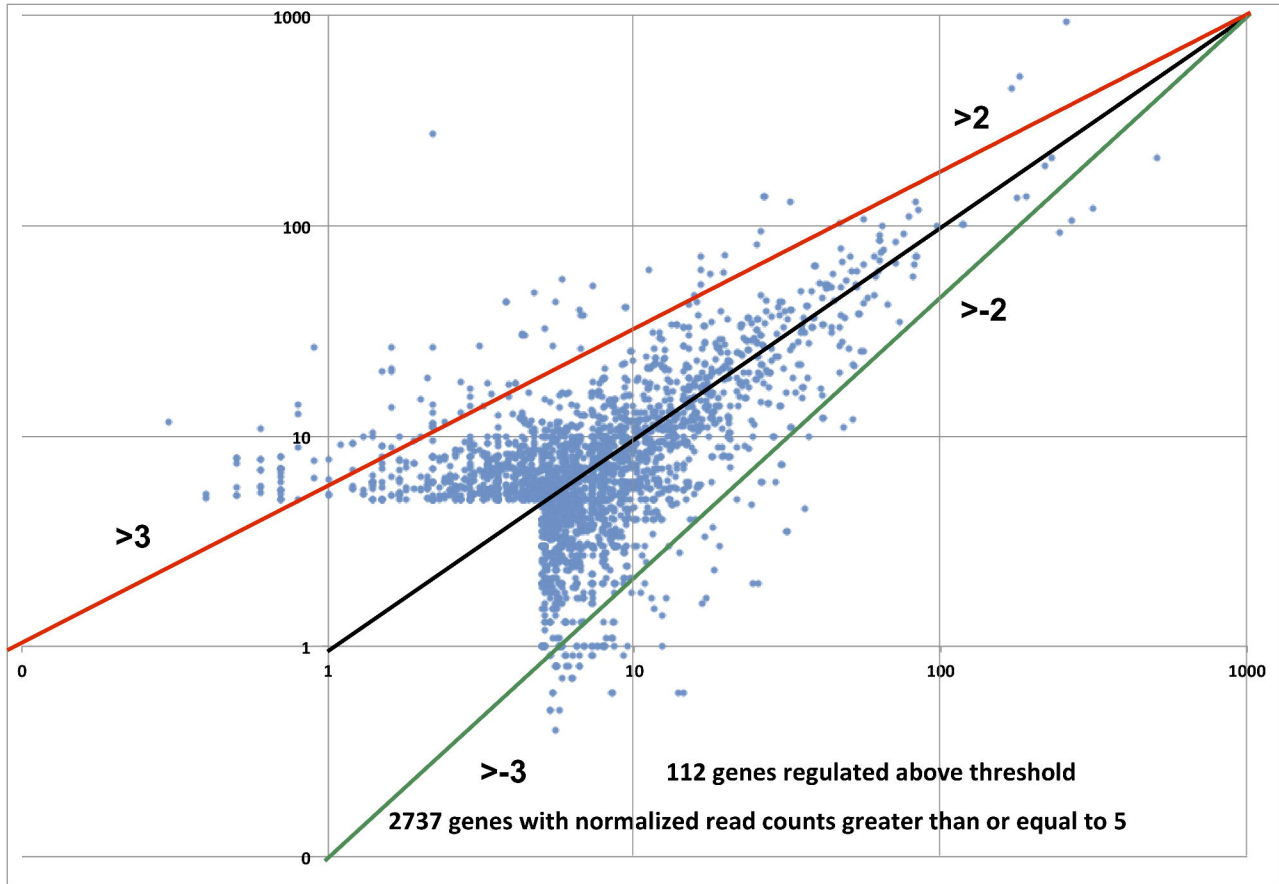








**Induced (unique read counts normalized)**



**Control (unique read counts normalized)**

**>3**

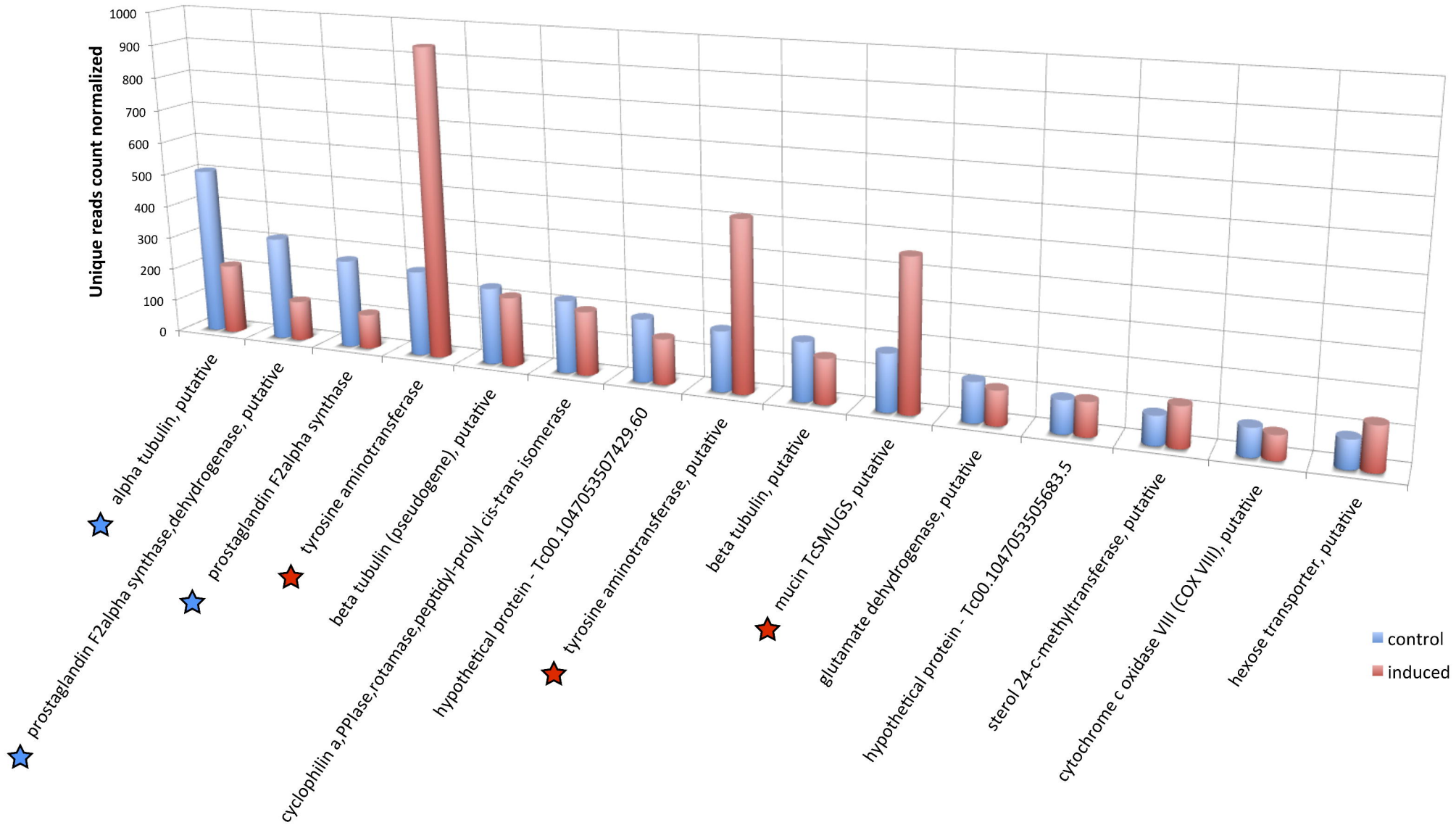
**>2**

**>-2**

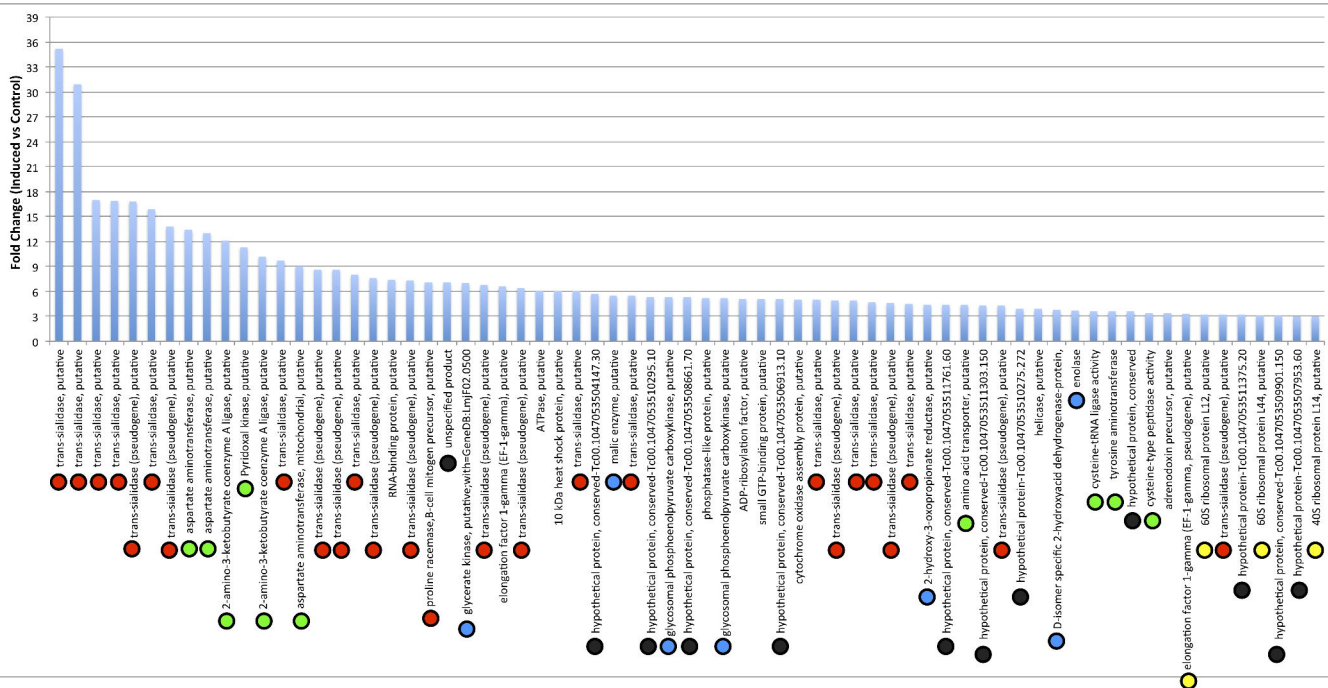
**>-3**

**112 genes regulated above threshold**

**2737 genes with normalized read counts greater than or equal to 5**

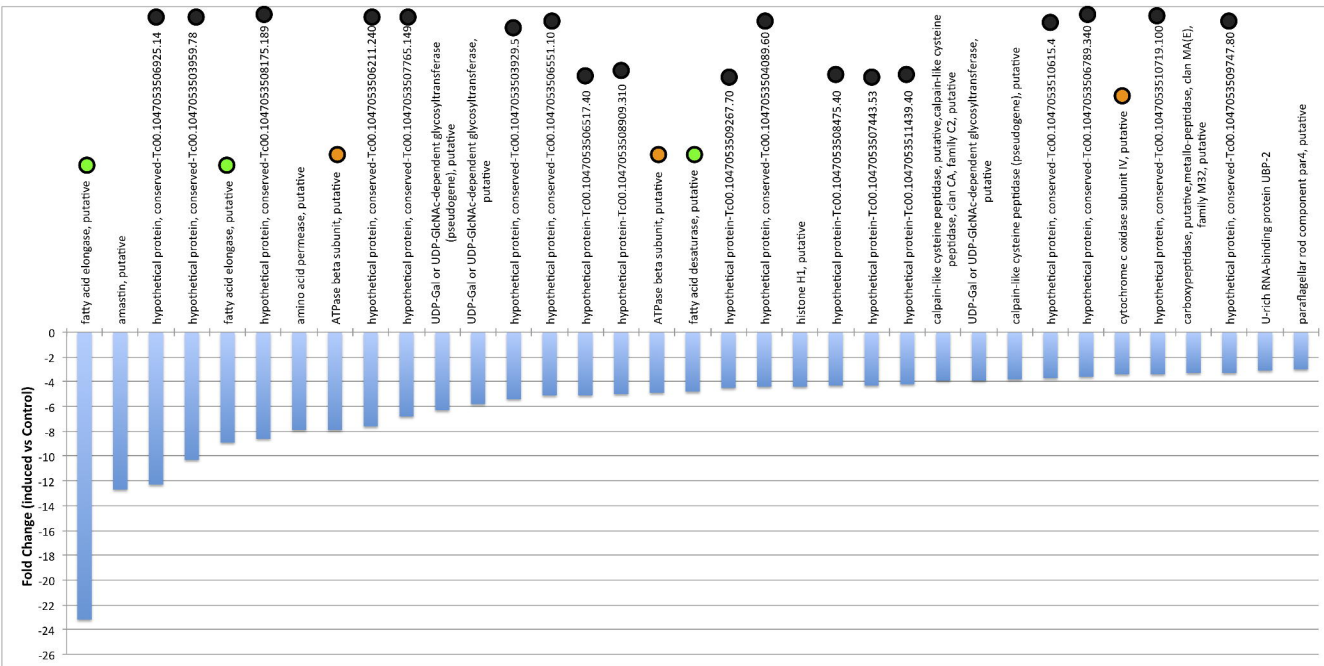


## Upregulation in ZFP1 induced (>3 fold Change)



- Infectivity and differentiation
- Hypothetical proteins
- Carbohydrate metabolism
- Amino acid metabolism
- Ribosomal function

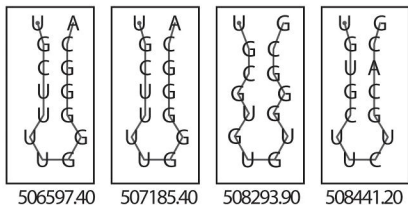
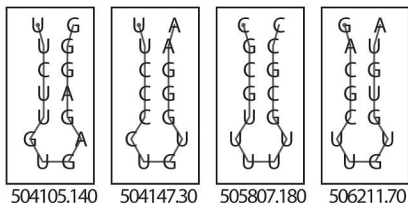
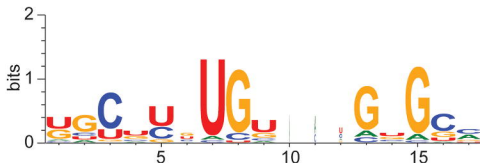
## Downregulation in ZFP1 induced (>3 fold Change)



- Hypothetical proteins
- Fatty acid biosynthesis
- Mitochondrial function

**a**

## Up-motif



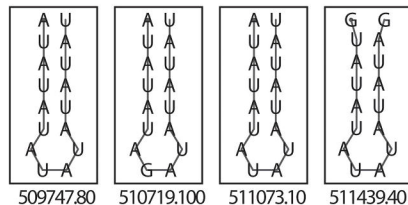
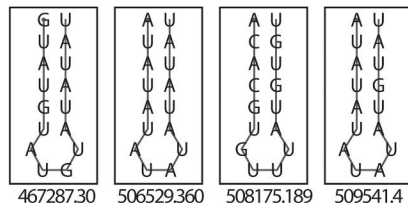
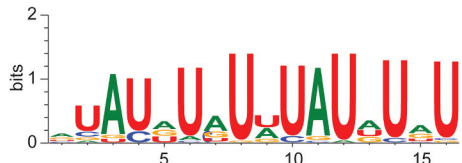
```

504105.140 UUCUUGUGA...GAGGG
#=SS      <<<<<.....>>>>>
504147.30  UUCCCCUGU...GGGAA
#=SS      <<<<<.....>>>>>
505807.180 CGCGUUUUU...GCGCC
#=SS      -<<<<<.....>>>>-
506211.70  GACGCUUG...GUGUA
#=SS      -<<<<<.....>>>>-
506597.40  UGCUUUUG...GGGCA
#=SS      <<<<<.....>>>>>
507185.40  UGCUUUUG...GGGCA
#=SS      <<<<<.....>>>>>
508293.90  UGCGUGUGU...GGGCG
#=SS      <<<<-<.....>>>>->
508441.20  UGUGCUUCU...GCACG
#=SS      <<<<<.....>>>>>
consensus  GGCcCcUGu...GcGCC
#=SS_cons  <<<<<____>>>>>

```

**b**

## Down-motif



```

467287.30  GU AUGUAUGUAUAUAU
#=SS      <<<<<<.....>>>>>
506529.360 AUUAUAUAUAUAUAU
#=SS      <<<<<<.....>>>>>
508175.189 ACACGUGUUUAUGUGU
#=SS      <<<<<<.....>>>>>
509541.4   AUUAUAUAUAUAUGUAU
#=SS      <<<<<<.....>>>>>
509747.80  AUUAUAUAUAUAUAUAU
#=SS      <<<<<<.....>>>>>
510719.100 AUUAUAUAGAUUAUAUAU
#=SS      <<<<<<.....>>>>>
511073.10  AUUAUAUAUAUAUAUAU
#=SS      <<<<<<.....>>>>>
511439.40  GUUAUAUAUAUAUAUAG
#=SS      -<<<<<<.....>>>>-
consensus  aUAUaUAUaUAUAUAU
#=SS_cons  <<<<<<.....>>>>>

```