# Four layer multi-omics reveals molecular responses to aneuploidy in *Leishmania*

Cuypers B.[1,2], Meysman P.[2], Erb I.[3], Bittremieux W.[4], Valkenborg D.[5,6], Baggerman G.[6,7], Mertens I.[6,7], Shyam Sundar[8], Basudha Khanal[9], Notredame C.[3], Dujardin, J-C.[1,10#], Domagalska M.A.[1*#], Laukens K.[2*#]

1. Molecular Parasitology Unit, Department of Biomedical Sciences, Institute of Tropical Medicine, Antwerp, Belgium.

2. Adrem Data Lab, Department of Computer Science, University of Antwerp, Antwerp, Belgium.

3. Centre for Genomic Regulation (CRG), The Barcelona Centre of Science and Technology, Barcelona, Spain.

4. Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA 92093, USA.

5. Center for Statistics - University of Hasselt, Hasselt, Belgium.

6. Center for Proteomics, University of Antwerp, Antwerp, Belgium.

7. VITO, Mol, Belgium.

8. Institute of Medical Sciences, Banaras Hindu University, Varanasi, India

9. BP Koirala Institute of Health Sciences, Dharan, Nepal

10. Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium.

* These authors share senior authorship

# Corresponding authors

# Abstract

Aneuploidy causes system-wide disruptions in the stochiometric balances of transcripts, proteins, and metabolites, often resulting in detrimental effects for the organism. The protozoan parasite *Leishmania* has an unusually high tolerance for aneuploidy, but its molecular consequences remain poorly understood. Here, we present the first integrated analysis of the genome, transcriptome, proteome, and metabolome of highly aneuploid *Leishmania donovani* strains. Our analysis unambiguously establishes that, despite this tolerance, aneuploidy in *Leishmania* globally and proportionally impacts the transcriptome and proteome of affected chromosomes, ultimately explaining the degree of metabolic differences between strains. This impact was present in both proliferative and infectious life stages. Protein complex subunits, secreted proteins, and non-cytoplasmic proteins, responded less or even not at all to aneuploidy-induced dosage changes. In contrast to other Eukaryotes, widespread transcript regulation modulating aneuploidy-related stress was absent. Further, the majority of regulated proteins was encoded by non-aneuploid chromosomes and had no underlying transcript change, suggesting that aneuploidy induces extensive post-transcriptional protein-level modulation. This makes *Leishmania* a unique Eukaryotic model for elucidating post-transcriptional protein-abundance modulation in the context of aneuploidy.

# Introduction

Aneuploidy is the presence of an abnormal number of chromosomes in a cell. All genes encoded by aneuploid chromosomes undergo a shift in their copy number or "gene dosage". This large number of unbalanced dosage changes typically leads to broad-scale disruptions in the stoichiometric balances of transcripts, proteins, and metabolites, making aneuploidy most widely known for its detrimental effects. Indeed, aneuploidy can cause oxidative stress, energetic stress, misfolded protein stress (proteotoxicity), and metabolic stress at the cellular level [1,2]. In multicellular organisms, where cells are closely fine-tuned to work together, this often leads to growth delays or lethality [1].

Contrastingly, aneuploidy can also result in fitness gains. In fungi, specific karyotypes confer resistance against drugs, toxicants, or increase survival under nutrient limiting conditions [3-6]. Likewise, in cancer cells, aneuploidy can be tumor growth-promoting [7]. By altering the copy number of many genes at once, aneuploidy enables cells to explore a wide fitness landscape [8]. Particularly when facing very novel or competitive conditions, aneuploidy might allow the rapid evaluation of many divergent phenotypes [9]. Fully understanding this complex trade-off between aneuploidy fitness gains and losses requires a systems biological comprehension of its molecular impact.

In this context, an organism with a remarkably high tolerance for aneuploidy is the protozoan parasite *Leishmania* (Euglenozoa, Kinetoplastida). Transmitted by its sand-fly vector, *Leishmania* causes the disease leishmaniasis in humans and various vertebrate animals. The parasite displays exceptional genome plasticity in experimental conditions. A genomic analysis of 204 *in vitro* cultured *Leishmania donovani* strains uncovered up to 22 aneuploid chromosomes (out of 36) within a single strain [10]. Moreover, all 36 chromosomes have the capacity to become aneuploid and lead to viable parasites [11]. Strikingly, parasite populations do not have uniform karyotypes but display a mosaic of them. Single-cell sequencing of 1560 *L. donovani* parasites from a clonal population identified 128 different karyotypes, of which four were dominant [12]. This genomic instability and mosaicism could explain the parasite's ability to rapidly select more advantageous karyotypes upon encountering novel *in vivo* or *in vitro* environments [13].

*Leishmania* also features an unusual gene expression system that is markedly different from most other Eukaryotes. Transcription of protein-coding genes cannot be controlled individually as these lack separate promoters. Instead, they are transcribed constitutively in long polycistronic units that span 10's to 100's of functionally unrelated genes [14]. The abundance of individual mRNAs and proteins is

68 still regulated, but entirely post-transcriptionally [13,15]. Thus, the most straightforward evolutionary
69 path towards altered gene expression might be altered gene dosage, which could explain the high
70 incidence of aneuploidy [13]. While aneuploidy in *Leishmania* is known to directly affect the transcription
71 of encoded transcripts, its molecular impact on the proteome and metabolome remains poorly
72 understood [13,16].

73 Altogether, *Leishmania* provides a unique perspective into the molecular impact of aneuploidy as (1)
74 a system where transcriptional control of individual protein-coding genes is absent, and as (2) an
75 organism with a remarkably high aneuploidy tolerance. Here we present the first study that generated
76 and integrated genome, transcriptome, proteome, and metabolome data of highly aneuploid *L.*
77 *donovani* strains. This multi-omics analysis unambiguously demonstrates that aneuploidy in
78 *Leishmania* globally and drastically impacts the transcriptome and proteome of affected
79 chromosomes, ultimately explaining the degree of metabolic differences between strains. This impact
80 appears to be present throughout the parasite's life cycle as we observed it in both the parasite's
81 proliferative and infectious life stages. Interestingly, not all proteins responded equally to dosage
82 changes; through dosage compensation, particular proteins such as protein complex subunits,
83 secreted proteins, and non-cytoplasmic proteins responded less or even not at all to aneuploidy.
84 Interestingly, we did not observe the widespread regulation at the transcript level that typically
85 modulates the effects of aneuploidy in other Eukaryotes. Instead, *Leishmania* features a surprisingly
86 high number of regulated proteins from non-aneuploid chromosomes, which are likely post-
87 transcriptionally modulated.
88

# Methods

## Parasites

The experiments in this study were performed on six cloned strains of *Leishmania donovani*: BPK275/0 cl18 (BPK275), BPK173/0 cl3 (BPK173), BPK282 cl4 (BPK282), BPK288/0 cl7 (BPK288), BPK178/0 cl3 (BPK178), BHU575/0 cl2 (BHU575) (**Supplementary Table S1.A**). A strain is a parasite population derived by serial passage *in vitro* from a primary isolate (from patient) with some degree of characterization, in our case bulk genome sequencing (definition modified from [17]). These strains were cloned to ensure genomic homogeneity. Parasites were cultured in HOMEM (Gibco) supplemented with 20% heat-inactivated fetal bovine serum (Biochrom), three days for replicative forms (LOG growth phase, 4-6 replicates) and six days for non-replicative forms (STAT, 4 replicates). Parasites from all replicates were washed twice by centrifugation at 18,000 x g at 0 °C, resuspension in ice-cold PBS buffer (PBS; pH 7.4 – Invitrogen), and pelleted at 18,000 x g at 0 °C. DNA, RNA, protein, and metabolite extractions were carried out in parallel (see further).

## DNA/RNA extraction, library preparation, and sequencing

DNA and RNA were extracted from the parasite pellets with the AllPrep DNA/RNA Mini Kit (Qiagen). The RNA integrity was then verified with the RNA 6000 Nano Kit using a Bioanalyzer 2100 (Agilent). DNA was quantified with the Qubit dsDNA BR assay (Thermo Fisher Scientific) and RNA with the Qubit RNA HS Assay.

Library preparation and sequencing of the DNA samples were performed at the Beijing Genomics Institute (BGI). Libraries were prepared with the TruSeq Nano DNA HT sample prep kit (Illumina) and 2 x 151 bp sequenced on the Illumina HiSeq 4000 platform. RNA sequencing libraries were prepared using the Spliced leader sequencing (SL-Seq) method as described in Cuypers et al. (2017) [15]. This protocol makes use of the presence of the affixed 39 nucleotide sequence spliced-leader (SL) that is present at the 5' end of all functional *Leishmania* mRNAs. RNAs containing a SL are selectively amplified with the protocol, and adapters required for Illumina sequencing are ligated. The SL-Seq libraries were 1 X 50 bp sequenced with the HiSeq 1500 platform of the Center of Medical Genetics Antwerp (Belgium).

## Genome and Transcriptome data analysis

The quality of the raw sequencing data was verified with FASTQC v0.11.4 [18]. Reads were quality trimmed using Trimmomatic v035 [19] with the settings 'LEADING:20 TRAILING:20

119 ILLUMINACLIP:adapter.fa:2:40:15 SLIDINGWINDOW:4:20 MINLEN:36'. Reads were subsequently
120 aligned to the *L. donovani* LdBPKv2 [20] reference genome using BWA Mem version 0.7.15 [21] with default
121 parameters. The resulting bam files were sorted and indexed with samtools [22]. Mapping reports were
122 generated using samtools flagstat.

123 Somy estimations and gene copy numbers were calculated as described in Downing (2011) [23]. Briefly,
124 the somy value of a chromosome (S) was calculated by dividing its median sequencing depth (i.e. the
125 number of times a nucleotide is read during sequencing, $d_{ch}$) by the median $d_{ch}$ of all 36 chromosomes
126 ($d_{mch}$) and multiplying this value by two (for a predominantly diploid organism). Gene copy number per
127 haploid genome (d) was defined as a raw depth for that gene ($d_r$), divided by the median depth of its
128 chromosome ($d_{ch}$): so that $d = d_r/d_{ch}$. Gene copy number estimations per cell (full depth, gene dosage)
129 were then calculated by multiplying d with S. SNPs were previously characterized in [10]. RNA-Seq gene
130 read count tables were generated with the HTSeq tool version 0.6.1 [24]. Count data were normalized
131 for library size by dividing each transcript count by the geometric mean of all counts for that library [25].
132 Differential gene expression between LOG and STAT life stages was calculated with DESeq2 v1.30.1[26].

133 **Proteomics**

134 *Sample preparation*

135 Cell pellets were efficiently lysed using 200 µl RIPA buffer and 1x HALT protease inhibitor (Thermo
136 Scientific), combined with a 30 sec sonication (Branson Sonifier SLPe ultrasonic homogenizer,
137 Labequip, Ontario, Canada) with an amplitude of 50% on ice. After centrifugation of the samples for
138 15 min at 10,000 g at 4 °C, the cell pellet was discarded. Next, the protein concentration was
139 determined using the Pierce BCA protein Assay kit in combination with a NanoDrop 2000
140 photospectrometer (Thermo Scientific).

141 For each sample, 25 µg of proteins were reduced using 2 µl of 200 mM tris(2-carboxyethyl) phosphine,
142 in a volume of 20 µl 200 mM triethylammonium bicarbonate (TEAB), and incubated for 1 h at 55°C.
143 After alkylation of the proteins with 2µL of 375 mM iodoacetamide for 30 min protected from light, 6
144 volumes of ice-cold acetone were added, and the samples were incubated overnight at -20°C. The next
145 day, samples were centrifuged for 10 min at 10.000 g at 4°C, the acetone was removed, and pellets
146 were resolved in 20 µl of 200 mM TEAB. Proteins were then digested with trypsin (Promega) overnight
147 at 37°C with an enzyme trypsin ratio of 1:50. Before LC-MS/MS analysis, the samples were desalted
148 with Pierce C18 spin columns according to the manufacturer's instructions (Thermo Scientific).

149 *Nano reverse-phase liquid chromatography and mass spectrometry*

150 Each of the digested samples was separated by nano reverse phase C18 (RP-C18) chromatography on
151 an Easy-nLC 1000 system (Thermo Scientific, San Jose, CA) using an Acclaim C18 PepMap®100 column
152 (75 µm x 2 cm, 3 µm particle size) connected to an Acclaim PepMap™ RSLC C18 analytical column (50
153 µm x 15 cm, 2 µm particle size) (Thermo Scientific, San Jose, CA). Of each sample, a total of 1µg of
154 peptides were loaded on the column. Before loading, digests were dissolved in mobile phase A,
155 containing 2% acetonitrile and 0.1% formic acid, at a concentration of 1µg/10µL and spiked with 20
156 fmol Glu-1-fibrinopeptide B (Glu-fib, Protea biosciences, Morgantown, WV). A linear gradient of mobile
157 phase B (0.1% formic acid in 100% acetonitrile) from 0 to 45% in 90 min, followed by a steep increase
158 to 100% mobile phase B in 10 min, was used at a flow rate of 300 nL/min. Liquid Chromatography was
159 followed by MS (LC-MS/MS) and was performed on a Q-Exactive Plus mass spectrometer equipped
160 with a nanospray ion source (Thermo Fisher, Waltham, MA, USA). The high-resolution mass
161 spectrometer was set up in an MS/MS mode where a full scan spectrum (350 – 1850 m/z, resolution
162 70,000) was followed by a high energy collision activated dissociation (HCD) tandem mass spectra (100
163 – 2000 m/z, resolution 17,500). Peptide ions were selected for further interrogation by tandem MS as
164 the twenty most intense peaks of a full scan mass spectrum. The normalized collision energy used was
165 set at 28%. A dynamic exclusion list of 20 sec for the data-dependent acquisition was applied.

166 *Proteomic Data Analysis*

167 Thermo raw files were converted to mzML files using MSConvert v3.0. Label-free protein quantification
168 (LFQ) was carried out with MaxQuant version 1.6.0.16 [27] using the following settings (other settings
169 were kept default): Oxidation (M) and acetyl (Protein N-term) were indicated as variable modifications,
170 carbamidomethyl (C) was indicated as a fixed modification, digestion with trypsin with maximum two
171 missed cleavages, match between runs = yes, dependent peptides = yes. The search database was the
172 LdBPKV2 proteome previously published by our group [13], and reversed proteins were used as decoys.

173 **Metabolomics**

174 *Sample preparation*

175 Sample pellets were extracted using a 1:3:1 (v:v:v) chloroform/methanol/water extraction, subjected
176 to liquid chromatography using a 2.1 mm ZIC-HILIC column (Sequant), and analyzed using the Orbitrap
177 Exactive (Thermo Fisher Scientific) platform at the Glasgow Polyomics Center (University of Glasgow,
178 Glasgow, Scotland) exactly as described elsewhere [28,29]. Both negative and positive ion modes were

179 run in parallel with rapid polarity switching. Samples were analyzed in randomized order and the same
180 analytical batch. Additionally, to aid accurate metabolite identification, verify LC-MS stability, and
181 identify contaminants, the following control samples were included as well: (1) solvents blanks, (2)
182 serial dilutions of a pooled sample (Undiluted, 1/2, 1/4, 1/8, and 1/16), (3) Authentic standard mixes
183 containing in total 217 metabolites (50–400 Da) and representing a wide array of metabolic classes
184 and pathways, (4) An amino acid standard mix (Sigma Product No. A9906).

185 *Metabolomic data analysis*

186 Data preprocessing was carried out with the XCMS 1.42.0 [30] and mzMatch 1.0.1 [31] packages in R,
187 exactly as described elsewhere [32]. Briefly, raw spectra (mzXML format) were first subjected to
188 retention time alignment with ObiWarp [33], after which peak detection was carried out with the
189 centWave algorithm [34]. Peaks were then filtered: 1) Corresponding peaks from the four biological
190 replicates were allowed a maximum reproducibility standard deviation (RSD) of 0.5, 2) Peak shape
191 (CoDA-DW > 0.8), 3) Detection in at least 3 out of 4 replicates of a single strain and 4) Minimal peak
192 intensity was set to 3000. The peaks were putatively identified using sequentially the LeishCyc
193 database [35], LIPID MAPS [36], KEGG [37], the peptide database included in mzMatch, and the Human
194 Metabolome Database [38]. Finally, the filtered peaks were normalized for total ion count and subjected
195 to manual quality screening.

196 **Integrated multi -omic analysis**

197 For each comparison between two strains, we calculated the $Log_2$ fold change ($Log_2FC$) for genes (gene
198 dosage), transcripts, proteins and metabolites. These $Log_2FC$ values were obtained by taking the $Log_2$
199 ratio of the average abundance between two strains. Associated *p*-values were obtained by performing
200 a Student *t*-test for each comparison and subsequently correcting it with the Benjamini-Hochberg
201 algorithm to limit the False Discovery Rate (FDR) to 5%. A Log2FC cutoff of 1 and adjusted *p*-value
202 cutoff of 0.05 was used for all 'omic layers, but for gene dosage, an additional criterium was applied;
203 only copy number differences of at least 0.5/haploid genome were considered as biologically
204 significant [39]. All plots were generated using the ggplot v 2_2.1.0 package [40], except the Circos plots
205 which were created with Circos v0.69 [41]. Gene ontology enrichments were calculated with the Fisher
206 Exact test using Python 3.8.5 and the matplotlib v3.3.4 [42] and SciPy v1.6.1 [43] libraries. GO annotations
207 were obtained from the gene ontology consortium (http://www.geneontology.org/).

208 **Dosage compensation analysis**

209 The following protein features were checked for their relation with protein dosage compensation:
210 degree (i.e., number of protein interactions as transferred from STRING), interacting (yes/no), stress
211 centrality (number of shortest paths between all node pairs in the protein-protein interaction network,
212 that pass through the specific protein), length (as a proxy for transcript and protein molecular size),
213 GRAVY (grand average of hydropathy), solubility (soluble versus membrane protein) and subcellular
214 location. We added transcript abundance as a covariate to the model, as the compensatory effect can
215 be expected to become larger at higher differential transcript levels.

216

# Results

## Aneuploidy globally impacts transcriptome and proteome of affected chromosomes

We examined the impact of aneuploidy in *L. donovani* on transcriptome and proteome by integrated multi-omic analysis of 6 *L. donovani* strains from India and Nepal (**Suppl. Table S1.A**). These closely related strains are almost identical at the sequence level but differ greatly in their aneuploidy [44]. Specifically, the entire genome between any two strains of this set differs at most by 64 non-synonymous SNPs and one indel in coding regions (**Suppl. Table. S1.B)**. Thus, they are attractive model strains for studying the impact of aneuploidy on transcriptome and proteome with minimal interference of DNA sequence variation.

Genome, transcriptome, and proteome were obtained from the replicative promastigote stage (LOG) of the parasite. All genomes reached an average sequencing coverage of at least 26.5 fold, and 97.9 % of the sequenced sites reached 10 X coverage or more. Transcript samples had on average 4.71 million reads and obtained an average mapping efficiency of 97%. Detailed mapping and coverage statistics for both genome and transcriptome data are available in **Suppl. Table S1.C/S1.D.** Identified indels and and estimations of chromosomal somy (i.e., chromosome copy number) and dosage of each gene in **Suppl. Table S2**. Identified transcripts and their counts are available in **Suppl. Table S3**. Proteome analysis identified in total 3145 proteins across all samples (**Suppl. Table S4**).

Our 6 *L. donovani* strains showed high but varying degrees of aneuploidy (**Fig 1A**). The most aneuploid strain was BPK288 with 12 aneuploid chromosomes, followed by BPK275 (10), BPK173 (10), BHU575 (10), BPK178 (9), and BPK282 (6). In *L. donovani,* chromosomes are numbered by increasing size. The baseline somy of chromosomes in *L. donovani* is 2N, except for chromosome 31, which is always 4N or more. Smaller chromosomes were more frequently aneuploid (Ld01, Ld05-Ld09, Ld11-Ld16), than larger chromosomes (Ld20, Ld23, Ld26, Ld31, Ld33, Ld35). Chromosome Ld31 was tetrasomic in all strains, and chromosome Ld23 always trisomic. Other aneuploid chromosomes had a variable somy. In bulk genomic analysis, somy values represent the average of the somy of individual cells present in a given population. Because of mosaicism, i.e. the presence of different kayotypes among individual cells of the parasite population in culture, many somy values were not discrete. The highest somy values (apart from Ld31) were observed for chromosomes Ld02 (somy = 3.9), Ld08 (3.9), and Ld33 (3.5) in BPK275, and for chromosome Ld05 (3.6) in BPK288.

246 The transcript abundance of aneuploid chromosomes mirrored almost perfectly their underlying somy
247 (**Fig 1B, Suppl. Fig S1A, S1B**). Specifically, the average abundance of a transcript encoded by a
248 chromosome was proportional with a factor of 0.94 with the chromosomal somy (**Fig 1C**: 95% CI: 0.92
249 – 0.96 and $p < 2 \cdot 10^{-16}$, $R^2$: 0.98). The average abundance of the proteins encoded by a chromosome
250 was also highly proportional to the aneuploidy of the chromosome ($p < 2 \cdot 10^{-16}$), albeit with a
251 decreased slope of 0.74 (95% CI: 0.69 - 0.76, $R^2$: 0.90). Thus, aneuploidy results in a linear, but not
252 equivalent, increase in protein abundance. Circos plots clearly showed that transcript- and protein
253 abundance of aneuploid chromosomes was affected globally and not restricted to specific
254 chromosomal regions (**Fig 1B, Suppl. Fig S1A, S1B**).

255 Interestingly, chromosome 35 exhibited only a partial amplification in strain BPK275, the first part
256 being disomic (somy = 2.2) and the second part, starting from gene LdBPK_350040900, being trisomic
257 (somy = 3.1) **(Fig 1B).** This phenomenon of partial chromosomal amplification has previously been
258 observed for chromosome 23 as well [45,46]. The difference in both average transcript and protein
259 abundance between the first and the second chromosome part was significant (t-test *p* Trans = $4.0 \times 10^{-}$
260 $^{82}$, Prot = $7.0 \times 10^{-7}$), the second part having on average 1.43x more transcript abundance and 1.22x
261 more protein abundance. This demonstrates that also partial chromosomal amplifications affect
262 transcript and protein abundance proportionally in the amplified region.

263

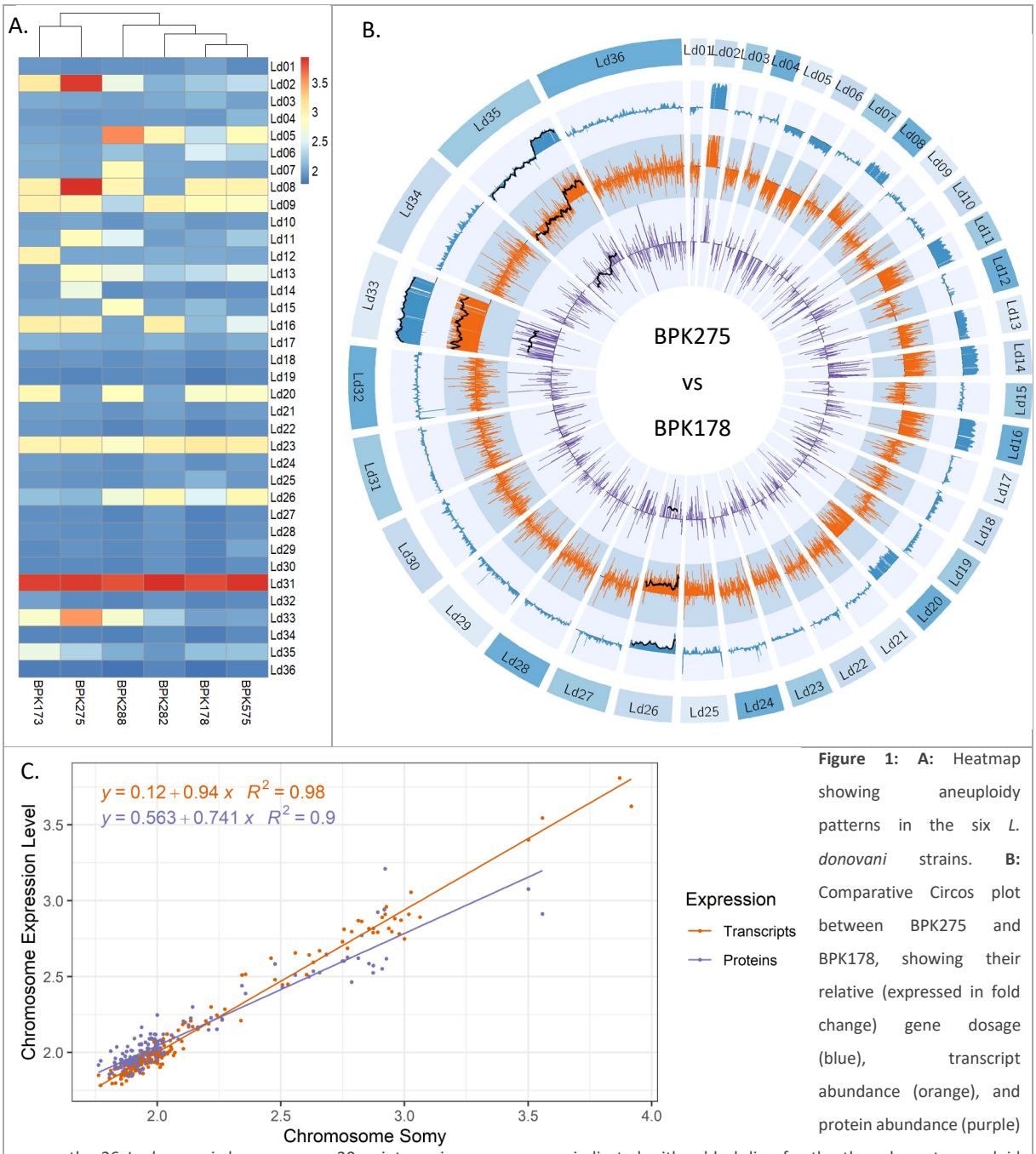



**Figure 1: A:** Heatmap showing aneuploidy patterns in the six *L. donovani* strains. **B:** Comparative Circos plot between BPK275 and BPK178, showing their relative (expressed in fold change) gene dosage (blue), transcript abundance (orange), and protein abundance (purple) across the 36 *L. donovani* chromosomes. 20 point moving averages are indicated with a black line for the three largest aneuploid chromosomes. We show this strain pair because of the partial chromosomal amplification on chromosome Ld35. However, Circos plots between BPK173 and BPK282, and BHU575 and BPK288 show similar agreement between genome, transcriptome, and proteome and can be found in supplementary figures (Suppl. Fig S1A, S1B). **C**: Impact of chromosome aneuploidy on average chromosome transcript (red) and protein (blue) abundance levels. Each dot represents 1 chromosome in one of the six studied aneuploid *L. donovani* strains in the replicateive (LOG) phase. Abundance levels were normalized so that two transcript/protein abundance units match the abundance level of a disomic chromosome. Somy values are often not discrete due to mosaic aneuploidy (i.e. presence of different kayotypes among individual cells present in 1 strain) in *Leishmania*.

264 **Aneuploidy does not impact all proteins equally: dosage compensation in *Leishmania***

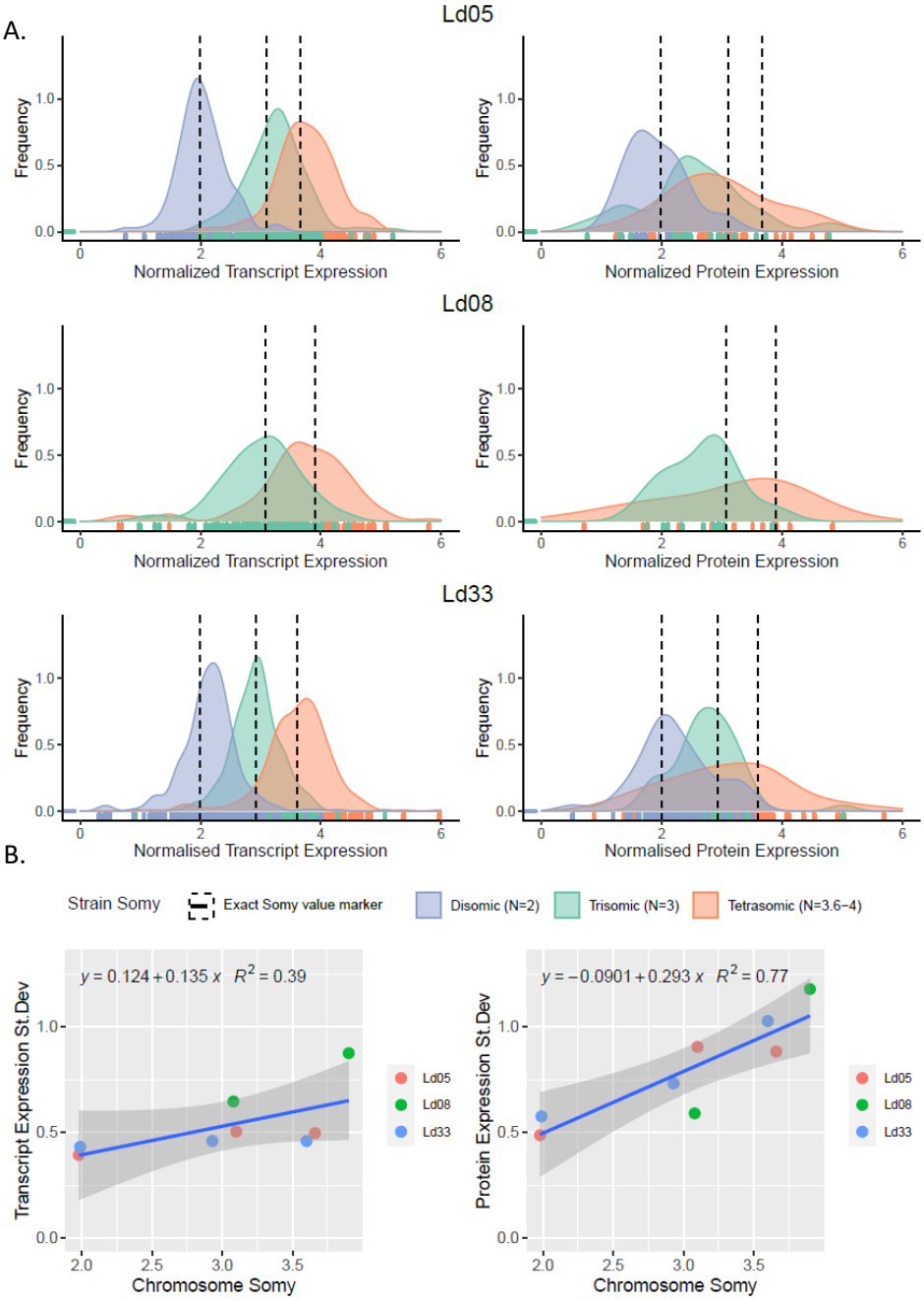
265 Our observation that a somy increase does not lead to an equivalent increase in protein abundance,
266 but only 0.74x the somy change, suggested either of two possible explanations. A first possibility is that
267 all proteins encoded by an aneuploid chromosome are modulated equally and exhibit only a partial
268 protein abundance increase. We further refer to this explanation as the 'dampening hypothesis'. Such
269 a general 'dampening' effect would suggest an underlying process related to intrinsic mechanisms or
270 limitations in the parasite's translation system. Alternatively, different proteins could respond
271 differently to increased gene dosage and transcript abundance, pointing to compensatory effects
272 linked to the identity, function, and physicochemical properties of those specific proteins or certain
273 functional protein groups. We refer to this explanation as the 'compensation hypothesis'.

274 To evaluate both hypotheses, the differential abundance patterns of proteins were compared to their
275 encoding chromosome differences from disomy to trisomy and tetrasomy. Suppose all proteins of a
276 chromosome are affected equally by a shift in somy, following the 'dampening hypothesis'. In this case,
277 the average of the protein abundance distribution should increase with increased somy, but the
278 distribution shape should remain unaltered. In our study, three chromosomes (Ld05, Ld08, and Ld33)
279 were at least once disomic, trisomic, and tetrasomic in one of our six study strains (**Fig 1B**), allowing
280 for such comparative analysis. Ld02 also matched these criteria but was excluded from the analysis as
281 only nine Ld02 proteins were detected. For reference, the same analysis was conducted for transcripts.

282 The average abundance of transcripts increased with somy, and the shape of the abundance
283 distribution was found to be similar (**Fig 2A**). Specifically, the transcript abundance standard deviation
284 did not significantly change with somy (Regression *p*: 0.10, **Fig 2B**), and the distributions were
285 symmetric. This suggests that a somy change affects most transcripts equally, and dosage
286 compensation at the transcript level is absent. Chromosome Ld08 had a higher variance in transcript
287 abundance than the two other chromosomes. Our data is not conclusive as to whether this is due to
288 the shift in somy or due to an intrinsically higher transcript abundance variance for this chromosome.

289 The average abundance of proteins also increased with somy, but the distribution shape was not
290 consistent. This was particularly evident when comparing the disomic versus the tetrasomic
291 chromosomes. We observed a clear left-tailing of the protein abundance distribution on chromosomes
292 Ld08 and Ld33 chromosomes towards the lower, more conservative fold change values. Thus, while a
293 subset of proteins follows the aneuploidy-induced gene dosage changes, another subset is maintained

294 at conservatively lower levels. This dosage compensation was also reflected in a general and significant
295 increase in protein abundance variance for all three chromosomes (Regression *p*: 0.004).




296

297 **Figure 2**: **A**: Normalised transcript (left) and protein (right) abundance patterns of disomic Ld05, Ld08, and Ld33 in their disomic (excl. chr
298 8), trisomic and tetrasomic state. Vertical dotted lines indicate the corresponding somy, derived from the genomic data. The small vertical
299 lines at the base of the plot indicate individual datapoints for transcripts and proteins, from which the distributions were generated. The
300 abundance of transcripts and proteins was normalized to their abundance in another strain that was disomic for their encoding
301 chromosome (normalized abundance at disomy = 2). This is the reason why we do not have disomic distributions for Ld08 since Ld08 was
302 exclusively disomic in BPK282. **B.** Regression between the somy and transcript (left) or protein (right) standard deviation for chromosomes
303 Ld05, Ld08, and Ld33 in different somy states.

## Compensated proteins are primarily protein complex subunits, secreted or non-cytoplasmic

We investigated if specific protein properties were associated with the dosage compensation observed in our dataset. For this analysis, we defined a protein's compensation level as its $Log_2$ transcript fold change (i.e., its transcript level on the aneuploid chromosome divided by its transcript level on disomic chromosome), minus its $Log_2$ protein fold change. A particular challenge in this analysis was that protein fold changes, and hence, the observed compensation levels, were generally very small and close to measurement noise levels. For example, for a fully compensated protein from a trisomic chromosome, the compensation level would only be 0.5 $Log_2$FC units ($Log_2$FC transcript ≈ 0.5, $Log_2$FC protein = 0). Therefore, we focused on a set of 85 transcripts and proteins, which had very accurate protein measurements (SD between biological replicates < 0.25 $Log_2$FC units). These covered a broad range of compensation levels ranging from 1.18 $Log_2$FC units, to values close to zero, or even some proteins that had a slight overamplification compared to the transcript level (-0.40 $Log_2$FC units) (**Suppl. Table S1.E**).

We noted that within the top 5 most compensated proteins in this high-accuracy set, 3 coded for subunits of protein complexes: e.g., the cleavage and Polyadenylation Specificity Factor (CPSF) subunit A, the dynein light chain, and a Qc-SNARE protein had respectively 1.12, 0.96, and 0.77 Log2FC units lower protein than transcript abundance. Accordingly, we hypothesized that protein complexes might be more likely to be dosage compensated. To test this systematically, all 85 proteins were screened for their involvement in heteromeric protein complexes. 47 proteins could be annotated (subunit/not a subunit) of which 16 were compensated (>= 0.30 Log2FC units), and 31 not (< 0.30 Log2FC units) (**Suppl. Table S1.E**). Strikingly, complex subunits accounted for 50.0 % (8 out of 16) of the compensated proteins which was significantly more than the 6.5% (2 out of 31) we observed for the uncompensated proteins ($Chi^2$ $p$ = 0.00034). For the majority of compensated subunits (6 out of 8), we could identify another subunit from the same complex, located on a different chromosome.

Next, we looked if there was a relation between dosage compensation and a set of biological and physicochemical protein properties to identify more subtle patterns. The following protein properties were checked for their relationship with protein dosage compensation using linear regression: the number of protein interactions, length (a proxy for transcript and protein size), hydrophobicity, membrane protein (yes/no), and subcellular location. Transcript abundance was added as a covariate to the model, as dosage compensation was larger at higher differential transcript levels ($p = 5.5 \times 10^{-4}$).

335 The degree of dosage compensation of a protein was dependent on its cellular compartement **(Suppl.**
336 **Table S1.E.** Cytoplasmatic proteins (n = 39) were significantly less compensated compared to all other
337 proteins ($p$ = 0.0054). Thus, cytoplasmatic proteins follow aneuploidy-induced dosage changes more
338 closely than proteins from other cellular compartments. In contrast, proteins associated with the Golgi
339 apparatus (n = 2) had a significantly higher compensation than all other proteins ($p$ = 0.048). These two
340 proteins were a vesicle-associated membrane protein and a small GTP-binding protein Rab18, both
341 known to be secreted. Proteins from mitochondria (n = 18), nucleus (n = 13), peroxisome (n = 6) and
342 other (n = 5) did not have significantly different compensation rates, and hence showed an
343 intermediate level of dosage compensation. Other tested protein properties did not show a significant
344 relation with dosage compensation.

## 345 The impact of aneuploid chromosomes on the gene products of their euploid 346 counterparts

347 In the previous sections, we reported how the aneuploidy of a chromosome affects the abundance of
348 its encoded transcripts and proteins. However, through perturbation of regulatory networks, these
349 changes might also affect transcripts and proteins encoded by genomic regions unaffected by
350 aneuploidy. These were previously defined as 'trans-effects' and have been observed on the
351 transcriptional level in diverse species, including humans, yeast, maize, and *Arabidopsis* [47-50]. Here we
352 investigated the extend of trans-effects in *Leishmania* by comparing and characterizing the fractions
353 of differential transcripts and proteins coming from aneuploid chromosomes, i.e. primary or cis-
354 transcripts and cis-proteins [50], versus those coming from euploid chromosomes, i.e. trans-transcripts
355 and trans-proteins.

356 We carried out differential transcript and protein abundance analyses (FDR adjusted $p < 0.05$) between
357 all possible pairs of the six *L. donovani* study strains (**Table 1**). On average, 71.8 % of differentially
358 expressed transcripts were cis-transcripts, while cis-proteins constituted only 40.7 %. The remaining
359 28.2 % of differentially expressed transcripts and 59.3 % of proteins were thus encoded by euploid
360 chromosomes (trans-) and have therefore been regulated by another mechanism than gene dosage.
361 Interestingly, only 20.0 % of these trans-proteins had a corresponding transcript change (|Log2FC| >
362 0.5), indicating the vast majority is regulated at protein level directly and not through modulation of
363 mRNA levels. Importantly, none of these trans-transcripts nor trans-proteins were associated with
364 local copy number variants: these were rare and excluded from this analysis. Hence, the most probable
365 explanation is that aneuploidy induces direct changes to cis-transcripts and cis-proteins, which in their

366 turn directly (e.g., protein-protein interaction) or indirectly drive abundance changes in trans-
367 transcripts and mostly, trans-proteins.

368 To investigate this hypothesis further, we looked deeper into the functions of trans-transcripts and
369 proteins. After pooling the trans-transcripts and trans-proteins from all comparisons, we identified a
370 set of 307 unique trans-transcripts and 180 unique trans-proteins (duplicates and hypothetical
371 proteins removed, **Suppl. Tables S1.G & S1.H**). Overrepresentation analysis showed that no GO terms
372 were significantly enriched in the pooled list of trans-transcripts or trans-proteins from all comparisons
373 (FDR adjusted *p* threshold 0.05), but it must be noted here that GO annotations in *Leishmania* are very
374 poor due to its distinct evolutionary position compared to better studied model organisms. Only 48.7%
375 of the *L. donovani* genes have at least one GO term, but they are often of the highest hierarchy
376 (unspecific). Strikingly, 12.2% (22/180) of all trans-proteins concerned subunits of macromolecular
377 complexes. A striking 82% (18/22) of these subunits had at least 1 other subunit located on an
378 aneuploid chromosome. This suggests that subunit trans-proteins are a result of cis-protein subunits
379 that alter the stability of all members of the macromolecular complex. Further, we found many trans-
380 proteins to be involved in protein metabolism, transport and metabolic functions.

381 **Table 1:** Numbers of differential (FDR adjusted $p < 0.05$) transcripts and proteins between all possible pairs of our 6 *L.donovani* strains. Each
382 time, it is indicated what proportion of these transcripts or proteins originate from chromosomes that have somy changes (because of
383 aneuploidy) in the strain pair. We typically noted more differential transcripts and proteins in strains with more somy differences. Five
384 comparisons are not shown because they contained less than 15 differential transcripts or proteins. These are shown in Supplementary table
385 1F.

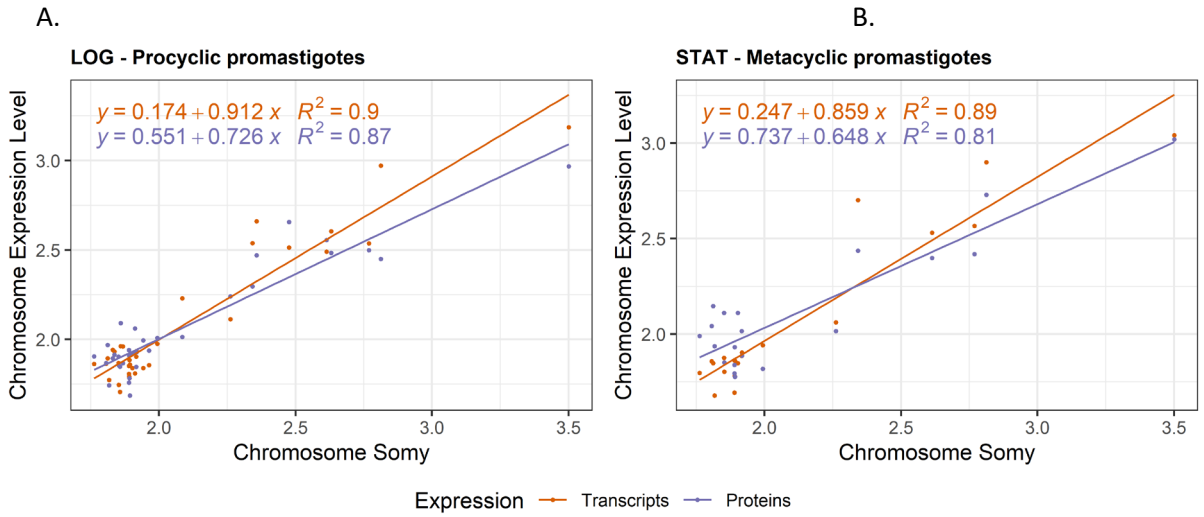
| | **DIFFERENTIAL TRANSCRIPTS** | | | **DIFFERENTIAL PROTEINS** | | |
|---|---|---|---|---|---|---|
| **STRAIN COMPARISON** | Cis | Trans | N | Cis | Trans | N |
| **275vs178** | 79.8 | 20.2 | 386 | 44.8 | 55.2 | 105 |
| **275vs173** | 86.6 | 13.4 | 134 | 40.7 | 59.3 | 91 |
| **275vs282** | 77.6 | 22.4 | 67 | 46.4 | 53.6 | 84 |
| **288vs282** | 60.9 | 39.1 | 23 | 39.3 | 60.7 | 61 |
| **275vs575** | 84.8 | 15.2 | 145 | 58.9 | 41.1 | 56 |
| **173vs288** | 86.3 | 13.8 | 80 | 55.6 | 44.4 | 27 |
| **173vs282** | 74.4 | 25.6 | 39 | 48.1 | 51.9 | 27 |
| **173vs178** | 55.0 | 45.0 | 40 | 26.3 | 73.7 | 19 |
| **575vs282** | 41.2 | 58.8 | 34 | 6.3 | 93.8 | 16 |
| | | | | | | |
| **AVERAGE** | 71.8 | 28.2 | | 40.7 | 59.3 | |

386

387 **Aneuploidy globally affects transcriptome and proteome regardless of life stage**

388 So far, this study discussed results from the logarithmic growth stage (LOG), which essentially consists
389 of *Leishmania* procyclic promastigotes (replicating, non-infectious). However, despite the absence of
390 transcriptional regulation of individual protein-coding genes, transcripts and proteins can undergo
391 post-transcriptional regulation during parasite differentiation. This raised the question whether the
392 correlation between chromosome somy, average transcript abundance, and average protein
393 abundance is also present in other life stages. Hence, for three strains, we also included the stationary
394 growth phase (STAT). STAT predominantly consists of metacyclic promastigotes, which are non-
395 replicating, have a distinct morphology, are metabolically different, and infectious to the human host
396 [51]. This was validated with metacyclogenesis markers META1 and HASPB, which were significantly
397 upregulated in STAT as they are expected to be (Adj. $p < 10^{-13}$, **Suppl. Table 1.H**). Transcriptome and
398 proteome measurements and analysis were carried out exactly as for the LOG phase.

399 For both LOG and STAT, a chromosome's somy was proportional to its average transcript expression
400 level, and protein expression level (**Fig 3**, $p < 2 * 10^{-16}$ for transcript and protein in both LOG and STAT).
401 The smaller number of data points led to slightly less stringent correlations than in our six strain
402 analysis for LOG (Fig 1). Dosage compensation was apparent at the protein level, in both LOG and STAT
403 growth phases. This suggests that the gene dosage changes associated with aneuploidy remain
404 consistent throughout the parasite's life cycle.




405

406 **Figure 3:** Impact of chromosome aneuploidy on average chromosome transcript (red) and protein (blue) output levels in LOG
407 phase (A) and STAT phase (B). Each dot represents 1 chromosome in one of the 3 studied aneuploid *L. donovani* strains
408 (BPK275, BPK178, and BPK173).

409 **Aneuploidy results in metabolic differences between *L. donovani* strains**

410 Among both cis- and trans-proteins we observed many responsible for metabolic functions. Thus, we
411 ultimately aimed to understand if aneuploidy in *L. donovani* is reflected in the parasite's metabolome.
412 Therefore, we carried out LC-MS metabolomics and confidently identified 78 metabolites across the
413 six aneuploid study strains (**Suppl. Table S5**). In contrast to transcripts and proteins, there is no
414 straightforward relation between a metabolite and a specific gene. Genes might affect metabolites
415 indirectly, affect multiple metabolites, or affect no metabolite at all. Therefore, we turned to an
416 indirect strategy to integrate the metabolomic data with our previous observations.

417 We determined if differences in aneuploidy between strains correlate with their metabolic differences.
418 As such, we calculated the pairwise genomic (somy), transcriptomic, proteomic, and metabolic
419 Euclidean distance between each possible strain pair, and correlated these distances (**Fig 4**). The larger
420 the distance between these 2 strains, the more different their respective genomes, transcriptomes,
421 proteomes, or metabolomes are from one another. Consistent with our other analyses, we found
422 strong correlations between the genomic and transcriptomic distance, between the transcriptomic
423 and proteomic distance, and between the genomic and proteomic distance ($p \leq 0.001$ for all
424 comparisons). However, the metabolome initially did not show any correlation with genome,
425 transcriptome, and proteome. We suspected this might be caused by the high intrinsic levels of
426 random variation of a subset of metabolites (by either technical or biological causes). Therefore, in the
427 second phase of the analysis, we filtered on metabolites that had at least once a significant fold change
428 (adjusted $p < 0.05$) between two strains. This strategy selects only metabolites that are stable enough
429 to reflect true differences in metabolite levels between strains. The resulting 'differentiating
430 metabolome' consisting of 34 metabolites correlated strongly with all omic layers ($p < 0.05$, Fig 4). This
431 demonstrates that the degree of karyotype difference between two strains correlates with their
432 degree of metabolic differences. Specifically, the differentiating metabolome included 9 amino acids
433 (alanine, aspartate, cysteine, leucine/isoleucine (not distinguishable), methionine, phenylalanine,
434 tryptophan, tyrosine, and valine), suggesting the amino acid metabolism is an important molecular
435 difference between these aneuploid parasites.

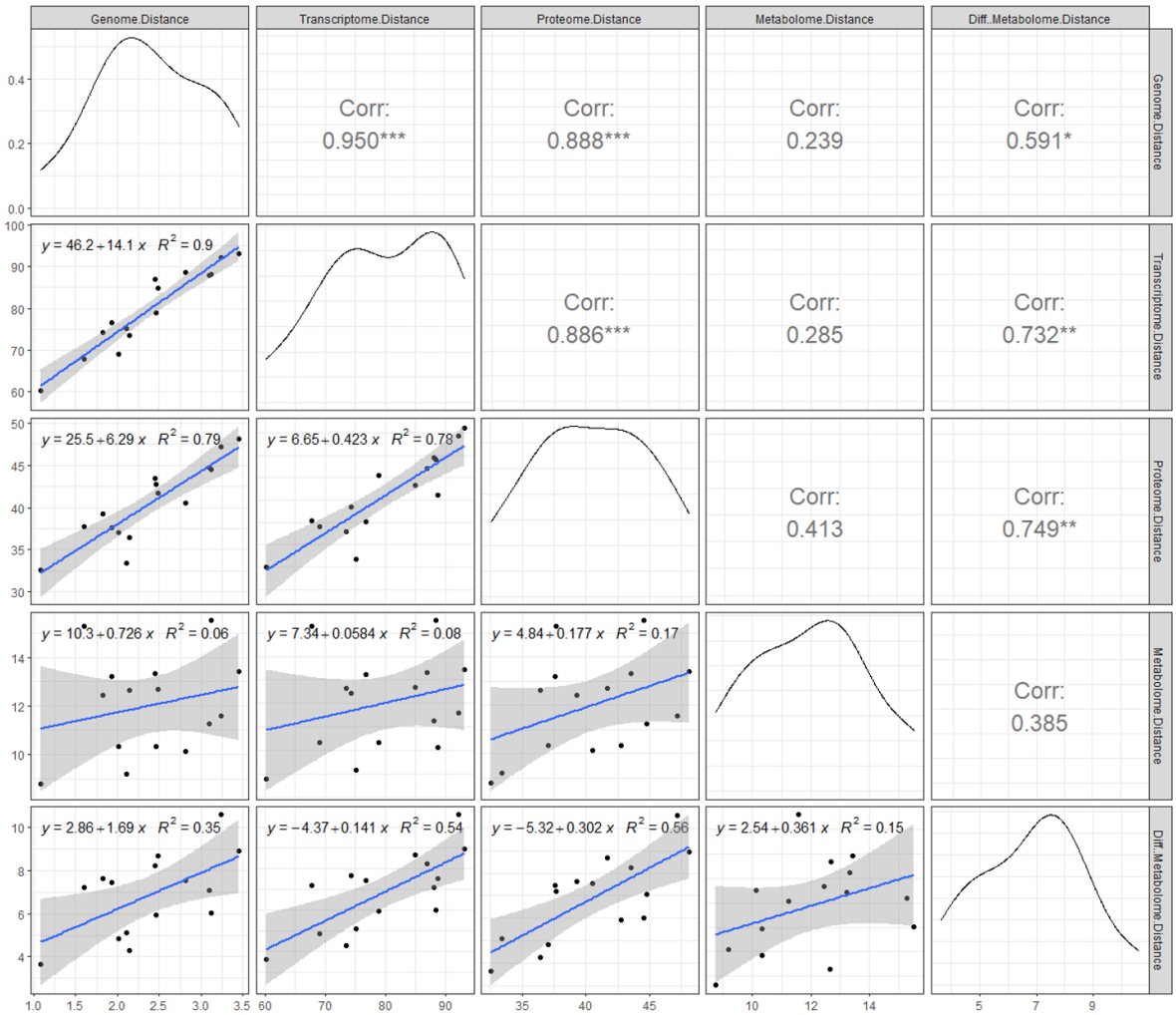



**Figure 4**: How differences between two *L. donovani* strains on 1 'omic layer correlate with their differences on another omic layer. The plots show the pairwise correlations between the genome distance (aneuploidy), transcriptome distance, proteome distance, metabolome distance (all 78 metabolites), and differentiating metabolome distance (34 metabolites), each time calculated between 2 *L. donovani* strains. Each dot represents 1 comparison between two strains. With 6 strains in our experiment, there were 15 possible pairwise comparisons. The distance metric that was used is the Euclidean distance. Somy values were used directly, but transcript abundance, protein abundance or metabolite abundance were first Z-scored before calculating the distance. The top-right panels show the Pearson correlation coefficients with * = Pval <0.05, ** Pval <0.01, *** = Pval <0.001, while the bottom-left panels show the linear regressions.

# Discussion

*Leishmania* displays a remarkably high tolerance for aneuploidy, with any chromosome having the capacity to become polysomic and produce a viable parasite. The high incidence of aneuploidy in experimental conditions stresses its central importance for the parasite to adapt to novel environments [10]. However, it remained unclear how this large genomic disruption affects the parasite as a system. Moreover, *Leishmania*'s distinct phylogenetic status and unique gene expression system disallow extrapolation from model systems. Therefore, we investigated the systemic impact of aneuploidy in *L. donovani* by integrating, for the first time, genome, transcriptome, proteome, and metabolome profiling data of highly aneuploid strains. Our study offers a comprehensive picture of the molecular changes associated with aneuploidy in this unique Eukaryotic system that lacks transcriptional regulation of individual protein-coding genes.

We confirmed that the transcript abundance of aneuploid chromosomes is globally affected. This observation is in line with previous work that has also found a strong link between aneuploidy-induced gene dosage changes and transcript abundance in *Leishmania* [13,16,52]. We further showed that this relation is almost equivalent, with transcription levels closely mirroring chromosome copy number changes (transcript abundance = 0.94x the somy). The impact of aneuploidy on the *L. donovani* proteome was also linear and chromosome-wide, but not equivalent. For every chromosome copy, the protein abundance levels of that chromosome increased (on average) by a factor 0.74x the somy.

This reduced impact of gene dosage on the protein level was due to the compensation of a subset of proteins. While some proteins did follow the aneuploidy-induced gene dosage changes, the dosage of a subset was compensated. This compensation ranged from partial to complete compensation to disomic levels. Our data thus suggest that these proteins are somehow post-transcriptionally attenuated. We made two observations concerning the identity of compensated proteins that might explain this attenuation. Firstly, compensated proteins were enriched (50.0%) for subunits of macromolecular protein complexes. This striking observation confirms several observations in aneuploid human- and yeast cell lines [47,48,53,54] that also detected protein complex subunits to be highly prevalent amongst compensated proteins. It can be explained by the fact that individual subunits of macromolecular complexes are typically unstable unless assembled into a stable complex [47]. As subunits of a single complex are often encoded on different chromosomes and produced in balanced, stoichiometric amounts, an increase in an individual subunit (here: by aneuploidy) will often not lead to more complex formation [47,48,55]. Instead, the protein finds itself without binding partner(s), remains

477 unstable, and rapidly degrades. Secondly, we observed a correlation between the degree of
478 compensation and the cellular destination of the compensated protein. Two secreted proteins
479 associated with the Golgi apparatus were the most compensated (i.e. their expression was close to
480 disomic levels), which could be explained by the fact that secreted proteins do not accumulate in the
481 cell. Correspondingly, dosage compensation of secreted proteins was also observed in a recent study
482 of human fibroblasts with trisomy 21 [47]. Proteins from mitochondria, the nucleus, and the peroxisome
483 were also significantly more compensated than cytoplasmatic proteins, which had the lowest level of
484 compensation (i.e., cytoplasmic proteins followed gene dosage the best). A potential explanation
485 might be that mitochondrial, nucleic, and peroxisome proteins require active transport into their
486 destined organelle. This transport requires extensive interaction with other proteins, potentially
487 explaining why their abundance scales less with increased protein production. In aneuploid (disomic
488 II) strains of *S. cerevisiae*, increased translation was found of genes related to protein trafficking
489 pathways of the endomembrane system and nucleus, also suggesting dependency on this system [56].

490 Partial chromosomal amplification can occur in *Leishmania*. We have previously observed this
491 phenomenon for chromosome Ld23 [45,46] and in the present study for Ld35. Interestingly, gene dosage
492 proportionally affected transcript and protein abundance for these subchromosomal amplifications as
493 well. This suggests that our findings can be extrapolated from aneuploidy to the concept of gene
494 dosage in general. Further, we demonstrated that these gene dosage changes remain important
495 throughout the life cycle of the parasite. Indeed, we observed an identical impact of gene dosage on
496 transcriptome and proteome in either LOG (enriched for procyclic promastigotes, proliferative stage)
497 or STAT (enriched for metacyclic promastigotes, infectious stage) growth phases. This is striking, as the
498 parasite featuring extensive transcriptome and proteome changes during its life cycle. For example,
499 transcriptomics studies have shown that up to 30-50% of transcripts are differentially expressed
500 between procyclic and metacyclic promastigotes [15,57]. Similarly, 10-40% of proteins are differentially
501 expressed between the parasite's promastigote and amastigote life stages [13,58].

502 A limited proportion of differential transcripts between aneuploid strains (23.6%) and the majority of
503 differential proteins (56.9%) originated from chromosomes without a change in somy, so-called trans-
504 transcripts, and trans-proteins. As we excluded the few genes with SNPs and INDELs from this analysis,
505 this suggests that the molecular impact of aneuploidy in *Leishmania* extends beyond the genes
506 encoded on the aneuploid chromosomes. It further appears that trans-protein abundances are
507 modulated at the translation or protein-stability level, as the majority did not have underlying changes
508 in transcript level. Interestingly, these results contrast with published findings for most other

509 Eukaryotes. High abundances of trans-transcripts have been reported in several other aneuploidy
510 studies on cells from humans, yeast, maize, and *Arabidopsis* [47-50]. These trans-transcriptomes were
511 typically enriched for transcription factors, genes related to metabolic processes, protein metabolism,
512 or cellular stress responses. It supports the idea that the cellular system mediates the primary impact
513 of aneuploidy by inducing transcriptional changes to trans-genes. We did not observe this coordinated
514 transcriptomic response in *Leishmania*. This matches the constitutive gene abundance model of the
515 parasite, where the transcription of individual protein-coding genes can not be controlled. Instead, we
516 found a large proportion of trans-proteins, which might be modulated directly at the protein level. This
517 modulation may also be present in other Eukaryotes, but could be overshadowed by the transcription
518 factor-induced changes. For this reason, we propose *Leishmania* as an attractive model to study
519 processes of post-transcriptional modulation of protein levels in the context of aneuploidy.

520 Finally, we showed that the degree of aneuploidy differences between strains matched the degree of
521 transcriptome, proteome, and differential metabolome changes. This is an important observation as it
522 univocally demonstrates that aneuploidy in *Leishmania* has the capacity to drive metabolic variation,
523 which is closely linked to the phenotype. It also matches our previous observations where we linked *L.*
524 *donovani* gene copy number variants directly to changes in metabolic pathways [32]. Ultimately, it
525 supports the view that aneuploidy can be adaptive and drive the metabolic changes that the parasite
526 needs to survive in a novel environment.

527 *Conclusion*

528 In summary, our study shows that aneuploidy in *Leishmania* globally and drastically impacts the
529 transcriptome and proteome of affected chromosomes throughout the parasite's life cycle. Ultimately,
530 these aneuploidy-induced changes can drive metabolomic differences between *Leishmania* strains.
531 Similarly as in other Eukaryotes, transcripts and proteins follow gene dosage, with the notable
532 exception of macromolecular subunits and secreted proteins. While most aneuploid Eukaryotes
533 display extensive transcriptional regulation of trans-transcripts to mediate the effects of aneuploidy,
534 we did not observe this in *Leishmania*. Instead, *Leishmania* features a surprisingly high number of
535 trans-proteins, which are likely post-transcriptionally modulated. This modulation may also be present
536 in other Eukaryotes, but could be overshadowed by the transcription factor-induced gene regulation
537 that is absent in *Leishmania.* For this reason, we believe that *Leishmania* is an attractive model to study
538 processes of post-transcriptional modulation of protein levels in the context of aneuploidy and
539 beyond.

# References


1 Siegel, J. J. & Amon, A. New Insights into the Troubles of Aneuploidy. *Annual review of cell and developmental biology* **28**, 189-214, doi:10.1146/annurev-cellbio-101011-155807 (2012).

2 Zhu, J., Tsai, H.-J., Gordon, M. R. & Li, R. Cellular Stress Associated with Aneuploidy. *Dev Cell* **44**, 420-431, doi:10.1016/j.devcel.2018.02.002 (2018).

3 Chen, G., Bradford, W. D., Seidel, C. W. & Li, R. Hsp90 stress potentiates rapid cellular adaptation through induction of aneuploidy. *Nature* **482**, 246, doi:10.1038/nature10795 (2012).

4 Selmecki, A., Forche, A. & Berman, J. Genomic plasticity of the human fungal pathogen Candida albicans. *Eukaryotic cell* **9**, 991-1008, doi:10.1128/ec.00060-10 (2010).

5 Gresham, D. *et al.* The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS genetics* **4**, e1000303, doi:10.1371/journal.pgen.1000303 (2008).

6 Gerstein, A. C. *et al.* Polyploid titan cells produce haploid and aneuploid progeny to promote stress adaptation. *mBio* **6**, e01340-01315, doi:10.1128/mBio.01340-15 (2015).

7 Seton-Rogers, S. Fitness penalties of aneuploidy. *Nature Reviews Cancer* **17**, 142-143, doi:10.1038/nrc.2017.9 (2017).

8 Rancati, G. *et al.* Aneuploidy underlies rapid adaptive evolution of yeast cells deprived of a conserved cytokinesis motor. *Cell* **135**, 879-893, doi:10.1016/j.cell.2008.09.039 (2008).

9 Sunshine, A. B. *et al.* The Fitness Consequences of Aneuploidy Are Driven by Condition-Dependent Gene Effects. *PLOS Biology* **13**, e1002155, doi:10.1371/journal.pbio.1002155 (2015).

10 Imamura, H. *et al.* Evolutionary genomics of epidemic visceral leishmaniasis in the Indian subcontinent. *eLife* **5**, e12613 (2016).

11 Franssen, S. U. *et al.* Global genome diversity of the Leishmania donovani complex. *eLife* **9**, e51243, doi:10.7554/eLife.51243 (2020).

12 Negreira, G. H. *et al.* Exploring the evolution and adaptive role of mosaic aneuploidy in a clonal Leishmania donovani population using high throughput single cell genome sequencing. *bioRxiv*, 2020.2003.2005.976233, doi:10.1101/2020.03.05.976233 (2020).

13 Dumetz, F. *et al.* Modulation of Aneuploidy in Leishmania donovani during Adaptation to Different In Vitro and In Vivo Environments and Its Impact on Gene Expression. *mBio* **8**, e00599-00517 (2017).

14 Ivens, A. C. *et al.* The genome of the kinetoplastid parasite, Leishmania major. *Science (New York, N.Y.)* **309**, 436-442, doi:10.1126/science.1112680 (2005).

15 Cuypers, B. *et al.* Multiplexed Spliced-Leader Sequencing: A high-throughput, selective method for RNA-seq in Trypanosomatids. *Scientific Reports* **7**, 3725, doi:10.1038/s41598-017-03987-0 (2017).

16 Prieto Barja, P. *et al.* Haplotype selection as an adaptive mechanism in the protozoan pathogen Leishmania donovani. *Nature ecology & evolution*, doi:10.1038/s41559-017-0361-x (2017).

17 WHO. Proposals for the nomenclature of salivarian trypanosomes and for the maintenance of reference collections*. *Bulletin of the World Health Organization* **56**, 467-480 (1978).

18 Andrews, S. FastQC: a quality control tool for high throughput sequence data. (2010).

19 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).

585 20 Dumetz, F. *et al.* Modulation of Aneuploidy in Leishmania donovani during Adaptation to
586 Different In Vitro and In Vivo Environments and Its Impact on Gene Expression. *mBio* **8**,
587 doi:10.1128/mBio.00599-17 (2017).
588 21 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*
589 *preprint arXiv:1303.3997* (2013).
590 22 Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and
591 population genetical parameter estimation from sequencing data. *Bioinformatics (Oxford,*
592 *England)* **27**, 2987-2993, doi:10.1093/bioinformatics/btr509 (2011).
593 23 Downing, T. *et al.* Whole genome sequencing of multiple Leishmania donovani clinical isolates
594 provides insights into population structure and mechanisms of drug resistance. *Genome*
595 *research* **21**, 2143-2156, doi:10.1101/gr.123430.111 (2011).
596 24 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput
597 sequencing data. *Bioinformatics (Oxford, England)* **31**, 166-169,
598 doi:10.1093/bioinformatics/btu638 (2015).
599 25 Quinn, T. P., Erb, I., Richardson, M. F. & Crowley, T. M. Understanding sequencing data as
600 compositions: an outlook and review. *Bioinformatics (Oxford, England)* **34**, 2870-2878,
601 doi:10.1093/bioinformatics/bty175 (2018).
602 26 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for
603 RNA-seq data with DESeq2. *Genome biology* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
604 27 Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-
605 based shotgun proteomics. *Nature Protocols* **11**, 2301-2319, doi:10.1038/nprot.2016.136
606 (2016).
607 28 Berg, M. *et al.* Experimental resistance to drug combinations in Leishmania donovani:
608 metabolic and phenotypic adaptations. *Antimicrobial agents and chemotherapy* **59**, 2242-2255
609 (2015).
610 29 Berg, M. *et al.* Metabolic adaptations of Leishmania donovani in relation to differentiation,
611 drug resistance, and drug pressure. *Mol Microbiol* **90**, 428-442, doi:10.1111/mmi.12374
612 (2013).
613 30 Benton, H. P., Want, E. J. & Ebbels, T. M. Correction of mass calibration gaps in liquid
614 chromatography-mass spectrometry metabolomics data. *Bioinformatics (Oxford, England)* **26**,
615 2488-2489, doi:10.1093/bioinformatics/btq441 (2010).
616 31 Scheltema, R. A., Jankevics, A., Jansen, R. C., Swertz, M. A. & Breitling, R. PeakML/mzMatch: a
617 file format, Java library, R library, and tool-chain for mass spectrometry data analysis.
618 *Analytical chemistry* **83**, 2786-2793, doi:10.1021/ac2000994 (2011).
619 32 Cuypers, B. *et al.* Integrated genomic and metabolomic profiling of ISC1, an emerging
620 Leishmania donovani population in the Indian subcontinent. *Infection, Genetics and Evolution*
621 **62**, 170-178, doi:10.1016/j.meegid.2018.04.021 (2018).
622 33 Prince, J. T. & Marcotte, E. M. Chromatographic alignment of ESI-LC-MS proteomics data sets
623 by ordered bijective interpolated warping. *Anal Chem* **78**, 6140-6152, doi:10.1021/ac0605344
624 (2006).
625 34 Tautenhahn, R., Böttcher, C. & Neumann, S. Highly sensitive feature detection for high
626 resolution LC/MS. *BMC Bioinformatics* **9**, 504, doi:10.1186/1471-2105-9-504 (2008).
627 35 Doyle, M. A. *et al.* LeishCyc: a biochemical pathways database for Leishmania major. *BMC Syst*
628 *Biol* **3**, 57-57, doi:10.1186/1752-0509-3-57 (2009).
629 36 Fahy, E. *et al.* Update of the LIPID MAPS comprehensive classification system for lipids. *J Lipid*
630 *Res* **50 Suppl**, S9-S14, doi:10.1194/jlr.R800095-JLR200 (2009).
631 37 Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on
632 genomes, pathways, diseases and drugs. *Nucleic acids research* **45**, D353-d361,
633 doi:10.1093/nar/gkw1092 (2017).

38 Wishart, D. S. *et al.* HMDB 4.0: the human metabolome database for 2018. *Nucleic acids research* **46**, D608-d617, doi:10.1093/nar/gkx1089 (2018).
39 Tihon, E. *et al.* Genomic analysis of Isometamidium Chloride resistance in Trypanosoma congolense. *International Journal for Parasitology: Drugs and Drug Resistance* **7**, 350-361, doi:10.1016/j.ijpddr.2017.10.002 (2017).
40 Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).
41 Krzywinski, M. I. *et al.* Circos: An information aesthetic for comparative genomics. *Genome research*, doi:10.1101/gr.092759.109 (2009).
42 Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9**, 90-95, doi:10.1109/MCSE.2007.55 (2007).
43 Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17**, 261-272, doi:10.1038/s41592-019-0686-2 (2020).
44 Imamura, H. *et al.* Evolutionary genomics of epidemic visceral leishmaniasis in the Indian subcontinent. *eLife* **5**, doi:10.7554/eLife.12613 (2016).
45 Dumetz, F. *et al.* Molecular Preadaptation to Antimony Resistance in Leishmania donovani on the Indian Subcontinent. *mSphere* **3**, e00548-00517, doi:10.1128/mSphere.00548-17 (2018).
46 Laffitte, M.-C. N. *et al.* Formation of Linear Amplicons with Inverted Duplications in Leishmania Requires the MRE11 Nuclease. *PLoS genetics* **10**, e1004805, doi:10.1371/journal.pgen.1004805 (2014).
47 Hwang, S. *et al.* Consequences of aneuploidy in human fibroblasts with trisomy 21. *Proceedings of the National Academy of Sciences* **118**, e2014723118, doi:10.1073/pnas.2014723118 (2021).
48 Torres, E. M. *et al.* Effects of Aneuploidy on Cellular Physiology and Cell Division in Haploid Yeast. *Science* **317**, 916-924, doi:10.1126/science.1142210 (2007).
49 Makarevitch, I., Phillips, R. L. & Springer, N. M. Profiling expression changes caused by a segmental aneuploid in maize. *BMC Genomics* **9**, 7, doi:10.1186/1471-2164-9-7 (2008).
50 Huettel, B., Kreil, D. P., Matzke, M. & Matzke, A. J. M. Effects of Aneuploidy on Genome Structure, Expression, and Interphase Organization in Arabidopsis thaliana. *PLoS genetics* **4**, e1000226, doi:10.1371/journal.pgen.1000226 (2008).
51 Berg, M. *et al.* Metabolic adaptations of Leishmania donovani in relation to differentiation, drug resistance, and drug pressure. *Molecular microbiology* **90**, 428-442, doi:10.1111/mmi.12374 (2013).
52 Iantorno, S. A. *et al.* Gene Expression in Leishmania Is Regulated Predominantly by Gene Dosage. *mBio* **8**, doi:10.1128/mBio.01393-17 (2017).
53 Dephoure, N. *et al.* Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. *eLife* **3**, e03023, doi:10.7554/eLife.03023 (2014).
54 Stingele, S. *et al.* Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. *Mol Syst Biol* **8**, 608, doi:10.1038/msb.2012.40 (2012).
55 Veitia, R. A., Bottani, S. & Birchler, J. A. Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. *Trends in Genetics* **24**, 390-397, doi:10.1016/j.tig.2008.05.005 (2008).
56 Larrimore, K. E., Barattin-Voynova, N. S., Reid, D. W. & Ng, D. T. W. Aneuploidy-induced proteotoxic stress can be effectively tolerated without dosage compensation, genetic mutations, or stress responses. *BMC Biology* **18**, 117, doi:10.1186/s12915-020-00852-x (2020).
57 Dillon, L. A. L. *et al.* Transcriptomic profiling of gene expression and RNA processing during Leishmania major differentiation. *Nucleic acids research* **43**, 6799-6813, doi:10.1093/nar/gkv656 (2015).

58 Lynn, M. A., Marr, A. K. & McMaster, W. R. Differential quantitative proteomic profiling of Leishmania infantum and Leishmania mexicana density gradient separated membranous fractions. *J Proteomics* **82**, 179-192, doi:10.1016/j.jprot.2013.02.010 (2013).

# Funding

This work was supported by the Research Foundation Flanders [11O1614N to B.C.]; the Interuniversity Attraction Poles Program of Belgian Science Policy [P7/41 to JC.D.], the InBev Baillet-Latour foundation and the Department of Economy, Science and Innovation in Flanders [ITM-SOFIB, SINGLE to JC.D.]

# Acknowledgements

We thank the Center of Medical Genetics at the University of Antwerp for hosting the NGS facility on which our RNA-Seq experiments were performed. The computational resources used for this work were provided by the VSC (Flemish Supercomputer Center) at the University of Antwerp.

# Data Availability Statement

The raw genomic and transcriptomic data in this study is available at the NCBI Sequence Read Archive (SRA) under the accession code: PRJNA762444. Raw proteomic data has been submitted to the PRIDE database under reference 1-20210911-141001. Raw metabolomic data was submitted to the MetaboLights database under accession code MTBLS1612.

# Conflict of Interest

The authors declare that they have no conflict of interest.

700 # Supplementary Tables

701 **Supplementary Table S1.A**

702 Detailed description of the *L. donovani* isolates used in this study.

703 **Supplementary Table S1.B**

704 Number of total SNPs, non-synonmous SNPs and homozygous non-synonymous SNPs between any
705 pair of *L. donovani* strains in this study.

706 **Supplementary Table S1.C**

707 Sequencing quality and mapping statistics of all 6 sequenced *L. donovani* genomes in this study.

708 **Supplementary Table S1.D**

709 Sequencing quality and mapping statistics of all sequenced *L. donovani* transcriptomes in this study.

710 **Supplementary Table S1.E**

711 Dosage compensation model data, coefficients and *p*-values.
712

713 **Supplementary Table S1.F**

714 Summary of all cis- and trans- transcripts and proteins of all possible pairwise comparisons between
715 the 6 *L. donovani* strains in this study.

716 **Supplementary Table S1.G**

717 Pooled trans-transcripts from all possible pairwise comparisons between the 6 *L. donovani* strains in
718 this study.

719 **Supplementary Table S1.H**

720 Pooled trans-proteins from all possible pairwise comparisons between the 6 *L. donovani* strains in this
721 study.

722 **Supplementary Table S1. I**

723 Upregulated metacyclogenis markers in STAT growth phase. Output generated by DESeq2.

724 **Supplementary Table S2**

725 Structure (Somy and local CNVs) of all sequenced *L. donovani* genomes in this study.

726 **Supplementary Table S3**

727 Differential RNA abundance analyses of the different biological comparisons made in this study.

728 **Supplementary Table S4**

729 Detailed statistics about all proteins detected in this study. Output generated by MaxQuant.

730 **Supplementary Table S5**

731 Detailed statistics about all metabolites detected in this study.

# 732 Supplementary Figures

733 **Supplementary Figure S1.A & S1.B**

734 Comparative Circos plot between BPK173 and BPK288 (S1.A) and between BHU575 and BPK282 (S1.B),
735 showing their relative (expressed in fold change) gene dosage (blue), transcript abundance (orange),
736 and protein abundance (purple) across the 36 *L. donovani* chromosomes.