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 the molecular and functional consequences for the pathogen remain poorly understood. Here, we addressed this question *in vitro* and present the first integrated analysis of the genome, transcriptome, proteome, and metabolome of highly aneuploid *Leishmania donovani* strains. Our analyses unambiguously establish that aneuploidy in *Leishmania* proportionally impacts the average transcript- and protein abundance levels of affected chromosomes, ultimately correlating with the degree of metabolic differences between strains. This proportionality was present in both proliferative and non-proliferative *in vitro* promastigotes. However, protein complex subunits and non-cytoplasmic proteins, showed dosage compensation, responding less or even not at all to aneuploidy-induced dosage changes. In contrast to other Eukaryotes, we did not observe the widespread regulation at the transcript level that typically modulates some of the negative effects of aneuploidy. Further, the majority of differentially expressed proteins between aneuploid strains were encoded by non-aneuploid chromosomes and were not driven by a significant underlying transcript change, suggesting that aneuploidy is accompanied by 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 selection of a beneficial phenotype from a pool 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, a new and unique Eukaryote model is emerging, *Leishmania*. This protozoan parasite (Euglenozoa, Kinetoplastida) that is transmitted by sand flies and infects humans and various vertebrate animals shows a remarkably high tolerance for aneuploidy. 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], while *in vivo*, aneuploidy is less pronounced [11]. All 36 chromosomes have the capacity to become aneuploid and lead to viable parasites [12]. Strikingly, clonal populations of the parasite do not have uniform karyotypes but display a mosaic of them[13]. Single-cell sequencing of two clonal promastigote populations of *L. donovani* identified respectively 208 different karyotypes in 1516 sequenced cells (that were generated in approximately 126 generations) and 117 karyotypes in 2378 cells (56 generations). This high degree of 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 [11].

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 RNA polymerase II 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 still regulated, but essentially post-transcriptionally [11,15]. Thus, the most straightforward evolutionary path towards altered gene expression might be altered gene dosage, which could explain the high incidence of aneuploidy in this organism [11]. While aneuploidy in *Leishmania* is known to directly affect the transcription of encoded transcripts, its molecular impact further downstream remains poorly understood [11,16].

Here we present the first study that generated and integrated genome, transcriptome, proteome, and metabolome data of highly aneuploid *L. donovani* strains, *in vitro*. This multi-omics analysis unambiguously demonstrates that aneuploidy in *Leishmania* globally and drastically impacts the transcriptome and proteome of affected chromosomes, ultimately correlating with the degree of metabolic differences between strains. This impact is not restricted to a single life stage as we observed it *in vitro* in both the parasite's proliferative and non-proliferative/infectious promastigote life stages. Interestingly, not all proteins responded equally to dosage changes; through dosage compensation, protein complex subunits and non-cytoplasmic proteins responded less or even not at all to aneuploidy. In parallel, aneuploid *Leishmania* strains here studied also feature a surprisingly high number of regulated proteins from non-aneuploid chromosomes, which are likely post-transcriptionally modulated. Through this integrated 4-'omic approach, we provide a systems-wide view of the molecular changes associated with aneuploidy in *Leishmania*.

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 $d = d_r/d_{ch}$. Somy values in our bulk-sequencing analyses are often not
129 discrete due to mosaic aneuploidy (i.e. presence of different kayotypes among individual cells present
130 in 1 strain) in *Leishmania*. Gene copy number estimations per cell (full depth, gene dosage) were then
131 calculated by multiplying d with S. SNPs were previously characterized in [10]. RNA-Seq gene read count
132 tables were generated with the HTSeq tool version 0.6.1 [24]. Count data were normalized for library size
133 by dividing each transcript count by the geometric mean of all counts for that library [25].

134 Proteomics

135 *Sample preparation*

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

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

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

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

167 *Proteomic Data Analysis*

168 Thermo raw files were converted to mzML files using MSConvert v3.0. Label-free protein quantification
169 (LFQ) was carried out with MaxQuant version 1.6.0.16 [26] using the following settings (other settings
170 were kept default): Oxidation (M) and acetyl (Protein N-term) were indicated as variable modifications,
171 carbamidomethyl (C) was indicated as a fixed modification, digestion with trypsin with maximum two
172 missed cleavages, match between runs = yes, dependent peptides = yes. The search database was the
173 LdBPKV2 proteome previously published by our group [11], and reversed proteins were used as decoys.

174 **Metabolomics**

175 *Sample preparation*

176 Sample pellets were extracted using a 1:3:1 (v:v:v) chloroform/methanol/water extraction, subjected
177 to liquid chromatography using a 2.1 mm ZIC-HILIC column (Sequant), and analyzed using the Orbitrap
178 Exactive (Thermo Fisher Scientific) platform at the Glasgow Polyomics Center (University of Glasgow,
179 Glasgow, Scotland) exactly as described elsewhere [27,28]. Both negative and positive ion modes were

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

186 *Metabolomic data analysis*

187 Data preprocessing was carried out with the XCMS 1.42.0 [29] and mzMatch 1.0.1 [30] packages in R,
188 exactly as described elsewhere [31]. Briefly, raw spectra (mzXML format) were first subjected to
189 retention time alignment with ObiWarp [32], after which peak detection was carried out with the
190 centWave algorithm [33]. Peaks were then filtered: 1) Corresponding peaks from the four biological
191 replicates were allowed a maximum reproducibility standard deviation (RSD) of 0.5, 2) Peak shape
192 (CoDA-DW > 0.8), 3) Detection in at least 3 out of 4 replicates of a single strain and 4) Minimal peak
193 intensity was set to 3000. The peaks were putatively identified using sequentially the LeishCyc
194 database [34], LIPID MAPS [35], KEGG [36], the peptide database included in mzMatch, and the Human
195 Metabolome Database [37]. Finally, the filtered peaks were normalized for total ion count and subjected
196 to manual quality screening.

197 **Integrated multi -omic analysis**

198 Pairwise comparisons between aneuploid strains: For each comparison between two strains, we
199 calculated the Log_2 fold change (Log_2FC) for genes (gene dosage), transcripts, proteins and metabolites.
200 These Log_2FC values were obtained by taking the Log_2 ratio of the average abundance between two
201 strains. Associated *p*-values were obtained by performing a Student *t*-test for each comparison and
202 subsequently correcting it with the Benjamini-Hochberg algorithm to limit the False Discovery Rate
203 (FDR) to 5%. A Log2FC cutoff of 1 and adjusted *p*-value cutoff of 0.05 was used for all 'omic layers,
204 except for gene dosage, for which only copy number differences of at least 0.5/haploid genome were
205 considered as biologically significant [38].

206 Average transcript and protein abundance levels per chromosome per strain (Fig 1 C): First, the median
207 was taken of the Log_2 normalized transcript counts (see respective section) and Log_2 LFQ protein
208 values, per chromosome per strain. This yielded an expression value for each chromosome in each
209 strain, for both transcript and protein expression. Then, for each chromosome, the median of the
210 expression values of all strains that were disomic (determined based on DNA sequencing data) for that

211 chromosome was selected as the 'disomic reference value' for that chromosome (i.e., the expression
212 value that corresponds to the median expression of a transcript/protein on a disomic chromosome).
213 The remaining per-strain per-chromosome expression were then divided by this disomic reference
214 value, multiplied by 2, and plotted in Figure 1C. The code for these normalizations and analyses is
215 available at (https://github.com/CuypersBart/Leishmania_Multiomics).

216 All plots were generated using the ggplot v 2_2.1.0 package [39], except the Circos plots which were
217 created with Circos v0.69 [40]. Gene ontology enrichments were calculated with the Fisher Exact test
218 using Python 3.8.5 and the matplotlib v3.3.4 [41] and SciPy v1.6.1 [42] libraries. GO annotations were
219 obtained from the gene ontology consortium (http://www.geneontology.org/).

220 Dosage compensation analysis

221 Dosage compensation analysis (Figure 2): This analysis was carried out on the only three chromosomes
222 (Ld05, Ld08, and Ld33) that were at least once disomic (somy of 1.7-2.3), trisomic (somy of 2.7-3.3),
223 and tetrasomic (somy of 3.5 or more) in our six study strains and had more than 10 detected proteins.
224 The abundance values of transcripts (corrected for library size, see above) and proteins (LFQ) were
225 divided by their abundance in the strain that was the most disomic for their encoding chromosome,
226 according to the genome sequencing data. These expression values were then multiplied by 2 so that
227 the normalized abundance at disomy = 2. Strains were always normalized by another strain than
228 themselves to avoid underestimation of the standard deviation. Therefore, we do not have disomic
229 distributions for Ld08 since Ld08 was exclusively disomic in BPK282, and thus already was used as
230 reference strain.

231 Functional dosage compensation analysis: For this analysis, we defined a protein's compensation level
232 as its Log_2 transcript fold change (i.e., its transcript level on the aneuploid chromosome divided by its
233 transcript level on disomic chromosome), minus its Log_2 protein fold change. We included only 85
234 transcripts and proteins for this analysis, which had very accurate protein measurements (SD between
235 biological replicates < 0.25 Log_2FC units). The following protein features were checked for their relation
236 with protein dosage compensation: degree (i.e., number of protein interactions as transferred from
237 STRING), length (as a proxy for transcript and protein molecular size), GRAVY (grand average of
238 hydropathy), solubility (soluble versus membrane protein) and subcellular location. We added
239 transcript abundance as a covariate to the model, as the compensatory effect can be expected to
240 become larger at higher differential transcript levels. The code these normalizations and analyses is
241 available at (https://github.com/CuypersBart/Leishmania_Multiomics).

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 [43]. 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 *in vitro* 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 estimations of chromosomal somy (i.e., average chromosome copy number in the population of cells) 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 confirmed 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 karyotypes 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.

272 The transcript abundance of aneuploid chromosomes mirrored almost perfectly their underlying somy
273 (**Fig 1B, Suppl. Fig S1A, S1B**). Specifically, the average abundance of a transcript encoded by a
274 chromosome was proportional with a factor of 0.94 with the chromosomal somy (**Fig 1C**: 95% CI: 0.92
275 – 0.96 and $p < 2 \cdot 10^{-16}$, R^2: 0.98). The average abundance of the proteins encoded by a chromosome
276 was also highly proportional to the aneuploidy of the chromosome ($p < 2 \cdot 10^{-16}$), albeit with a
277 decreased slope of 0.76 (95% CI: 0.70 - 0.77, R^2: 0.92). Thus, aneuploidy results in a linear, but not
278 equivalent, increase in protein abundance. Circos plots clearly showed that transcript- and protein
279 abundance of aneuploid chromosomes was affected globally and not restricted to specific
280 chromosomal regions (**Fig 1B, Suppl. Fig S1A, S1B**).

281 Interestingly, chromosome 35 exhibited only a partial amplification in strain BPK275, the first part
282 being disomic (somy = 2.2) and the second part, starting from gene LdBPK_350040900, being trisomic
283 (somy = 3.1) **(Fig 1B).** This phenomenon of partial chromosomal amplification has previously been
284 observed for chromosome 23 as well [44,45]. The difference in both average transcript and protein
285 abundance between the first and the second chromosome part was significant (t-test *p* Trans = $4.0 \times 10^{-}$
286 ^{82}, Prot = 7.0×10^{-7}), the second part having on average 1.43x more transcript abundance and 1.22x
287 more protein abundance. This demonstrates that also partial chromosomal amplifications affect
288 transcript and protein abundance proportionally in the amplified region.

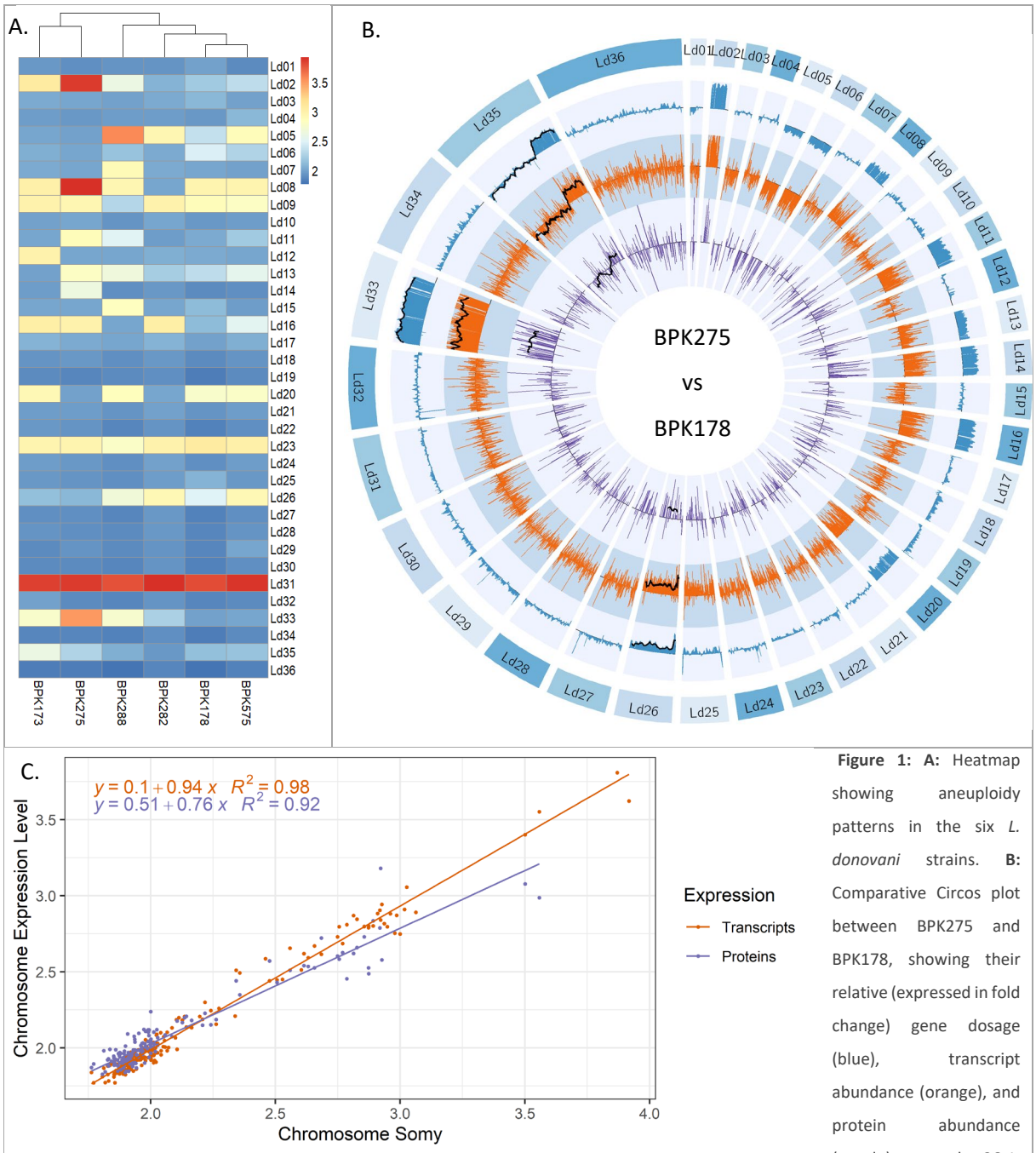


Figure 1: A: Heatmap showing aneuploidy patterns in the six *L. donovani* strains. **B:** Comparative Circos plot between BPK275 and BPK178, showing their relative (expressed in fold change) gene dosage (blue), transcript abundance (orange), and protein abundance (purple) across the 36 *L. donovani* 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 replicative (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*.

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

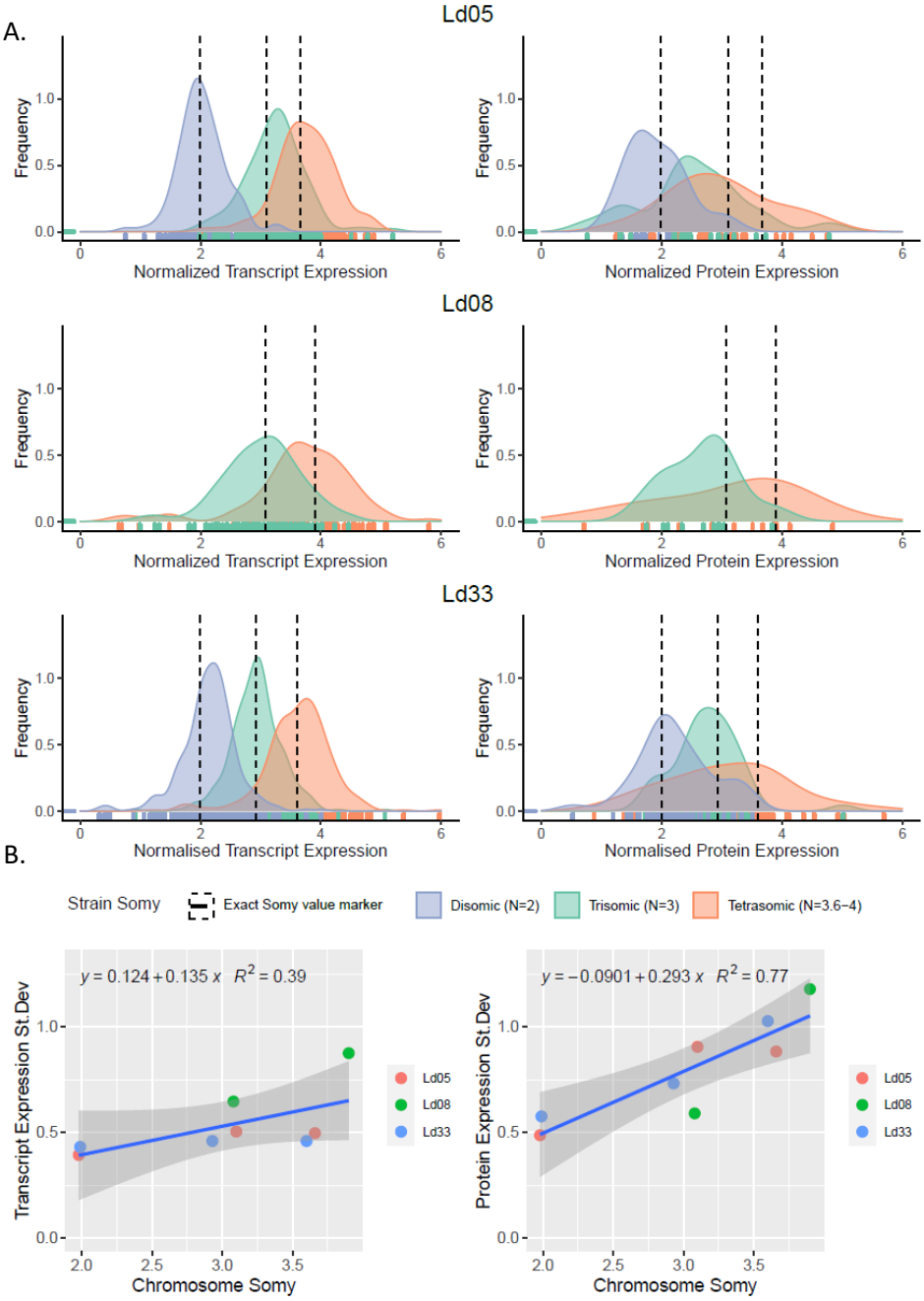
290 Our observation that a somy increase does not lead to an equivalent increase in protein abundance,
291 but 0.76x the somy change, suggested either of two possible explanations. A first possibility is that all
292 proteins encoded by an aneuploid chromosome are modulated equally and exhibit only a partial
293 protein abundance increase. We further refer to this explanation as the 'dampening hypothesis'. Such
294 a general 'dampening' effect would suggest an underlying process related to intrinsic mechanisms or
295 limitations in the parasite's translation system. Alternatively, different proteins could respond
296 differently to increased gene dosage and transcript abundance, pointing to compensatory effects
297 linked to the identity, function, and physicochemical properties of those specific proteins or certain
298 functional protein groups. We refer to this explanation as the 'compensation hypothesis'.

299 To evaluate both hypotheses, the differential abundance patterns of proteins were compared to their
300 encoding chromosome differences from disomy to trisomy and tetrasomy. Following the 'dampening
301 hypothesis', the average of the protein abundance distribution should increase with increased somy,
302 but the distribution shape should remain unaltered. In our study, three chromosomes (Ld05, Ld08, and
303 Ld33) were at least once disomic, trisomic, and tetrasomic in one of our six study strains (**Fig 1B**),
304 allowing for such comparative analysis. Ld02 also matched these criteria but was excluded from the
305 analysis as only nine Ld02 proteins were detected. For reference, the same analysis was conducted for
306 transcripts.

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

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

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



322

323 **Figure 2**: **A**: Normalized transcript (left) and protein (right) abundance distributions of chromosomes Ld05, Ld08, and Ld33 in their disomic
324 (excl. chr 8), trisomic and tetrasomic states, each time observed in one of our 6 study strains. Vertical dotted lines indicate the corresponding
325 somy, derived from the genomic data. The small vertical lines at the base of the plot indicate individual datapoints for transcripts and
326 proteins, from which the distributions were generated. The abundance values of transcripts and proteins were normalized so that the
327 normalized abundance at disomy = 2. This normalization was carried out with a strain that was disomic for the chromosome under
328 investigation. We do not have disomic distributions for Ld08 since Ld08 was exclusively disomic in BPK282, and thus already was used as

329 reference strain. **B.** Regression between the somy and transcript (left) or protein (right) standard deviation for chromosomes Ld05, Ld08, and
330 Ld33 in different somy states.

331 Compensated proteins are primarily protein complex subunits or non-cytoplasmic

332 We investigated if specific protein properties were associated with the dosage compensation observed
333 in our dataset. For this analysis, we defined a protein's compensation level as its Log_2 transcript fold
334 change (i.e., its transcript level on the aneuploid chromosome divided by its transcript level on disomic
335 chromosome), minus its Log_2 protein fold change. A particular challenge in this analysis was that
336 protein fold changes, and hence, the observed compensation levels, were generally very small and
337 close to measurement noise levels. For example, for a fully compensated protein from a trisomic
338 chromosome, the compensation level would only be 0.5 Log_2FC units (Log_2FC transcript ≈ 0.5, Log_2FC
339 protein = 0). Therefore, we focused on a set of 85 transcripts and proteins, which had very accurate
340 protein measurements (SD between biological replicates < 0.25 Log_2FC units). These covered a broad
341 range of compensation levels ranging from 1.18 Log_2FC units, to values close to zero, or even some
342 proteins that had a slight overamplification compared to the transcript level (-0.40 Log_2FC units)
343 (**Suppl. Table S1.E**).

344 We noted that within the top 5 most compensated proteins in this high-accuracy set, 2 coded for
345 subunits of protein complexes: e.g., the cleavage and Polyadenylation Specificity Factor (CPSF) subunit
346 A and the dynein light chain had respectively 1.12 and 0.96 Log2FC units lower protein than transcript
347 abundance. Accordingly, we hypothesized that protein complexes might be more likely to be dosage
348 compensated (>= 0.30 Log2FC units). To test this systematically, all 85 proteins were screened for their
349 involvement in heteromeric protein complexes (using annotation keywords 'subunit' and 'chain',
350 **Suppl. Table S1.E**). Strikingly, complex subunits accounted for 15.6 % (5 out of 32) of the compensated,
351 proteins which was significantly more than the 1.8% (1 out of 53) we observed for the uncompensated
352 proteins (Fischer exact p = 0.0265).

353 Next, we looked if there was a relation between dosage compensation and a set of biological and
354 physicochemical protein properties to identify more subtle patterns. Specifically, we checked 5 protein
355 properties for their relationship with protein dosage compensation using AN(C)OVA **(Suppl. Table**
356 **S1.E)**. Transcript abundance was added as a covariate to the model, as dosage compensation was
357 larger at higher differential transcript levels (Adj $p = 2.4 \times 10^{-4}$). The variables 'membrane protein' (yes
358 or no) (FDR adjusted p = 0.696), the number of protein interactions (Adj. p = 1), protein length (a proxy
359 for transcript and protein size) (Adj. p = 1) and hydrophobicity (Adj. p = 1) had no significant relationship
360 with dosage compensation. In contrast, the subcellular location (Adj. p = 0.032) of a protein was

361 significantly linked to its degree of dosage compensation. To investigate this further, we investigated
362 the relation between each subcellular compartment and the compensation rate of its proteins **(Suppl.**
363 **Table S1.E)**. Cytoplasmatic proteins (n = 39) were significantly less compensated compared to all other
364 proteins (Adj. p = 0.011). Thus, cytoplasmatic proteins follow aneuploidy-induced dosage changes
365 more closely than proteins from other cellular compartments. In contrast, proteins associated with the
366 Golgi apparatus (n = 2) had a significantly higher compensation than all other proteins (Adj. p = 0.0065).
367 However, this result needs to be interpreted with caution as the sample size was only 2 proteins. These
368 two proteins were a vesicle-associated membrane protein and a small GTP-binding protein Rab18,
369 both known to be secreted. Proteins from mitochondria (Adj. p = 0.32, n = 18), nucleus (Adj. p = 0.50,
370 n = 13), peroxisome (Adj. p = 0.81, n = 6) and ER (Adj. p = 0.74, n = 2) did not have significantly different
371 compensation rates compared to proteins from all other compartments.

372 The impact of aneuploid chromosomes on the gene products of their euploid 373 counterparts

374 In the previous sections, we reported how the aneuploidy of a chromosome affects the abundance of
375 its encoded transcripts and proteins. However, through perturbation of regulatory networks, these
376 changes might also affect transcripts and proteins encoded by genomic regions unaffected by
377 aneuploidy. These were previously defined as 'trans-effects' and have been observed on the
378 transcriptional level in diverse species, including humans, yeast, maize, and *Arabidopsis* [46-49]. Here we
379 investigated the extent of trans-effects in *Leishmania* by comparing and characterizing the fractions of
380 differential transcripts and proteins coming from aneuploid chromosomes, i.e. primary or cis-
381 transcripts and cis-proteins [49], versus those coming from euploid chromosomes, i.e. trans-transcripts
382 and trans-proteins.

383 We carried out differential transcript and protein abundance analyses (FDR adjusted $p < 0.05$) between
384 all possible pairs of the six *L. donovani* study strains (**Table 1**). On average, 71.8 % of differentially
385 expressed transcripts were cis-transcripts, while cis-proteins constituted only 40.7 %. The remaining
386 28.2 % of differentially expressed transcripts and 59.3 % of proteins were thus encoded by euploid
387 chromosomes (trans-) and have therefore been regulated by another mechanism than gene dosage.
388 Interestingly, only 20.0 % of these trans-proteins had a corresponding transcript change (|Log2FC| >
389 0.5), indicating the vast majority is regulated at protein level directly and not through modulation of
390 mRNA levels. Importantly, none of these trans-transcripts nor trans-proteins were associated with
391 local copy number variants: these were rare and excluded from this analysis. Also sequence variation

392 between strain pairs is unlikely to have resulted in the observed proportions of differential trans-
393 transcripts and trans-proteins. Only 11 trans-transcripts contained SNPs between our strains (6
394 missense, 1 synonymous and 4 upstream SNPs) and 2 proteins (1 upstream and 1 downstream SNP)
395 (Supplementary Tables 1G & 1H). Therefore, our hypothesis is that aneuploidy induces direct changes
396 to cis-transcripts and cis-proteins, which in their turn directly (e.g., protein-protein interaction) or
397 indirectly drive abundance changes in trans-transcripts and mostly, trans-proteins.

398 Next, we looked deeper into the functions of trans-transcripts and proteins. After pooling the trans-
399 transcripts and trans-proteins from all comparisons, we identified a set of 541 unique trans-transcripts
400 and 203 unique trans-proteins (duplicates and hypothetical proteins removed, **Suppl. Tables S1.G &**
401 **S1.H**). GO-overrepresentation analysis showed that no GO terms were significantly enriched in the
402 pooled list of trans-transcripts or trans-proteins from all comparisons (FDR adjusted *p* threshold 0.05).
403 However, it must be noted here that GO- and other annotations in *Leishmania* are often relatively
404 unspecific due to the parasite's distinct evolutionary position, and therefore low gene homology with
405 better studied model organisms. For example, only 48.7% of the *L. donovani* genes could be annotated
406 with at least one GO term, and even these are often of the highest hierarchies (low specificity)[11]. As an
407 alternative enrichment analysis approach using the full genomic annotations directly, we first manually
408 checked if any frequent keywords could be observed occurring in annotations of the trans-proteins
409 and identified 10. We then compared if these keywords were significantly enriched versus all detected
410 proteins with a Fisher Exact test. 7 of these were indeed significantly enriched of which 3 related to
411 protein metabolism: peptidase activity (Adj. $p = 9.4 \times 10^{-5}$), heat-shock protein (Adj. $p = 2.3 \times 10^{-2}$) and
412 chaperon(in) (Adj. $p = 2.1 \times 10^{-2}$), 2 related to metabolism: glutamate (Adj. $p = 1.9 \times 10^{-2}$) and long-
413 chain-fatty-acid-CoA ligase (Adj. $p = 5.4 \times 10^{-3}$), and two others: mitochondrial (Adj. $p = 3.0 \times 10^{-2}$) and
414 hypothetical protein (FDR adjusted $p = 3.4 \times 10^{-7}$). At RNA-level none of the keywords were significant.
415 This demonstrates that differential trans-proteins associated with aneuploidy are not a random
416 occurrences, but related to specific functional classes and processes. Particularly relevant in this
417 context are the enrichment of heat-shock proteins and chaperones, as aneuploidy is known to burden
418 protein folding pathways[50].

419 **Table 1:** Total numbers (N) of differential (FDR adjusted $p < 0.05$) transcripts and proteins between all possible pairs of our 6 *L.donovani*
420 strains. Each time, it is indicated what proportion of these transcripts or proteins originate from chromosomes that have somy changes
421 (because of aneuploidy) in the strain pair (Cis) or chromosomes that have identical somies between the two strains (Trans). Five comparisons
422 are not shown because they contained less than 15 differential transcripts or proteins. These are shown in Supplementary table 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	

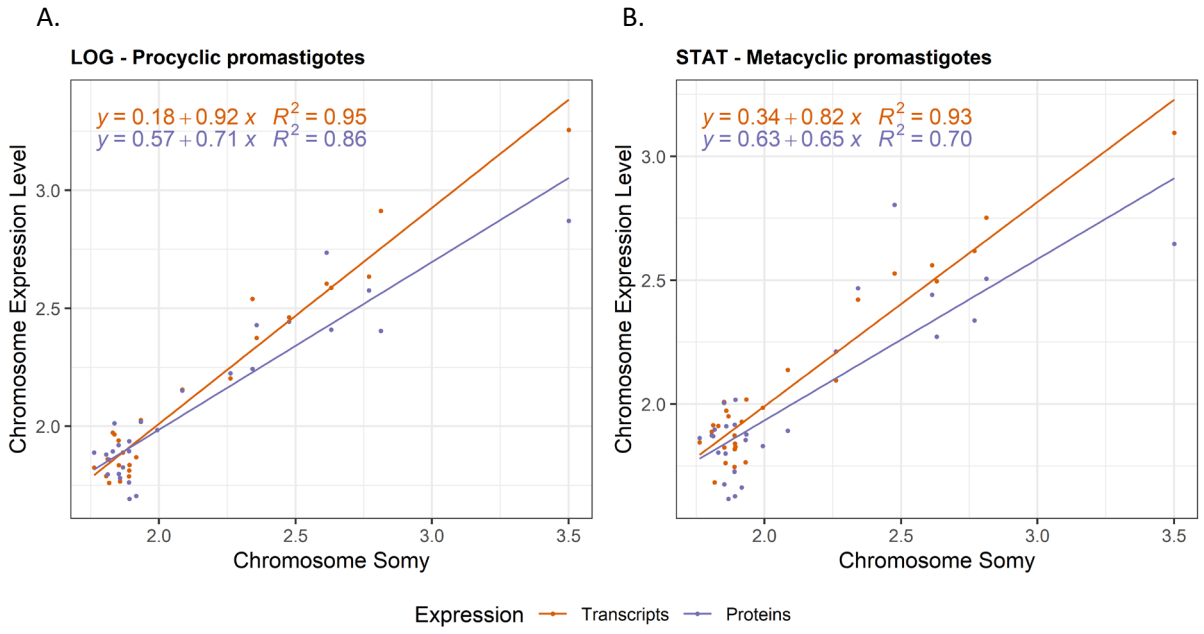
423

424 ## Aneuploidy globally affects transcriptome and proteome regardless of life stage

425 So far, this study discussed results from the logarithmic growth stage (LOG), which essentially consists
426 of *Leishmania* procyclic promastigotes (replicating, non-infectious). However, despite the absence of
427 transcriptional regulation of individual protein-coding genes, transcripts and proteins can undergo
428 post-transcriptional regulation during parasite differentiation. This raised the question whether the
429 correlation between chromosome somy, average transcript abundance, and average protein
430 abundance is also present in other life stages. Hence, for three strains, we compared the LOG growth
431 with the STAT growth phase. STAT is highly enriched for metacyclic promastigotes, which are non-
432 replicating, have a distinct morphology, are metabolically different, and infectious to the human host
433 [51]. This was validated with metacyclogenesis markers META1 and HASPB, which were significantly
434 upregulated in STAT as they are expected to be (Adj. $p < 10^{-13}$, **Suppl. Table 1.H**). Transcriptome and
435 proteome measurements and analyses were carried out exactly as for the LOG phase.

436 For both LOG and STAT, a chromosome's somy was proportional to its average transcript expression
437 level, and protein expression level (**Fig 3**, $p < 2 * 10^{-16}$ for transcript and protein in both LOG and STAT).
438 Dosage compensation was apparent at the protein level, in both LOG and STAT growth phases. This

439 suggests that the gene dosage effects associated with aneuploidy remain present throughout the
440 parasite's life cycle.



441

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

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

446 We ultimately aimed to understand if aneuploidy in *L. donovani* is reflected in the parasite's
447 metabolome. Therefore, we carried out LC-MS metabolomics and confidently identified 78
448 metabolites across the six aneuploid study strains (**Suppl. Table S5**). In contrast to transcripts and
449 proteins, there is no straightforward relation between a metabolite and a specific gene. Genes might
450 affect metabolites directly (e.g. an enzyme), indirectly, affect multiple metabolites, or affect no
451 metabolite at all. Therefore, we turned to an indirect strategy to integrate the metabolomic data with
452 our previous observations.

453 We determined if differences in aneuploidy between strains in LOG correlate with their metabolic
454 differences. As such, we calculated the pairwise genomic (somy), transcriptomic, proteomic, and
455 metabolic Euclidean distance (based on all measured chromosomes, transcripts, proteins and
456 metabolites) between each possible strain pair, and correlated these distances (**Fig 4**). The larger the
457 distance between these 2 strains, the more different their respective genomes, transcriptomes,

458 proteomes, or metabolomes are from one another. Consistent with our other analyses, we found
459 strong correlations between the genomic and transcriptomic distance, between the transcriptomic
460 and proteomic distance, and between the genomic and proteomic distance ($p \leq 0.001$ for all
461 comparisons). However, the metabolome initially did not show any correlation with genome,
462 transcriptome, and proteome. We suspected this might be caused by the high intrinsic levels of
463 random variation of a subset of metabolites (by either technical or biological causes). Therefore, in the
464 second phase of the analysis, we filtered on metabolites that had at least once a significant fold change
465 (adjusted $p < 0.05$) between any pair of two strains. This strategy selects only metabolites that are
466 stable enough to reflect true differences in metabolite levels between strains. The resulting
467 'differentiating metabolome' consisting of 34 metabolites correlated strongly with all omic layers ($p <$
468 0.05, Fig 4). This demonstrates that the degree of karyotype difference between two strains correlates
469 with their degree of metabolic differences. Specifically, the differentiating metabolome included 9
470 amino acids (alanine, aspartate, cysteine, leucine/isoleucine (not distinguishable), methionine,
471 phenylalanine, tryptophan, tyrosine, and valine), suggesting the amino acid metabolism is an
472 important molecular difference between these aneuploid parasites.

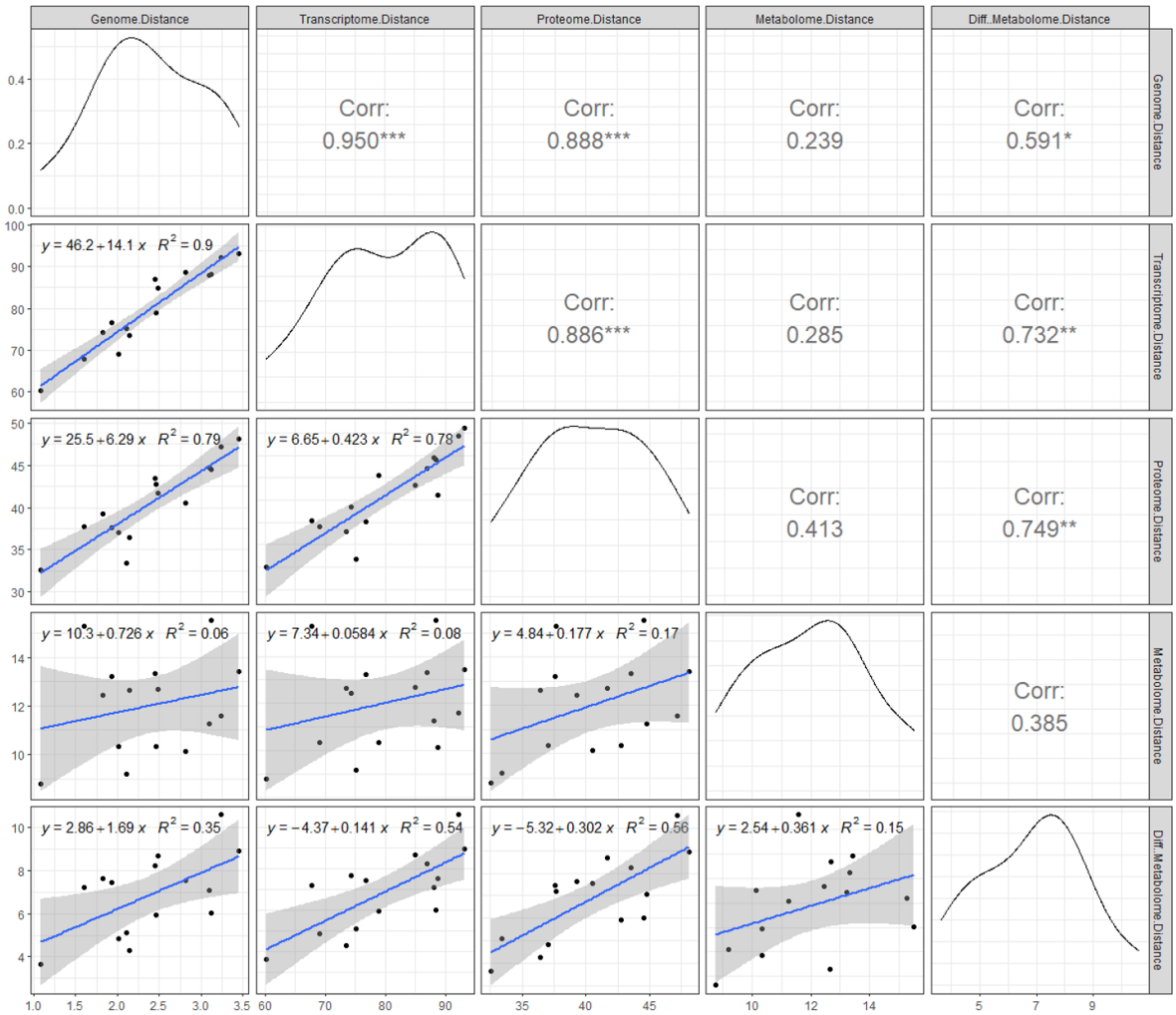


Figure 4: How differences between two *L. donovani* strains on 1 'omic layer correlate with their differences on another omic layer. The plots show the pairwise correlations between the genome distance (aneuploidy), transcriptome distance, proteome distance, metabolome distance (all 78 metabolites), 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 karyotype dynamicity and the drastic karyotype changes between experimental conditions stress the central importance of aneuploidy for the parasite to adapt to novel environments [10]. However, it remained unclear how this large genomic unstability affects the parasite as a system. Moreover, *Leishmania*'s distinct phylogenetic status and unique gene expression system disallows 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 model that lacks transcriptional regulation of individual protein-coding genes.

We confirmed that transcript abundance is affected across the entirety of aneuploid chromosomes. 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* [11,16,52]. We further showed that this relation is almost equivalent, with transcription levels closely mirroring chromosome copy number changes (transcript abundance = 0.94x the somy). It is important to note here that to establish this 0.94x equivalence, we compared the same chromosome to its own disomic state as the reference transcript expression level at disomy. This does not exclude that chromosomes (or, on a smaller scale, cistrons) can have inherently different transcription levels. Indeed, it has previously been shown that chromosome 31 has the expression of a disomic chromosome in its default tetratomic state[16]. 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 0.76x the somy.

This reduced impact of gene dosage on the protein level was due to the compensation of a subset of proteins. While many 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 the abundance of these proteins is somehow post-transcriptionally attenuated in response to a dosage change. We made two observations concerning the identity of compensated proteins that might explain this attenuation. Firstly, compensated proteins were enriched for subunits of macromolecular protein complexes (15.6% versus 1.8% in the background set). This observation matches several observations in aneuploid yeast and human

514 (including cancer) cell lines [46,47,53-56] that also detected protein complex subunits to be highly prevalent
515 amongst compensated proteins. It can be explained by the fact that individual subunits of
516 macromolecular complexes are typically unstable unless assembled into a stable complex [46]. As
517 subunits of a single complex are often encoded on different chromosomes and produced in balanced,
518 stoichiometric amounts, an increase in an individual subunit (here: by aneuploidy) will often not lead
519 to more complex formation [46,47,57]. Instead, the protein will not have an available binding partner(s),
520 remains unstable, and rapidly degrades. Possibly, this stochiometric imbalance caused by aneuploidy
521 is somehow detrimental to the parasite. One argument in this direction is that we and previous studies
522 found smaller *Leishmania* chromosomes to be more frequently aneuploid, which matches observation
523 in yeast and humans (chromosome 21 has the smallest number of genes in humans and the least
524 severe autosomal aneuploid) [10,23,58]. Additionally, it has been shown in yeast that genes that are
525 deleterious when overexpressed are enriched for protein complex subunits[59]. Secondly, we observed
526 a correlation between the degree of compensation and the cellular destination of the compensated
527 protein. Two secreted proteins associated with the Golgi apparatus were the most compensated (i.e.
528 their expression was close to disomic levels), which could be explained by the fact that secreted
529 proteins do not accumulate in the cell. Correspondingly, dosage compensation of secreted proteins
530 was also observed in a recent study of human fibroblasts with trisomy 21 [46]. Proteins from
531 mitochondria, the nucleus, and the peroxisome were also significantly more compensated than
532 cytoplasmatic proteins, which had the lowest level of compensation (i.e., cytoplasmic proteins
533 followed gene dosage the best). A potential explanation might be that mitochondrial, nucleic, and
534 peroxisome proteins require active transport into their destined organelle. This transport requires
535 extensive interaction with other proteins, potentially explaining why their abundance scales less with
536 increased protein production. In aneuploid (disomic II) strains of *S. cerevisiae*, increased translation
537 was found of genes related to protein trafficking pathways of the endomembrane system and nucleus,
538 also suggesting dependency on this system [60].

539 Partial chromosomal amplification can occur in *Leishmania*. We have previously observed this
540 phenomenon for chromosome Ld23 [44,45] and in the present study for Ld35. Interestingly, gene dosage
541 proportionally affected transcript and protein abundance for these subchromosomal amplification as
542 well. This suggests that our findings can be extrapolated to segmental aneuploidy as well.

543 Further, we demonstrated that these gene dosage changes remain important throughout at least
544 several life stages of the parasite. Indeed, we observed an similar impact of gene dosage on
545 transcriptome and proteome in either LOG (enriched for procyclic promastigotes, proliferative stage)

546 or STAT (enriched for metacyclic promastigotes, infectious stage) growth phases. This is striking, as the
547 parasite featuring extensive transcriptome and proteome changes during its life cycle. For example,
548 transcriptomics studies have shown that up to 30-50% of transcripts are differentially expressed
549 between procyclic and metacyclic promastigotes [15,61]. Similarly, 10-40% of proteins are differentially
550 expressed between the parasite's promastigote and amastigote life stages [11,62].

551 A limited proportion of differential transcripts between aneuploid strains (23.6%) and the majority of
552 differential proteins (56.9%) originated from chromosomes without a change in somy, so-called trans-
553 transcripts, and trans-proteins. As we excluded genes with CNVS between strains from our analysis
554 and SNPs and INDELs were very rare, we hypothesize that the observed trans-effects could result from
555 aneuploidy, although our study does not allow to test for a causal relationship. High abundances of
556 trans-transcripts have been reported in many other aneuploidy studies on cells from humans, yeast,
557 maize, *Drosophila* and *Arabidopsis*, as well as on patients with Turner and Klinefelter syndrome [46-50,63-
558 66]. These trans-transcriptomes were typically enriched for transcription factors, genes related to
559 metabolic processes, protein metabolism, or cellular stress responses. It supports the idea that the
560 cellular system mediates the primary impact of aneuploidy by inducing transcriptional changes to
561 trans-genes. Strikingly, we did not observe this coordinated transcriptomic response in *Leishmania*,
562 with relatively low proportions of differential trans-transcripts between aneuploid strains and no
563 significant enrichment of specific functional classes amongst these. On the one hand, this matches the
564 constitutive gene abundance model of the parasite, where the transcription of individual protein-
565 coding genes can not be controlled by transcription factors. On the other hand, *Leish*mania does
566 feature substantial post-transcriptional regulation of transcript levels (e.g. between life stages), so in
567 theory, extensive regulation in response to aneuploidy could have occurred at this level, but we did
568 not observe this. Instead, we did find a surprisingly high proportion of trans-proteins. It appears that
569 trans-protein abundances are directly modulated at the translation or protein-stability level, as the
570 majority did not have underlying transcript-level changes. Interestingly, these trans-proteins were
571 significantly enriched for specific protein groups, including protein metabolism-related chaperones
572 and chaperonins, peptidases, and heat-shock proteins. This is relevant as aneuploidy is known to put
573 an extra burden on the protein-folding system and suggests a similar mediation of aneuploidy-induced
574 effects as in other Eukaryotes, but instead regulated directly at the protein level. This direct
575 modulation at protein-level (without underlying transcriptional change) may in fact also be present in
576 other Eukaryotes, but could be overshadowed by the transcription factor-induced changes. For this

577 reason, we propose *Leishmania* as an attractive model to study processes of post-transcriptional
578 modulation of protein levels in the context of aneuploidy.

579 Finally, we showed that the degree of aneuploidy differences between strains matched the degree of
580 transcriptome, proteome, and differential metabolome changes. This is an important observation as it
581 suggests that aneuploidy in *Leishmania* has the capacity to drive metabolic variation, which is closely
582 linked to the phenotype. It also matches our previous observations where we linked *L. donovani* gene
583 copy number variants directly to changes in metabolic pathways [31]. Ultimately, it supports the view
584 that aneuploidy can be adaptive and drive the metabolic changes that the parasite needs to survive in
585 a novel environment.

586 *Conclusion*

587 In summary, our study shows that aneuploidy in *Leishmania* globally and proportionally impacts the
588 transcriptome and proteome of affected chromosomes throughout the parasite's life cycle *in vitro*. In
589 *Leishmania* strains with a closely related genetic background, the degree of these aneuploidy-induced
590 changes ultimately correlates with the degree of metabolomic difference. Similarly as in other
591 Eukaryotes, we observed dosage compensation at protein level for protein-complex subunits and
592 secreted proteins. Additionally, we found non-cytoplasmic proteins to be more compensated that
593 cytoplasmic ones. While most aneuploid Eukaryotes display extensive regulation of trans-transcripts
594 to mediate the effects of aneuploidy, we did not observe this to the same extend in *Leishmania*. This
595 suggests that the post-transcriptional mRNA regulation system of the parasite does not mediate the
596 impact of aneuploidy. Instead, *Leishmania* features a surprisingly high number of trans-proteins, which
597 are likely post-transcriptionally modulated. It thus seems that aneuploidy is accompanied by trans-
598 proteomic changes in Eukaryotes, regardless of the system of regulation. Post-transcriptional
599 modulation of trans-proteins may also be present in other Eukaryotes, but could be overshadowed by
600 the transcription factor-induced gene regulation that is absent in *Leishmania.* For this reason, we
601 believe that *Leishmania* is an attractive model to study processes of post-transcriptional modulation
602 of protein levels in the context of aneuploidy and beyond.

603

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. *Developmental 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 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).

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

13 Sterkers, Y., Lachaud, L., Crobu, L., Bastien, P. & Pagès, M. FISH analysis reveals aneuploidy and continual generation of chromosomal mosaicism in Leishmania major. *Cell Microbiol* **13**, 274-283, doi:10.1111/j.1462-5822.2010.01534.x (2011).

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).

649 20 Dumetz, F. *et al.* Modulation of Aneuploidy in Leishmania donovani during Adaptation to
650 Different In Vitro and In Vivo Environments and Its Impact on Gene Expression. *MBio* **8**,
651 doi:10.1128/mBio.00599-17 (2017).
652 21 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*
653 *preprint arXiv:1303.3997* (2013).
654 22 Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and
655 population genetical parameter estimation from sequencing data. *Bioinformatics (Oxford,*
656 *England)* **27**, 2987-2993, doi:10.1093/bioinformatics/btr509 (2011).
657 23 Downing, T. *et al.* Whole genome sequencing of multiple Leishmania donovani clinical isolates
658 provides insights into population structure and mechanisms of drug resistance. *Genome*
659 *research* **21**, 2143-2156, doi:10.1101/gr.123430.111 (2011).
660 24 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput
661 sequencing data. *Bioinformatics (Oxford, England)* **31**, 166-169,
662 doi:10.1093/bioinformatics/btu638 (2015).
663 25 Quinn, T. P., Erb, I., Richardson, M. F. & Crowley, T. M. Understanding sequencing data as
664 compositions: an outlook and review. *Bioinformatics (Oxford, England)* **34**, 2870-2878,
665 doi:10.1093/bioinformatics/bty175 (2018).
666 26 Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-
667 based shotgun proteomics. *Nature Protocols* **11**, 2301-2319, doi:10.1038/nprot.2016.136
668 (2016).
669 27 Berg, M. *et al.* Experimental resistance to drug combinations in Leishmania donovani:
670 metabolic and phenotypic adaptations. *Antimicrobial agents and chemotherapy* **59**, 2242-2255
671 (2015).
672 28 Berg, M. *et al.* Metabolic adaptations of Leishmania donovani in relation to differentiation,
673 drug resistance, and drug pressure. *Mol Microbiol* **90**, 428-442, doi:10.1111/mmi.12374
674 (2013).
675 29 Benton, H. P., Want, E. J. & Ebbels, T. M. Correction of mass calibration gaps in liquid
676 chromatography-mass spectrometry metabolomics data. *Bioinformatics (Oxford, England)* **26**,
677 2488-2489, doi:10.1093/bioinformatics/btq441 (2010).
678 30 Scheltema, R. A., Jankevics, A., Jansen, R. C., Swertz, M. A. & Breitling, R. PeakML/mzMatch: a
679 file format, Java library, R library, and tool-chain for mass spectrometry data analysis.
680 *Analytical chemistry* **83**, 2786-2793, doi:10.1021/ac2000994 (2011).
681 31 Cuypers, B. *et al.* Integrated genomic and metabolomic profiling of ISC1, an emerging
682 Leishmania donovani population in the Indian subcontinent. *Infection, Genetics and Evolution*
683 **62**, 170-178, doi:https://doi.org/10.1016/j.meegid.2018.04.021 (2018).
684 32 Prince, J. T. & Marcotte, E. M. Chromatographic alignment of ESI-LC-MS proteomics data sets
685 by ordered bijective interpolated warping. *Anal Chem* **78**, 6140-6152, doi:10.1021/ac0605344
686 (2006).
687 33 Tautenhahn, R., Böttcher, C. & Neumann, S. Highly sensitive feature detection for high
688 resolution LC/MS. *BMC Bioinformatics* **9**, 504, doi:10.1186/1471-2105-9-504 (2008).
689 34 Doyle, M. A. *et al.* LeishCyc: a biochemical pathways database for Leishmania major. *BMC Syst*
690 *Biol* **3**, 57-57, doi:10.1186/1752-0509-3-57 (2009).
691 35 Fahy, E. *et al.* Update of the LIPID MAPS comprehensive classification system for lipids. *J Lipid*
692 *Res* **50 Suppl**, S9-S14, doi:10.1194/jlr.R800095-JLR200 (2009).
693 36 Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on
694 genomes, pathways, diseases and drugs. *Nucleic acids research* **45**, D353-d361,
695 doi:10.1093/nar/gkw1092 (2017).
696 37 Wishart, D. S. *et al.* HMDB 4.0: the human metabolome database for 2018. *Nucleic acids*
697 *research* **46**, D608-d617, doi:10.1093/nar/gkx1089 (2018).

38 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:https://doi.org/10.1016/j.ijpddr.2017.10.002 (2017).
39 Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).
40 Krzywinski, M. I. *et al.* Circos: An information aesthetic for comparative genomics. *Genome research*, doi:10.1101/gr.092759.109 (2009).
41 Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9**, 90-95, doi:10.1109/MCSE.2007.55 (2007).
42 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).
43 Imamura, H. *et al.* Evolutionary genomics of epidemic visceral leishmaniasis in the Indian subcontinent. *Elife* **5**, doi:10.7554/eLife.12613 (2016).
44 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).
45 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).
46 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).
47 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).
48 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).
49 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).
50 Sheltzer, J. M., Torres, E. M., Dunham, M. J. & Amon, A. Transcriptional consequences of aneuploidy. *Proc Natl Acad Sci U S A* **109**, 12644-12649, doi:10.1073/pnas.1209227109 (2012).
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 McShane, E. *et al.* Kinetic Analysis of Protein Stability Reveals Age-Dependent Degradation. *Cell* **167**, 803-815.e821, doi:10.1016/j.cell.2016.09.015 (2016).
56 Schukken, K. M. & Sheltzer, J. M. Extensive protein dosage compensation in aneuploid human cancers. *bioRxiv*, 2021.2006.2018.449005, doi:10.1101/2021.06.18.449005 (2021).
57 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).
58 Gilchrist, C. & Stelkens, R. Aneuploidy in yeast: Segregation error or adaptation mechanism? *Yeast* **36**, 525-539, doi:10.1002/yea.3427 (2019).

59 Robinson, D., Place, M., Hose, J., Jochem, A. & Gasch, A. P. Natural variation in the consequences of gene overexpression and its implications for evolutionary trajectories. *eLife* **10**, e70564, doi:10.7554/eLife.70564 (2021).
60 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).
61 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).
62 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).
63 Sun, L. *et al.* Dosage compensation and inverse effects in triple X metafemales of Drosophila. *Proceedings of the National Academy of Sciences* **110**, 7383-7388, doi:doi:10.1073/pnas.1305638110 (2013).
64 Hou, J. *et al.* Global impacts of chromosomal imbalance on gene expression in Arabidopsis and other taxa. *Proc Natl Acad Sci U S A* **115**, E11321-e11330, doi:10.1073/pnas.1807796115 (2018).
65 Raznahan, A. *et al.* Sex-chromosome dosage effects on gene expression in humans. *Proceedings of the National Academy of Sciences* **115**, 7398-7403, doi:doi:10.1073/pnas.1802889115 (2018).
66 Zhang, X. *et al.* Integrated functional genomic analyses of Klinefelter and Turner syndromes reveal global network effects of altered X chromosome dosage. *Proceedings of the National Academy of Sciences* **117**, 4864-4873, doi:doi:10.1073/pnas.1910003117 (2020).

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.

785 # Supplementary Tables

786 **Supplementary Table S1.A**

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

788 **Supplementary Table S1.B**

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

791 **Supplementary Table S1.C**

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

793 **Supplementary Table S1.D**

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

795 **Supplementary Table S1.E**

796 Dosage compensation model data, coefficients and *p*-values.
797

798 **Supplementary Table S1.F**

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

801 **Supplementary Table S1.G**

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

804 **Supplementary Table S1.H**

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

807 **Supplementary Table S1. I**

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

809 **Supplementary Table S2**

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

811 **Supplementary Table S3**

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

813 **Supplementary Table S4**

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

815 **Supplementary Table S5**

816 Detailed statistics about all metabolites detected in this study.

817 Supplementary Figures

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

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