Title: Replication stress increases de novo CNVs across the malaria parasite genome

Short title: Stress-induced de novo CNVs

Authors: Noah Brown[1], Aleksander Luniewski[1], Xuanxuan Yu[2,3], Michelle Warthan[1], Shiwei Liu[1,4], Julia Zulawinska[1], Syed Ahmad[1], Nadia Prasad[1], Molly Congdon[5], Webster Santos[5], Feifei Xiao[2], Jennifer L Guler[1*]

Affiliations:

[1]University of Virginia, Department of Biology, Charlottesville, VA, USA

[2]University of Florida, Department of Biostatistics, College of Public Health and Health Professions, Gainesville, FL, USA

[3]University of Florida, Department of Surgery, College of Medicine, Gainesville, FL, USA

[4]Current affiliation: Indiana University School of Medicine, Indianapolis, IN, USA

[5]Virginia Tech, Department of Chemistry, Blacksburg, VA, USA

*****Corresponding Author

Email: jlg5fw@virginia.edu

ABSTRACT

Changes in the copy number of large genomic regions, termed copy number variations (CNVs), contribute to important phenotypes. CNVs are readily identified using conventional approaches when present in a large fraction of the cell population. However, CNVs in only a few genomes are often overlooked but important; if beneficial, a de novo CNV that arises in a single genome can expand during selection to create a population of cells with novel characteristics. While single cell methods for studying de novo CNVs are increasing, we continue to lack information about CNV dynamics in rapidly evolving microbial populations. Here, we investigated de novo CNVs in the genome of the *Plasmodium* parasite that causes human malaria. The highly AT-rich *P. falciparum* genome readily accumulates CNVs that facilitate rapid adaptation. We employed low-input genomics and specialized computational tools to evaluate the impact of sub-lethal stress on the de novo CNV rate. We observed a significant increase in genome-wide de novo CNVs following treatment with an antimalarial compound that inhibits replication. De novo CNVs encompassed genes from various cellular pathways participating in human infection. This snapshot of CNV dynamics emphasizes the connection between replication stress, DNA repair, and CNV generation in this important microbial pathogen.

INTRODUCTION

Changes in the copy number of large genomic regions, termed copy number variations (CNVs), are a source of phenotypic diversity for many organisms (as reviewed in (1-3)), including rapidly evolving microbes such as bacteria, yeast, and viruses (4-6). CNVs also contribute to many diseases from cancer to blood, metabolic, neurological, and infectious diseases (7-19). Increased access to genome sequencing has facilitated the identification of these important genomic rearrangements, especially following selection. CNVs that are identified using standard “bulk” analysis approaches (e.g. read coverage methods) are present in a large fraction of the cell population (>50% (2)). However, those in a minority of

45 genomes, or even a few genomes, are "averaged" away during analysis steps. This artifact limits our ability
46 to assess how CNVs arise and contribute to the genomic diversity of individual cells within a population.
47 In order to observe a genome's evolutionary potential in the absence of selection, we require approaches
48 specifically designed to detect CNVs that are not present in a predecessor "parental" genome. Due to their
49 rare and novel nature, these events are termed "de novo" CNVs (3,20-24). Early experimental progress
50 detecting de novo CNVs involved cloning individual cells, which takes time, is prone to contamination, and
51 prevents detection of detrimental CNVs (20-22). Misalignment of short-reads to reference genomes (i.e.
52 discordant or split reads) can also indicate the presence of de novo CNVs, but false positives are common
53 if matched normal samples are not available (reviewed in (25)). Recent single cell-based methods have
54 successfully assessed de novo CNVs during experimental evolution, disease progression, and tissue
55 development (as reviewed in (25-29)). While the reach of single cell analysis is expanding, we continue to
56 lack information about de novo CNVs in microbes and their dynamics in evolving pathogen populations.
57 The *Plasmodium* parasite that causes malaria readily accumulates kb-sized CNVs in its genome (30-33).
58 This protozoan propagates within the mosquito vector and then the human host, where it infects the liver
59 and cyclically invades red blood cells (erythrocytic cycle, **Fig. 1A**). CNVs impact parasite survival, allowing
60 evasion of clinical detection (34), expansion of beneficial gene families (35), invasion of new host cells
61 (36), and development of antimalarial resistance (37-39). We are specifically interested in the genetic
62 diversity of one species of malaria, *P. falciparum*, since this may explain its rapid adaptation to new drugs
63 and host environments (40). Due to its relatively small, AT-rich genome (23 Mb, 19.4% GC (41)), low-input
64 genomics is challenging in this single cell protozoan. However, we previously optimized a single cell
65 genomics approach for erythrocytic stage *P. falciparum* parasites and recently developed novel
66 computational tools to evaluate de novo CNVs in challenging genomes (42-44).
67 Here, we combined these advancements and made additional improvements to directly investigate de
68 novo CNV formation in the *P. falciparum* genome. Since various types of cellular stress can induce genetic

69 change (reviewed in (2,45-47)), we also evaluated the impacts of sub-lethal stress on de novo CNV
70 development. Using a low-input genomics pipeline, we observed that replication stress increased the
71 number of de novo CNVs across the parasite genome. This finding provides key evidence for adaptive
72 amplification in an important pathogen that has evolved strategies to encourage CNV formation.
73

74 **MATERIALS AND METHODS**

75 **Parasite Lines, Compounds, and Treatments**

76 We acquired *Dd2* (MRA-156) and *FCR3* (MRA-731) parasite lines from Bei Resources (ATCC, Manassas,
77 VA). In this study, we were interested in detecting sub-clonal levels of genomic diversity that occur
78 naturally in cell culture (i.e. untreated conditions); therefore, we did not re-clone parasite lines prior to
79 treatment. For low-input genomics, we grew parasites in complete RPMI 1640 with HEPES (Thermo Fisher
80 Scientific, Waltham, MA) supplemented 0.5% Albumax II Lipid-Rich BSA (MilliporeSigma, Burlington, MA)
81 and 50 mg/L hypoxanthine (Thermo Fisher Scientific) and donor A+ human erythrocytes (BioIVT,
82 Hicksville, NY). We grew all cultures at 3% hematocrit at 37°C and individually flushed flasks with 5%
83 oxygen, 5% carbon dioxide, and 90% nitrogen gas. We diluted cultures with uninfected erythrocytes and
84 changed the culture medium every other day to keep parasitemia below 2% during maintenance. We
85 confirmed that all cultures were negative for mycoplasma contamination approximately monthly using a
86 LookOut Mycoplasma PCR detection kit (MilliporeSigma). We synthesized DSM1 as in previous studies
87 (48,49) and purchased aphidicolin (MilliporeSigma).
88

89 *Pilot assessments of DSM1 treatment*

90 We conducted multiple independent pilot assessments comparing DSM1 treatment with a replication
91 inhibitor, aphidicolin (see general scheme in **Fig. S1A**). We first acquired high ring-stage cultures (>60%)
92 by synchronizing mixed cultures of parasites twice with 5% sorbitol, ~48hrs apart. For these pilot

93 assessments, we used sorbitol synchronization to increase the level of ring stage in the cultures. We chose
94 this approach instead of magnetic purification or Percoll gradients in order to limit the time parasites were
95 outside of culture conditions. We then applied 1µM DSM1, solvent control (dimethylsulfoxide, DMSO), or
96 4.4µM aphidicolin to ring-dominant parasite populations for 12hrs. Following treatment, we washed
97 parasites with sterile 1x phosphate-buffered saline (PBS, Thermo Fisher Scientific), returned them to
98 complete RPMI, and allowed parasites to recover for >36hrs, depending on the experiment (detailed in
99 **Fig. S1B-E**). We tracked parasitemia and parasite viability on an Accuri C6 flow cytometer (BD Biosciences,
100 Franklin Lakes, NJ) as previously performed (42,48,50). At various time points (start at 0h, post-treatment
101 at 12h, or during recovery, **Fig. S1**), we stained parasites with 1x SYBR Green (Thermo Fisher Scientific,
102 stains the parasite nucleus) to assess the proportion of infected erythrocytes (parasitemia) and stage of
103 the parasite development cycle; we also stained parasites with 10nM MitoProbe DiIC1 (5) (Thermo Fisher
104 Scientific, stains active parasite mitochondria) to indicate the proportion of the parasites that were viable
105 after recovery. All pilot experiments were assessed from multiple flasks (>2, **Fig. S1**).
106
107 *DSM1 treatment for low input genomics*
108 For parasite treatment for low-input genomics, we synchronized parasites (as above), applied 1µM DSM1
109 or the DMSO control for 12hrs in duplicate flasks, and allowed recovery for 28-31hrs (**Fig. S2A**). We
110 staggered sample sorting in order to limit the impact on parasite health by removal from culture
111 conditions and allowed parasites to reinvade to produce rings. We removed treatments and tracked
112 parasite number and health as described above (**Fig. S2B-G**). Following reinvasion, we harvested viable 1n
113 ring stage parasites for low-input genomics using flow sorting (details below).
114
115 **Parasite Flow Sorting for Low-input Genomics**

116 *Cell sorter calibration & accuracy assessments.* We calibrated the flow sorter (SH800, Sony Biotechnology,
117 San Jose, CA) using the manufacturer's calibration beads. We accounted for overlaps in the
118 excitation/emission wavelengths using the integrated compensation panel matrix calculation in the SH800
119 software according to the manufacturer's procedure. We also manually calibrated the droplet sorting to
120 the nearest 0.2mm, as recommended by the manufacturer, using the 96-well plate setting (Armadillo
121 high- performance 96-well plate, Thermo Fisher Scientific). We evaluated SH800 sorting accuracy prior to
122 low-input harvest using a colorimetric assay as previously described (51). Briefly, we mixed SYBR
123 Green+/MitoProbe+ parasites (see staining details in *Parasite lines, Compounds, & Treatments*) with
124 horseradish peroxidase enzyme (Thermo Fisher Scientific) at a final concentration of 2.5mg/ml. We then
125 sorted parasites into a 96-well plate filled with TMB-ELIZA substrate (Thermo Fisher Scientific) using the
126 single cell (3 drops) instrument setting, in triplicate plates (**Fig. S3A & B**). Formation of a color in the well
127 (blue, green, or yellow) indicates the successful sorting of the enzyme, and therefore parasites, into the
128 well with the substrate. This assessment allowed us to evaluate the accuracy of SH800 sorting (through
129 the evaluation of success for 1- versus 2-cell wells, **Fig. S3C**), the consistency of sorting (through the
130 evaluation of replicates), and the best plate positions for sorting (through the evaluation of performance
131 in different plate rows/columns). Based on these evaluations, we proceeded with isolating 2- and 10-cells
132 per well and avoided sorting into the top 2 rows and the first and last column of the 96-well plate (**Fig.
133 S3D**).
134
135 *Parasite isolation & storage*. We stained parasites with SYBR Green and MitoProbe DiIC1 (5) in complete
136 RPMI as above (see staining details in *Parasite Lines, Compounds, & Treatments*), gassed the tubes with
137 5% CO_2, 5% 0_2, 90% N, and placed sample on ice to ensure viability prior to flow sorting of viable , ring-
138 stage parasites (SH800, Sony Biotechnology Inc., San Jose, CA). We used a final concentration of 1 x 10^7
139 parasites/ml diluted in sterile 1x PBS (Thermo Fisher Scientific) as input for sorting at the "single-cell

140 setting" (3 drop) into a 96-well plate (Armadillo high performance 96-well plate, Thermo Fisher Scientific)
141 with each well containing 2.375µl of cell lysis buffer (0.025M Tris Ph8.8 (Roche Diagnostics, Indianapolis,
142 IN), 0.01M NaCl (MilliporeSigma), 0.01M KCl (MilliporeSigma), 0.01M $(NH_4)_2SO_4$ (Thermo Fisher Scientific),
143 0.001M EDTA (Promega, Madison, WI), and 10% Triton X-100 (MilliporeSigma)). We gated viable 1n ring-
144 stage parasites (**Fig. S4**) and sorted into wells containing cell lysis buffer with an approximate sorting time
145 of 15 min. After sorting, we centrifuged for 30 seconds in a plate centrifuge (MPS1000, Labnet
146 International, Madison, NJ). We immediately overlaid samples with one drop (approx. 25µl) of light
147 mineral oil (BioReagent grade for molecular biology, Millipore Sigma) and sealed the plates with
148 Microamp® Clear Adhesive Film (Applied Biosystems, Waltham, MA) before storage at -80°C until whole
149 genome amplification. All samples were isolated on the same day and stored frozen for ~1 month before
150 amplification (*FCR3*: 33 days, treated *Dd2*: 35 days, and untreated *Dd2*: 40 days).
151
152 **MALBAC Whole Genome Amplification for Low-input Genomics**
153 We handled each plate separately in order to limit cross-contamination during amplification steps. First,
154 we thawed each plate containing sorted parasites (see *Parasite Isolation & Storage*) and added 1mg/ml
155 Proteinase K in sterile 1x PBS (Thermo Fisher Scientific) to a final volume of 2.5µl per well. We heated the
156 plate in a PCR cycler (C1000, Bio-Rad Laboratories, Hercules, CA) at 50°C for 3hrs, followed by 75°C for 20
157 min and 80°C for 5 min for proteinase k digestion. We amplified parasite genomes using the Multiple
158 Annealing and Looping Based Amplification Cycles (MALBAC) method essentially as previously described
159 ((42), Version 1) with some modifications (Version 2, **Fig. S5**). In summary, 1) we modified the pre-
160 amplification random primer by adding 5 additional degenerate bases with 20% GC-content to increase
161 annealing to AT-rich genome sequences (5'GTGAGTGATGGTTGAGGTAGTGTGGAGNNNNNNNNNNTTT 3');
162 2) we performed 19 of the 21 total linear cycles with the *Bsu* DNA Polymerase (Large Fragment, New
163 England Biosciences), which has a lower optimal reaction temperature (37°C) to improve the amplification

164 on AT-rich sequences (52); 3) we lowered the extension temperature from 40/50°C to 37°C during the
165 linear amplification cycles that used the *Bsu* enzyme (see full cycling parameters in **Fig. S5** and impacts of
166 these changes in **Table S1**); and 4) we integrated robotic pipetting (Mosquito LV, SPT Labtech, Melbourn,
167 UK) to increase the throughput of our assays (from 23 samples in Version 1 to 90 samples in the current
168 Version 2) and limit contamination potential.
169 Overall, we performed 21 total linear cycles (19 cycles with *Bsu* polymerase and 2 cycles with *Bst*
170 polymerase, New England Biolabs, Ipswitch, MA) and 17 total exponential amplification cycles using
171 Herculase II Fusion DNA polymerase (Agilent Technologies, Santa Clara, CA). During amplification steps,
172 we employed standard steps to limit contamination (42). For automated pipetting of the enzyme solution
173 during linear cycles, tips were changed after each round of pipetting. Post-amplification, we purified
174 amplified DNA with Zymo DNA Clean & Concentrator-5columns (Zymo Research, Irvine, CA) according to
175 the protocol and ran 2µl of all samples on 1% agarose gels to check for the presence of DNA (generally, if
176 >30ng/µl, samples could be visualized with a size range of 100 to >1500bp).
177
178 **Assessments of Amplification Success for Low-input Genomics**
179 *DNA quantification.* We quantified the MALBAC-amplified DNA using a Qubit fluorimeter (Qubit 1X dsDNA
180 High Sensitivity Assay Kit, Thermo Fisher Scientific) (**Fig. S6**).
181
182 *Droplet Digital PCR.* To confirm the presence of parasite DNA in MALBAC-amplified samples, we
183 performed droplet digital PCR (ddPCR) as described previously using the QX2000 droplet generator, C1000
184 thermocycler, and QX2000 droplet reader (Bio-Rad Laboratories) (48,53) (**Fig. S7**). We used duplex assays
185 to evaluate two parasite genomic loci concurrently (*pfmdr1:* Forward- TGCCCACAGAATTGCATCTA;
186 Reverse- ACCCTGATCGAAATGGAACCT; Probe - TCGTGTGTTCCATGTGACTG; *pfhsp70:* Forward-
187 TGCTGTCATTACCGTTCCAG; Reverse - AGATGCTGGTACAATTGCAGGA; Probe -

188 AGCAGCTGCAGTAGGTTCATT (Integrated DNA technologies, Newark, NJ). The reaction master mix
189 contained 600nm of forward and reverse primers, 50nm probes, 10µl of ddPCR Supermix for Probes (2x,
190 Bio-Rad Laboratories), 3µl of nuclease-free water (QIAGEN), and 1.5ng (5µl) of template DNA per assay
191 (total of 20µl). We used the following cycling conditions for PCR amplification: 10 min at 95°C initial
192 denaturation step, 1 min at 95°C second denaturation step, and 2 min at 58°C annealing and extension
193 step (ramp rate of 1°C per second), the second denaturation step and the annealing/extension step
194 repeated 60 times, and then 10 min at 98°C to halt the reaction (53). In addition to running amplified
195 samples to assess amplification success, we ran ddPCR with bulk genomic DNA as a positive control, no
196 template controls (water replaced DNA), and material from "no cell" wells to assess cross-well
197 contamination. We considered the samples positive for parasite DNA if there were more than 50 total
198 positive droplets in target-positive clusters. For comparisons to DNA amounts (**Fig. S6**), we plotted total
199 nanograms against total positive droplets (from both *pfmdr1* and *pfhsp70* loci) and performed a simple
200 linear regression in PRISM (GraphPad, La Jolla, CA).
201
202 *High Resolution Melting Assay.* To assess potential contamination between MALBAC-amplified samples,
203 we performed asymmetric PCR amplification of the *pfdhps* locus followed by high-resolution melting
204 (HRM) as described previously (54,55) (**Fig. S8**). The *pfdhps* locus at codon 613 is distinct in *Dd2* and *FCR3*
205 parasite lines (*Dd2*: Ser-613 and *FCR3*: Ala-613, (56)). Each 20µL reaction contained 8µl of the 2.5x
206 LightScanner Master mix (BioFire™ Defense, Salt Lake City, Utah, USA), 1/10µM of forward/reverse
207 primers and 8µM probes targeting the *pfdhps* gene position 613: Forward -
208 CTCTTACAAAATATACATGTATATGATGAGTATCCACTT; Reverse-
209 CATGTAATTTTTGTTGTGTATTTATTACAACATTTTGA; Probe - AAGATTTATTGCCCATTGCATGA/3SpC3,
210 (Integrated DNA technologies), 7µl of nuclease free water, and 3µl of DNA (~0.05ng total). We used the
211 following cycling conditions for PCR amplification with the Rotor-Gene Q instrument with a 72-well rotor

212 (QIAGEN): 95°C for 5 min, 45 cycles of 95°C for 10s, 55°C for 30s, and 72°C for 10s, followed by a pre-melt
213 at 55°C for 90s, and a HRM ramp from 65°C to 90°C, with an increase of 0.1°C every 2s. We plotted the
214 change in fluorescence versus temperature (dF/T) using Rotor-Gene Q software (version 2.3.5, build 1;
215 QIAGEN) and compared HRM peaks of amplified samples to bulk genomic DNA and plasmid controls.
216

217 **Bulk DNA Extraction for Short-Read Sequencing**

218 We extracted bulk DNA for short-read sequencing as previously performed (42). Briefly, we lysed
219 erythrocytes with 1.5% saponin and washed the parasite pellet 3 times with 1x PBS (Thermo Fisher
220 Scientific), before resuspension in a buffered solution (150mM NaCl (MilliporeSigma), 10mM EDTA
221 (Promega Corporation, Madison, WI), and 50mM Tris pH7.5 (Roche Diagnostics)) to a total volume of
222 500µl. We then lysed the parasites with 10% sarkosyl (MilliporeSigma) and 20mg/ml proteainase K
223 (Thermo Fisher Scientific) at 37°C overnight before DNA purification using standard
224 phenol/chloroform/isoamyl alcohol extraction and chloroform washing steps (2 times each, (42)). Finally,
225 we precipitated DNA using 100% ethanol with 100mM of sodium acetate overnight in DNA-lo bind tubes
226 (Eppendorf, Enfield, CT) and then washed twice with 70% ethanol before resuspension in 50µl nuclease
227 free water (QIAGEN). We stored bulk genomic DNA at -20°C until sequencing library preparation.
228

229 **Low-input Genomics Sample Selection & Short-Read Sequencing**

230 *Low-input sample selection.* We selected 16 low-input samples from the *FCR3* plate, 36 samples from the
231 untreated *Dd2* plate, and 38 samples from the treated *Dd2* plate for short-read Illumina sequencing. We
232 based our selection on the quantity of the MALBAC amplified DNA and presence of parasite DNA using
233 ddPCR (**Figs. S6 & S7**). In summary, the majority of *FCR3* and untreated *Dd2* samples yielded quantifiable
234 parasite DNA following MALBAC amplification (53/60 *FCR3* samples and 60/60 untreated *Dd2* samples);
235 in these conditions, we chose samples randomly to proceed with sequencing (indicated in **Fig. S6**). For

236 treated *Dd2* samples, we chose samples for sequencing if they had adequate DNA quantity (>10ng total,
237 30 samples, **Fig. S6**) or had ddPCR results showing the presence of parasite DNA (an additional 8 samples).
238
239 *Short-read sequencing.* Before short-read sequencing, we sheared bulk samples and low-input samples
240 using Covaris M220 Focused Ultrasonicator for 150s and 130s, respectively, to generate fragment sizes of
241 ~350bp as evaluated by an Agilent 2100 Bioanalyzer using the High Sensitivity DNA kit (Agilent, Santa
242 Clara). We adjusted the volume of sheared samples with nuclease-free water up to 50µl. For samples with
243 >100ng (**Table S2**, including bulk and MALBAC amplified samples), we diluted them to 1.2-2ng/µl; for
244 samples <100ng, we proceeded with no dilution. We used NEBNext Ultra II kit (Illumina Inc., San Diego,
245 CA) to prepare libraries for sequencing with 3 cycles of PCR amplification, as performed previously (42).
246 We quantified the resulting libraries using NEBNext Library Quant Kit (Illumina Inc.) before sequencing on
247 the Illumina Nextseq 550 using 150 bp paired-end cycles.
248
249 **Short-Read Sequence Processing & Analysis**
250 *Read processing and alignment*. We performed short-read quality control steps as described previously
251 (40,42). Briefly, we reordered and removed singletons and subsequently interleaved paired reads using
252 BBMap, filtered for high-quality reads (>Q30), trimmed the MALBAC common sequence, PhiX, and
253 Illumina adapters from the remaining reads with the BBDuk tool within BBMap, and aligned reads to the
254 *pf3D7-62_v3* reference genome using the Speedseq genome aligner (57). We removed reads that mapped
255 to VAR regions from bam files according to previously defined genomic coordinates (58). We filtered out
256 reads with low mapping quality (mapq<30) and duplicated reads using SAMtools (59). We calculated the
257 percentage of reads that map to the *P. falciparum* genome by dividing the number of mapq>30 mapped
258 reads by the number of total unaligned reads (mapped reads/total reads). We employed Qualimap to
259 report mean coverage and standard deviation across the genome (60). Using non-overlapping 20kb size

260 bins, we calculated the coefficient of variation of read coverage by dividing the standard deviation of
261 coverage within a bin by the mean across a sample and multiplying by 100 (61,62) (R version 4.4.2). Short-
262 read data is available at NCBI Sequence Reads Archives under project PRJNA1201106.
263
264 *BLAST-based evaluation of read composition.* We blasted reads from low-input genomics samples
265 essentially as previously described (42). We also performed an additional BLAST analysis to understand
266 read composition. We randomly downsampled and extracted 5,000 reads from either 1) all aligned reads
267 (mapq>1) or 2) unmapped reads (mapq>30, **Fig. S9**). To ensure that BLAST alignments consisted of high-
268 quality reads, we manually filtered out low complexity (%N>50) and short sequences (length < 100bp).
269 We used these reads as input for a custom pipeline, where we blasted reads against a database consisting
270 of common genomes downloaded from NCBI (i.e., human, mouse, mycoplasma, bacterial genera, etc).
271 For the BLAST step, we employed a Conda installable BLASTN v2.16.0+ for analyzing the reads (63). In
272 order to count as a hit, the sequence must have 90% of its bases aligned and have an e-value of <e-20.
273 The pipeline is designed to return the top 10 blast hits with these criteria, and either call the most
274 significant hit (lowest e-value), or if alignments from the *Plasmodium* genus were present in the result
275 (but not the most significant hit), call the most significant *Plasmodium* hit. We included this step to
276 account for possible mismatching due to the abundance of highly repetitive sequences in the *Plasmodium*
277 genome. We also used a custom script to quantify the abundance of hits based on the genus related to
278 each hit's accession number. If the pipeline identified a hit, but the genus was not represented in the
279 custom database, the hit was called 'unknown'. The code for the pipeline, including construction of the
280 database, is available at https://doi.org/10.6084/m9.figshare.29885879.
281
282 *Single nucleotide polymorphism analysis*. We performed SNP genotyping and analysis as previously (42),
283 based on the MalariaGen *P. falciparum* Community Project V6.0 pipeline (64-67) using the *pf3D7-62_v3*

284 reference genome (**Fig. S10**). Briefly, we applied GATK's Base Quality Score Recalibration using default
285 settings. We detected potential SNPs using GATK's HaplotypeCaller and subsequently genotyped the SNPs
286 using CombineGVCFs and GenotypeGVCFs. We employed GATK's VariantRecalibrator using previously
287 validated SNP datasets (68) and applied GATK's ApplyRecalibration to assign VQSLOD scores (67). We
288 filtered the resulting SNPs for those with VQSLOD > 6 and for a GT quality metric >20 to ensure high-
289 quality variant calling. We only selected variants flagged as Bi-allelic to simplify the analysis. For SNP
290 Principal Component Analysis (PCA), we merged experiment-wide SNP data (described above) into a single
291 file. Then we merged the VCF into a large matrix and converted the genomic data into numeric
292 information using the 'vcfr' package in R (Version 4.2.3) (https://CRAN.R-project.org/package=vcfR). We
293 excluded individual SNPs if >25% of the samples lacked a call in this position or if all calls were the same
294 for each sample in that position). We scaled the remaining SNPs around the origin using the 'scale' R
295 function. We then calculated the principal components using the 'prcomp' R function and scored the
296 dataset using the 'scores' function from the 'vegan' R package (https://CRAN.R-
297 project.org/package=vegan).

298 **Copy Number Variation Analysis**

299 *CNV calling in bulk samples.* We performed CNV detection for bulk samples similar to as previously
300 described (40,42). Briefly, we called CNVs independently using two methods, CNVnator (read depth based
301 calling, (69)) and LUMPY (split and discordant read based calling, (70)). To Identify CNVs called in both
302 methods, we used SVCROWS to define overlapping CNV regions relative to their size (43). Briefly,
303 SVCROWS uses a reciprocal-overlap-based approach (i.e. two CNVs must be overlapping each other at, or
304 greater than, a defined threshold) to determine if two CNVs are close enough in their genomic position to
305 be called the same. The program utilizes adaptive thresholds for overlap based on the sizes of the CNVs
306 being compared, ensuring that we account for shifts in CNV position, particularly in regions of high

307 variance. For known CNV calls, we used the following SVCROWS input parameters: ExpandRORegion =
308 FALSE, BPfactor = TRUE, DefaultSizes = FALSE, xs = 3000, xl = 10,000, y1s = 300, y1l = 1000, y2s = 50% and
309 y2l = 80%; based on the average size of a *P. falciparum* gene (~2.3kb) and intergenic region (~2kb). Similar
310 to our previous study (42), we identified 3 known CNVs that were called by both LUMPY and CNVnator
311 methods in the core genome of bulk samples (untreated and treated). We determined known CNV
312 boundaries using SVCROWS: *pfmdr1* (*Pf3D7_05_v3*, 888001-970000, 82kb), *pf11-1* (*Pf3D7_10_v3*,
313 1521345-1541576, 20kb), and *pf332* (*Pf3D7_11_v3*, 1950201-1962400, 12kb).
314
315 *Downsampling.* For CNV analysis that assessed downsampled data, we first converted the processed bam
316 files back into FASTQ files using SAMtools and then used the reformat.sh option of BBtools to select 1.3M
317 reads from each FASTQ (represented the fewest number of reads from a sample that passed quality
318 filtering from the final dataset). We then realigned files to the reference genome (*pf3D7-62_v3*).

319 *CNV calling in 2-cell samples.* We employed two methods for CNV detection in the core genome of low-
320 input samples; LUMPY is a split/discordant read strategy with high sensitivity (70), and HapCNV is a read
321 coverage-based strategy designed for haploid genomes (44). We ran LUMPY as part of Speedseq with
322 default parameters as previously described (40). We filtered resulting structural variants to include only
323 duplications (DUP, >1 copy of a region) and deletions (DEL, one less copy of region than the reference).
324 We then filtered those calls for those GQ > 20 to ensure high-fidelity calls. In HapCNV, we used a quality
325 control and bias correction procedure to exclude bins of poor quality and remove bias introduced by GC
326 content and mappability variation. We then constructed a pseudo-reference for each *Dd2* low-input
327 sample using within-*Dd2* information, which enabled control of background noise while preserving CNV
328 signals after normalization. Finally, we used a circular binary segmentation algorithm (CBS, (71)) to detect
329 copy number change points followed by a Gaussian Mixture Model (GMM, (72)) for CNV identification.

330 For statistics, we used PRISM (GraphPad, La Jolla, CA), using unpaired parametric T-tests with Welch's
331 correction. We calculated standard error in Microsoft Excel.
332
333 *Defining CNV regions/determination of "rarity" in CNV calling.* Small differences in sequence quality
334 surrounding a read can lead to shifted breakpoint determination for biologically identical CNVs, which is
335 especially true for low-input genomics datasets (28). To account for this, we used SVCROWS to determine
336 whether two CNV signals were the same within and between samples (see *CNV calling in bulk samples* for
337 input parameters (43)). We assigned the categorizations of "rare" and "common" by assessing the CNV
338 region frequency within datasets as performed in other studies (44,73). "Rare" CNVs were defined as
339 occurring in <10% of the samples within a treatment group; "common" CNVs were defined as occurring
340 in ≥10% of samples within a treatment group; "known" CNVs were defined by CNVs called in bulk samples
341 (see above, *CNV calling in bulk samples*).
342
343 *High-confidence CNV region identification.* To identify "high-confidence" CNV regions called by both
344 HapCNV and LUMPY methods, we compared the 'consensus list' generated by SVCROWS (43) for each
345 detection method by combining them into a single input file. Because there is a large disparity of average
346 CNV region lengths between the two methods (HapCNV = ~40kb, LUMPY = ~4.3kb), we relaxed the
347 stringency of the SVCROWS parameters for this analysis. Our input parameters to generate the "high-
348 confidence" list were as follows: xs = 3000, xl = 6000, y1s = 500, y1l = 1500, y2s = 30% and y2l = 60%. We
349 defined high-confidence CNV regions as those that had >1 match from both HapCNV and LUMPY that was
350 the same type (i.e. either duplication or deletion or mixed). For Venn diagram generation, we calculated
351 overlaps using SVCROWS "Scavenge" mode (43), input parameters: ExpandRORegion = FALSE, BPfactor =
352 TRUE, DefaultSizes = FALSE, xs = 3000, xl = 6000, y1s = 500, y1l = 1500, y2s = 30% and y2l = 60%). We
353 systematically compared lists for each overlap comparison, and if regions had at least one match in an

354 opposing dataset, we considered it a match. We used the *draw.quad.venn* function in the
355 "VennDiagram" R package (R 4.2.1) to generate the diagram.
356
357 *Breakpoint analysis of high-confidence CNV regions.* We assessed features of breakpoints flanking high-
358 confidence CNVs as performed previously (40). Briefly, we used the CNV breakpoint coordinates either
359 set by LUMPY or HapCNV to extract a 2000bp window around the 5' or 3' CNV end (±1000bp from the
360 breakpoint). Although Huckaby et al. solely used LUMPY to determine breakpoints (40), we employed CNV
361 boundary information from both CNV callers (LUMPY and HAPCNV) merged using the Expand-RO function
362 in SVCROWS to construct the conserved CNV region ((43), see parameters in *High-confidence CNV region*
363 *identification*). In this analysis, breakpoints are called using either and/or both HapCNV and LUMPY
364 boundaries on either end, which captures variance but can limit accuracy. For this reason, we only
365 identified breakpoints of high-confidence CNVs, which are conserved between CNV-callers. With these
366 defined breakpoints, we then used Vienna 2.1.9 folding prediction software with Mathews 2004 DNA
367 folding parameters, wherein G-quadruplexes, GU pairing, and lonely base pairs were disallowed to predict
368 stable structures in that region based on Gibbs free energy estimates in sliding 50bp windows. We called
369 stable hairpins by determining local minima of Gibbs free energy across the 2000bp region using the same
370 50bp windows. We reported the position of the stable structure closest to the breakpoint using the
371 threshold of -5.8kCal/mol, which we previously reported as the cutoff for stable secondary structure
372 formation in this genome. If no stable hairpins could be determined within 2kb around the breakpoint,
373 we recorded that breakpoint as "unresolved". We also calculated the GC content of the breakpoints from
374 the 100bp window surrounding the stable hairpin, whether this position was genic or intergenic, and the
375 size of the resulting CNV as the distance between the two breakpoints.
376

377 *Gene Ontology enrichment and protein class identification from high-confidence CNV regions.* We used
378 the online Gene Ontology Resource (geneontology.org) to perform GO enrichment analysis using the
379 PANTHER Classification System (74,75). Since a large portion of the *P. falciparum* genome remains
380 unannotated (PlasmoDB, ~30%) and the majority of molecular functions remain unclassified (92.8%), we
381 used the Panther Protein class assessment (version 19.0, only 55.9% remained unclassified) with default
382 statistics (Fisher's test with FDR adjusted p value of <0.05 for significance, which is recommended for small
383 counts and overlaps between classes). We used the web tool to represent protein classes on pie charts.
384
385 *Clinical importance of high-confidence CNV regions.* We assessed the clinical importance of genes within
386 high-confidence CNVs by first collecting gene IDs from high-confidence duplications and deletions. To
387 generate this gene list, we used SVCROWS "Hunt" mode ((43), input parameters: BPfactor = TRUE,
388 DefaultSizes = FALSE, xs = 3000, xl = 6000, y1s = 300, y1l = 600, y2s = 30, y2l = 60), which takes a
389 secondary input list of known genes (Pf3D7_62_v3, Plasmodb.org) and overlaps CNVs to the gene list
390 using a size-weighted scale. We then performed a literature search to identify genes within the list that
391 are essential for erythrocytic stage, mosquito stage (including gametocyte, gamete, ookinete, sporozoite
392 formation), or liver stage *P. falciparum* or *P. berghei* development. We also included genes that produced
393 known parasite antigens important for immune recognition/vaccine escape or contributed to clinical drug
394 resistance. We noted the Mutagenesis Index Score (MIS) for each gene, which estimates essentiality in
395 erythrocytic stage parasites (76), from the PlasmoDB database (77). To assess overlap of high-confidence
396 CNVs with a large catalogue of *P. falciparum* CNVs from clinical isolates, we collected the gene list from
397 high-frequency CNVs (>1%, >300bp) identified in clinical isolates (Supplementary Table 3 from (38))
398 and manually reformatted the list to match SVCROWS input style guidelines (43). We then assessed
399 how often these genes overlapped with our high-confidence gene list using the SVCROWS "Hunt"
400 mode described above.

401

402 **RESULTS**

403 **Refined low-input genomics pipeline increased efficiency and accuracy**

404 We developed a robust pipeline for assessing the frequency of de novo CNVs in the *P. falciparum* genome.
405 We adapted our single cell genomics method to improve parasite isolation and whole genome
406 amplification steps over our previous study (42). In this modified protocol, we used fluorescence-activated
407 cell sorting (FACS) to isolate viable, 1n parasites to improve efficiency and modified aspects of our whole
408 genome amplification method to improve coverage (**Table S1**, **Fig. S5**). We also sorted low-cell
409 populations to increase accuracy (e.g. 2-cells per well, **Fig. S3**), and added quality control PCR-based steps
410 that confirmed our samples were of high quality prior to short-read sequencing (**Fig. S7 & S8**). The
411 resulting low-input genomics pipeline consisted of basic steps including parasite isolation using flow
412 sorting, whole genome amplification using a modified MALBAC-based approach, quality control
413 confirmation, short-read sequencing, and CNV analysis (**Fig. 1B**).

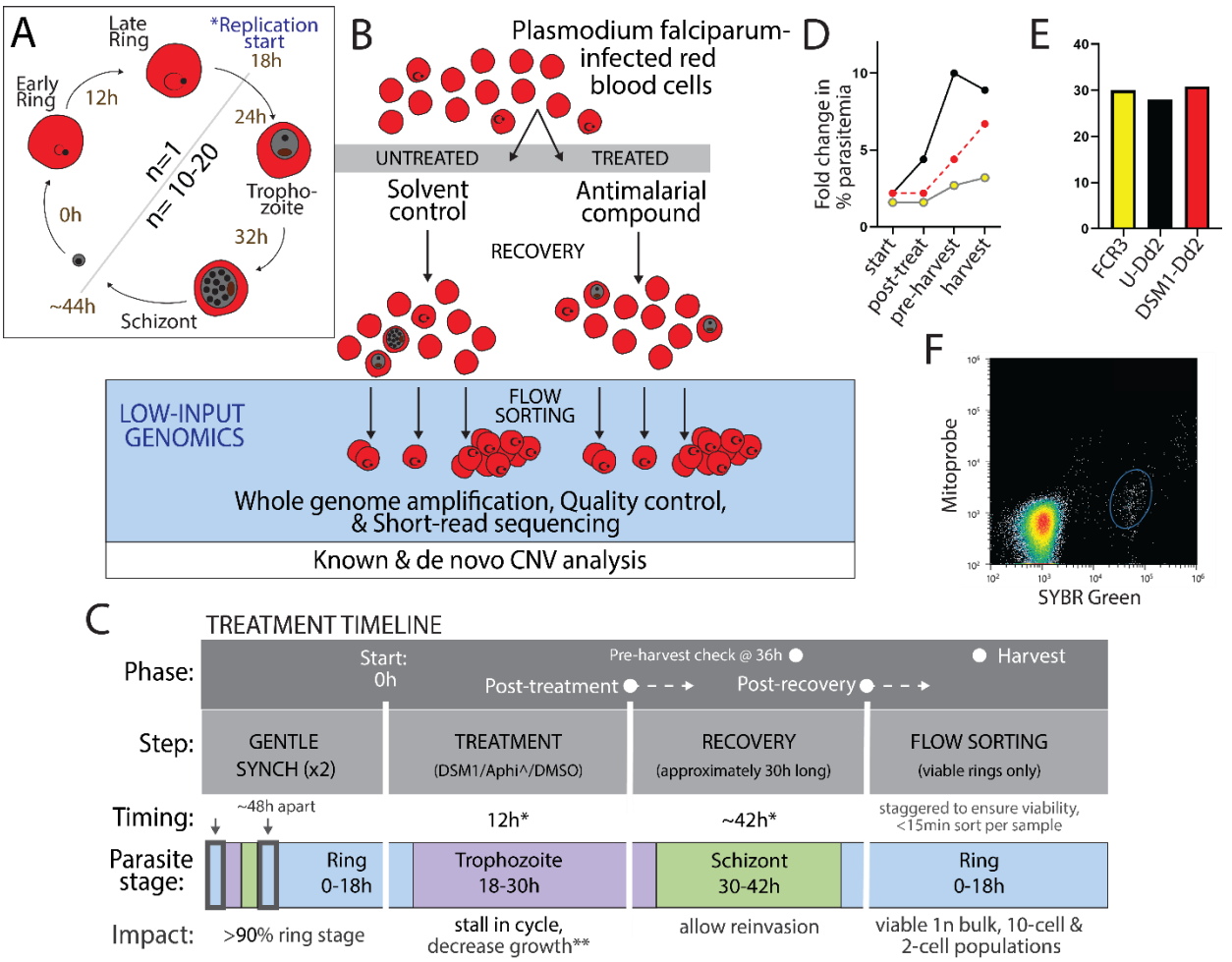


Figure 1. Low-input genomics approach for analysis of malaria parasite genomes under stress. **A**. *P. falciparum* erythrocytic cycle including approximate timing (h) and genome number (n). Cycle length is 44h total for the *Dd2* parasite line (78,79). Ring stage parasites have a single genome, replication begins at trophozoite stage, and schizonts develop with 10-20 individual parasites within a single erythrocyte. **B**. Overview of low-input genomics procedures. Early erythrocytic stage *P. falciparum* parasites grown in vitro were treated with an antimalarial compound (DSM1) or the solvent control (dimethylsulfoxide, DMSO) followed by recovery and reinvasion to produce a new round of ring stage parasites (see details in panel C). Parasite-infected erythrocytes were isolated using flow sorting (10-cell control or 2-cell samples) and genomes were amplified using a modified MALBAC-based whole genome amplification approach. Quality control steps involved DNA quantification to assess amplification success, droplet digital PCR to assess parasite genome amplification, and PCR-high resolution melting to assess sample cross-contamination. After Illumina short-read sequencing, reads were filtered, trimmed and used as input for copy number variation analysis using two approaches, HapCNV and LUMPY. **C**. Timeline of parasite treatment and recovery. For all experiments (pilot and low-input genomics), parasites were synchronized to enrich for rings before short-term treatment was applied. Parasite growth and viability were tracked from start, to post-treatment, to post-recovery, and harvest. The recovery period allowed for completion of the erythrocytic cycle and reinvasion to produce rings before harvest for low-input genomics (see more details on pilot treatments in **Fig. S1A** and low-input genomics in **Fig. S2A**). *, cumulative time from start (0h). **treatment led to a decrease in parasitemia and delay in ring progression and therefore staging is approximate. Aphi, aphidicolin. **D.** Mean growth rate of parasite lines and treatments for low-input genomics experiment: *FCR3* (yellow circles), untreated *Dd2* (black circles), and treated *Dd2* parasites (red circles). Fold change in % parasitemia was calculated by dividing the ending parasitemia from the starting parasitemia for the period being measured. Cumulative timing is as follows: start (0h), post treat (12h), pre-harvest (38h), harvest (+40-43h). N=2. Cumulative parasitemia and ring percentage shown in **Fig. S2B**; raw flow cytometry plots tracking viability and parasitemia shown in **Fig. S2C-G. E.** Recovery time for parasite lines and treatments: *FCR3* (yellow), untreated *Dd2* (black), and DSM1-treated *Dd2* parasites (red). Recovery is defined as the period just after treatment until harvest for down-stream processing. **F.** FACS plots at parasite isolation step for DSM1-treated *Dd2* samples. MitoProbe DiIC1 (5) (stains viable parasites)

439 vs SYBR Green (stains parasite genome) indicates the proportion of erythrocytes (heat map of density) that contain viable ring
440 stage parasites (blue circle: 0.38% ring stage parasitemia). Plots for other parasite lines are depicted in **Fig. S4**.
441

442 **Replication stress followed by a recovery period led to isolation of viable parasites**

443 To explore the impact of antimalarial stress on de novo CNV generation, we treated the parasites with a
444 compound that targets *Plasmodium* dihydroorotate dehydrogenase (DSM1, (80)). The application of
445 DSM1 for an extended period kills parasites (>48hrs at 10x the EC_{50}, (81)). However, we found that when
446 applied to ring-stage parasites for a short period of time (see general timeline in **Fig. 1C**), DSM1 stalls
447 parasite progression through the life cycle in a reproducible and reversible manner (**Fig. S1B, C & E**); this
448 effect was similar to parasite treatment with the replication inhibitor, aphidicolin ((82,83), **Fig. S1D & E**).
449 While aphidicolin stalls replication through inhibition of B-family DNA polymerases (83,84), DSM1 likely
450 impacts replication through depletion of pyrimidine pools as is observed for other forms of dNTP
451 depletion (85,86). For this reason, we now refer to DSM1 treatment as "replication stress".
452 For the low-input genomics harvest, we applied DSM1 to ring-stage parasites for 12h (**Fig. 1C & S2A**).
453 Similar to our pilot experiments (**Fig. S1B, C & E**), we observed that DSM1-treated parasites exhibited a
454 slightly decreased growth rate and altered stage progression compared to untreated parasites (**Figs. 1D**
455 **& S2B-G**). We harvested viable parasites after a recovery period (an additional ~30hrs, **Fig. 1E**, or ~42
456 hours from the treatment start, **Fig. 1C**), where we allowed parasites to complete an additional round of
457 replication and erythrocyte invasion to produce those that have a single, haploid genome (ring stage, **Fig.**
458 **S2E, G**).
459 As a control for cross-sample contamination between isolation wells, we isolated untreated parasites with
460 different genetic backgrounds (*FCR3* vs *Dd2*, **Table 1, Fig. 1D & E**). Prior to isolation of low-cell populations,
461 we confirmed that *FCR3*, untreated *Dd2*, and DSM1-treated *Dd2* parasites were at a similar life cycle stage
462 and viability (**Table 1**); we saved a portion of these samples for parasite population sequencing (i.e. bulk

463 samples). We then proceeded to isolate small populations of viable, ring stage parasites using FACS (**Fig.**

464 **S4**) for whole genome amplification and sequencing.

465

466 **Table 1: Parasite density, staging, and health at isolation for low-input genomics.**

Line/Treatment*	Mean % Parasitemia[$]	Mean % Rings**	Mean % Viability[@]
Untreated *FCR3*^	0.6%	90%	91%
Untreated *Dd2*^	0.8%	87%[&]	94%
DSM1-treated *Dd2*	0.6%	77%[&]	95%

467 *Treatment conditions: 1μM DSM1 (~10x EC_{50}) for 12hrs prior to 31hrs of recovery. Untreated samples were incubated with
468 DMSO as a solvent control for 12hrs and allowed to recover for 28-30hrs (**Fig. 1E**). [$]Parasitemia was determined by calculating
469 the number of infected erythrocytes (SYBR Green+) compared to uninfected erythrocytes (SYBR Green-) (**Fig. S2C-G**). **Mean %
470 Rings value was collected from "post-recovery" plots in **Fig. S2E/G** (R1 SYBR green plots). [@]Viability of parasites was determined
471 by measuring the mitochondrial membrane potential (Mitoprobe+) and calculating % of total SYBR Green-based parasitemia
472 (plots presented in **Fig. S2E, S2G,** and **Fig. S4**). ^*FCR3* (Africa) and *Dd2* (Southeast Asia) parasite lines are from distinct geographic
473 origins (87,88). [&]The initial % rings were 91.8% for both untreated and treated conditions (**Fig. S2C**).
474

475 **Quality assessments showed effective isolation and amplification of low-input samples**

476 We sorted ring-stage low-cell populations from each parasite group into 60 wells of a 96-well plate (*FCR3*,

477 untreated *Dd2,* and treated *Dd2*), including 6 wells with 10-cells and 56 wells with 2-cells. Ten-cell wells

478 served as positive controls for the whole genome amplification step and 2-cell wells provided the optimal

479 balance between sorting accuracy (**Fig. S3**) and de novo CNV detection. Zero cells were sorted into the

480 top 2 rows of the plate ("no-cell" wells). After parasite lysis and applying *Pf*MALBAC version-2 whole

481 genome amplification (**Fig. S5**), we assessed amplification success using three approaches. First, we

482 measured the resulting DNA quantity across 80% of the amplified wells (**Fig. S6**). On average, MALBAC

483 amplification in wells that contained sorted parasites yielded ~120ng of total DNA per reaction, with a

484 ~10% increase in DNA for 10- vs 2-cell samples (mean of 127ng versus 116ng total, respectively). We did

485 not detect position-based bias across plates or appreciable amplification from no-cell wells, but we

486 observed that treated *Dd2* wells had ~3-fold lower levels of amplification than other samples (mean of

487 51ng versus 151ng total, respectively). Of note, there was little difference in mean amplified DNA amounts
488 between the two untreated sample groups (*FCR3* at 146ng and untreated *Dd2* at 152ng).
489 Second, we performed droplet digital PCR (ddPCR) for parasite-specific genes on approximately one-third
490 of wells to confirm the amplification of the parasite genome. DdPCR for *pfmdr1* and *pfhsp70* displayed
491 that wells with measurable DNA contained amplified parasite DNA (**Fig. S6** and **S7**). Additionally, we
492 confirmed that 2-cell wells with very low total DNA amounts (**Fig. S6**) were positive for parasite genomes
493 while "no-cell" wells did not show evidence of parasite material (**Fig. S7A-C**). We observed a significant
494 correlation between total DNA and positive ddPCR droplet counts (black dotted line, p value of 0.0004)
495 but also identified some samples where DNA quantity was low (<50ng) when measured using Qubit but
496 parasite-specific signal was high (red points, all treated *Dd2*, **Fig. S7D**). This observation indicated that
497 DNA quantification is not the best way to measure amplification success; target-based methods like ddPCR
498 are more helpful to estimate the quantity of amplified material.
499 Finally, we employed high-resolution melting analysis to profile a drug resistance marker that differs
500 between *FCR3* and *Dd2* parasites (56). By assessing the *pfdhps* SNP profile of amplified genomes and
501 comparing it to the parental profile in ~10% of samples, we confirmed that there was no evidence of cross-
502 sample contamination during the preparation and amplification steps (**Fig. S8**). Therefore, we proceeded
503 to sequence the amplified bulk and low-input samples (**Table S2**).
504
505 **Coverage deviation and SNP profiles exhibited expected trends in low-input samples**
506 We sequenced 3 bulk samples and 90 low-input samples using Illumina short-read sequencing (**Table 2**
507 and **S3**). Overall, sequencing proceeded well as indicated by coverage and coverage deviation of the bulk
508 samples, as well as an equivalent mean mapping quality across all samples (**Table 2**). Because we noticed
509 that treated *Dd2* wells had lower levels of DNA following amplification (**Fig. S6**), we sequenced higher
510 amounts of material for this condition; this choice impacted mean coverage levels where treated *Dd2*

511 samples had ~4-times higher coverage than untreated *Dd2* samples (**Table 2**). As expected, based on
512 previous studies (42), coverage deviation was ~3-fold higher in low-input samples when compared to bulk
513 samples, reflecting the bias of the whole genome amplification step to over- or under-amplify specific
514 genomic regions.
515 The percentage of high-quality reads that mapped to the *P. falciparum* genome was high across all
516 samples (mean of 67% of all reads). To identify any contaminating reads, we also blasted a subset of high-
517 quality reads (%N <50, length >100bp, and read q score >30) from each sample to a database of genomes
518 from other organisms. While the majority yielded matches to the *P. falciparum* genome (97.5%), a very
519 small number of high-quality reads matched sequences from different organisms (**Fig. S9**). Together, this
520 information shows efficient amplification of the parasite genome and little contribution of environmental
521 contamination during sample preparation. We removed five low-input samples from further analysis
522 based on low coverage levels; on average, excluded samples had ~7-times lower coverage than other low-
523 input samples (**Table S3**). Of the remaining samples, 18 were 10-cell samples and 57 were 2-cell samples.
524 Although mean coverage was ~2-fold higher for 2-cell samples due to the higher coverage of treated low-
525 input samples, the mean normalized deviation was similar between 10- and 2-cell samples (3.5 and 3.1,
526 respectively).
527 To once again check for cross-sample contamination, we tracked SNPs in the low-input samples compared
528 to bulk samples. Despite some variation due to the non-clonal nature of parasite lines (see *Materials and*
529 *Methods*), low-input SNP profiles were similar to their corresponding bulk sample as compared using PCA
530 (**Fig. S10A**). After normalizing total SNPs to mapped reads, we detected a lower rate of SNPs in treated
531 samples compared to untreated counterparts (p value of 0.0001, **Fig. S10B**). While we did not detect a
532 correlation between normalized total SNPs and amplification quality (**Fig. S10C**, R_2: untreated *Dd2*, 0.01;
533 treated *Dd2,* 0.01), we did observe a positive correlation between SNP number and coverage depth in
534 treated versus untreated *Dd2* samples (**Fig. S10D**, R_2: untreated *Dd2*, 0.66; treated *Dd2,* 0.27). This latter

535 finding indicates that the difference in SNP numbers between sample groups is likely due to varying
536 sensitivity at different levels of read coverage (89,90).
537
538 **Table 2: Sequencing Summary for low-input samples and paired bulk samples**

		No. of samples **	Mean no. mapped reads per sample	Mean coverage per sample	Mean Coefficient of Variation (CV#)	Mean mapping quality
Bulk	Untreated *FCR3*	1	9,966,137	54.9	59.5	58.2
Low-input	Untreated *FCR3*	16	2,694,209	13.7	105.1	58.2
Bulk	Untreated *Dd2*	1	5,735,107	33.6	33.3	58.5
Low-input	Untreated *Dd2*	33	2,579,674	13.2	82.6	58.3
Bulk	Treated *Dd2*	1	72,380,215	416.3	35.2	58.5
Low-input	Treated *Dd2**	36	9,936,903	53.7	86.5	58.4

539 *Four times more material was loaded on the flow cell for the treated samples than the untreated samples due to lower initial
540 amplification yields (**Fig. S6**).
541 **Excludes samples that were removed due to low coverage. For low-input samples, analysis includes both 10- and 2-cell samples.
542 #CV is the coefficient of variation of normalized read abundance as in (42)
543
544
545 **Experimental and computational advances improved known CNV calls across low-input samples**
546 For the current study, we employed two different CNV calling methods in low-input samples. HapCNV is
547 a novel read coverage-based CNV calling method specifically designed for low-cell data from haploid
548 genomes (44). In contrast to traditional methods that arbitrarily select reference samples for CNV data
549 normalization, HapCNV constructs a genomic location (or bin)-based pseudo-reference as a comparison
550 baseline. This step systematically alleviates amplification bias for the identification of de novo CNVs. On
551 the other hand, LUMPY is an established CNV calling method that exhibits high sensitivity due to the
552 incorporation of multiple CNV signals (i.e. split and discordant reads) generated from short-read

553 sequencing. It is particularly well-suited for detecting low-frequency variants; however, as for many CNV
554 callers, high sensitivity leads to higher false positives (70,91,92).
555 Using these two distinct CNV calling methods, combined with a recently developed CNV counting
556 approach (SVCROWS (43)), we evaluated the presence of known CNVs in our 2-cell samples (**Fig. 2A & B**).
557 The identification of known CNVs (i.e. those identified in the bulk sample, see *Materials and Methods*)
558 within low-input samples displays the utility of the specific CNV calling methods for different size CNVs in
559 specific genome locations. In our previous study, we identified 2 of the 3 known CNVs in ~10% of single
560 cell genomes (42). In the current study, we identified the *pfmdr1* amplicon in 100% of 2-cell samples using
561 HapCNV (57/57) and 79% of samples using LUMPY (45/57). We did not identify the *pf11-1* amplicon in any
562 2-cell samples using HapCNV (0/57) but detected this locus in 75% of samples using LUMPY (43/57).
563 Finally, we identified *pf332* CNVs in 46% of 2-cell samples using HapCNV (26/57) and 100% of samples
564 using LUMPY (57/57). Although our two studies used different CNV calling methods and are not directly
565 comparable, the overall improvement in the detection of known CNVs in the current study is likely due to
566 advances in both the whole genome amplification method to limit amplification bias (**Table S1**) and
567 recently developed analysis approaches. When we evaluated the detection of three known CNVs
568 specifically in *Dd2* low-input samples, we observed a somewhat higher rate of known CNV detection by
569 either method in treated samples (**Fig. 2B**, HapCNV: mean of 1.2 out of 3 total CNVs for untreated and 1.7
570 for treated samples (increase of 42%), LUMPY: mean of 2.1 out of 3 total CNVs for untreated and 2.9 for
571 treated samples (increase of 38%, **Table S4**).
572
573 **De novo CNVs in low-input samples consisted of rare and common CNVs**
574 We next sought to quantify de novo CNVs in low-input samples. We defined de novo CNVs as those that
575 are not present in the bulk sample and categorized them based on their frequency in low cell samples. As
576 defined in other studies (44,73), “common” CNVs were present in a larger number of genomes (≥10% of

577 the same sample type, i.e. untreated or treated), and "rare" CNVs were those that occurred in a small
578 proportion of samples (<10% of the sample type) (**Fig. 2A**). Overall, LUMPY detected more total de novo
579 CNVs than HapCNV across all samples (~12-fold), and the proportion of rare versus common CNVs varied
580 depending on the method (18% vs 82% for HapCNV, 61% vs 39% for LUMPY, respectively, **Table S4**).
581 Additionally, de novo CNVs were more often identified as duplications than deletions for both CNV calling
582 methods (**Fig. 2C**).
583 To understand the nature of common and rare CNVs, we also assessed how often their locations
584 overlapped across the two sample types (untreated and treated, **Fig. S11**). This analysis is useful for
585 tracking common/rare category utility and relevance. For utility, this step acts as a sanity check since, by
586 definition, we do not expect rare CNV locations to overlap as often as common CNVs. For relevance, de
587 novo CNVs with conserved locations across sample types are less likely to represent true CNVs newly
588 arising in a genome. As expected, we identified common CNVs with conserved locations between
589 untreated and treated samples (22% for HapCNV and 52% for LUMPY, **Fig. S11A**). The lower rate of
590 overlapping calls across samples for HapCNV is likely due to the bin-based normalization strategy to
591 remove amplification artifacts (44). Conversely, we found that rare CNVs were predominantly called in
592 unique genome locations (94% for HapCNV and 97% for LUMPY, **Fig. S11A**), supporting their novel nature.
593 This pattern was consistent when we randomly downsampled all sequencing data to the lowest read
594 coverage prior to CNV calling (1.3 million reads, **Fig. S11B**). This comparison not only highlights the
595 suitability of the common and rare categories but also the difference between the CNV calling methods.
596 Based on these observations, we assessed common and rare CNVs using both methods to capture the
597 broadest view of stress effects on CNV generation.
598
599 **Genome-wide de novo CNVs increased following replication stress**

600 When we compared de novo CNVs in genomes with and without replication stress, we found that results
601 were consistent regardless of the CNV calling method (**Fig. 2**). While the proportion of duplication and
602 deletions somewhat changed with treatment (p value of 0.02 for decreased duplications in treated
603 samples for LUMPY, **Fig. 2C**), we identified large differences in numbers de novo CNVs between treated
604 and untreated 2-cell samples (p value of 0.0002 for HapCNV and <0.0001 for LUMPY, **Fig. 2D**). This pattern
605 was consistent when we downsampled all sequencing data (p value of 0.001 for HapCNV and <0.0001 for
606 LUMPY, **Fig. S12A & B**), indicating that the difference in de novo CNV counts between treatments was not
607 due to read coverage. When we assessed common and rare CNV categories, we once again observed
608 highly significant differences between treated and untreated 2-cell samples in common CNVs using both
609 methods (common: p value 0.0004 for HapCNV and <0.0001 for LUMPY, **Fig. 2E;** rare: p value of 0.008 for
610 HapCNV and <0.0001 for LUMPY, **Fig. 2F**).

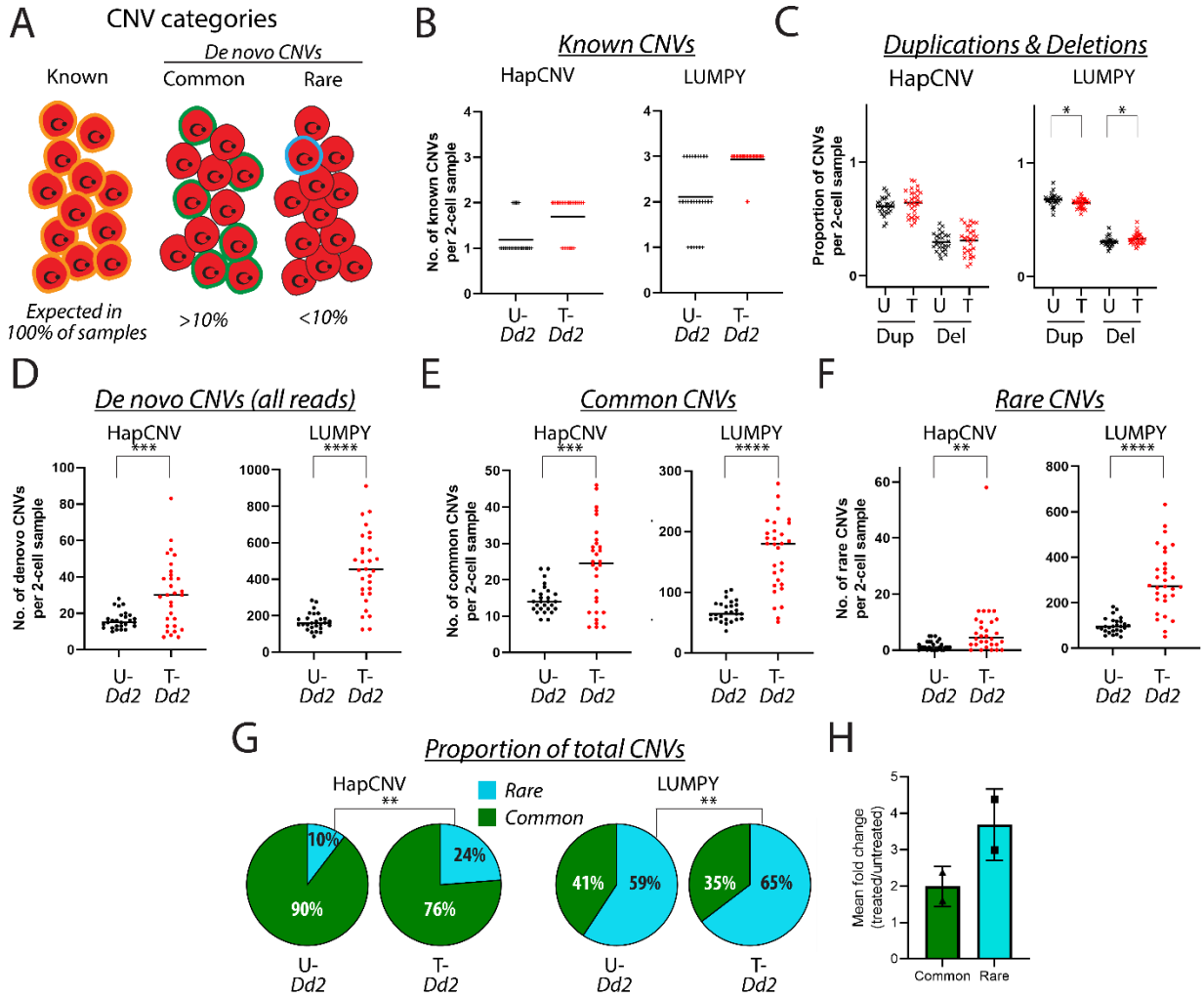


Figure 2. Low cell genomics displays an increase in de novo CNVs following replication stress. Number of CNVs from untreated (U-*Dd2*) and treated (T-*Dd2*) 2-cell samples from two CNV analysis methods: HapCNV and LUMPY. Statistics for all plots use an unpaired T-test with two tailed Welch's correction (****:p value <0.0001; ***: <0.001; **: <0.01; *:<0.05; no stars: not significant). Analyses include all reads. Line at mean value for each dataset. **A**. Depiction of CNV categories used in the analysis. Known CNVs (orange) are detected in bulk samples and present in all low cell samples. De novo CNVs are not present in bulk samples are considered common (green, >10%) or rare (teal, <10%) depending on their frequency across the 2-cell samples. **B**. Detection of three known CNVs in 2-cell samples. Known CNVs were identified in *Dd2* bulk sequence (either *pfmdr1, pf11-1, or pf332*). 0: no known CNVs were detected in 2-cell sample; 1/2/3: one/two/or three known CNVs were detected in 2-cell sample (see **Table S4** for sample counts). **C.** Proportion of total CNVs detected as duplications (Dup) or deletions (Del) in untreated (U) or treated (T) 2-cell *Dd2* samples (p values: 0.02 for Dup and 0.03 for Del from LUMPY). **D**. Detection of de novo CNVs (common and rare combined) from all reads (p values: 0.0002 for HapCNV, <0.0001 for LUMPY). **E**. Detection of common CNVs from all reads (p values: 0.0004 for HapCNV, <0.0001 for LUMPY). **F**. Detection of rare CNVs from all reads (p values: 0.008 for HapCNV, <0.0001 for LUMPY). **G**. Proportion of total CNVs detected as rare and common from all reads; pie charts show the mean but statistics are calculated using all data points from the rare CNV category (p values: 0.003 for HapCNV, 0.009 for LUMPY). Pie chart size does not represent total de novo CNV numbers (~12x higher for LUMPY, **Table S4**). **H**. Mean fold change between rare and common CNVs detected by HapCNV and LUMPY in untreated and treated *Dd2* 2-cell samples (see **Table 3**).

When we compared the proportion of de novo CNVs relative to total CNVs per sample, rare CNVs were significantly increased over common CNVs (p value of 0.003 for HapCNV and 0.009 for LUMPY, **Fig. 2G**).

631 This difference persisted regardless of downsampling (p value of 0.02 for HapCNV and 0.004 for LUMPY,
632 **Fig. S12B & C**) and is in line with our assessment above that rare CNVs sit in unique genome locations and
633 are more likely to be novel in nature (**Fig. S11**). Overall, we detected a ~2-3-fold increase of de novo CNVs
634 in treated samples, regardless of the CNV calling method (**Table 3**). Once again, rare CNVs displayed the
635 largest increase following treatment (~3-4-fold, **Table 3, Fig. 2H**).
636
637 **Table 3. Mean CNV counts per 2-cell sample using two analysis methods.**

CNV detection method$	Condition (2-cell only)	Rare CNVs per sample (SE)*	Common CNVs per sample (SE)*	Combined de novo CNVs per sample (SE)^
HapCNV	treated *Dd2*	7 (1.9)	23 (2.1)	15 (3.3)
	untreated *Dd2*	1.7 (0.4)	15 (0.8)	8 (1.0)
	fold change	**4.4**	**1.6**	**1.9**
LUMPY	treated *Dd2*	297 (25)	163 (11)	230 (35)
	untreated *Dd2*	99 (6.5)	68 (3.1)	84 (9.3)
	fold change	**3.0**	**2.4**	**2.7**

638 *Includes both duplications and deletions. Values are calculated by taking the mean of CNV counts per sample within each
639 category. ^De novo CNV counts combined the subcategories of rare (<10% of samples) and common CNVs (>10% of samples,
640 absent in bulk). $CNV analysis performed using all reads (downsampled data presented in **Fig. S12**).
641

642 **High-confidence de novo CNVs across the genome share consistent breakpoint characteristics**

643 When we compared overlaps between the HapCNV and LUMPY (**Fig. 3A**), we detected a set of CNV regions
644 that was consistent within sample groups (5 for untreated and 38 for treated, **Table S5**). The frequency of
645 these "high-confidence" CNV regions also displayed an increase following replication stress (with knowns
646 excluded, ~12-fold increase in treated samples, **Fig. 3B**). Overall, high-confidence CNV regions
647 represented both duplications and deletions (**Fig. 3B**) and were located on the majority of chromosomes
648 (12 of 14, **Fig. 3C**). Of note, approximately half of these regions were identified as "rare" by both CNV

calling methods across treated samples (17/36, 47%; **Table S5**), indicating that novel CNVs were stimulated in parasite genomes under stress.

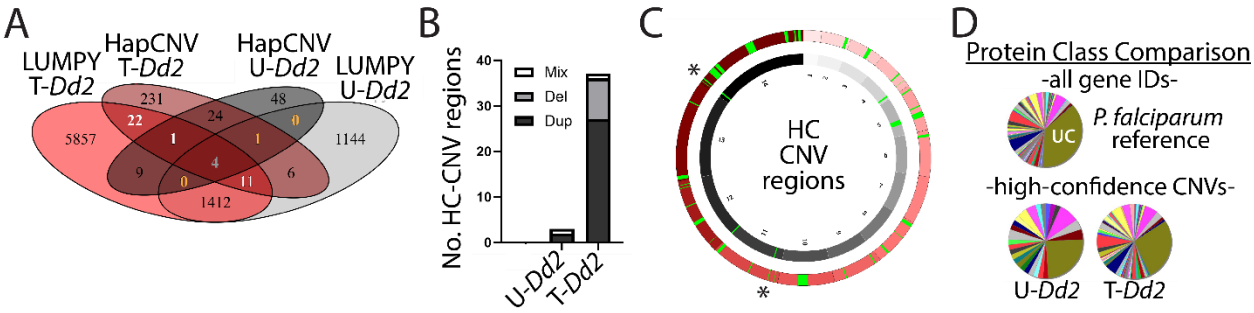


Figure 3: High-confidence CNVs are located across the genome and represent diverse protein classes. A. Comparison of CNV calls showing the number of CNV regions consistent across the two CNV calling methods. High-confidence CNV regions in untreated samples (U-*Dd2*, in yellow text); high-confidence CNV regions in treated samples (T-*Dd2*, in white text). Central number (grey): CNV regions consistent across all samples and calling methods (includes 2 known CNVs and 2 de novo CNVs, **Table S5**). **B**. Summary of number of high-confidence (HC) CNV regions in untreated (U-*Dd2*) and treated (T-*Dd2*) samples including duplication (Dup), deletion (Del), and mixed calls (sub-regions were called as duplications and deletions across a single CNV region by the same CNV calling method, i.e. HapCNV or LUMPY). **C**. Chromosomal location of high-confidence CNV regions identified by both HapCNV and LUMPY methods (green) from untreated (black) and treated (red) parasites. Only core regions of the genome are included in the representation; subtelomeric regions as defined by Otto, et al. were omitted. Each CNV region was increased in length by a factor of 2 to facilitate visualization relative to the rest of the genome. *, de novo high-confidence CNV regions identified in both untreated and treated samples. **D**. Panther classification system v19 protein class comparison. Top chart: protein classes from all annotated *P. falciparum* genes. Bottom charts: protein classes represented by high-confidence CNV regions in untreated (U-*Dd2*) and treated (T-*Dd2*) samples. UC: unclassified proteins (green). Other colors are randomly assigned by the program to represent diverse protein classes.

We also identified the proposed boundaries for the start and end of high-confidence CNV regions (termed "breakpoints") and assessed their properties. Although CNV calling methods vary in their ability to precisely call CNV breakpoints (93), we were interested in whether these sequences shared characteristics with *P. falciparum* those identified using conventional methods (i.e. high AT content and nearby stable DNA secondary structures, (40)).). Overall, we found that features of breakpoints from high-confidence CNVs (12 for untreated and 66 for treated conditions, **Table S6**) were consistent with what was previously reported. Most breakpoints (>75%) had nearby stable secondary structures predicted to form hairpins with an average free energy of folding of -7 to -8 kcal/mol (depending on the treatment, **Table S6**). Most of these regions were situated in intergenic areas of the genome (56% for untreated and 75% for treated). The breakpoint regions also exhibited a higher AT-content compared to the genomic average (85-86%,

677 **Table S6**, versus 80.6% genome-wide (41)). Based on breakpoint positions, we estimated the size of the
678 high-confidence CNV regions; in general, this list represented relatively large regions (>4kb) that had the
679 potential to cover multiple genes (maximum of 115kb, average of 23kb, **Table S6**).
680
681 **High-confidence de novo CNVs represented diverse cellular pathways with potential clinical benefits**
682 When we searched for genes that are covered by these regions, we identified 26 genes (across 3 de novo
683 CNV regions) and 198 genes (across 37 de novo CNV regions) in untreated and treated *Dd2* samples,
684 respectively (**Table S5**). Genes encompassed by the CNV regions represented diverse protein classes (**Fig.
685 3D**) and no gene ontology (GO) categories were significantly enriched in this list of genes (using an FDR
686 adjusted p value of 0.05, **Table S7**). Along with the distribution across the genome (**Fig. 3C**), the lack of GO
687 term enrichment emphasizes the random nature of the de novo CNVs and the absence of selection during
688 sample preparation.
689 High-confidence CNV regions encompassed genes that were previously reported as important for clinical
690 malaria infections (**Table 4**). Several genes play essential roles in transmission of the parasite from the
691 human blood to the mosquito and back. We identified CNV regions that included genes for two TRAP-
692 related proteins, CTRP and S6, which are important for infection of the mosquito midgut and salivary
693 glands, respectively (94,95). Additionally, CNV regions encompassed two PHIST genes (Pfg14-744 and 748)
694 that are important for early gametocyte development (96,97). We detected CNV regions covering 3 genes
695 important for liver stage development including two genetically-attenuated vaccine candidates, UIS3 and
696 SLARP (98-101); the latter of which achieved equivalent protection levels to whole sporozoite vaccines
697 (PfSPZ-GA1, (102,103)). Two CNV regions carried genes important for replication of the erythrocytic stage
698 (SEA1 and eif4A) and another two for virulence and antigen expression (PTP6 and SURFIN 14.1). One CNV
699 region encompassed an aminophospholipid transporter previously identified as important for artemisinin
700 resistance in a GWAS study (104). Finally, some of the genes from high confidence CNV regions were

701 shown to be essential in the erythrocytic stage transposon mutagenesis screen (Pfg14-744, Pfg14-748,

702 eif4A, and PTP6), indicating their multiple important roles across the parasite developmental cycle.

703

704 **Table 4. High-confidence CNV regions with clinical importance.**

Gene ID	DUP/DEL	Chrom.	Product description	Clinical importance	Citation
PF3D7_0315200	DUP	3	circumsporozoite- and TRAP-related protein (CTRP)	transmission	(94)
PF3D7_0315600	DUP	3	male development protein (MD3)	transmission	(105)
PF3D7_1403800	DUP	14	nuclear formin-like protein MISFIT	transmission	(106)
PF3D7_1442600	DUP	14	TRAP-like protein (S6)	transmission	(95)
PF3D7_1477700*	DEL	14	Plasmodium exported protein, PHISTa (Pfg14-748)	transmission**	(96,97)
PF3D7_1477300*	DEL	14	Plasmodium exported protein, PHIST (Pfg14-744)	transmission**	(96,97)
PF3D7_1474900	DEL	14	trailer hitch homolog (CITH)	transmission^	(107)
PF3D7_1223700	DEL	12	vacuolar iron transporter (VIT)	liver stage ^	(108)
PF3D7_1147000	DEL	11	sporozoite and liver stage asparagine-rich protein (SLARP)	liver stage	(99,100)
PF3D7_1302200	DEL	13	upregulated in infectious sporozoites 3 (UIS3)	liver stage	(98,101)
PF3D7_1021800	DUP	10	schizont egress antigen-1 (SEA1)	erythrocytic stage replication***	(109)
PF3D7_1468700*	DEL	14	eukaryotic initiation factor 4A (eif4A, PfH45)	erythrocytic stage replication	(110,111)
PF3D7_1302000*	DEL	13	EMP1-trafficking protein (PTP6)	virulence	(112)
PF3D7_1477600	DEL	14	surface-associated interspersed protein 14.1 (SURFIN 14.1)	antigen#	(113)
PF3D7_1468600	DEL	14	aminophospholipid transporter/flippase	drug resistance^^	(104)

705 *Essential in erythrocytic stage parasites based on MIS score <0.3 ((76), as reported in PlasmdDB (77)).
706 **While essential in erythrocytic parasites, these genes are expressed during early gametocytogenesis and categorized as
707 impactful for transmission to mosquitos.

***Not essential in erythrocytic stage but categorized as impactful for replication due to role in schizont egress. Antibodies to SEA1 have been detected in malaria-immune blood but the implication of this observation remains unknown (114).
^Studies performed with *P. berghei* orthologues of *P. falciparum* genes.
#The SURFIN gene family is located in subtelomeric regions of the genome and expressed on the surface of erythrocytes (115); combined with indications of selection in an in-silico study, this gene is a likely antigen target for protective immunity.
^^ Although this transporter class is likely important for parasite biology (reviewed in (116)), the role of this locus in artemisinin resistance has not been validated beyond the initial GWAS study.

As a final analysis, we evaluated whether genes from high-confidence CNV regions specifically from our treated samples have been detected in parasites from clinical infections. We performed this analysis by searching the largest catalogue of clinically relevant CNVs to date; this list of high-frequency CNVs was previously called using genomes from 2855 parasite isolates from 21 malaria-endemic countries and represented genes from larger (>300bp), high-quality, core genome variants (38). Assuming ~5000 genes in the core *P. falciparum* genome (58), we found that genes from the two lists were ~2-times more likely to overlap than by chance (chi-square odds ratio of 2.5, Fisher's exact p value <0.0001 for HapCNV-treated CNV list and 1.6, 0.03 for LUMPY-treated CNV list). Therefore, stress-induced de novo CNVs have the potential to be beneficial in the clinical environment.

DISCUSSION

Low-input genomics for studying de novo CNVs

Non-parental or "de novo" CNVs are not detected when analyzing a population of parasites predominantly because their signal (e.g. extra reads that align to that region or reads that span breakpoints) is negated by the overwhelming signal from normal copy number at that genome location. For this reason, assessments of fewer cells are necessary to investigate de novo CNV generation. While de novo CNVs have been previously tracked using flow cytometry of single yeast genomes (23), this sensitive approach provides a limited view of CNV dynamics by focusing on few specific loci that express fluorescent reporters. Inferring gene copy number through single cell transcriptomics can identify Mb-sized structural

737 changes across genomes from heterogeneous tumor samples (117-120) but this approach is not
738 applicable for smaller sized de novo CNVs. Single-cell genomics, where individual genomes are isolated
739 and amplified to a level that can be sequenced, has been used to directly quantify de novo CNV rates in
740 brain tissue, cancer cells, and *Leishmania* parasites (121-128). Early single-cell genomics studies of *P.*
741 *falciparum* have been promising but so far have had limited success with CNV identification (42,129,130).
742 Here, we optimized a low-input analysis pipeline and successfully isolated, amplified, and sequenced *P.*
743 *falciparum* samples for CNV analysis. With experimental and computational improvements, we were able
744 to increase our rate of parental, or "known", CNV calling over our prior study (42). Importantly, we
745 detected de novo CNVs across the parasite genome and replication stress significantly increased their rate
746 of formation. By analyzing ~45 low-input samples per condition, we are limited to observing a subset of
747 the population. However, this study is one of few that have directly assessed de novo CNVs in microbes
748 and our findings demonstrate that replication stress readily drives the rapid generation of de novo CNVs
749 in an important pathogen. Below, we cover the rationale behind our experimental/analysis choices and
750 integrate our findings into an overarching model of *P. falciparum* adaptation.
751
752 *Application of stress without evidence of selection to explore CNV dynamics*
753 To evaluate the impact of replication stress on de novo CNV generation, we applied treatment to parasites
754 just prior to replication. We observed replication stall and then resume post-treatment, which provided
755 evidence that we successfully applied sub-lethal stress (**Figs. 1D**, **S1**, and **S2**). Following this step, we
756 allowed the parasites to complete replication and reinvade new erythrocytes; we reasoned that this
757 "recovery phase" enabled the repair of the resulting DNA damage, which is likely to be replication-
758 dependent (reviewed in (1,131). Additionally, reinvasion facilitated the isolation of haploid parasite
759 genomes (1n, **Fig. 1A**), encouraging the detection of de novo CNVs due to limited contrasting signal (26).

760 Because of the reinvasion step, which involves an expansion in parasite number (~3-fold, **Fig. 1D**), there
761 was potential to select for beneficial DNA changes across the population of parasites. However, we did
762 not detect evidence of strong selection from SNP profiles (**Fig. S10A**) or high-confidence CNV regions (**Fig.**
763 **3**, **Table S5**). Specifically, we did not observe an enrichment of CNVs that encompassed the *pfdhodh* gene,
764 which contributes directly to DSM1 resistance (132), or genes from DNA-related pathways (**Table S7**).
765 While studies of numerous microbes observe CNVs arise under strong selective conditions like nutrient
766 starvation or drug treatment (as for *Saccharomyces* yeast: (5,133,134), *Salmonella* bacteria: reviewed in
767 (135), as well as *Plasmodium*: (132,136-142)), we now show that mild, transient conditions can stimulate
768 CNVs that have the potential to increase parasite survival.
769
770 *Relative comparison using multiple CNV calling methods to appreciate the impact of stress*
771 De novo CNV estimates using single cell methods from yeast, neurons, and human cancer cells vary greatly
772 and are difficult to standardize due to the use of different experimental techniques and CNV calling
773 methods (23,24,143). For these reasons, we are not attempting to compare the rate of *P. falciparum* CNV
774 formation from this study to those from other organisms. Additionally, this lack of standardization in the
775 field led us to use two distinct CNV calling methods in our analysis. Due to the strengths and weaknesses
776 of HapCNV and LUMPY, we observed differences in both known (**Fig. 2B**) and de novo CNV (**Table S4**)
777 calling using the two methods. LUMPY identifies reads that cover breakpoint regions to sensitively detect
778 CNVs (70); because we are counting regions with few reads as support, both sensitivity and the number
779 of false positives are high in this analysis. High known calling rates along with high numbers of de novo
780 CNVs in our studies exemplified this feature of LUMPY. On the other hand, HapCNV uses a genome-specific
781 pseudo-reference for normalization, which removes repeated patterns of over-and under-amplification
782 ((44) and *Materials and Methods*); because we require read coverage to span 3 consecutive 1kb bins,
783 small CNVs in lower coverage genomic neighborhoods are excluded in HapCNV analysis. This limited the

784 detection of smaller known amplicons (*pf11-1* and *pf332*), led to fewer de novo CNV calls using this
785 method, and likely contributed to the larger size of high-confidence CNVs (**Table S6**). Given the high
786 abundance of small CNVs (<300bp) in the parasite genome (38,144), HapCNV is likely underestimating the
787 impact of small CNVs in our study.

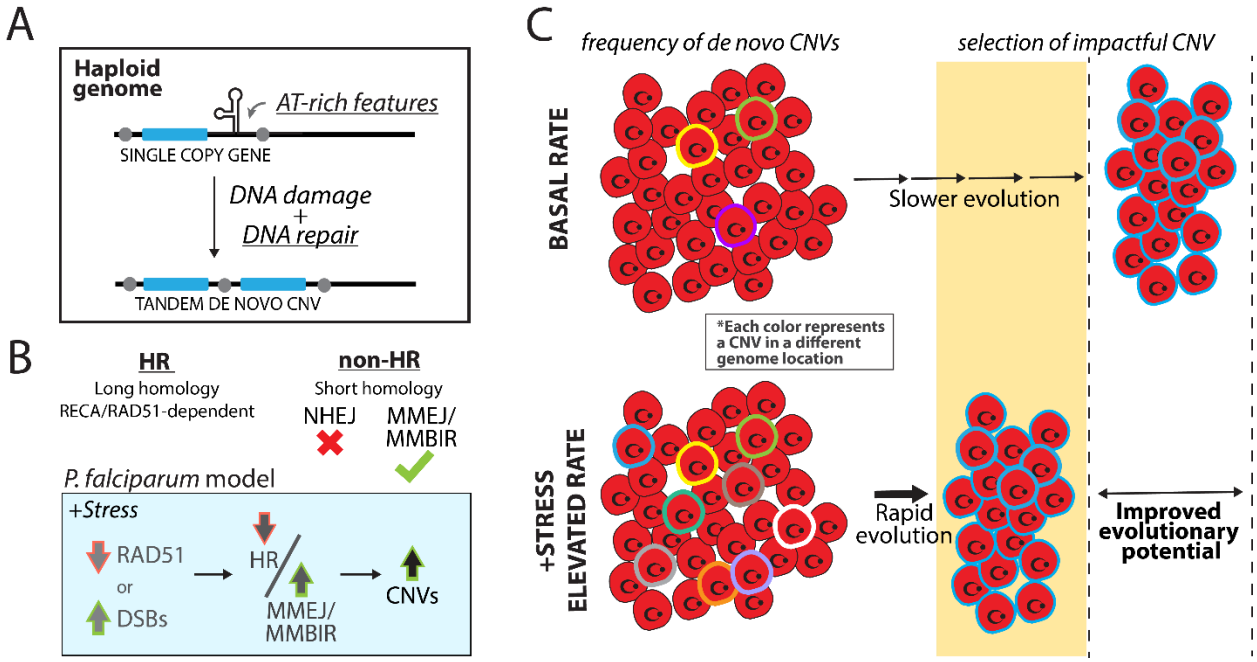
788 Ultimately, the value of our study is in the relative comparison of treated and untreated samples. During
789 sample preparation, we acknowledge that the apparent 3-fold difference in amplification efficiencies
790 between untreated and treated sample groups (**Fig. S6**) could compromise this comparison. In order to
791 prevent cross-contamination, 2-cell samples were sorted into different plates and consequently, they
792 were amplified separately. Although separated by just a few days (see *Materials and Methods*), alterations
793 in reagent performance from batch to batch could contribute to differences in amplification efficiency.
794 Additionally, we speculate that there may have been a general under-quantification of the treated *Dd2*
795 plate; we increased the amount of each sample sequenced by 4-fold to compensate for low DNA amounts
796 and ended up achieving ~4-fold higher coverage levels for treated samples (**Table 2**). This under-
797 quantification is supported by our ddPCR assays, where samples with small amounts of quantified DNA
798 still yielded abundant PCR positive droplets (**Fig. S7D**), indicating adequate *P. falciparum* genome
799 amplification for this round. Because sequencing quality was similar between sample groups (**Table 2**, CV
800 and mapping quality), downsampled data showed a consistent result with analyses using all data (**Figs. 2**,
801 **S11**, and **S12**), and treatment had a consistent impact on CNV categories using two distinct CNV calling
802 programs (**Table 3** & **Fig. S12B**), we are confident that coverage differences are not dramatically impacting
803 our overall conclusions.

804 During our analysis, we assumed that false positive CNVs occur at a similar rate across both treated and
805 untreated groups (within each CNV caller and coverage level), which allowed us to confidently assess the
806 impact of stress despite some experimental variation. Variations in known and de novo CNV calling
807 described above emphasize that no CNV calling method is perfect and combining them can improve

808 confidence in results (42,145-147). Therefore, we investigated de novo CNV patterns using the two
809 individual methods (**Figs. 2**, **S11**, & **S12**) as well as those that overlapped between HapCNV and LUMPY
810 (**Fig. 3**, **Table S5**). Importantly, these "high-confidence" CNVs reflected increases after stress previously
811 detected using the individual tools, albeit at a greater level (~3- versus 12-fold increase). Additionally, we
812 speculate that newly arising CNVs would have distinct locations across samples; thus, the rare nature and
813 unique locations of high-confidence CNVs emphasized their potential to be novel (**Table S5**).
814
815 *De novo CNV categories highlighting existing and novel genome variation*
816 During our investigations, we identified two types of de novo CNVs; those detected in one or a few
817 genomes (rare, <10%) and those detected in more than a few (common, >10%). While others have used
818 these CNV categories (73), there is no precedence for them in the context of *Plasmodium* biology (i.e. a
819 haploid parasite with asynchronous replication and schizogeny (148)). However, we propose that tracking
820 these categories helps us to understand the biological relevance of de novo CNVs in our analysis.
821 Based on their frequency, common CNVs are either artifacts of low-input procedures/CNV analysis or
822 represent minor variants that preexisted in the population or arose early in the replication cycle. For the
823 former, bias during the whole genome amplification step (i.e. the repeated pattern of over/under-
824 amplification that occurs in a reproducible pattern across the parasite genome (42)) and PCR during library
825 construction have the potential to skew gene copy number and increase the false positive rate (149,150).
826 However, we chose experimental and computational methods designed to limit the contribution of
827 amplification bias. First, MALBAC amplification itself limits the over-amplification of certain genomic
828 regions by avoiding exponential amplification at the earliest steps (149) and we used limited PCR cycles
829 during library preparation (3 cycles, (42)). These efforts are most clearly shown through the reduction in
830 CV following MALBAC optimization in both of our studies (by ~39% after modifying the amplification
831 primer (42) and by 43% after switching to the *Bsu* polymerase, **Table S1**). Second, LUMPY is not dependent

832 on read coverage and HapCNV specifically addresses amplification artifacts by removing repeated signal
833 present in all samples (44,70). Overall, we detected few CNVs with conserved genomic locations across
834 low-input samples, which provides evidence that our methods limit the effect of amplification bias on the
835 results; we only identified two high-confidence CNV regions that had conserved locations across multiple
836 *Dd2* and *FCR3* 2-cell samples (**Fig. 3C**, **Table S5**). In the future, amplification-free methods (151), or
837 visualization of single long-reads (49,152), may offer advantages in distinguishing amplification bias from
838 minor variants and de novo CNVs.
839 Rare CNVs, on the other hand, represent either random noise or true signal from novel CNVs arising in the
840 genome. We assert that most noise is removed through normalization procedures, especially with
841 HapCNV, and the impact of remaining false positives are minimized by the relative comparison of our
842 studies (see above). We identified the majority of rare CNVs in unique genome locations across sample
843 types, providing evidence that they are not a result of amplification bias where the same CNVs are
844 repeatedly detected in each sample. Additionally, the greater impact of stress on rare CNVs than common
845 CNVs (**Table 3** and **Figs. 2F, 2G**) supports their replication dependence. The random nature of de novo
846 CNVs, as well as the capacity to encompass any gene across the genome, ensures that CNVs can alter all
847 aspects of parasite biology in response to the host environment. Further, our finding that stress-induced
848 de novo CNVs tended to exhibit altered copy number in clinical isolates combined with the high frequency
849 of unique CNVs in previous genome-wide CNV studies (32,38), directly illustrate the expansive
850 evolutionary potential of this pathogen.
851
852 *Adaptations that encourage de novo CNV formation*
853 The current model of CNV formation in asexual erythrocytic *P. falciparum* parasites is that AT-rich
854 sequences form hairpins, disrupt replication, and eventually lead to double-strand breaks that are
855 repaired by error-prone pathways (40). The evolution of CNVs in this organism is especially interesting

856 because of its unique genome architecture and alternative repertoire of CNV-generating repair pathways
857 (41,153). Although they arise at many locations across the genome ((30,31,38), **Fig. 3C**), *P. falciparum*
858 CNVs that contribute to adaptation are commonly gene duplications with a relatively simple structure.
859 Many impactful duplications form in tandem head-to-tail orientation ((40,132,142,154), **Fig. 4A**), which is
860 likely due to a limited repertoire of DNA repair pathways; *P. falciparum* lacks the canonical
861 nonhomologous end-joining (NHEJ) pathway that is a major contributor to CNV formation in other
862 organisms (1,153). Instead, parasites use pathways that employ varying lengths of sequencing homology
863 (i.e. homologous recombination, or HR, and microhomology-mediated repair, **Fig. 4B**). This repair
864 repertoire, along with an especially high AT-content genome that facilitates CNV formation (40,132), and
865 a lack of cell cycle checkpoints that control replication forks during times of stress (reviewed in (155)),
866 likely represent adaptations that assist haploid *P. falciparum* parasites in accumulating CNVs across their
867 genome (**Fig. 4A**). One major question that remains is whether these same processes are active in other
868 rapidly replicating parasite stages such as oocysts in the mosquito midgut or schizonts in the human liver.



869

870 **Figure 4: Potential connection between replication stress, DNA repair, and CNV generation in the malaria genome**. **A**.
871 Adaptations that encourage CNV formation in the *P. falciparum* genome (underlined, (40)). **B**. Proposed model of how replication
872 stress can impact DNA repair pathways based on prior studies (HR, homologous recombination; NHEJ, non-homologous end-

joining; MMEJ, microhomology-mediated end joining; MMBIR, microhomology-mediated break-induced repair; DSB, double-strand breaks). Model is suggested based on studies showing 1) stress can lead to reduction of RAD51 and increased DSBs reduced HR activity (156,157), 2) stress can trigger pathways that use microhomology sequences for DNA repair (158,159), and 3) *P. falciparum* uses microhomologous sequences to generate CNVs (40). **C.** Potential benefits of a diverse parasite population for evolutionary potential. Stress elevates the frequency of de novo CNVs across the population, which leads to more rapid evolution of beneficial CNVs (blue cells).

Updating the model of P. falciparum genome adaptation

In conjunction with previous studies in other organisms, our findings in *P. falciparum* support the connection between replication stress, DNA repair, and CNV generation (**Fig. 4B**). Studies from bacteria to cancer cells have shown that stress can either alter levels of proteins essential for HR-based repair or increase the frequency of DNA breaks (159). In cancer cells, hypoxic stress causes a decrease in HR activity (156) and this may lead to an increased reliance on alternative error-prone pathways to repair DNA damage (160). In bacteria (158,161), starvation drives initial amplification steps at microhomologous sequences. The nature, degree, and timing of applied stress likely matter because some studies also show the dose-dependent induction of HR proteins under some conditions (162). Studying these processes in *P. falciparum* under various conditions will be particularly important due to the lack of NHEJ repair (see *Adaptations that encourage de novo CNV formation*); when NHEJ is deficient in cancer cells, DNA repair becomes more error-prone and contributes to CNV formation (163,164).

Prior to the current study, the predominant evidence connecting DNA repair and CNV generation in *P. falciparum* was the detection of microhomology-mediated pathway signatures in CNV breakpoints (40). Microhomology-mediated pathways require less homology and therefore, are more likely to interact with diverse sequences up- and downstream of a DNA break to generate various length CNVs. Our current observation of stress-induced de novo CNV formation (**Table 3** and **Fig. 2H**) further supports this model. We originally set out to understand how sub-lethal antimalarial treatment impacts CNV generation because this is a condition that the parasite encounters during clinical infection (165). However, because this compound targets pyrimidine biosynthesis and its impacts mirror aphidicolin treatment (**Fig. S1**), we speculate that the effects we observed are likely due to replication inhibition. Our result is consistent with

901 studies on diverse organisms, like humans, mice, and flies, where replication stress leads to CNV formation
902 (20-22,166,167). Interestingly, the degree of de novo CNV stimulation is also consistent across these
903 studies; we detected a ~3-fold increase in *P. falciparum* de novo CNVs (**Table 3**) compared to a ~3-5-fold
904 increase in mammalian de novo CNVs (20-22). Compared to this work, our study highlights that despite
905 the smaller genome and divergent biology of *Plasmodium*, as well as the use of different experimental
906 approaches, there are likely conserved processes that increase genetic diversity under stress. Future
907 studies will address gaps in knowledge including assessment of conditions that stimulate the *P. falciparum*
908 adaptive amplification response, the timing of this response, and the requirement and regulation of
909 specific mechanistic players.
910
911 *Support for adaptive amplification in microbial populations*
912 The experimental evolution approach to evaluate the impact of selective conditions on microbial
913 populations across relatively long time scales has provided many insights that include the important role
914 of gene amplification in genomic diversity (e.g. (5,134)). In bacteria, the term "adaptive amplification" was
915 established to describe how environmental conditions can induce genetic change to help cells adapt to
916 stress (161). De novo CNVs have been directly investigated in bacteria and *Leishmania* parasites using
917 single read assessments and single-cell isolation, respectively (128,152); to our knowledge, prior studies
918 have not assessed the effect of non-selective short-term conditions on de novo CNV rates in microbes. By
919 continuing to improve genome amplification and CNV detection methods, we anticipate that future
920 studies of genome dynamics will lead to new insight on how stress stimulates microbial evolution.
921 Already, studies from a variety or organisms are showing that an increased rate of CNV formation can
922 have an impact on a population of organisms; if beneficial under specific conditions, a CNV that arises in
923 a single genome can expand during selection into a larger population of cells with novel characteristics
924 (135,168). This rapid expansion is exemplified when minor bacterial populations with higher gene copy

925 numbers confer "heteroresistance" during clinical antibiotic selection (152,169-171). In another example,
926 higher levels of intra-tumor heterogeneity in gene copy number predict a poorer cancer prognosis
927 (127,172,173).
928 For *Plasmodium*, even with a change in the copy number of a single region per parasite, the genomic
929 diversity within a single infected human is expansive due to the sheer numbers of parasites (estimated to
930 reach 10^8 parasites when symptomatic and $>10^{11}$ in severe *P. falciparum* infection (174)). This diversity
931 can provide an obvious advantage as a heterogeneous population prepares asexual parasites to respond
932 to diverse stressors (**Fig. 4C**). While our current studies point to a role of adaptive amplification in *P.*
933 *falciparum*, we do note a low level of de novo CNVs across the parasite genome under normal conditions
934 (**Fig. 2D** and **S12A**). Since some antimalarials act rapidly (81), some level of de novo CNVs already present
935 in a few parasites across the population would increase the chances of survival during drug treatment. It
936 is also important to understand how parasites respond to stressful environments during infection,
937 including changes in nutrient composition in different hosts, drug treatment during symptomatic
938 infection, or attack from the human immune system. While the current study focused on one antimalarial
939 compound, it will be important for future studies to evaluate the impact of other sources of stress on *P.*
940 *falciparum* CNV formation. For example, hypoxia stimulates CNV formation in cancer cells (156) and a
941 proteotoxic drug stimulates genetic change in yeast (175). As a eukaryotic microbe, *P. falciparum* occupies
942 special niche that can be exploited as a model to study eukaryotic mechanisms of stress-induced genome
943 dynamics.
944
945 *Clinical implications & future questions*
946 *P. falciparum* causes the majority of malaria deaths worldwide and readily acquires antimalarial resistance
947 (176,177). Resistance-conferring CNVs that encompass multiple genes have been identified in clinical
948 infections (30,38,178-182). While a gene from one of our high-confidence CNVs may contribute to

949 artemisinin resistance, we also identified genes that participate in many processes important for malaria
950 transmission and infection (**Table 4**). This observation combined with the diverse genome location and
951 protein classes represented (**Fig. 3**) illustrates the enormous adaptive potential of the *P. falciparum*
952 genome.

953 Despite their direct contributions to various phenotypes, CNVs may also facilitate the acquisition of point
954 mutations in haploid *P. falciparum*; strong evidence for the close relationship comes from the observation
955 of point mutations *within* amplifications selected in vitro (137,139,142,183,184). Once de novo CNVs form
956 during replication of the asexual erythrocytic stage (**Fig. 4**), meiotic recombination during the sexual phase
957 in the mosquito can streamline beneficial CNVs to balance fitness costs (68). Given the importance of
958 CNVs in *P. falciparum* adaptation, it is not surprising that this organism has evolved strategies to
959 encourage CNV formation (as described in *Adaptations that encourage de novo CNV formation*).
960 Additionally, parasites from specific regions of the world may have an increased propensity to develop
961 drug resistance (185). Evaluating whether the CNV rate correlates with the parasite background will help
962 to define the evolutionary potential of this successful pathogen.

963 Antimalarial therapies and vaccines targeting the *Plasmodium* parasite are in danger due to drug
964 resistance and access challenges (186,187). A strategy to impede genome evolution may be required to
965 control malaria infections. "Evolution-proof" therapies have been explored in response to antibiotic and
966 anticancer resistance (reviewed in (188,189)). Of particular note is a recent study targeting NHEJ repair to
967 prevent resistance in melanoma cell lines (190). While antimalarial combination therapies were originally
968 suggested to limit recrudescence following artemisinin-based drug treatment (191), the parasite has
969 gained CNV-based strategies to overcome partner drugs ((181,182) and reviewed in (192)). Therefore,
970 identifying and targeting unique aspects of *Plasmodium* biology that facilitate CNV formation (e.g. DNA
971 repair, AT-dependent mechanisms, or replication control) may limit parasite evolution and increase
972 treatment efficacy.

973 **Data Availability**

974 The source code for the BLAST pipeline and SVCROWS can be accessed at
975 https://doi.org/10.6084/m9.figshare.29885879 and https://doi.org/10.6084/m9.figshare.29099801.v1.
976 Short read data is available at NCBI Sequence Reads Archives under project PRJNA1201106.

977

978 **Supplementary Data are available at *NAR* Online**

979

980 **Acknowledgments**

981 We would like to thank Dr. John Campbell for the use of the Mosquito LV instrument and assay reagents
982 for the colorimetric assessment of the SH800. We would also like to thank Dr. Ali Guler for statistical
983 consultation and Dr. Heidi Seears from the University of Virginia Department of Biology Genomics Core
984 Facility for sequencing assistance. We also thank the Vector and Eukaryotic Pathogen Genomics Database
985 Resource (VEuPathDB) and specifically PlasmoDB (https://plasmodb.org) for hosting data used in some
986 aspects of this study.

987

988 **Funding**

989 This work was supported by the National Institutes of Health (1R01AI150856 to J.L.G.); and the National
990 Science Foundation (NRT-ROL 2021791 to N.J.B.).

991

992 **References**

993 1. Hastings, P.J., Lupski, J.R., Rosenberg, S.M. and Ira, G. (2009) Mechanisms of change in gene
994 copy number. *Nat Rev Genet*, **10**, 551-564.
995 2. Lauer, S. and Gresham, D. (2019) An evolving view of copy number variants. *Curr Genet*.
996 3. Pos, O., Radvanszky, J., Buglyo, G., Pos, Z., Rusnakova, D., Nagy, B. and Szemes, T. (2021) DNA
997 copy number variation: Main characteristics, evolutionary significance, and pathological aspects.
998 *Biomed J*, **44**, 548-559.
999 4. Roth, J.R. (2011) The joys and terrors of fast adaptation: new findings elucidate antibiotic
1000 resistance and natural selection. *Mol Microbiol*, **79**, 279-282.

1001 5. Payen, C., Di Rienzi, S.C., Ong, G.T., Pogachar, J.L., Sanchez, J.C., Sunshine, A.B., Raghuraman,
1002 M.K., Brewer, B.J. and Dunham, M.J. (2014) The dynamics of diverse segmental amplifications in
1003 populations of Saccharomyces cerevisiae adapting to strong selection. *G3 (Bethesda)*, **4**, 399-
1004 409.
1005 6. Elde, N.C., Child, S.J., Eickbush, M.T., Kitzman, J.O., Rogers, K.S., Shendure, J., Geballe, A.P. and
1006 Malik, H.S. (2012) Poxviruses deploy genomic accordions to adapt rapidly against host antiviral
1007 defenses. *Cell*, **150**, 831-841.
1008 7. Leary, R.J., Lin, J.C., Cummins, J., Boca, S., Wood, L.D., Parsons, D.W., Jones, S., Sjoblom, T., Park,
1009 B.H., Parsons, R. *et al.* (2008) Integrated analysis of homozygous deletions, focal amplifications,
1010 and sequence alterations in breast and colorectal cancers. *Proc Natl Acad Sci U S A*, **105**, 16224-
1011 16229.
1012 8. Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J.,
1013 Boehm, J.S., Dobson, J., Urashima, M. *et al.* (2010) The landscape of somatic copy-number
1014 alteration across human cancers. *Nature*, **463**, 899-905.
1015 9. Zhang, L., Yuan, Y., Lu, K.H. and Zhang, L. (2016) Identification of recurrent focal copy number
1016 variations and their putative targeted driver genes in ovarian cancer. *BMC Bioinformatics*, **17**,
1017 222.
1018 10. Peng, H., Lu, L., Zhou, Z., Liu, J., Zhang, D., Nan, K., Zhao, X., Li, F., Tian, L., Dong, H. *et al.* (2019)
1019 CNV Detection from Circulating Tumor DNA in Late Stage Non-Small Cell Lung Cancer Patients.
1020 *Genes (Basel)*, **10**, 926.
1021 11. Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A., Sanchez, R., Catano, G., Nibbs, R.J.,
1022 Freedman, B.I., Quinones, M.P., Bamshad, M.J. *et al.* (2005) The influence of CCL3L1 gene-
1023 containing segmental duplications on HIV-1/AIDS susceptibility. *Science*, **307**, 1434-1440.
1024 12. Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S.,
1025 Krasnitz, A., Kendall, J. *et al.* (2007) Strong association of de novo copy number mutations with
1026 autism. *Science*, **316**, 445-449.
1027 13. Glessner, J.T., Bradfield, J.P., Wang, K., Takahashi, N., Zhang, H., Sleiman, P.M., Mentch, F.D.,
1028 Kim, C.E., Hou, C., Thomas, K.A. *et al.* (2010) A genome-wide study reveals copy number variants
1029 exclusive to childhood obesity cases. *Am J Hum Genet*, **87**, 661-666.
1030 14. Olsson, L.M. and Holmdahl, R. (2012) Copy number variation in autoimmunity--importance
1031 hidden in complexity? *Eur J Immunol*, **42**, 1969-1976.
1032 15. Butchbach, M.E. (2016) Copy Number Variations in the Survival Motor Neuron Genes:
1033 Implications for Spinal Muscular Atrophy and Other Neurodegenerative Diseases. *Front Mol*
1034 *Biosci*, **3**, 7.
1035 16. Sekar, A., Bialas, A.R., de Rivera, H., Davis, A., Hammond, T.R., Kamitaki, N., Tooley, K.,
1036 Presumey, J., Baum, M., Van Doren, V. *et al.* (2016) Schizophrenia risk from complex variation of
1037 complement component 4. *Nature*, **530**, 177-183.
1038 17. Zekavat, S.M., Ruotsalainen, S., Handsaker, R.E., Alver, M., Bloom, J., Poterba, T., Seed, C., Ernst,
1039 J., Chaffin, M., Engreitz, J. *et al.* (2018) Deep coverage whole genome sequences and plasma
1040 lipoprotein(a) in individuals of European and African ancestries. *Nat Commun*, **9**, 2606.
1041 18. Farashi, S. and Harteveld, C.L. (2018) Molecular basis of alpha-thalassemia. *Blood Cells Mol Dis*,
1042 **70**, 43-53.
1043 19. Wisniowiecka-Kowalnik, B. and Nowakowska, B.A. (2019) Genetics and epigenetics of autism
1044 spectrum disorder-current evidence in the field. *J Appl Genet*, **60**, 37-47.
1045 20. Arlt, M.F., Mulle, J.G., Schaibley, V.M., Ragland, R.L., Durkin, S.G., Warren, S.T. and Glover, T.W.
1046 (2009) Replication stress induces genome-wide copy number changes in human cells that
1047 resemble polymorphic and pathogenic variants. *Am J Hum Genet*, **84**, 339-350.

1048 21. Arlt, M.F., Ozdemir, A.C., Birkeland, S.R., Wilson, T.E. and Glover, T.W. (2011) Hydroxyurea
1049 induces de novo copy number variants in human cells. *Proc Natl Acad Sci U S A*, **108**, 17360-
1050 17365.
1051 22. Arlt, M.F., Rajendran, S., Birkeland, S.R., Wilson, T.E. and Glover, T.W. (2012) De novo CNV
1052 formation in mouse embryonic stem cells occurs in the absence of Xrcc4-dependent
1053 nonhomologous end joining. *PLoS Genet*, **8**, e1002981.
1054 23. Lauer, S., Avecilla, G., Spealman, P., Sethia, G., Brandt, N., Levy, S.F. and Gresham, D. (2018)
1055 Single-cell copy number variant detection reveals the dynamics and diversity of adaptation. *PLoS*
1056 *Biol*, **16**, e3000069.
1057 24. Oketch, D.J.A., Giulietti, M. and Piva, F. (2023) Copy Number Variations in Pancreatic Cancer:
1058 From Biological Significance to Clinical Utility. *Int J Mol Sci*, **25**, 391.
1059 25. Yi, D., Nam, J.W. and Jeong, H. (2023) Toward the functional interpretation of somatic structural
1060 variations: bulk- and single-cell approaches. *Brief Bioinform*, **24**, 297.
1061 26. Macaulay, I.C. and Voet, T. (2014) Single cell genomics: advances and future perspectives. *PLoS*
1062 *Genet*, **10**, e1004126.
1063 27. Wang, Y. and Navin, N.E. (2015) Advances and applications of single-cell sequencing
1064 technologies. *Mol Cell*, **58**, 598-609.
1065 28. Gawad, C., Koh, W. and Quake, S.R. (2016) Single-cell genome sequencing: current state of the
1066 science. *Nat Rev Genet*, **17**, 175-188.
1067 29. Wang, X., Chen, H. and Zhang, N.R. (2018) DNA copy number profiling using single-cell
1068 sequencing. *Brief Bioinform*, **19**, 731-736.
1069 30. Kidgell, C., Volkman, S.K., Daily, J., Borevitz, J.O., Plouffe, D., Zhou, Y., Johnson, J.R., Le Roch, K.,
1070 Sarr, O., Ndir, O. *et al.* (2006) A systematic map of genetic variation in Plasmodium falciparum.
1071 *PLoS Pathog*, **2**, e57.
1072 31. Ribacke, U., Mok, B.W., Wirta, V., Normark, J., Lundeberg, J., Kironde, F., Egwang, T.G., Nilsson,
1073 P. and Wahlgren, M. (2007) Genome wide gene amplifications and deletions in Plasmodium
1074 falciparum. *Molecular and Biochemical Parasitology*, **155**, 33-44.
1075 32. Cheeseman, I., Gomez-Escobar, N., Carret, C., Ivens, A., Stewart, L., Tetteh, K. and Conway, D.
1076 (2009) Gene copy number variation throughout the Plasmodium falciparum genome. *BMC*
1077 *Genomics*, **10**, 353.
1078 33. Bopp, S.E., Manary, M.J., Bright, A.T., Johnston, G.L., Dharia, N.V., Luna, F.L., McCormack, S.,
1079 Plouffe, D., McNamara, C.W., Walker, J.R. *et al.* (2013) Mitotic evolution of Plasmodium
1080 falciparum shows a stable core genome but recombination in antigen families. *PLoS Genet*, **9**,
1081 e1003293.
1082 34. Gendrot, M., Fawaz, R., Dormoi, J., Madamet, M. and Pradines, B. (2019) Genetic diversity and
1083 deletion of Plasmodium falciparum histidine-rich protein 2 and 3: a threat to diagnosis of P.
1084 falciparum malaria. *Clin Microbiol Infect*, **25**, 580-585.
1085 35. Matesanz, F., Tellez, M. and Alcina, A. (2003) The Plasmodium falciparum fatty acyl-CoA
1086 synthetase family (PfACS) and differential stage-specific expression in infected erythrocytes. *Mol*
1087 *Biochem Parasitol*, **126**, 109-112.
1088 36. Dankwa, S., Lim, C., Bei, A.K., Jiang, R.H., Abshire, J.R., Patel, S.D., Goldberg, J.M., Moreno, Y.,
1089 Kono, M., Niles, J.C. *et al.* (2016) Ancient human sialic acid variant restricts an emerging zoonotic
1090 malaria parasite. *Nat Commun*, **7**, 11187.
1091 37. Price, R.N., Uhlemann, A.-C., Brockman, A., McGready, R., Ashley, E., Phaipun, L., Patel, R., Laing,
1092 K., Looareesuwan, S., White, N.J. *et al.* (2004) Mefloquine resistance in Plasmodium falciparum
1093 and increased pfmdr1 gene copy number. *The Lancet*, **364**, 438-447.

1094 38. Ravenhall, M., Benavente, E.D., Sutherland, C.J., Baker, D.A., Campino, S. and Clark, T.G. (2019)
1095 An analysis of large structural variation in global Plasmodium falciparum isolates identifies a
1096 novel duplication of the chloroquine resistance associated gene. *Sci Rep*, **9**, 8287.
1097 39. Heinberg, A., Siu, E., Stern, C., Lawrence, E.A., Ferdig, M.T., Deitsch, K.W. and Kirkman, L.A.
1098 (2013) Direct evidence for the adaptive role of copy number variation on antifolate susceptibility
1099 in Plasmodium falciparum. *Mol Microbiol*, **88**, 702-712.
1100 40. Huckaby, A.C., Granum, C.S., Carey, M.A., Szlachta, K., Al-Barghouthi, B., Wang, Y.H. and Guler,
1101 J.L. (2019) Complex DNA structures trigger copy number variation across the Plasmodium
1102 falciparum genome. *Nucleic acids research*, **47**, 1615-1627.
1103 41. Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A.,
1104 Nelson, K.E., Bowman, S. *et al.* (2002) Genome sequence of the human malaria parasite
1105 Plasmodium falciparum. *Nature*, **419**, 498-511.
1106 42. Liu, S., Huckaby, A.C., Brown, A.C., Moore, C.C., Burbulis, I., McConnell, M.J. and Guler, J.L.
1107 (2021) Single-cell sequencing of the small and AT-skewed genome of malaria parasites. *Genome*
1108 *Med*, **13**, 75.
1109 43. Brown, N., Danis, C., Ahmedjanova, V. and Guler, J.L. (2025) SVCROWS: A User-Defined Tool for
1110 Interpreting Significant Structural Variants in Heterogeneous Datasets. *bioRxiv*.
1111 44. Yu, X., Qin, F., Liu, S., Brown, N.J., Lu, Q., Cai, G., Guler, J.L. and Xiao, F. (2025) HapCNV: A
1112 Comprehensive Framework for CNV Detection in Low-input DNA Sequencing Data. *bioRxiv*.
1113 45. Galhardo, R.S., Hastings, P.J. and Rosenberg, S.M. (2007) Mutation as a stress response and the
1114 regulation of evolvability. *Crit Rev Biochem Mol Biol*, **42**, 399-435.
1115 46. Arlt, M.F., Wilson, T.E. and Glover, T.W. (2012) Replication stress and mechanisms of CNV
1116 formation. *Curr Opin Genet Dev*, **22**, 204-210.
1117 47. Glover, T.W., Wilson, T.E. and Arlt, M.F. (2017) Fragile sites in cancer: more than meets the eye.
1118 *Nat Rev Cancer*, **17**, 489-501.
1119 48. McDaniels, J.M., Huckaby, A.C., Carter, S.A., Lingeman, S., Francis, A., Congdon, M., Santos, W.,
1120 Rathod, P.K. and Guler, J.L. (2021) Extrachromosomal DNA amplicons in antimalarial-resistant
1121 Plasmodium falciparum. *Mol Microbiol*, **115**, 574-590.
1122 49. Liu, S., Zulawinska, J., Ebel, E.R., Luniewski, A., Danis, C., Simpson, M.L., Kim, J., Ene, N.,
1123 Braukmann, T.W.A., Congdon, M. *et al.* (2025) Direct long-read visualization reveals hidden
1124 variation in GCH1 gene copy number and precise expansion steps. *BMC Genomics*, **26**, 671.
1125 50. Brown, A.C., Moore, C.C. and Guler, J.L. (2020) Cholesterol-dependent enrichment of
1126 understudied erythrocytic stages of human Plasmodium parasites. *Sci Rep*, **10**, 4591.
1127 51. Rodrigues, O.R. and Monard, S. (2016) A rapid method to verify single-cell deposition setup for
1128 cell sorters. *Cytometry A*, **89**, 594-600.
1129 52. Nam, D., Kim, S., Kim, J.H., Lee, S., Kim, D., Son, J., Kim, D., Cha, B.S., Lee, E.S. and Park, K.S.
1130 (2023) Low-Temperature Loop-Mediated Isothermal Amplification Operating at Physiological
1131 Temperature. *Biosensors (Basel)*, **13**.
1132 53. Brown, N., da Silva, C., Webb, C., Matias, D., Dias, B., Cancio, B., Silva, M., Viegas, R., Salvador,
1133 C., Chivale, N. *et al.* (2024) Antimalarial resistance risk in Mozambique detected by a novel
1134 quadruplex droplet digital PCR assay. *Antimicrob Agents Chemother*, **68**, e0034624.
1135 54. Daniels, R., Ndiaye, D., Wall, M., McKinney, J., Sene, P.D., Sabeti, P.C., Volkman, S.K., Mboup, S.
1136 and Wirth, D.F. (2012) Rapid, field-deployable method for genotyping and discovery of single-
1137 nucleotide polymorphisms associated with drug resistance in Plasmodium falciparum.
1138 *Antimicrob Agents Chemother*, **56**, 2976-2986.
1139 55. Kassaza, K., Long, A.C., McDaniels, J.M., Andre, M., Fredrickson, W., Nyehangane, D., Orikiriza,
1140 P., Operario, D.J., Bazira, J., Mwanga-Amumpaire, J.A. *et al.* (2021) Surveillance of Plasmodium

1141 falciparum pfcrt haplotypes in southwestern uganda by high-resolution melt analysis. *Malar J*,
1142 **20**, 114.
1143 56. Alker, A.P., Mwapasa, V. and Meshnick, S.R. (2004) Rapid real-time PCR genotyping of mutations
1144 associated with sulfadoxine-pyrimethamine resistance in Plasmodium falciparum. *Antimicrob*
1145 *Agents Chemother*, **48**, 2924-2929.
1146 57. Chiang, C., Layer, R.M., Faust, G.G., Lindberg, M.R., Rose, D.B., Garrison, E.P., Marth, G.T.,
1147 Quinlan, A.R. and Hall, I.M. (2015) SpeedSeq: ultra-fast personal genome analysis and
1148 interpretation. *Nature methods*, **12**, 966-968.
1149 58. Otto, T.D., Bohme, U., Sanders, M., Reid, A., Bruske, E.I., Duffy, C.W., Bull, P.C., Pearson, R.D.,
1150 Abdi, A., Dimonte, S. *et al.* (2018) Long read assemblies of geographically dispersed Plasmodium
1151 falciparum isolates reveal highly structured subtelomeres. *Wellcome Open Res*, **3**, 52.
1152 59. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G.,
1153 Durbin, R. and Genome Project Data Processing, S. (2009) The Sequence Alignment/Map format
1154 and SAMtools. *Bioinformatics*, **25**, 2078-2079.
1155 60. Garcia-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L.M., Gotz, S., Tarazona, S., Dopazo, J.,
1156 Meyer, T.F. and Conesa, A. (2012) Qualimap: evaluating next-generation sequencing alignment
1157 data. *Bioinformatics*, **28**, 2678-2679.
1158 61. Huang, L., Ma, F., Chapman, A., Lu, S. and Xie, X.S. (2015) Single-Cell Whole-Genome
1159 Amplification and Sequencing: Methodology and Applications. *Annu Rev Genomics Hum Genet*,
1160 **16**, 79-102.
1161 62. Chen, C., Xing, D., Tan, L., Li, H., Zhou, G., Huang, L. and Xie, X.S. (2017) Single-cell whole-
1162 genome analyses by Linear Amplification via Transposon Insertion (LIANTI). *Science*, **356**, 189-
1163 194.
1164 63. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment
1165 search tool. *J Mol Biol*, **215**, 403-410.
1166 64. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K.,
1167 Altshuler, D., Gabriel, S., Daly, M. *et al.* (2010) The Genome Analysis Toolkit: a MapReduce
1168 framework for analyzing next-generation DNA sequencing data. *Genome Res*, **20**, 1297-1303.
1169 65. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del
1170 Angel, G., Rivas, M.A., Hanna, M. *et al.* (2011) A framework for variation discovery and
1171 genotyping using next-generation DNA sequencing data. *Nat Genet*, **43**, 491-498.
1172 66. Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A.,
1173 Jordan, T., Shakir, K., Roazen, D., Thibault, J. *et al.* (2013) From FastQ data to high confidence
1174 variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*, **43**,
1175 1110.
1176 67. MalariaGen, Ahouidi, A., Ali, M., Almagro-Garcia, J., Amambua-Ngwa, A., Amaratunga, C.,
1177 Amato, R., Amenga-Etego, L., Andagalu, B., Anderson, T.J.C. *et al.* (2021) An open dataset of
1178 Plasmodium falciparum genome variation in 7,000 worldwide samples. *Wellcome Open Res*, **6**,
1179 42.
1180 68. Miles, A., Iqbal, Z., Vauterin, P., Pearson, R., Campino, S., Theron, M., Gould, K., Mead, D., Drury,
1181 E., O'Brien, J. *et al.* (2016) Indels, structural variation, and recombination drive genomic diversity
1182 in Plasmodium falciparum. *Genome Res*, **26**, 1288-1299.
1183 69. Abyzov, A., Urban, A.E., Snyder, M. and Gerstein, M. (2011) CNVnator: an approach to discover,
1184 genotype, and characterize typical and atypical CNVs from family and population genome
1185 sequencing. *Genome Res*, **21**, 974-984.
1186 70. Layer, R.M., Chiang, C., Quinlan, A.R. and Hall, I.M. (2014) LUMPY: a probabilistic framework for
1187 structural variant discovery. *Genome Biol*, **15**, R84.

1188 71. Venkatraman, E.S. and Olshen, A.B. (2007) A faster circular binary segmentation algorithm for
1189 the analysis of array CGH data. *Bioinformatics*, **23**, 657-663.
1190 72. Xiao, F., Niu, Y., Hao, N., Xu, Y., Jin, Z. and Zhang, H. (2017) modSaRa: a computationally efficient
1191 R package for CNV identification. *Bioinformatics*, **33**, 2384-2385.
1192 73. Jiang, Y., Wang, R., Urrutia, E., Anastopoulos, I.N., Nathanson, K.L. and Zhang, N.R. (2018)
1193 CODEX2: full-spectrum copy number variation detection by high-throughput DNA sequencing.
1194 *Genome Biol*, **19**, 202.
1195 74. Mi, H., Ebert, D., Muruganujan, A., Mills, C., Albou, L.P., Mushayamaha, T. and Thomas, P.D.
1196 (2021) PANTHER version 16: a revised family classification, tree-based classification tool,
1197 enhancer regions and extensive API. *Nucleic acids research*, **49**, D394-D403.
1198 75. Singh, G. and Gupta, D. (2022) In-Silico Functional Annotation of Plasmodium falciparum
1199 Hypothetical Proteins to Identify Novel Drug Targets. *Front Genet*, **13**, 821516.
1200 76. Zhang, M., Wang, C., Otto, T.D., Oberstaller, J., Liao, X., Adapa, S.R., Udenze, K., Bronner, I.F.,
1201 Casandra, D., Mayho, M. *et al.* (2018) Uncovering the essential genes of the human malaria
1202 parasite Plasmodium falciparum by saturation mutagenesis. *Science*, **360**, eaap7847.
1203 77. Aurrecoechea, C., Brestelli, J., Brunk, B.P., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A.,
1204 Grant, G., Harb, O.S. *et al.* (2009) PlasmoDB: a functional genomic database for malaria
1205 parasites. *Nucleic acids research*, **37**, D539-543.
1206 78. Reilly, H.B., Wang, H., Steuter, J.A., Marx, A.M. and Ferdig, M.T. (2007) Quantitative dissection of
1207 clone-specific growth rates in cultured malaria parasites. *Int J Parasitol*, **37**, 1599-1607.
1208 79. Reilly Ayala, H.B., Wacker, M.A., Siwo, G. and Ferdig, M.T. (2010) Quantitative trait loci mapping
1209 reveals candidate pathways regulating cell cycle duration in Plasmodium falciparum. *BMC*
1210 *Genomics*, **11**, 577.
1211 80. Phillips, M.A., Gujjar, R., Malmquist, N.A., White, J., El Mazouni, F., Baldwin, J. and Rathod, P.K.
1212 (2008) Triazolopyrimidine-Based Dihydroorotate Dehydrogenase Inhibitors with Potent and
1213 Selective Activity against the Malaria Parasite Plasmodium falciparum. *Journal of Medicinal*
1214 *Chemistry*, **51**, 3649-3653.
1215 81. Sanz, L.M., Crespo, B., De-Cozar, C., Ding, X.C., Llergo, J.L., Burrows, J.N., Garcia-Bustos, J.F. and
1216 Gamo, F.J. (2012) P. falciparum in vitro killing rates allow to discriminate between different
1217 antimalarial mode-of-action. *PloS one*, **7**, e30949.
1218 82. Inselburg, J. and Banyal, H.S. (1984) Synthesis of DNA during the asexual cycle of Plasmodium
1219 falciparum in culture. *Mol Biochem Parasitol*, **10**, 79-87.
1220 83. Inselburg, J. and Banyal, H.S. (1984) Plasmodium falciparum: synchronization of asexual
1221 development with aphidicolin, a DNA synthesis inhibitor. *Exp Parasitol*, **57**, 48-54.
1222 84. Sugino, A. and Nakayama, K. (1980) DNA polymerase alpha mutants from a Drosophila
1223 melanogaster cell line. *Proc Natl Acad Sci U S A*, **77**, 7049-7053.
1224 85. Petermann, E., Orta, M.L., Issaeva, N., Schultz, N. and Helleday, T. (2010) Hydroxyurea-stalled
1225 replication forks become progressively inactivated and require two different RAD51-mediated
1226 pathways for restart and repair. *Mol Cell*, **37**, 492-502.
1227 86. Mannava, S., Moparthy, K.C., Wheeler, L.J., Natarajan, V., Zucker, S.N., Fink, E.E., Im, M.,
1228 Flanagan, S., Burhans, W.C., Zeitouni, N.C. *et al.* (2013) Depletion of deoxyribonucleotide pools
1229 is an endogenous source of DNA damage in cells undergoing oncogene-induced senescence. *Am*
1230 *J Pathol*, **182**, 142-151.
1231 87. Jensen, J.B. and Trager, W. (1978) Plasmodium falciparum in culture: establishment of additional
1232 strains. *The American journal of tropical medicine and hygiene*, **27**, 743-746.
1233 88. Dolan, S.A., Miller, L.H. and Wellems, T.E. (1990) Evidence for a switching mechanism in the
1234 invasion of erythrocytes by Plasmodium falciparum. *J Clin Invest*, **86**, 618-624.

1235 89. Tian, S., Yan, H., Neuhauser, C. and Slager, S.L. (2016) An analytical workflow for accurate variant
1236 discovery in highly divergent regions. *BMC Genomics*, **17**, 703.
1237 90. Sanderson, N.D., Swann, J., Barker, L., Kavanagh, J., Hoosdally, S., Crook, D., GonFast
1238 Investigators, G., Street, T.L. and Eyre, D.W. (2020) High precision Neisseria gonorrhoeae variant
1239 and antimicrobial resistance calling from metagenomic Nanopore sequencing. *Genome Res*, **30**,
1240 1354-1363.
1241 91. Cameron, D.L., Di Stefano, L. and Papenfuss, A.T. (2019) Comprehensive evaluation and
1242 characterisation of short read general-purpose structural variant calling software. *Nat Commun*,
1243 **10**, 3240.
1244 92. Gabrielaite, M., Torp, M.H., Rasmussen, M.S., Andreu-Sanchez, S., Vieira, F.G., Pedersen, C.B.,
1245 Kinalis, S., Madsen, M.B., Kodama, M., Demircan, G.S. *et al.* (2021) A Comparison of Tools for
1246 Copy-Number Variation Detection in Germline Whole Exome and Whole Genome Sequencing
1247 Data. *Cancers (Basel)*, **13**.
1248 93. Kosugi, S., Momozawa, Y., Liu, X., Terao, C., Kubo, M. and Kamatani, Y. (2019) Comprehensive
1249 evaluation of structural variation detection algorithms for whole genome sequencing. *Genome*
1250 *Biol*, **20**, 117.
1251 94. Dessens, J.T., Beetsma, A.L., Dimopoulos, G., Wengelnik, K., Crisanti, A., Kafatos, F.C. and Sinden,
1252 R.E. (1999) CTRP is essential for mosquito infection by malaria ookinetes. *EMBO J*, **18**, 6221-
1253 6227.
1254 95. Steinbuechel, M. and Matuschewski, K. (2009) Role for the Plasmodium sporozoite-specific
1255 transmembrane protein S6 in parasite motility and efficient malaria transmission. *Cell Microbiol*,
1256 **11**, 279-288.
1257 96. Eksi, S., Haile, Y., Furuya, T., Ma, L., Su, X. and Williamson, K.C. (2005) Identification of a
1258 subtelomeric gene family expressed during the asexual-sexual stage transition in Plasmodium
1259 falciparum. *Mol Biochem Parasitol*, **143**, 90-99.
1260 97. Silvestrini, F., Lasonder, E., Olivieri, A., Camarda, G., van Schaijk, B., Sanchez, M., Younis Younis,
1261 S., Sauerwein, R. and Alano, P. (2010) Protein export marks the early phase of
1262 gametocytogenesis of the human malaria parasite Plasmodium falciparum. *Mol Cell Proteomics*,
1263 **9**, 1437-1448.
1264 98. Mueller, A.K., Labaied, M., Kappe, S.H. and Matuschewski, K. (2005) Genetically modified
1265 Plasmodium parasites as a protective experimental malaria vaccine. *Nature*, **433**, 164-167.
1266 99. Silvie, O., Goetz, K. and Matuschewski, K. (2008) A sporozoite asparagine-rich protein controls
1267 initiation of Plasmodium liver stage development. *PLoS Pathog*, **4**, e1000086.
1268 100. van Schaijk, B.C., Ploemen, I.H., Annoura, T., Vos, M.W., Foquet, L., van Gemert, G.J., Chevalley-
1269 Maurel, S., van de Vegte-Bolmer, M., Sajid, M., Franetich, J.F. *et al.* (2014) A genetically
1270 attenuated malaria vaccine candidate based on P. falciparum b9/slarp gene-deficient
1271 sporozoites. *Elife*, **3**, e03582.
1272 101. Real, E., Rodrigues, L., Cabal, G.G., Enguita, F.J., Mancio-Silva, L., Mello-Vieira, J., Beatty, W.,
1273 Vera, I.M., Zuzarte-Luis, V., Figueira, T.N. *et al.* (2018) Plasmodium UIS3 sequesters host LC3 to
1274 avoid elimination by autophagy in hepatocytes. *Nat Microbiol*, **3**, 17-25.
1275 102. Roestenberg, M., Walk, J., van der Boor, S.C., Langenberg, M.C.C., Hoogerwerf, M.A., Janse, J.J.,
1276 Manurung, M., Yap, X.Z., Garcia, A.F., Koopman, J.P.R. *et al.* (2020) A double-blind, placebo-
1277 controlled phase 1/2a trial of the genetically attenuated malaria vaccine PfSPZ-GA1. *Sci Transl*
1278 *Med*, **12**, eaaz5629.
1279 103. Franke-Fayard, B., Marin-Mogollon, C., Geurten, F.J.A., Chevalley-Maurel, S., Ramesar, J., Kroeze,
1280 H., Baalbergen, E., Wessels, E., Baron, L., Soulard, V. *et al.* (2022) Creation and preclinical
1281 evaluation of genetically attenuated malaria parasites arresting growth late in the liver. *NPJ*
1282 *Vaccines*, **7**, 139.

1283 104. Van Tyne, D., Park, D.J., Schaffner, S.F., Neafsey, D.E., Angelino, E., Cortese, J.F., Barnes, K.G.,
1284 Rosen, D.M., Lukens, A.K., Daniels, R.F. *et al.* (2011) Identification and functional validation of
1285 the novel antimalarial resistance locus PF10_0355 in Plasmodium falciparum. *PLoS Genet*, **7**,
1286 e1001383.
1287 105. Farrukh, A., Musabyimana, J.P., Distler, U., Mahlich, V.J., Mueller, J., Bick, F., Tenzer, S., Pradel,
1288 G. and Ngwa, C.J. (2024) The Plasmodium falciparum CCCH zinc finger protein MD3 regulates
1289 male gametocytogenesis through its interaction with RNA-binding proteins. *Mol Microbiol*, **121**,
1290 543-564.
1291 106. Bushell, E.S., Ecker, A., Schlegelmilch, T., Goulding, D., Dougan, G., Sinden, R.E., Christophides,
1292 G.K., Kafatos, F.C. and Vlachou, D. (2009) Paternal effect of the nuclear formin-like protein
1293 MISFIT on Plasmodium development in the mosquito vector. *PLoS Pathog*, **5**, e1000539.
1294 107. Mair, G.R., Lasonder, E., Garver, L.S., Franke-Fayard, B.M., Carret, C.K., Wiegant, J.C., Dirks, R.W.,
1295 Dimopoulos, G., Janse, C.J. and Waters, A.P. (2010) Universal features of post-transcriptional
1296 gene regulation are critical for Plasmodium zygote development. *PLoS Pathog*, **6**, e1000767.
1297 108. Slavic, K., Krishna, S., Lahree, A., Bouyer, G., Hanson, K.K., Vera, I., Pittman, J.K., Staines, H.M.
1298 and Mota, M.M. (2016) A vacuolar iron-transporter homologue acts as a detoxifier in
1299 Plasmodium. *Nat Commun*, **7**, 10403.
1300 109. Perrin, A.J., Bisson, C., Faull, P.A., Renshaw, M.J., Lees, R.A., Fleck, R.A., Saibil, H.R., Snijders,
1301 A.P., Baker, D.A. and Blackman, M.J. (2021) Malaria Parasite Schizont Egress Antigen-1 Plays an
1302 Essential Role in Nuclear Segregation during Schizogony. *mBio*, **12**, e03377-03320.
1303 110. Pradhan, A., Hussain, E.M. and Tuteja, R. (2008) Characterization of replication fork and
1304 phosphorylation stimulated Plasmodium falciparum helicase 45. *Gene*, **420**, 66-75.
1305 111. Evans, L., Gowers, D., Firman, K. and Youell, J. (2012) Enhanced purification and characterization
1306 of the PfeIF4A (PfH45) helicase from Plasmodium falciparum using a codon-optimised clone.
1307 *Protein Expr Purif*, **85**, 1-8.
1308 112. Maier, A.G., Rug, M., O'Neill, M.T., Brown, M., Chakravorty, S., Szestak, T., Chesson, J., Wu, Y.,
1309 Hughes, K., Coppel, R.L. *et al.* (2008) Exported proteins required for virulence and rigidity of
1310 Plasmodium falciparum-infected human erythrocytes. *Cell*, **134**, 48-61.
1311 113. Ajibola, O., Diop, M.F., Ghansah, A., Amenga-Etego, L., Golassa, L., Apinjoh, T.,
1312 Randrianarivelojosia, M., Maiga-Ascofare, O., Yavo, W., Bouyou-Akotet, M. *et al.* (2021) In silico
1313 characterisation of putative Plasmodium falciparum vaccine candidates in African malaria
1314 populations. *Sci Rep*, **11**, 16215.
1315 114. Raj, D.K., Nixon, C.P., Nixon, C.E., Dvorin, J.D., DiPetrillo, C.G., Pond-Tor, S., Wu, H.W., Jolly, G.,
1316 Pischel, L., Lu, A. *et al.* (2014) Antibodies to PfSEA-1 block parasite egress from RBCs and protect
1317 against malaria infection. *Science*, **344**, 871-877.
1318 115. Winter, G., Kawai, S., Haeggstrom, M., Kaneko, O., von Euler, A., Kawazu, S., Palm, D.,
1319 Fernandez, V. and Wahlgren, M. (2005) SURFIN is a polymorphic antigen expressed on
1320 Plasmodium falciparum merozoites and infected erythrocytes. *J Exp Med*, **201**, 1853-1863.
1321 116. Weiner, J. and Kooij, T.W. (2016) Phylogenetic profiles of all membrane transport proteins of the
1322 malaria parasite highlight new drug targets. *Microb Cell*, **3**, 511-521.
1323 117. Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P.,
1324 Nahed, B.V., Curry, W.T., Martuza, R.L. *et al.* (2014) Single-cell RNA-seq highlights intratumoral
1325 heterogeneity in primary glioblastoma. *Science*, **344**, 1396-1401.
1326 118. Xu, Q.H., Chen, S.H., Hu, Y.B. and Huang, W. (2021) Single-cell RNA transcriptome reveals the
1327 intra-tumoral heterogeneity and regulators underlying tumor progression in metastatic
1328 pancreatic ductal adenocarcinoma. *Cell Death Discov*, **7**.

1329 119. Gao, R.L., Bai, S.S., Henderson, Y.C., Lin, Y.Y., Schalck, A., Yan, Y., Kumar, T., Hu, M., Sei, E., Davis,
1330 A. *et al.* (2021) Delineating copy number and clonal substructure in human tumors from single-
1331 cell transcriptomes. *Nature Biotechnology*, **39**, 599-608.
1332 120. Mahdipour-Shirayeh, A., Erdmann, N., Leung-Hagesteijn, C. and Tiedemann, R.E. (2022) sciCNV:
1333 high-throughput paired profiling of transcriptomes and DNA copy number variations at single-
1334 cell resolution. *Briefings in Bioinformatics*, **23**, bbab413.
1335 121. Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., Cook, K., Stepansky, A.,
1336 Levy, D., Esposito, D. *et al.* (2011) Tumour evolution inferred by single-cell sequencing. *Nature*,
1337 **472**, 90-94.
1338 122. McConnell, M.J., Lindberg, M.R., Brennand, K.J., Piper, J.C., Voet, T., Cowing-Zitron, C.,
1339 Shumilina, S., Lasken, R.S., Vermeesch, J.R., Hall, I.M. *et al.* (2013) Mosaic copy number variation
1340 in human neurons. *Science*, **342**, 632-637.
1341 123. Cai, X., Evrony, G.D., Lehmann, H.S., Elhosary, P.C., Mehta, B.K., Poduri, A. and Walsh, C.A.
1342 (2014) Single-cell, genome-wide sequencing identifies clonal somatic copy-number variation in
1343 the human brain. *Cell Rep*, **8**, 1280-1289.
1344 124. Francis, J.M., Zhang, C.Z., Maire, C.L., Jung, J., Manzo, V.E., Adalsteinsson, V.A., Homer, H.,
1345 Haidar, S., Blumenstiel, B., Pedamallu, C.S. *et al.* (2014) EGFR variant heterogeneity in
1346 glioblastoma resolved through single-nucleus sequencing. *Cancer Discov*, **4**, 956-971.
1347 125. Eirew, P., Steif, A., Khattra, J., Ha, G., Yap, D., Farahani, H., Gelmon, K., Chia, S., Mar, C., Wan, A.
1348 *et al.* (2015) Dynamics of genomic clones in breast cancer patient xenografts at single-cell
1349 resolution. *Nature*, **518**, 422-426.
1350 126. Rohrback, S., April, C., Kaper, F., Rivera, R.R., Liu, C.S., Siddoway, B. and Chun, J. (2018)
1351 Submegabase copy number variations arise during cerebral cortical neurogenesis as revealed by
1352 single-cell whole-genome sequencing. *Proc Natl Acad Sci U S A*, **115**, 10804-10809.
1353 127. Baslan, T., Kendall, J., Volyanskyy, K., McNamara, K., Cox, H., D'Italia, S., Ambrosio, F., Riggs, M.,
1354 Rodgers, L., Leotta, A. *et al.* (2020) Novel insights into breast cancer copy number genetic
1355 heterogeneity revealed by single-cell genome sequencing. *Elife*, **9**, e51480.
1356 128. Negreira, G.H., Monsieurs, P., Imamura, H., Maes, I., Kuk, N., Yagoubat, A., Van den Broeck, F.,
1357 Sterkers, Y., Dujardin, J.C. and Domagalska, M.A. (2022) High throughput single-cell genome
1358 sequencing gives insights into the generation and evolution of mosaic aneuploidy in Leishmania
1359 donovani. *Nucleic acids research*, **50**, 293-305.
1360 129. Nair, S., Nkhoma, S.C., Serre, D., Zimmerman, P.A., Gorena, K., Daniel, B.J., Nosten, F., Anderson,
1361 T.J. and Cheeseman, I.H. (2014) Single-cell genomics for dissection of complex malaria
1362 infections. *Genome Res*, **24**, 1028-1038.
1363 130. Trevino, S.G., Nkhoma, S.C., Nair, S., Daniel, B.J., Moncada, K., Khoswe, S., Banda, R.L., Nosten, F.
1364 and Cheeseman, I.H. (2017) High-Resolution Single-Cell Sequencing of Malaria Parasites.
1365 *Genome Biol Evol*, **9**, 3373-3383.
1366 131. Saxena, S. and Zou, L. (2022) Hallmarks of DNA replication stress. *Mol Cell*, **82**, 2298-2314.
1367 132. Guler, J.L., Freeman, D.L., Ahyong, V., Patrapuvich, R., White, J., Gujjar, R., Phillips, M.A., DeRisi,
1368 J. and Rathod, P.K. (2013) Asexual populations of the human malaria parasite, Plasmodium
1369 falciparum, use a two-step genomic strategy to acquire accurate, beneficial DNA amplifications.
1370 *PLoS Pathog*, **9**, e1003375.
1371 133. Dunham, M.J., Badrane, H., Ferea, T., Adams, J., Brown, P.O., Rosenzweig, F. and Botstein, D.
1372 (2002) Characteristic genome rearrangements in experimental evolution of Saccharomyces
1373 cerevisiae. *Proc Natl Acad Sci U S A*, **99**, 16144-16149.
1374 134. Gresham, D., Desai, M.M., Tucker, C.M., Jenq, H.T., Pai, D.A., Ward, A., DeSevo, C.G., Botstein, D.
1375 and Dunham, M.J. (2008) The repertoire and dynamics of evolutionary adaptations to controlled
1376 nutrient-limited environments in yeast. *PLoS Genet*, **4**, e1000303.

1377 135. Sandegren, L. and Andersson, D.I. (2009) Bacterial gene amplification: implications for the
1378 evolution of antibiotic resistance. *Nat Rev Microbiol*, **7**, 578-588.
1379 136. Dharia, N., Sidhu, A., Cassera, M., Westenberger, S., Bopp, S., Eastman, R., Plouffe, D., Batalov,
1380 S., Park, D., Volkman, S. *et al.* (2009) Use of high-density tiling microarrays to identify mutations
1381 globally and elucidate mechanisms of drug resistance in Plasmodium falciparum. *Genome*
1382 *Biology*, **10**, R21.
1383 137. Rottmann, M., McNamara, C., Yeung, B.K., Lee, M.C., Zou, B., Russell, B., Seitz, P., Plouffe, D.M.,
1384 Dharia, N.V., Tan, J. *et al.* (2010) Spiroindolones, a potent compound class for the treatment of
1385 malaria. *Science*, **329**, 1175-1180.
1386 138. Ross, L.S., Gamo, F.J., Lafuente-Monasterio, M.J., Singh, O.M., Rowland, P., Wiegand, R.C. and
1387 Wirth, D.F. (2014) In Vitro Resistance Selections for Plasmodium falciparum Dihydroorotate
1388 Dehydrogenase Inhibitors Give Mutants with Multiple Point Mutations in the Drug-binding Site
1389 and Altered Growth. *J Biol Chem*, **289**, 17980-17995.
1390 139. Phillips, M.A., Lotharius, J., Marsh, K., White, J., Dayan, A., White, K.L., Njoroge, J.W., El Mazouni,
1391 F., Lao, Y., Kokkonda, S. *et al.* (2015) A long-duration dihydroorotate dehydrogenase inhibitor
1392 (DSM265) for prevention and treatment of malaria. *Sci Transl Med*, **7**, 296ra111.
1393 140. Cowell, A. and Winzeler, E. (2018) Exploration of the Plasmodium falciparum Resistome and
1394 Druggable Genome Reveals New Mechanisms of Drug Resistance and Antimalarial Targets.
1395 *Microbiol Insights*, **11**.
1396 141. Rocamora, F., Zhu, L., Liong, K.Y., Dondorp, A., Miotto, O., Mok, S. and Bozdech, Z. (2018)
1397 Oxidative stress and protein damage responses mediate artemisinin resistance in malaria
1398 parasites. *PLoS Pathog*, **14**, e1006930.
1399 142. Luth, M.R., Godinez-Macias, K.P., Chen, D., Okombo, J., Thathy, V., Cheng, X., Daggupati, S.,
1400 Davies, H., Dhingra, S.K., Economy, J.M. *et al.* (2024) Systematic in vitro evolution in Plasmodium
1401 falciparum reveals key determinants of drug resistance. *Science*, **386**, eadk9893.
1402 143. Evrony, G.D., Lee, E., Park, P.J. and Walsh, C.A. (2016) Resolving rates of mutation in the brain
1403 using single-neuron genomics. *Elife*, **5**, e12966.
1404 144. Mills, R.E., Walter, K., Stewart, C., Handsaker, R.E., Chen, K., Alkan, C., Abyzov, A., Yoon, S.C., Ye,
1405 K., Cheetham, R.K. *et al.* (2011) Mapping copy number variation by population-scale genome
1406 sequencing. *Nature*, **470**, 59-65.
1407 145. Zhao, M., Wang, Q., Wang, Q., Jia, P. and Zhao, Z. (2013) Computational tools for copy number
1408 variation (CNV) detection using next-generation sequencing data: features and perspectives.
1409 *BMC Bioinformatics*, **14 Suppl 11**, S1.
1410 146. Guan, P. and Sung, W.K. (2016) Structural variation detection using next-generation sequencing
1411 data: A comparative technical review. *Methods*, **102**, 36-49.
1412 147. Gabrielaite, M., Torp, M.H., Rasmussen, M.S., Andreu-Sanchez, S., Vieira, F.G., Pedersen, C.B.,
1413 Kinalis, S., Madsen, M.B., Kodama, M., Demircan, G.S. *et al.* (2021) A Comparison of Tools for
1414 Copy-Number Variation Detection in Germline Whole Exome and Whole Genome Sequencing
1415 Data. *Cancers (Basel)*, **13**, 6283.
1416 148. Arnot, D.E., Ronander, E. and Bengtsson, D.C. (2011) The progression of the intra-erythrocytic
1417 cell cycle of Plasmodium falciparum and the role of the centriolar plaques in asynchronous
1418 mitotic division during schizogony. *Int J Parasitol*, **41**, 71-80.
1419 149. Zong, C., Lu, S., Chapman, A.R. and Xie, X.S. (2012) Genome-wide detection of single-nucleotide
1420 and copy-number variations of a single human cell. *Science*, **338**, 1622-1626.
1421 150. Lasken, R.S. (2013) Single-cell sequencing in its prime. *Nat Biotechnol*, **31**, 211-212.
1422 151. Laks, E., McPherson, A., Zahn, H., Lai, D., Steif, A., Brimhall, J., Biele, J., Wang, B., Masud, T., Ting,
1423 J. *et al.* (2019) Clonal Decomposition and DNA Replication States Defined by Scaled Single-Cell
1424 Genome Sequencing. *Cell*, **179**, 1207-1221 e1222.

1425 152. Kupke, J., Brombach, J., Fang, Y., Wolf, S.A., Thrukonda, L., Ghazisaeedi, F., Kuropka, B., Hanke,
1426 D., Semmler, T., Nordholt, N. *et al.* (2025) Heteroresistance in Enterobacter cloacae complex
1427 caused by variation in transient gene amplification events. *NPJ Antimicrob Resist*, **3**, 13.
1428 153. Kirkman, L.A., Lawrence, E.A. and Deitsch, K.W. (2014) Malaria parasites utilize both
1429 homologous recombination and alternative end joining pathways to maintain genome integrity.
1430 *Nucleic acids research*, **42**, 370-379.
1431 154. Triglia, T., Foote, S.J., Kemp, D.J. and Cowman, A.F. (1991) Amplification of the multidrug
1432 resistance gene pfmdr1 in Plasmodium falciparum has arisen as multiple independent events.
1433 *Mol. Cell. Biol.*, **11**, 5244-5250.
1434 155. Matthews, H., Duffy, C.W. and Merrick, C.J. (2018) Checks and balances? DNA replication and
1435 the cell cycle in Plasmodium. *Parasit Vectors*, **11**, 216.
1436 156. Bindra, R.S., Schaffer, P.J., Meng, A., Woo, J., Maseide, K., Roth, M.E., Lizardi, P., Hedley, D.W.,
1437 Bristow, R.G. and Glazer, P.M. (2004) Down-regulation of Rad51 and decreased homologous
1438 recombination in hypoxic cancer cells. *Mol Cell Biol*, **24**, 8504-8518.
1439 157. Saleh-Gohari, N., Bryant, H.E., Schultz, N., Parker, K.M., Cassel, T.N. and Helleday, T. (2005)
1440 Spontaneous homologous recombination is induced by collapsed replication forks that are
1441 caused by endogenous DNA single-strand breaks. *Mol Cell Biol*, **25**, 7158-7169.
1442 158. Slack, A., Thornton, P.C., Magner, D.B., Rosenberg, S.M. and Hastings, P.J. (2006) On the
1443 mechanism of gene amplification induced under stress in Escherichia coli. *PLoS Genet*, **2**, e48.
1444 159. Hastings, P.J., Ira, G. and Lupski, J.R. (2009) A microhomology-mediated break-induced
1445 replication model for the origin of human copy number variation. *PLoS Genet*, **5**, e1000327.
1446 160. Valerie, K. and Povirk, L.F. (2003) Regulation and mechanisms of mammalian double-strand
1447 break repair. *Oncogene*, **22**, 5792-5812.
1448 161. Hastings, P.J., Bull, H.J., Klump, J.R. and Rosenberg, S.M. (2000) Adaptive amplification: an
1449 inducible chromosomal instability mechanism. *Cell*, **103**, 723-731.
1450 162. Cole, G.M., Schild, D., Lovett, S.T. and Mortimer, R.K. (1987) Regulation of RAD54- and RAD52-
1451 lacZ gene fusions in Saccharomyces cerevisiae in response to DNA damage. *Mol Cell Biol*, **7**,
1452 1078-1084.
1453 163. Raghavan, D., Shipley, W.U., Garnick, M.B., Russell, P.J. and Richie, J.P. (1990) Biology and
1454 management of bladder cancer. *N Engl J Med*, **322**, 1129-1138.
1455 164. Bentley, J., Diggle, C.P., Harnden, P., Knowles, M.A. and Kiltie, A.E. (2004) DNA double strand
1456 break repair in human bladder cancer is error prone and involves microhomology-associated
1457 end-joining. *Nucleic acids research*, **32**, 5249-5259.
1458 165. White, N.J. (2013) Pharmacokinetic and pharmacodynamic considerations in antimalarial dose
1459 optimization. *Antimicrob Agents Chemother*, **57**, 5792-5807.
1460 166. Durkin, S.G., Ragland, R.L., Arlt, M.F., Mulle, J.G., Warren, S.T. and Glover, T.W. (2008)
1461 Replication stress induces tumor-like microdeletions in FHIT/FRA3B. *Proc Natl Acad Sci U S A*,
1462 **105**, 246-251.
1463 167. Chen, L., Zhou, W., Zhang, C., Lupski, J.R., Jin, L. and Zhang, F. (2015) CNV instability associated
1464 with DNA replication dynamics: evidence for replicative mechanisms in CNV mutagenesis. *Hum*
1465 *Mol Genet*, **24**, 1574-1583.
1466 168. Craven, S.H. and Neidle, E.L. (2007) Double trouble: medical implications of genetic duplication
1467 and amplification in bacteria. *Future Microbiol*, **2**, 309-321.
1468 169. Hjort, K., Nicoloff, H. and Andersson, D.I. (2016) Unstable tandem gene amplification generates
1469 heteroresistance (variation in resistance within a population) to colistin in Salmonella enterica.
1470 *Mol Microbiol*, **102**, 274-289.
1471 170. Anderson, S.E., Sherman, E.X., Weiss, D.S. and Rather, P.N. (2018) Aminoglycoside
1472 Heteroresistance in Acinetobacter baumannii AB5075. *mSphere*, **3**, e00271-00218.

1473 171. Nicoloff, H., Hjort, K., Levin, B.R. and Andersson, D.I. (2019) The high prevalence of antibiotic
1474 heteroresistance in pathogenic bacteria is mainly caused by gene amplification. *Nat Microbiol*, **4**,
1475 504-514.
1476 172. van Dijk, E., van den Bosch, T., Lenos, K.J., El Makrini, K., Nijman, L.E., van Essen, H.F.B., Lansu,
1477 N., Boekhout, M., Hageman, J.H., Fitzgerald, R.C. *et al.* (2021) Chromosomal copy number
1478 heterogeneity predicts survival rates across cancers. *Nat Commun*, **12**, 3188.
1479 173. Sobral, D., Martins, M., Kaplan, S., Golkaram, M., Salmans, M., Khan, N., Vijayaraghavan, R.,
1480 Casimiro, S., Fernandes, A., Borralho, P. *et al.* (2022) Genetic and microenvironmental intra-
1481 tumor heterogeneity impacts colorectal cancer evolution and metastatic development. *Commun*
1482 *Biol*, **5**, 937.
1483 174. White, N.J. (2017) Malaria parasite clearance. *Malar J*, **16**, 88.
1484 175. Shor, E., Fox, C.A. and Broach, J.R. (2013) The yeast environmental stress response regulates
1485 mutagenesis induced by proteotoxic stress. *PLoS Genet*, **9**, e1003680.
1486 176. WHO. (2023) World Malaria Report 2023.
1487 177. Okombo, J. and Fidock, D.A. (2025) Towards next-generation treatment options to combat
1488 Plasmodium falciparum malaria. *Nat Rev Microbiol*, **23**, 178-191.
1489 178. Wilson, C., Serrano, A., Wasley, A., Bogenschutz, M., Shankar, A. and Wirth, D. (1989)
1490 Amplification of a gene related to mammalian mdr genes in drug-resistant Plasmodium
1491 falciparum. *Science*, **244**, 1184-1186.
1492 179. Nair, S., Nash, D., Sudimack, D., Jaidee, A., Barends, M., Uhlemann, A.-C., Krishna, S., Nosten, F.
1493 and Anderson, T.J.C. (2007) Recurrent Gene Amplification and Soft Selective Sweeps during
1494 Evolution of Multidrug Resistance in Malaria Parasites. *Mol Biol Evol*, **24**, 562-573.
1495 180. Ravenhall, M., Benavente, E.D., Mipando, M., Jensen, A.T., Sutherland, C.J., Roper, C.,
1496 Sepulveda, N., Kwiatkowski, D.P., Montgomery, J., Phiri, K.S. *et al.* (2016) Characterizing the
1497 impact of sustained sulfadoxine/pyrimethamine use upon the Plasmodium falciparum
1498 population in Malawi. *Malar J*, **15**, 575.
1499 181. Leroy, D., Macintyre, F., Adoke, Y., Ouoba, S., Barry, A., Mombo-Ngoma, G., Ndong Ngomo, J.M.,
1500 Varo, R., Dossou, Y., Tshefu, A.K. *et al.* (2019) African isolates show a high proportion of multiple
1501 copies of the Plasmodium falciparum plasmepsin-2 gene, a piperaquine resistance marker.
1502 *Malar J*, **18**, 126.
1503 182. Bopp, S., Magistrado, P., Wong, W., Schaffner, S.F., Mukherjee, A., Lim, P., Dhorda, M.,
1504 Amaratunga, C., Woodrow, C.J., Ashley, E.A. *et al.* (2018) Plasmepsin II-III copy number accounts
1505 for bimodal piperaquine resistance among Cambodian Plasmodium falciparum. *Nat Commun*, **9**,
1506 1769.
1507 183. Thaithong, S., Ranford-Cartwright, L.C., Siripoon, N., Harnyuttanakorn, P., Kanchanakhan, N.S.,
1508 Seugorn, A., Rungsihirunrat, K., Cravo, P.V. and Beale, G.H. (2001) Plasmodium falciparum: gene
1509 mutations and amplification of dihydrofolate reductase genes in parasites grown in vitro in
1510 presence of pyrimethamine. *Exp Parasitol*, **98**, 59-70.
1511 184. Istvan, E.S., Dharia, N.V., Bopp, S.E., Gluzman, I., Winzeler, E.A. and Goldberg, D.E. (2011)
1512 Validation of isoleucine utilization targets in Plasmodium falciparum. *Proc Natl Acad Sci U S A*,
1513 **108**, 1627-1632.
1514 185. Rathod, P.K., McErlean, T. and Lee, P.C. (1997) Variations in frequencies of drug resistance in
1515 Plasmodium falciparum. *Proc Natl Acad Sci U S A*, **94**, 9389-9393.
1516 186. Olawade, D.B., Wada, O.Z., Ezeagu, C.N., Aderinto, N., Balogun, M.A., Asaolu, F.T. and David-
1517 Olawade, A.C. (2024) Malaria vaccination in Africa: A mini-review of challenges and
1518 opportunities. *Medicine (Baltimore)*, **103**, e38565.
1519 187. Theodoridis, L. and Carvalho, T.G. (2025) Antimalarial drug resistance and drug discovery:
1520 learning from the past to innovate the future. *Int J Parasitol Drugs Drug Resist*, **28**, 100602.

1521 188. Bell, G. and MacLean, C. (2018) The Search for 'Evolution-Proof' Antibiotics. *Trends Microbiol*,
1522 **26**, 471-483.
1523 189. Zhang, L., Ma, J., Liu, L., Li, G., Li, H., Hao, Y., Zhang, X., Ma, X., Chen, Y., Wu, J. *et al.* (2023)
1524 Adaptive therapy: a tumor therapy strategy based on Darwinian evolution theory. *Crit Rev Oncol*
1525 *Hematol*, **192**, 104192.
1526 190. Dharanipragada, P., Zhang, X., Liu, S., Lomeli, S.H., Hong, A., Wang, Y., Yang, Z., Lo, K.Z., Vega-
1527 Crespo, A., Ribas, A. *et al.* (2023) Blocking Genomic Instability Prevents Acquired Resistance to
1528 MAPK Inhibitor Therapy in Melanoma. *Cancer Discov*, **13**, 880-909.
1529 191. Li, G.Q., Arnold, K., Guo, X.B., Jian, H.X. and Fu, L.C. (1984) Randomised comparative study of
1530 mefloquine, qinghaosu, and pyrimethamine-sulfadoxine in patients with falciparum malaria.
1531 *Lancet*, **2**, 1360-1361.
1532 192. Siddiqui, F.A., Liang, X. and Cui, L. (2021) Plasmodium falciparum resistance to ACTs: Emergence,
1533 mechanisms, and outlook. *Int J Parasitol Drugs Drug Resist*, **16**, 102-118.
1534