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

Authors: Noah Brown[1], Aleksander Luniewski[1], Xuanxuan Yu[2,3], Michelle Warthan[1], Shiwei Liu[1,4], Julia Zulawinska[1], Syed Ahmad[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]Unifersity of Florida, Department of Biostatistics, Gainesville, FL, USA

[3]Unifersity 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 in many organisms. CNVs are readily identified using conventional approaches when present in a large fraction of the cell population. However, CNVs that are present in only a few genomes across a population are often overlooked but important; if beneficial under specific conditions, a de novo CNV that arises in a single genome can expand during selection to create a larger population of cells with novel characteristics. While the reach of single cell methods to study de novo CNVs is 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 to new drugs and host environments. We employed a low-input genomics approach optimized for this unique genome as well as specialized computational tools to evaluate the de novo CNV rate both before and after the application of stress. We observed a significant increase in genome-wide de novo CNVs following treatment with a replication inhibitor. These stress-induced de novo CNVs encompassed genes that contribute to various cellular pathways and tended to be altered in clinical parasite genomes. 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]). The structure of CNVs and their formation is complex; they can involve a few base pairs or whole genes, encompass many structural forms (e.g. tandem arrays, inversions, translocations), and be generated by a range of DNA repair pathways (e.g. recombination and end-joining pathways). CNVs are especially important for rapidly evolving microbes such as bacteria, yeast, and viruses by contributing to antimicrobial resistance, nutrient acquisition, and pathogenesis [4-6]. CNVs also contribute to cancer growth and progression ([7-10] and reviewed in [11, 12]). Further, we are beginning to appreciate CNVs as drivers of other human disorders and disease susceptibility including blood, metabolic, neurological, and infectious diseases [13-21].

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 genomes, or even a few genomes, are “averaged” away during analysis steps. This artifact limits our ability to assess how CNVs arise and contribute to the genomic diversity of individual cells within a population. Such diversity is important; if beneficial under specific conditions, a CNV that arises in a single genome can expand during selection into a larger population of cells with novel characteristics [22, 23]. This rapid expansion is exemplified when minor bacterial populations with higher gene copy numbers confer “heteroresistance” during clinical antibiotic selection [24, 25, 26]. In another example, higher levels of intra-tumor heterogeneity in gene copy number predict a poorer cancer prognosis [27-29].

In order to observe a genome's evolutionary potential in the absence of selection, we require approaches specifically designed to detect CNVs that are not present in a predecessor "parental" genome. Due to their rare and novel nature, these events are commonly termed "de novo" CNVs [3, 30-34]. Early experimental progress detecting de novo CNVs involved cloning individual cells, which takes time, is prone to contamination, and prevents detection of detrimental CNVs [30-32]. Misalignment of short-reads to reference genomes (i.e. discordant or split reads) can also indicate the presence of de novo CNVs, but false positives are common if matched normal samples are not available (reviewed in [35]).

Recent methods that isolate single cells have been successful at assessing de novo CNVs during experimental evolution, disease progression, and tissue development (as reviewed in [35-39]). De novo CNVs have been tracked in evolving yeast genomes using flow cytometry to quantify fluorescent reporters integrated into a specific selectable locus [33]. This approach is sensitive but provides a limited view of CNV dynamics by focusing on a single or few specific loci. Single cell transcriptomics, which infers gene copy number using mRNA abundance, can identify large structural changes across genomes from heterogeneous tumor samples [40-43]. This approach averages relative expression over Mb-sized genomic regions and thus, is not applicable to identify smaller de novo CNVs. Single cell genomics, where individual genomes or nuclei are isolated and amplified to a level that can be sequenced, has been used to directly quantify de novo CNV rates in brain tissue and cancer cells [27, 44-49]. A recent approach, termed direct library preparation, is free from amplification steps, thus limiting genome skewing; however, this method is less accessible due to the requirement for specialized cell dispensers (microfluidics or piezoelectric) and has not been tested with CNVs smaller than 500kb [50]. While the reach of single cell methods is expanding, we continue to lack information about de novo CNVs in microbes and their dynamics in evolving populations.

The *Plasmodium* parasite that causes malaria readily accumulates kb-sized CNVs in its genome [51-54]. CNVs impact the survival of malaria, allowing this parasite to evade clinical detection [55], expand beneficial gene families [56], invade new host cells [57], and develop clinical antimalarial resistance [58-60]. We are specifically interested in the genetic diversity of one species of malaria, *P. falciparum*, since this may explain its rapid adaptation to new drugs and host environments [61]. Due to its relatively small, AT-rich genome (23 Mb, 19.4% GC [62]), low-input genomics is challenging in this single cell protozoan. However, we previously optimized a single cell genomics approach and developed novel computational tools to evaluate de novo CNVs in the *P. falciparum* genome [63] (see *Materials and Methods* for tool details).

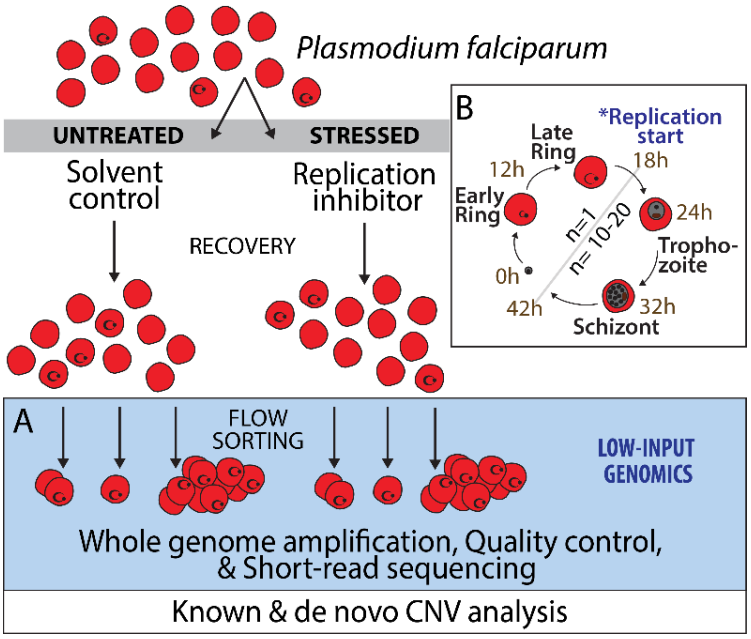
Here, we combined these advancements and made additional improvements to investigate de novo CNV formation in the *P. falciparum* genome. Since various types of cellular stress can induce genetic change (reviewed in [2, 64-66]), we also evaluated the impacts of replication stress on de novo CNV identification. Using our low-input genomics pipeline, we observed that replication stress increased the number of de novo CNVs across the parasite genome. This study of genome dynamics, along with improved tools, increases our understanding of how stress can stimulate rapid microbial evolution.

RESULTS

Refined low-input genomics pipeline increased efficiency and accuracy

99 We developed a robust pipeline for assessing the frequency of de novo CNVs in the *Plasmodium* genome
100 based on our prior studies (**Fig. 1**). We adapted our single cell genomics method to improve parasite
101 isolation and whole genome amplification steps ([63], **Fig. 1A, Fig. S1**). In this modified protocol, we
102 used fluorescence-activated cell sorting (FACS) to isolate parasites to improve efficiency and modified
103 aspects of our whole genome amplification method to improve coverage (**Table S1**, **Fig. S1**). We also
104 sorted low-cell populations to increase accuracy (e.g. 2-cells per well, **Fig. S2**), and added quality control
105 steps that confirmed our samples were of high quality prior to short-read sequencing. The resulting low-
106 input genomics pipeline consisted of basic steps including parasite isolation using flow sorting, whole
107 genome amplification using a modified MALBAC-based approach, quality control confirmation, short-
108 read sequencing, and CNV analysis (**Fig. 1A**).

109



110

111 **Figure 1. Low-input genomics approach for analysis of malaria parasite genomes under stress**. Early erythrocytic stage *P.*
112 *falciparum* parasites grown in vitro were treated with a replication inhibitor (DSM1) or the solvent control (dimethylsulfoxide)
113 followed by recovery and reinvasion to produce a new round of ring stage parasites. **A**. For low-input genomics, parasite-
114 infected erythrocytes were isolated using flow sorting (10-cell control or 2-cell samples). Parasite genomes were amplified using
115 a modified MALBAC-based whole genome amplification approach (**Fig. S1**). Quality control steps involved DNA quantification to
116 assess amplification success, droplet digital PCR to assess parasite genome amplification, and PCR-high resolution melting to
117 assess sample cross-contamination. After Illumina short-read sequencing, reads are filtered, trimmed and used as input for
118 copy number variation analysis using HapCNV and LUMPY. **B.** Parasite erythrocytic life cycle including approximate timing (h),
119 genome number (n), and effect of treatment.

120

121 **Replication stress followed by a recovery period led to isolation of healthy parasites**

122 To explore the impact of replication stress on de novo CNV generation, we treated the parasites with an
123 antimalarial that inhibits DNA replication by limiting pyrimidine pools; DSM1 targets *Plasmodium* the
124 dihydroorotate dehydrogenase enzyme [67]. Application of DSM1 for an extended period kills parasites
125 by perturbing DNA synthesis (>48hrs at 10x the EC_{50}, [68]). However, short-term treatment reversibly

126 stalls parasite replication similar to another replication inhibitor, aphidicolin (**Fig. S3**). For low-input
127 genomics, we applied DSM1 to ring-stage parasites for a brief time (12hr, **Fig. S4A and B**). Similar to our
128 pilot experiment (**Fig. S3A**), we observed that treated parasites stalled prior to replication (**Fig. S4C and**
129 **E**) and slightly decreased their growth rate compared to untreated parasites (**Fig. S4G**). We harvested
130 viable parasites after a recovery period (~30hrs, **Fig. S4H**), where we allowed parasites to complete an
131 additional round of replication and erythrocyte invasion to produce those that have a single, haploid
132 genome (rings, **Fig. S4D and F**). The recovery stage was essential as it allowed parasites to repair the
133 DNA damage accumulated during treatment. As a control for cross-sample contamination between
134 isolation wells, we isolated and amplified untreated parasites with different genetic backgrounds (*FCR3*
135 vs *Dd2*, **Table 1, Fig. S5**).

136 Prior to isolation of low-cell populations, we confirmed that untreated and DSM1-treated parasites were
137 at a similar life cycle stage and viability using staining for mitochondrial membrane potential (**Table 1,**
138 **Fig. S5A**); we saved a portion of these samples for parasite population sequencing (i.e. bulk samples).
139 We then proceeded to isolate small populations of viable, ring stage parasites using FACS (**Fig. S5B**) for
140 whole genome amplification.

141

142 **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%

143 *Treatment conditions: 1µM DSM1 (~10x EC50) for 12hrs prior to 30.8hrs of recovery. Untreated samples were incubated with
144 DMSO as a solvent control for 12hrs and allowed to recover for 28-30hrs (**Fig. S4H**). $Parasitemia was determined by calculating
145 the number of infected erythrocytes (SYBR Green+) compared to uninfected erythrocytes (SYBR Green-) (**Fig. S4**).@Viability of
146 parasites was determined by measuring the mitochondrial membrane potential (Mitoprobe+) and calculating % of total SYBR
147 Green-based parasitemia (plots presented in **Fig. S4D, S4F,** and **Fig. S5**). ^*FCR3* (Africa) and *Dd2* (Southeast Asia) parasite lines
148 are from distinct geographic origins [69, 70].
149

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

151 We sorted low-cell populations from each parasite group into 60 wells of a 96-well plate (*FCR3*,
152 untreated *Dd2,* and treated *Dd2*), including 6 wells with 10-cells and 56 wells with 2-cells. Ten-cell wells
153 served as positive controls for the whole genome amplification step and 2-cell wells provided the
154 optimal balance between sorting accuracy (**Fig. S2**) and de novo CNV detection. Zero cells were sorted
155 into the top 2 rows of the plate ("no-cell" wells). After parasite lysis and applying *Pf*MALBAC version-2
156 whole genome amplification (**Fig. S1**), we assessed amplification success using three approaches. First,
157 we measured the resulting DNA quantity across 80% of the amplified wells (**Fig. S6**). On average,
158 MALBAC amplification in wells that contained sorted parasites yielded ~120ng of total DNA per reaction,
159 with a ~10% increase in DNA for 10- vs 2-cell samples (mean of 127ng versus 116ng total, respectively).
160 We did not detect position-based bias across plates or appreciable amplification from no-cell wells, but
161 we did observe that treated *Dd2* wells had ~3-fold lower levels of amplification than other samples

162 (mean of 51ng versus 151ng total, respectively). There was little difference in mean amplified DNA
163 amounts between the two untreated sample groups (*FCR3* at 146ng and untreated-*Dd2* at 152ng).

164 Second, we performed droplet digital PCR (ddPCR) for parasite-specific genes on approximately one-
165 third of wells post isolation to confirm the amplification of the parasite genome. DdPCR for *pfmdr1* and
166 *pfhsp70* displayed that wells with measurable DNA contained amplified parasite DNA (**Fig. S6** and **S7**).
167 Additionally, we confirmed that 2-cell wells with very low total DNA amounts (**Fig. S6**) were positive for
168 parasite genomes while "no-cell" wells did not show evidence of parasite material (**Fig. S7**).

169 Finally, we employed high-resolution melting analysis to profile a drug resistance marker that differs
170 between *FCR3* and *Dd2* parasites. By assessing the *pfdhps* SNP profile of amplified genomes and
171 comparing it to the parental profile in ~10% of samples, we confirmed that there was no evidence of
172 cross-sample contamination during the preparation and amplification steps (**Fig. S8**). Therefore, we
173 proceeded to sequence the amplified bulk and low-input samples (**Table S2**).

174

175 **Coverage deviation and SNP profiles exhibited expected trends in low-input samples**

176 We sequenced 3 bulk samples and 90 low-input samples using Illumina short-read sequencing (**Table 2**
177 and **Table S3**). Overall, sequencing proceeded well as indicated by coverage and coverage deviation of
178 the bulk samples, as well as an equivalent mean mapping quality across all samples (**Table 2**). Because
179 we noticed that treated *Dd2* wells had lower levels of DNA following amplification (**Fig. S6**), we
180 sequenced higher amounts of material for this condition; this choice impacted mean coverage levels
181 where treated *Dd2* samples had ~4-times higher coverage than untreated *Dd2* samples (**Table 1**). The
182 percent of total reads that mapped to the *P. falciparum* genome was high across all samples (mean of
183 67%), indicating efficient amplification of the parasite genome and little contribution of environmental
184 contamination during sample amplification. As expected based on previous studies [63], coverage
185 deviation was ~3-fold higher in low-input samples when compared to bulk samples, reflecting the bias of
186 the whole genome amplification step to over- or under-amplify specific genomic regions.

187 We removed five low-input samples from further analysis based on low coverage levels; on average,
188 excluded samples had ~7-times lower coverage than other low-input samples (**Table S3**). Of the
189 remaining samples, 18 were 10-cell samples and 57 were 2-cell samples. Although mean coverage was
190 ~2-fold higher for 2-cell samples (due to the higher coverage of treated low-input samples), mean
191 normalized deviation was similar between 10- and 2-cell samples (3.5 and 3.1, respectively).

192 To evaluate the quality of the sequencing data, we tracked SNPs in the low-input samples compared to
193 bulk samples. Despite some variation due to the non-clonal nature of parasite lines (see *Materials and*
194 *Methods*), low-input SNP profiles were similar to their corresponding bulk sample (**Fig. S9A**). After
195 normalizing total SNPs to mapped reads, we detected a lower rate of SNPs in treated samples compared
196 to untreated counterparts (p value of 0.0001, **Fig. S9B**). We did not detect a correlation between
197 normalized total SNPs and amplification quality (**Fig. S9C**, R_2: untreated *Dd2*, 0.01; treated *Dd2,* 0.01);
198 we did observe a positive correlation between SNP number and coverage depth in treated versus
199 untreated *Dd2* samples (**Fig. S9D**, R_2: untreated *Dd2*, 0.66; treated *Dd2,* 0.27), indicating that the
200 difference in SNP numbers is likely due to varying sensitivity at different levels of read coverage [71, 72].

201

202
203
204 **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

205 *On average, ~4x more material was loaded on the flow cell for the treated samples than the untreated samples due to lower
206 initial amplification yields (**Fig. S6**).
207 **Excludes samples that were removed due to low coverage. For low-input samples, includes both 10- and 2-cell samples.
208 [#]CV is the coefficient of variation of normalized read abundance as in [63]
209
210

211 **Experimental and computational advances improved known CNV calls across low-input samples**

212 For the current study, we employed two different CNV calling methods in low-input samples. HapCNV is
213 a novel read coverage-based CNV calling method specifically designed for low cell data from haploid
214 genomes [73]. In contrast to traditional methods that arbitrarily select reference samples for CNV data
215 normalization, HapCNV constructs a genomic location (or bin)-based pseudo-reference as a comparison
216 baseline. This step systematically alleviates amplification bias for the identification of de novo CNVs.
217 LUMPY is an established CNV calling method that exhibits high sensitivity due to the incorporation of
218 multiple CNV signals (i.e. split and discordant reads) generated from short-read sequencing. It is
219 particularly well-suited for detecting low-frequency variants in low-coverage datasets; however, for
220 many CNV callers, high sensitivity leads to higher false positives [74-76].

221 Using these two distinct CNV calling methods, combined with a recently developed CNV counting
222 approach, we evaluated the presence of known CNVs in our 2-cell samples (**Fig. 2A**). The identification of
223 known CNVs (i.e. those identified in the bulk sample, see *Materials and Methods*) within low-input
224 samples displays the utility of the specific CNV calling method for different size CNVs in specific genome
225 locations. In our previous study, we identified 2 of the 3 known CNVs in ~10% of single cell genomes
226 [63]. In the current study, we identified the *pfmdr1* amplicon in 100% of 2-cell samples using HapCNV
227 (57/57) and 79% of samples using LUMPY (45/57). We did not identify the *pf11-1* amplicon in any 2-cell
228 samples using HapCNV (0/57) but detected this locus in 75% of samples using LUMPY (43/57). Finally,
229 we identified *pf332* CNVs in 46% of 2-cell samples using HapCNV (26/57) and 100% of samples using
230 LUMPY (57/57). Although our two studies used different CNV calling methods and are not directly

231 comparable, the overall improvement in the detection of known CNVs in the current study is likely due
232 to advances in both the whole genome amplification method to limit amplification bias (**Table S1**) and
233 recently developed analyses approaches. When we evaluated the detection of three known CNVs in *Dd2*
234 low-input samples, we observed a somewhat higher rate of known CNV detection by either method in
235 treated samples (**Fig. 2B**, HapCNV: mean of 1.2 out of 3 total CNVs for untreated and 1.7 for treated
236 (increase of 42%), LUMPY: mean of 2.1 out of 3 total CNVs for untreated and 2.9 for treated (increase of
237 38%, **Table S4**).

238 **De novo CNVs in low-input samples consisted of rare and common CNVs**

239 We next sought to quantify de novo CNVs in low-input samples. We defined de novo CNVs as those not
240 present in the bulk sample and we categorized them based on their frequency in low cell samples.
241 "Common" CNVs were present in a larger number of genomes (≥10% of the same sample type, i.e.
242 untreated or treated), and "rare" CNVs were those that occurred in a small proportion of samples (<10%
243 of the sample type) (**Fig. 2A**). Overall, LUMPY detected more total de novo CNVs than HapCNV across all
244 samples (12-fold), and the proportion of rare versus common CNVs varied depending on the method
245 (18% vs 82% for HapCNV, 61% vs 39% for LUMPY, respectively, **Table S4**). Additionally, de novo CNVs
246 were more often identified as duplications than deletions for both CNV calling methods (**Fig. 2C**).

247 To understand the nature of common and rare CNVs, we also assessed how often their locations
248 overlapped across the two sample types (untreated and treated, **Fig. S10**). This analysis is useful for
249 tracking common/rare category utility and relevance. For utility, this step acts as a sanity check since, by
250 definition, we do not expect rare CNV locations to overlap as often as common CNVs. For relevance, de
251 novo CNVs with conserved locations across sample types are less likely to represent true CNVs newly
252 arising in a genome. As expected, we identified many common CNVs with conserved locations (22% for
253 HapCNV and 52% for LUMPY, **Fig. S10A**). The lower rate of overlapping calls across samples for HapCNV
254 is likely due to the bin-based normalization strategy to remove amplification artifacts [73]. Conversely,
255 we found that rare CNVs were predominantly called in unique genome locations (94% for HapCNV and
256 97% for LUMPY, **Fig. S10A**), supporting their novel nature. This pattern was consistent when we
257 randomly down-sampled all sequencing data to the lowest read coverage prior to CNV calling (1.3
258 million reads, **Fig. S10B**). This comparison not only highlights the suitability of the common and rare
259 categories but also the difference between the CNV calling methods. Based on these observations, for
260 the following analysis, we assessed common and rare CNVs using both methods to capture the broadest
261 view of stress effects on CNV generation.

262 **Genome-wide de novo CNVs increased following replication stress**

263 When we compared de novo CNVs in genomes with and without replication stress, we found that results
264 were consistent regardless of the CNV calling method (**Fig. 2**). While the proportion of duplication and
265 deletions did not change appreciably with treatment (**Fig. 2C**), we identified a highly significant
266 difference in de novo CNVs between treated and untreated 2-cell samples (p value of 0.0002 for
267 HapCNV and <0.0001 for LUMPY, **Fig. 2D**). This pattern was consistent when we down-sampled all
268 sequencing data (p value of 0.001 for HapCNV and <0.0001 for LUMPY, **Fig. 2E**), indicating that the
269 difference in de novo CNV counts between treatments was not due to read coverage. When we
270 assessed common and rare CNV categories, we once again observed highly significant differences
271 between treated and untreated 2-cell samples in common CNVs using both methods (common: p value

272 0.0004 for HapCNV and <0.0001 for LUMPY, **Fig. 2F;** rare: p value of 0.008 for HapCNV and <0.0001 for
273 LUMPY, **Fig. 2G**).

274 When we compared the proportion of de novo CNVs relative to total CNVs per sample, rare CNVs were
275 significantly increased over common CNVs (p value of 0.003 for HapCNV and 0.009 for LUMPY, **Fig. 2H**).
276 This difference persisted regardless of down-sampling (p value of 0.02 for HapCNV and 0.004 for LUMPY,
277 **Fig. S11**) and is in line with our assessment above that rare CNVs are more likely to be novel in nature
278 (**Fig. S10**). Overall, we detected a ~2-3-fold increase of de novo CNVs in treated samples, regardless of
279 the CNV calling method (**Table 3**). Once again, rare CNVs displayed the largest increase following
280 treatment (~3-4-fold, **Table 3, Fig. 2I**).

281 **De novo CNVs represented diverse cellular pathways with clinical benefits**

282 When we compared overlaps between the HapCNV and LUMPY (**Fig. 3A**), we detected a set of CNV
283 regions that was consistent within sample groups (5 for untreated and 38 for treated, **Table S5**). The
284 frequency of these "high-confidence" CNVs also reflected the increase in CNVs following replication
285 stress (with knowns excluded, ~15-fold increase in treated samples). Overall, high-confidence CNVs were
286 located on the majority of chromosomes (12 of 14, **Fig. 3B**) and represented both duplications and
287 deletions (**Fig. 3C**). Of note, approximately half of these regions were identified as "rare" by both CNV
288 calling methods across treated samples (17/36, 47%; **Table S5**), indicating that novel CNVs were
289 stimulated in parasite genomes under stress. When we searched for genes that are covered by these
290 regions, we identified 26 genes (across 3 de novo CNV regions) and 198 genes (across 37 de novo CNV
291 regions) in untreated and treated *Dd2* samples, respectively (**Table S5**). Emphasizing their random
292 nature, genes encompassed by the CNV regions represented diverse protein classes (**Fig. 3D**) and no
293 gene ontology (GO) categories were significantly enriched (using an FDR adjusted p value of 0.05, **Table
294 S6**).

295 Finally, we evaluated whether CNVs that arose under stress were important for parasite survival by
296 comparing genes covered by our treated CNV regions to those from the largest catalogue of clinically
297 relevant CNVs to date. This list of high frequency CNVs was previously called using genomes from 2855
298 parasite isolates from 21 malaria-endemic countries and represented genes from larger (>300bp), high-
299 quality, core genome variants [59]. Assuming ~5000 genes in the core *P. falciparum* genome [77], we
300 found that genes from the two lists were ~2-times more likely to overlap than by chance (chi-square
301 odds ratio of 2.5 for HapCNV-treated CNV list and 1.6 for LUMPY-treated CNV list). This significant
302 association between CNV locations suggests that stress-induced de novo CNVs have the potential to be
303 beneficial in the clinical environment (Fisher's exact p value <0.0001 for HapCNV-treated CNV list and
304 0.03 for LUMPY-treated CNV list).

305

306 **Table 3. Mean CNV counts per 2-cell sample using two analysis methods.**

CNV detection method$	Condition (2-cell only)	Rare CNVs per sample*	Common CNVs per sample*	Combined de novo CNVs per sample^

HapCNV	treated *Dd2*	7	23	15
	untreated *Dd2*	1.7	15	8
	fold change	**4.4**	**1.6**	**1.9**
LUMPY	treated *Dd2*	297	163	230
	untreated *Dd2*	99	68	84
	fold change	**3.0**	**2.4**	**2.7**

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

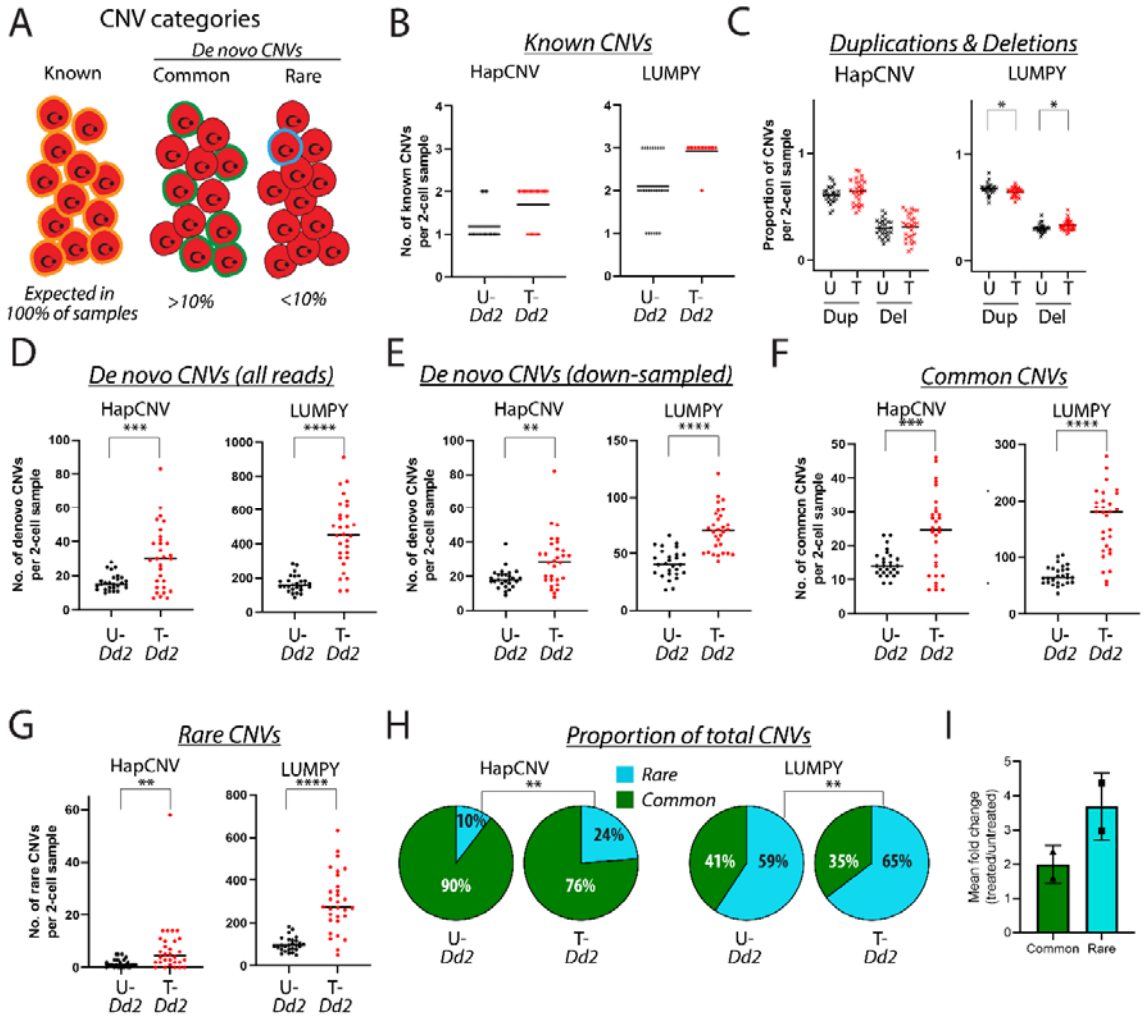


Figure 2. Low cell genomics displays an increase in de novo CNVs following replication stress. Number of CNVs from untreated (U-*Dd2*) and treated (T-*Dd2*) 2-cell samples from two CNV analysis methods: HapCNV and LUMPY. Statistics for all plots use an unpaired T-test with two tailed Welch's correction (****:p value <0.0001; ***: <0.001; **: <0.01; *:<0.05; no stars: not significant). Analyses include all reads, unless otherwise indicated (i.e. panel E is down-sampled). 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 de novo CNVs (common and rare combined) from down-sampled reads (p values: 0.0014 for HAPCNV and <0.0001 for LUMPY). **F**. Detection of common CNVs from all reads (p values: 0.0004 for HapCNV, <0.0001 for LUMPY). **G**. Detection of rare CNVs from all reads (p values: 0.008 for HapCNV, <0.0001 for LUMPY). **H**. Proportion of total CNVs detected as rare and common from all reads; pie charts plot 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**) I. Mean fold change between rare and common CNVs detected by HapCNV and LUMPY in untreated and treated *Dd2* 2-cell samples (see **Table 3**).

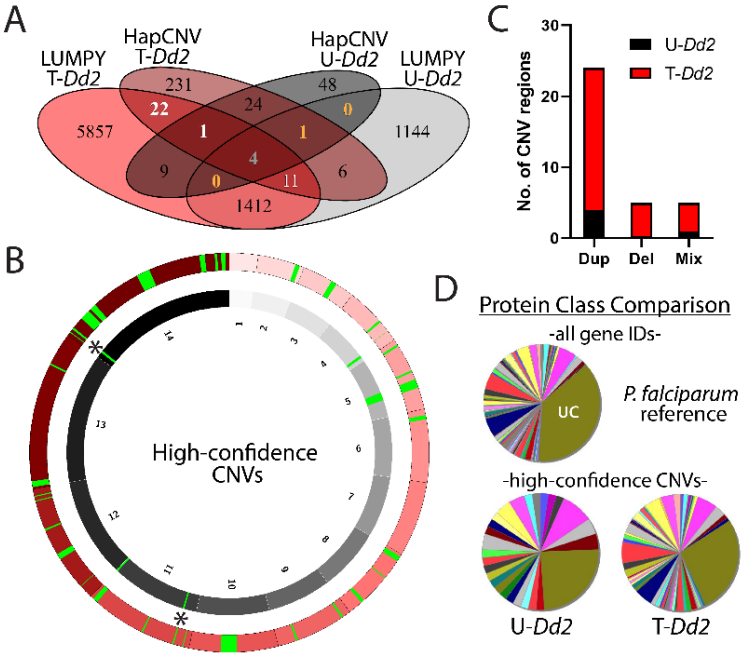


Figure 3: High-confidence CNVs are located across the genome and represent diverse protein classes. A. Comparison of CNV calls showing the number of CNV regions consistent across the two CNV calling methods. High-confidence CNVs in untreated samples (U-*Dd2*, in yellow text); high-confidence CNVs in treated samples (T-*Dd2*, in white text). Central number (grey): CNVs consistent across all samples and calling methods (includes 2 known CNVs and 2 de novo CNVs, **Table S5**). **B**. Chromosomal location of high-confidence CNVs 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 by a factor of 2 to facilitate visualization relative to the rest of the genome. *, de novo high-confidence CNVs identified in both untreated and treated samples. **C**. Summary of duplications and deletions represented in high-confidence CNV list. U-*Dd2*, untreated; T-*Dd2*, treated. Mix: sub-regions were called as duplications and deletions across a single CNV region by the same CNV calling method (i.e. HapCNV or LUMPY). **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.

Discussion:

346 De novo CNVs are not detected when analyzing a population of parasites predominantly because their
347 signal (e.g. extra reads that align to that region or reads that span breakpoints) is negated by the
348 overwhelming signal from normal copy number at that genome location. For this reason, assessments of
349 fewer cells are necessary to investigate de novo CNV generation. Here, we optimized a low-input
350 analysis pipeline and successfully isolated, amplified, and sequenced *P. falciparum* samples. With
351 experimental and computational improvements, we were able to increase our rate of parental, or
352 "known", CNV calling over our prior study [63]. Importantly, we detected non-parental, or "de novo",
353 CNVs across the parasite genome and replication stress significantly increased their rate formation. By
354 analyzing ~45 low-input samples per condition, we are limited to observing a subset of the population.
355 However, our findings demonstrate that replication stress readily drives the rapid generation of de novo
356 CNVs. Below we cover how this study contributes to our understanding of genome evolution and
357 integrate it with an overarching model of *P. falciparum* adaptation.

358 *Application of sub-lethal stress without evidence of selection to explore CNV dynamics*

359 To evaluate the impact of replication stress on de novo CNV generation, we applied sub-lethal treatment
360 to parasites just prior to replication. We observed replication stall and then resume post-treatment,
361 which provided evidence that we successfully applied non-lethal stress (**Fig. S3 and S4**). Following this
362 step, we allowed the parasites to complete replication and reinvade new erythrocytes. We reasoned
363 that this "recovery phase" enabled the repair of the resulting DNA damage, which is likely to be
364 replication-dependent (reviewed in [1, 78]. Additionally, reinvasion facilitated the isolation of haploid
365 parasite genomes (1n, **Fig. 1B** and **S4B**), encouraging the detection of de novo CNVs due to limited
366 contrasting signal [36]. Because of the reinvasion step, which involves an expansion in parasite number
367 (~3-fold, **Fig. S4G**), there was a potential to select for beneficial DNA changes across the population of
368 parasites. However, we did not detect evidence of strong selection from SNP profiles (**Fig. S9A**) or high-
369 confidence CNV regions (**Fig. 3**, **Table S5**). Specifically, we did not observe a preference for CNVs that
370 encompassed the dihydroorotate dehydrogenase gene (*pfdhodh*), which contributes directly to DSM1
371 resistance [79], or enrichment of CNVs that include genes from DNA-related pathways (**Table S6**).

372 *Relative comparison using multiple CNV calling methods to appreciate the impact of stress*

373 De novo CNV estimates using single cell methods from neurons, yeast, and human cancer vary greatly
374 and are difficult to standardize due to the use of different experimental techniques and CNV calling
375 methods [33, 34, 80]. For these reasons, we are not attempting to compare the rate of *P. falciparum*
376 CNV formation from this study to those from other organisms. Additionally, this lack of standardization
377 in the field led us to use multiple CNV calling methods in our analysis. Due to the strengths and
378 weaknesses of HapCNV and LUMPY, we observed differences in both known (**Fig. 2B**) and de novo CNV
379 (**Table S4**) calling using the two methods. LUMPY identifies reads that cover breakpoint regions to
380 sensitively detect CNVs [74]; because we are counting regions with few reads as support, both sensitivity
381 and the number of false positives are high in this analysis. High known calling rates along with high
382 numbers of de novo CNVs in our studies exemplified this feature of LUMPY. On the other hand, HapCNV
383 uses a genome-specific pseudo-reference for normalization, which removes repeated patterns of over-
384 and under-amplification ([73], and *Materials and Methods*); because we require read coverage to span
385 3 consecutive 1kb bins, small CNVs in lower coverage genomic neighborhoods are excluded in HapCNV
386 anaylsis. This limited the detection of smaller known amplicons (*pf11-1* and *pf332*) and led to fewer de
387 novo CNV calls using this method. Given the high abundance of small CNVs (<300bp) in the parasite
388 genome [59, 81], HapCNV is likely underestimating their impact in our studies.

389 Ultimately, the value of our study is in the relative comparison of treated and untreated samples by
390 both methods. In this case, we assume that false positive CNVs occur at a similar rate across both
391 groups, which allowed us to confidently assess the impact of stress. Variations in known and de novo
392 CNV calling described above serve to remind us that no CNV calling method is perfect and combining
393 them can improve confidence in results [63, 76, 82, 83]. Therefore, we investigated de novo CNV
394 patterns using the two individual methods (**Figs. 2**, **S10**, **S11**) as well as those that overlapped between
395 HapCNV and LUMPY (**Fig. 3**, **Table S5**). Importantly, these "high-confidence" CNVs reflected increases
396 after stress previously detected using the individual tools, albeit at a greater level (3- vs >10-fold
397 increase). Additionally, we speculate that newly arising CNVs would have distinct locations across
398 samples; thus, the rare nature and unique locations of high-confidence CNVs emphasized their potential
399 to be novel (**Table S5**).

400 *De novo CNV categories highlighting existing and novel genome variation*

401 During our investigations, we identified two types of de novo CNVs; those detected in one or a few
402 genomes (rare) and those detected in more than a few (common). There is no precedence for these CNV
403 categories in the context of *Plasmodium* biology (i.e. a haploid parasite with asynchronous replication
404 and schizogeny [84]). However, we propose that tracking these categories helps us to understand the
405 biological relevance of de novo CNVs in our analysis.

406 Based on their frequency, common CNVs are either artifacts of low-input procedures/CNV analysis or
407 represent minor variants that preexist in the population or arise early in the replication cycle. For the
408 former, bias during the whole genome amplification step (i.e. the repeated pattern of over/under-
409 amplification that occurs in a reproducible pattern across the parasite genome [63]) and PCR during
410 library construction have the potential to skew gene copy number and increase the false positive rate
411 [85, 86]. However, we chose experimental and computational methods designed to limit the
412 contribution of amplification bias. First, MALBAC amplification itself limits the over-amplification of
413 genomic regions by avoiding exponential amplification at the earliest steps [85] and we used limited PCR
414 cycles during library preparation (3 cycles, [63]). These efforts are most clearly shown through the
415 reduction in CV following MALBAC optimization in both of our studies (by ~39% after modifying the
416 amplification primer [63] and by 43% after switching to the *Bsu* polymerase, **Table S1**). Second, LUMPY
417 is not dependent on read coverage and HapCNV specifically addresses amplification artifacts by
418 removing repeated signal present in all samples [73, 74]. Overall, we detected very few CNVs with
419 conserved genomic locations across low-input samples, which provides evidence that our methods limit
420 the effect of amplification bias on the final results; we only identified two high-confidence CNV regions
421 that had conserved locations across multiple *Dd2* and *FCR3* 2-cell samples (**Fig. 3B**, **Table S5**). In the
422 future, single-read visualization of long-reads may offer advantages in distinguishing amplification bias
423 from minor variants and de novo CNVs [87].

424 Rare CNVs, on the other hand, represent either random noise or true signal from novel CNVs arising in
425 the genome. We assert that most noise is removed through normalization procedures, especially with
426 HapCNV, and the impact of remaining false positives are minimized by the relative comparison of our
427 studies (see above). We identified the majority of rare CNVs in unique genome locations across sample
428 types, providing evidence that they are not a result of amplification bias where the same CNVs are
429 repeatedly detected in each sample. Additionally, the greater impact of stress on rare CNVs than
430 common CNVs (**Table 3** and **Figs. 2G, 2I**) supports their replication dependence. The random nature of

431 de novo CNVs, as well as the capacity to encompass any gene across the genome, ensures that CNVs can
432 alter all aspects of parasite biology in response to the host environment. Further, our finding that stress-
433 induced de novo CNVs tended to exhibit altered copy number in clinical isolates combined with the high
434 frequency of unique CNVs in previous genome-wide CNV studies [53, 59], directly illustrate the
435 expansive evolutionary potential of this organism.

436 *Adaptations that encourage de novo CNV formation*

437 The current model of CNV formation in asexual erythrocytic *P. falciparum* parasites is that AT-rich
438 sequences form hairpins, disrupt replication, and eventually lead to double-strand breaks that are
439 repaired by error-prone pathways [61]. The evolution of CNVs in this organism is especially interesting
440 because of its unique genome architecture and alternative repertoire of CNV-generating repair
441 pathways [62, 88]. Although they arise at many locations across the genome ([51, 52, 59], **Fig. 3A**), *P.*
442 *falciparum* CNVs that contribute to adaptation are commonly gene duplications with a relatively simple
443 structure. Many impactful duplications form in tandem head-to-tail orientation ([61, 79, 89, 90], **Fig.**
444 **4A**), which is likely due to a limited repertoire of DNA repair pathways; *P. falciparum* lacks the canonical
445 nonhomologous end-joining pathway that contributes to CNV formation in other organisms [1, 88].
446 Instead, parasites use pathways that employ varying lengths of sequencing homology (i.e. homologous
447 recombination, or HR, and microhomology-mediated repair, **Fig. 4B**). This repair repertoire, along with
448 an especially high AT-content genome that facilitates CNV formation [61, 79] and a lack of cell cycle
449 checkpoints that control replication forks during times of stress (reviewed in [91]), likely represent
450 adaptations that assist haploid *P. falciparum* parasites in accumulating CNVs across their genome (**Fig.**
451 **4A**).

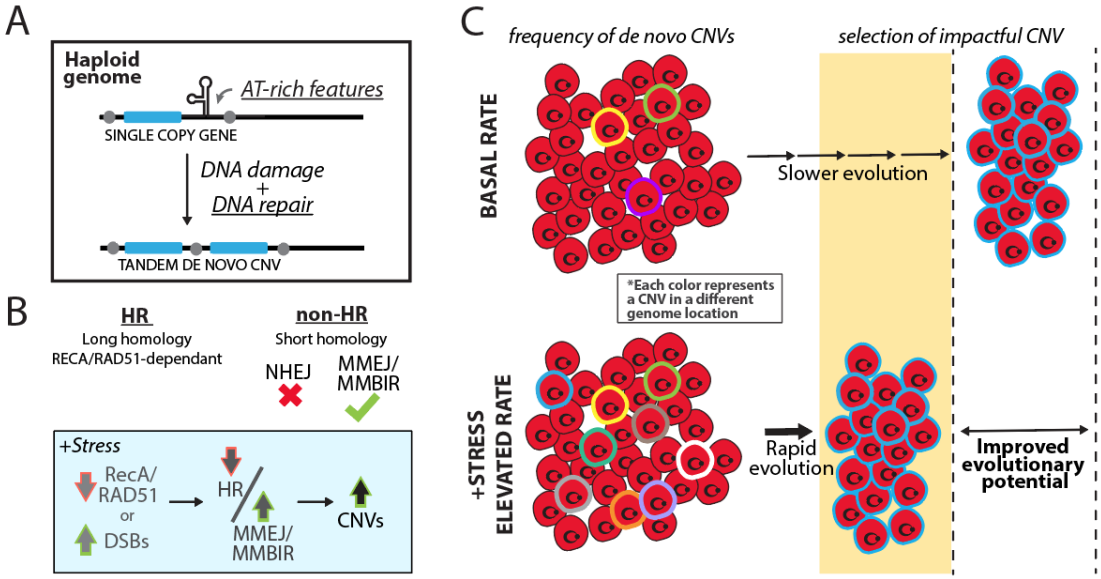
452 *Updating the model of P. falciparum genome adaptation*

453 By combining insights from the current study with previous knowledge about CNV formation [1, 61, 65],
454 we propose a connection between replication stress, DNA repair, and CNV generation in *P. falciparum*
455 (**Fig. 4B**). Prior studies have shown that stress can either alter levels of proteins essential for HR-based
456 repair or increase the frequency of DNA breaks [92-94]. With a decrease in HR activity in particular, the
457 parasite may increase its reliance on alternative error-prone pathways to repair DNA damage.
458 Microhomology-mediated pathways require less homology and therefore, are more likely to interact
459 with diverse sequences up- and downstream of a DNA break to generate various length CNVs. So far, the
460 predominant evidence for this model in *P. falciparum* was the detection of microhomology-mediated
461 pathway signatures in CNV breakpoints [61]. Our observation of stress-induced de novo CNV formation
462 (**Table 3** and **Fig. 2I**) further supports this model in *P. falciparum* and is consistent with studies on
463 diverse organisms [30-32, 95, 96]. Interestingly, the level of de novo CNV stimulation is consistent across
464 organisms; treatment of mammalian cells with replication inhibitors also leads to a ~3-5-fold increase in
465 de novo CNVs [30-32].

466 Even with a change in the copy number of a single region per parasite, the genomic diversity within a
467 single infected human is expansive due to the sheer numbers of *P. falciparum* parasites (estimated to
468 reach 10^8 parasites when symptomatic and $>10^{11}$ in severe infection, [97]). This diversity becomes an
469 obvious advantage as a heterogeneous population prepares asexual parasites to respond to diverse
470 stressors (**Fig. 4C**). However, one question has been whether random CNVs constitutively arise across
471 the *P. falciparum* genome or only when under stress. In the former, random amplicons within individual
472 parasites would position the population to respond rapidly to selection (e.g. antimalarial exposure). In
473 the latter, specific stressors would stimulate CNV formation to increase genomic diversity. Since some

474 antimalarials act rapidly [68], we hypothesize that beneficial CNVs must already be present in a few
475 parasites across the population to increase the chances of survival. Indeed, we observed a low level of
476 de novo CNVs across the parasite genome under normal conditions (**Fig. 2D** and **2E**). However, it is also
477 important to understand how parasites respond to stressful environments during infection, including
478 changes in nutrient composition in different hosts, drug treatment during symptomatic infection, or
479 attack from the human immune system. While the current study focused on replication stress, it will be
480 important to evaluate the impact of other sources of stress on *P. falciparum* CNV formation. For
481 example, hypoxia stimulates CNV formation in cancer cells [92] and a proteotoxic drug stimulates
482 genetic change in yeast [98].

483



484

485 **Figure 4: Connection between replication stress, DNA repair, and CNV generation in the malaria genome**. **A**. Adaptations that
486 encourage CNV formation in the *P. falciparum* genome (underlined). **B**. Summary of how replication stress impacts DNA repair
487 pathways. HR, homologous recombination; NHEJ, non-homologous end-joining; MMEJ, microhomology-mediated end joining;
488 MMBIR, microhomology-mediated break-induced repair; DSB, double-strand breaks. **C.** Benefits of a diverse parasite
489 population for evolutionary potential. Stress elevates the frequency of de novo CNVs across the population, which leads to
490 more rapid evolution of beneficial CNVs (blue cells).

491 *Clinical implications & future questions*

492 *P. falciparum* causes the majority of worldwide malaria deaths and readily acquires antimalarial
493 resistance [99, 100]. Resistance-conferring CNVs that encompass multiple genes have been identified in
494 both clinical infections [51, 59, 101-105] and laboratory selections [79, 90, 106-111]. Despite their direct
495 contribution to resistance, CNVs may also facilitate the acquisition of point mutations in haploid *P.*
496 *falciparum*; strong evidence for the close relationship comes from the observation of point mutations
497 *within* amplifications selected in vitro [90, 107, 109, 112, 113]. Once de novo CNVs form during
498 replication of the asexual erythrocytic stage (**Fig. 4**), meiotic recombination during the sexual phase in
499 the mosquito can streamline beneficial CNVs to balance fitness costs [114]. Given the importance of
500 CNVs in *P. falciparum* adaptation, it is not surprising that this organism has evolved strategies to
501 encourage CNV formation (as described in *Adaptations that encourage de novo CNV formation*).
502 Additionally, parasites from specific regions of the world may have an increased propensity to develop

503 drug resistance [115]. Evaluating whether the CNV rate correlates with the parasite background will help
504 to define the evolutionary potential of this successful pathogen. Despite some success with antimalarial
505 therapies and vaccines targeting the *Plasmodium* parasite, a strategy to impede genome evolution may
506 be required to control malaria infections.

507

508 **Materials and Methods:**

509 **Parasite Lines, Compounds, and Treatments**

510 We acquired *Dd2* (MRA-156) and *FCR3* (MRA-731) parasite lines from Bei Resources (ATCC, Manassas,
511 VA). In this study, we were interested in detecting sub-clonal levels of genomic diversity that occur
512 naturally in cell culture (i.e. untreated conditions); therefore, we did not re-clone parasite lines prior to
513 treatment. For low-input genomics, we grew parasites in complete RPMI 1640 with HEPES (Thermo
514 Fisher Scientific, Waltham, MA) supplemented 0.5% Albumax II Lipid-Rich BSA (MilliporeSigma,
515 Burlington, MA) and 50 mg/L hypoxanthine (Thermo Fisher Scientific) and donor A+ human
516 erythrocytes (BioIVT, Hicksville, NY). We grew all cultures at 3% hematocrit at 37°C and individually
517 flushed flasks with 5% oxygen, 5% carbon dioxide, and 90% nitrogen gas. We diluted cultures with
518 uninfected erythrocytes and changed the culture medium every other day to keep parasitemia below
519 2% during maintenance. We confirmed that all cultures were negative for mycoplasma contamination
520 approximately monthly using a LookOut Mycoplasma PCR detection kit (MilliporeSigma).

521 We synthesized DSM1 as in previous studies [87, 116] and purchased aphidicolin (MilliporeSigma).
522 DSM1 targets *P. falciparum dhodh*, which contributes to pyrimidine biosynthesis [67]. Aphidicolin
523 inhibits B-Family DNA polymerases, and consequently, *P. falciparum* replication [117, 118]. Both
524 compounds, when applied to ring-stage parasites (**Fig. 1B**), inhibit DNA replication and stall parasites at
525 trophozoite stage (**Fig. S3A** and **S4G**, [118]).

526 To assess the effects of short-term DSM1 treatment, we acquired high ring-stage cultures (>85%) by
527 synchronizing parasites twice with 5% sorbitol, 48hrs apart. We then applied 1µM DSM1, solvent control
528 (dimethylsulfoxide, DMSO), or replication inhibition control (4.4µM aphidicolin) for 12hrs. Following
529 treatment, we washed parasites with sterile 1x phosphate-buffered saline (PBS, Thermo Fisher
530 Scientific), returned them to complete RPMI, and allowed parasites to complete their life cycle and
531 reinvade new erythrocytes for an additional 29.5-34.5hrs (**Fig. S3B**). We tracked parasitemia and
532 parasite viability on an Accuri C6 flow cytometer (BD Biosciences, Franklin Lakes, NJ) as previously
533 performed [63, 116, 119]. We stained parasites with 1x SYBR Green (Thermo Fisher Scientific, stains the
534 parasite nucleus) to assess the proportion of infected erythrocytes (parasitemia) and stage of the
535 parasite development cycle (**Fig. 1B**) and 10nM MitoProbe DiIC1 (5) (Thermo Fisher Scientific, stains
536 active parasite mitochondria) to indicate the proportion of the parasites that are viable over time (**Fig.
537 S3C**).

538 For parasite treatment for low-input genomics, we synchronized parasites (as above), applied 1µM
539 DSM1 or the DMSO control for 12hrs, and allowed recovery for 28-31hrs (**Fig. S4H**). We removed
540 treatments and tracked parasite number and health as described above. Following reinvasion, we
541 harvested viable 1n ring stage parasites for low-input genomics using flow sorting (details in *Parasite
542 Flow Sorting for Low-input Genomics*).

543 **Parasite Flow Sorting for Low-input Genomics**

544 *Cell sorter calibration & accuracy assessments.* We calibrated the flow sorter (SH800, Sony
545 Biotechnology, San Jose, CA) using the manufacturer's calibration beads. We accounted for overlaps in
546 the excitation/emission wavelengths using the integrated compensation panel matrix calculation in the
547 SH800 software according to the manufacturer's procedure. We also manually calibrated the droplet
548 sorting to the nearest 0.2mm, as recommended by the manufacturer, using the 96-well plate setting
549 (Armadillo high- performance 96-well plate, Thermo Fisher Scientific). We evaluated SH800 sorting
550 accuracy prior to low-input harvest using a colorimetric assay as previously described [120]. Briefly, we
551 mixed SYBR Green+/MitoProbe+ parasites (see staining details in *Parasite lines, Compounds, &*
552 *Treatments*) with horseradish peroxidase enzyme (Thermo Fisher Scientific) at a final concentration of
553 2.5mg/ml. We then sorted parasites into a 96-well plate filled with TMB-ELIZA substrate (Thermo Fisher
554 Scientific) using the single cell (3 drops) instrument setting, in triplicate plates (**Fig. S2**). Formation of a
555 color in the well (blue, green, or yellow) indicates the successful sorting of the enzyme, and therefore
556 parasites, into the well with the substrate. This assessment allowed us to evaluate the accuracy of
557 SH800 sorting (through the evaluation of success for 1- versus 2-cell wells, **Fig. S2C**), the consistency of
558 sorting (through the evaluation of replicates), and the best plate positions for sorting (through the
559 evaluation of performance in different plate rows/columns). Based on these evaluations, we proceeded
560 with isolating 2- and 10-cells per well (**Fig. S2D**) and avoided sorting into the top 2 rows and the first and
561 last column of the 96-well plate (**Figs. S2D** and **S6**).

562 *Parasite isolation & storage*. We stained parasites with SYBR Green and MitoProbe DilC1 (5) in complete
563 RPMI as above (see staining details in *Parasite Lines, Compounds, & Treatments*), gassed the tubes with
564 5% CO_2, 5% 0_2, 90% N, and placed sample on ice to ensure viability prior to flow sorting within 15min
565 (SH800, Sony Biotechnology Inc., San Jose, CA). We used a final concentration of 1 x 10^7 parasites/ml
566 diluted in sterile 1x PBS (Thermo Fisher Scientific) as input for sorting at the "single-cell setting" (3 drop)
567 into a 96-well plate (Armadillo high performance 96-well plate, Thermo Fisher Scientific) with each well
568 containing 2.375µl of cell lysis buffer (0.025M Tris Ph8.8 (Roche Diagnostics, Indianapolis, IN), 0.01M
569 NaCl (MilliporeSigma), 0.01M KCl (MilliporeSigma), 0.01M $(NH_4)_2SO_4$ (Thermo Fisher Scientific), 0.001M
570 EDTA (Promega, Madison, WI), and 10% Triton X-100 (MilliporeSigma)). We gated viable 1n ring-stage
571 parasites (**Fig. S5**) and sorted into the wells containing cell lysis buffer with an approximate sorting time
572 of 10min. After sorting, we centrifuged for 30 seconds in a plate centrifuge (MPS1000, Labnet
573 International, Madison, NJ). We immediately overlaid samples with one drop (approx. 25µl) of light
574 mineral oil (BioReagent grade for molecular biology, MilliporeSigma) and sealed the plates with
575 Microamp® Clear Adhesive Film (Applied Biosystems, Waltham, MA) before storage at -80°C until whole
576 genome amplification.

577 **MALBAC Whole Genome Amplification for Low-input Genomics**

578 Before whole genome amplification, we thawed the plates containing sorted parasites (see *Parasite*
579 *Isolation & Storage*) and added 1mg/ml Proteinase K in sterile 1x PBS (Thermo Fisher Scientific) to a final
580 volume of 2.5µl per well. We heated the plates in a PCR cycler (C1000, Bio-Rad Laboratories, Hercules,
581 CA) at 50°C for 3hrs, followed by 75°C for 20 min and 80°C for 5 min for proteinase k inactivation. We
582 amplified the parasite genome using the Multiple Annealing and Looping Based Amplification Cycles
583 (MALBAC) method essentially as previously described ([63], Version 1 in **Fig. S1**) with some
584 modifications (Version 2, **Fig. S1**). In summary, 1) we modified the pre-amplification random primer by

585 adding 5 additional degenerate bases with 20% GC-content to increase annealing to AT-rich genome
586 sequences (5'GTGAGTGATGGTTGAGGTAGTGTGGAGNNNNNNNNNNTTT 3'); 2) we performed 19 of the
587 21 total linear cycles with the *Bsu* DNA Polymerase (Large Fragment, New England Biosciences), which
588 has a lower optimal reaction temperature (37°C) to improve the amplification on AT-rich sequences
589 [121]; 3) we lowered the extension temperature from 40/50°C to 37°C during the linear amplification
590 cycles that used the *Bsu* enzyme (see full cycling parameters in **Fig. S1**); and 4) we integrated robotic
591 pipetting (Mosquito LV, SPT Labtech, Melbourn, UK) to increase the throughput of our assays (from 23
592 samples in Version 1 to 90 samples in the current Version 2) and limit contamination potential.

593 Overall, we performed 21 total linear cycles (19 cycles with *Bsu* polymerase and 2 cycles with *Bst*
594 polymerase, New England Biolabs, Ipswitch, MA) and 17 total exponential amplification cycles using
595 Herculase II Fusion DNA polymerase (Agilent Technologies, Santa Clara, CA). During amplification steps,
596 we employed standard steps to limit contamination [63]. For automated pipetting of the enzyme
597 solution during linear cycles, tips were changed after each round of pipetting. Post-amplification, we
598 purified amplified DNA with Zymo DNA Clean & Concentrator-5columns (Zymo Research, Irvine, CA)
599 according to the protocol and ran 2µl of all samples on 1% agarose gels to check for the presence of
600 DNA (generally, if >30ng/µl, samples could be visualized with a size range of 100 to >1500bp).

601 **Assessments of Amplification Success for Low-input Genomics**

602 *DNA quantification.* We quantified the MALBAC-amplified DNA using a Qubit fluorimeter (Qubit 1X
603 dsDNA High Sensitivity Assay Kit, Thermo Fisher Scientific).

604 *Droplet Digital PCR.* To confirm the presence of parasite DNA in MALBAC-amplified samples, we
605 performed droplet digital PCR (ddPCR) as described previously using the QX2000 droplet generator,
606 C1000 thermocycler, and QX2000 droplet reader (Bio-Rad Laboratories) [116, 122]. We used duplex
607 assays to evaluate two parasite genomic loci concurrently (*pfmdr1:* Forward-
608 TGCCCACAGAATTGCATCTA; Reverse- ACCCTGATCGAAATGGAACCT; Probe -
609 TCGTGTGTTCCATGTGACTG; *pfhsp70:* Forward- TGCTGTCATTACCGTTCCAG; Reverse -
610 AGATGCTGGTACAATTGCAGGA; Probe - AGCAGCTGCAGTAGGTTCATT (Integrated DNA
611 technologies, Newark, NJ). The reaction master mix contained 600nm of forward and reverse primers,
612 50nm probes, 10µl of ddPCR Supermix for Probes (2x, Bio-Rad Laboratories), 3µl of nuclease-free water
613 (QIAGEN), and 1.5ng (5µl) of template DNA per assay (total of 20µl). We used the following cycling
614 conditions for PCR amplification: 10 min at 95°C initial denaturation step, 1 min at 95°C second
615 denaturation step, and 2 min at 58°C annealing and extension step (ramp rate of 1°C per second), the
616 second denaturation step and the annealing/extension step repeated 60 times, and then 10 min at 98°C
617 to halt the reaction [122]. In addition to running amplified samples to assess amplification success, we
618 ran ddPCR with bulk genomic DNA as a positive control, no template controls (water replaced DNA), and
619 material from "no cell" wells to assess cross-well contamination. We considered the samples positive for
620 parasite DNA if there were more than 50 total positive droplets in target-positive clusters.

621 *High Resolution Melting Assay.* To assess potential contamination between MALBAC-amplified samples,
622 we performed asymmetric PCR amplification of the *pfdhps* locus followed by high-resolution melting
623 (HRM) as described previously [123, 124]. The *pfdhps* locus at codon 613 is distinct in *Dd2* and *FCR3*
624 parasite lines (*Dd2*: Ser-613 and *FCR3*: Ala-613, [125]). Each 20µL reaction contained 8µl of the 2.5x
625 LightScanner Master mix (BioFire™ Defense, Salt Lake City, Utah, USA), 1/10µM of forward/reverse
626 primers and 8µM probes targeting the *pfdhps* gene position 613: Forward -

627 CTCTTACAAAATATACATGTATATGATGAGTATCCACTT; Reverse-
628 CATGTAATTTTTGTTGTGTATTTATTACAACATTTTGA; Probe - AAGATTTATTGCCCATTGCATGA/3SpC3,
629 (Integrated DNA technologies), 7µl of nuclease free water, and 3µl of DNA (~0.05ng total). We used the
630 following cycling conditions for PCR amplification with the Rotor-Gene Q instrument with a 72-well rotor
631 (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-
632 melt 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
633 the change in fluorescence versus temperature (dF/T) using Rotor-Gene Q software (version 2.3.5, build
634 1; QIAGEN) and compared HRM peaks of amplified samples to bulk genomic DNA and plasmid controls.

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

636 We extracted bulk DNA for short-read sequencing as previously performed [63]. Briefly, we lysed
637 erythrocytes with 1.5% saponin and washed the parasite pellet 3 times with 1x PBS (Thermo Fisher
638 Scientific), before resuspension in a buffered solution (150mM NaCl (MilliporeSigma), 10mM EDTA
639 (Promega Corporation, Madison, WI), and 50mM Tris pH7.5 (Roche Diagnostics)) to a total volume of
640 500µl. We then lysed the parasites with 10% sarkosyl (MilliporeSigma) and 20mg/ml proteainase K
641 (Thermo Fisher Scientific) at 37°C overnight before DNA purification using standard
642 phenol/chloroform/isoamyl alcohol extraction and chloroform washing steps (2 times each, [63]).
643 Finally, we precipitated DNA using 100% ethanol with 100mM of sodium acetate overnight in DNA-lo
644 bind tubes (Eppendorf, Enfield, CT) and then washed twice with 70% ethanol before resuspension in
645 50µl nuclease free water (QIAGEN). We stored bulk genomic DNA at -20°C until sequencing library
646 preparation.

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

648 *Low-input sample selection.* We selected 16 low-input samples from the *FCR3* plate, 36 samples from
649 the untreated *Dd2* plate, and 38 samples from the treated *Dd2* plate for short-read Illumina sequencing.
650 We based our selection on the quantity of the MALBAC amplified DNA and presence of parasite DNA
651 using ddPCR (**Figs. S6 and S7**). In summary, the majority of *FCR3* and untreated *Dd2* samples yielded
652 quantifiable parasite DNA following MALBAC amplification (53/60 *FCR3* samples and 60/60 untreated
653 *Dd2* samples); in these conditions, we chose samples randomly to proceed with sequencing (indicated in
654 **Fig. S6**). For treated *Dd2* samples, we chose samples for sequencing if they had adequate DNA quantity
655 (>10ng total, 30 samples, **Fig. S6**) or had ddPCR results showing the presence of parasite DNA (an
656 additional 8 samples).

657 *Short-read sequencing.* Before short-read sequencing, we sheared bulk samples and low-input samples
658 using Covaris M220 Focused Ultrasonicator for 150s and 130s, respectively, to generate fragment sizes
659 of ~350bp as evaluated by an Agilent 2100 Bioanalyzer using the High Sensitivity DNA kit (Agilent, Santa
660 Clara). We adjusted the volume of sheared samples with nuclease-free water up to 50µl. For samples
661 with >100ng (**Table S2,** including bulk and MALBAC amplified samples), we diluted them to 1.2-2ng/µl;
662 for samples <100ng, we proceeded with no dilution. We used NEBNext Ultra II kit (Illumina Inc., San
663 Diego, CA) to prepare libraries for sequencing with 3 cycles of PCR amplification, as performed
664 previously [63]. We quantified the resulting libraries using NEBNext Library Quant Kit (Illumina Inc.)
665 before sequencing on the Illumina Nextseq 550 using 150 bp paired-end cycles.

666 **Short-Read Sequence Processing & Analysis**

667 *Read processing and alignment*. We performed short-read quality control steps as described previously
668 [61, 63]. Briefly, we reordered and removed singletons and subsequently interleaved paired reads using
669 BBMap, trimmed the MALBAC common sequence, PhiX, and Illumina adapters from the remaining reads
670 with the BBDuk tool within BBMap, and aligned reads to the *pf3D7-62_v3* reference genome using the
671 Speedseq genome aligner [126]. We removed reads that map to VAR regions from bam files according
672 to previously defined genomic coordinates [77]. We filtered out reads with low mapping quality (<q30)
673 and duplicated reads using SAMtools [127]. We employed Qualimap to report mean coverage and
674 standard deviation across the genome [128]. Using non-overlapping 20kb size bins, we calculated the
675 coefficient of variation of read coverage by dividing the standard deviation of coverage within a bin by
676 the mean across a sample and multiplying by 100 [129, 130] (R version 4.4.2).

677 *Down-sampling.* For analysis that assessed down-sampled data (**Figs. 2, S10, and S11**), we first
678 converted the processed bam files back into FASTQ files using SAMtools and then used the reformat.sh
679 option of BBtools to select 1.3M reads from each FASTQ (represented the fewest number of reads from
680 a sample that passed quality filtering from the final dataset). We then realigned files to the reference
681 genome (*pf3D7-62_v3*).

682 *Single nucleotide polymorphism analysis*. We performed SNP genotyping and analysis as previously [63],
683 based on the MalariaGen *P. falciparum* Community Project V6.0 pipeline [131-134] using the *pf3D7-*
684 *62_v3* reference genome. Briefly, we applied GATK's Base Quality Score Recalibration using default
685 settings. We detected potential SNPs using GATK's HaplotypeCaller and subsequently genotyped the
686 SNPs using CombineGVCFs and GenotypeGVCFs. Then we employed GATK's VariantRecalibrator using
687 previously validated SNP datasets [114]. We then applied GATK's ApplyRecalibration to assign a VQSLOD
688 score [134]. We filtered the resulting SNPs for those with VQSLOD > 6 and for a GT quality metric >20 to
689 ensure high-quality variant calling. We only selected variants flagged as Bi-allelic to simplify the analysis.
690 For SNP Principle Component Analysis (PCA), we merged experiment-wide SNP data (described above)
691 into a single file. Then we merged the VCF into a large matrix and converted the genomic data into
692 numeric information using the 'vcfr' package in R (Version 4.2.3) (https://CRAN.R-
693 project.org/package=vcfR). We excluded individual SNPs if >25% of the samples lacked a call in this
694 position or if all calls were the same for each sample in that position). We scaled the remaining SNPs
695 around the origin using the 'scale' R function. We then calculated the principal components using the
696 'prcomp' R function, and scored the dataset using the 'scores' function from the 'vegan' R package
697 (https://CRAN.R-project.org/package=vegan).

698 **Copy Number Variation Analysis**

699 *CNV calling in bulk samples.* We performed CNV detection for bulk samples similar to as previously
700 described [61, 63]. Briefly, we called CNVs independently using two methods, CNVnator (read depth
701 based calling, [135]) and LUMPY (split and discordant read based calling, [74]). To Identify CNVs called in
702 both methods, we used SVCROWS to define overlapping CNV regions relative to their size. Briefly,
703 SVCROWS uses a reciprocal-overlap-based approach (i.e. two CNVs must be overlapping each other at,
704 or greater than, a defined threshold) to determine if two CNVs are close enough in their genomic
705 position to be called the same. The source code for SVCROWS can be accessed at https://github.com/A-
706 Crow-Nowhere/SVCROWS.git. The program utilizes different thresholds for overlap based on the sizes of
707 the CNVs being compared, ensuring that we account for shifts in CNV position. For known CNV calls, we

708 used the following SVCROWs input parameters: ExpandRORegion = FALSE, BPfactor = TRUE, DefaultSizes
709 = FALSE, xs = 3000, xl = 10,000, y1s = 300, y1l = 1000, y2s = 50% and y2l = 80%; based on the average size
710 of a *P. falciparum* gene (~2.3kb) and intergenic region (~2kb). Similar to our previous study [63], we
711 identified 3 known CNVs that were called by both LUMPY and CNVnator methods in the core genome of
712 bulk samples (untreated and treated). We determined known CNV boundaries using SVCROWS: *pfmdr1*
713 (*Pf3D7_05_v3*, 888001-970000, 82kb), *pf11-1* (*Pf3D7_10_v3*, 1521345-1541576, 20kb), and *pf332*
714 (*Pf3D7_11_v3*, 1950201-1962400, 12kb).

715 *CNV calling in 2-cell samples.* We employed two methods for CNV detection in the core genome of low-
716 input samples; LUMPY is a split/discordant read strategy with high sensitivity [74], and HapCNV is a read
717 coverage-based strategy designed for haploid genomes [73]. We ran LUMPY as part of Speedseq with
718 default parameters as previously described [61]. We filtered resulting structural variants to include only
719 duplications (DUP, >1 copy of a region) and deletions (DEL, one less copy of region than the reference).
720 We then filtered those calls for those GQ > 20 to ensure high-fidelity calls. In HapCNV, we used a quality
721 control and bias correction procedure to exclude bins of poor quality and remove bias introduced by GC
722 content and mappability variation. We then constructed a pseudo-reference for each *Dd2* low-input
723 sample using within-*Dd2* information, which enabled control of background noise while preserving CNV
724 signals after normalization. Finally, we used a circular binary segmentation algorithm (CBS, [136]) to
725 detect copy number change points followed by a Gaussian Mixture Model (GMM, [137]) for CNV
726 identification. The source code for HapCNV and examples of real data application can be accessed at
727 https://github.com/FeifeiXiao-lab/HapCNV. For statistics, we used PRISM (GraphPad Software, La Jolla,
728 CA), using unpaired parametric T-tests with Welch's correction.

729 *Defining CNV regions/determination of "rarity" in CNV calling.* Small differences in sequence quality
730 surrounding a read can lead to shifted breakpoint determination for biologically identical CNVs, which is
731 especially true for low-input genomics datasets [38]. To account for this, we used SVCROWS to
732 determine whether two CNV signals were the same within and between samples (see *CNV calling in bulk*
733 *samples* for input parameters). We assigned the categorizations of "rare" and "common" by assessing
734 the CNV region frequency within datasets. "Rare" CNVs were defined as occurring in <10% of the
735 samples within a treatment group; "common" CNVs were defined as occurring in ≥10% of samples
736 within a treatment group; "known" CNVs were defined by CNVs called in bulk samples (see above, *CNV*
737 *calling in bulk samples*).

738 *High-confidence CNV region identification.* To identify "high-confidence" CNV regions called by both
739 HapCNV and LUMPY methods, we compared the 'consensus list' generated by SVCROWS for each
740 detection method by combining them into a single SVCROWS input file. Because 1) HapCNV generates
741 imprecise breakpoints, and 2) there is a large disparity of average CNV region lengths between the two
742 methods (HapCNV = ~40kb, LUMPY = ~4.3kb), we relaxed the stringency of the SVCROWS parameters.
743 Our input parameters to generate the "high-confidence" list were as follows: xs = 3000, xl = 6000, y1s =
744 500, y1l = 1500, y2s = 30% and y2l = 60%. We defined "high-confidence" CNV regions as those that had
745 >1 match from both HapCNV and LUMPY that was the same type (i.e. either duplication or deletion or
746 mixed). For Venn diagram generation, we calculated overlaps using SVCROWS "Scavenge" mode (input
747 parameters: ExpandRORegion = FALSE, BPfactor = TRUE, DefaultSizes = FALSE, xs = 3000, xl = 6000, y1s =
748 500, y1l = 1500, y2s = 30% and y2l = 60%). We systematically compared lists for each overlap
749 comparison, and if regions had at least one match in an opposing dataset, we considered it a match. We
750 used the *draw.quad.venn* function in the "VennDiagram" R package (R 4.2.1) to generate the diagram.

751 *Gene Ontology enrichment and protein class identification.* We used the online Gene Ontology Resource
752 (geneontology.org) to perform GO enrichment analysis using the PANTHER Classification System [138,
753 139]. Since a large portion of the *P. falciparum* genome remains unannotated (PlasmoDB, 30%) and the
754 majority of molecular functions remain unclassified (92.8%), we used the Panther Protein class
755 assessment (version 19.0, only 55.9% remained unclassified) with default statistics (Fisher's test with
756 FDR adjusted p value of <0.05 for significance, which is recommended for small counts and overlaps
757 between classes). We used the web tool to represent protein classes on pie charts.

758 *Comparison to clinical CNV dataset.* To determine genes covered by de novo CNVs from HapCNV and
759 LUMPY, we used SVCROWS "Hunt" mode (input parameters: BPfactor = TRUE, DefaultSizes = FALSE,
760 xs = 3000, xl = 6000, y1s = 300, y1l = 600, y2s = 30, y2l = 60), which takes a secondary input list of
761 known genes (Pf3D7_62_v3, Plasmodb.org) against which to compare CNV regions. For the clinical
762 dataset, we used genes covered by >300bp CNV regions present in high frequency in clinical isolates
763 (Supplementary Table 3 from [59]), and manually reformatted to match SVCROWS input style
764 guidelines before rerunning as above.

765

766 **Acknowledgments:**

767 We would like to thank Dr. John Campbell for use of the Mosquito LV instrument and assay reagents for
768 the colorimetric assessment of the SH800. We would also like to thank Dr. Ali Guler for statistical
769 consultation and Dr. Heidi Seears from the University of Virginia, Department of Biology Genomics Core
770 Facility for sequencing assistance. This work was funded by 1R01AI150856-01A1 (to JLG).

771

772 **References:**

773 1. Hastings, P.J., J.R. Lupski, S.M. Rosenberg, and G. Ira, *Mechanisms of change in gene copy*
774 *number.* Nat Rev Genet, 2009. 10(8): p. 551-64.
775 2. Lauer, S. and D. Gresham, *An evolving view of copy number variants.* Curr Genet, 2019.
776 3. Pos, O., J. Radvanszky, G. Buglyo, Z. Pos, D. Rusnakova, B. Nagy, and T. Szemes, *DNA copy*
777 *number variation: Main characteristics, evolutionary significance, and pathological aspects.*
778 Biomed J, 2021. 44(5): p. 548-559.
779 4. Roth, J.R., *The joys and terrors of fast adaptation: new findings elucidate antibiotic resistance*
780 *and natural selection.* Mol Microbiol, 2011. 79(2): p. 279-82.
781 5. Payen, C., S.C. Di Rienzi, G.T. Ong, J.L. Pogachar, J.C. Sanchez, A.B. Sunshine, M.K. Raghuraman,
782 B.J. Brewer, and M.J. Dunham, *The dynamics of diverse segmental amplifications in populations*
783 *of Saccharomyces cerevisiae adapting to strong selection.* G3 (Bethesda), 2014. 4(3): p. 399-409.
784 6. Elde, N.C., S.J. Child, M.T. Eickbush, J.O. Kitzman, K.S. Rogers, J. Shendure, A.P. Geballe, and H.S.
785 Malik, *Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses.*
786 Cell, 2012. 150(4): p. 831-41.
787 7. Leary, R.J., J.C. Lin, J. Cummins, S. Boca, L.D. Wood, D.W. Parsons, S. Jones, T. Sjoblom, B.H. Park,
788 R. Parsons, J. Willis, D. Dawson, J.K. Willson, T. Nikolskaya, Y. Nikolsky, L. Kopelovich, N.
789 Papadopoulos, L.A. Pennacchio, T.L. Wang, S.D. Markowitz, G. Parmigiani, K.W. Kinzler, B.
790 Vogelstein, and V.E. Velculescu, *Integrated analysis of homozygous deletions, focal*
791 *amplifications, and sequence alterations in breast and colorectal cancers.* Proc Natl Acad Sci U S
792 A, 2008. 105(42): p. 16224-9.

8. Beroukhim, R., C.H. Mermel, D. Porter, G. Wei, S. Raychaudhuri, J. Donovan, J. Barretina, J.S. Boehm, J. Dobson, M. Urashima, K.T. Mc Henry, R.M. Pinchback, A.H. Ligon, Y.J. Cho, L. Haery, H. Greulich, M. Reich, W. Winckler, M.S. Lawrence, B.A. Weir, K.E. Tanaka, D.Y. Chiang, A.J. Bass, A. Loo, C. Hoffman, J. Prensner, T. Liefeld, Q. Gao, D. Yecies, S. Signoretti, E. Maher, F.J. Kaye, H. Sasaki, J.E. Tepper, J.A. Fletcher, J. Tabernero, J. Baselga, M.S. Tsao, F. Demichelis, M.A. Rubin, P.A. Janne, M.J. Daly, C. Nucera, R.L. Levine, B.L. Ebert, S. Gabriel, A.K. Rustgi, C.R. Antonescu, M. Ladanyi, A. Letai, L.A. Garraway, M. Loda, D.G. Beer, L.D. True, A. Okamoto, S.L. Pomeroy, S. Singer, T.R. Golub, E.S. Lander, G. Getz, W.R. Sellers, and M. Meyerson, *The landscape of somatic copy-number alteration across human cancers.* Nature, 2010. 463(7283): p. 899-905.
9. Zhang, L., Y. Yuan, K.H. Lu, and L. Zhang, *Identification of recurrent focal copy number variations and their putative targeted driver genes in ovarian cancer.* BMC Bioinformatics, 2016. 17(1): p. 222.
10. Peng, H., L. Lu, Z. Zhou, J. Liu, D. Zhang, K. Nan, X. Zhao, F. Li, L. Tian, H. Dong, and Y. Yao, *CNV Detection from Circulating Tumor DNA in Late Stage Non-Small Cell Lung Cancer Patients.* Genes (Basel), 2019. 10(11).
11. Krijgsman, O., B. Carvalho, G.A. Meijer, R.D. Steenbergen, and B. Ylstra, *Focal chromosomal copy number aberrations in cancer-Needles in a genome haystack.* Biochim Biophys Acta, 2014. 1843(11): p. 2698-2704.
12. Carey-Smith, S.L., R.S. Kotecha, L.C. Cheung, and S. Malinge, *Insights into the Clinical, Biological and Therapeutic Impact of Copy Number Alteration in Cancer.* Int J Mol Sci, 2024. 25(13).
13. Gonzalez, E., H. Kulkarni, H. Bolivar, A. Mangano, R. Sanchez, G. Catano, R.J. Nibbs, B.I. Freedman, M.P. Quinones, M.J. Bamshad, K.K. Murthy, B.H. Rovin, W. Bradley, R.A. Clark, S.A. Anderson, J. O'Connell R, B.K. Agan, S.S. Ahuja, R. Bologna, L. Sen, M.J. Dolan, and S.K. Ahuja, *The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility.* Science, 2005. 307(5714): p. 1434-40.
14. Sebat, J., B. Lakshmi, D. Malhotra, J. Troge, C. Lese-Martin, T. Walsh, B. Yamrom, S. Yoon, A. Krasnitz, J. Kendall, A. Leotta, D. Pai, R. Zhang, Y.H. Lee, J. Hicks, S.J. Spence, A.T. Lee, K. Puura, T. Lehtimaki, D. Ledbetter, P.K. Gregersen, J. Bregman, J.S. Sutcliffe, V. Jobanputra, W. Chung, D. Warburton, M.C. King, D. Skuse, D.H. Geschwind, T.C. Gilliam, K. Ye, and M. Wigler, *Strong association of de novo copy number mutations with autism.* Science, 2007. 316(5823): p. 445-9.
15. Glessner, J.T., J.P. Bradfield, K. Wang, N. Takahashi, H. Zhang, P.M. Sleiman, F.D. Mentch, C.E. Kim, C. Hou, K.A. Thomas, M.L. Garris, S. Deliard, E.C. Frackelton, F.G. Otieno, J. Zhao, R.M. Chiavacci, M. Li, J.D. Buxbaum, R.I. Berkowitz, H. Hakonarson, and S.F. Grant, *A genome-wide study reveals copy number variants exclusive to childhood obesity cases.* Am J Hum Genet, 2010. 87(5): p. 661-6.
16. Olsson, L.M. and R. Holmdahl, *Copy number variation in autoimmunity--importance hidden in complexity?* Eur J Immunol, 2012. 42(8): p. 1969-76.
17. Butchbach, M.E., *Copy Number Variations in the Survival Motor Neuron Genes: Implications for Spinal Muscular Atrophy and Other Neurodegenerative Diseases.* Front Mol Biosci, 2016. 3: p. 7.
18. Sekar, A., A.R. Bialas, H. de Rivera, A. Davis, T.R. Hammond, N. Kamitaki, K. Tooley, J. Presumey, M. Baum, V. Van Doren, G. Genovese, S.A. Rose, R.E. Handsaker, C. Schizophrenia Working Group of the Psychiatric Genomics, M.J. Daly, M.C. Carroll, B. Stevens, and S.A. McCarroll, *Schizophrenia risk from complex variation of complement component 4.* Nature, 2016. 530(7589): p. 177-83.
19. Zekavat, S.M., S. Ruotsalainen, R.E. Handsaker, M. Alver, J. Bloom, T. Poterba, C. Seed, J. Ernst, M. Chaffin, J. Engreitz, G.M. Peloso, A. Manichaikul, C. Yang, K.A. Ryan, M. Fu, W.C. Johnson, M. Tsai, M. Budoff, R.S. Vasan, L.A. Cupples, J.I. Rotter, S.S. Rich, W. Post, B.D. Mitchell, A. Correa, A. Metspalu, J.G. Wilson, V. Salomaa, M. Kellis, M.J. Daly, B.M. Neale, S. McCarroll, I. Surakka, T.

841 Esko, A. Ganna, S. Ripatti, S. Kathiresan, P. Natarajan, and N.T.L.W. Group, *Deep coverage whole*
842 *genome sequences and plasma lipoprotein(a) in individuals of European and African ancestries.*
843 Nat Commun, 2018. 9(1): p. 2606.
844 20. Farashi, S. and C.L. Harteveld, *Molecular basis of alpha-thalassemia.* Blood Cells Mol Dis, 2018.
845 70: p. 43-53.
846 21. Wisniowiecka-Kowalnik, B. and B.A. Nowakowska, *Genetics and epigenetics of autism spectrum*
847 *disorder-current evidence in the field.* J Appl Genet, 2019. 60(1): p. 37-47.
848 22. Craven, S.H. and E.L. Neidle, *Double trouble: medical implications of genetic duplication and*
849 *amplification in bacteria.* Future Microbiol, 2007. 2(3): p. 309-21.
850 23. Sandegren, L. and D.I. Andersson, *Bacterial gene amplification: implications for the evolution of*
851 *antibiotic resistance.* Nat Rev Microbiol, 2009. 7(8): p. 578-88.
852 24. Hjort, K., H. Nicoloff, and D.I. Andersson, *Unstable tandem gene amplification generates*
853 *heteroresistance (variation in resistance within a population) to colistin in Salmonella enterica.*
854 Mol Microbiol, 2016. 102(2): p. 274-289.
855 25. Anderson, S.E., E.X. Sherman, D.S. Weiss, and P.N. Rather, *Aminoglycoside Heteroresistance in*
856 *Acinetobacter baumannii AB5075.* mSphere, 2018. 3(4).
857 26. Nicoloff, H., K. Hjort, B.R. Levin, and D.I. Andersson, *The high prevalence of antibiotic*
858 *heteroresistance in pathogenic bacteria is mainly caused by gene amplification.* Nat Microbiol,
859 2019. 4(3): p. 504-514.
860 27. Baslan, T., J. Kendall, K. Volyanskyy, K. McNamara, H. Cox, S. D'Italia, F. Ambrosio, M. Riggs, L.
861 Rodgers, A. Leotta, J. Song, Y. Mao, J. Wu, R. Shah, R. Gularte-Merida, K. Chadalavada, G.
862 Nanjangud, V. Varadan, A. Gordon, C. Curtis, A. Krasnitz, N. Dimitrova, L. Harris, M. Wigler, and J.
863 Hicks, *Novel insights into breast cancer copy number genetic heterogeneity revealed by single-*
864 *cell genome sequencing.* Elife, 2020. 9.
865 28. van Dijk, E., T. van den Bosch, K.J. Lenos, K. El Makrini, L.E. Nijman, H.F.B. van Essen, N. Lansu,
866 M. Boekhout, J.H. Hageman, R.C. Fitzgerald, C.J.A. Punt, J.B. Tuynman, H.J.G. Snippert, G. Kops,
867 J.P. Medema, B. Ylstra, L. Vermeulen, and D.M. Miedema, *Chromosomal copy number*
868 *heterogeneity predicts survival rates across cancers.* Nat Commun, 2021. 12(1): p. 3188.
869 29. Sobral, D., M. Martins, S. Kaplan, M. Golkaram, M. Salmans, N. Khan, R. Vijayaraghavan, S.
870 Casimiro, A. Fernandes, P. Borralho, C. Ferreira, R. Pinto, C. Abreu, A.L. Costa, S. Zhang, T.
871 Pawlowski, J. Godsey, A. Mansinho, D. Macedo, S. Lobo-Martins, P. Filipe, R. Esteves, J.
872 Coutinho, P.M. Costa, A. Ramires, F. Aldeia, A. Quintela, A. So, L. Liu, A.R. Grosso, and L. Costa,
873 *Genetic and microenvironmental intra-tumor heterogeneity impacts colorectal cancer evolution*
874 *and metastatic development.* Commun Biol, 2022. 5(1): p. 937.
875 30. Arlt, M.F., J.G. Mulle, V.M. Schaibley, R.L. Ragland, S.G. Durkin, S.T. Warren, and T.W. Glover,
876 *Replication stress induces genome-wide copy number changes in human cells that resemble*
877 *polymorphic and pathogenic variants.* Am J Hum Genet, 2009. 84(3): p. 339-50.
878 31. Arlt, M.F., A.C. Ozdemir, S.R. Birkeland, T.E. Wilson, and T.W. Glover, *Hydroxyurea induces de*
879 *novo copy number variants in human cells.* Proc Natl Acad Sci U S A, 2011. 108(42): p. 17360-5.
880 32. Arlt, M.F., S. Rajendran, S.R. Birkeland, T.E. Wilson, and T.W. Glover, *De novo CNV formation in*
881 *mouse embryonic stem cells occurs in the absence of Xrcc4-dependent nonhomologous end*
882 *joining.* PLoS Genet, 2012. 8(9): p. e1002981.
883 33. Lauer, S., G. Avecilla, P. Spealman, G. Sethia, N. Brandt, S.F. Levy, and D. Gresham, *Single-cell*
884 *copy number variant detection reveals the dynamics and diversity of adaptation.* PLoS Biol, 2018.
885 16(12): p. e3000069.
886 34. Oketch, D.J.A., M. Giulietti, and F. Piva, *Copy Number Variations in Pancreatic Cancer: From*
887 *Biological Significance to Clinical Utility.* Int J Mol Sci, 2023. 25(1).

35. Yi, D., J.W. Nam, and H. Jeong, *Toward the functional interpretation of somatic structural variations: bulk- and single-cell approaches.* Brief Bioinform, 2023. 24(5).
36. Macaulay, I.C. and T. Voet, *Single cell genomics: advances and future perspectives.* PLoS Genet, 2014. 10(1): p. e1004126.
37. Wang, Y. and N.E. Navin, *Advances and applications of single-cell sequencing technologies.* Mol Cell, 2015. 58(4): p. 598-609.
38. Gawad, C., W. Koh, and S.R. Quake, *Single-cell genome sequencing: current state of the science.* Nat Rev Genet, 2016. 17(3): p. 175-88.
39. Wang, X., H. Chen, and N.R. Zhang, *DNA copy number profiling using single-cell sequencing.* Brief Bioinform, 2018. 19(5): p. 731-736.
40. Patel, A.P., I. Tirosh, J.J. Trombetta, A.K. Shalek, S.M. Gillespie, H. Wakimoto, D.P. Cahill, B.V. Nahed, W.T. Curry, R.L. Martuza, D.N. Louis, O. Rozenblatt-Rosen, M.L. Suvà, A. Regev, and B.E. Bernstein, *Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma.* Science, 2014. 344(6190): p. 1396-1401.
41. Xu, Q.H., S.H. Chen, Y.B. Hu, and W. Huang, *Single-cell RNA transcriptome reveals the intra-tumoral heterogeneity and regulators underlying tumor progression in metastatic pancreatic ductal adenocarcinoma.* Cell Death Discovery, 2021. 7(1).
42. Gao, R.L., S.S. Bai, Y.C. Henderson, Y.Y. Lin, A. Schalck, Y. Yan, T. Kumar, M. Hu, E. Sei, A. Davis, F. Wang, S.F. Shaitelman, J.R. Wang, K. Chen, S. Moulder, S.Y. Lai, and N.E. Navin, *Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes.* Nature Biotechnology, 2021. 39(5): p. 599-608.
43. Mahdipour-Shirayeh, A., N. Erdmann, C. Leung-Hagesteijn, and R.E. Tiedemann, *sciCNV: high-throughput paired profiling of transcriptomes and DNA copy number variations at single-cell resolution.* Briefings in Bioinformatics, 2022. 23(1).
44. Navin, N., J. Kendall, J. Troge, P. Andrews, L. Rodgers, J. McIndoo, K. Cook, A. Stepansky, D. Levy, D. Esposito, L. Muthuswamy, A. Krasnitz, W.R. McCombie, J. Hicks, and M. Wigler, *Tumour evolution inferred by single-cell sequencing.* Nature, 2011. 472(7341): p. 90-4.
45. McConnell, M.J., M.R. Lindberg, K.J. Brennand, J.C. Piper, T. Voet, C. Cowing-Zitron, S. Shumilina, R.S. Lasken, J.R. Vermeesch, I.M. Hall, and F.H. Gage, *Mosaic copy number variation in human neurons.* Science, 2013. 342(6158): p. 632-7.
46. Cai, X., G.D. Evrony, H.S. Lehmann, P.C. Elhosary, B.K. Mehta, A. Poduri, and C.A. Walsh, *Single-cell, genome-wide sequencing identifies clonal somatic copy-number variation in the human brain.* Cell Rep, 2014. 8(5): p. 1280-9.
47. Francis, J.M., C.Z. Zhang, C.L. Maire, J. Jung, V.E. Manzo, V.A. Adalsteinsson, H. Homer, S. Haidar, B. Blumenstiel, C.S. Pedamallu, A.H. Ligon, J.C. Love, M. Meyerson, and K.L. Ligon, *EGFR variant heterogeneity in glioblastoma resolved through single-nucleus sequencing.* Cancer Discov, 2014. 4(8): p. 956-71.
48. Eirew, P., A. Steif, J. Khattra, G. Ha, D. Yap, H. Farahani, K. Gelmon, S. Chia, C. Mar, A. Wan, E. Laks, J. Biele, K. Shumansky, J. Rosner, A. McPherson, C. Nielsen, A.J. Roth, C. Lefebvre, A. Bashashati, C. de Souza, C. Siu, R. Aniba, J. Brimhall, A. Oloumi, T. Osako, A. Bruna, J.L. Sandoval, T. Algara, W. Greenwood, K. Leung, H. Cheng, H. Xue, Y. Wang, D. Lin, A.J. Mungall, R. Moore, Y. Zhao, J. Lorette, L. Nguyen, D. Huntsman, C.J. Eaves, C. Hansen, M.A. Marra, C. Caldas, S.P. Shah, and S. Aparicio, *Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution.* Nature, 2015. 518(7539): p. 422-6.
49. Rohrback, S., C. April, F. Kaper, R.R. Rivera, C.S. Liu, B. Siddoway, and J. Chun, *Submegabase copy number variations arise during cerebral cortical neurogenesis as revealed by single-cell whole-genome sequencing.* Proc Natl Acad Sci U S A, 2018. 115(42): p. 10804-10809.

935 50. Laks, E., A. McPherson, H. Zahn, D. Lai, A. Steif, J. Brimhall, J. Biele, B. Wang, T. Masud, J. Ting, D.
936 Grewal, C. Nielsen, S. Leung, V. Bojilova, M. Smith, O. Golovko, S. Poon, P. Eirew, F. Kabeer, T.
937 Ruiz de Algara, S.R. Lee, M.J. Taghiyar, C. Huebner, J. Ngo, T. Chan, S. Vatrt-Watts, P. Walters, N.
938 Abrar, S. Chan, M. Wiens, L. Martin, R.W. Scott, T.M. Underhill, E. Chavez, C. Steidl, D. Da Costa,
939 Y. Ma, R.J.N. Coope, R. Corbett, S. Pleasance, R. Moore, A.J. Mungall, C. Mar, F. Cafferty, K.
940 Gelmon, S. Chia, C.I.G.C. Team, M.A. Marra, C. Hansen, S.P. Shah, and S. Aparicio, *Clonal*
941 *Decomposition and DNA Replication States Defined by Scaled Single-Cell Genome Sequencing.*
942 Cell, 2019. 179(5): p. 1207-1221 e22.
943 51. Kidgell, C., S.K. Volkman, J. Daily, J.O. Borevitz, D. Plouffe, Y. Zhou, J.R. Johnson, K. Le Roch, O.
944 Sarr, O. Ndir, S. Mboup, S. Batalov, D.F. Wirth, and E.A. Winzeler, *A systematic map of genetic*
945 *variation in Plasmodium falciparum.* PLoS Pathog, 2006. 2(6): p. e57.
946 52. Ribacke, U., B.W. Mok, V. Wirta, J. Normark, J. Lundeberg, F. Kironde, T.G. Egwang, P. Nilsson,
947 and M. Wahlgren, *Genome wide gene amplifications and deletions in Plasmodium falciparum.*
948 Molecular and Biochemical Parasitology, 2007. 155(1): p. 33-44.
949 53. Cheeseman, I., N. Gomez-Escobar, C. Carret, A. Ivens, L. Stewart, K. Tetteh, and D. Conway, *Gene*
950 *copy number variation throughout the Plasmodium falciparum genome.* BMC Genomics, 2009.
951 10(1): p. 353.
952 54. Bopp, S.E., M.J. Manary, A.T. Bright, G.L. Johnston, N.V. Dharia, F.L. Luna, S. McCormack, D.
953 Plouffe, C.W. McNamara, J.R. Walker, D.A. Fidock, E.L. Denchi, and E.A. Winzeler, *Mitotic*
954 *evolution of Plasmodium falciparum shows a stable core genome but recombination in antigen*
955 *families.* PLoS Genet, 2013. 9(2): p. e1003293.
956 55. Gendrot, M., R. Fawaz, J. Dormoi, M. Madamet, and B. Pradines, *Genetic diversity and deletion*
957 *of Plasmodium falciparum histidine-rich protein 2 and 3: a threat to diagnosis of P. falciparum*
958 *malaria.* Clin Microbiol Infect, 2019. 25(5): p. 580-585.
959 56. Matesanz, F., M. Tellez, and A. Alcina, *The Plasmodium falciparum fatty acyl-CoA synthetase*
960 *family (PfACS) and differential stage-specific expression in infected erythrocytes.* Mol Biochem
961 Parasitol, 2003. 126(1): p. 109-12.
962 57. Dankwa, S., C. Lim, A.K. Bei, R.H. Jiang, J.R. Abshire, S.D. Patel, J.M. Goldberg, Y. Moreno, M.
963 Kono, J.C. Niles, and M.T. Duraisingh, *Ancient human sialic acid variant restricts an emerging*
964 *zoonotic malaria parasite.* Nat Commun, 2016. 7: p. 11187.
965 58. Price, R.N., A.-C. Uhlemann, A. Brockman, R. McGready, E. Ashley, L. Phaipun, R. Patel, K. Laing,
966 S. Looareesuwan, N.J. White, F. Nosten, and S. Krishna, *Mefloquine resistance in Plasmodium*
967 *falciparum and increased pfmdr1 gene copy number.* The Lancet, 2004. 364(9432): p. 438-447.
968 59. Ravenhall, M., E.D. Benavente, C.J. Sutherland, D.A. Baker, S. Campino, and T.G. Clark, *An*
969 *analysis of large structural variation in global Plasmodium falciparum isolates identifies a novel*
970 *duplication of the chloroquine resistance associated gene.* Sci Rep, 2019. 9(1): p. 8287.
971 60. Heinberg, A., E. Siu, C. Stern, E.A. Lawrence, M.T. Ferdig, K.W. Deitsch, and L.A. Kirkman, *Direct*
972 *evidence for the adaptive role of copy number variation on antifolate susceptibility in*
973 *Plasmodium falciparum.* Mol Microbiol, 2013. 88(4): p. 702-12.
974 61. Huckaby, A.C., C.S. Granum, M.A. Carey, K. Szlachta, B. Al-Barghouthi, Y.H. Wang, and J.L. Guler,
975 *Complex DNA structures trigger copy number variation across the Plasmodium falciparum*
976 *genome.* Nucleic Acids Res, 2019. 47(4): p. 1615-1627.
977 62. Gardner, M.J., N. Hall, E. Fung, O. White, M. Berriman, R.W. Hyman, J.M. Carlton, A. Pain, K.E.
978 Nelson, S. Bowman, I.T. Paulsen, K. James, J.A. Eisen, K. Rutherford, S.L. Salzberg, A. Craig, S.
979 Kyes, M.-S. Chan, V. Nene, S.J. Shallom, B. Suh, J. Peterson, S. Angiuoli, M. Pertea, J. Allen, J.
980 Selengut, D. Haft, M.W. Mather, A.B. Vaidya, D.M.A. Martin, A.H. Fairlamb, M.J. Fraunholz, D.S.
981 Roos, S.A. Ralph, G.I. McFadden, L.M. Cummings, G.M. Subramanian, C. Mungall, J.C. Venter,

982 D.J. Carucci, S.L. Hoffman, C. Newbold, R.W. Davis, C.M. Fraser, and B. Barrell, *Genome sequence*
983 *of the human malaria parasite Plasmodium falciparum.* Nature, 2002. 419(6906): p. 498-511.
984 63. Liu, S., A.C. Huckaby, A.C. Brown, C.C. Moore, I. Burbulis, M.J. McConnell, and J.L. Guler, *Single-*
985 *cell sequencing of the small and AT-skewed genome of malaria parasites.* Genome Med, 2021.
986 13(1): p. 75.
987 64. Galhardo, R.S., P.J. Hastings, and S.M. Rosenberg, *Mutation as a stress response and the*
988 *regulation of evolvability.* Crit Rev Biochem Mol Biol, 2007. 42(5): p. 399-435.
989 65. Arlt, M.F., T.E. Wilson, and T.W. Glover, *Replication stress and mechanisms of CNV formation.*
990 Curr Opin Genet Dev, 2012. 22(3): p. 204-10.
991 66. Glover, T.W., T.E. Wilson, and M.F. Arlt, *Fragile sites in cancer: more than meets the eye.* Nat Rev
992 Cancer, 2017. 17(8): p. 489-501.
993 67. Phillips, M.A., R. Gujjar, N.A. Malmquist, J. White, F. El Mazouni, J. Baldwin, and P.K. Rathod,
994 *Triazolopyrimidine-Based Dihydroorotate Dehydrogenase Inhibitors with Potent and Selective*
995 *Activity against the Malaria Parasite Plasmodium falciparum.* Journal of Medicinal Chemistry,
996 2008. 51(12): p. 3649-3653.
997 68. Sanz, L.M., B. Crespo, C. De-Cozar, X.C. Ding, J.L. Llergo, J.N. Burrows, J.F. Garcia-Bustos, and F.J.
998 Gamo, *P. falciparum in vitro killing rates allow to discriminate between different antimalarial*
999 *mode-of-action.* PLoS One, 2012. 7(2): p. e30949.
1000 69. Jensen, J.B. and W. Trager, *Plasmodium falciparum in culture: establishment of additional*
1001 *strains.* Am J Trop Med Hyg, 1978. 27(4): p. 743-6.
1002 70. Dolan, S.A., L.H. Miller, and T.E. Wellems, *Evidence for a switching mechanism in the invasion of*
1003 *erythrocytes by Plasmodium falciparum.* J Clin Invest, 1990. 86(2): p. 618-24.
1004 71. Tian, S., H. Yan, C. Neuhauser, and S.L. Slager, *An analytical workflow for accurate variant*
1005 *discovery in highly divergent regions.* BMC Genomics, 2016. 17(1): p. 703.
1006 72. Sanderson, N.D., J. Swann, L. Barker, J. Kavanagh, S. Hoosdally, D. Crook, G. GonFast
1007 Investigators, T.L. Street, and D.W. Eyre, *High precision Neisseria gonorrhoeae variant and*
1008 *antimicrobial resistance calling from metagenomic Nanopore sequencing.* Genome Res, 2020.
1009 30(9): p. 1354-1363.
1010 73. Yu, X., F. Qin, S. Liu, N.J. Brown, Q. Lu, G. Cai, J.L. Guler, and F. Xiao, *HapCNV: A Comprehensive*
1011 *Framework for CNV Detection in Low-input DNA Sequencing Data.* bioRxiv, 2024: p.
1012 2024.12.19.629494.
1013 74. Layer, R.M., C. Chiang, A.R. Quinlan, and I.M. Hall, *LUMPY: a probabilistic framework for*
1014 *structural variant discovery.* Genome Biol, 2014. 15(6): p. R84.
1015 75. Cameron, D.L., L. Di Stefano, and A.T. Papenfuss, *Comprehensive evaluation and*
1016 *characterisation of short read general-purpose structural variant calling software.* Nat Commun,
1017 2019. 10(1): p. 3240.
1018 76. Gabrielaite, M., M.H. Torp, M.S. Rasmussen, S. Andreu-Sanchez, F.G. Vieira, C.B. Pedersen, S.
1019 Kinalis, M.B. Madsen, M. Kodama, G.S. Demircan, A. Simonyan, C.W. Yde, L.R. Olsen, R.L. Marvig,
1020 O. Ostrup, M. Rossing, F.C. Nielsen, O. Winther, and F.O. Bagger, *A Comparison of Tools for*
1021 *Copy-Number Variation Detection in Germline Whole Exome and Whole Genome Sequencing*
1022 *Data.* Cancers (Basel), 2021. 13(24).
1023 77. Otto, T.D., U. Bohme, M. Sanders, A. Reid, E.I. Bruske, C.W. Duffy, P.C. Bull, R.D. Pearson, A.
1024 Abdi, S. Dimonte, L.B. Stewart, S. Campino, M. Kekre, W.L. Hamilton, A. Claessens, S.K. Volkman,
1025 D. Ndiaye, A. Amambua-Ngwa, M. Diakite, R.M. Fairhurst, D.J. Conway, M. Franck, C.I. Newbold,
1026 and M. Berriman, *Long read assemblies of geographically dispersed Plasmodium falciparum*
1027 *isolates reveal highly structured subtelomeres.* Wellcome Open Res, 2018. 3: p. 52.
1028 78. Saxena, S. and L. Zou, *Hallmarks of DNA replication stress.* Mol Cell, 2022. 82(12): p. 2298-2314.

1029 79. Guler, J.L., D.L. Freeman, V. Ahyong, R. Patrapuvich, J. White, R. Gujjar, M.A. Phillips, J. DeRisi,
1030 and P.K. Rathod, *Asexual populations of the human malaria parasite, Plasmodium falciparum,*
1031 *use a two-step genomic strategy to acquire accurate, beneficial DNA amplifications.* PLoS
1032 Pathog, 2013. 9(5): p. e1003375.
1033 80. Evrony, G.D., E. Lee, P.J. Park, and C.A. Walsh, *Resolving rates of mutation in the brain using*
1034 *single-neuron genomics.* Elife, 2016. 5.
1035 81. Mills, R.E., K. Walter, C. Stewart, R.E. Handsaker, K. Chen, C. Alkan, A. Abyzov, S.C. Yoon, K. Ye,
1036 R.K. Cheetham, A. Chinwalla, D.F. Conrad, Y. Fu, F. Grubert, I. Hajirasouliha, F. Hormozdiari, L.M.
1037 lakoucheva, Z. Iqbal, S. Kang, J.M. Kidd, M.K. Konkel, J. Korn, E. Khurana, D. Kural, H.Y. Lam, J.
1038 Leng, R. Li, Y. Li, C.Y. Lin, R. Luo, X.J. Mu, J. Nemesh, H.E. Peckham, T. Rausch, A. Scally, X. Shi,
1039 M.P. Stromberg, A.M. Stutz, A.E. Urban, J.A. Walker, J. Wu, Y. Zhang, Z.D. Zhang, M.A. Batzer, L.
1040 Ding, G.T. Marth, G. McVean, J. Sebat, M. Snyder, J. Wang, K. Ye, E.E. Eichler, M.B. Gerstein, M.E.
1041 Hurles, C. Lee, S.A. McCarroll, J.O. Korbel, and P. Genomes, *Mapping copy number variation by*
1042 *population-scale genome sequencing.* Nature, 2011. 470(7332): p. 59-65.
1043 82. Zhao, M., Q. Wang, Q. Wang, P. Jia, and Z. Zhao, *Computational tools for copy number variation*
1044 *(CNV) detection using next-generation sequencing data: features and perspectives.* BMC
1045 Bioinformatics, 2013. 14 Suppl 11(Suppl 11): p. S1.
1046 83. Guan, P. and W.K. Sung, *Structural variation detection using next-generation sequencing data: A*
1047 *comparative technical review.* Methods, 2016. 102: p. 36-49.
1048 84. Arnot, D.E., E. Ronander, and D.C. Bengtsson, *The progression of the intra-erythrocytic cell cycle*
1049 *of Plasmodium falciparum and the role of the centriolar plaques in asynchronous mitotic division*
1050 *during schizogony.* Int J Parasitol, 2011. 41(1): p. 71-80.
1051 85. Zong, C., S. Lu, A.R. Chapman, and X.S. Xie, *Genome-wide detection of single-nucleotide and*
1052 *copy-number variations of a single human cell.* Science, 2012. 338(6114): p. 1622-6.
1053 86. Lasken, R.S., *Single-cell sequencing in its prime.* Nat Biotechnol, 2013. 31(3): p. 211-2.
1054 87. Liu, S., E.R. Ebel, A. Luniewski, J. Zulawinska, M.L. Simpson, J. Kim, N. Ene, T.W.A. Braukmann, M.
1055 Congdon, W. Santos, E. Yeh, and J.L. Guler, *Direct long read visualization reveals metabolic*
1056 *interplay between two antimalarial drug targets.* bioRxiv, 2023.
1057 88. Kirkman, L.A., E.A. Lawrence, and K.W. Deitsch, *Malaria parasites utilize both homologous*
1058 *recombination and alternative end joining pathways to maintain genome integrity.* Nucleic Acids
1059 Res, 2014. 42(1): p. 370-9.
1060 89. Triglia, T., S.J. Foote, D.J. Kemp, and A.F. Cowman, *Amplification of the multidrug resistance*
1061 *gene pfmdr1 in Plasmodium falciparum has arisen as multiple independent events.* Mol. Cell.
1062 Biol., 1991. 11(10): p. 5244-5250.
1063 90. Luth, M.R., K.P. Godinez-Macias, D. Chen, J. Okombo, V. Thathy, X. Cheng, S. Daggupati, H.
1064 Davies, S.K. Dhingra, J.M. Economy, R.C.S. Edgar, M.G. Gomez-Lorenzo, E.S. Istvan, J.C. Jado,
1065 G.M. LaMonte, B. Melillo, S. Mok, S.K. Narwal, T. Ndiaye, S. Ottilie, S. Palomo Diaz, H. Park, S.
1066 Pena, F. Rocamora, T. Sakata-Kato, J.L. Small-Saunders, R.L. Summers, P.K. Tumwebaze, M.
1067 Vanaerschot, G. Xia, T. Yeo, A. You, F.J. Gamo, D.E. Goldberg, M.C.S. Lee, C.W. McNamara, D.
1068 Ndiaye, P.J. Rosenthal, S.L. Schreiber, G. Serra, J.L. De Siqueira-Neto, T.S. Skinner-Adams, A.C.
1069 Uhlemann, N. Kato, A.K. Lukens, D.F. Wirth, D.A. Fidock, and E.A. Winzeler, *Systematic in vitro*
1070 *evolution in Plasmodium falciparum reveals key determinants of drug resistance.* Science, 2024.
1071 386(6725): p. eadk9893.
1072 91. Matthews, H., C.W. Duffy, and C.J. Merrick, *Checks and balances? DNA replication and the cell*
1073 *cycle in Plasmodium.* Parasit Vectors, 2018. 11(1): p. 216.
1074 92. Bindra, R.S., P.J. Schaffer, A. Meng, J. Woo, K. Maseide, M.E. Roth, P. Lizardi, D.W. Hedley, R.G.
1075 Bristow, and P.M. Glazer, *Down-regulation of Rad51 and decreased homologous recombination*
1076 *in hypoxic cancer cells.* Molecular and Cellular Biology, 2004. 24(19): p. 8504-8518.

1077 93. Saleh-Gohari, N., H.E. Bryant, N. Schultz, K.M. Parker, T.N. Cassel, and T. Helleday, *Spontaneous*
1078 *homologous recombination is induced by collapsed replication forks that are caused by*
1079 *endogenous DNA single-strand breaks.* Mol Cell Biol, 2005. 25(16): p. 7158-69.
1080 94. Hastings, P.J., G. Ira, and J.R. Lupski, *A microhomology-mediated break-induced replication*
1081 *model for the origin of human copy number variation.* PLoS Genet, 2009. 5(1): p. e1000327.
1082 95. Durkin, S.G., R.L. Ragland, M.F. Arlt, J.G. Mulle, S.T. Warren, and T.W. Glover, *Replication stress*
1083 *induces tumor-like microdeletions in FHIT/FRA3B.* Proc Natl Acad Sci U S A, 2008. 105(1): p. 246-
1084 51.
1085 96. Chen, L., W. Zhou, C. Zhang, J.R. Lupski, L. Jin, and F. Zhang, *CNV instability associated with DNA*
1086 *replication dynamics: evidence for replicative mechanisms in CNV mutagenesis.* Hum Mol Genet,
1087 2015. 24(6): p. 1574-83.
1088 97. White, N.J., *Malaria parasite clearance.* Malar J, 2017. 16(1): p. 88.
1089 98. Shor, E., C.A. Fox, and J.R. Broach, *The yeast environmental stress response regulates*
1090 *mutagenesis induced by proteotoxic stress.* PLoS Genet, 2013. 9(8): p. e1003680.
1091 99. WHO, *World Malaria Report 2023.* 2023.
1092 100. Okombo, J. and D.A. Fidock, *Towards next-generation treatment options to combat Plasmodium*
1093 *falciparum malaria.* Nat Rev Microbiol, 2024.
1094 101. Wilson, C., A. Serrano, A. Wasley, M. Bogenschutz, A. Shankar, and D. Wirth, *Amplification of a*
1095 *gene related to mammalian mdr genes in drug-resistant Plasmodium falciparum.* Science, 1989.
1096 244(4909): p. 1184-1186.
1097 102. Nair, S., D. Nash, D. Sudimack, A. Jaidee, M. Barends, A.-C. Uhlemann, S. Krishna, F. Nosten, and
1098 T.J.C. Anderson, *Recurrent Gene Amplification and Soft Selective Sweeps during Evolution of*
1099 *Multidrug Resistance in Malaria Parasites.* Mol Biol Evol, 2007. 24(2): p. 562-573.
1100 103. Ravenhall, M., E.D. Benavente, M. Mipando, A.T. Jensen, C.J. Sutherland, C. Roper, N. Sepulveda,
1101 D.P. Kwiatkowski, J. Montgomery, K.S. Phiri, A. Terlouw, A. Craig, S. Campino, H. Ocholla, and
1102 T.G. Clark, *Characterizing the impact of sustained sulfadoxine/pyrimethamine use upon the*
1103 *Plasmodium falciparum population in Malawi.* Malar J, 2016. 15(1): p. 575.
1104 104. Leroy, D., F. Macintyre, Y. Adoke, S. Ouoba, A. Barry, G. Mombo-Ngoma, J.M. Ndong Ngomo, R.
1105 Varo, Y. Dossou, A.K. Tshefu, T.T. Duong, B.Q. Phuc, B. Laurijssens, R. Klopper, N. Khim, E.
1106 Legrand, and D. Menard, *African isolates show a high proportion of multiple copies of the*
1107 *Plasmodium falciparum plasmepsin-2 gene, a piperaquine resistance marker.* Malar J, 2019.
1108 18(1): p. 126.
1109 105. Bopp, S., P. Magistrado, W. Wong, S.F. Schaffner, A. Mukherjee, P. Lim, M. Dhorda, C.
1110 Amaratunga, C.J. Woodrow, E.A. Ashley, N.J. White, A.M. Dondorp, R.M. Fairhurst, F. Ariey, D.
1111 Menard, D.F. Wirth, and S.K. Volkman, *Plasmepsin II-III copy number accounts for bimodal*
1112 *piperaquine resistance among Cambodian Plasmodium falciparum.* Nat Commun, 2018. 9(1): p.
1113 1769.
1114 106. Dharia, N., A. Sidhu, M. Cassera, S. Westenberger, S. Bopp, R. Eastman, D. Plouffe, S. Batalov, D.
1115 Park, S. Volkman, D. Wirth, Y. Zhou, D. Fidock, and E. Winzeler, *Use of high-density tiling*
1116 *microarrays to identify mutations globally and elucidate mechanisms of drug resistance in*
1117 *Plasmodium falciparum.* Genome Biology, 2009. 10(2): p. R21.
1118 107. Rottmann, M., C. McNamara, B.K. Yeung, M.C. Lee, B. Zou, B. Russell, P. Seitz, D.M. Plouffe, N.V.
1119 Dharia, J. Tan, S.B. Cohen, K.R. Spencer, G.E. Gonzalez-Paez, S.B. Lakshminarayana, A. Goh, R.
1120 Suwanarusk, T. Jegla, E.K. Schmitt, H.P. Beck, R. Brun, F. Nosten, L. Renia, V. Dartois, T.H. Keller,
1121 D.A. Fidock, E.A. Winzeler, and T.T. Diagana, *Spiroindolones, a potent compound class for the*
1122 *treatment of malaria.* Science, 2010. 329(5996): p. 1175-80.
1123 108. Ross, L.S., F.J. Gamo, M.J. Lafuente-Monasterio, O.M. Singh, P. Rowland, R.C. Wiegand, and D.F.
1124 Wirth, *In Vitro Resistance Selections for Plasmodium falciparum Dihydroorotate Dehydrogenase*

1125 *Inhibitors Give Mutants with Multiple Point Mutations in the Drug-binding Site and Altered*
1126 *Growth.* J Biol Chem, 2014. 289(26): p. 17980-95.
1127 109. Phillips, M.A., J. Lotharius, K. Marsh, J. White, A. Dayan, K.L. White, J.W. Njoroge, F. El Mazouni,
1128 Y. Lao, S. Kokkonda, D.R. Tomchick, X. Deng, T. Laird, S.N. Bhatia, S. March, C.L. Ng, D.A. Fidock,
1129 S. Wittlin, M. Lafuente-Monasterio, F.J. Benito, L.M. Alonso, M.S. Martinez, M.B. Jimenez-Diaz,
1130 S.F. Bazaga, I. Angulo-Barturen, J.N. Haselden, J. Louttit, Y. Cui, A. Sridhar, A.M. Zeeman, C.
1131 Kocken, R. Sauerwein, K. Dechering, V.M. Avery, S. Duffy, M. Delves, R. Sinden, A. Ruecker, K.S.
1132 Wickham, R. Rochford, J. Gahagen, L. Iyer, E. Riccio, J. Mirsalis, I. Bathhurst, T. Rueckle, X. Ding,
1133 B. Campo, D. Leroy, M.J. Rogers, P.K. Rathod, J.N. Burrows, and S.A. Charman, *A long-duration*
1134 *dihydroorotate dehydrogenase inhibitor (DSM265) for prevention and treatment of malaria.* Sci
1135 Transl Med, 2015. 7(296): p. 296ra111.
1136 110. Cowell, A. and E. Winzeler, *Exploration of the Plasmodium falciparum Resistome and Druggable*
1137 *Genome Reveals New Mechanisms of Drug Resistance and Antimalarial Targets.* Microbiol
1138 Insights, 2018. 11: p. 1178636118808529.
1139 111. Rocamora, F., L. Zhu, K.Y. Liong, A. Dondorp, O. Miotto, S. Mok, and Z. Bozdech, *Oxidative stress*
1140 *and protein damage responses mediate artemisinin resistance in malaria parasites.* PLoS Pathog,
1141 2018. 14(3): p. e1006930.
1142 112. Thaithong, S., L.C. Ranford-Cartwright, N. Siripoon, P. Harnyuttanakorn, N.S. Kanchanakhan, A.
1143 Seugorn, K. Rungsihirunrat, P.V. Cravo, and G.H. Beale, *Plasmodium falciparum: gene mutations*
1144 *and amplification of dihydrofolate reductase genes in parasites grown in vitro in presence of*
1145 *pyrimethamine.* Exp Parasitol, 2001. 98(2): p. 59-70.
1146 113. Istvan, E.S., N.V. Dharia, S.E. Bopp, I. Gluzman, E.A. Winzeler, and D.E. Goldberg, *Validation of*
1147 *isoleucine utilization targets in Plasmodium falciparum.* Proc Natl Acad Sci U S A, 2011. 108(4): p.
1148 1627-32.
1149 114. Miles, A., Z. Iqbal, P. Vauterin, R. Pearson, S. Campino, M. Theron, K. Gould, D. Mead, E. Drury, J.
1150 O'Brien, V. Ruano Rubio, B. MacInnis, J. Mwangi, U. Samarakoon, L. Ranford-Cartwright, M.
1151 Ferdig, K. Hayton, X.Z. Su, T. Wellems, J. Rayner, G. McVean, and D. Kwiatkowski, *Indels,*
1152 *structural variation, and recombination drive genomic diversity in Plasmodium falciparum.*
1153 Genome Res, 2016. 26(9): p. 1288-99.
1154 115. Rathod, P.K., T. McErlean, and P.C. Lee, *Variations in frequencies of drug resistance in*
1155 *Plasmodium falciparum.* Proc Natl Acad Sci U S A, 1997. 94(17): p. 9389-93.
1156 116. McDaniels, J.M., A.C. Huckaby, S.A. Carter, S. Lingeman, A. Francis, M. Congdon, W. Santos, P.K.
1157 Rathod, and J.L. Guler, *Extrachromosomal DNA amplicons in antimalarial-resistant Plasmodium*
1158 *falciparum.* Mol Microbiol, 2021. 115(4): p. 574-590.
1159 117. Sugino, A. and K. Nakayama, *DNA polymerase alpha mutants from a Drosophila melanogaster*
1160 *cell line.* Proc Natl Acad Sci U S A, 1980. 77(12): p. 7049-53.
1161 118. Inselburg, J. and H.S. Banyal, *Plasmodium falciparum: synchronization of asexual development*
1162 *with aphidicolin, a DNA synthesis inhibitor.* Exp Parasitol, 1984. 57(1): p. 48-54.
1163 119. Brown, A.C., C.C. Moore, and J.L. Guler, *Cholesterol-dependent enrichment of understudied*
1164 *erythrocytic stages of human Plasmodium parasites.* Sci Rep, 2020. 10(1): p. 4591.
1165 120. Rodrigues, O.R. and S. Monard, *A rapid method to verify single-cell deposition setup for cell*
1166 *sorters.* Cytometry A, 2016. 89(6): p. 594-600.
1167 121. Nam, D., S. Kim, J.H. Kim, S. Lee, D. Kim, J. Son, D. Kim, B.S. Cha, E.S. Lee, and K.S. Park, *Low-*
1168 *Temperature Loop-Mediated Isothermal Amplification Operating at Physiological Temperature.*
1169 Biosensors (Basel), 2023. 13(3).
1170 122. Brown, N., C. da Silva, C. Webb, D. Matias, B. Dias, B. Cancio, M. Silva, R. Viegas, C. Salvador, N.
1171 Chivale, S. Luis, P. Arnaldo, J. Zulawinska, C.C. Moore, F. Nogueira, and J.L. Guler, *Antimalarial*

1172 *resistance risk in Mozambique detected by a novel quadruplex droplet digital PCR assay.*
1173 Antimicrob Agents Chemother, 2024. 68(7): p. e0034624.
1174 123. Daniels, R., D. Ndiaye, M. Wall, J. McKinney, P.D. Sene, P.C. Sabeti, S.K. Volkman, S. Mboup, and
1175 D.F. Wirth, *Rapid, field-deployable method for genotyping and discovery of single-nucleotide*
1176 *polymorphisms associated with drug resistance in Plasmodium falciparum.* Antimicrob Agents
1177 Chemother, 2012. 56(6): p. 2976-86.
1178 124. Kassaza, K., A.C. Long, J.M. McDaniels, M. Andre, W. Fredrickson, D. Nyehangane, P. Orikiriza,
1179 D.J. Operario, J. Bazira, J.A. Mwanga-Amumpaire, C.C. Moore, J.L. Guler, and Y. Boum, 2nd,
1180 *Surveillance of Plasmodium falciparum pfcrt haplotypes in southwestern uganda by high-*
1181 *resolution melt analysis.* Malar J, 2021. 20(1): p. 114.
1182 125. Alker, A.P., V. Mwapasa, and S.R. Meshnick, *Rapid real-time PCR genotyping of mutations*
1183 *associated with sulfadoxine-pyrimethamine resistance in Plasmodium falciparum.* Antimicrob
1184 Agents Chemother, 2004. 48(8): p. 2924-9.
1185 126. Chiang, C., R.M. Layer, G.G. Faust, M.R. Lindberg, D.B. Rose, E.P. Garrison, G.T. Marth, A.R.
1186 Quinlan, and I.M. Hall, *SpeedSeq: ultra-fast personal genome analysis and interpretation.* Nat
1187 Methods, 2015. 12(10): p. 966-8.
1188 127. Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin,
1189 and S. Genome Project Data Processing, *The Sequence Alignment/Map format and SAMtools.*
1190 Bioinformatics, 2009. 25(16): p. 2078-9.
1191 128. Garcia-Alcalde, F., K. Okonechnikov, J. Carbonell, L.M. Cruz, S. Gotz, S. Tarazona, J. Dopazo, T.F.
1192 Meyer, and A. Conesa, *Qualimap: evaluating next-generation sequencing alignment data.*
1193 Bioinformatics, 2012. 28(20): p. 2678-9.
1194 129. Huang, L., F. Ma, A. Chapman, S. Lu, and X.S. Xie, *Single-Cell Whole-Genome Amplification and*
1195 *Sequencing: Methodology and Applications.* Annu Rev Genomics Hum Genet, 2015. 16: p. 79-
1196 102.
1197 130. Chen, C., D. Xing, L. Tan, H. Li, G. Zhou, L. Huang, and X.S. Xie, *Single-cell whole-genome analyses*
1198 *by Linear Amplification via Transposon Insertion (LIANTI).* Science, 2017. 356(6334): p. 189-194.
1199 131. McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D.
1200 Altshuler, S. Gabriel, M. Daly, and M.A. DePristo, *The Genome Analysis Toolkit: a MapReduce*
1201 *framework for analyzing next-generation DNA sequencing data.* Genome Res, 2010. 20(9): p.
1202 1297-303.
1203 132. DePristo, M.A., E. Banks, R. Poplin, K.V. Garimella, J.R. Maguire, C. Hartl, A.A. Philippakis, G. del
1204 Angel, M.A. Rivas, M. Hanna, A. McKenna, T.J. Fennell, A.M. Kernytsky, A.Y. Sivachenko, K.
1205 Cibulskis, S.B. Gabriel, D. Altshuler, and M.J. Daly, *A framework for variation discovery and*
1206 *genotyping using next-generation DNA sequencing data.* Nat Genet, 2011. 43(5): p. 491-8.
1207 133. Van der Auwera, G.A., M.O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-Moonshine, T.
1208 Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K.V. Garimella, D. Altshuler, S. Gabriel, and
1209 M.A. DePristo, *From FastQ data to high confidence variant calls: the Genome Analysis Toolkit*
1210 *best practices pipeline.* Curr Protoc Bioinformatics, 2013. 43(1110): p. 11 10 1-11 10 33.
1211 134. MalariaGen, A. Ahouidi, M. Ali, J. Almagro-Garcia, A. Amambua-Ngwa, C. Amaratunga, R. Amato,
1212 L. Amenga-Etego, B. Andagalu, T.J.C. Anderson, V. Andrianaranjaka, T. Apinjoh, C. Ariani, E.A.
1213 Ashley, S. Auburn, G.A. Awandare, H. Ba, V. Baraka, A.E. Barry, P. Bejon, G.I. Bertin, M.F. Boni, S.
1214 Borrmann, T. Bousema, O. Branch, P.C. Bull, G.B.J. Busby, T. Chookajorn, K. Chotivanich, A.
1215 Claessens, D. Conway, A. Craig, U. D'Alessandro, S. Dama, N.P. Day, B. Denis, M. Diakite, A.
1216 Djimde, C. Dolecek, A.M. Dondorp, C. Drakeley, E. Drury, P. Duffy, D.F. Echeverry, T.G. Egwang,
1217 B. Erko, R.M. Fairhurst, A. Faiz, C.A. Fanello, M.M. Fukuda, D. Gamboa, A. Ghansah, L. Golassa, S.
1218 Goncalves, W.L. Hamilton, G.L.A. Harrison, L. Hart, C. Henrichs, T.T. Hien, C.A. Hill, A. Hodgson, C.
1219 Hubbart, M. Imwong, D.S. Ishengoma, S.A. Jackson, C.G. Jacob, B. Jeffery, A.E. Jeffreys, K.J.

1220 Johnson, D. Jyothi, C. Kamaliddin, E. Kamau, M. Kekre, K. Kluczynski, T. Kochakarn, A. Konate,
1221 D.P. Kwiatkowski, M.P. Kyaw, P. Lim, C. Lon, K.M. Loua, O. Maiga-Ascofare, C. Malangone, M.
1222 Manske, J. Marfurt, K. Marsh, M. Mayxay, A. Miles, O. Miotto, V. Mobegi, O.A. Mokuolu, J.
1223 Montgomery, I. Mueller, P.N. Newton, T. Nguyen, T.N. Nguyen, H. Noedl, F. Nosten, R.
1224 Noviyanti, A. Nzila, L.I. Ochola-Oyier, H. Ocholla, A. Oduro, I. Omedo, M.A. Onyamboko, J.B.
1225 Ouedraogo, K. Oyebola, R.D. Pearson, N. Peshu, A.P. Phyo, C.V. Plowe, R.N. Price, S.
1226 Pukrittayakamee, M. Randrianarivelojosia, J.C. Rayner, P. Ringwald, K.A. Rockett, K. Rowlands, L.
1227 Ruiz, D. Saunders, A. Shayo, P. Siba, V.J. Simpson, J. Stalker, X.Z. Su, C. Sutherland, S. Takala-
1228 Harrison, L. Tavul, V. Thathy, A. Tshefu, F. Verra, J. Vinetz, T.E. Wellems, J. Wendler, N.J. White, I.
1229 Wright, W. Yavo and H. Ye, *An open dataset of Plasmodium falciparum genome variation in*
1230 *7,000 worldwide samples.* Wellcome Open Res, 2021. 6: p. 42.
1231 135. Abyzov, A., A.E. Urban, M. Snyder, and M. Gerstein, *CNVnator: an approach to discover,*
1232 *genotype, and characterize typical and atypical CNVs from family and population genome*
1233 *sequencing.* Genome Res, 2011. 21(6): p. 974-84.
1234 136. Venkatraman, E.S. and A.B. Olshen, *A faster circular binary segmentation algorithm for the*
1235 *analysis of array CGH data.* Bioinformatics, 2007. 23(6): p. 657-63.
1236 137. Xiao, F., Y. Niu, N. Hao, Y. Xu, Z. Jin, and H. Zhang, *modSaRa: a computationally efficient R*
1237 *package for CNV identification.* Bioinformatics, 2017. 33(15): p. 2384-2385.
1238 138. Mi, H., D. Ebert, A. Muruganujan, C. Mills, L.P. Albou, T. Mushayamaha, and P.D. Thomas,
1239 *PANTHER version 16: a revised family classification, tree-based classification tool, enhancer*
1240 *regions and extensive API.* Nucleic Acids Res, 2021. 49(D1): p. D394-D403.
1241 139. Singh, G. and D. Gupta, *In-Silico Functional Annotation of Plasmodium falciparum Hypothetical*
1242 *Proteins to Identify Novel Drug Targets.* Front Genet, 2022. 13: p. 821516.

1243