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

We developed a robust pipeline for assessing the frequency of de novo CNVs in the *Plasmodium* genome based on our prior studies (**Fig. 1**). We adapted our single cell genomics method to improve parasite

isolation and whole genome amplification steps ([63], **Fig. 1A, Fig. S1**). In this modified protocol, we used fluorescence-activated cell sorting (FACS) to isolate parasites to improve efficiency and modified aspects of our whole genome amplification method to improve coverage (**Table S1**, **Fig. S1**). We also sorted low-cell populations to increase accuracy (e.g. 2-cells per well, **Fig. S2**), and added quality control steps that confirmed our samples were of high quality prior to short-read sequencing. The resulting low-input genomics pipeline consisted of basic steps including parasite isolation using flow sorting, whole genome amplification using a modified MALBAC-based approach, quality control confirmation, short-read sequencing, and CNV analysis (**Fig. 1A**).

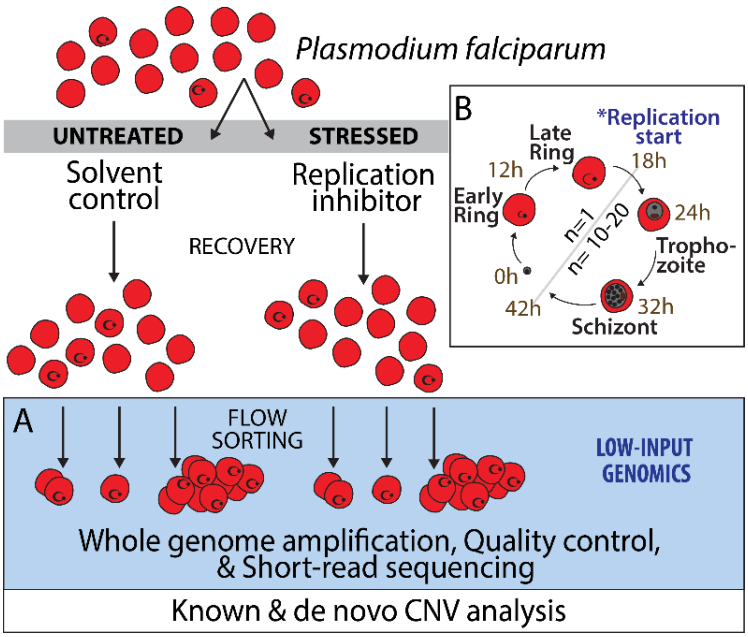

Figure 1. Low-input genomics approach for analysis of malaria parasite genomes under stress. Early erythrocytic stage *P. falciparum* parasites grown in vitro were treated with a replication inhibitor (DSM1) or the solvent control (dimethylsulfoxide) followed by recovery and reinvasion to produce a new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

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

To explore the impact of replication stress on de novo CNV generation, we treated the parasites with an antimalarial that inhibits DNA replication by limiting pyrimidine pools; DSM1 targets *Plasmodium* the dihydroorotate dehydrogenase enzyme [67]. Application of DSM1 for an extended period kills parasites by perturbing DNA synthesis (>48hrs at 10x the EC_{50}, [68]). However, short-term treatment reversibly stalls parasite replication similar to another replication inhibitor, aphidicolin (**Fig. S3**). For low-input genomics, we applied DSM1 to ring-stage parasites for a brief time (12hr, **Fig. S4A and B**). Similar to our

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

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

139

140 **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%

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

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

149 We sorted low-cell populations from each parasite group into 60 wells of a 96-well plate (*FCR3*, untreated
150 *Dd2,* and treated *Dd2*), including 6 wells with 10-cells and 56 wells with 2-cells. Ten-cell wells served as
151 positive controls for the whole genome amplification step and 2-cell wells provided the optimal balance
152 between sorting accuracy (**Fig. S2**) and de novo CNV detection. Zero cells were sorted into the top 2 rows
153 of the plate ("no-cell" wells). After parasite lysis and applying *Pf*MALBAC version-2 whole genome
154 amplification (**Fig. S1**), we assessed amplification success using three approaches. First, we measured the
155 resulting DNA quantity across 80% of the amplified wells (**Fig. S6**). On average, MALBAC amplification in
156 wells that contained sorted parasites yielded ~120ng of total DNA per reaction, with a ~10% increase in
157 DNA for 10- vs 2-cell samples (mean of 127ng versus 116ng total, respectively). We did not detect position-
158 based bias across plates or appreciable amplification from no-cell wells, but we did observe that treated
159 *Dd2* wells had ~3-fold lower levels of amplification than other samples (mean of 51ng versus 151ng total,
160 respectively). There was little difference in mean amplified DNA amounts between the two untreated
161 sample groups (*FCR3* at 146ng and untreated-*Dd2* at 152ng).

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

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

172

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

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

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

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

199

200 **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

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

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

For the current study, we employed two different CNV calling methods in low-input samples. HapCNV is a novel read coverage-based CNV calling method specifically designed for low cell data from haploid genomes. In contrast to traditional methods that arbitrarily select reference samples for CNV data normalization, HapCNV constructs a genomic location (or bin)-based pseudo-reference as a comparison baseline. This step systematically alleviates amplification bias for the identification of de novo CNVs. LUMPY is an established CNV calling method that exhibits high sensitivity due to the incorporation of multiple CNV signals (i.e. split and discordant reads) generated from short-read sequencing. It is particularly well-suited for detecting low-frequency variants in low-coverage datasets; however, for many CNV callers, high sensitivity leads to higher false positives [73-75].

Using these two distinct CNV calling methods, combined with a recently developed CNV counting approach, we evaluated the presence of known CNVs in our 2-cell samples (**Fig. 2A**). The identification of known CNVs (i.e. those identified in the bulk sample, see *Materials and Methods*) within low-input samples displays the utility of the specific CNV calling method for different size CNVs in specific genome locations. In our previous study, we identified 2 of the 3 known CNVs in ~10% of single cell genomes [63]. In the current study, we identified the *pfmdr1* amplicon in 100% of 2-cell samples using HapCNV (57/57) and 79% of samples using LUMPY (45/57). We did not identify the *pf11-1* amplicon in any 2-cell samples using HapCNV (0/57) but detected this locus in 75% of samples using LUMPY (43/57). Finally, we identified *pf332* CNVs in 46% of 2-cell samples using HapCNV (26/57) and 100% of samples using LUMPY (57/57). Although our two studies used different CNV calling methods and are not directly comparable, the overall improvement in the detection of known CNVs in the current study is likely due to advances in both the whole genome amplification method to limit amplification bias (**Table S1**) and recently developed analyses approaches. When we evaluated the detection of three known CNVs in *Dd2* low-input samples,

230 we observed a somewhat higher rate of known CNV detection by either method in treated samples (**Fig.**
231 **2B**, HapCNV: mean of 1.2 out of 3 total CNVs for untreated and 1.7 for treated (increase of 42%), LUMPY:
232 mean of 2.1 out of 3 total CNVs for untreated and 2.9 for treated (increase of 38%, **Table S4**).

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

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

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

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

258 When we compared de novo CNVs in genomes with and without replication stress, we found that results
259 were consistent regardless of the CNV calling method (**Fig. 2**). While the proportion of duplication and
260 deletions did not change appreciably with treatment (**Fig. 2C**), we identified a highly significant difference
261 in de novo CNVs between treated and untreated 2-cell samples (p value of 0.0002 for HapCNV and <0.0001
262 for LUMPY, **Fig. 2D**). This pattern was consistent when we down-sampled all sequencing data (p value of
263 0.001 for HapCNV and <0.0001 for LUMPY, **Fig. 2E**), indicating that the difference in de novo CNV counts
264 between treatments was not due to read coverage. When we assessed common and rare CNV categories,
265 we once again observed highly significant differences between treated and untreated 2-cell samples in
266 common CNVs using both methods (common: p value 0.0004 for HapCNV and <0.0001 for LUMPY, **Fig.**
267 **2F;** rare: p value of 0.008 for HapCNV and <0.0001 for LUMPY, **Fig. 2G**).

268 When we compared the proportion of de novo CNVs relative to total CNVs per sample, rare CNVs were
269 significantly increased over common CNVs (p value of 0.003 for HapCNV and 0.009 for LUMPY, **Fig. 2H**).
270 This difference persisted regardless of down-sampling (p value of 0.02 for HapCNV and 0.004 for LUMPY,

271 **Fig. S11**) and is in line with our assessment above that rare CNVs are more likely to be novel in nature (**Fig.**
272 **S10**). Overall, we detected a ~2-3-fold increase of de novo CNVs in treated samples, regardless of the CNV
273 calling method (**Table 3**). Once again, rare CNVs displayed the largest increase following treatment (~3-4-
274 fold, **Table 3, Fig. 2I**).

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

276 When we compared overlaps between the HapCNV and LUMPY (**Fig. 3A**), we detected a set of CNV regions
277 that was consistent within sample groups (5 for untreated and 38 for treated, **Table S5**). The frequency of
278 these "high-confidence" CNVs also reflected the increase in CNVs following replication stress (with
279 knowns excluded, ~15-fold increase in treated samples). Overall, high-confidence CNVs were located on
280 the majority of chromosomes (12 of 14, **Fig. 3B**) and represented both duplications and deletions (**Fig.**
281 **3C**). Of note, approximately half of these regions were identified as "rare" by both CNV calling methods
282 across treated samples (17/36, 47%; **Table S5**), indicating that novel CNVs were stimulated in parasite
283 genomes under stress. When we searched for genes that are covered by these regions, we identified 26
284 genes (across 3 de novo CNV regions) and 198 genes (across 37 de novo CNV regions) in untreated and
285 treated *Dd2* samples, respectively (**Table S5**). Emphasizing their random nature, genes encompassed by
286 the CNV regions represented diverse protein classes (**Fig. 3D**) and no gene ontology (GO) categories were
287 significantly enriched (using an FDR adjusted p value of 0.05, **Table S6**).

288 Finally, we evaluated whether CNVs that arose under stress were important for parasite survival by
289 comparing genes covered by our treated CNV regions to those from the largest catalogue of clinically
290 relevant CNVs to date. This list of high frequency CNVs was previously called using genomes from 2855
291 parasite isolates from 21 malaria-endemic countries and represented genes from larger (>300bp), high-
292 quality, core genome variants [59]. Assuming ~5000 genes in the core *P. falciparum* genome [76], we
293 found that genes from the two lists were ~2-times more likely to overlap than by chance (chi-square odds
294 ratio of 2.5 for HapCNV-treated CNV list and 1.6 for LUMPY-treated CNV list). This significant association
295 between CNV locations suggests that stress-induced de novo CNVs have the potential to be beneficial in
296 the clinical environment (Fisher's exact p value <0.0001 for HapCNV-treated CNV list and 0.03 for LUMPY-
297 treated CNV list).

298

299 **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.

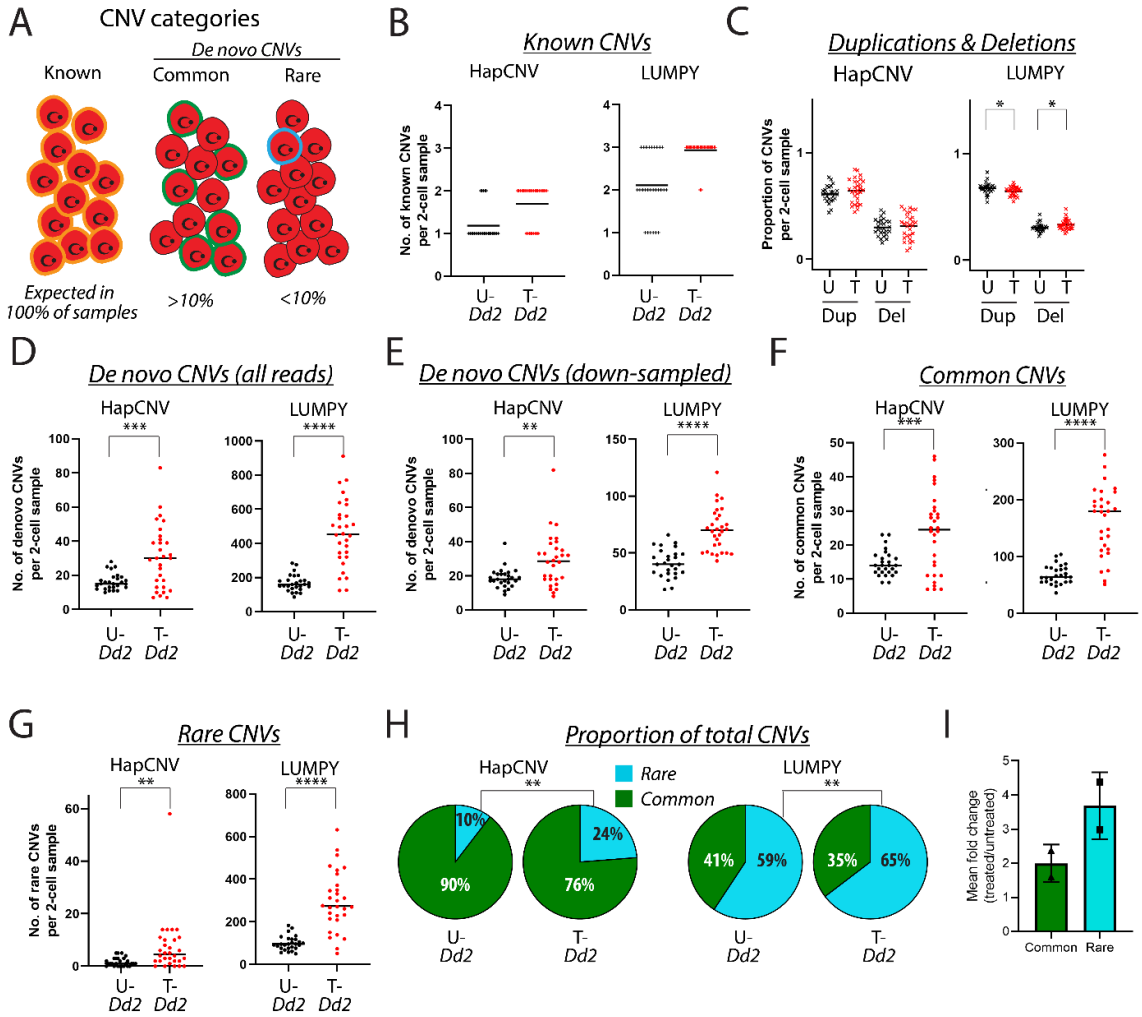

Figure 2. Low cell genomics displays an increase in de novo CNVs following replication stress. Number of CNVs from untreated (U-*Dd2*) and treated (T-*Dd2*) 2-cell samples from two CNV analysis methods: HapCNV and LUMPY. Statistics for all plots use an unpaired T-test with two tailed Welch's correction (****: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**).

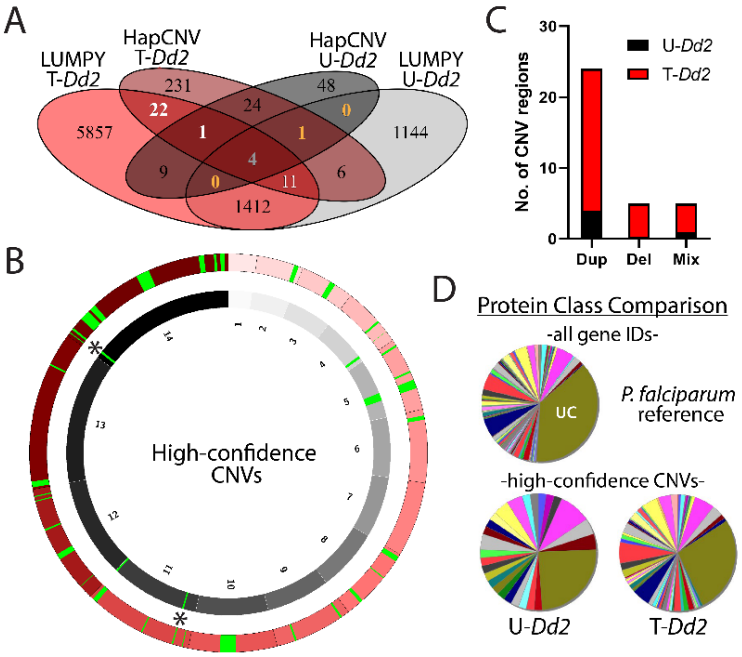

Figure 3: High-confidence CNVs are located across the genome and represent diverse protein classes. A. Comparison of CNV calls showing the number of CNV regions consistent across the two CNV calling methods. High-confidence CNVs in untreated samples (U-*Dd2*, in yellow text); high-confidence CNVs in 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:

De novo CNVs are not detected when analyzing a population of parasites predominantly because their signal (e.g. extra reads that align to that region or reads that span breakpoints) is negated by the overwhelming signal from normal copy number at that genome location. For this reason, assessments of fewer cells are necessary to investigate de novo CNV generation. Here, we optimized a low-input analysis pipeline and successfully isolated, amplified, and sequenced *P. falciparum* samples. With experimental and computational improvements, we were able to increase our rate of parental, or “known”, CNV calling over our prior study [63]. Importantly, we detected non-parental, or “de novo“, CNVs across the parasite genome and replication stress significantly increased their rate formation. By analyzing ~45 low-input

347 samples per condition, we are limited to observing a subset of the population. However, our findings
348 demonstrate that replication stress readily drives the rapid generation of de novo CNVs. Below we cover
349 how this study contributes to our understanding of genome evolution and integrate it with an overarching
350 model of *P. falciparum* adaptation.

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

352 To evaluate the impact of replication stress on de novo CNV generation, we applied sub-lethal treatment
353 to parasites just prior to replication. We observed replication stall and then resume post-treatment, which
354 provided evidence that we successfully applied non-lethal stress (**Fig. S3 and S4**). Following this step, we
355 allowed the parasites to complete replication and reinvade new erythrocytes. We reasoned that this
356 "recovery phase" enabled the repair of the resulting DNA damage, which is likely to be replication-
357 dependent (reviewed in [1, 77]. Additionally, reinvasion facilitated the isolation of haploid parasite
358 genomes (1n, **Fig. 1B** and **S4B**), encouraging the detection of de novo CNVs due to limited contrasting
359 signal [36]. Because of the reinvasion step, which involves an expansion in parasite number (~3-fold, **Fig.**
360 **S4G**), there was a potential to select for beneficial DNA changes across the population of parasites.
361 However, we did not detect evidence of strong selection from SNP profiles (**Fig. S9A**) or high-confidence
362 CNV regions (**Fig. 3**, **Table S5**). Specifically, we did not observe a preference for CNVs that encompassed
363 the dihydroorotate dehydrogenase gene (*pfdhodh*), which contributes directly to DSM1 resistance [78],
364 or enrichment of CNVs that include genes from DNA-related pathways (**Table S6**).

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

366 De novo CNV estimates using single cell methods from neurons, yeast, and human cancer vary greatly and
367 are difficult to standardize due to the use of different experimental techniques and CNV calling methods
368 [33, 34, 79]. For these reasons, we are not attempting to compare the rate of *P. falciparum* CNV formation
369 from this study to those from other organisms. Additionally, this lack of standardization in the field led us
370 to use multiple CNV calling methods in our analysis. Due to the strengths and weaknesses of HapCNV and
371 LUMPY, we observed differences in both known (**Fig. 2B**) and de novo CNV (**Table S4**) calling using the two
372 methods. LUMPY identifies reads that cover breakpoint regions to sensitively detect CNVs [73]; because
373 we are counting regions with few reads as support, both sensitivity and the number of false positives are
374 high in this analysis. High known calling rates along with high numbers of de novo CNVs in our studies
375 exemplified this feature of LUMPY. On the other hand, HapCNV uses a genome-specific pseudo-reference
376 for normalization, which removes repeated patterns of over-and under-amplification (see *Materials and*
377 *Methods* for details); because we require read coverage to span 3 consecutive 1kb bins, small CNVs in
378 lower coverage genomic neighborhoods are excluded in HapCNV anaylsis. This limited the detection of
379 smaller known amplicons (*pf11-1* and *pf332*) and led to fewer de novo CNV calls using this method. Given
380 the high abundance of small CNVs (<300bp) in the parasite genome [59, 80], HapCNV is likely
381 underestimating their impact in our studies.

382 Ultimately, the value of our study is in the relative comparison of treated and untreated samples by both
383 methods. In this case, we assume that false positive CNVs occur at a similar rate across both groups, which
384 allowed us to confidently assess the impact of stress. Variations in known and de novo CNV calling
385 described above serve to remind us that no CNV calling method is perfect and combining them can
386 improve confidence in results [63, 75, 81, 82]. Therefore, we investigated de novo CNV patterns using the
387 two individual methods (**Figs. 2**, **S10**, **S11**) as well as those that overlapped between HapCNV and LUMPY
388 (**Fig. 3**, **Table S5**). Importantly, these "high-confidence" CNVs reflected increases after stress previously

389 detected using the individual tools, albeit at a greater level (3- vs >10-fold increase). Additionally, we
390 speculate that newly arising CNVs would have distinct locations across samples; thus, the rare nature and
391 unique locations of high-confidence CNVs emphasized their potential to be novel (**Table S5**).

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

393 During our investigations, we identified two types of de novo CNVs; those detected in one or a few
394 genomes (rare) and those detected in more than a few (common). There is no precedence for these CNV
395 categories in the context of *Plasmodium* biology (i.e. a haploid parasite with asynchronous replication and
396 schizogeny [83]). However, we propose that tracking these categories helps us to understand the
397 biological relevance of de novo CNVs in our analysis.

398 Based on their frequency, common CNVs are either artifacts of low-input procedures/CNV analysis or
399 represent minor variants that preexist in the population or arise early in the replication cycle. For the
400 former, bias during the whole genome amplification step (i.e. the repeated pattern of over/under-
401 amplification that occurs in a reproducible pattern across the parasite genome [63]) and PCR during library
402 construction have the potential to skew gene copy number and increase the false positive rate [84, 85].
403 However, we chose experimental and computational methods designed to limit the contribution of
404 amplification bias. First, MALBAC amplification itself limits the over-amplification of genomic regions by
405 avoiding exponential amplification at the earliest steps [84] and we used limited PCR cycles during library
406 preparation (3 cycles, [63]). These efforts are most clearly shown through the reduction in CV following
407 MALBAC optimization in both of our studies (by ~39% after modifying the amplification primer [63] and
408 by 43% after switching to the *Bsu* polymerase, **Table S1**). Second, LUMPY is not dependent on read
409 coverage and HapCNV specifically addresses amplification artifacts by removing repeated signal present
410 in all samples [73]. Overall, we detected very few CNVs with conserved genomic locations across low-
411 input samples, which provides evidence that our methods limit the effect of amplification bias on the final
412 results; we only identified two high-confidence CNV regions that had conserved locations across multiple
413 *Dd2* and *FCR3* 2-cell samples (**Fig. 3B**, **Table S5**). In the future, single-read visualization of long-reads may
414 offer advantages in distinguishing amplification bias from minor variants and de novo CNVs [86].

415 Rare CNVs, on the other hand, represent either random noise or true signal from novel CNVs arising in the
416 genome. We assert that most noise is removed through normalization procedures, especially with
417 HapCNV, and the impact of remaining false positives are minimized by the relative comparison of our
418 studies (see above). We identified the majority of rare CNVs in unique genome locations across sample
419 types, providing evidence that they are not a result of amplification bias where the same CNVs are
420 repeatedly detected in each sample. Additionally, the greater impact of stress on rare CNVs than common
421 CNVs (**Table 3** and **Figs. 2G, 2I**) supports their replication dependence. The random nature of de novo
422 CNVs, as well as the capacity to encompass any gene across the genome, ensures that CNVs can alter all
423 aspects of parasite biology in response to the host environment. Further, our finding that stress-induced
424 de novo CNVs tended to exhibit altered copy number in clinical isolates combined with the high frequency
425 of unique CNVs in previous genome-wide CNV studies [53, 59], directly illustrate the expansive
426 evolutionary potential of this organism.

427 *Adaptations that encourage de novo CNV formation*

428 The current model of CNV formation in asexual erythrocytic *P. falciparum* parasites is that AT-rich
429 sequences form hairpins, disrupt replication, and eventually lead to double-strand breaks that are

430 repaired by error-prone pathways [61]. The evolution of CNVs in this organism is especially interesting
431 because of its unique genome architecture and alternative repertoire of CNV-generating repair pathways
432 [62, 87]. Although they arise at many locations across the genome ([51, 52, 59], **Fig. 3A**), *P. falciparum*
433 CNVs that contribute to adaptation are commonly gene duplications with a relatively simple structure.
434 Many impactful duplications form in tandem head-to-tail orientation ([61, 78, 88, 89], **Fig. 4A**), which is
435 likely due to a limited repertoire of DNA repair pathways; *P. falciparum* lacks the canonical
436 nonhomologous end-joining pathway that contributes to CNV formation in other organisms [1, 87].
437 Instead, parasites use pathways that employ varying lengths of sequencing homology (i.e. homologous
438 recombination, or HR, and microhomology-mediated repair, **Fig. 4B**). This repair repertoire, along with an
439 especially high AT-content genome that facilitates CNV formation [61, 78] and a lack of cell cycle
440 checkpoints that control replication forks during times of stress (reviewed in [90]), likely represent
441 adaptations that assist haploid *P. falciparum* parasites in accumulating CNVs across their genome (**Fig.
442 4A**).

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

444 By combining insights from the current study with previous knowledge about CNV formation [1, 61, 65],
445 we propose a connection between replication stress, DNA repair, and CNV generation in *P. falciparum*
446 (**Fig. 4B**). Prior studies have shown that stress can either alter levels of proteins essential for HR-based
447 repair or increase the frequency of DNA breaks [91-93]. With a decrease in HR activity in particular, the
448 parasite may increase its reliance on alternative error-prone pathways to repair DNA damage.
449 Microhomology-mediated pathways require less homology and therefore, are more likely to interact with
450 diverse sequences up- and downstream of a DNA break to generate various length CNVs. So far, the
451 predominant evidence for this model in *P. falciparum* was the detection of microhomology-mediated
452 pathway signatures in CNV breakpoints [61]. Our observation of stress-induced de novo CNV formation
453 (**Table 3** and **Fig. 2I**) further supports this model in *P. falciparum* and is consistent with studies on diverse
454 organisms [30-32, 94, 95]. Interestingly, the level of de novo CNV stimulation is consistent across
455 organisms; treatment of mammalian cells with replication inhibitors also leads to a ~3-5-fold increase in
456 de novo CNVs [30-32].

457 Even with a change in the copy number of a single region per parasite, the genomic diversity within a
458 single infected human is expansive due to the sheer numbers of *P. falciparum* parasites (estimated to
459 reach 10^8 parasites when symptomatic and $>10^{11}$ in severe infection, [96]). This diversity becomes an
460 obvious advantage as a heterogeneous population prepares asexual parasites to respond to diverse
461 stressors (**Fig. 4C**). However, one question has been whether random CNVs constitutively arise across the
462 *P. falciparum* genome or only when under stress. In the former, random amplicons within individual
463 parasites would position the population to respond rapidly to selection (e.g. antimalarial exposure). In the
464 latter, specific stressors would stimulate CNV formation to increase genomic diversity. Since some
465 antimalarials act rapidly [68], we hypothesize that beneficial CNVs must already be present in a few
466 parasites across the population to increase the chances of survival. Indeed, we observed a low level of de
467 novo CNVs across the parasite genome under normal conditions (**Fig. 2D** and **2E**). However, it is also
468 important to understand how parasites respond to stressful environments during infection, including
469 changes in nutrient composition in different hosts, drug treatment during symptomatic infection, or
470 attack from the human immune system. While the current study focused on replication stress, it will be
471 important to evaluate the impact of other sources of stress on *P. falciparum* CNV formation. For example,
472 hypoxia stimulates CNV formation in cancer cells [91] and a proteotoxic drug stimulates genetic change
473 in yeast [97].

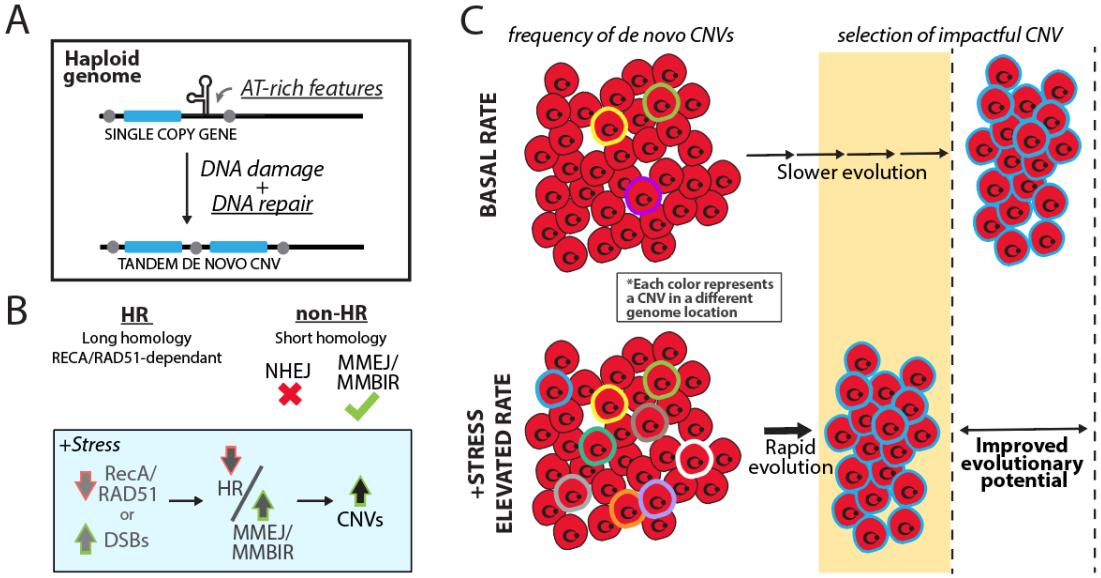

Figure 4: Connection between replication stress, DNA repair, and CNV generation in the malaria genome. **A**. Adaptations that encourage CNV formation in the *P. falciparum* genome (underlined). **B**. Summary of how replication stress impacts DNA repair pathways. HR, homologous recombination; NHEJ,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

Clinical implications & future questions

P. falciparum causes the majority of worldwide malaria deaths and readily acquires antimalarial resistance [98, 99]. Resistance-conferring CNVs that encompass multiple genes have been identified in both clinical infections [51, 59, 100-104] and laboratory selections [78, 89, 105-110]. Despite their direct contribution to resistance, CNVs may also facilitate the acquisition of point mutations in haploid *P. falciparum*; strong evidence for the close relationship comes from the observation of point mutations *within* amplifications selected in vitro [89, 106, 108, 111, 112]. Once de novo CNVs form during replication of the asexual erythrocytic stage (**Fig. 4**), meiotic recombination during the sexual phase in the mosquito can streamline beneficial CNVs to balance fitness costs [113]. Given the importance of CNVs in *P. falciparum* adaptation, it is not surprising that this organism has evolved strategies to encourage CNV formation (as described in *Adaptations that encourage de novo CNV formation*). Additionally, parasites from specific regions of the world may have an increased propensity to develop drug resistance [114]. Evaluating whether the CNV rate correlates with the parasite background will help to define the evolutionary potential of this successful pathogen. Despite some success with antimalarial therapies and vaccines targeting the *Plasmodium* parasite, a strategy to impede genome evolution may be required to control malaria infections.

Materials and Methods:

Parasite Lines, Compounds, and Treatments

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

512 We synthesized DSM1 as in previous studies [86, 115] and purchased aphidicolin (MilliporeSigma). DSM1
513 targets *P. falciparum dhodh*, which contributes to pyrimidine biosynthesis [67]. Aphidicolin inhibits B-
514 Family DNA polymerases, and consequently, *P. falciparum* replication [116, 117]. Both compounds, when
515 applied to ring-stage parasites (**Fig. 1B**), inhibit DNA replication and stall parasites at trophozoite stage
516 (**Fig. S3A** and **S4G**, [117]).

517 To assess the effects of short-term DSM1 treatment, we acquired high ring-stage cultures (>85%) by
518 synchronizing parasites twice with 5% sorbitol, 48hrs apart. We then applied 1µM DSM1, solvent control
519 (dimethylsulfoxide, DMSO), or replication inhibition control (4.4µM aphidicolin) for 12hrs. Following
520 treatment, we washed parasites with sterile 1x phosphate-buffered saline (PBS, Thermo Fisher Scientific),
521 returned them to complete RPMI, and allowed parasites to complete their life cycle and reinvade new
522 erythrocytes for an additional 29.5-34.5hrs (**Fig. S3B**). We tracked parasitemia and parasite viability on an
523 Accuri C6 flow cytometer (BD Biosciences, Franklin Lakes, NJ) as previously performed [63, 115, 118]. We
524 stained parasites with 1x SYBR Green (Thermo Fisher Scientific, stains the parasite nucleus) to assess the
525 proportion of infected erythrocytes (parasitemia) and stage of the parasite development cycle (**Fig. 1B**)
526 and 10nM MitoProbe DiIC1 (5) (Thermo Fisher Scientific, stains active parasite mitochondria) to indicate
527 the proportion of the parasites that are viable over time (**Fig. S3C**).

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

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

534 *Cell sorter calibration & accuracy assessments.* We calibrated the flow sorter (SH800, Sony Biotechnology,
535 San Jose, CA) using the manufacturer's calibration beads. We accounted for overlaps in the
536 excitation/emission wavelengths using the integrated compensation panel matrix calculation in the SH800
537 software according to the manufacturer's procedure. We also manually calibrated the droplet sorting to
538 the nearest 0.2mm, as recommended by the manufacturer, using the 96-well plate setting (Armadillo
539 high- performance 96-well plate, Thermo Fisher Scientific). We evaluated SH800 sorting accuracy prior to
540 low-input harvest using a colorimetric assay as previously described [119]. Briefly, we mixed SYBR
541 Green+/MitoProbe+ parasites (see staining details in *Parasite lines, Compounds, & Treatments*) with
542 horseradish peroxidase enzyme (Thermo Fisher Scientific) at a final concentration of 2.5mg/ml. We then

543 sorted parasites into a 96-well plate filled with TMB-ELIZA substrate (Thermo Fisher Scientific) using the
544 single cell (3 drops) instrument setting, in triplicate plates (**Fig. S2**). Formation of a color in the well (blue,
545 green, or yellow) indicates the successful sorting of the enzyme, and therefore parasites, into the well
546 with the substrate. This assessment allowed us to evaluate the accuracy of SH800 sorting (through the
547 evaluation of success for 1- versus 2-cell wells, **Fig. S2C**), the consistency of sorting (through the evaluation
548 of replicates), and the best plate positions for sorting (through the evaluation of performance in different
549 plate rows/columns). Based on these evaluations, we proceeded with isolating 2- and 10-cells per well
550 (**Fig. S2D**) and avoided sorting into the top 2 rows and the first and last column of the 96-well plate (**Figs.**
551 **S2D** and **S6**).

552 *Parasite isolation & storage*. We stained parasites with SYBR Green and MitoProbe DiIC1 (5) in complete
553 RPMI as above (see staining details in *Parasite Lines, Compounds, & Treatments*), gassed the tubes with
554 5% CO_2, 5% 0_2, 90% N, and placed sample on ice to ensure viability prior to flow sorting within 15min
555 (SH800, Sony Biotechnology Inc., San Jose, CA). We used a final concentration of 1 x 10^7 parasites/ml
556 diluted in sterile 1x PBS (Thermo Fisher Scientific) as input for sorting at the "single-cell setting" (3 drop)
557 into a 96-well plate (Armadillo high performance 96-well plate, Thermo Fisher Scientific) with each well
558 containing 2.375µl of cell lysis buffer (0.025M Tris Ph8.8 (Roche Diagnostics, Indianapolis, IN), 0.01M NaCl
559 (MilliporeSigma), 0.01M KCl (MilliporeSigma), 0.01M $(NH_4)_2SO_4$ (Thermo Fisher Scientific), 0.001M EDTA
560 (Promega, Madison, WI), and 10% Triton X-100 (MilliporeSigma)). We gated viable 1n ring-stage parasites
561 (**Fig. S5**) and sorted into the wells containing cell lysis buffer with an approximate sorting time of 10min.
562 After sorting, we centrifuged for 30 seconds in a plate centrifuge (MPS1000, Labnet International,
563 Madison, NJ). We immediately overlaid samples with one drop (approx. 25µl) of light mineral oil
564 (BioReagent grade for molecular biology, MilliporeSigma) and sealed the plates with Microamp® Clear
565 Adhesive Film (Applied Biosystems, Waltham, MA) before storage at -80°C until whole genome
566 amplification.

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

568 Before whole genome amplification, we thawed the plates containing sorted parasites (see *Parasite*
569 *Isolation & Storage*) and added 1mg/ml Proteinase K in sterile 1x PBS (Thermo Fisher Scientific) to a final
570 volume of 2.5µl per well. We heated the plates in a PCR cycler (C1000, Bio-Rad Laboratories, Hercules,
571 CA) at 50°C for 3hrs, followed by 75°C for 20 min and 80°C for 5 min for proteinase k inactivation. We
572 amplified the parasite genome using the Multiple Annealing and Looping Based Amplification Cycles
573 (MALBAC) method essentially as previously described ([63], Version 1 in **Fig. S1**) with some modifications
574 (Version 2, **Fig. S1**). In summary, 1) we modified the pre-amplification random primer by adding 5
575 additional degenerate bases with 20% GC-content to increase annealing to AT-rich genome sequences
576 (5'GTGAGTGATGGTTGAGGTAGTGTGGAG<u>NNNNNNNNNN</u>TTT 3'); 2) we performed 19 of the 21 total linear
577 cycles with the *Bsu* DNA Polymerase (Large Fragment, New England Biosciences), which has a lower
578 optimal reaction temperature (37°C) to improve the amplification on AT-rich sequences [120]; 3) we
579 lowered the extension temperature from 40/50°C to 37°C during the linear amplification cycles that used
580 the *Bsu* enzyme (see full cycling parameters in **Fig. S1**); and 4) we integrated robotic pipetting (Mosquito
581 LV, SPT Labtech, Melbourn, UK) to increase the throughput of our assays (from 23 samples in Version 1 to
582 90 samples in the current Version 2) and limit contamination potential.

583 Overall, we performed 21 total linear cycles (19 cycles with *Bsu* polymerase and 2 cycles with *Bst*
584 polymerase, New England Biolabs, Ipswitch, MA) and 17 total exponential amplification cycles using

585 Herculase II Fusion DNA polymerase (Agilent Technologies, Santa Clara, CA). During amplification steps,
586 we employed standard steps to limit contamination [63]. For automated pipetting of the enzyme solution
587 during linear cycles, tips were changed after each round of pipetting. Post-amplification, we purified
588 amplified DNA with Zymo DNA Clean & Concentrator-5columns (Zymo Research, Irvine, CA) according to
589 the protocol and ran 2µl of all samples on 1% agarose gels to check for the presence of DNA (generally, if
590 >30ng/µl, samples could be visualized with a size range of 100 to >1500bp).

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

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

594 *Droplet Digital PCR.* To confirm the presence of parasite DNA in MALBAC-amplified samples, we
595 performed droplet digital PCR (ddPCR) as described previously using the QX2000 droplet generator, C1000
596 thermocycler, and QX2000 droplet reader (Bio-Rad Laboratories) [115, 121]. We used duplex assays to
597 evaluate two parasite genomic loci concurrently (*pfmdr1:* Forward- TGCCCACAGAATTGCATCTA;
598 Reverse- ACCCTGATCGAAATGGAACCT; Probe - TCGTGTGTTCCATGTGACTG; *pfhsp70:* Forward-
599 TGCTGTCATTACCGTTCCAG; Reverse - AGATGCTGGTACAATTGCAGGA; Probe -
600 AGCAGCTGCAGTAGGTTCATT (Integrated DNA technologies, Newark, NJ). The reaction master mix
601 contained 600nm of forward and reverse primers, 50nm probes, 10µl of ddPCR Supermix for Probes (2x,
602 Bio-Rad Laboratories), 3µl of nuclease-free water (QIAGEN), and 1.5ng (5µl) of template DNA per assay
603 (total of 20µl). We used the following cycling conditions for PCR amplification: 10 min at 95°C initial
604 denaturation step, 1 min at 95°C second denaturation step, and 2 min at 58°C annealing and extension
605 step (ramp rate of 1°C per second), the second denaturation step and the annealing/extension step
606 repeated 60 times, and then 10 min at 98°C to halt the reaction [121]. In addition to running amplified
607 samples to assess amplification success, we ran ddPCR with bulk genomic DNA as a positive control, no
608 template controls (water replaced DNA), and material from "no cell" wells to assess cross-well
609 contamination. We considered the samples positive for parasite DNA if there were more than 50 total
610 positive droplets in target-positive clusters.

611 *High Resolution Melting Assay.* To assess potential contamination between MALBAC-amplified samples,
612 we performed asymmetric PCR amplification of the *pfdhps* locus followed by high-resolution melting
613 (HRM) as described previously [122, 123]. The *pfdhps* locus at codon 613 is distinct in *Dd2* and *FCR3*
614 parasite lines (*Dd2*: Ser-613 and *FCR3*: Ala-613, [124]). Each 20µL reaction contained 8µl of the 2.5x
615 LightScanner Master mix (BioFire™ Defense, Salt Lake City, Utah, USA), 1/10µM of forward/reverse
616 primers and 8µM probes targeting the *pfdhps* gene position 613: Forward -
617 CTCTTACAAAATATACATGTATATGATGAGTATCCACTT; Reverse-
618 CATGTAATTTTTGTTGTGTATTTATTACAACATTTTGA; Probe - AAGATTTATTGCCCATTGCATGA/3SpC3,
619 (Integrated DNA technologies), 7µl of nuclease free water, and 3µl of DNA (~0.05ng total). We used the
620 following cycling conditions for PCR amplification with the Rotor-Gene Q instrument with a 72-well rotor
621 (QIAGEN): 95°C for 5 min, 45 cycles of 95°C for 10s, 55°C for 30s, and 72°C for 10s, followed by a pre-melt
622 at 55°C for 90s, and a HRM ramp from 65°C to 90°C, with an increase of 0.1°C every 2s. We plotted the
623 change in fluorescence versus temperature (dF/T) using Rotor-Gene Q software (version 2.3.5, build 1;
624 QIAGEN) and compared HRM peaks of amplified samples to bulk genomic DNA and plasmid controls.

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

626 We extracted bulk DNA for short-read sequencing as previously performed [63]. Briefly, we lysed
627 erythrocytes with 1.5% saponin and washed the parasite pellet 3 times with 1x PBS (Thermo Fisher
628 Scientific), before resuspension in a buffered solution (150mM NaCl (MilliporeSigma), 10mM EDTA
629 (Promega Corporation, Madison, WI), and 50mM Tris pH7.5 (Roche Diagnostics)) to a total volume of
630 500µl. We then lysed the parasites with 10% sarkosyl (MilliporeSigma) and 20mg/ml proteainase K
631 (Thermo Fisher Scientific) at 37°C overnight before DNA purification using standard
632 phenol/chloroform/isoamyl alcohol extraction and chloroform washing steps (2 times each, [63]). Finally,
633 we precipitated DNA using 100% ethanol with 100mM of sodium acetate overnight in DNA-lo bind tubes
634 (Eppendorf, Enfield, CT) and then washed twice with 70% ethanol before resuspension in 50µl nuclease
635 free water (QIAGEN). We stored bulk genomic DNA at -20°C until sequencing library preparation.

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

637 *Low-input sample selection.* We selected 16 low-input samples from the *FCR3* plate, 36 samples from the
638 untreated *Dd2* plate, and 38 samples from the treated *Dd2* plate for short-read Illumina sequencing. We
639 based our selection on the quantity of the MALBAC amplified DNA and presence of parasite DNA using
640 ddPCR (**Figs. S6 and S7**). In summary, the majority of *FCR3* and untreated *Dd2* samples yielded quantifiable
641 parasite DNA following MALBAC amplification (53/60 *FCR3* samples and 60/60 untreated *Dd2* samples);
642 in these conditions, we chose samples randomly to proceed with sequencing (indicated in **Fig. S6**). For
643 treated *Dd2* samples, we chose samples for sequencing if they had adequate DNA quantity (>10ng total,
644 30 samples, **Fig. S6**) or had ddPCR results showing the presence of parasite DNA (an additional 8 samples).

645 *Short-read sequencing.* Before short-read sequencing, we sheared bulk samples and low-input samples
646 using Covaris M220 Focused Ultrasonicator for 150s and 130s, respectively, to generate fragment sizes of
647 ~350bp as evaluated by an Agilent 2100 Bioanalyzer using the High Sensitivity DNA kit (Agilent, Santa
648 Clara). We adjusted the volume of sheared samples with nuclease-free water up to 50µl. For samples with
649 >100ng (**Table S2,** including bulk and MALBAC amplified samples), we diluted them to 1.2-2ng/µl; for
650 samples <100ng, we proceeded with no dilution. We used NEBNext Ultra II kit (Illumina Inc., San Diego,
651 CA) to prepare libraries for sequencing with 3 cycles of PCR amplification, as performed previously [63].
652 We quantified the resulting libraries using NEBNext Library Quant Kit (Illumina Inc.) before sequencing on
653 the Illumina Nextseq 550 using 150 bp paired-end cycles.

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

655 *Read processing and alignment*. We performed short-read quality control steps as described previously
656 [61, 63]. Briefly, we reordered and removed singletons and subsequently interleaved paired reads using
657 BBMap, trimmed the MALBAC common sequence, PhiX, and Illumina adapters from the remaining reads
658 with the BBDuk tool within BBMap, and aligned reads to the *pf3D7-62_v3* reference genome using the
659 Speedseq genome aligner [125]. We removed reads that map to VAR regions from bam files according to
660 previously defined genomic coordinates [76]. We filtered out reads with low mapping quality (<q30) and
661 duplicated reads using SAMtools [126]. We employed Qualimap to report mean coverage and standard
662 deviation across the genome [127]. Using non-overlapping 20kb size bins, we calculated the coefficient of
663 variation of read coverage by dividing the standard deviation of coverage within a bin by the mean across
664 a sample and multiplying by 100 [128, 129] (R version 4.4.2).

665 *Down-sampling.* For analysis that assessed down-sampled data (**Figs. 2, S10, and S11**), we first converted
666 the processed bam files back into FASTQ files using SAMtools and then used the reformat.sh option of

667 BBtools to select 1.3M reads from each FASTQ (represented the fewest number of reads from a sample
668 that passed quality filtering from the final dataset). We then realigned files to the reference genome
669 (*pf3D7-62_v3*).

670 *Single nucleotide polymorphism analysis*. We performed SNP genotyping and analysis as previously [63],
671 based on the MalariaGen *P. falciparum* Community Project V6.0 pipeline [130-133] using the *pf3D7-62_v3*
672 reference genome. Briefly, we applied GATK's Base Quality Score Recalibration using default settings. We
673 detected potential SNPs using GATK's HaplotypeCaller and subsequently genotyped the SNPs using
674 CombineGVCFs and GenotypeGVCFs. Then we employed GATK's VariantRecalibrator using previously
675 validated SNP datasets [113]. We then applied GATK's ApplyRecalibration to assign a VQSLOD score [133].
676 We filtered the resulting SNPs for those with VQSLOD > 6 and for a GT quality metric >20 to ensure high-
677 quality variant calling. We only selected variants flagged as Bi-allelic to simplify the analysis. For SNP
678 Principle Component Analysis (PCA), we merged experiment-wide SNP data (described above) into a single
679 file. Then we merged the VCF into a large matrix and converted the genomic data into numeric
680 information using the 'vcfr' package in R (Version 4.2.3) (https://CRAN.R-project.org/package=vcfR). We
681 excluded individual SNPs if >25% of the samples lacked a call in this position or if all calls were the same
682 for each sample in that position). We scaled the remaining SNPs around the origin using the 'scale' R
683 function. We then calculated the principal components using the 'prcomp' R function, and scored the
684 dataset using the 'scores' function from the 'vegan' R package (https://CRAN.R-
685 project.org/package=vegan).

686 **Copy Number Variation Analysis**

687 *CNV calling in bulk samples.* We performed CNV detection for bulk samples similar to as previously
688 described [61, 63]. Briefly, we called CNVs independently using two methods, CNVnator (read depth based
689 calling, [134]) and LUMPY (split and discordant read based calling, [73]). To Identify CNVs called in both
690 methods, we used SVCROWS to define overlapping CNV regions relative to their size. Briefly, SVCROWS
691 uses a reciprocal-overlap-based approach (i.e. two CNVs must be overlapping each other at, or greater
692 than, a defined threshold) to determine if two CNVs are close enough in their genomic position to be
693 called the same. The source code for SVCROWS can be accessed at https://github.com/A-Crow-
694 Nowhere/SVCROWS.git. The program utilizes different thresholds for overlap based on the sizes of the
695 CNVs being compared, ensuring that we account for shifts in CNV position. For known CNV calls, we used
696 the following SVCROWs input parameters: ExpandRORegion = FALSE, BPfactor = TRUE, DefaultSizes =
697 FALSE, xs = 3000, xl = 10,000, y1s = 300, y1l = 1000, y2s = 50% and y2l = 80%; based on the average size of
698 a *P. falciparum* gene (~2.3kb) and intergenic region (~2kb). Similar to our previous study [63], we identified
699 3 known CNVs that were called by both LUMPY and CNVnator methods in the core genome of bulk
700 samples (untreated and treated). We determined known CNV boundaries using SVCROWS: *pfmdr1*
701 (*Pf3D7_05_v3*, 888001-970000, 82kb), *pf11-1* (*Pf3D7_10_v3*, 1521345-1541576, 20kb), and *pf332*
702 (*Pf3D7_11_v3*, 1950201-1962400, 12kb).

703 *CNV calling in 2-cell samples.* We employed two methods for CNV detection in the core genome of low-
704 input samples; LUMPY is a split/discordant read strategy with high sensitivity [73], and HapCNV is a read
705 coverage-based strategy designed for haploid genomes. We ran LUMPY as part of Speedseq with default
706 parameters as previously described [61]. We filtered resulting structural variants to include only
707 duplications (DUP, >1 copy of a region) and deletions (DEL, one less copy of region than the reference).

708 We then filtered those calls for those GQ > 20 to ensure high-fidelity calls. In HapCNV, we used a quality
709 control and bias correction procedure to exclude bins of poor quality and remove bias introduced by GC
710 content and mappability variation. We then constructed a pseudo-reference for each *Dd2* low-input
711 sample using within-*Dd2* information, which enabled control of background noise while preserving CNV
712 signals after normalization. Finally, we used a circular binary segmentation algorithm (CBS, [135]) to
713 detect copy number change points followed by a Gaussian Mixture Model (GMM, [136]) for CNV
714 identification. The source code for HapCNV and examples of real data application can be accessed at
715 https://github.com/FeifeiXiao-lab/HapCNV. For statistics, we used PRISM (GraphPad Software, La Jolla,
716 CA), using unpaired parametric T-tests with Welch's correction.

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

726 *High-confidence CNV region identification.* To identify "high-confidence" CNV regions called by both
727 HapCNV and LUMPY methods, we compared the 'consensus list' generated by SVCROWS for each
728 detection method by combining them into a single SVCROWS input file. Because 1) HapCNV generates
729 imprecise breakpoints, and 2) there is a large disparity of average CNV region lengths between the two
730 methods (HapCNV = ~40kb, LUMPY = ~4.3kb), we relaxed the stringency of the SVCROWS parameters.
731 Our input parameters to generate the "high-confidence" list were: xs = 3000, xl = 6000, y1s = 500, y1l =
732 1500, y2s = 30% and y2l = 60%. We defined "high-confidence" CNV regions as those that had >1 match
733 from both HapCNV and LUMPY that was the same type (i.e. either duplication or deletion or mixed). For
734 Venn diagram generation, we calculated overlaps using SVCROWS "Scavenge" mode (input parameters:
735 ExpandRORegion = FALSE, BPfactor = TRUE, DefaultSizes = FALSE, xs = 3000, xl = 6000, y1s = 500, y1l =
736 1500, y2s = 30% and y2l = 60%). We systematically compared lists for each overlap comparison, and if
737 regions had at least one match in an opposing dataset, we considered it a match. We used the
738 *draw.quad.venn* function in the "VennDiagram" R package (R 4.2.1) to generate the diagram.

739 *Gene Ontology enrichment and protein class identification.* We used the online Gene Ontology Resource
740 (geneontology.org) to perform GO enrichment analysis using the PANTHER Classification System [137,
741 138]. Since a large portion of the *P. falciparum* genome remains unannotated (PlasmoDB, 30%) and the
742 majority of molecular functions remain unclassified (92.8%), we used the Panther Protein class
743 assessment (version 19.0, only 55.9% remained unclassified) with default statistics (Fisher's test with FDR
744 adjusted p value of <0.05 for significance, which is recommended for small counts and overlaps between
745 classes). We used the web tool to represent protein classes on pie charts.

746 *Comparison to clinical CNV dataset.* To determine genes covered by de novo CNVs from HapCNV and
747 LUMPY, we used SVCROWS "Hunt" mode (input parameters: BPfactor = TRUE, DefaultSizes = FALSE,
748 xs = 3000, xl = 6000, y1s = 300, y1l = 600, y2s = 30, y2l = 60), which takes a secondary input list of
749 known genes (Pf3D7_62_v3, Plasmodb.org) against which to compare CNV regions. For the clinical

750 dataset, we used genes covered by >300bp CNV regions present in high frequency in clinical isolates
751 (Supplementary Table 3 from [59]), and manually reformatted to match SVCROWS input style
752 guidelines.

753

754 **Acknowledgments:**

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

759

760 **References:**

761 1. Hastings, P.J., J.R. Lupski, S.M. Rosenberg, and G. Ira, *Mechanisms of change in gene copy*
762 *number.* Nat Rev Genet, 2009. 10(8): p. 551-64.
763 2. Lauer, S. and D. Gresham, *An evolving view of copy number variants.* Curr Genet, 2019.
764 3. Pos, O., J. Radvanszky, G. Buglyo, Z. Pos, D. Rusnakova, B. Nagy, and T. Szemes, *DNA copy*
765 *number variation: Main characteristics, evolutionary significance, and pathological aspects.*
766 Biomed J, 2021. 44(5): p. 548-559.
767 4. Roth, J.R., *The joys and terrors of fast adaptation: new findings elucidate antibiotic resistance*
768 *and natural selection.* Mol Microbiol, 2011. 79(2): p. 279-82.
769 5. Payen, C., S.C. Di Rienzi, G.T. Ong, J.L. Pogachar, J.C. Sanchez, A.B. Sunshine, M.K. Raghuraman,
770 B.J. Brewer, and M.J. Dunham, *The dynamics of diverse segmental amplifications in populations*
771 *of Saccharomyces cerevisiae adapting to strong selection.* G3 (Bethesda), 2014. 4(3): p. 399-409.
772 6. Elde, N.C., S.J. Child, M.T. Eickbush, J.O. Kitzman, K.S. Rogers, J. Shendure, A.P. Geballe, and H.S.
773 Malik, *Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses.*
774 Cell, 2012. 150(4): p. 831-41.
775 7. Leary, R.J., J.C. Lin, J. Cummins, S. Boca, L.D. Wood, D.W. Parsons, S. Jones, T. Sjoblom, B.H. Park,
776 R. Parsons, J. Willis, D. Dawson, J.K. Willson, T. Nikolskaya, Y. Nikolsky, L. Kopelovich, N.
777 Papadopoulos, L.A. Pennacchio, T.L. Wang, S.D. Markowitz, G. Parmigiani, K.W. Kinzler, B.
778 Vogelstein, and V.E. Velculescu, *Integrated analysis of homozygous deletions, focal*
779 *amplifications, and sequence alterations in breast and colorectal cancers.* Proc Natl Acad Sci U S
780 A, 2008. 105(42): p. 16224-9.
781 8. Beroukhim, R., C.H. Mermel, D. Porter, G. Wei, S. Raychaudhuri, J. Donovan, J. Barretina, J.S.
782 Boehm, J. Dobson, M. Urashima, K.T. Mc Henry, R.M. Pinchback, A.H. Ligon, Y.J. Cho, L. Haery, H.
783 Greulich, M. Reich, W. Winckler, M.S. Lawrence, B.A. Weir, K.E. Tanaka, D.Y. Chiang, A.J. Bass, A.
784 Loo, C. Hoffman, J. Prensner, T. Liefeld, Q. Gao, D. Yecies, S. Signoretti, E. Maher, F.J. Kaye, H.
785 Sasaki, J.E. Tepper, J.A. Fletcher, J. Tabernero, J. Baselga, M.S. Tsao, F. Demichelis, M.A. Rubin,
786 P.A. Janne, M.J. Daly, C. Nucera, R.L. Levine, B.L. Ebert, S. Gabriel, A.K. Rustgi, C.R. Antonescu, M.
787 Ladanyi, A. Letai, L.A. Garraway, M. Loda, D.G. Beer, L.D. True, A. Okamoto, S.L. Pomeroy, S.
788 Singer, T.R. Golub, E.S. Lander, G. Getz, W.R. Sellers, and M. Meyerson, *The landscape of*
789 *somatic copy-number alteration across human cancers.* Nature, 2010. 463(7283): p. 899-905.
790 9. Zhang, L., Y. Yuan, K.H. Lu, and L. Zhang, *Identification of recurrent focal copy number variations*
791 *and their putative targeted driver genes in ovarian cancer.* BMC Bioinformatics, 2016. 17(1): p.
792 222.

793 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*
794 *Detection from Circulating Tumor DNA in Late Stage Non-Small Cell Lung Cancer Patients.* Genes
795 (Basel), 2019. 10(11).
796 11. Krijgsman, O., B. Carvalho, G.A. Meijer, R.D. Steenbergen, and B. Ylstra, *Focal chromosomal copy*
797 *number aberrations in cancer-Needles in a genome haystack.* Biochim Biophys Acta, 2014.
798 1843(11): p. 2698-2704.
799 12. Carey-Smith, S.L., R.S. Kotecha, L.C. Cheung, and S. Malinge, *Insights into the Clinical, Biological*
800 *and Therapeutic Impact of Copy Number Alteration in Cancer.* Int J Mol Sci, 2024. 25(13).
801 13. Gonzalez, E., H. Kulkarni, H. Bolivar, A. Mangano, R. Sanchez, G. Catano, R.J. Nibbs, B.I.
802 Freedman, M.P. Quinones, M.J. Bamshad, K.K. Murthy, B.H. Rovin, W. Bradley, R.A. Clark, S.A.
803 Anderson, J. O'Connell R, B.K. Agan, S.S. Ahuja, R. Bologna, L. Sen, M.J. Dolan, and S.K. Ahuja,
804 *The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility.*
805 Science, 2005. 307(5714): p. 1434-40.
806 14. Sebat, J., B. Lakshmi, D. Malhotra, J. Troge, C. Lese-Martin, T. Walsh, B. Yamrom, S. Yoon, A.
807 Krasnitz, J. Kendall, A. Leotta, D. Pai, R. Zhang, Y.H. Lee, J. Hicks, S.J. Spence, A.T. Lee, K. Puura, T.
808 Lehtimaki, D. Ledbetter, P.K. Gregersen, J. Bregman, J.S. Sutcliffe, V. Jobanputra, W. Chung, D.
809 Warburton, M.C. King, D. Skuse, D.H. Geschwind, T.C. Gilliam, K. Ye, and M. Wigler, *Strong*
810 *association of de novo copy number mutations with autism.* Science, 2007. 316(5823): p. 445-9.
811 15. Glessner, J.T., J.P. Bradfield, K. Wang, N. Takahashi, H. Zhang, P.M. Sleiman, F.D. Mentch, C.E.
812 Kim, C. Hou, K.A. Thomas, M.L. Garris, S. Deliard, E.C. Frackelton, F.G. Otieno, J. Zhao, R.M.
813 Chiavacci, M. Li, J.D. Buxbaum, R.I. Berkowitz, H. Hakonarson, and S.F. Grant, *A genome-wide*
814 *study reveals copy number variants exclusive to childhood obesity cases.* Am J Hum Genet, 2010.
815 87(5): p. 661-6.
816 16. Olsson, L.M. and R. Holmdahl, *Copy number variation in autoimmunity--importance hidden in*
817 *complexity?* Eur J Immunol, 2012. 42(8): p. 1969-76.
818 17. Butchbach, M.E., *Copy Number Variations in the Survival Motor Neuron Genes: Implications for*
819 *Spinal Muscular Atrophy and Other Neurodegenerative Diseases.* Front Mol Biosci, 2016. 3: p. 7.
820 18. Sekar, A., A.R. Bialas, H. de Rivera, A. Davis, T.R. Hammond, N. Kamitaki, K. Tooley, J. Presumey,
821 M. Baum, V. Van Doren, G. Genovese, S.A. Rose, R.E. Handsaker, C. Schizophrenia Working
822 Group of the Psychiatric Genomics, M.J. Daly, M.C. Carroll, B. Stevens, and S.A. McCarroll,
823 *Schizophrenia risk from complex variation of complement component 4.* Nature, 2016.
824 530(7589): p. 177-83.
825 19. Zekavat, S.M., S. Ruotsalainen, R.E. Handsaker, M. Alver, J. Bloom, T. Poterba, C. Seed, J. Ernst,
826 M. Chaffin, J. Engreitz, G.M. Peloso, A. Manichaikul, C. Yang, K.A. Ryan, M. Fu, W.C. Johnson, M.
827 Tsai, M. Budoff, R.S. Vasan, L.A. Cupples, J.I. Rotter, S.S. Rich, W. Post, B.D. Mitchell, A. Correa,
828 A. Metspalu, J.G. Wilson, V. Salomaa, M. Kellis, M.J. Daly, B.M. Neale, S. McCarroll, I. Surakka, T.
829 Esko, A. Ganna, S. Ripatti, S. Kathiresan, P. Natarajan, and N.T.L.W. Group, *Deep coverage whole*
830 *genome sequences and plasma lipoprotein(a) in individuals of European and African ancestries.*
831 Nat Commun, 2018. 9(1): p. 2606.
832 20. Farashi, S. and C.L. Harteveld, *Molecular basis of alpha-thalassemia.* Blood Cells Mol Dis, 2018.
833 70: p. 43-53.
834 21. Wisniowiecka-Kowalnik, B. and B.A. Nowakowska, *Genetics and epigenetics of autism spectrum*
835 *disorder-current evidence in the field.* J Appl Genet, 2019. 60(1): p. 37-47.
836 22. Craven, S.H. and E.L. Neidle, *Double trouble: medical implications of genetic duplication and*
837 *amplification in bacteria.* Future Microbiol, 2007. 2(3): p. 309-21.
838 23. Sandegren, L. and D.I. Andersson, *Bacterial gene amplification: implications for the evolution of*
839 *antibiotic resistance.* Nat Rev Microbiol, 2009. 7(8): p. 578-88.

840 24. Hjort, K., H. Nicoloff, and D.I. Andersson, *Unstable tandem gene amplification generates*
841 *heteroresistance (variation in resistance within a population) to colistin in Salmonella enterica.*
842 Mol Microbiol, 2016. 102(2): p. 274-289.
843 25. Anderson, S.E., E.X. Sherman, D.S. Weiss, and P.N. Rather, *Aminoglycoside Heteroresistance in*
844 *Acinetobacter baumannii AB5075.* mSphere, 2018. 3(4).
845 26. Nicoloff, H., K. Hjort, B.R. Levin, and D.I. Andersson, *The high prevalence of antibiotic*
846 *heteroresistance in pathogenic bacteria is mainly caused by gene amplification.* Nat Microbiol,
847 2019. 4(3): p. 504-514.
848 27. Baslan, T., J. Kendall, K. Volyanskyy, K. McNamara, H. Cox, S. D'Italia, F. Ambrosio, M. Riggs, L.
849 Rodgers, A. Leotta, J. Song, Y. Mao, J. Wu, R. Shah, R. Gularte-Merida, K. Chadalavada, G.
850 Nanjangud, V. Varadan, A. Gordon, C. Curtis, A. Krasnitz, N. Dimitrova, L. Harris, M. Wigler, and J.
851 Hicks, *Novel insights into breast cancer copy number genetic heterogeneity revealed by single-*
852 *cell genome sequencing.* Elife, 2020. 9.
853 28. van Dijk, E., T. van den Bosch, K.J. Lenos, K. El Makrini, L.E. Nijman, H.F.B. van Essen, N. Lansu,
854 M. Boekhout, J.H. Hageman, R.C. Fitzgerald, C.J.A. Punt, J.B. Tuynman, H.J.G. Snippert, G. Kops,
855 J.P. Medema, B. Ylstra, L. Vermeulen, and D.M. Miedema, *Chromosomal copy number*
856 *heterogeneity predicts survival rates across cancers.* Nat Commun, 2021. 12(1): p. 3188.
857 29. Sobral, D., M. Martins, S. Kaplan, M. Golkaram, M. Salmans, N. Khan, R. Vijayaraghavan, S.
858 Casimiro, A. Fernandes, P. Borralho, C. Ferreira, R. Pinto, C. Abreu, A.L. Costa, S. Zhang, T.
859 Pawlowski, J. Godsey, A. Mansinho, D. Macedo, S. Lobo-Martins, P. Filipe, R. Esteves, J.
860 Coutinho, P.M. Costa, A. Ramires, F. Aldeia, A. Quintela, A. So, L. Liu, A.R. Grosso, and L. Costa,
861 *Genetic and microenvironmental intra-tumor heterogeneity impacts colorectal cancer evolution*
862 *and metastatic development.* Commun Biol, 2022. 5(1): p. 937.
863 30. Arlt, M.F., J.G. Mulle, V.M. Schaibley, R.L. Ragland, S.G. Durkin, S.T. Warren, and T.W. Glover,
864 *Replication stress induces genome-wide copy number changes in human cells that resemble*
865 *polymorphic and pathogenic variants.* Am J Hum Genet, 2009. 84(3): p. 339-50.
866 31. Arlt, M.F., A.C. Ozdemir, S.R. Birkeland, T.E. Wilson, and T.W. Glover, *Hydroxyurea induces de*
867 *novo copy number variants in human cells.* Proc Natl Acad Sci U S A, 2011. 108(42): p. 17360-5.
868 32. Arlt, M.F., S. Rajendran, S.R. Birkeland, T.E. Wilson, and T.W. Glover, *De novo CNV formation in*
869 *mouse embryonic stem cells occurs in the absence of Xrcc4-dependent nonhomologous end*
870 *joining.* PLoS Genet, 2012. 8(9): p. e1002981.
871 33. Lauer, S., G. Avecilla, P. Spealman, G. Sethia, N. Brandt, S.F. Levy, and D. Gresham, *Single-cell*
872 *copy number variant detection reveals the dynamics and diversity of adaptation.* PLoS Biol, 2018.
873 16(12): p. e3000069.
874 34. Oketch, D.J.A., M. Giulietti, and F. Piva, *Copy Number Variations in Pancreatic Cancer: From*
875 *Biological Significance to Clinical Utility.* Int J Mol Sci, 2023. 25(1).
876 35. Yi, D., J.W. Nam, and H. Jeong, *Toward the functional interpretation of somatic structural*
877 *variations: bulk- and single-cell approaches.* Brief Bioinform, 2023. 24(5).
878 36. Macaulay, I.C. and T. Voet, *Single cell genomics: advances and future perspectives.* PLoS Genet,
879 2014. 10(1): p. e1004126.
880 37. Wang, Y. and N.E. Navin, *Advances and applications of single-cell sequencing technologies.* Mol
881 Cell, 2015. 58(4): p. 598-609.
882 38. Gawad, C., W. Koh, and S.R. Quake, *Single-cell genome sequencing: current state of the science.*
883 Nat Rev Genet, 2016. 17(3): p. 175-88.
884 39. Wang, X., H. Chen, and N.R. Zhang, *DNA copy number profiling using single-cell sequencing.* Brief
885 Bioinform, 2018. 19(5): p. 731-736.
886 40. Patel, A.P., I. Tirosh, J.J. Trombetta, A.K. Shalek, S.M. Gillespie, H. Wakimoto, D.P. Cahill, B.V.
887 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.

50. Laks, E., A. McPherson, H. Zahn, D. Lai, A. Steif, J. Brimhall, J. Biele, B. Wang, T. Masud, J. Ting, D. Grewal, C. Nielsen, S. Leung, V. Bojilova, M. Smith, O. Golovko, S. Poon, P. Eirew, F. Kabeer, T. Ruiz de Algara, S.R. Lee, M.J. Taghiyar, C. Huebner, J. Ngo, T. Chan, S. Vatrt-Watts, P. Walters, N. Abrar, S. Chan, M. Wiens, L. Martin, R.W. Scott, T.M. Underhill, E. Chavez, C. Steidl, D. Da Costa, Y. Ma, R.J.N. Coope, R. Corbett, S. Pleasance, R. Moore, A.J. Mungall, C. Mar, F. Cafferty, K. Gelmon, S. Chia, C.I.G.C. Team, M.A. Marra, C. Hansen, S.P. Shah, and S. Aparicio, *Clonal Decomposition and DNA Replication States Defined by Scaled Single-Cell Genome Sequencing.* Cell, 2019. 179(5): p. 1207-1221 e22.

51. Kidgell, C., S.K. Volkman, J. Daily, J.O. Borevitz, D. Plouffe, Y. Zhou, J.R. Johnson, K. Le Roch, O. Sarr, O. Ndir, S. Mboup, S. Batalov, D.F. Wirth, and E.A. Winzeler, *A systematic map of genetic variation in Plasmodium falciparum.* PLoS Pathog, 2006. 2(6): p. e57.

934 52. Ribacke, U., B.W. Mok, V. Wirta, J. Normark, J. Lundeberg, F. Kironde, T.G. Egwang, P. Nilsson,
935 and M. Wahlgren, *Genome wide gene amplifications and deletions in Plasmodium falciparum.*
936 Molecular and Biochemical Parasitology, 2007. 155(1): p. 33-44.
937 53. Cheeseman, I., N. Gomez-Escobar, C. Carret, A. Ivens, L. Stewart, K. Tetteh, and D. Conway, *Gene*
938 *copy number variation throughout the Plasmodium falciparum genome.* BMC Genomics, 2009.
939 10(1): p. 353.
940 54. Bopp, S.E., M.J. Manary, A.T. Bright, G.L. Johnston, N.V. Dharia, F.L. Luna, S. McCormack, D.
941 Plouffe, C.W. McNamara, J.R. Walker, D.A. Fidock, E.L. Denchi, and E.A. Winzeler, *Mitotic*
942 *evolution of Plasmodium falciparum shows a stable core genome but recombination in antigen*
943 *families.* PLoS Genet, 2013. 9(2): p. e1003293.
944 55. Gendrot, M., R. Fawaz, J. Dormoi, M. Madamet, and B. Pradines, *Genetic diversity and deletion*
945 *of Plasmodium falciparum histidine-rich protein 2 and 3: a threat to diagnosis of P. falciparum*
946 *malaria.* Clin Microbiol Infect, 2019. 25(5): p. 580-585.
947 56. Matesanz, F., M. Tellez, and A. Alcina, *The Plasmodium falciparum fatty acyl-CoA synthetase*
948 *family (PfACS) and differential stage-specific expression in infected erythrocytes.* Mol Biochem
949 Parasitol, 2003. 126(1): p. 109-12.
950 57. Dankwa, S., C. Lim, A.K. Bei, R.H. Jiang, J.R. Abshire, S.D. Patel, J.M. Goldberg, Y. Moreno, M.
951 Kono, J.C. Niles, and M.T. Duraisingh, *Ancient human sialic acid variant restricts an emerging*
952 *zoonotic malaria parasite.* Nat Commun, 2016. 7: p. 11187.
953 58. Price, R.N., A.-C. Uhlemann, A. Brockman, R. McGready, E. Ashley, L. Phaipun, R. Patel, K. Laing,
954 S. Looareesuwan, N.J. White, F. Nosten, and S. Krishna, *Mefloquine resistance in Plasmodium*
955 *falciparum and increased pfmdr1 gene copy number.* The Lancet, 2004. 364(9432): p. 438-447.
956 59. Ravenhall, M., E.D. Benavente, C.J. Sutherland, D.A. Baker, S. Campino, and T.G. Clark, *An*
957 *analysis of large structural variation in global Plasmodium falciparum isolates identifies a novel*
958 *duplication of the chloroquine resistance associated gene.* Sci Rep, 2019. 9(1): p. 8287.
959 60. Heinberg, A., E. Siu, C. Stern, E.A. Lawrence, M.T. Ferdig, K.W. Deitsch, and L.A. Kirkman, *Direct*
960 *evidence for the adaptive role of copy number variation on antifolate susceptibility in*
961 *Plasmodium falciparum.* Mol Microbiol, 2013. 88(4): p. 702-12.
962 61. Huckaby, A.C., C.S. Granum, M.A. Carey, K. Szlachta, B. Al-Barghouthi, Y.H. Wang, and J.L. Guler,
963 *Complex DNA structures trigger copy number variation across the Plasmodium falciparum*
964 *genome.* Nucleic Acids Res, 2019. 47(4): p. 1615-1627.
965 62. Gardner, M.J., N. Hall, E. Fung, O. White, M. Berriman, R.W. Hyman, J.M. Carlton, A. Pain, K.E.
966 Nelson, S. Bowman, I.T. Paulsen, K. James, J.A. Eisen, K. Rutherford, S.L. Salzberg, A. Craig, S.
967 Kyes, M.-S. Chan, V. Nene, S.J. Shallom, B. Suh, J. Peterson, S. Angiuoli, M. Pertea, J. Allen, J.
968 Selengut, D. Haft, M.W. Mather, A.B. Vaidya, D.M.A. Martin, A.H. Fairlamb, M.J. Fraunholz, D.S.
969 Roos, S.A. Ralph, G.I. McFadden, L.M. Cummings, G.M. Subramanian, C. Mungall, J.C. Venter,
970 D.J. Carucci, S.L. Hoffman, C. Newbold, R.W. Davis, C.M. Fraser, and B. Barrell, *Genome sequence*
971 *of the human malaria parasite Plasmodium falciparum.* Nature, 2002. 419(6906): p. 498-511.
972 63. Liu, S., A.C. Huckaby, A.C. Brown, C.C. Moore, I. Burbulis, M.J. McConnell, and J.L. Guler, *Single-*
973 *cell sequencing of the small and AT-skewed genome of malaria parasites.* Genome Med, 2021.
974 13(1): p. 75.
975 64. Galhardo, R.S., P.J. Hastings, and S.M. Rosenberg, *Mutation as a stress response and the*
976 *regulation of evolvability.* Crit Rev Biochem Mol Biol, 2007. 42(5): p. 399-435.
977 65. Arlt, M.F., T.E. Wilson, and T.W. Glover, *Replication stress and mechanisms of CNV formation.*
978 Curr Opin Genet Dev, 2012. 22(3): p. 204-10.
979 66. Glover, T.W., T.E. Wilson, and M.F. Arlt, *Fragile sites in cancer: more than meets the eye.* Nat Rev
980 Cancer, 2017. 17(8): p. 489-501.

981 67. Phillips, M.A., R. Gujjar, N.A. Malmquist, J. White, F. El Mazouni, J. Baldwin, and P.K. Rathod,
982 *Triazolopyrimidine-Based Dihydroorotate Dehydrogenase Inhibitors with Potent and Selective*
983 *Activity against the Malaria Parasite Plasmodium falciparum.* Journal of Medicinal Chemistry,
984 2008. 51(12): p. 3649-3653.
985 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.
986 Gamo, *P. falciparum in vitro killing rates allow to discriminate between different antimalarial*
987 *mode-of-action.* PLoS One, 2012. 7(2): p. e30949.
988 69. Jensen, J.B. and W. Trager, *Plasmodium falciparum in culture: establishment of additional*
989 *strains.* Am J Trop Med Hyg, 1978. 27(4): p. 743-6.
990 70. Dolan, S.A., L.H. Miller, and T.E. Wellems, *Evidence for a switching mechanism in the invasion of*
991 *erythrocytes by Plasmodium falciparum.* J Clin Invest, 1990. 86(2): p. 618-24.
992 71. Tian, S., H. Yan, C. Neuhauser, and S.L. Slager, *An analytical workflow for accurate variant*
993 *discovery in highly divergent regions.* BMC Genomics, 2016. 17(1): p. 703.
994 72. Sanderson, N.D., J. Swann, L. Barker, J. Kavanagh, S. Hoosdally, D. Crook, G. GonFast
995 Investigators, T.L. Street, and D.W. Eyre, *High precision Neisseria gonorrhoeae variant and*
996 *antimicrobial resistance calling from metagenomic Nanopore sequencing.* Genome Res, 2020.
997 30(9): p. 1354-1363.
998 73. Layer, R.M., C. Chiang, A.R. Quinlan, and I.M. Hall, *LUMPY: a probabilistic framework for*
999 *structural variant discovery.* Genome Biol, 2014. 15(6): p. R84.
1000 74. Cameron, D.L., L. Di Stefano, and A.T. Papenfuss, *Comprehensive evaluation and*
1001 *characterisation of short read general-purpose structural variant calling software.* Nat Commun,
1002 2019. 10(1): p. 3240.
1003 75. Gabrielaite, M., M.H. Torp, M.S. Rasmussen, S. Andreu-Sanchez, F.G. Vieira, C.B. Pedersen, S.
1004 Kinalis, M.B. Madsen, M. Kodama, G.S. Demircan, A. Simonyan, C.W. Yde, L.R. Olsen, R.L. Marvig,
1005 O. Ostrup, M. Rossing, F.C. Nielsen, O. Winther, and F.O. Bagger, *A Comparison of Tools for*
1006 *Copy-Number Variation Detection in Germline Whole Exome and Whole Genome Sequencing*
1007 *Data.* Cancers (Basel), 2021. 13(24).
1008 76. Otto, T.D., U. Bohme, M. Sanders, A. Reid, E.I. Bruske, C.W. Duffy, P.C. Bull, R.D. Pearson, A.
1009 Abdi, S. Dimonte, L.B. Stewart, S. Campino, M. Kekre, W.L. Hamilton, A. Claessens, S.K. Volkman,
1010 D. Ndiaye, A. Amambua-Ngwa, M. Diakite, R.M. Fairhurst, D.J. Conway, M. Franck, C.I. Newbold,
1011 and M. Berriman, *Long read assemblies of geographically dispersed Plasmodium falciparum*
1012 *isolates reveal highly structured subtelomeres.* Wellcome Open Res, 2018. 3: p. 52.
1013 77. Saxena, S. and L. Zou, *Hallmarks of DNA replication stress.* Mol Cell, 2022. 82(12): p. 2298-2314.
1014 78. Guler, J.L., D.L. Freeman, V. Ahyong, R. Patrapuvich, J. White, R. Gujjar, M.A. Phillips, J. DeRisi,
1015 and P.K. Rathod, *Asexual populations of the human malaria parasite, Plasmodium falciparum,*
1016 *use a two-step genomic strategy to acquire accurate, beneficial DNA amplifications.* PLoS
1017 Pathog, 2013. 9(5): p. e1003375.
1018 79. Evrony, G.D., E. Lee, P.J. Park, and C.A. Walsh, *Resolving rates of mutation in the brain using*
1019 *single-neuron genomics.* Elife, 2016. 5.
1020 80. Mills, R.E., K. Walter, C. Stewart, R.E. Handsaker, K. Chen, C. Alkan, A. Abyzov, S.C. Yoon, K. Ye,
1021 R.K. Cheetham, A. Chinwalla, D.F. Conrad, Y. Fu, F. Grubert, I. Hajirasouliha, F. Hormozdiari, L.M.
1022 Iakoucheva, Z. Iqbal, S. Kang, J.M. Kidd, M.K. Konkel, J. Korn, E. Khurana, D. Kural, H.Y. Lam, J.
1023 Leng, R. Li, Y. Li, C.Y. Lin, R. Luo, X.J. Mu, J. Nemesh, H.E. Peckham, T. Rausch, A. Scally, X. Shi,
1024 M.P. Stromberg, A.M. Stutz, A.E. Urban, J.A. Walker, J. Wu, Y. Zhang, Z.D. Zhang, M.A. Batzer, L.
1025 Ding, G.T. Marth, G. McVean, J. Sebat, M. Snyder, J. Wang, K. Ye, E.E. Eichler, M.B. Gerstein, M.E.
1026 Hurles, C. Lee, S.A. McCarroll, J.O. Korbel, and P. Genomes, *Mapping copy number variation by*
1027 *population-scale genome sequencing.* Nature, 2011. 470(7332): p. 59-65.

1028 81. Zhao, M., Q. Wang, Q. Wang, P. Jia, and Z. Zhao, *Computational tools for copy number variation*
1029 *(CNV) detection using next-generation sequencing data: features and perspectives.* BMC
1030 Bioinformatics, 2013. 14 Suppl 11(Suppl 11): p. S1.
1031 82. Guan, P. and W.K. Sung, *Structural variation detection using next-generation sequencing data: A*
1032 *comparative technical review.* Methods, 2016. 102: p. 36-49.
1033 83. Arnot, D.E., E. Ronander, and D.C. Bengtsson, *The progression of the intra-erythrocytic cell cycle*
1034 *of Plasmodium falciparum and the role of the centriolar plaques in asynchronous mitotic division*
1035 *during schizogony.* Int J Parasitol, 2011. 41(1): p. 71-80.
1036 84. Zong, C., S. Lu, A.R. Chapman, and X.S. Xie, *Genome-wide detection of single-nucleotide and*
1037 *copy-number variations of a single human cell.* Science, 2012. 338(6114): p. 1622-6.
1038 85. Lasken, R.S., *Single-cell sequencing in its prime.* Nat Biotechnol, 2013. 31(3): p. 211-2.
1039 86. Liu, S., E.R. Ebel, A. Luniewski, J. Zulawinska, M.L. Simpson, J. Kim, N. Ene, T.W.A. Braukmann, M.
1040 Congdon, W. Santos, E. Yeh, and J.L. Guler, *Direct long read visualization reveals metabolic*
1041 *interplay between two antimalarial drug targets.* bioRxiv, 2023.
1042 87. Kirkman, L.A., E.A. Lawrence, and K.W. Deitsch, *Malaria parasites utilize both homologous*
1043 *recombination and alternative end joining pathways to maintain genome integrity.* Nucleic Acids
1044 Res, 2014. 42(1): p. 370-9.
1045 88. Triglia, T., S.J. Foote, D.J. Kemp, and A.F. Cowman, *Amplification of the multidrug resistance*
1046 *gene pfmdr1 in Plasmodium falciparum has arisen as multiple independent events.* Mol. Cell.
1047 Biol., 1991. 11(10): p. 5244-5250.
1048 89. Luth, M.R., K.P. Godinez-Macias, D. Chen, J. Okombo, V. Thathy, X. Cheng, S. Daggupati, H.
1049 Davies, S.K. Dhingra, J.M. Economy, R.C.S. Edgar, M.G. Gomez-Lorenzo, E.S. Istvan, J.C. Jado,
1050 G.M. LaMonte, B. Melillo, S. Mok, S.K. Narwal, T. Ndiaye, S. Ottilie, S. Palomo Diaz, H. Park, S.
1051 Pena, F. Rocamora, T. Sakata-Kato, J.L. Small-Saunders, R.L. Summers, P.K. Tumwebaze, M.
1052 Vanaerschot, G. Xia, T. Yeo, A. You, F.J. Gamo, D.E. Goldberg, M.C.S. Lee, C.W. McNamara, D.
1053 Ndiaye, P.J. Rosenthal, S.L. Schreiber, G. Serra, J.L. De Siqueira-Neto, T.S. Skinner-Adams, A.C.
1054 Uhlemann, N. Kato, A.K. Lukens, D.F. Wirth, D.A. Fidock, and E.A. Winzeler, *Systematic in vitro*
1055 *evolution in Plasmodium falciparum reveals key determinants of drug resistance.* Science, 2024.
1056 386(6725): p. eadk9893.
1057 90. Matthews, H., C.W. Duffy, and C.J. Merrick, *Checks and balances? DNA replication and the cell*
1058 *cycle in Plasmodium.* Parasit Vectors, 2018. 11(1): p. 216.
1059 91. Bindra, R.S., P.J. Schaffer, A. Meng, J. Woo, K. Maseide, M.E. Roth, P. Lizardi, D.W. Hedley, R.G.
1060 Bristow, and P.M. Glazer, *Down-regulation of Rad51 and decreased homologous recombination*
1061 *in hypoxic cancer cells.* Molecular and Cellular Biology, 2004. 24(19): p. 8504-8518.
1062 92. Saleh-Gohari, N., H.E. Bryant, N. Schultz, K.M. Parker, T.N. Cassel, and T. Helleday, *Spontaneous*
1063 *homologous recombination is induced by collapsed replication forks that are caused by*
1064 *endogenous DNA single-strand breaks.* Mol Cell Biol, 2005. 25(16): p. 7158-69.
1065 93. Hastings, P.J., G. Ira, and J.R. Lupski, *A microhomology-mediated break-induced replication*
1066 *model for the origin of human copy number variation.* PLoS Genet, 2009. 5(1): p. e1000327.
1067 94. Durkin, S.G., R.L. Ragland, M.F. Arlt, J.G. Mulle, S.T. Warren, and T.W. Glover, *Replication stress*
1068 *induces tumor-like microdeletions in FHIT/FRA3B.* Proc Natl Acad Sci U S A, 2008. 105(1): p. 246-
1069 51.
1070 95. Chen, L., W. Zhou, C. Zhang, J.R. Lupski, L. Jin, and F. Zhang, *CNV instability associated with DNA*
1071 *replication dynamics: evidence for replicative mechanisms in CNV mutagenesis.* Hum Mol Genet,
1072 2015. 24(6): p. 1574-83.
1073 96. White, N.J., *Malaria parasite clearance.* Malar J, 2017. 16(1): p. 88.
1074 97. Shor, E., C.A. Fox, and J.R. Broach, *The yeast environmental stress response regulates*
1075 *mutagenesis induced by proteotoxic stress.* PLoS Genet, 2013. 9(8): p. e1003680.

1076 98. WHO, *World Malaria Report 2023.* 2023.
1077 99. Okombo, J. and D.A. Fidock, *Towards next-generation treatment options to combat Plasmodium*
1078 *falciparum malaria.* Nat Rev Microbiol, 2024.
1079 100. Wilson, C., A. Serrano, A. Wasley, M. Bogenschutz, A. Shankar, and D. Wirth, *Amplification of a*
1080 *gene related to mammalian mdr genes in drug-resistant Plasmodium falciparum.* Science, 1989.
1081 244(4909): p. 1184-1186.
1082 101. Nair, S., D. Nash, D. Sudimack, A. Jaidee, M. Barends, A.-C. Uhlemann, S. Krishna, F. Nosten, and
1083 T.J.C. Anderson, *Recurrent Gene Amplification and Soft Selective Sweeps during Evolution of*
1084 *Multidrug Resistance in Malaria Parasites.* Mol Biol Evol, 2007. 24(2): p. 562-573.
1085 102. Ravenhall, M., E.D. Benavente, M. Mipando, A.T. Jensen, C.J. Sutherland, C. Roper, N. Sepulveda,
1086 D.P. Kwiatkowski, J. Montgomery, K.S. Phiri, A. Terlouw, A. Craig, S. Campino, H. Ocholla, and
1087 T.G. Clark, *Characterizing the impact of sustained sulfadoxine/pyrimethamine use upon the*
1088 *Plasmodium falciparum population in Malawi.* Malar J, 2016. 15(1): p. 575.
1089 103. Leroy, D., F. Macintyre, Y. Adoke, S. Ouoba, A. Barry, G. Mombo-Ngoma, J.M. Ndong Ngomo, R.
1090 Varo, Y. Dossou, A.K. Tshefu, T.T. Duong, B.Q. Phuc, B. Laurijssens, R. Klopper, N. Khim, E.
1091 Legrand, and D. Menard, *African isolates show a high proportion of multiple copies of the*
1092 *Plasmodium falciparum plasmepsin-2 gene, a piperaquine resistance marker.* Malar J, 2019.
1093 18(1): p. 126.
1094 104. Bopp, S., P. Magistrado, W. Wong, S.F. Schaffner, A. Mukherjee, P. Lim, M. Dhorda, C.
1095 Amaratunga, C.J. Woodrow, E.A. Ashley, N.J. White, A.M. Dondorp, R.M. Fairhurst, F. Ariey, D.
1096 Menard, D.F. Wirth, and S.K. Volkman, *Plasmepsin II-III copy number accounts for bimodal*
1097 *piperaquine resistance among Cambodian Plasmodium falciparum.* Nat Commun, 2018. 9(1): p.
1098 1769.
1099 105. Dharia, N., A. Sidhu, M. Cassera, S. Westenberger, S. Bopp, R. Eastman, D. Plouffe, S. Batalov, D.
1100 Park, S. Volkman, D. Wirth, Y. Zhou, D. Fidock, and E. Winzeler, *Use of high-density tiling*
1101 *microarrays to identify mutations globally and elucidate mechanisms of drug resistance in*
1102 *Plasmodium falciparum.* Genome Biology, 2009. 10(2): p. R21.
1103 106. Rottmann, M., C. McNamara, B.K. Yeung, M.C. Lee, B. Zou, B. Russell, P. Seitz, D.M. Plouffe, N.V.
1104 Dharia, J. Tan, S.B. Cohen, K.R. Spencer, G.E. Gonzalez-Paez, S.B. Lakshminarayana, A. Goh, R.
1105 Suwanarusk, T. Jegla, E.K. Schmitt, H.P. Beck, R. Brun, F. Nosten, L. Renia, V. Dartois, T.H. Keller,
1106 D.A. Fidock, E.A. Winzeler, and T.T. Diagana, *Spiroindolones, a potent compound class for the*
1107 *treatment of malaria.* Science, 2010. 329(5996): p. 1175-80.
1108 107. Ross, L.S., F.J. Gamo, M.J. Lafuente-Monasterio, O.M. Singh, P. Rowland, R.C. Wiegand, and D.F.
1109 Wirth, *In Vitro Resistance Selections for Plasmodium falciparum Dihydroorotate Dehydrogenase*
1110 *Inhibitors Give Mutants with Multiple Point Mutations in the Drug-binding Site and Altered*
1111 *Growth.* J Biol Chem, 2014. 289(26): p. 17980-95.
1112 108. Phillips, M.A., J. Lotharius, K. Marsh, J. White, A. Dayan, K.L. White, J.W. Njoroge, F. El Mazouni,
1113 Y. Lao, S. Kokkonda, D.R. Tomchick, X. Deng, T. Laird, S.N. Bhatia, S. March, C.L. Ng, D.A. Fidock,
1114 S. Wittlin, M. Lafuente-Monasterio, F.J. Benito, L.M. Alonso, M.S. Martinez, M.B. Jimenez-Diaz,
1115 S.F. Bazaga, I. Angulo-Barturen, J.N. Haselden, J. Louttit, Y. Cui, A. Sridhar, A.M. Zeeman, C.
1116 Kocken, R. Sauerwein, K. Dechering, V.M. Avery, S. Duffy, M. Delves, R. Sinden, A. Ruecker, K.S.
1117 Wickham, R. Rochford, J. Gahagen, L. Iyer, E. Riccio, J. Mirsalis, I. Bathhurst, T. Rueckle, X. Ding,
1118 B. Campo, D. Leroy, M.J. Rogers, P.K. Rathod, J.N. Burrows, and S.A. Charman, *A long-duration*
1119 *dihydroorotate dehydrogenase inhibitor (DSM265) for prevention and treatment of malaria.* Sci
1120 Transl Med, 2015. 7(296): p. 296ra111.
1121 109. Cowell, A. and E. Winzeler, *Exploration of the Plasmodium falciparum Resistome and Druggable*
1122 *Genome Reveals New Mechanisms of Drug Resistance and Antimalarial Targets.* Microbiol
1123 Insights, 2018. 11: p. 1178636118808529.

1124 110. Rocamora, F., L. Zhu, K.Y. Liong, A. Dondorp, O. Miotto, S. Mok, and Z. Bozdech, *Oxidative stress*
1125 *and protein damage responses mediate artemisinin resistance in malaria parasites.* PLoS Pathog,
1126 2018. 14(3): p. e1006930.
1127 111. Thaithong, S., L.C. Ranford-Cartwright, N. Siripoon, P. Harnyuttanakorn, N.S. Kanchanakhan, A.
1128 Seugorn, K. Rungsihirunrat, P.V. Cravo, and G.H. Beale, *Plasmodium falciparum: gene mutations*
1129 *and amplification of dihydrofolate reductase genes in parasites grown in vitro in presence of*
1130 *pyrimethamine.* Exp Parasitol, 2001. 98(2): p. 59-70.
1131 112. Istvan, E.S., N.V. Dharia, S.E. Bopp, I. Gluzman, E.A. Winzeler, and D.E. Goldberg, *Validation of*
1132 *isoleucine utilization targets in Plasmodium falciparum.* Proc Natl Acad Sci U S A, 2011. 108(4): p.
1133 1627-32.
1134 113. Miles, A., Z. Iqbal, P. Vauterin, R. Pearson, S. Campino, M. Theron, K. Gould, D. Mead, E. Drury, J.
1135 O'Brien, V. Ruano Rubio, B. MacInnis, J. Mwangi, U. Samarakoon, L. Ranford-Cartwright, M.
1136 Ferdig, K. Hayton, X.Z. Su, T. Wellems, J. Rayner, G. McVean, and D. Kwiatkowski, *Indels,*
1137 *structural variation, and recombination drive genomic diversity in Plasmodium falciparum.*
1138 Genome Res, 2016. 26(9): p. 1288-99.
1139 114. Rathod, P.K., T. McErlean, and P.C. Lee, *Variations in frequencies of drug resistance in*
1140 *Plasmodium falciparum.* Proc Natl Acad Sci U S A, 1997. 94(17): p. 9389-93.
1141 115. McDaniels, J.M., A.C. Huckaby, S.A. Carter, S. Lingeman, A. Francis, M. Congdon, W. Santos, P.K.
1142 Rathod, and J.L. Guler, *Extrachromosomal DNA amplicons in antimalarial-resistant Plasmodium*
1143 *falciparum.* Mol Microbiol, 2021. 115(4): p. 574-590.
1144 116. Sugino, A. and K. Nakayama, *DNA polymerase alpha mutants from a Drosophila melanogaster*
1145 *cell line.* Proc Natl Acad Sci U S A, 1980. 77(12): p. 7049-53.
1146 117. Inselburg, J. and H.S. Banyal, *Plasmodium falciparum: synchronization of asexual development*
1147 *with aphidicolin, a DNA synthesis inhibitor.* Exp Parasitol, 1984. 57(1): p. 48-54.
1148 118. Brown, A.C., C.C. Moore, and J.L. Guler, *Cholesterol-dependent enrichment of understudied*
1149 *erythrocytic stages of human Plasmodium parasites.* Sci Rep, 2020. 10(1): p. 4591.
1150 119. Rodrigues, O.R. and S. Monard, *A rapid method to verify single-cell deposition setup for cell*
1151 *sorters.* Cytometry A, 2016. 89(6): p. 594-600.
1152 120. 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-*
1153 *Temperature Loop-Mediated Isothermal Amplification Operating at Physiological Temperature.*
1154 Biosensors (Basel), 2023. 13(3).
1155 121. Brown, N., C. da Silva, C. Webb, D. Matias, B. Dias, B. Cancio, M. Silva, R. Viegas, C. Salvador, N.
1156 Chivale, S. Luis, P. Arnaldo, J. Zulawinska, C.C. Moore, F. Nogueira, and J.L. Guler, *Antimalarial*
1157 *resistance risk in Mozambique detected by a novel quadruplex droplet digital PCR assay.*
1158 Antimicrob Agents Chemother, 2024. 68(7): p. e0034624.
1159 122. Daniels, R., D. Ndiaye, M. Wall, J. McKinney, P.D. Sene, P.C. Sabeti, S.K. Volkman, S. Mboup, and
1160 D.F. Wirth, *Rapid, field-deployable method for genotyping and discovery of single-nucleotide*
1161 *polymorphisms associated with drug resistance in Plasmodium falciparum.* Antimicrob Agents
1162 Chemother, 2012. 56(6): p. 2976-86.
1163 123. Kassaza, K., A.C. Long, J.M. McDaniels, M. Andre, W. Fredrickson, D. Nyehangane, P. Orikiriza,
1164 D.J. Operario, J. Bazira, J.A. Mwanga-Amumpaire, C.C. Moore, J.L. Guler, and Y. Boum, 2nd,
1165 *Surveillance of Plasmodium falciparum pfcrt haplotypes in southwestern uganda by high-*
1166 *resolution melt analysis.* Malar J, 2021. 20(1): p. 114.
1167 124. Alker, A.P., V. Mwapasa, and S.R. Meshnick, *Rapid real-time PCR genotyping of mutations*
1168 *associated with sulfadoxine-pyrimethamine resistance in Plasmodium falciparum.* Antimicrob
1169 Agents Chemother, 2004. 48(8): p. 2924-9.

125. Chiang, C., R.M. Layer, G.G. Faust, M.R. Lindberg, D.B. Rose, E.P. Garrison, G.T. Marth, A.R. Quinlan, and I.M. Hall, *SpeedSeq: ultra-fast personal genome analysis and interpretation.* Nat Methods, 2015. 12(10): p. 966-8.
126. Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and S. Genome Project Data Processing, *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. 25(16): p. 2078-9.
127. Garcia-Alcalde, F., K. Okonechnikov, J. Carbonell, L.M. Cruz, S. Gotz, S. Tarazona, J. Dopazo, T.F. Meyer, and A. Conesa, *Qualimap: evaluating next-generation sequencing alignment data.* Bioinformatics, 2012. 28(20): p. 2678-9.
128. Huang, L., F. Ma, A. Chapman, S. Lu, and X.S. Xie, *Single-Cell Whole-Genome Amplification and Sequencing: Methodology and Applications.* Annu Rev Genomics Hum Genet, 2015. 16: p. 79-102.
129. Chen, C., D. Xing, L. Tan, H. Li, G. Zhou, L. Huang, and X.S. Xie, *Single-cell whole-genome analyses by Linear Amplification via Transposon Insertion (LIANTI).* Science, 2017. 356(6334): p. 189-194.
130. McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M.A. DePristo, *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.* Genome Res, 2010. 20(9): p. 1297-303.
131. DePristo, M.A., E. Banks, R. Poplin, K.V. Garimella, J.R. Maguire, C. Hartl, A.A. Philippakis, G. del Angel, M.A. Rivas, M. Hanna, A. McKenna, T.J. Fennell, A.M. Kernytsky, A.Y. Sivachenko, K. Cibulskis, S.B. Gabriel, D. Altshuler, and M.J. Daly, *A framework for variation discovery and genotyping using next-generation DNA sequencing data.* Nat Genet, 2011. 43(5): p. 491-8.
132. Van der Auwera, G.A., M.O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K.V. Garimella, D. Altshuler, S. Gabriel, and M.A. DePristo, *From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline.* Curr Protoc Bioinformatics, 2013. 43(1110): p. 11 10 1-11 10 33.
133. MalariaGen, A. Ahouidi, M. Ali, J. Almagro-Garcia, A. Amambua-Ngwa, C. Amaratunga, R. Amato, L. Amenga-Etego, B. Andagalu, T.J.C. Anderson, V. Andrianaranjaka, T. Apinjoh, C. Ariani, E.A. Ashley, S. Auburn, G.A. Awandare, H. Ba, V. Baraka, A.E. Barry, P. Bejon, G.I. Bertin, M.F. Boni, S. Borrmann, T. Bousema, O. Branch, P.C. Bull, G.B.J. Busby, T. Chookajorn, K. Chotivanich, A. Claessens, D. Conway, A. Craig, U. D'Alessandro, S. Dama, N.P. Day, B. Denis, M. Diakite, A. Djimde, C. Dolecek, A.M. Dondorp, C. Drakeley, E. Drury, P. Duffy, D.F. Echeverry, T.G. Egwang, B. Erko, R.M. Fairhurst, A. Faiz, C.A. Fanello, M.M. Fukuda, D. Gamboa, A. Ghansah, L. Golassa, S. Goncalves, W.L. Hamilton, G.L.A. Harrison, L. Hart, C. Henrichs, T.T. Hien, C.A. Hill, A. Hodgson, C. Hubbart, M. Imwong, D.S. Ishengoma, S.A. Jackson, C.G. Jacob, B. Jeffery, A.E. Jeffreys, K.J. Johnson, D. Jyothi, C. Kamaliddin, E. Kamau, M. Kekre, K. Kluczynski, T. Kochakarn, A. Konate, D.P. Kwiatkowski, M.P. Kyaw, P. Lim, C. Lon, K.M. Loua, O. Maiga-Ascofare, C. Malangone, M. Manske, J. Marfurt, K. Marsh, M. Mayxay, A. Miles, O. Miotto, V. Mobegi, O.A. Mokuolu, J. Montgomery, I. Mueller, P.N. Newton, T. Nguyen, T.N. Nguyen, H. Noedl, F. Nosten, R. Noviyanti, A. Nzila, L.I. Ochola-Oyier, H. Ocholla, A. Oduro, I. Omedo, M.A. Onyamboko, J.B. Ouedraogo, K. Oyebola, R.D. Pearson, N. Peshu, A.P. Phyo, C.V. Plowe, R.N. Price, S. Pukrittayakamee, M. Randrianarivelojosia, J.C. Rayner, P. Ringwald, K.A. Rockett, K. Rowlands, L. Ruiz, D. Saunders, A. Shayo, P. Siba, V.J. Simpson, J. Stalker, X.Z. Su, C. Sutherland, S. Takala-Harrison, L. Tavul, V. Thathy, A. Tshefu, F. Verra, J. Vinetz, T.E. Wellems, J. Wendler, N.J. White, I. Wright, W. Yavo and H. Ye, *An open dataset of Plasmodium falciparum genome variation in 7,000 worldwide samples.* Wellcome Open Res, 2021. 6: p. 42.

1216 134. Abyzov, A., A.E. Urban, M. Snyder, and M. Gerstein, *CNVnator: an approach to discover,*
1217 *genotype, and characterize typical and atypical CNVs from family and population genome*
1218 *sequencing.* Genome Res, 2011. 21(6): p. 974-84.
1219 135. Venkatraman, E.S. and A.B. Olshen, *A faster circular binary segmentation algorithm for the*
1220 *analysis of array CGH data.* Bioinformatics, 2007. 23(6): p. 657-63.
1221 136. Xiao, F., Y. Niu, N. Hao, Y. Xu, Z. Jin, and H. Zhang, *modSaRa: a computationally efficient R*
1222 *package for CNV identification.* Bioinformatics, 2017. 33(15): p. 2384-2385.
1223 137. Mi, H., D. Ebert, A. Muruganujan, C. Mills, L.P. Albou, T. Mushayamaha, and P.D. Thomas,
1224 *PANTHER version 16: a revised family classification, tree-based classification tool, enhancer*
1225 *regions and extensive API.* Nucleic Acids Res, 2021. 49(D1): p. D394-D403.
1226 138. Singh, G. and D. Gupta, *In-Silico Functional Annotation of Plasmodium falciparum Hypothetical*
1227 *Proteins to Identify Novel Drug Targets.* Front Genet, 2022. 13: p. 821516.

1228