A coordinated transcriptional switching network mediates antigenic variation of human malaria parasites

Xu Zhang[1], Francesca Florini[1], Joseph E. Visone[1], Irina Leonardi[2], Mackensie R. Gross[1], Valay Patal[1] and Kirk W. Deitsch[1*]

[1] Department of Microbiology and Immunology, Weill Cornell Medical College, New York, NY, 10065, USA

[2] Jill Roberts Center for Inflammatory Bowel Disease, Weill Cornell Medical College, New York, New York, USA

* To whom correspondence should be addressed. Tel: 011-212-746-4976; Fax: 011-212-746-8587; Email: kwd2001@med.cornell.edu Orcid ID: http://orcid.org/0000-0002-9183-2480

Keywords: mutually exclusive expression, antigenic switching, malaria pathogenesis, immune evasion, gene expression

Abstract: Malaria parasites avoid immune clearance through their ability to systematically alter antigens exposed on the surface of infected red blood cells. This is accomplished by tightly regulated transcriptional control of individual members of a large, multicopy gene family called *var* and is the key to both the virulence and chronic nature of malaria infections. Expression of *var* genes is mutually exclusive and controlled epigenetically, however how large populations of parasites coordinate *var* gene switching to avoid premature exposure of the antigenic repertoire is unknown. Here we provide evidence for a transcriptional network anchored by a universally conserved gene called *var2csa* that coordinates the switching process. We describe a structured switching bias that shifts overtime and could shape the pattern of *var* expression over the course of a lengthy infection. Our results provide an explanation for a previously mysterious aspect of malaria infections and shed light on how parasites possessing a relatively small repertoire of variant antigen encoding genes can coordinate switching events to limit antigen exposure, thereby maintaining chronic infections.

INTRODUCTION

Malaria is a disease of enormous historical significance that continues to exert a heavy burden on health and economic development in many regions of the world (1). The most virulent of the human malaria parasites is *Plasmodium falciparum*, a mosquito-borne pathogen that infects the circulating red blood cells (RBCs) of its hosts. While resident within the RBC, the parasites modify the host cell cytoskeleton through the insertion into the RBC membrane of an adhesive protein called *Plasmodium falciparum* erythrocyte member protein 1 (PfEMP1) (2-4). This enables adhesion to the vascular endothelium (5), thereby avoiding filtration and destruction in the spleen (6, 7), however this also exposes PfEMP1 as a target for adaptive immunity (8, 9). To escape clearance by antibodies that recognize PfEMP1, parasites systematically change the form of PfEMP1 they express, thus continuously altering their antigenic signature and promoting a chronic infection that can extend over a year (reviewed in (10)).

Different forms of PfEMP1 are encoded by individual members of the *var* gene family, a collection of highly variable, paralogous genes numbering between 45 and 90 and distributed throughout the parasite's genome, primarily within subtelomeric regions or as tandemly arranged clusters within the interior of the chromosomes (11). Epigenetic mechanisms ensure that only a single gene is actively transcribed at a time (12, 13), a process with parallels to allelic exclusion of immunoglobulin genes (14) or mutually exclusive expression within the olfactory receptor gene family in mammals (15). However, unlike these systems in which selection of the active gene is part of the differentiation process and thus permanent, *var* gene activation and silencing are reversible, thereby enabling parasites to switch which *var* gene is expressed, thus

63 changing the form of PfEMP1 on the infected cell surface. Many attributes of the
64 epigenetic processes that control *var* gene expression, including the histone
65 modifications that mark a gene for activation or silencing, are shared with model
66 eukaryotes (16-18). However, the added complexity of switching which gene is active
67 while maintaining mutually exclusive expression is largely without precedent outside of
68 pathogenic organisms. Moreover, given that the number of *var* genes within the
69 parasite's genome is relatively limited, switching events must be coordinated to avoid
70 rapid exhaustion of the parasite's variant capacity. Specifically, uncoordinated, random
71 *var* gene switching by individual parasites within a circulating population that can
72 number in the billions would result in rapid exposure of the entire *var* repertoire. In
73 contrast, *P. falciparum* infections are characterized by rising and falling waves of
74 parasitemia, with each wave representing a population of parasites expressing a single
75 or small number of *var* genes (19-21). Modelling studies have suggested that in semi-
76 immune hosts, an underlying structure or coordination to the switching process would
77 be optimal for extending an infection (22-24) by limit activation to a single or small
78 number of genes at a time within the circulating parasite population. However, while
79 communication between parasites has been suggested (25, 26), there is no evidence
80 that this influences *var* gene expression patterns and there does not appear to be a
81 strict switching order within the *var* gene family, therefore how this is accomplished
82 remains entirely without explanation.

83 In 2011, Recker and colleagues explored this phenomenon using mathematical
84 modelling based on data from switching events observed in cultured parasites (22).
85 They proposed that parasites do not switch directly from one gene to another, but rather

86 that a population of parasites switches from a dominant *var* gene, through a switch-
87 intermediate state in which many genes are transiently expressed, to either a new
88 dominant transcript or back to the original. This model also predicted that specific genes
89 could serve as either “source” or “sink” nodes and provide additional structure to the
90 network. The model provided a compelling explanation for how a population of parasites
91 that can reach up to 10^{10} individuals can display seemingly coordinated switching
92 patterns without the need for communication between cells or a fixed order of gene
93 activation. Further, it described both how parasites can extend a chronic infection as
94 well as successfully re-infect individuals when partial immunity exists. However, to date
95 no experimental data have provided a molecular basis for such a structured network,
96 the hypothetical switch-intermediate state or the source/sink nodes predicted by this
97 model.

98 Here we describe molecular genetic evidence that fulfils many of the predictions
99 of the previously proposed mathematical model. We identified a universally conserved,
100 unique *var* gene that displays characteristics of a sink node and plays a key role in
101 coordinating transcriptional switching events. Deletion of this locus disrupts *var* gene
102 switching, resulting in parasites that have a drastically reduced ability to change *var*
103 gene expression, thus disabling the process of antigenic variation that is required to
104 maintain a chronic infection. In addition, we describe an underlying pattern of biased *var*
105 transcriptional activation which shifts overtime, providing the potential for a coordinated
106 pattern of expression switching that could prevent premature exposure of the parasite’s
107 variant repertoire over the length of an infection. These data extend our understanding

108 of antigenic variation beyond mathematical models and represent an important step
109 forward in understanding the pathogenic nature of human malaria.

110

111 RESULTS

112 *Clonal parasite lines display either "single" or "many"* var *gene expression profiles and*
113 *can transition between these two states*

114 The mathematical model derived by Recker and colleagues was based on observations
115 of *var* expression patterns observed in recently isolated clonal parasite populations (22).
116 The proposed transition from expression of a single *var* gene, through an intermediate
117 state in which many genes are expressed, then back to a single gene that becomes
118 dominant in the population is referred to as the "single-many-single" model for *var* gene
119 switching (Figure 1A). This is similar to the two-step model for olfactory receptor gene
120 activation in vertebrates, which has recently been shown to involve an initial, transient
121 state in which large numbers of genes are expressed at a low level, followed by
122 selection of a single gene for stable expression (27). Given the observations from this
123 model system and previous reports of *var* gene switching patterns, we attempted to
124 experimentally validate the SMS model of structured switching and to determine the
125 molecular requirements that underlie the hypothetical *var* gene network that forms the
126 foundation for this pathway.

127 The original proposal of the SMS model was based on analysis of *var* gene
128 expression by Northern blotting and quantitative real time RT-PCR (qRT-PCR) using the
129 IT and 3D7 genetic backgrounds (22). Since this original description, the method for

130 determining *var* expression patterns has become widely used for NF54 and 3D7
131 through the further refinement of a standardized qRT-PCR method developed
132 specifically for this genetic background (28). This standardized method enables rapid,
133 quantitative assessment of the expression level of each *var* gene in the parasite's
134 genome. We employed this method to examine *var* gene expression profiles obtained
135 from recently cloned parasite lines derived from a single parent population of 3D7.
136 Specifically, we determined the expression level of each *var* gene within each
137 subcloned population approximately four weeks after cloning, the earliest time point
138 from which we could obtain suitable parasite numbers. Interestingly, these subcloned
139 populations displayed two fundamentally different expression patterns. Clones
140 displayed either dominant, relatively high-level expression of a single *var* gene with low-
141 level expression of other members of the family, as exemplified by clone 1 (Figure 1B,
142 top panel), or alternatively they lacked expression of a dominant gene and instead only
143 displayed heterogeneous, low-level expression of a large portion of the *var* gene family,
144 as shown by clone 2 (Figure 1B, bottom panel). Measurement of the total level of *var*
145 gene expression from all members of the family detected significantly fewer *var*
146 transcripts in the expression profile of clone 2 (Figure 1C), consistent with the lack of a
147 dominantly expressed gene and indicating that overall *var* expression is lower in this
148 population in addition to being more heterogenous. Given that the number of
149 generations after cloning was identical for both populations, these differences in *var*
150 gene expression profiles presumably reflect differences in the initial *var* expression state
151 or in switching frequency. Similar, rather dramatic differences in total *var* expression
152 levels have been previously described for clonal lines derived from both the NF54 and

153 IT parasite isolates (29, 30), suggesting this phenomenon is typical of cultured parasites
154 of different genetic backgrounds. The two phenotypes, either high-level, stable
155 expression of a dominant *var* gene or highly diverse, low-level expression of many
156 genes, are consistent with the “single” and “many” expression states proposed in the
157 SMS model (22).

158 If parasites transition between the single and many states as part of *var*
159 expression switching as proposed in the SMS model, and if this model applies to the
160 two different *var* expression profiles we observed in our recently cloned lines, then the
161 phenotypes should be reversible. Specifically, it should be possible to obtain “single”
162 parasites from a population of parasites that cumulatively expresses “many” *var* genes,
163 and vice versa. To test this hypothesis, we used serial subcloning by limiting dilution to
164 isolate and examine a large number of additional subclones (Figure 1D). As predicted,
165 we again obtained parasite populations that either displayed low-level expression of
166 many *var* genes or high-level, stable expression of one or a small number of genes,
167 regardless of which phenotype was displayed by the population from which the clone
168 was derived (Figure 1D; additional clones are shown in Figure 2B). Thus, it appears that
169 parasites can transition between these two states, leading to populations that display
170 heterogenous expression of many *var* genes or relatively stable expression of a single,
171 dominant *var* gene. In total, 74 recently subcloned lines were isolated and examined,
172 and each subclone was determined to be primarily in either the “single” or “many” state.
173 When total *var* transcript levels were determined by qRT-PCR, levels were significantly
174 lower in the “many” lines (Figure 1E), consistent with the initial observations shown in
175 Figure 1B and the SMS model. To ensure that this phenomenon was not a property

176 unique to parasites that had been in continuous culture for decades, we repeated the
177 subcloning a *var* expression analysis with a line of NF54 (the original isolate from which
178 3D7 was derived) that had not been in culture for as many replicative cycles and that
179 maintains many characteristics of parasites recently isolated from the field, for example
180 gametocytogenesis and knob formation. These subcloned lines similarly displayed
181 either the “single” or “many” phenotypes, with the “singles” displayed much higher levels
182 of total *var* transcripts (Supplementary Figure 1).

183

184 *Distinct semi-conserved pattern of* var *gene transcription underlies the “single” and*
185 *“many” states and shifts slowly over time*

186 In our *var* expression analysis of recently cloned parasites populations, we observed
187 that populations in both the “many” and “single” states expressed a subset of the *var*
188 gene family at a low level (Figure 1C). Importantly, the entire *var* gene family is not
189 equally represented within the subset of minor transcripts, suggesting that this low-level
190 *var* gene activation is consistently biased toward certain *var* genes. If such intrinsic
191 switching biases shift over time, they could shape the trajectory of *var* gene expression
192 over the course of an infection and provide a model for how large populations of
193 parasites could systematically cycle through their complement of *var* genes. This would
194 lead to semi-coordinated *var* gene switching, thereby avoiding antibody mediated
195 clearance while protecting the majority of the variant repertoire. If true, this hypothesis
196 provides a key insight into how antigenic variation with a small repertoire of genes can

197 maintain a chronic infection, however direct evidence for shifting switching biases within
198 the context of the SMS model has been lacking.

199 To specifically examine minor transcript expression, we compared the patterns of
200 minor transcripts from five of the newly isolated clonal populations shown by asterisks in
201 Figure 1D that displayed both the “single” and “many” expression profiles. RNA was
202 obtained on days 0, 10, 35, 49, 75 and 105 after initiation of the experiment and the
203 expression level of each *var* gene determined by qRT-PCR. The dominant *var* transcript
204 was removed from the expression profile of “single” populations and the expression
205 levels of the remainder of the *var* gene family displayed in a heatmap, thus enabling
206 easy visualization of the minor transcript expression pattern. We used this method to
207 determine if a distinct pattern exists and if it shifts overtime (Figure 2A). The heatmap
208 shows that there are subsets of *var* genes that are more highly represented within the
209 profiles of minor transcripts, and this pattern is largely shared within this group of closely
210 related subclones. However, the pattern is not fixed and displays variation between
211 subclones and over time. Two additional parasite populations that are genetically
212 identical but had been cultured separately for several years (clones 4 and 5) displayed
213 entirely different patterns of minor transcripts, providing additional evidence that the
214 pattern is not fixed (Figure 2A). The similarity of the patterns displayed by the five
215 subclones examined here likely reflects that these populations originated from a
216 common original clonal population.

217 To further investigate the shifting pattern of *var* minor transcript expression, we
218 employed a parallel approach. We again performed serial subcloning starting from
219 subclone 3.1, thereby obtaining additional closely related populations. If the pattern of

220 minor transcripts is semi-conserved and shifts slowly over time, it should be largely
221 shared in all the subcloned populations, regardless of whether they display a “single” or
222 “many” *var* expression profile, or if they express different dominant *var* genes. Rather
223 than following individual populations overtime, we instead isolated three “subclone
224 generations” (Figure 2B) and determined both the dominant *var* transcript as well as the
225 pattern of minor *var* transcripts for each subcloned population. Similar to our previous
226 observations, the subcloned populations displayed both “single” and “many” expression
227 patterns, and the dominant *var* gene varied amongst the “single” subclones (Figure 2B).
228 Comparison of the *var* minor transcripts once again detected the existence of a distinct
229 expression pattern that was semi-conserved but not fixed within this group of closely
230 related populations (Figure 2C).

231

232 *Switching events converge overtime to* var2csa*, a conserved* var *locus*

233 Previously, Mok and colleagues observed that after extended growth in culture, most
234 parasite populations eventually converged to stable expression of a unique, highly
235 conserved *var* gene called *var2csa* (UpsE), proposing that this gene could represent a
236 default choice for *var* gene switching (31). We similarly previously observed
237 convergence to *var2csa* expression when we artificially induced accelerated *var* gene
238 switching by altering the activity of histone modifiers (32, 33), further suggesting that the
239 *var2csa* locus could occupy a unique position in the *var* switching hierarchy. To
240 determine if *var2csa* likewise displayed preferential activation in our recently subcloned
241 parasite populations, we more closely examined *var* transcript prevalence over time in

242 the expression profiles for the five clonal lines shown in Figure 2A. Consistent with
243 previous observations (31, 33), all the clones, regardless of their initial switching
244 frequency, eventually displayed significant expression levels of *var2csa* (Figure 3A). In
245 each instance, switching to *var2csa* was highly favoured, suggesting that this *var* gene
246 might occupy a unique position within the *var* gene switching hierarchy.

247

248 *The* var2csa *locus is required for efficient* var *gene switching*

249 With respect to the SMS model, *var2csa* displays the properties of a sink node, and is
250 by far the most likely *var* gene to become highly activated in our cultures (Figure 3A).
251 This gene also displays several other properties that make it an atypical member of the
252 family, including universal conservation extending to related species that infect
253 chimpanzees and gorillas (34, 35) and the presence of a unique upstream regulatory
254 region that includes an upstream open reading frame (uORF) resulting in translational
255 repression of the mRNA (36-38). The gene can therefore transcribe mRNA without
256 producing PfEMP1 and thus could be transcriptionally activated repeatedly over the
257 course of an infection without generating an antibody response by the host. Based on
258 these unusual properties, we hypothesized that *var2csa* represents the dominant sink
259 node at the center of the SMS network. If correct, we predicted that loss of *var2csa*
260 would significantly alter how parasites undergo *var* gene switching. We recently isolated
261 numerous independent clonal parasite lines in which the entire subtelomeric region
262 surrounding *var2csa* was deleted from the parasite's genome and the chromosome end
263 "healed" through de novo telomere addition (Figure 3B) (39). To complement this set of

264 deletions, we employed CRISPR/Cas9 mediated genome editing to replace the *var2csa*
265 upstream region with a plasmid construct that included a drug selectable marker,
266 thereby deleting the entire promoter region and rendering the gene non-functional
267 (Figure 3C). Several independent clones were obtained from independent transfections
268 of both 3D7 and NF54 that had this integration event and the structure of the resulting
269 locus was verified by whole genome sequencing. For all subsequent experiments, we
270 observed no differences in phenotypes between the clones resulting from telomeric
271 truncations and those derived from deletion of the *var2csa* promoter region.

272 To examine the effect of disruption of *var2csa* on *var* gene expression, *var*
273 expression profiles were examined over an extended time in culture using qRT-PCR,
274 and five representative clones (two derived from CRISPR-mediated disruption of the
275 locus and three from subtelomeric deletion) are shown in Figure 4A. For all the clones,
276 the initial *var* expression profiles resembled the two phenotypes previously observed
277 when wildtype parasite populations were examined (i.e. either “single” or “many”). For
278 example, two clonal lines (V2dis1 and ΔV2_3) displayed expression of a dominant *var*
279 gene while three other clones (V2dis2, ΔV2_1 and ΔV2_2) displayed a heterogeneous
280 mixture of numerous *var* transcripts (Figure 4A). These results suggest that the *var2csa*
281 locus is not required for *var* gene activation or expression, since all the isolated clones
282 were able to express *var* genes. To determine if *var2csa* has a role in *var* gene
283 switching, these 5 clones were cultured continuously in parallel for various periods of
284 time, ranging from 8 months to a year, monitoring *var* gene expression profiles
285 periodically. Given that wildtype parasite populations tend to converge to expression of
286 *var2csa* after extended time in culture (Figure 3A and (31)), we investigated what would

287 happen in the absence of this locus. We observed that all the *var2csa* deleted clones
288 remained remarkably stable in their expression profiles, with the two “single” lines
289 maintaining dominant expression of the same active *var* gene throughout the
290 experiment, a time period extending over a year (Figure 4A). Moreover, the three clones
291 that displayed the “many” state failed to converge to expression of any dominant *var*
292 gene and continued to display highly heterogenous expression profiles (Figure 4A).
293 These data indicate that convergent transcriptional activation is a unique property of
294 *var2csa* and that no other *var* gene within the 3D7 genome can easily substitute upon
295 *var2csa* deletion. In addition, the loss of *var2csa* appears to alter how efficiently
296 parasites undergo *var* gene switching, resulting in parasites that, while some changes in
297 expression are detectable, switch at a much slower rate. These results are consistent
298 with a role for *var2csa* as the dominant sink node at the center of the SMS network.

299

300 *Loss of* var2csa *disrupts the underlying* var *gene transcriptional activation network*

301 Given that the loss of *var2csa* altered *var* gene switching frequency (Figure 4A), we
302 were interested in whether it might also influence the underlying pattern of minor *var*
303 transcript expression that we observed in wildtype parasites (Figure 2). To investigate
304 this, we generated a set of *var2csa*-deleted subclones that were all derived from the
305 same parent clonal population but that displayed different *var* expression profiles
306 (Figure 4B, top). Similar to our previous analysis, we then examined the minor transcript
307 pattern over multiple time points of continuous culture (Figure 4B, heatmap). While
308 patterns of expression are detectable within individual subclones, these patterns were

309 much more diverse than displayed by the wildtype subclones and varied substantially
310 between subclones and over time. To verify this conclusion in an unbiased, empirical
311 fashion, we applied non-metric dimensional scaling based on Bray-Curtis dissimilarity
312 (NMDS), treating the expression levels of each *var* minor transcript as variables (Figure
313 4C), and we included both wildtype and *var2csa*-deleted lines for comparison. The
314 patterns from all the wildtype clones closely clustered within a small region of the plot
315 (shown in shades of light green), indicative of the similarity in the underlying expression
316 patterns. In sharp contrast, the clusters from the different *var2csa*-deleted lines are
317 scattered throughout the plot with only minor overlap, thus displaying a much-reduced
318 stability of the pattern of *var* minor transcripts upon deletion of *var2csa* (Figure 4C). The
319 increased variance within the clusters from the *var2csa*-deleted lines was confirmed by
320 measuring the degree of sample dispersion, which can be inferred by the area on the
321 plot (as defined by the mean distance of each time point from the group centroid)
322 (Figure 4D).

323 Taken together, these data are consistent with the conclusion that the underlying
324 pattern of *var* expression is semi-stable within closely related wildtype parasite
325 populations but is less stable and displays much greater variability in the *var2csa*-
326 deleted lines. If the patterns of minor transcripts analysed here reflect underlying on-
327 switching rates for individual genes, a model emerges in which *var* genes have differing
328 propensities for activation, and that these change overtime, thereby enabling
329 populations of parasites to systematically cycle through their repertoire of *var* genes
330 over the course of an infection. The greater variability displayed by the *var2csa*-deleted
331 lines indicates that the loss of the *var2csa* locus disrupted the organization of the

332 underlying pattern of *var* gene minor transcripts or its stability, which could in turn
333 disrupt or substantially alter the structured pattern of *var* gene transcriptional activation.

334

335 *The "single" and "many"* var *expression states are associated with differences in the*
336 *overall ring-stage transcriptome*

337 To gain additional insights into the two hypothetical *var* expression states, we employed
338 whole transcriptome analysis of both wildtype and *var2csa*-mutated lines. Specifically,
339 we performed RNA-seq from tightly synchronized ring stage (~16 hrs post-invasion) and
340 trophozoite stage (~34 hrs post-invasion) parasite populations. We compared
341 transcriptomes from 11 different lines, including 4 *var2csa*-mutated lines in the "single"
342 state, 3 *var2csa*-mutated lines in the "many" state, and 4 wildtype lines (2 "manys" and
343 2 "singles"). To control for possible effects from plasmid integration, two of the wildtype
344 lines (called WT_rif1 and WT_rif2) contain a plasmid integrated into a *rif* gene,
345 analogous to the integration used to disrupt *var2csa* as shown in Figure 3C. Because
346 we were interested in changes to expression of genes other than those that undergo
347 clonally variant expression (40), we excluded *var, rif, stevor* and *phist* genes and
348 performed the analysis on the remaining protein coding genes in the 3D7 genome.

349 As an initial assessment of gene expression in ring-stage parasites (16 hrs post-
350 invasion), we performed principal component analysis (PCA) based on the mRNA levels
351 of 5721 genes to determine if gene expression patterns are associated with the different
352 *var* expression states. This analysis demonstrated clear clustering between populations
353 that displayed either the "single" or "many" phenotype (Figure 4E). Further, the parasite

354 lines with the *var2csa*-mutant “single” or “many” expression profiles displayed the most
355 polarized patterns of gene expression while the wildtype lines presented intermediate
356 expression patterns. Permutational multivariate ANOVA (ADONIS) confirmed that the
357 *var* expression profiles contributed strongly to gene expression variability (R^2= 0.42768,
358 p=0.001***). This implies that the *var* gene expression state (either “single” or “many”)
359 is associated with changes in the overall transcriptome at the ring-stage of the asexual
360 cycle. Interestingly, this phenomenon was not observed in trophozoites (see below).
361 Our previous qRT-PCR analysis of wildtype “single” and “many” parasite lines found
362 that they display very different total *var* gene expression levels (Figure 1C and E). To
363 confirm this observation, we quantified the *var* transcripts from our RNA-seq analysis
364 from all sequenced lines. Consistent with our previous results, the wildtype “single” lines
365 displayed higher total *var* transcripts than the “many” lines, and these differences were
366 magnified in the *var2csa*-mutant lines (Figure 4F).

367

368 *Transcriptome analysis indicates that “single” and “many” parasites differ in cell cycle*
369 *progression*

370 Further analysis of the transcriptional differences between the populations
371 identified 611 differentially expressed genes (FDR<0.05) when comparing the *var2csa*-
372 mutant “singles” to the *var2csa*-mutant “manys” at the ring stage (Figure 5A, left;
373 Supplementary Table S1). Specifically, we found 562 genes upregulated in “singles”
374 and the remaining 49 genes upregulated in “manys”, and these differences largely
375 disappeared in trophozoites (Figure 5A, right; Supplementary Table S2). Functional and

376 Gene Ontology analyses of the differentially expressed genes only presented broad,
377 relatively unspecific terms and did not identify any enriched metabolic pathways or
378 functions (Supplementary Tables S3 and S4). This result prompted us to investigate
379 whether the observed transcriptional changes were indicative of a shift in cell cycle
380 progression rather than specific changes in cellular functions. We therefore applied two
381 independent methods as described by Poran et al. (41) and Lemieux et al. (42) to
382 estimate the approximate point in the replicative cycle of each of our ring-stage parasite
383 populations. We compared the transcriptomes obtained from our lines to hypothetical
384 transcriptomes derived from modeled parasite populations over a simulated asexual
385 replication time course. Both methods consistently indicated a shift in cell cycle
386 progression, with the *var2csa*-mutant "single" lines displaying an earlier point in the
387 asexual cycle than the *var2csa*-mutant "many" lines and with the wildtype lines
388 displaying an intermediate timepoint (Figure 5B and C). This continuum in cell cycle
389 progression was consistent for all 11 lines, despite the fact that all cultures were
390 synchronized to 16 hours post-invasion (hpi). Once again, the *var2csa*-mutant lines
391 occupied the extreme positions in this continuum, indicating that these lines display
392 more polarized phenotypes (Figure 5B and C, Supplementary Table S5).

393 To further investigate the possibility that the differences in gene expression
394 resulted from a shift in ring-stage progression, we used the PlasmoDB database to
395 examine the transcriptional timing of all the differentially expressed genes. We found
396 that of the 562 genes upregulated in "singles", 559 have higher transcription levels at 0-
397 10 hour hpi compared to 10-20 hpi and similarly, all 49 genes upregulated in "manys"
398 have higher transcription levels at 10-20 hpi compared to 0-10 hpi (Figure 5D and E,

399 Supplementary Tables S3 and S4), providing additional evidence that the transcriptional
400 differences of these populations likely reflect a shift in the replicative cycle. The
401 expression levels of the differentially expressed genes for all 11 lines again confirm that
402 the 4 wildtype lines display an intermediate level of transcription for all the differentially
403 expressed genes (Figure 5F). Taken together, these analyses indicate that being in the
404 "single" or "many" state correlates strongly with how quickly parasites progress through
405 the ring-stage of the cycle. In addition, the *var2csa*-mutant lines display more polarized
406 phenotypes, suggesting that these populations are more homogenous in being either
407 "single" or "many", a conclusion consistent with the hypothesis that in the absence of
408 *var2csa*, parasites are less efficient in transitioning between states. A similar change in
409 ring-stage progression was previously observed in response to treatment with a histone
410 methyltransferase inhibitor (43) that also altered *var* gene expression patterns (33), as
411 well as in parasites that display reduced sensitivity to artemisinin (44, 45), indicating that
412 how quickly parasites progress through the first half of their replicative cycle is
413 somewhat flexible. Interestingly, these differences appear to be largely limited to
414 progression through the ring-stage. Measurement of DNA content of synchronized
415 cultures by flowcytometry indicated that *var2csa*-mutant "single" and "many" populations
416 undergo DNA replication with similar timing as wildtype parasites (Figure 5G), and
417 growth assays did not detect any overall differences in the growth of the cultures
418 (Supplementary Figure 2), indicating that length of the complete replication cycle is not
419 detectably different. Thus, these experiments indicate that the "single" and "many"
420 states are distinct beyond just *var* gene expression, and likely reflect changes in the

421 nuclear or epigenetic organization of the parasite's genome that affect ring-stage
422 progression and associated gene expression patterns.

423

424 DISCUSSION

425 The SMS model for *var* gene transcriptional switching proposes that switching
426 expression from one *var* gene to another involves transition from transcribing a single
427 *var* gene, to many, then back to a single gene again (Figure 1A). This type of two-step
428 selection of a gene for activation is now well documented for the choice of a single
429 transcriptionally active olfactory receptor gene in mammals (46) and more recently for
430 expression of a single *vsg* in metacyclic African trypanosomes (47). In both cases this
431 involves the initial, low-level transcription of multiple genes, which is then limited to a
432 single gene in fully differentiated cells (27, 47-49). In olfactory neurons, this coincides
433 with the stepwise establishment of specific inter-chromosomal contacts and the
434 assembly of subnuclear compartments that limit expression to a single gene (50). Our
435 results indicate that a similar process occurs in malaria parasites during *var* gene
436 switching, a possibility that is consistent with the suggestion that individual parasites
437 can express multiple *var* genes, a phenomenon that was previously reported using
438 immunofluorescence and in situ hybridization (51). If correct, mutually exclusive *var*
439 gene expression is possibly more plastic than previously assumed, although additional
440 work at the single cell level will be required to fully characterize the extent of *var* gene
441 activation in individual parasites in the "many" state.

442 Populations of parasites in which *var2csa* was deleted or disrupted displayed
443 slower *var* gene switching rates (Figures 3) and appear more polarized with respect to
444 the “single” vs “many” states (Figure 5), indicating that they are much less efficient in
445 transitioning between these two states. How the *var2csa* locus contributes to
446 transcriptional switching is not known, although given that the *var2csa* promoter region
447 appears to be important for this function, it is tempting to propose a model of promoter
448 competition, as has been proposed for chromosome choice during X-inactivation in
449 mammals or *vsg* expression site choice in African trypanosomes (47, 52). Such models
450 typically propose that when in the many state, multiple loci compete or “race” for
451 activation, with each locus expressing a low level of transcripts. When one locus
452 reaches a threshold, transcription from the other loci is suppressed, resulting in stable
453 monoallelic expression. If this model applies to *var* gene expression, the unique nature
454 of the *var2csa* promoter could enable it to compete with an actively expressed *var* gene,
455 thereby lowering its expression level to below the threshold and thus allowing the
456 parasite to re-enter the “many” state. This would provide an explanation for why loss of
457 the *var2csa* locus results in stabilization of the active *var* gene and greatly reduces
458 switching. Previous work from Duffy et al (53, 54) showed that alterations to sequences
459 adjacent to the *var2csa* promoter can further increase its competitiveness, making it
460 hypercompetitive and stably expressed, thus suppressing switching rates in a way
461 consistent with this model. The propensities of cultured parasites to converge toward
462 expression of *var2csa* independent of genomic rearrangement is also consistent with its
463 highly competitive nature (31). To complete a switch to an alternative *var* gene,
464 activation of *var2csa* would have to be unstable, a property that could result from its

465 unique structure, including the presence of an upstream open reading frame and an
466 untranslated mRNA (36, 37).

467 The location of a large proportion of the *var* gene family within subtelomeric
468 regions makes them uniquely subject to frequent recombination and deletion events, an
469 inherent plasticity that is reflected in the variable number of *var* genes observed in the
470 genomes of different parasites isolates (55). Interestingly, despite the hyper-
471 recombinogenic nature of the subtelomeric region in which it resides, *var2csa* remains
472 conserved in all *P. falciparum* lines examined, and this conservation extends to the
473 related species *P. reichenowi* and *P. praefalciparum* (35). We hypothesize that its role
474 in mediating expression switching selects for its preservation within the genomes of
475 wildtypes parasites. This hypothesized additional role for *var2csa* is consistent with
476 detection of its transcription non-pregnant individuals (56-58) as well as activation of
477 *var2csa* coincident with *var* gene switching during an experimental human infection
478 (20). We previously documented frequent subtelomeric deletions that included *var*
479 genes on other chromosomes (59, 60), including the subtelomeric regions of
480 chromosomes 2, 3 and 14 in the 3D7 control lines employed here (59), however none of
481 these deletion events resulted in altered *var* switching phenotypes, indicating that the
482 properties we describe here are unique to the *var2csa* locus. The frequently described
483 convergence of cultured parasite populations to expression of *var2csa* could be a result
484 of a selective growth advantage parasites acquire if the *var2csa* transcript only
485 translates the uORF, thus saving the metabolic cost of translating a full-length PfEMP1.
486 While feasible, if true one would expect the “many” parasites to similarly display a

487 selective growth advantage given their very low rate of *var* gene transcription, which we
488 do not observe (Supplementary Figure 2).

489 In addition to changing the frequency of *var* expression switching, deletion of
490 *var2csa* also resulted in major changes in the pattern of minor *var* transcripts displayed
491 by parasite populations (Figure 3). This underlying pattern of transcription has been
492 proposed to reflect switching biases within the *var* gene network (22, 61), thereby
493 imposing a structure on the trajectory that *var* expression takes over the course of an
494 infection. This type of structured genetic network could coordinate *var* gene activation
495 throughout the parasite population without requiring communication between infected
496 cells and thus would enable lengthy infections despite a relatively small number of
497 variant antigen-encoding genes. The shifting pattern of *var* minor transcripts observed in
498 wildtype parasites and shown in Figure 2 is consistent with this hypothesis. The
499 disruption of the pattern of minor transcripts upon deletion of *var2csa* indicates that this
500 gene plays an important role in maintaining the structure of the *var* gene transcriptional
501 network, although the molecular mechanism underlying this phenomenon is not known.
502 Additional work focused on subnuclear genome organization, specifically with regard to
503 heterochromatic regions, and the role that *var2csa* plays in this organization are likely to
504 shed light on this process.

505 The switching biases observed in cultured parasites could help refine the
506 trajectory of *var* expression over the course of an infection, as suggested by the model
507 proposes by Recker et al (22). However, it is important to note that while *var* gene
508 expression patterns displayed by parasite cultures likely reflect inherent switching
509 biases, in a natural infection the host's immune system, in particular existing antibodies

510 against PfEMP1, are likely to exert a profound selective pressure that determines which
511 parasites can successfully establish a high parasitemia. Pre-existing anti-PfEMP1
512 antibodies that recognize any particular variant will select against parasites expressing
513 its encoding *var* gene, thus strongly shaping the *var* expression pattern over the course
514 of an infection. The importance of antibody selection and pre-existing immunity explains
515 why expression patterns cannot be truly hardwired and must maintain flexibility,
516 indicating why switching biases rather than a strict switching order are ideal for
517 maintaining chronic infections.

518 Given the lack of an *in vivo* experimental system for *P. falciparum*, it is difficult to
519 investigate the "single" and "many" states in natural infections. However, an interesting
520 case study of a semi-immune African immigrant to Europe might be informative (62). No
521 parasites were initially detected in this patient by microscopy or by malaria antigen
522 tests, although the individual displayed high titers of antibodies to malaria antigens
523 indicating a significant level of immunity. Upon splenectomy, parasitemia became
524 evident and increased to high levels, indicating a latent infection existed prior to removal
525 of the spleen. The parasites observed in the peripheral circulation represented all
526 stages of the asexual cycle, indicating a lack of cytoadherence, and in vitro binding
527 assays failed to detect adhesion to various endothelial receptors. Further, molecular
528 analysis indicated that these parasites did not express *var* genes at a detectable level.
529 However, upon transfer to in vitro culture, *var* gene expression became easily
530 detectable and the infected cells became cytoadherent. This is reminiscent of our
531 observations that parasites can transition between high-level expression of individual
532 *var* genes (the "single" state) and very low-level *var* gene expression (the "many" state).

533 A similar example of parasites lacking expression of surface antigens upon splenectomy
534 was observed in monkeys infected with *Plasmodium knowlesi* (7), suggesting that this
535 might be a more general phenomenon amongst malaria parasite species that display
536 cytoadherence. It is tempting to speculate that in the presence of high titers of anti-
537 PfEMP1 antibodies, parasites in the "many" state are selected for and thus represent
538 the dominant population found in semi-immune, asymptomatic infected individuals.
539 Consistent with this hypothesis, Kho et al. more recently reported that the vast majority
540 of the parasite biomass in asymptomatic infections is found in the spleen (63), indicative
541 of non-cytoadherent parasites and potentially reflecting lower expression of PfEMP1.
542 Similarly, a recent study of parasites obtained from asymptomatic patients at the end of
543 a dry season detected a lower level of *var* expression than observed in parasites from
544 symptomatic patients, although this difference was not statistically significant (64). Thus,
545 it is possible that in individuals with significant levels of immunity, parasites in the
546 "many" state can persist at a low level, escaping anti-PfEMP1 antibodies through
547 repressed *var* gene expression and maintaining very low parasitemias.

548

549 ACKNOWLEDGEMENTS

550 The Department of Microbiology and Immunology at Weill Medical College of Cornell
551 University acknowledges the support of the William Randolph Hearst Foundation. This
552 work was supported by grants AI 52390 and AI99327 from the National Institutes of
553 Health to KWD. JEV was supported by training grant T32GM008539 from the National
554 Institutes of Health to Weill Cornell Graduate School of Biomedical Sciences. FF

555 received support from Early Postdoc Mobility grant P2BEP3_191777 from the Swiss
556 National Science Foundation. KWD is a Stavros S. Niarchos Scholar and a recipient of
557 a William Randolf Hearst Endowed Faculty Fellowship. The funders had no role in the
558 study design, data collection and analysis, decision to publish, or preparation of the
559 manuscript.

560

561 AUTHOR CONTRIBUTIONS

562 Conceptualization, XZ and KWD; Methodology, XZ, FF, JEV, IL and KWD;
563 Investigation, XZ, FF, JEV, IL, VP and MRG; Writing – Original Draft, XZ and KWD;
564 Writing – Reviewing and Editing, XZ, FF, JEV, IL and KWD; Funding Acquisition, KWD
565 and FF; Supervision, XZ and KWD.

566

567 DECLARATION OF INTERESTS

568 The authors declare no competing interests.

569

570 MATERIALS AND METHODS

571 *Parasite culture*: All parasite strains were derived from the reference strain NF54/3D7
572 and maintained in RPMI 1640 supplemented with 0.5% AlbuMAX II (Invitrogen) in an
573 atmosphere of 5% O_2, 5% CO_2 and 90% N_2 at 37° C and 3% -5% hematocrit.
574 Transgenic lines were maintained under continuous drug-selection. Clonal parasite lines

575 were obtained by limiting dilution (65) in 96 well plates. For analysis of gene expression
576 in subcloned populations, parasites were grown for 5 weeks from the initial isolation of
577 individual parasites. Daily parasitemias were monitored using a Cytek DxP12 flow
578 cytometer.

579 *Generation of transgenic lines*: Parasites with chromosomal truncations that included
580 the *var2csa* locus were generated and described in a previous study (66). Specific
581 disruption of the *var2csa* locus was performed via CRISPR-targeted plasmid integration.
582 Briefly, parasites were transfected by electroporation (67, 68) using derivatives of the
583 plasmids pL6_eGFP and pUF1_Cas9 for CRISPR/Cas9-based genome editing as
584 described by Ghorbal and colleagues (69). Homology blocks were PCR amplified from
585 3D7 genomic DNA, using specific primers flanked by 15bp overlapping regions from
586 pL6-*var2csa*-promoter-deletion plasmid and pUF1_Cas9 (list of primers, see
587 Supplementary Table S8) that allowed the cloning by infusion cloning (Clontech, Takara
588 Bio USA, Mountain View, CA, USA). Plasmid integration was validated by PCR across
589 the site of integration and whole genome sequencing.

590 *RNA extraction, cDNA synthesis and Quantitative RT-PCR*: Parasite RNA was
591 extracted from synchronized late ring-stage parasites as described previously (70).
592 Briefly, ring-stage parasites were collected using a double synchronization approach.
593 Cultures were initially synchronized using 5% sorbitol to select for ring-stages. 36 hours
594 later, late-stage infected erythrocytes (42-48 hours post-invasion) were collected using
595 percoll/sorbitol gradient centrifugation and allowed to reinvade overnight. Ring-stage
596 parasites were collected for RNA extraction 18 hours after enrichment for late-stage
597 infected erythrocytes (12-18 hours post-invasion). After an additional 20 hours in

598 culture, trophozoites were collected for RNA isolation. Synchronization was verified by
599 microscopy and flow cytometry. RNA was extracted with TRiZol (Invitrogen) and purified
600 on PureLink (Invitrogen) columns following the manufacturer's protocols. Isolated RNA
601 was treated with Deoxyribonuclease I (DNase I) (Invitrogen) to degrade contaminating
602 genomic DNA. cDNA was synthesized from approximately 800 ng of RNA in a reaction
603 that included Superscript II RNase H reverse transcriptase (Invitrogen) as described by
604 the manufacturer. Control reactions in absence of reverse transcriptase were employed
605 to verify a lack of gDNA contamination. We employed the qRT-PCR *var* primer set of
606 Salanti et al to detect transcript levels from all *var* genes (71). This primer set was
607 specifically designed and tested to enable absolute quantification of transcript levels.
608 Additional primers were designed and applied to increase the accuracy (list of primers,
609 see Supplementary Table S7). Quantitative PCR was performed using a Quant Studio 6
610 Flex 489 realtime PCR machine (Thermo Fisher) using iTaq Sybr Green (Bio-Rad).
611 Quantities were normalized to seryl-tRNA synthetase (PF3D7_0717700). ΔCT for each
612 individual primer pair was determined by subtracting the individual CT value from the
613 CT value of the control and converting to relative copy numbers with the formula $2^{\Delta CT}$.
614 All Q-RT-PCR assays were performed in a 384-well format PCR machine enabling
615 duplicate or triplicate runs performed simultaneously. Biological replicates were
616 prepared from independent RNA extractions. Relative copy numbers for each *var* gene
617 were determined in Microsoft Excel and transcriptional profiles of individual genes are
618 presented as pie charts or as bar graphs. In bar graphs, standard deviations from
619 biological replicates are shown with error bars.

620 *RNA sequencing and analysis*: Parasite-infected RBCs from highly synchronous
621 cultures containing ring or trophozoites stage parasites were collected for RNAseq.
622 Following RNA isolation, RNA concentrations were measured using the NanoDrop
623 system (Thermo Fisher Scientific, Inc., Waltham, MA). Total RNA integrity was checked
624 using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). rRNA removal,
625 preparation of an RNA sample library and final cDNA library were completed and the
626 libraries pooled for sequenced using an Illumina HiSeq4000 sequencer with paired-end
627 protocol performed by the Genomics Core Laboratory at Weill Cornell Medicine. The
628 raw reads that passed quality control were mapped to the *Plasmodium falciparum*
629 genome (PlasmoDB-9.0_Pfalciparum3D7_Genome) by aligner HISAT2(v.2.2.0) (72,
630 73). Based on the alignment, featureCounts (Package Rsubread v.2.0.1) (74) was used
631 to generate a raw counts matrix for differential expression analysis (DE). We
632 independently performed both DESeq2 (v.3.14) (75) and EdgeR (v.3.14) (76-78) for the
633 differential expression analysis. Genes with a false discovery rate of <=0.05 with a
634 mean FPKM >5 in at least one strain were called significant. To exclude the obvious
635 impact of antigenic variant genes, we excluded *var*, *rif*, *stevor* and *phist* genes from the
636 differential expression analysis. The analysis was conducted with R (v.4.1.0) in Rstudio
637 (1.4.1717). RNA sequencing was performed by the Weill Cornell Genomic core with the
638 following parameters. All samples must have a RIN value greater than 8 for library
639 synthesis. Ribosomal RNA was removed from total RNA using the Ribo Zero Gold for
640 human/mouse/rat kit (Illumina). Using the TruSeq RNA Sample Library Preparation v.2
641 kit (Illumina), RNA was fragmented into small pieces using divalent cations under
642 elevated temperature. Cleaved RNA fragments were copied into first-strand cDNA using

643 reverse transcriptase and random primers. Second-strand cDNA synthesis followed,
644 using DNA Polymerase I and RNaseH. The cDNA fragments then went through an end
645 repair process, the addition of a single 'A' base and ligation of adaptors. The products
646 were then purified and enriched with PCR to create the final cDNA library. Libraries
647 were pooled and sequenced on an Illumina HiSeq4000 sequencer with 100bp Paired
648 End Sequencing (PE100). Total reads and mapped reads for each sample are shown in
649 Supplemental Table S9.

650 *Statistical analysis:* Statistical analysis was performed in R (v.4.1.0). We assessed beta
651 diversity of our samples' *var genes* expression profiles using non-metric multi-
652 dimensional scaling (NMDS) ordination. NMDS plots and Beta dispersion were
653 generated in vegan: Community Ecology Package (v.2.5-7) using Bray-Curtis
654 dissimilarity. Permutational multivariate analysis of variance (PERMANOVA) was
655 performed using the 'adonis' function in vegan. Heatmaps were generated by
656 heatmap.2 from gplots (v. 2.3.0).

657 *Flow cytometry and assays for cell cycle progression:* Progression through the cell cycle
658 was determined by flow cytometric analysis of parasite DNA content as previously
659 described (79). Briefly, parasites were tightly synchronized were stained at 37°C with 16
660 μM Hoechst 33342 for 30 min at 1% hematocrit in incomplete media followed by a
661 single wash in PBS. Cells were then diluted to 0.1% hematocrit in PBS and analyzed
662 using a Cytek DxP11 flow cytometer for Hoechst 33342 DNA-staining (375nm laser,
663 450/50 emission filter).

664

665 *Growth assays:* Cultures were adjusted to 0.05% - 0.5% parasitemia, 5% hematocrit in
666 a total culture volume of 5 ml. Parasitemia was obtained daily by flow cytometry and
667 verified by thin smear stained with Giemsa (Sigma). Parasitemia was allowed to
668 increase over time until reaching a level greater than 0.5% - 1%, at which time cultures
669 were diluted 1 in 10 in culture media containing uninfected RBCs. Changes in
670 parasitemia were calculated by multiplying the daily parasitemia by the exponential
671 dilution factor and the data were graphed on a log scale over time.
672

673 *Materials Availability Statement:* Data Deposition: Whole genome sequence and
674 transcriptome data are available at the BioProject database of the NCBI. The genome
675 sequencing data can be accessed at this link:
676 http://www.ncbi.nlm.nih.gov/bioproject/515738. The RNAseq data can be accessed at
677 this link: https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA802886. All genetically
678 modified parasite lines are available from the authors upon request.

679 REFERENCES
680

681 1. WHO, World Malaria Report 2021. (2021).
682 2. D. I. Baruch *et al.*, Cloning the *P. falciparum* gene encoding PfEMP1, a malarial variant antigen
683 and adherence receptor on the surface of parasitized human erythrocytes. *Cell* **82**, 77-87 (1995).
684 3. J. D. Smith *et al.*, Switches in expression of *Plasmodium falciparum var* genes correlate with
685 changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell* **82**, 101-110
686 (1995).
687 4. X. Su *et al.*, A large and diverse gene family (*var*) encodes 200-350 kD proteins implicated in the
688 antigenic variation and cytoadherence of *Plasmodium falciparum*-infected erythrocytes. *Cell* **82**,
689 89-100 (1995).
690 5. J. P. Gardner, R. A. Pinches, D. J. Roberts, C. I. Newbold, Variant antigens and endothelial
691 receptor adhesion in *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences*
692 *USA* **93**, 3503-3508 (1996).
693 6. P. H. David, M. Hommel, L. H. Miller, I. J. Udeinya, L. D. Oligino, Parasite sequestration in
694 *Plasmodium falciparum* malaria: Spleen and antibody modulation of cytoadherence of infected
695 erythrocytes. *Proceedings of the National Academy of Sciences USA* **80**, 5075-5079 (1983).
696 7. J. W. Barnwell, R. J. Howard, H. G. Coon, L. H. Miller, Splenic requirement for antigenic variation
697 and expression of the variant antigen on the erythrocyte membrane in cloned Plasmodium
698 knowlesi malaria. *Infect Immun* **40**, 985-994 (1983).
699 8. P. C. Bull *et al.*, Parasite antigens on the infected red cell surface are targets for naturally
700 acquired immunity to malaria. *Nat. Med* **4**, 358-360 (1998).
701 9. H. A. Giha *et al.*, Antibodies to variable Plasmodium falciparum-infected erythrocyte surface
702 antigens are associated with protection from novel malaria infections. *Immunol. Lett* **71**, 117-
703 126 (2000).
704 10. K. W. Deitsch, R. Dzikowski, Variant Gene Expression and Antigenic Variation by Malaria
705 Parasites. *Annu. Rev. Microbiol* **71**, 625-641 (2017).
706 11. T. D. Otto *et al.*, Long read assemblies of geographically dispersed Plasmodium falciparum
707 isolates reveal highly structured subtelomeres. *Wellcome. Open. Res* **3**, 52 (2018).
708 12. A. Scherf *et al.*, Antigenic variation in malaria: *in situ* switching, relaxed and mutually exclusive
709 transcription of *var* genes during intra-erythrocytic development in *Plasmodium falciparum*.
710 *EMBO J* **17**, 5418-5426 (1998).
711 13. K. W. Deitsch, A. del Pinal, T. E. Wellems, Intra-cluster recombination and var transcription
712 switches in the antigenic variation of Plasmodium falciparum. *Mol. Biochem. Parasitol* **101**, 107-
713 116 (1999).
714 14. A. E. Corcoran, Immunoglobulin locus silencing and allelic exclusion. *Semin. Immunol* **17**, 141-
715 154 (2005).
716 15. I. Rodriguez, Singular expression of olfactory receptor genes. *Cell* **155**, 274-277 (2013).
717 16. T. Chookajorn *et al.*, Epigenetic memory at malaria virulence genes. *Proc. Natl. Acad. Sci. U. S. A*
718 **104**, 899-902 (2007).
719 17. J. J. Lopez-Rubio, L. Mancio-Silva, A. Scherf, Genome-wide analysis of heterochromatin
720 associates clonally variant gene regulation with perinuclear repressive centers in malaria
721 parasites. *Cell Host. Microbe* **5**, 179-190 (2009).
722 18. C. Flueck *et al.*, Plasmodium falciparum heterochromatin protein 1 marks genomic loci linked to
723 phenotypic variation of exported virulence factors. *PLoS. Pathog* **5**, e1000569 (2009).
724 19. A. Bachmann *et al.*, Highly co-ordinated var gene expression and switching in clinical
725 Plasmodium falciparum isolates from non-immune malaria patients. *Cell Microbiol* **13**, 1397-
726 1409 (2011).

727 20. A. Bachmann *et al.*, Controlled human malaria infection with Plasmodium falciparum
728 demonstrates impact of naturally acquired immunity on virulence gene expression. *PLoS Pathog*
729 **15**, e1007906 (2019).
730 21. M. Kaestli, A. Cortes, M. Lagog, M. Ott, H. P. Beck, Longitudinal assessment of Plasmodium
731 falciparum var gene transcription in naturally infected asymptomatic children in Papua New
732 Guinea. *J. Infect. Dis* **189**, 1942-1951 (2004).
733 22. M. Recker *et al.*, Antigenic variation in Plasmodium falciparum malaria involves a highly
734 structured switching pattern. *PLoS. Pathog* **7**, e1001306 (2011).
735 23. R. Noble, M. Recker, A statistically rigorous method for determining antigenic switching
736 networks. *PLoS. ONE* **7**, e39335 (2012).
737 24. R. Noble *et al.*, The antigenic switching network of Plasmodium falciparum and its implications
738 for the immuno-epidemiology of malaria. *Elife. (Cambridge)* **2**, e01074 (2013).
739 25. P. Y. Mantel *et al.*, Malaria-infected erythrocyte-derived microvesicles mediate cellular
740 communication within the parasite population and with the host immune system. *Cell Host*
741 *Microbe* **13**, 521-534 (2013).
742 26. N. Regev-Rudzki *et al.*, Cell-cell communication between malaria-infected red blood cells via
743 exosome-like vesicles. *Cell* **153**, 1120-1133 (2013).
744 27. L. Tan, Q. Li, X. S. Xie, Olfactory sensory neurons transiently express multiple olfactory receptors
745 during development. *Mol Syst Biol* **11**, 844 (2015).
746 28. A. Salanti *et al.*, Selective upregulation of a single distinctly structured var gene in chondroitin
747 sulphate A-adhering Plasmodium falciparum involved in pregnancy-associated malaria.
748 *Molecular Microbiology* **49**, 179-191 (2003).
749 29. C. J. Merrick *et al.*, Functional analysis of sirtuin genes in multiple Plasmodium falciparum
750 strains. *PLoS. ONE* **10**, e0118865 (2015).
751 30. J. H. Janes *et al.*, Investigating the host binding signature on the Plasmodium falciparum PfEMP1
752 protein family. *PLoS Pathog* **7**, e1002032 (2011).
753 31. B. W. Mok *et al.*, Default Pathway of var2csa switching and translational repression in
754 Plasmodium falciparum. *PLoS. ONE* **3**, e1982 (2008).
755 32. U. E. Ukaegbu *et al.*, Recruitment of PfSET2 by RNA polymerase II to variant antigen encoding
756 loci contributes to antigenic variation in P. falciparum. *PLoS. Pathog* **10**, e1003854 (2014).
757 33. U. E. Ukaegbu *et al.*, A Unique Virulence Gene Occupies a Principal Position in Immune Evasion
758 by the Malaria Parasite Plasmodium falciparum. *PLoS. Genet* **11**, e1005234 (2015).
759 34. M. M. Zilversmit *et al.*, Hypervariable antigen genes in malaria have ancient roots. *BMC. Evol.*
760 *Biol* **13**, 110 (2013).
761 35. M. R. Gross, R. Hsu, K. W. Deitsch, Evolution of transcriptional control of antigenic variation and
762 virulence in human and ape malaria parasites. *BMC Ecol Evol* **21**, 139 (2021).
763 36. B. Amulic, A. Salanti, T. Lavstsen, M. A. Nielsen, K. W. Deitsch, An upstream open reading frame
764 controls translation of var2csa, a gene implicated in placental malaria. *PLoS. Pathog* **5**, e1000256
765 (2009).
766 37. C. Bancells, K. W. Deitsch, A molecular switch in the efficiency of translation reinitiation controls
767 expression of var2csa, a gene implicated in pregnancy-associated malaria. *Mol. Microbiol*
768 (2013).
769 38. S. Chan *et al.*, Regulation of PfEMP1-VAR2CSA translation by a Plasmodium translation-
770 enhancing factor. *Nat Microbiol* **2**, 17068 (2017).
771 39. X. Zhang *et al.*, Rapid antigen diversification through mitotic recombination in the human
772 malaria parasite Plasmodium falciparum. *PLoS. Biol* **17**, e3000271 (2019).
773 40. A. Cortes, K. W. Deitsch, Malaria Epigenetics. *Cold Spring Harb. Perspect. Med*
774 cshperspect.a025528 [pii];10.1101/cshperspect.a025528 [doi] (2017).

775 41. A. Poran *et al.*, Single-cell RNA sequencing reveals a signature of sexual commitment in malaria
776 parasites. *Nature* **551**, 95-99 (2017).
777 42. J. E. Lemieux *et al.*, Statistical estimation of cell-cycle progression and lineage commitment in
778 Plasmodium falciparum reveals a homogeneous pattern of transcription in ex vivo culture. *Proc*
779 *Natl Acad Sci U S A* **106**, 7559-7564 (2009).
780 43. A. Chan, A. Dziedziech, L. A. Kirkman, K. W. Deitsch, J. Ankarklev, A Histone Methyltransferase
781 Inhibitor Can Reverse Epigenetically Acquired Drug Resistance in the Malaria Parasite
782 Plasmodium falciparum. *Antimicrob Agents Chemother* **64** (2020).
783 44. S. Mok *et al.*, Artemisinin resistance in Plasmodium falciparum is associated with an altered
784 temporal pattern of transcription. *BMC. Genomics* **12**, 391 (2011).
785 45. S. Mok *et al.*, Drug resistance. Population transcriptomics of human malaria parasites reveals the
786 mechanism of artemisinin resistance. *Science* **347**, 431-435 (2015).
787 46. S. Serizawa, K. Miyamichi, H. Sakano, One neuron-one receptor rule in the mouse olfactory
788 system. *Trends Genet* **20**, 648-653 (2004).
789 47. S. Hutchinson *et al.*, The establishment of variant surface glycoprotein monoallelic expression
790 revealed by single-cell RNA-seq of Trypanosoma brucei in the tsetse fly salivary glands. *PLoS*
791 *Pathog* **17**, e1009904 (2021).
792 48. N. K. Hanchate *et al.*, Single-cell transcriptomics reveals receptor transformations during
793 olfactory neurogenesis. *Science* **350**, 1251-1255 (2015).
794 49. L. R. Saraiva *et al.*, Hierarchical deconstruction of mouse olfactory sensory neurons: from whole
795 mucosa to single-cell RNA-seq. *Sci Rep* **5**, 18178 (2015).
796 50. K. Monahan, A. Horta, S. Lomvardas, LHX2- and LDB1-mediated trans interactions regulate
797 olfactory receptor choice. *Nature* **565**, 448-453 (2019).
798 51. L. Joergensen *et al.*, Surface co-expression of two different PfEMP1 antigens on single
799 plasmodium falciparum-infected erythrocytes facilitates binding to ICAM1 and PECAM1. *PLoS.*
800 *Pathog* **6**, e1001083 (2010).
801 52. M. Constancia, B. Pickard, G. Kelsey, W. Reik, Imprinting mechanisms. *Genome Res* **8**, 881-900
802 (1998).
803 53. M. F. Duffy, T. J. Byrne, C. Carret, A. Ivens, G. V. Brown, Ectopic recombination of a malaria var
804 gene during mitosis associated with an altered var switch rate. *J. Mol. Biol* **389**, 453-469 (2009).
805 54. M. F. Duffy *et al.*, Activation and clustering of a Plasmodium falciparum var gene are affected by
806 subtelomeric sequences. *FEBS J* **284**, 237-257 (2017).
807 55. T. Otto *et al.*, Evolutionary analysis of the most polymorphic gene family in falciparum malaria.
808 *Wellcome Open Res* **4**, 193 (2019).
809 56. M. F. Duffy *et al.*, Transcribed var genes associated with placental malaria in Malawian women.
810 *Infect. Immun* **74**, 4875-4883 (2006).
811 57. J. G. Beeson *et al.*, Antibodies among men and children to placental-binding Plasmodium
812 falciparum-infected erythrocytes that express var2csa. *Am J Trop Med Hyg* **77**, 22-28 (2007).
813 58. E. Rovira-Vallbona *et al.*, Transcription of var genes other than var2csa in Plasmodium
814 falciparum parasites infecting Mozambican pregnant women. *J Infect Dis* **204**, 27-35 (2011).
815 59. S. F. Calhoun *et al.*, Chromosome End Repair and Genome Stability in Plasmodium falciparum.
816 *MBio* **8** (2017).
817 60. J. Reed, L. A. Kirkman, B. F. Kafsack, C. E. Mason, K. W. Deitsch, Telomere length dynamics in
818 response to DNA damage in malaria parasites. *iScience* **24**, 102082 (2021).
819 61. P. Horrocks, R. Pinches, Z. Christodoulou, S. Kyes, C. Newbold, Variable var transition rates
820 underlie antigenic variation in malaria. *Proc. Natl. Acad. Sci. U. S. A* (2004).

821 62. A. Bachmann *et al.*, Absence of erythrocyte sequestration and lack of multicopy gene family
822 expression in Plasmodium falciparum from a splenectomized malaria patient. *PLoS. ONE* **4**,
823 e7459 (2009).
824 63. S. Kho *et al.*, Hidden Biomass of Intact Malaria Parasites in the Human Spleen. *N Engl J Med* **384**,
825 2067-2069 (2021).
826 64. C. M. Andrade *et al.*, Increased circulation time of Plasmodium falciparum underlies persistent
827 asymptomatic infection in the dry season. *Nat Med* **26**, 1929-1940 (2020).
828 65. L. A. Kirkman, X. Z. Su, T. E. Wellems, Plasmodium falciparum: isolation of large numbers of
829 parasite clones from infected blood samples. *Exp Parasitol* **83**, 147-149 (1996).
830 66. X. Zhang *et al.*, Rapid antigen diversification through mitotic recombination in the human
831 malaria parasite Plasmodium falciparum. *PLoS Biol* **17**, e3000271 (2019).
832 67. Y. Wu, L. A. Kirkman, T. E. Wellems, Transformation of Plasmodium falciparum malaria parasites
833 by homologous integration of plasmids that confer resistance to pyrimethamine. *Proc Natl Acad*
834 *Sci U S A* **93**, 1130-1134 (1996).
835 68. K. Deitsch, C. Driskill, T. Wellems, Transformation of malaria parasites by the spontaneous
836 uptake and expression of DNA from human erythrocytes. *Nucleic Acids Res* **29**, 850-853 (2001).
837 69. M. Ghorbal *et al.*, Genome editing in the human malaria parasite Plasmodium falciparum using
838 the CRISPR-Cas9 system. *Nat Biotechnol* **32**, 819-821 (2014).
839 70. R. Dzikowski, M. Frank, K. Deitsch, Mutually exclusive expression of virulence genes by malaria
840 parasites is regulated independently of antigen production. *PLoS Pathog* **2**, e22 (2006).
841 71. A. Salanti *et al.*, Selective upregulation of a single distinctly structured var gene in chondroitin
842 sulphate A-adhering Plasmodium falciparum involved in pregnancy-associated malaria. *Mol*
843 *Microbiol* **49**, 179-191 (2003).
844 72. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome alignment and
845 genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907-915 (2019).
846 73. E. Afgan *et al.*, The Galaxy platform for accessible, reproducible and collaborative biomedical
847 analyses: 2018 update. *Nucleic Acids Res* **46**, W537-W544 (2018).
848 74. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning
849 sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
850 75. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-
851 seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
852 76. Y. Chen, A. T. Lun, G. K. Smyth, From reads to genes to pathways: differential expression analysis
853 of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Res* **5**,
854 1438 (2016).
855 77. D. J. McCarthy, Y. Chen, G. K. Smyth, Differential expression analysis of multifactor RNA-Seq
856 experiments with respect to biological variation. *Nucleic Acids Res* **40**, 4288-4297 (2012).
857 78. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential
858 expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
859 79. B. T. Grimberg, Methodology and application of flow cytometry for investigation of human
860 malaria parasites. *J. Immunol. Methods* **367**, 1-16 (2011).
861 80. R. Bartfai *et al.*, H2A.Z demarcates intergenic regions of the plasmodium falciparum epigenome
862 that are dynamically marked by H3K9ac and H3K4me3. *PLoS. Pathog* **6**, e1001223 (2010).
863 81. Z. Bozdech *et al.*, The transcriptome of the intraerythrocytic developmental cycle of Plasmodium
864 falciparum. *PLoS. Biol* **1**, E5 (2003).

865

866 FIGURE LEGENDS

867 Figure 1: Detection of “single” and “many” *var* expression profiles in cultured parasite
868 populations. A) Schematic representation of the single-many-single (SMS) model for *var*
869 gene transcriptional switching. Each circle represents expression of an individual *var*
870 gene and arrows represent switches in expression. Switching events are hypothesized
871 to transition from activation of a single gene (top) to a broad range of genes (middle) to
872 a different single gene (bottom) or back to the original gene (reverse arrow). B) *var* gene
873 expression profiles for two clonal populations that display the single phenotype (clone 1,
874 top) or the many phenotype (clone 2, bottom). *var* transcription levels were determined
875 using a standardized qRT-PCR protocol with the expression of each individual *var* gene
876 displayed as relative copy number in the histogram. Error bars represent standard
877 deviation of three biological replicates. C) Total *var* expression for the two subclones
878 shown in B, with transcripts from each individual *var* gene shown in a different colour.
879 For clone 1, transcripts from the dominant gene are marked, while both clones display
880 similar levels of minor transcripts. D) Clone tree of wildtype 3D7 parasites. Pie charts
881 display the *var* expression profile for each subcloned population with each slice of the
882 pie representing the expression level of a single *var* gene. Vertical and horizontal lines
883 delineate sequential rounds of subcloning by limiting dilution. For parasite populations
884 that display a dominantly expressed *var* gene, the annotation number is shown below
885 the pie chart with the *var* type shown in parenthesis. The five subclones marked with an
886 asterisk are further analyzed in Figure 2. D) Total *var* expression levels as determined
887 by qRT-PCR for 74 subclones. The median +/- 95% confidence interval is shown, and
888 an unpaired t-test indicates a p:<0.0001 (***).

889

890 Figure 2: Detection of *var* gene minor transcript expression profiles in clonal populations
891 overtime. A) Heatmap of *var* minor transcripts for the subclones 3.1, 3.2, 3.3, 3.4 and
892 3.5. The clones were all derived from the same parent population as shown in Figure
893 1D. Pie charts show the initial *var* gene expression profiles above the heatmap. The
894 annotation numbers for each *var* gene are shown to the left of the heatmap and
895 organized according to *var* type (left, *var2csa* is considered separately in Figure 3). 6
896 timepoints are included for each clone. Two genetically identical parasite populations
897 (clones 4 and 5) also originally derived from 3D7 but grown separately for several years
898 are shown for comparison. For subclones 3.4 and 3.5, the dominant transcript
899 (PF3D7_1240600 and PF3D7_0421100, respectively, marked by black boxes in the
900 heatmap) was removed from the analysis to enable visualization of the minor
901 transcripts. B) Clone tree of wildtype 3D7. The tree is organized into 3 “subclone
902 generations” derived from initial Subclone 3.1 (top row). Pie charts display the *var*
903 expression profile for each subclone. C) Heatmap of *var* minor transcripts for the
904 individual clones from each generation shown in B with the pattern of the parent
905 population (subclone 3.1) shown at the left for comparison. The annotation numbers for
906 each *var* gene are shown to the left of the heatmap, and the gene order was organized
907 according to *var* type. The order from left to right of each column in the heatmap
908 corresponds to the order from left to right of the pie charts for each subclone generation
909 shown in B. For parasites expressing the “single” phenotype, the dominant transcript
910 (marked by black boxes in the heatmap) was removed to enable visualization of the
911 minor transcripts.

912

913 Figure 3: Convergence to *var2csa* expression in long-term cultures and targeted
914 deletion of the *var2csa* locus. A) Changes in *var* expression over time in five clonal
915 parasite populations, three “manys” (subclones 3.1, 3.2 and 3.3) and two “singles”
916 (subclones 3.4 and 3.5), over several months of continuous culture. Expression of
917 *var2csa* is shown in red. B) Schematic diagram showing the truncation of the end of
918 chromosome 12 by telomere healing. Telomere repeats are shown as slanted lines and
919 the telomere associated repeat elements (TARES) are shown as an open box. The
920 interior of the chromosome is displayed as a dashed line. A ~60 kb deletion, including
921 three *var* genes (blue), three *rif* genes (green) and the *acs7* gene is shown. C)
922 Schematic diagram showing the *var2csa* locus and the plasmid containing homology
923 blocks for targeted integration (hb1 and hb2). The crossed lines linking the plasmid to
924 the chromosome signify sites of double cross over recombination leading to deletion of
925 approximately 2.5 kb upstream of the *va2csa* gene, including the promoter.

926

927 Figure 4: Alterations in *var* gene expression after deletion or disruption of *var2csa*. A)
928 Changes in *var* expression profile over time in 5 clonal parasite populations containing
929 deletions or disruption of the *var2csa* locus, including two in which the locus was
930 disrupted by plasmid integration (V2dis1 and V2dis2, top) and three in which the locus
931 was deleted by chromosomal truncation (ΔV2_1, ΔV2_2, ΔV2_3). V2dis1 and ΔV2_3
932 display a “single” phenotype while V2dis2, ΔV2_1, ΔV2_2 express many *var* genes. B)
933 Heatmap displaying the pattern of *var* minor transcripts over several timepoints for 5

934 *var2csa*-deleted lines. The clones were all derived from the same parent population,
935 and pie charts showing the initial *var* gene expression profiles are shown above the
936 heatmap. The annotation numbers for each *var* gene are shown to the left of the
937 heatmap, organized according to *var* type. For ΔV2_3, the dominant transcript
938 (Pf3D7_0712600, marked by black boxes in the heatmap) was removed to enable
939 visualization of the minor transcripts. C) NMDS plot based on Bray-Curtis dissimilarities
940 displaying the expression profiles of the *var* minor transcripts for the 6 *var2csa*-deleted
941 lines alongside 5 wildtype lines. D) The variability within each cluster shown in C is
942 inferred by its area on the plot (as defined by the mean distance from the group centroid
943 for each time point) and displayed in a boxplot. E) Transcriptomic analysis of wildtype
944 and *var2csa*-mutated parasite lines. Principal Component Analysis (PCA) is shown
945 based on the normalized expression level of 5721 genes log10(FPKM+1). RNAseq
946 derived whole transcriptomes were obtained from eleven clonal parasite populations:
947 two wildtype lines (WT 3.1 and WT 7) and two WT lines with a control plasmid
948 integration into a *rif* gene (WT_rif1 and WT_rif2) are displayed in orange; four *var2csa*-
949 mutated lines displaying a “single” phenotype (V2dis1, V2dis2, V2dis3 and ΔV2_3) are
950 displayed in blue; and three *var2csa*-mutated lines displaying a “many” phenotype
951 (ΔV2_1, ΔV2_5 and ΔV2_7) are displayed in green. All analyses were performed using
952 data obtained from synchronized populations ~16 hours after red cell invasion. F)
953 Comparisons of normalized *var* transcript counts from *var2csa*-mutated and wildtype
954 lines displaying either the single (left) or many (right) phenotypes.

955

956 Figure 5: Parasites displaying “single” vs “many” *var* expression profiles differ in
957 replication cycle progression. A) Volcano plots showing differentially expressed genes
958 derived from whole transcriptome comparisons of *var2csa*-mutated “single” and “many”
959 parasite lines. Differentially expressed genes are shown in red. For ring-stage parasites
960 (left) 562 transcripts displayed higher expression levels in the single lines, while 49 were
961 higher in the many lines. For trophozoite-stage parasites, only 51 genes were
962 differentially expressed. B) Estimation of cell cycle position using the method of Poran
963 et al (42). Using the transcriptome profiles in Supplementary Table S6 (80), the
964 likelihood (vertical axis) of each sample having been derived from a particular timepoint
965 of the cycle (horizontal axis) is displayed. 11 parasite lines are shown, with wildtype
966 parasites in orange, *var2csa*-mutated, “single” parasites in blue and *var2csa*-mutated,
967 “many” parasites in green type. C) Table showing estimates of replication cycle
968 progression for all eleven parasite lines. Estimates of the approximate time point within
969 the 48-hour asexual cycle for each population were obtained by comparison with a
970 modeled parasite population over a simulated infection time course using established
971 datasets (41). The best time point match is shown for each population along with the
972 correlation coefficient. Wildtype parasites are designated with orange type, *var2csa*-
973 mutated, “single” parasites with blue type and *var2csa*-mutated, “many” parasites with
974 green type. D) Heatmap displaying changes in expression over the asexual replicative
975 cycle for 611 differentially expressed genes according to the 48h asexual cycle of the *P.*
976 *falciparum* HB3 transcriptome datasets defined by Bozdech et al (81)(Supplementary
977 Table S7). Genes upregulated in *var2csa*-mutated “single” lines are shown on top and
978 those from *var2csa*-mutated “many” lines are displayed on the bottom. E) Average

979 expression levels across the 48h asexual cycle of HB3 for 611 differentially expressed
980 genes. Genes upregulated in *var2csa*-mutated, “single” lines are shown on top and
981 those upregulated in *var2csa*-mutated, “many” lines are shown on the bottom. F) Left
982 panel, transcript expression levels for 559 differentially expressed genes that display
983 higher expression in “single” parasites. Average, normalized transcript levels were
984 obtained from *var2csa*-mutated, “single” parasites (blue), wildtype “single” and “many”
985 parasites (orange, left and right, respectively), and *var2csa*-mutated, “many” parasites
986 (green). Box and whisker plots display the mean and standard deviation for each
987 dataset. Right panel, analysis of transcript expression levels for 49 differentially
988 expressed genes that display higher expression in “many” parasites. Pairwise
989 comparisons using t tests with pooled standard deviation. P value adjustment method is
990 from Bonferroni. (Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’). G) Monitoring of DNA
991 content of infected cells by flow cytometry. Populations of *var2csa*-mutant “singles”
992 (top), wildtype (middle) and *var2csa*-mutant “manys” (bottom) were tightly synchronized
993 then assayed for DNA content by flow cytometry using Hoechst 33342 staining.
994 Parasites were assayed at 16 and 36 hours after invasion.

995

Figure 1

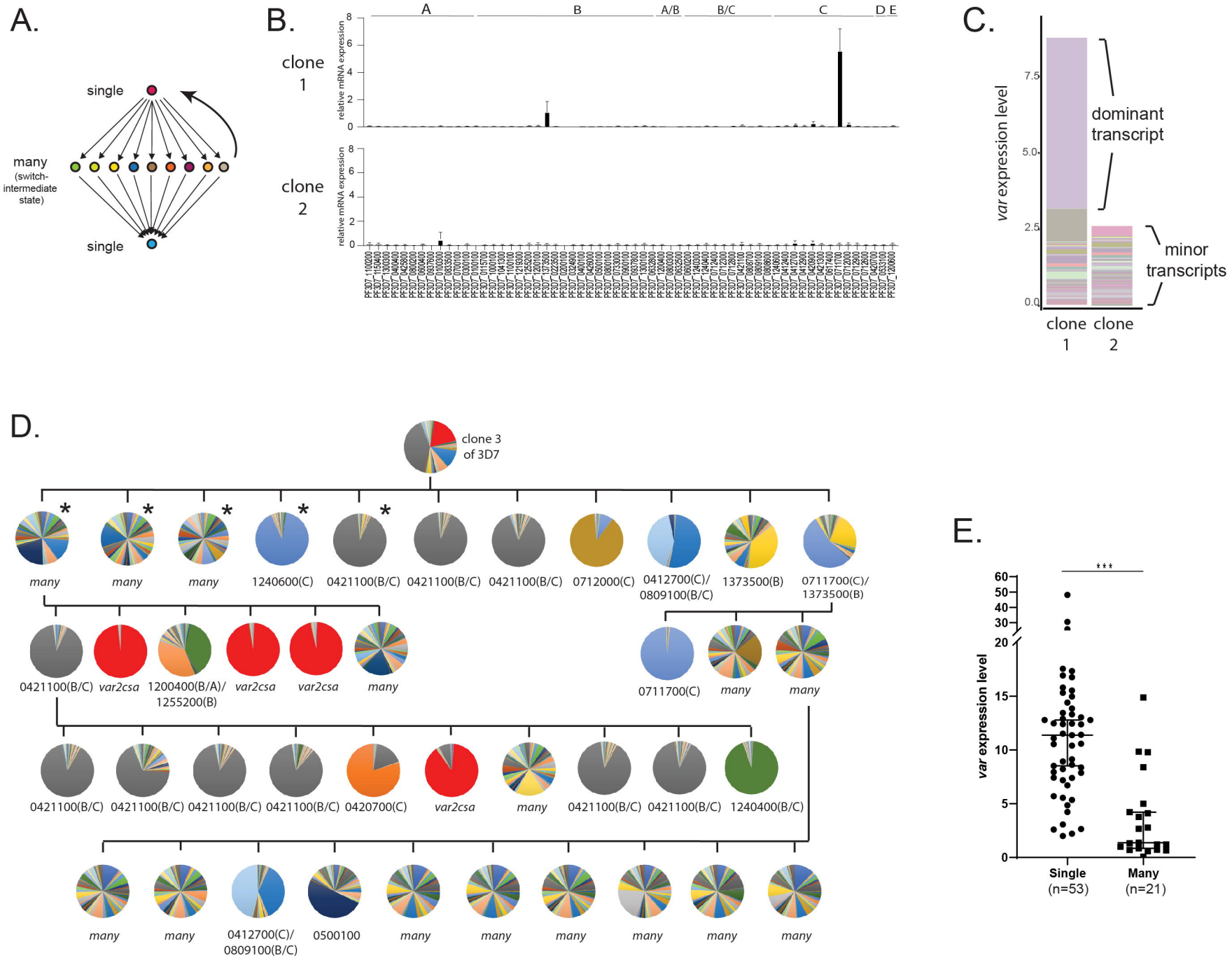

A.
single
many
(switch-intermediate state)
single
B.
A
B
A/B
B/C
C
D E
clone 1
clone 2
relative mRNA expression
C.
var expression level
dominant transcript
minor transcripts
clone 1
clone 2
D.
clone 3 of 3D7
many
many
many
1240600(C)
0421100(B/C)
0421100(B/C)
0421100(B/C)
0712000(C)
0412700(C)/ 0809100(B/C)
1373500(B)
0711700(C)/ 1373500(B)
0421100(B/C)
var2csa
1200400(B/A)/ 1255200(B)
var2csa
var2csa
many
0711700(C)
many
many
0421100(B/C)
0421100(B/C)
0421100(B/C)
0421100(B/C)
0420700(C)
var2csa
many
0421100(B/C)
0421100(B/C)
1240400(B/C)
many
many
0412700(C)/ 0809100(B/C)
0500100
many
many
many
many
many
many
E.
var expression level
Single (n=53)
Many (n=21)

Figure 2

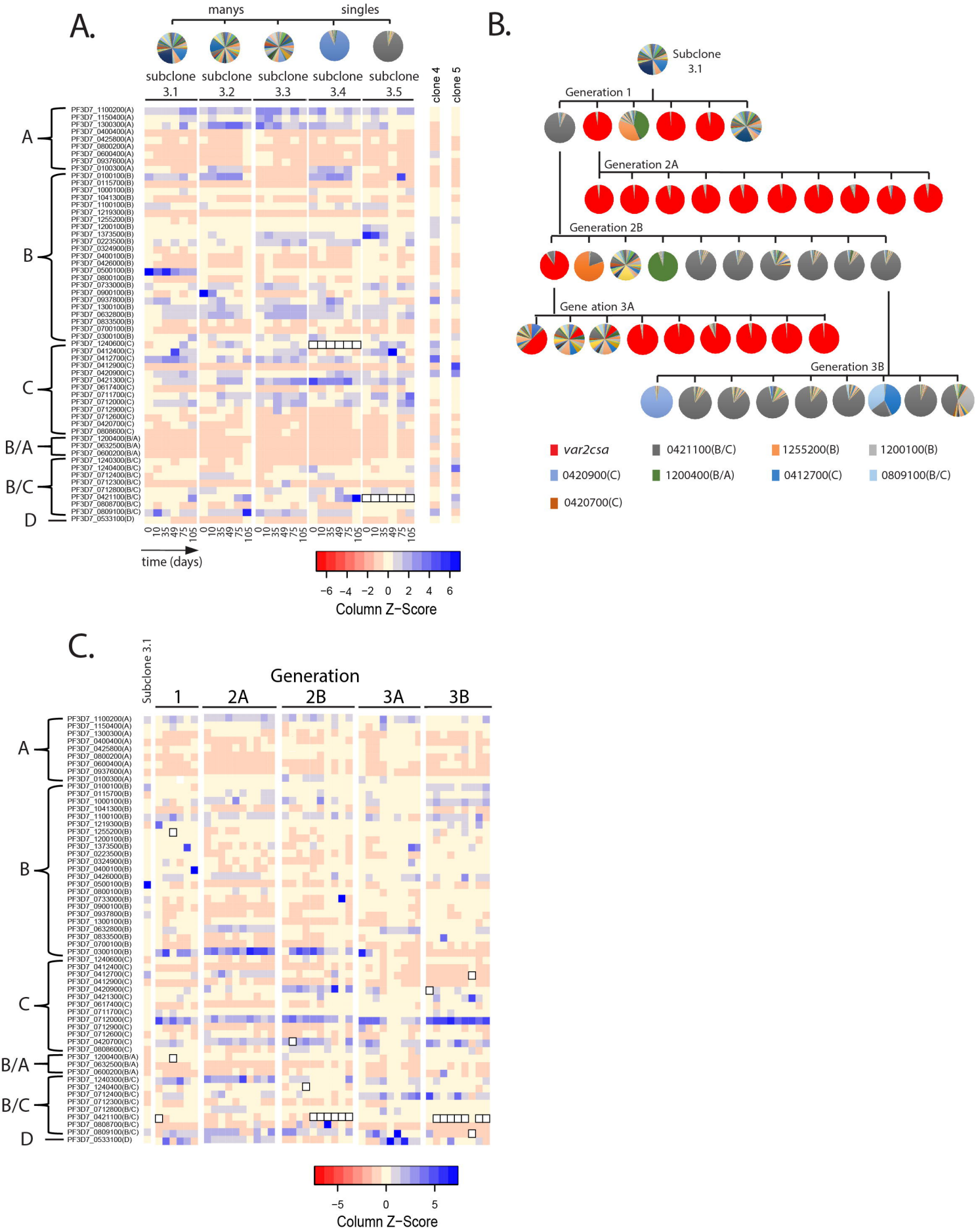
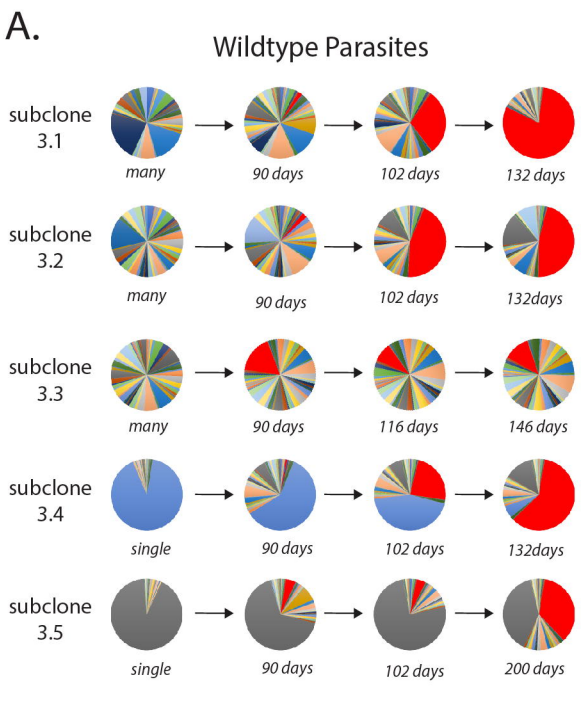
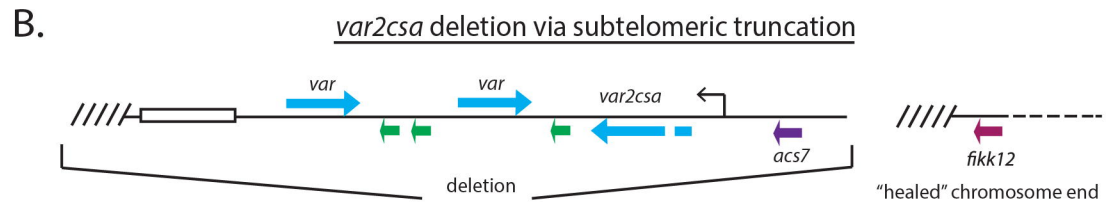

A.
manys
singles
subclone 3.1
subclone 3.2
subclone 3.3
subclone 3.4
subclone 3.5
clone 4
clone 5
A
B
C
B/A
B/C
D
time (days)
Column Z-Score
B.
Subclone 3.1
Generation 1
Generation 2A
Generation 2B
Gene ation 3A
Generation 3B
var2csa
0421100(B/C)
1255200(B)
1200100(B)
0420900(C)
1200400(B/A)
0412700(C)
0809100(B/C)
0420700(C)
C.
Subclone 3.1
Generation
1
2A
2B
3A
3B
Column Z-Score

Figure 3


A.
Wildtype Parasites
subclone 3.1
many
90 days
102 days
132 days
subclone 3.2
many
90 days
102 days
132days
subclone 3.3
many
90 days
116 days
146 days
subclone 3.4
single
90 days
102 days
132days
subclone 3.5
single
90 days
102 days
200 days
var2csa


B.
var2csa deletion via subtelomeric truncation
var
var
var2csa
acs7
fikk12
deletion
"healed" chromosome end

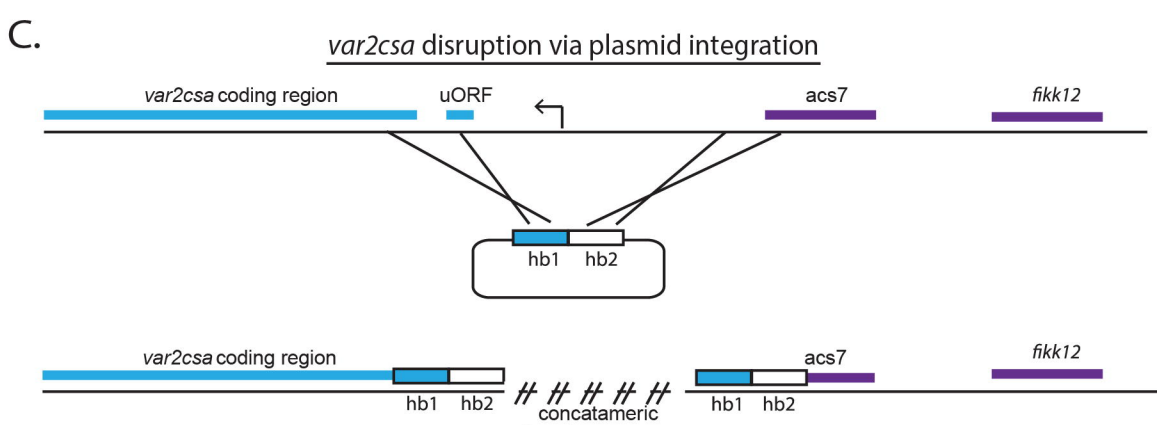

C.
var2csa disruption via plasmid integration
var2csa coding region
uORF
acs7
fikk12
hb1
hb2
var2csa coding region
hb1
hb2
concatameric
plasmid sequences
acs7
hb1
hb2
fikk12

Figure 4

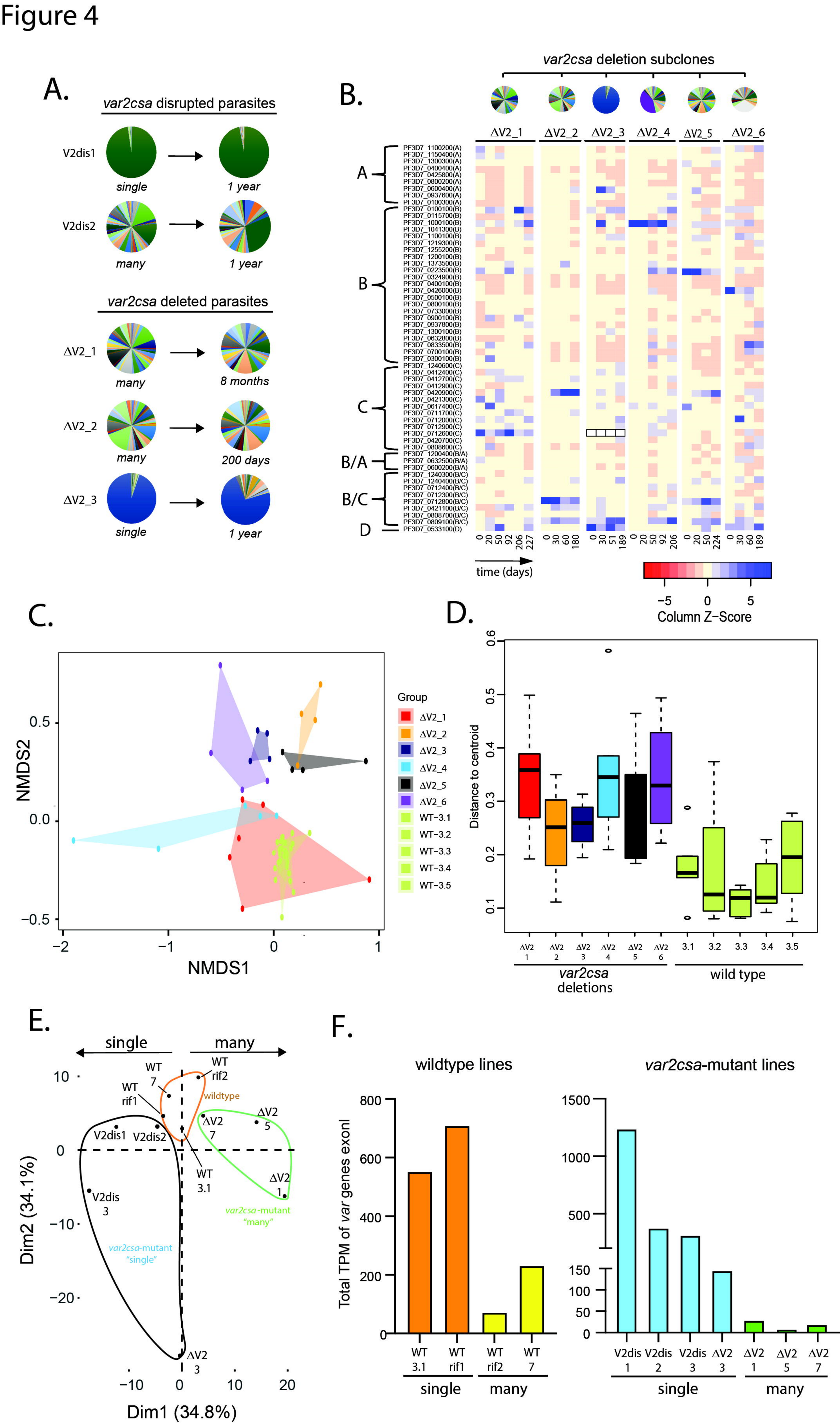
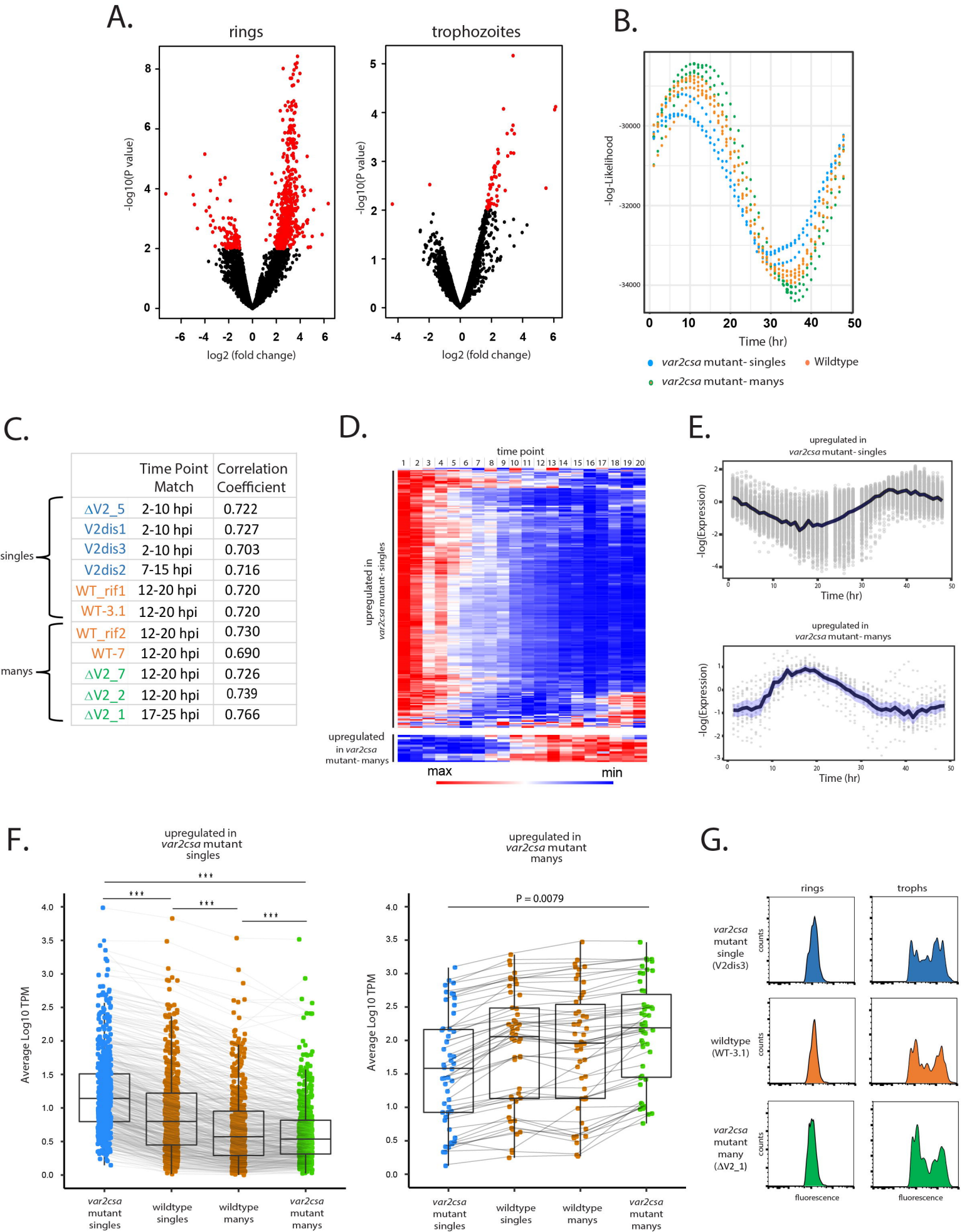

A.
var2csa disrupted parasites
V2dis1
single
1 year
V2dis2
many
1 year
var2csa deleted parasites
ΔV2_1
many
8 months
ΔV2_2
many
200 days
ΔV2_3
single
1 year
B.
var2csa deletion subclones
ΔV2_1
ΔV2_2
ΔV2_3
ΔV2_4
ΔV2_5
ΔV2_6
A
B
C
B/A
B/C
D
time (days)
−5
0
5
Column Z-Score
C.
NMDS2
NMDS1
Group
ΔV2_1
ΔV2_2
ΔV2_3
ΔV2_4
ΔV2_5
ΔV2_6
WT−3.1
WT−3.2
WT−3.3
WT−3.4
WT−3.5
D.
Distance to centroid
var2csa deletions
wild type
E.
single
many
wildtype
var2csa-mutant "many"
var2csa-mutant "single"
Dim1 (34.8%)
Dim2 (34.1%)
F.
wildtype lines
var2csa-mutant lines
Total TPM of var genes exonl
single
many

Figure 5



	Time Point Match	Correlation Coefficient
ΔV2_5	2-10 hpi	0.722
V2dis1	2-10 hpi	0.727
V2dis3	2-10 hpi	0.703
V2dis2	7-15 hpi	0.716
WT_rif1	12-20 hpi	0.720
WT-3.1	12-20 hpi	0.720
WT_rif2	12-20 hpi	0.730
WT-7	12-20 hpi	0.690
ΔV2_7	12-20 hpi	0.726
ΔV2_2	12-20 hpi	0.739
ΔV2_1	17-25 hpi	0.766