Supplementary Tables

Supplementary Table 1. Summary of genome sequence analysis of clones containing a single copy of the *GAP1* CNV reporter. Estimated copy number of the *GAP1* gene and inserted GFP gene of sequenced clones from five 1-copy-GFP minor subpopulations of the WT genome architecture strain. Copy number estimation is defined as the read depth of the target gene relative to the average read depth of the chromosome XI. Populations 1, 2, 4, 5 contain clones harboring GAP1 *CNVs* but only 1 copy of GFP. Clones from population 3 and 5 harbor 1 copy each of *GAP1* and GFP suggesting these lineages have beneficial mutations elsewhere in the genome, allowing coexistence with the *GAP1* CNV major subpopulation. CN, copy number, CNV = copy number variant, *GAP1* = general amino acid permease gene

Sample	Generation	Chemostat	Population	Background Strain	*GAP1* CN	GFP CN	Left CNV Boundary Feature	Right CNV Boundary Feature	CNV Mechanism
3150	182	H03	1	WT	5	1	Between *GFP* and *GAP1*	DYN1	ODIRA
3171	182	H03	1	WT	3	1	Between *GFP* and *GAP1*	TIF1	ODIRA
3172	182	H03	1	WT	3	1	Between *GFP* and *GAP1*	DYN1	ODIRA
3173	182	H03	1	WT	3	1	Between *GFP* and *GAP1*	DYN1	ODIRA
3174	182	H03	1	WT	3	1	kanamycin CDS	NUP133	ODIRA
3151	153	G04	2	WT	3	1	Between *GFP* and *GAP1*	RPF2	ODIRA
3152	153	G04	2	WT	3	1	Between *GFP* and *GAP1*	GLG1	ODIRA
3153	153	G04	2	WT	3	1	kanamycin promoter	MRS4	ODIRA
3154	153	G04	2	WT	3	1	Between *GFP* and *GAP1*	RPF2	ODIRA
3155	153	G04	2	WT	2	1	Between *GFP* and *GAP1*	LTR YKRC□12 or tRNA	unresolved
3156	182	H05	3	WT	1	1	No ChrXI CNV	No ChrXI CNV	NA
3157	182	H05	3	WT	1	1	No ChrXI CNV	No ChrXI CNV	NA
3158	182	H05	3	WT	1	1	No ChrXI CNV	No ChrXI CNV	NA
3175	182	H05	3	WT	1	1	No ChrXI CNV	No ChrXI CNV	NA
3176	182	H05	3	WT	1	1	No ChrXI CNV	No ChrXI CNV	NA
3177	182	H05	3	WT	1	1	No ChrXI CNV	No ChrXI CNV	NA
3178	182	H05	3	WT	1	1	No ChrXI CNV	No ChrXI CNV	NA
3179	182	H05	3	WT	1	1	No ChrXI CNV	No ChrXI CNV	NA
3180	182	H05	3	WT	1	1	No ChrXI CNV	No ChrXI CNV	NA
3181	182	H05	3	WT	1	1	No ChrXI CNV	No ChrXI CNV	NA
3182	182	H05	3	WT	1	1	No ChrXI CNV	No ChrXI CNV	NA
3159	166	G06	4	WT	3	1	kanamycin promoter	Between YKR041W and UTH1	ODIRA
3160	166	G06	4	WT	3	1	kanamycin promoter	Between YKR041W and UTH1	ODIRA
3162	166	G06	4	WT	3	1	kanamycin promoter	Between YKR041W and UTH1	ODIRA
3163	166	G06	4	WT	3	1	kanamycin CDS	ARS1118	ODIRA
3164	182	H07	5	WT	1	1	No *GAP1* CNV	No *GAP1* CNV	NA
3165	182	H07	5	WT	3	1	kanamycin CDS	MRS4	ODIRA
3166	182	H07	5	WT	3	1	Between *GFP* and *GAP1*	UIP5	ODIRA
3167	182	H07	5	WT	3	1	kanamycin CDS	MRS4	ODIRA
3168	182	H07	5	WT	3	1	Between *GFP* and *GAP1*	UIP5	ODIRA

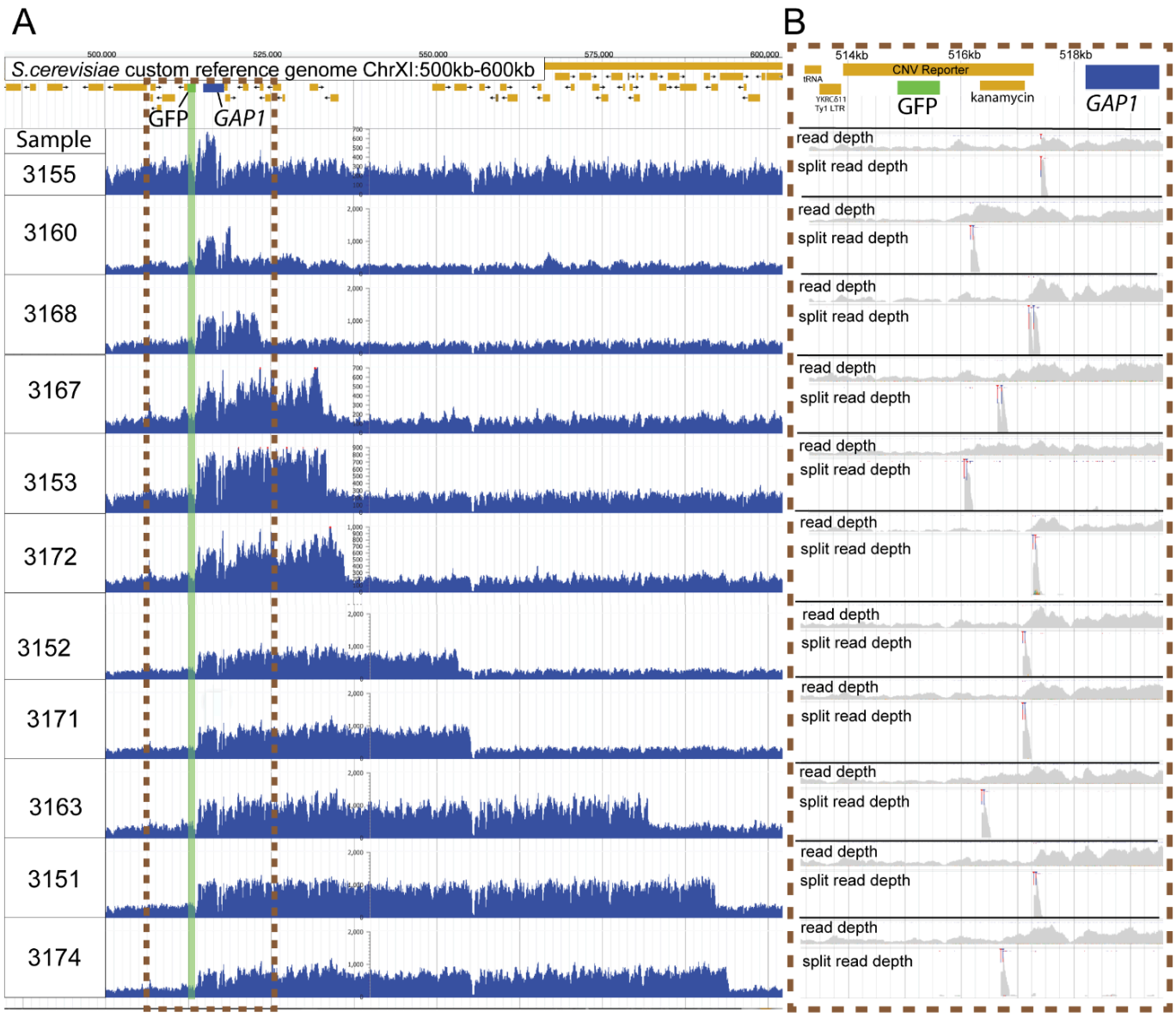
Supplementary Table 2. Estimation of network confidence. The coverage, defining the probability that the true parameter falls within the 95% highest density interval (HDI) of the posterior distribution, for 829 synthetic simulations in which the final reported *GAP1* CNV proportion is at least 0.3. 95% HDI was calculated for each simulation using 200 posterior samples. Our neural density estimator is slightly over-confident for φ (coverage of 0.934), and under-confident for *GAP1* CNV selection coefficient and formation rate (coverage of 0.992 for s_C and 0.995 for δ_C). Despite this under-confidence, the posterior distributions are narrow in biological terms: the 95% HDI represents less than an order of magnitude for both s_C and δ_C. Thus, we did not apply post-training adjustments to the neural density estimator, such as calibration (Cook et al., 2006) or ensembles (Caspi et al., 2023; Hermans et al., 2022).

Parameter	Coverage
s_C	0.992
δ_C	0.995
φ	0.934

Supplementary Table 3. Inferred CNV mechanisms by strain. Counts of inferred CNV mechanisms for each sequenced clone, n=177, separated by strain.

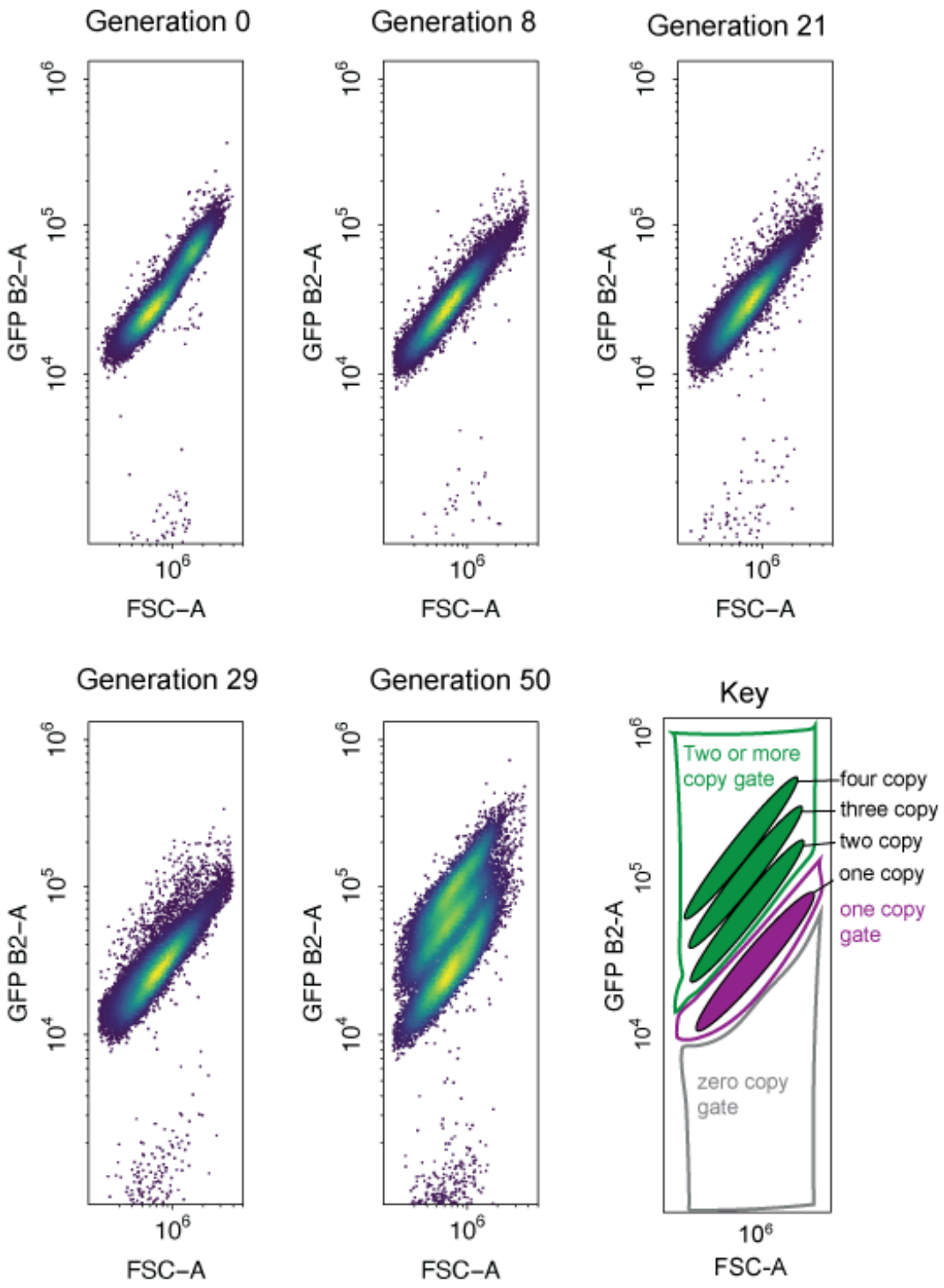
CNV Mechanism	WT	LTRΔ	ARSΔ	ALLΔ	Total
Aneuploid	1	5	0	0	6
Complex CNV	3	3	3	14	23
LTR NAHR	11	0	27	0	38
NAHR	0	2	1	1	4
ODIRA	22	42	11	12	87
Transposon-mediated	0	0	0	19	19
Total	37	52	42	46	177

Supplementary Figures

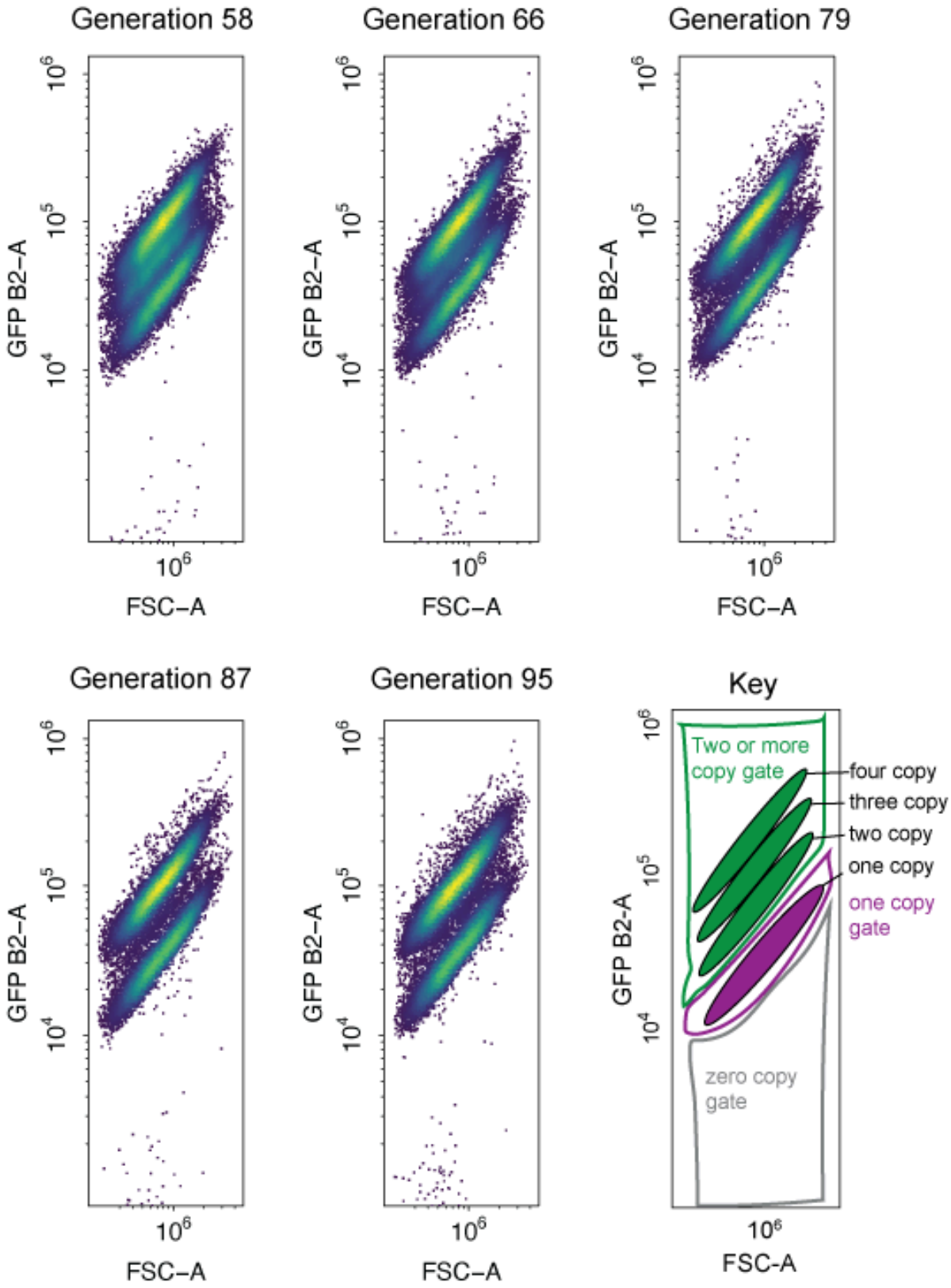


Supplementary Figure 1. **Independent *GAP1* amplifications lacking CNV reporter amplification**. **(A)** Read depth plots of the *GAP1* CNV reporter locus, ChrXI:500-600kb, of sequenced clones from 1-copy-*GFP* subpopulations isolated across five chemostats. Identification of eleven distinct CNVs, shown above, indicate the occurrence of at least eleven independent amplifications of *GAP1* without *GFP* co-amplification. Sequences were aligned to a custom reference genome containing the CNV reporter upstream of the *GAP1* gene. The CNV reporter comprises a GFP gene and kanamycin resistance gene. *GFP* reference gene - green rectangle, *GAP1* reference gene - blue rectangle. **(B)** Inset of the left-most CNV junction at the *GFP*, kanamycin, and *GAP1* region, ChrXI: 513193-519171 with genome read depth and split read depth tracks for each sample. The location of the split reads pileup (blue and red clipping marks) show the precise CNV breakpoint which is downstream of the *GFP* gene and upstream of the *GAP1* coding sequence for every clone indicating each lack an amplification of the inserted GFP gene.

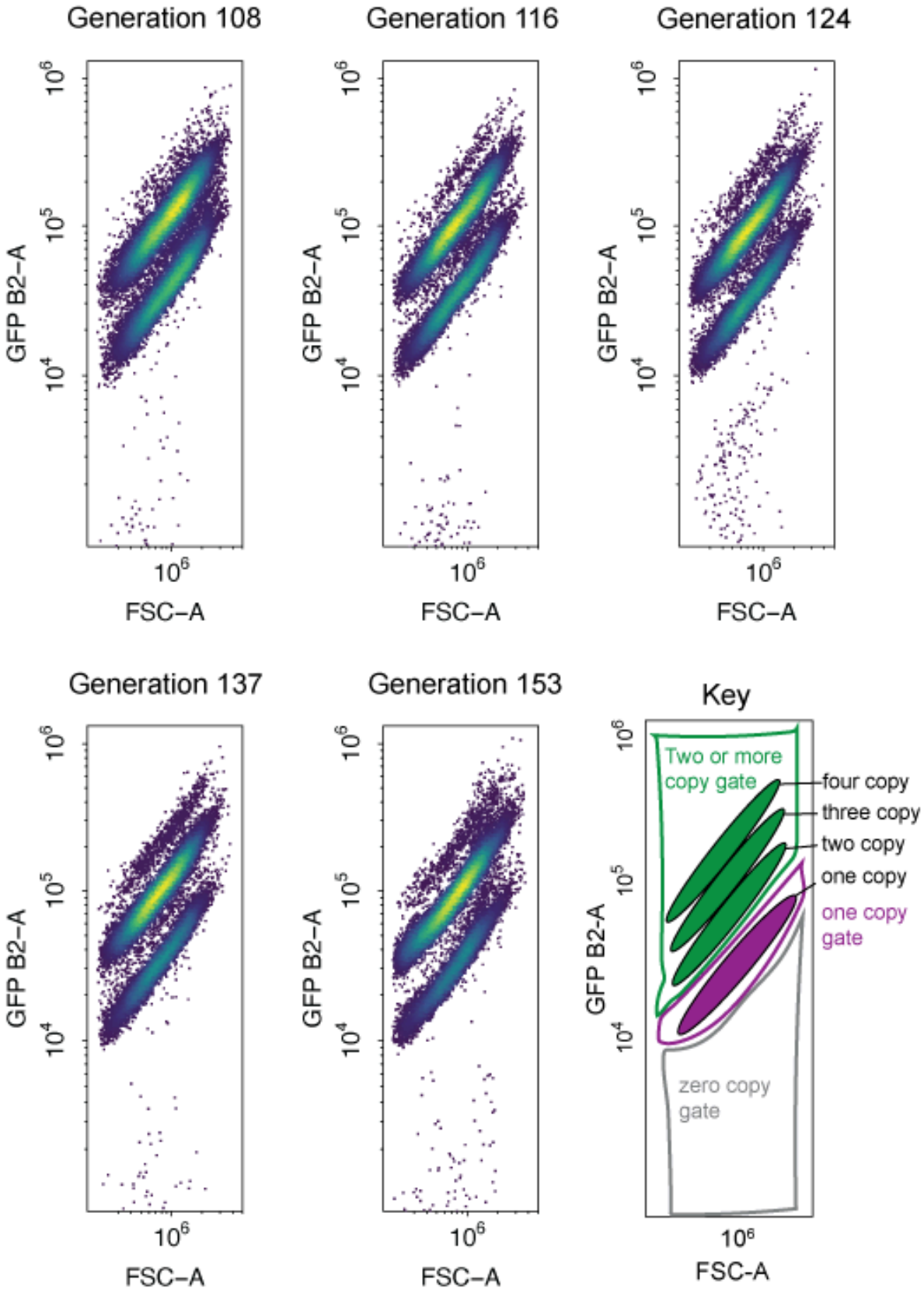
Wildtype population 1



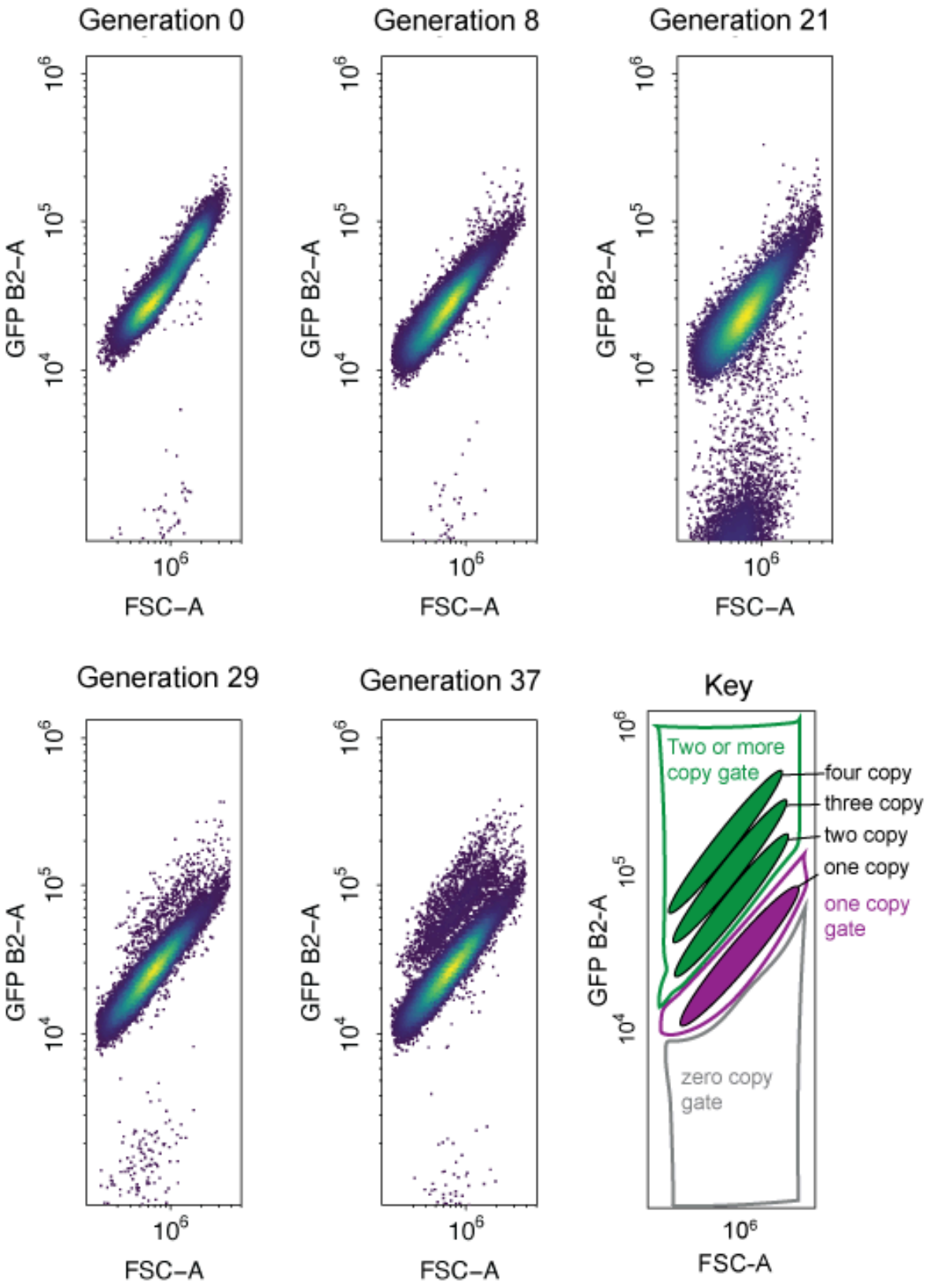
Wildtype population 1



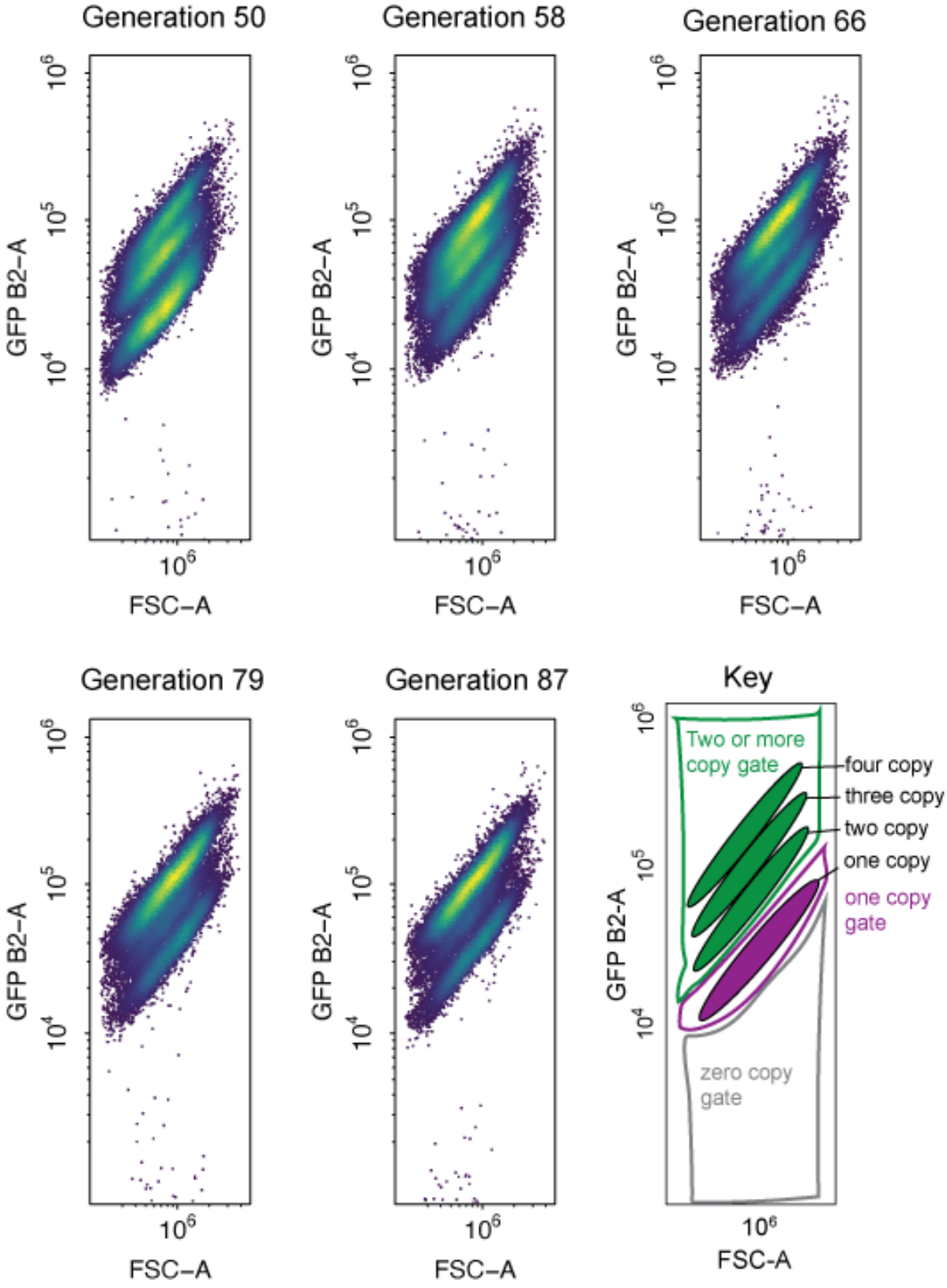
Wildtype population 1



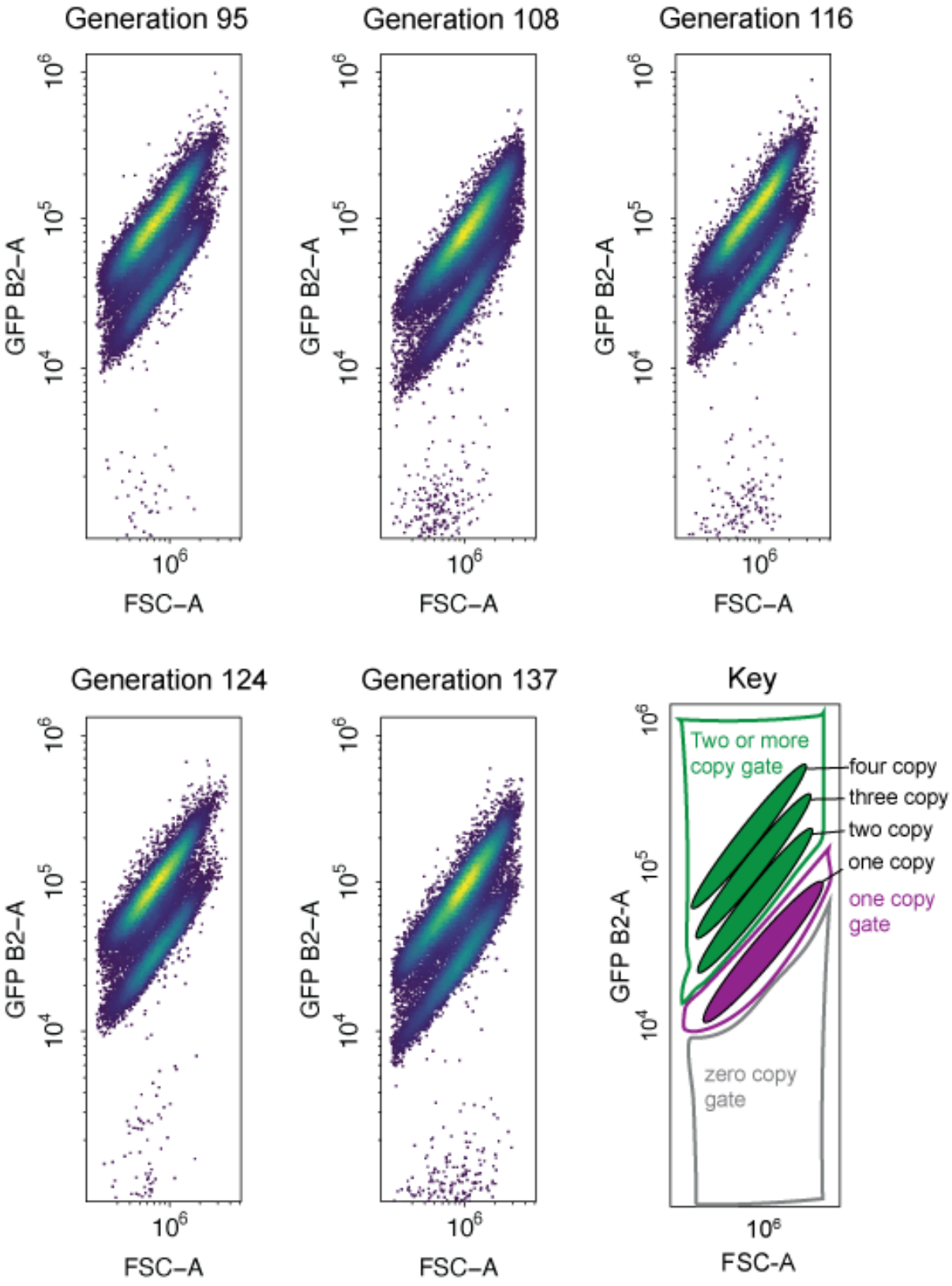
Wildtype population 2

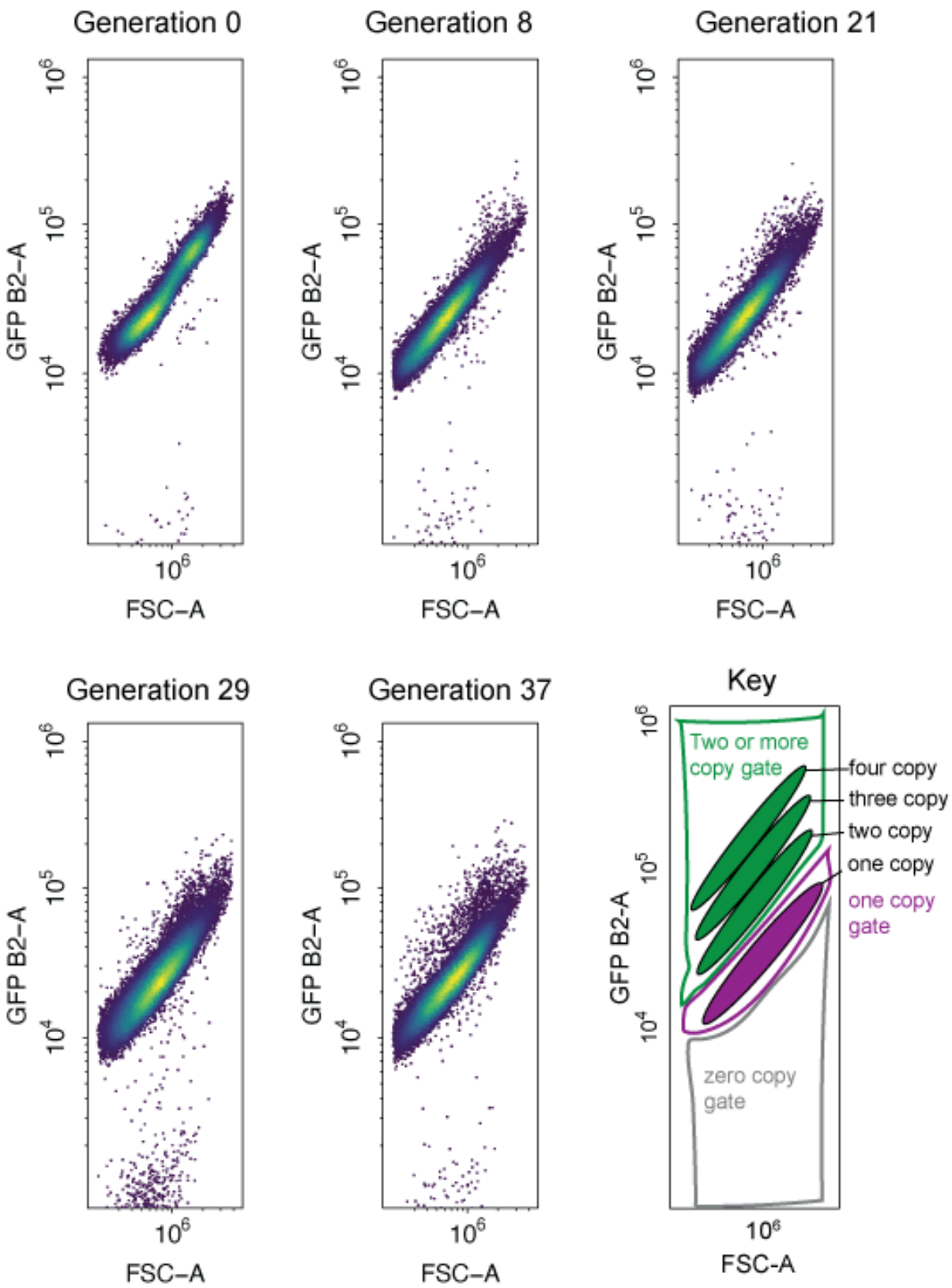


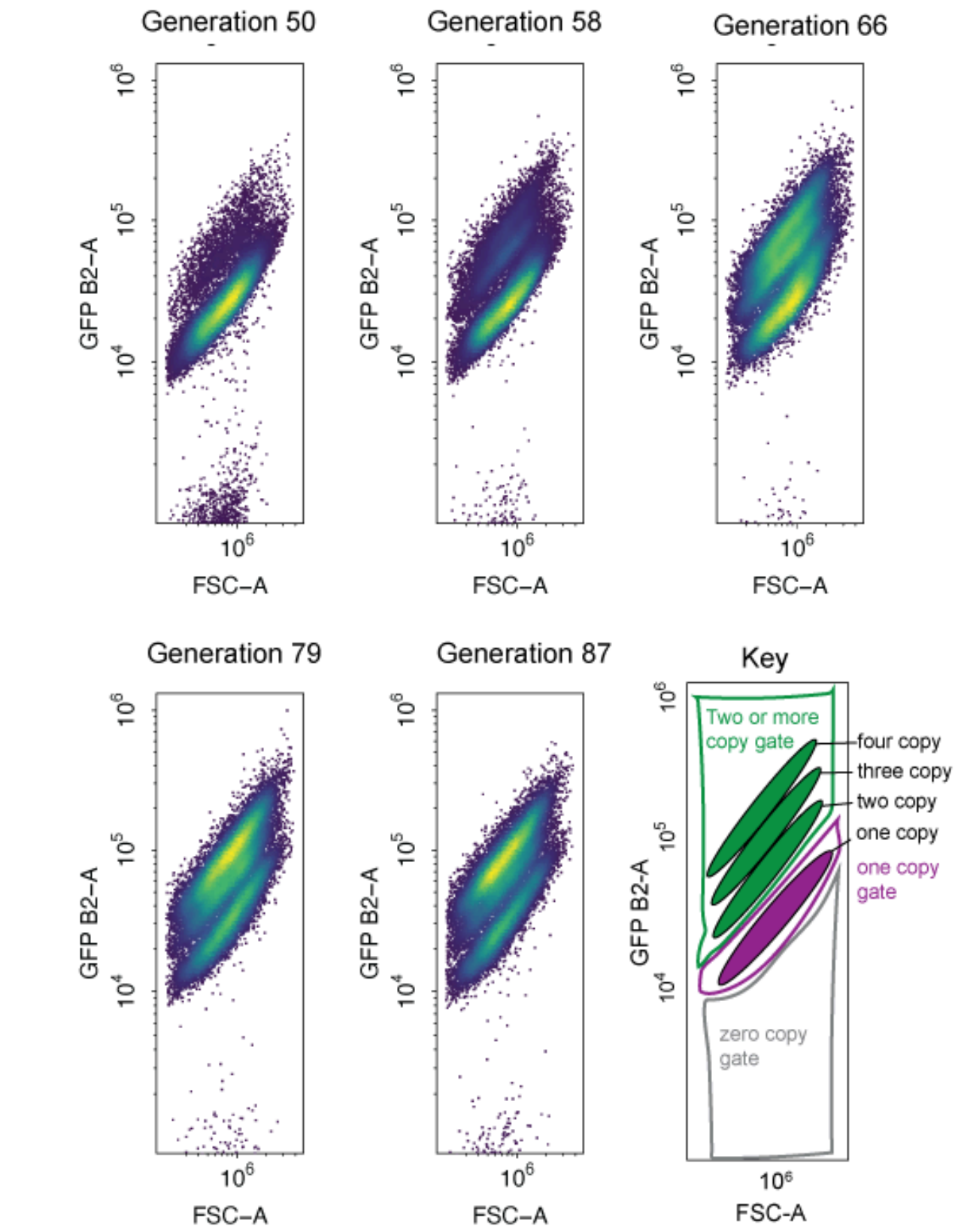
Wildtype population 2



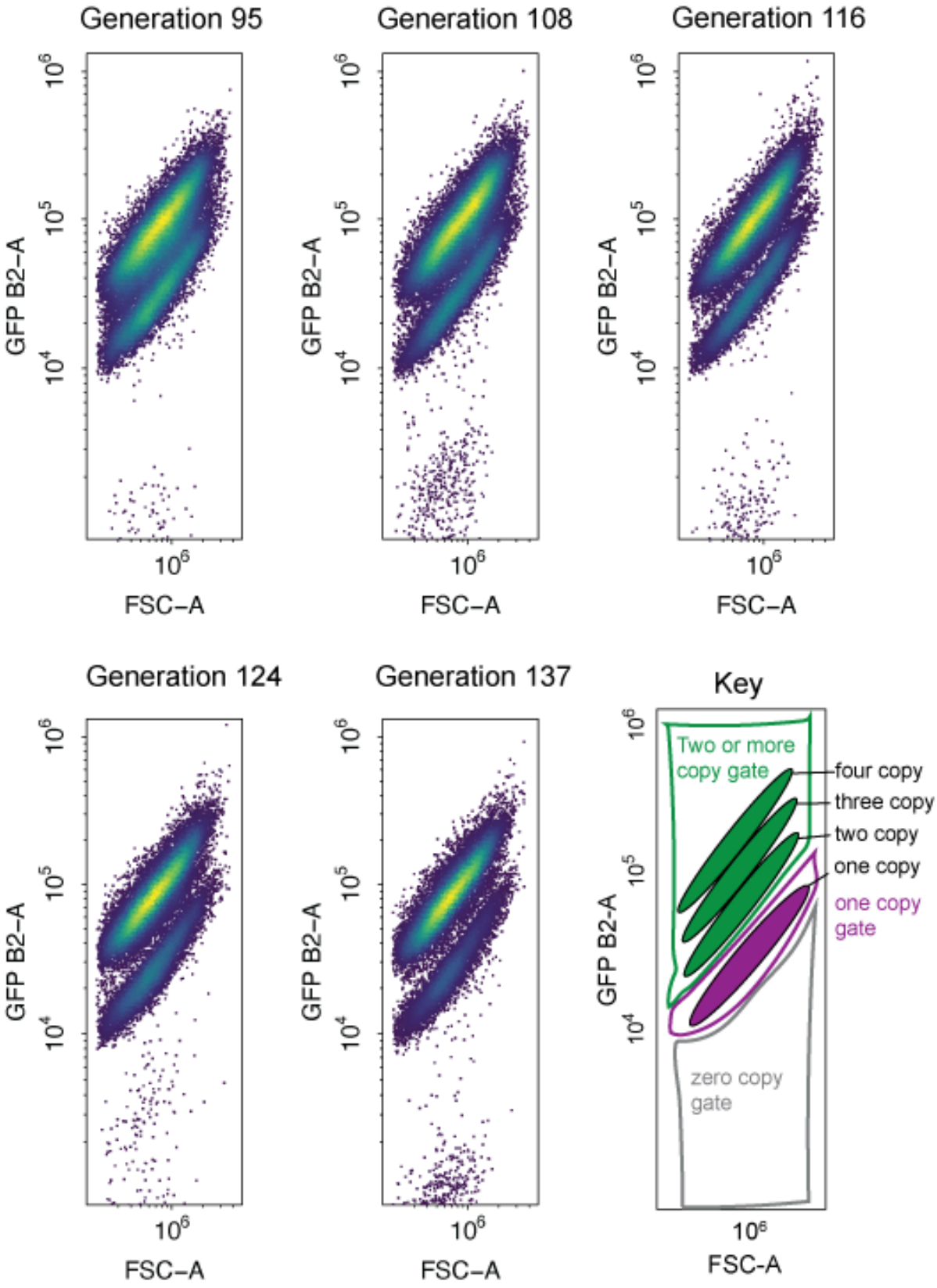
Wildtype population 2



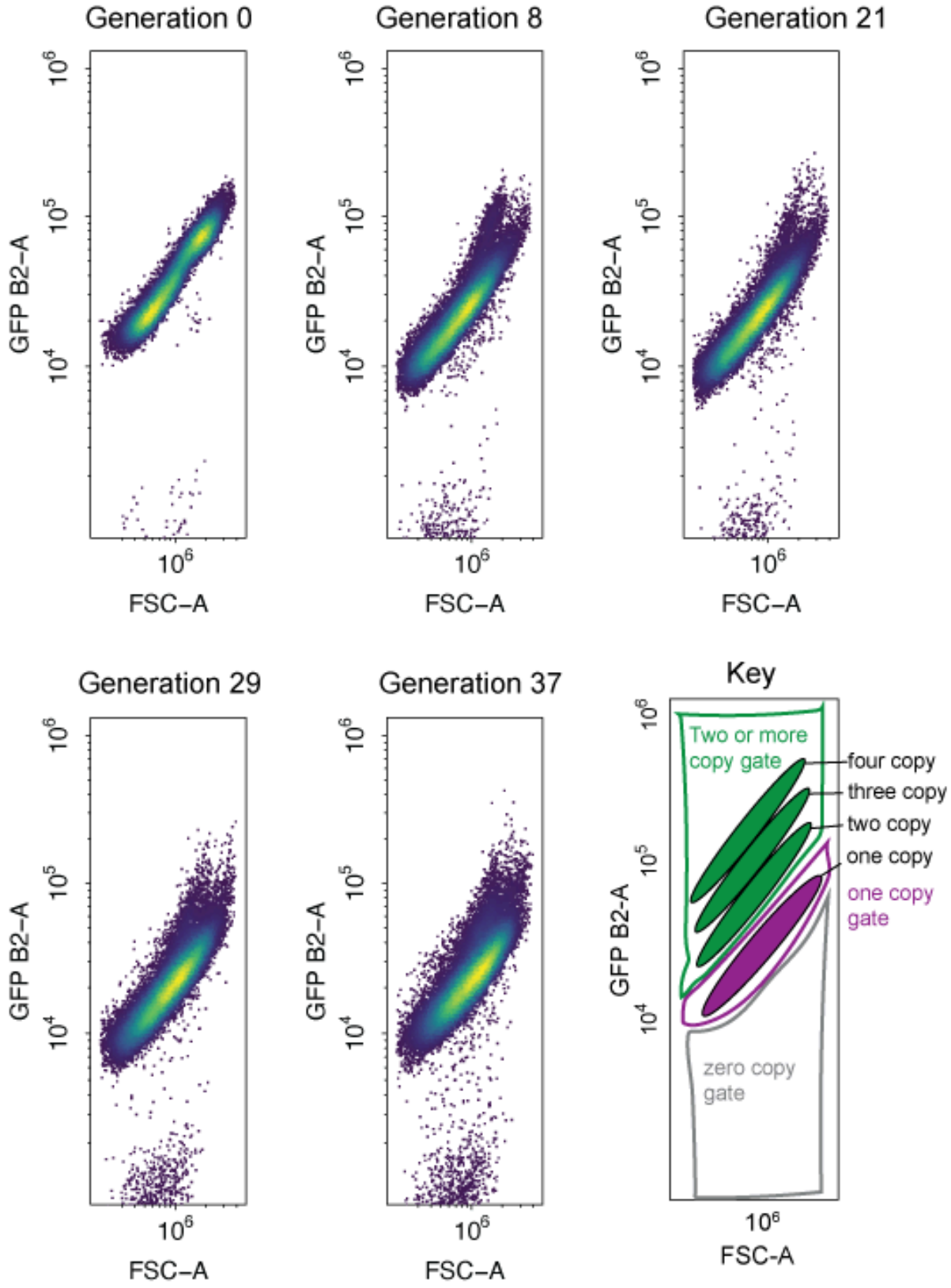




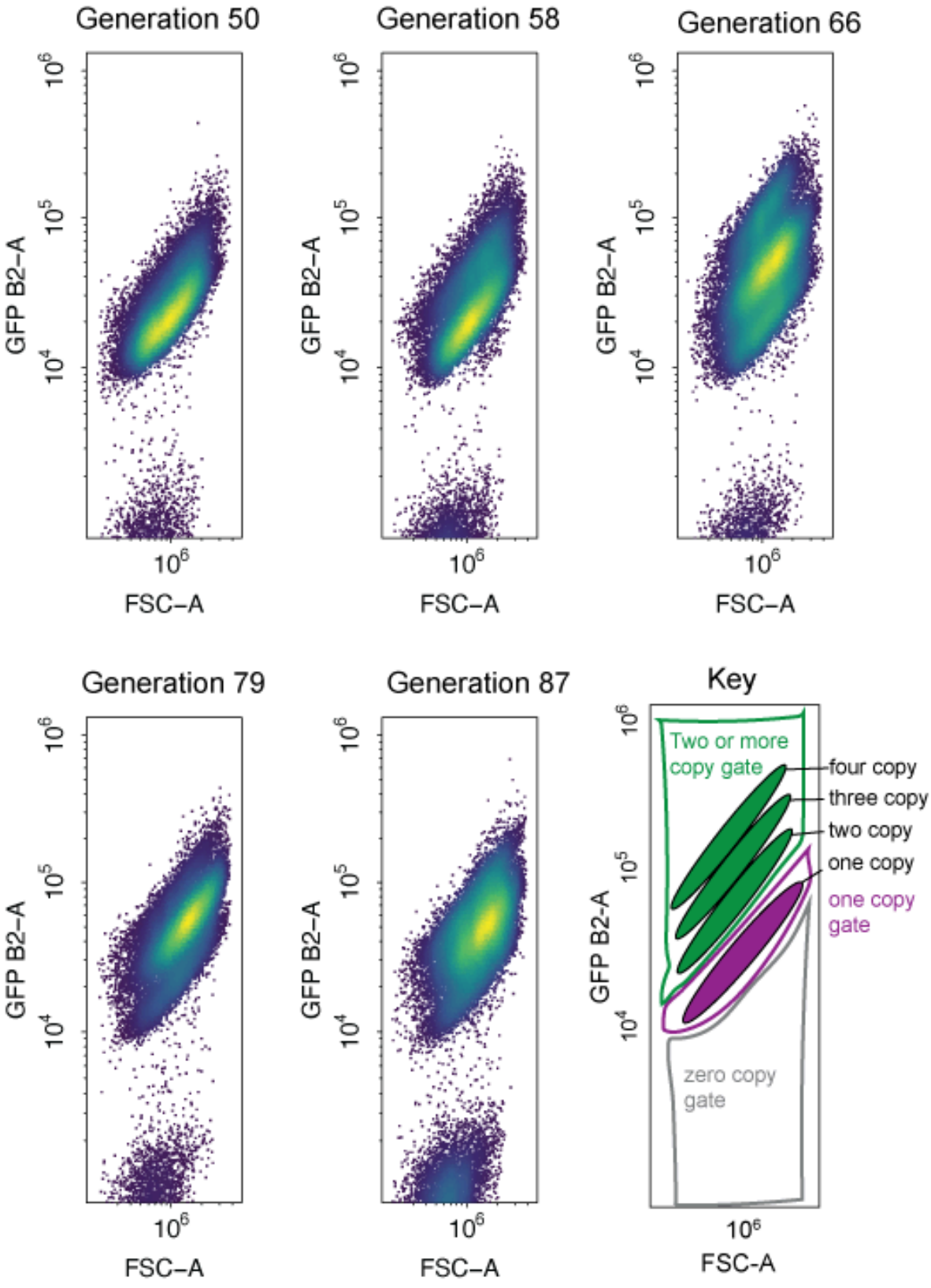
Wildtype population 3



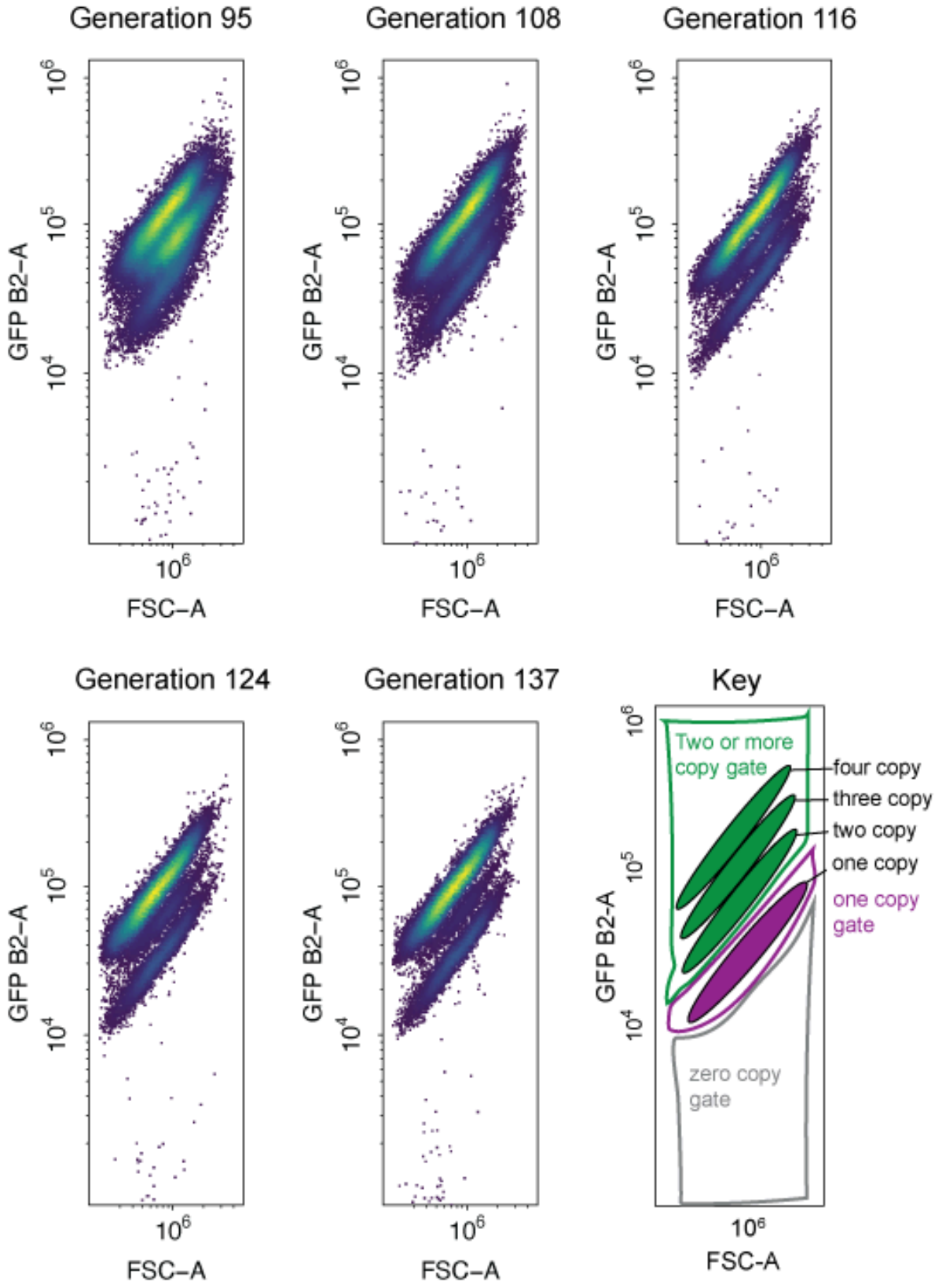
Wildtype population 4



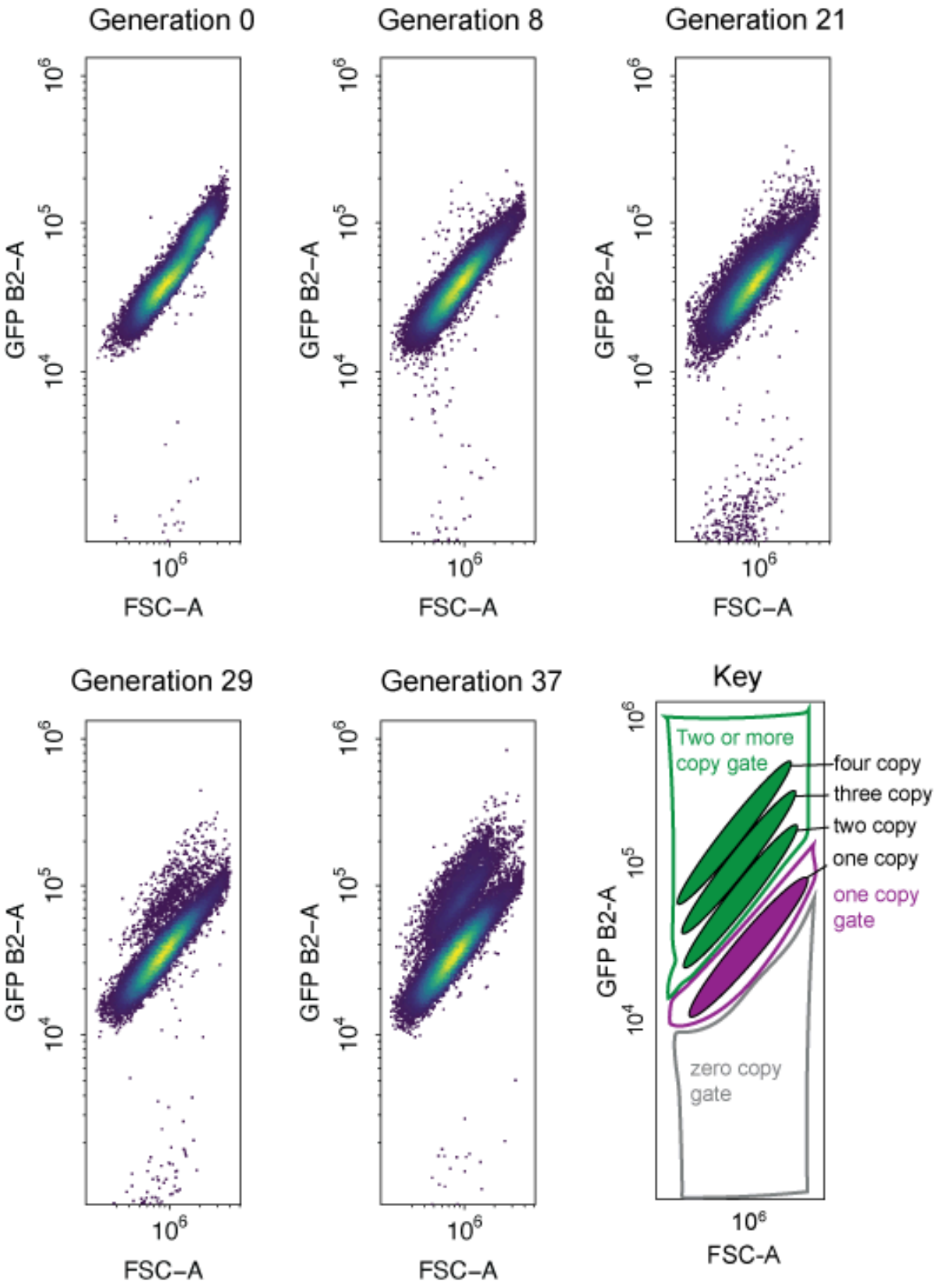
Wildtype population 4



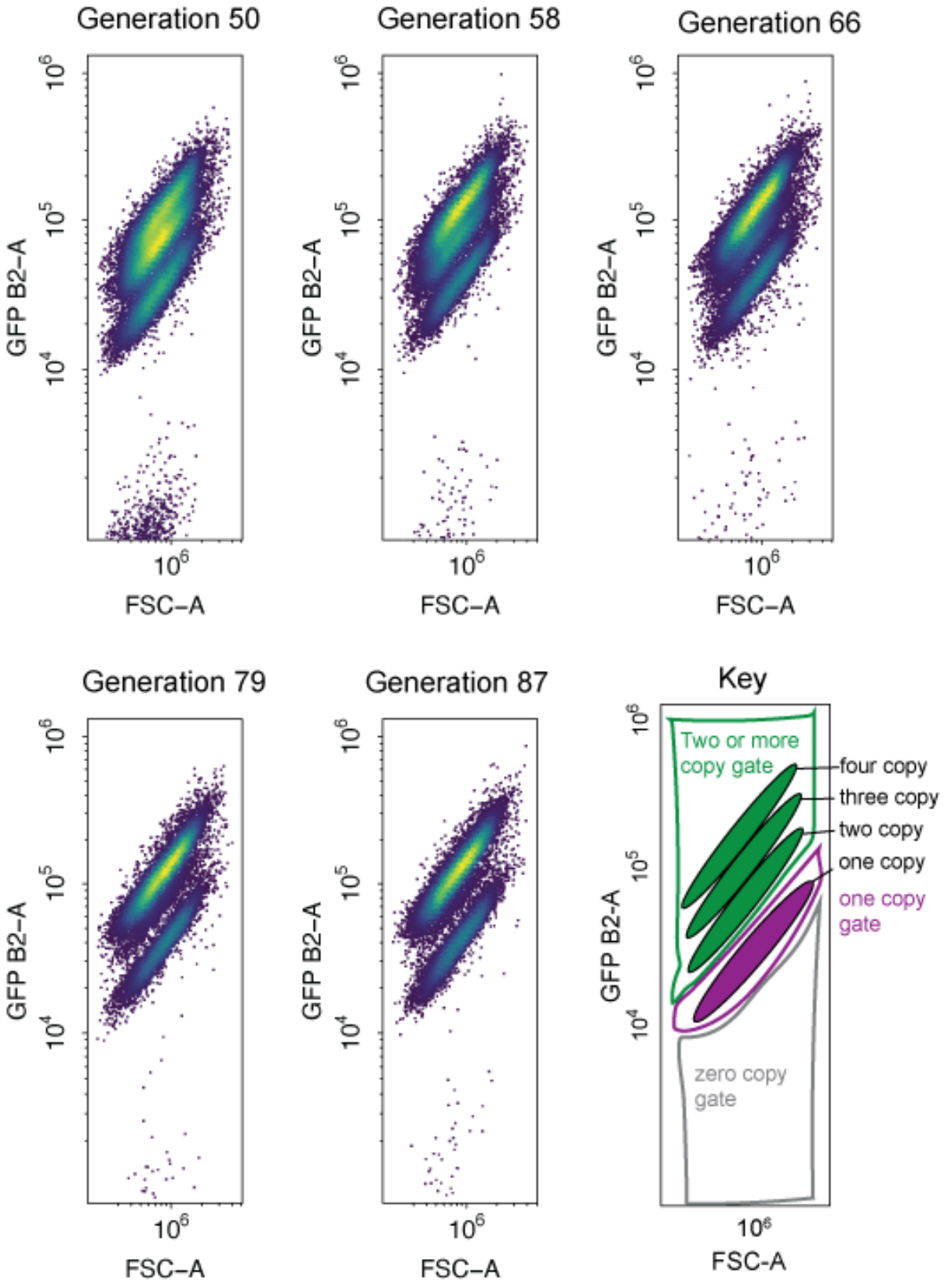
Wildtype population 4



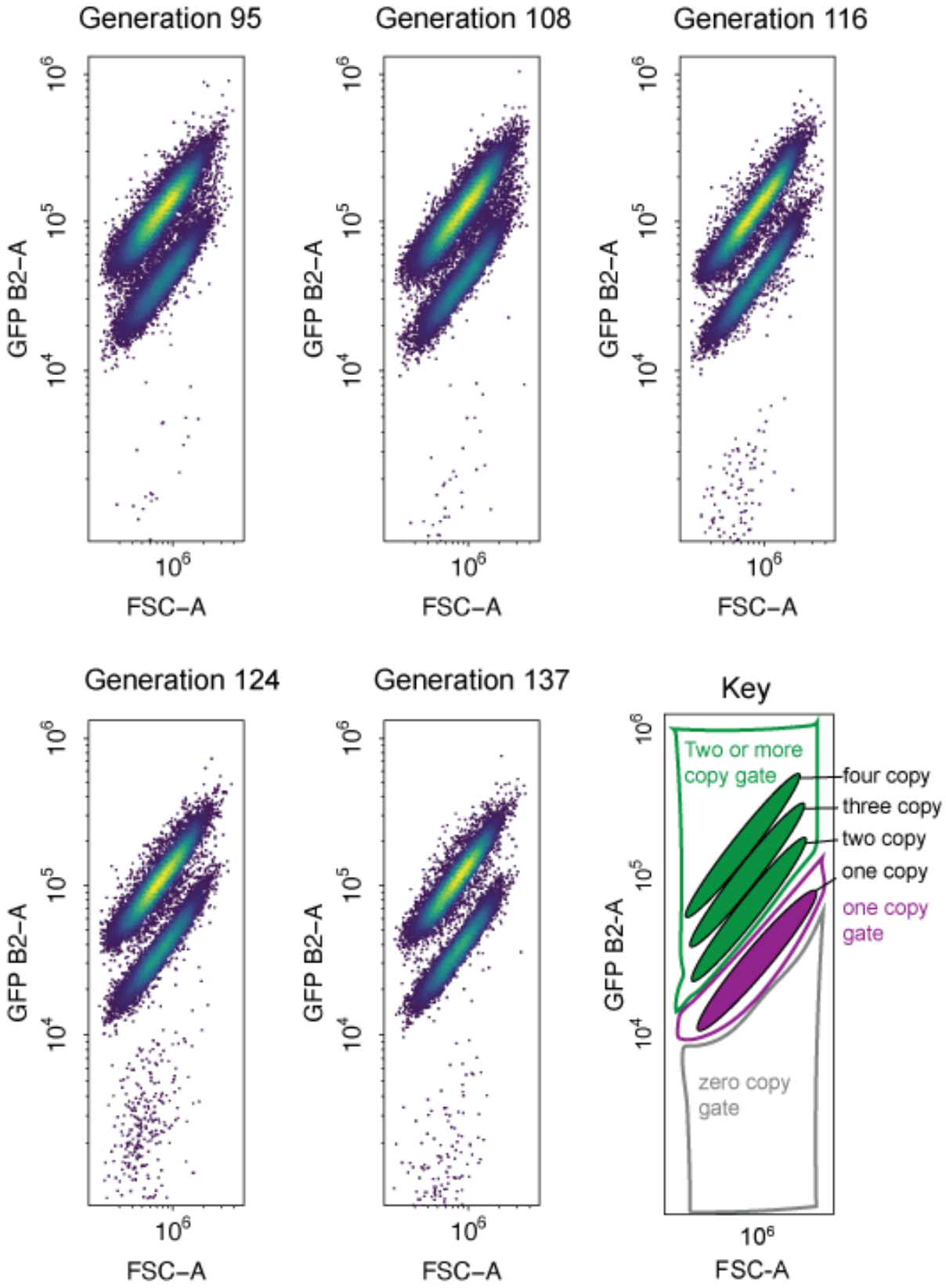
Wildtype population 5



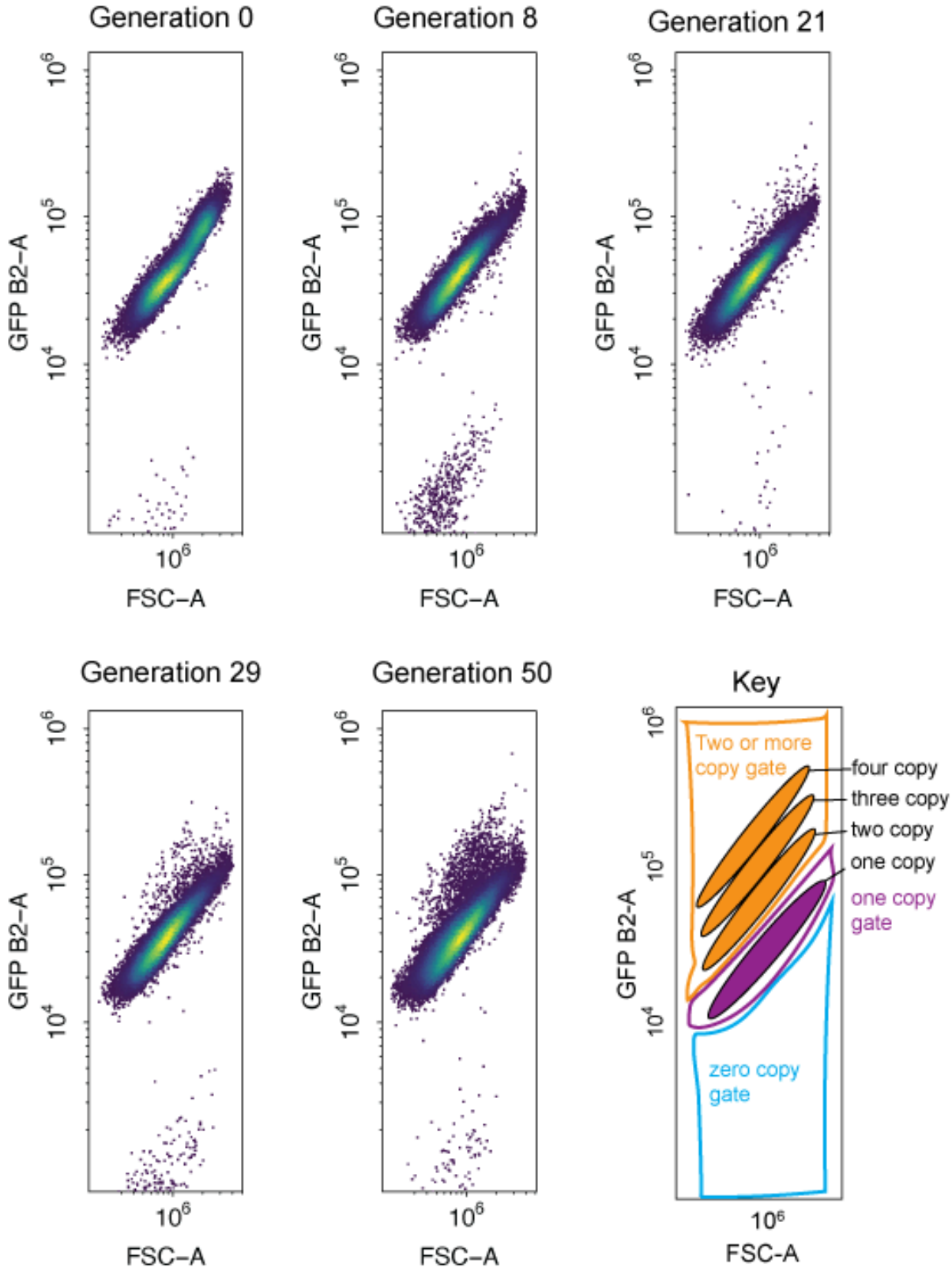
Wildtype population 5



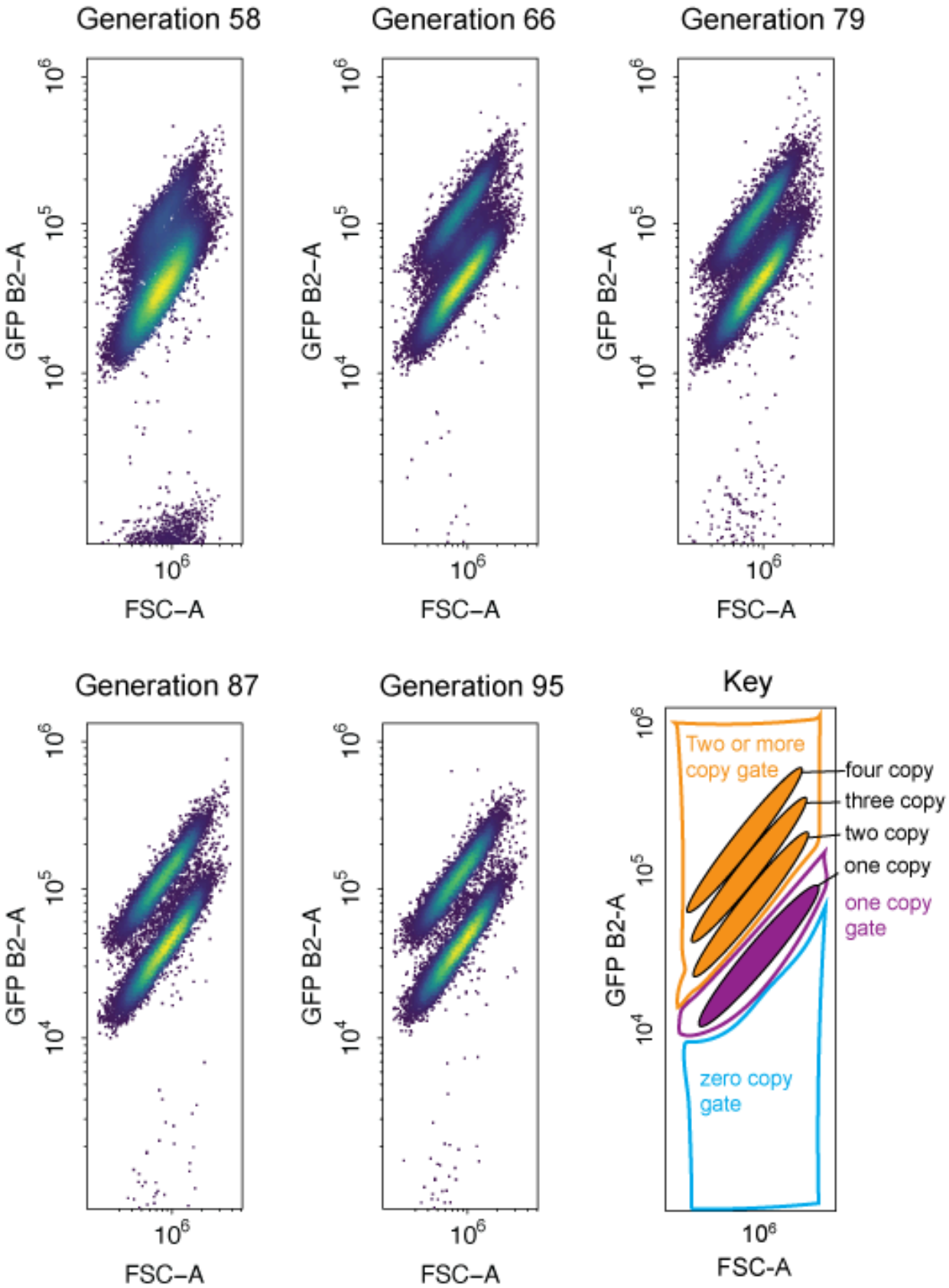
Wildtype population 5



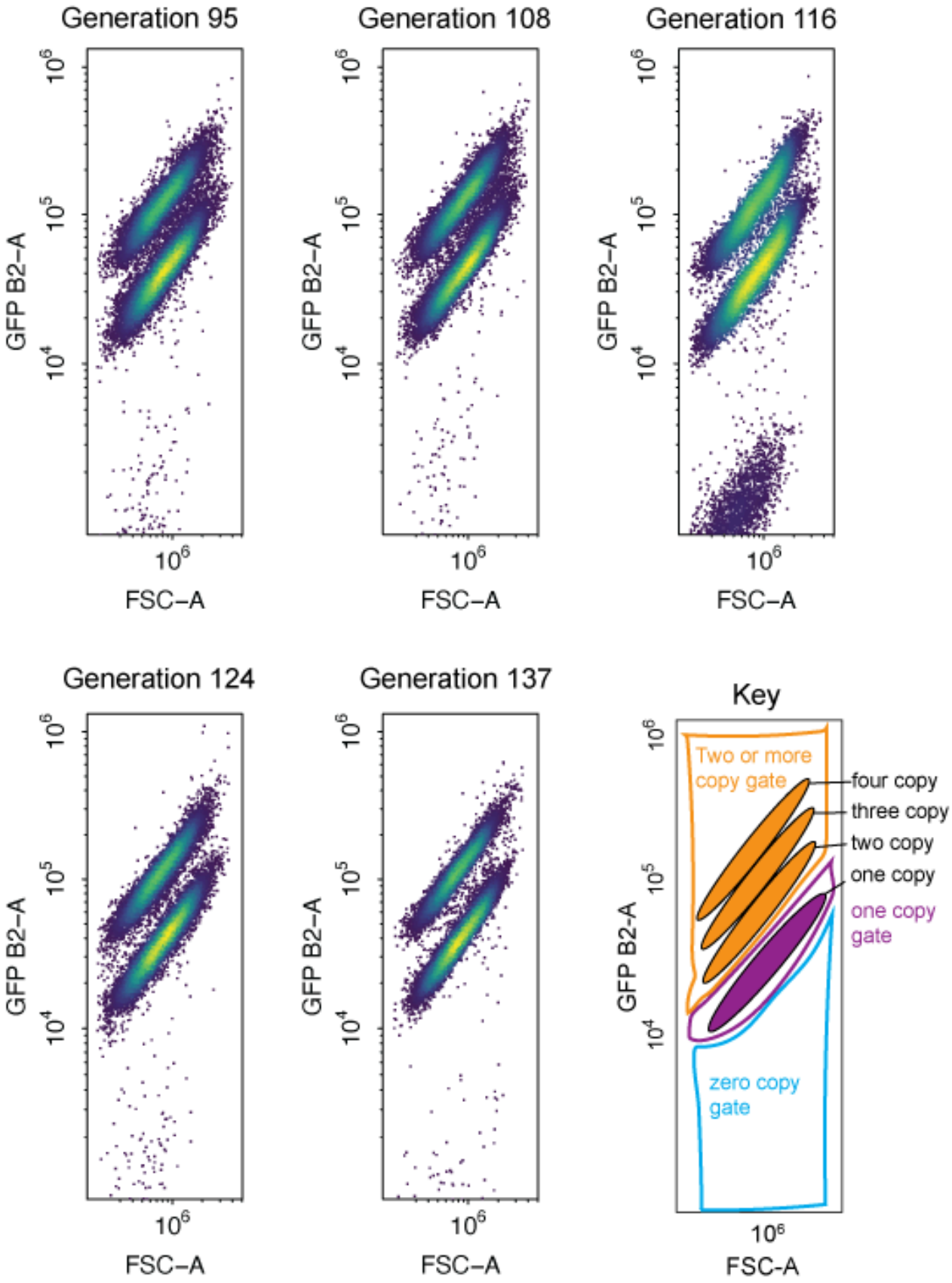
LTRΔ population 1



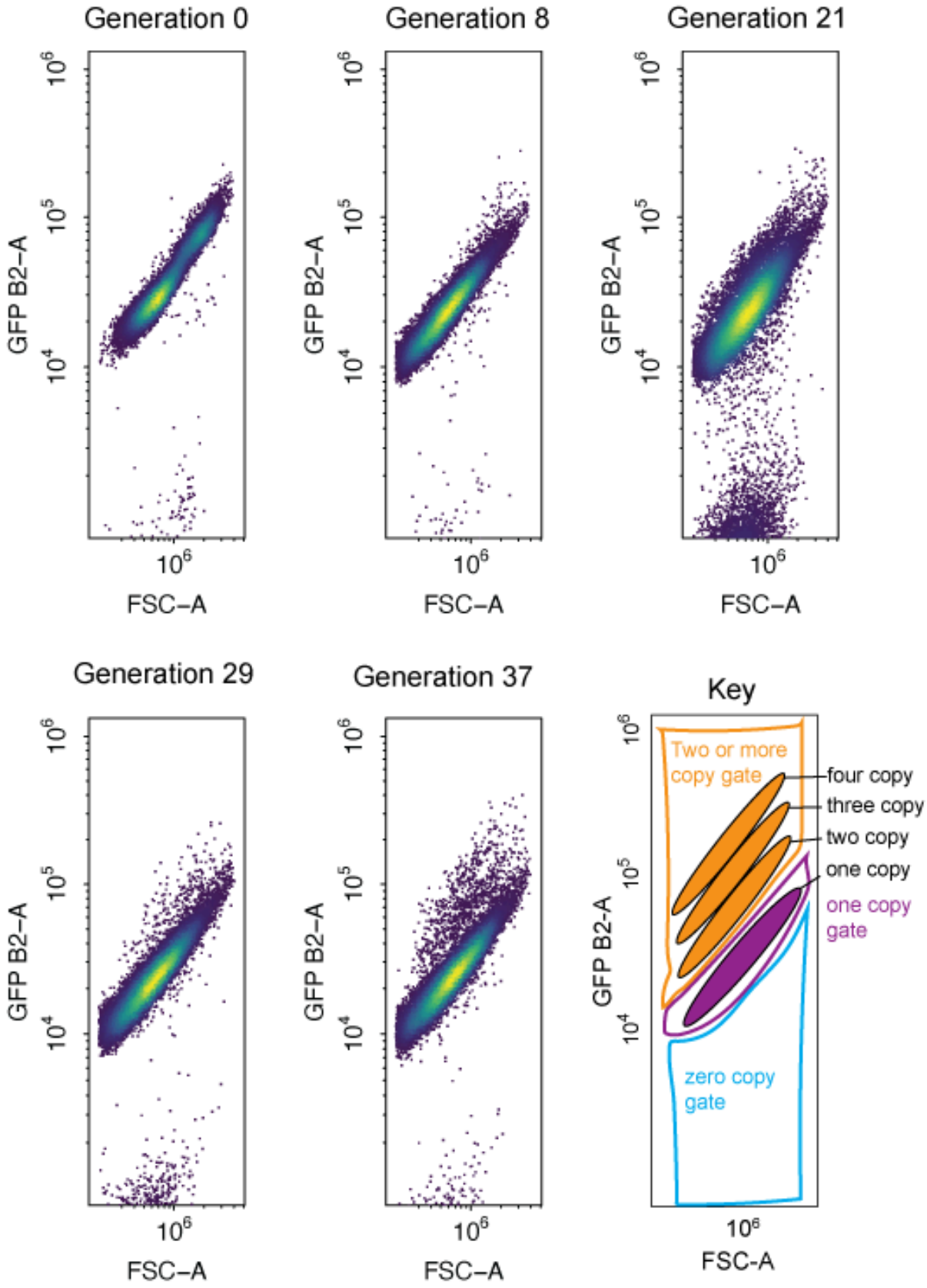
LTRΔ population 1



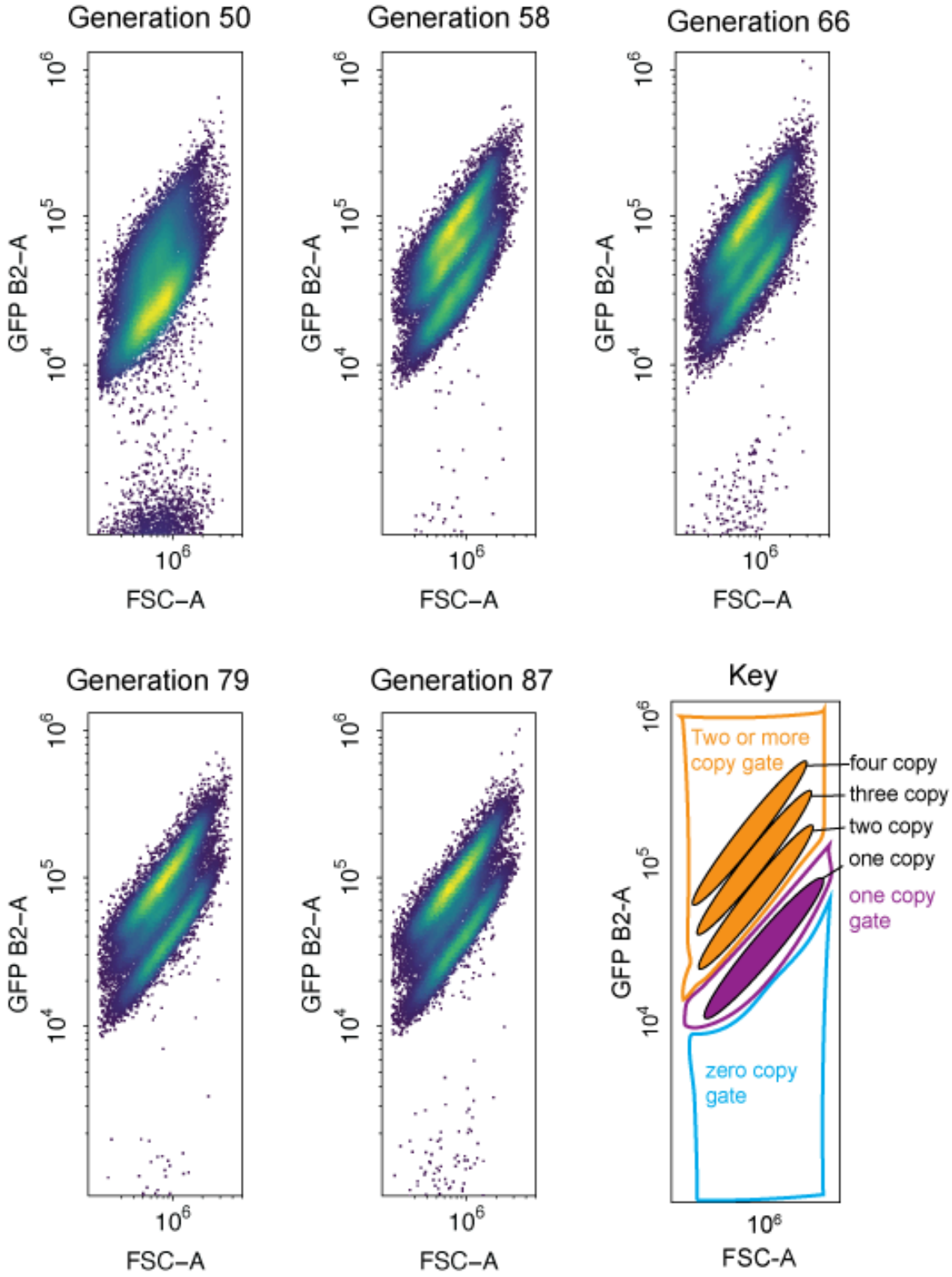
LTRΔ population 1



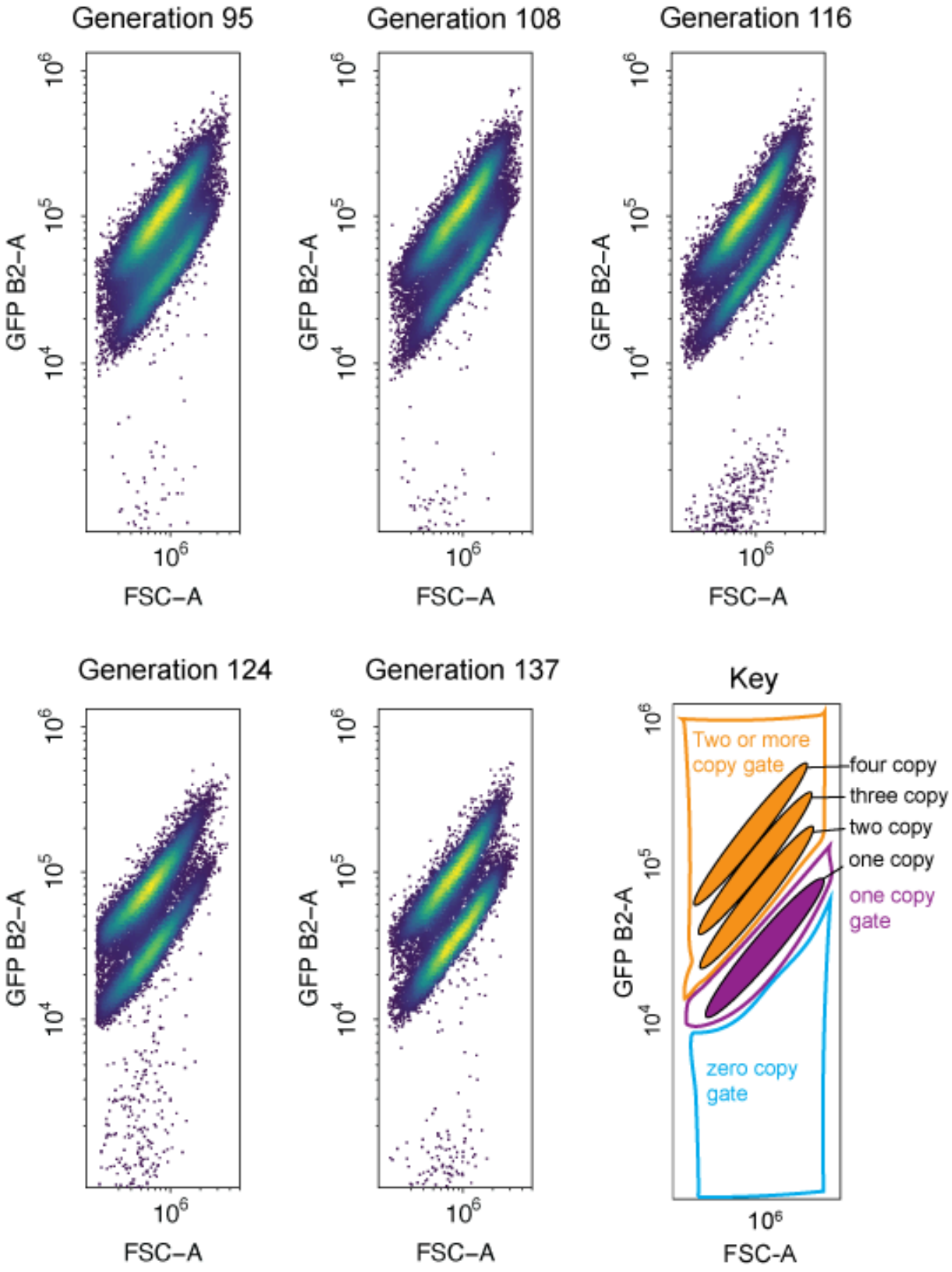
LTRΔ population 3



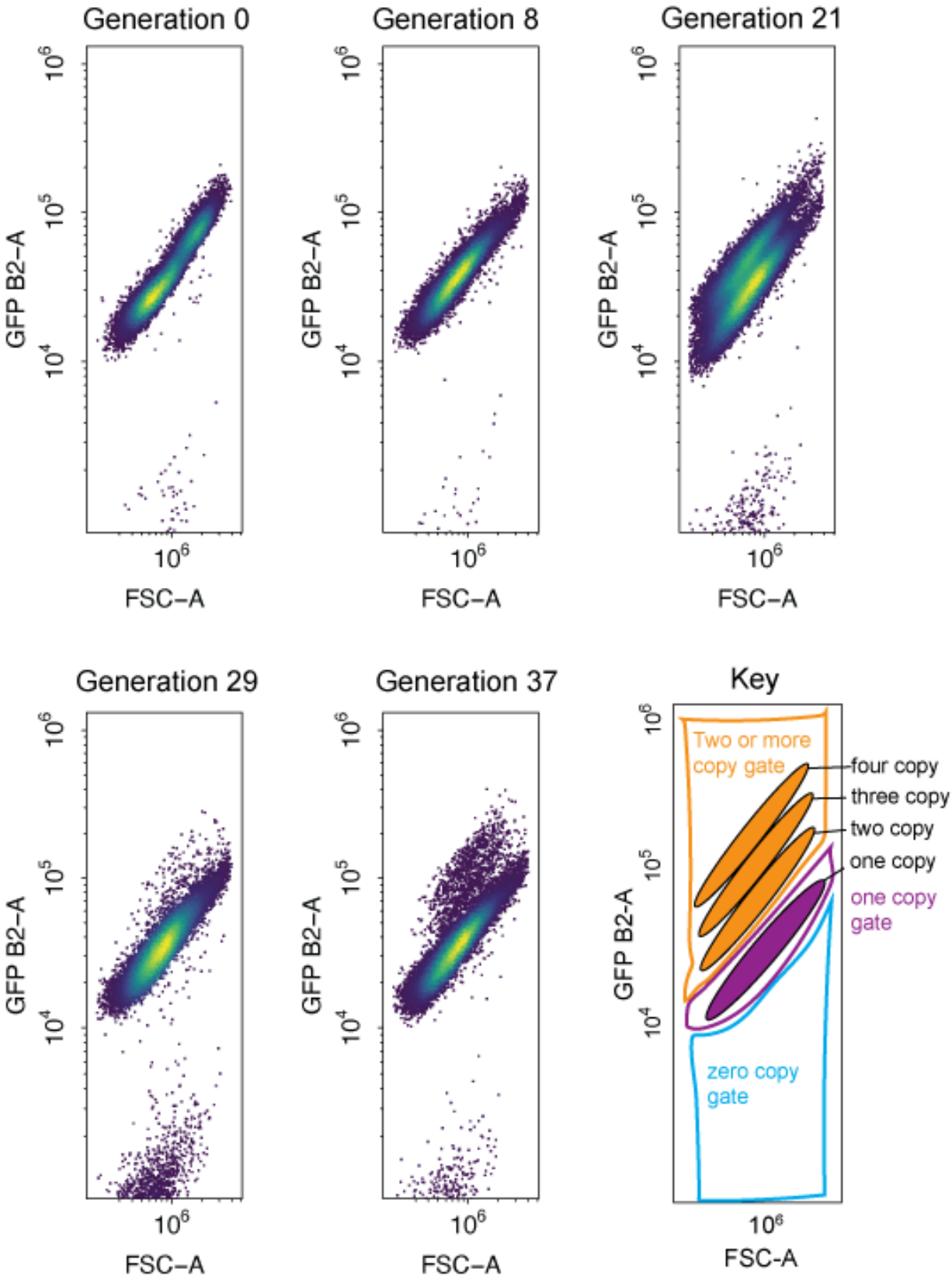
LTR∆ population 3



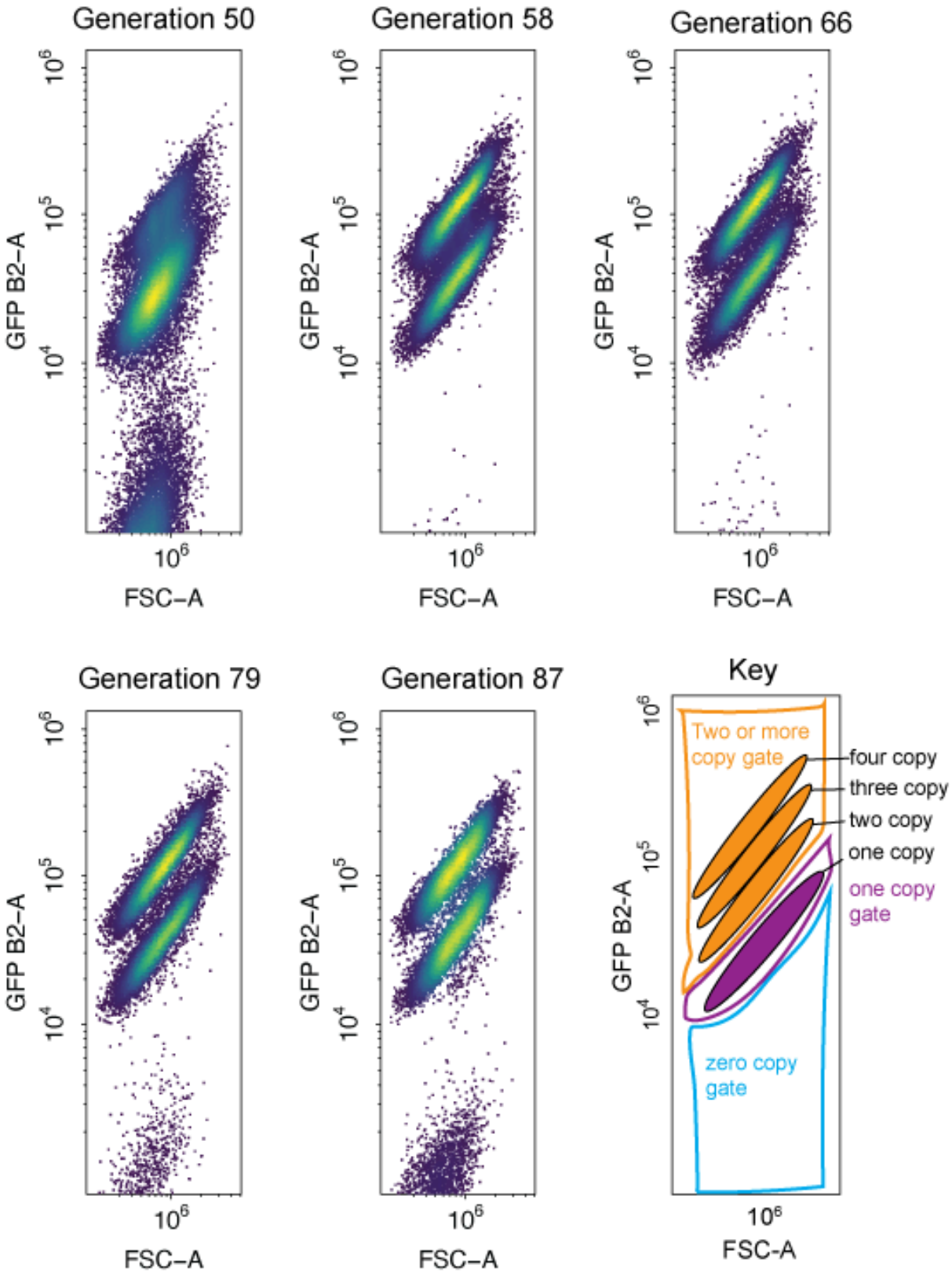
LTRΔ population 3



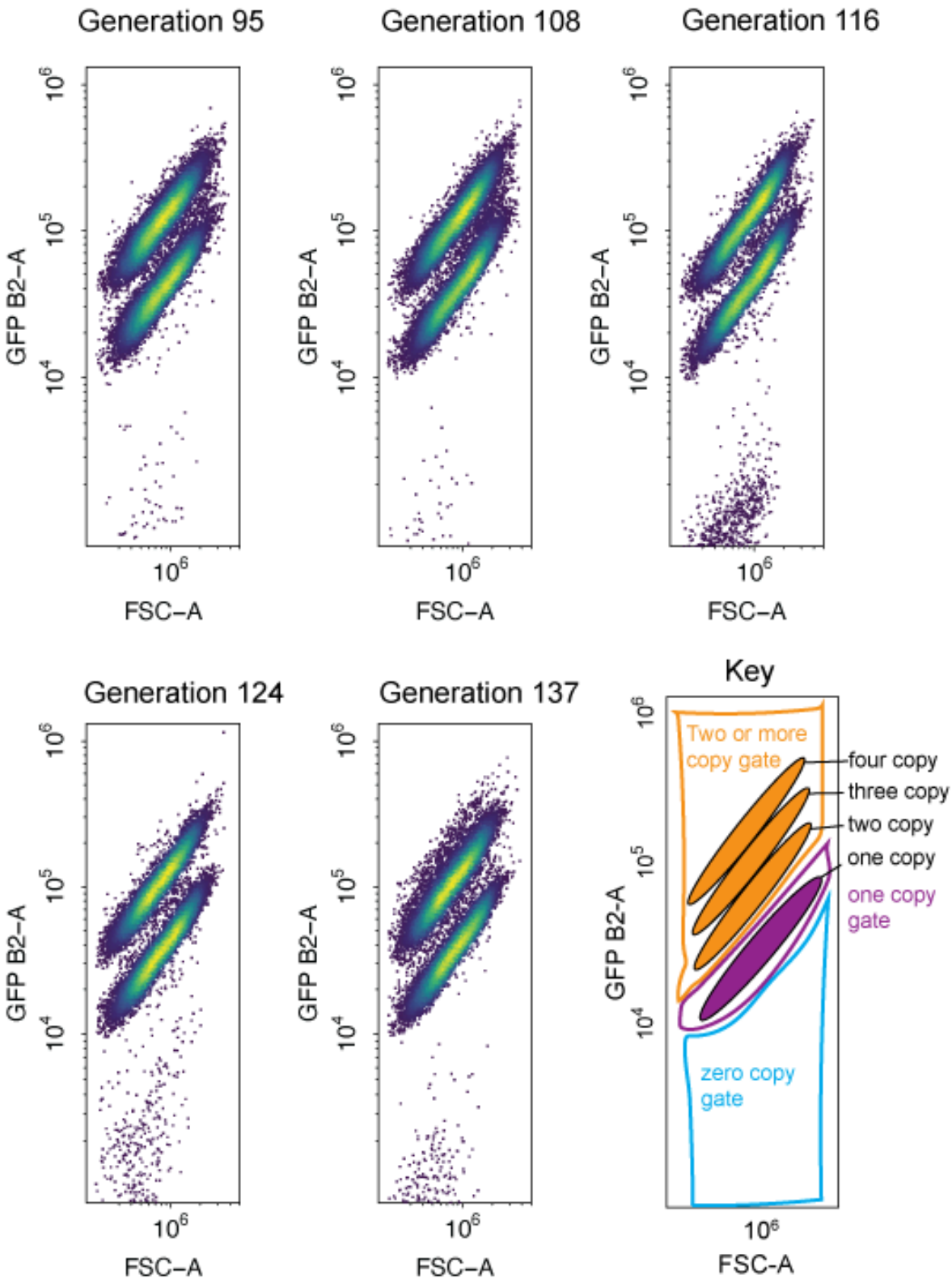
LTRΔ population 4

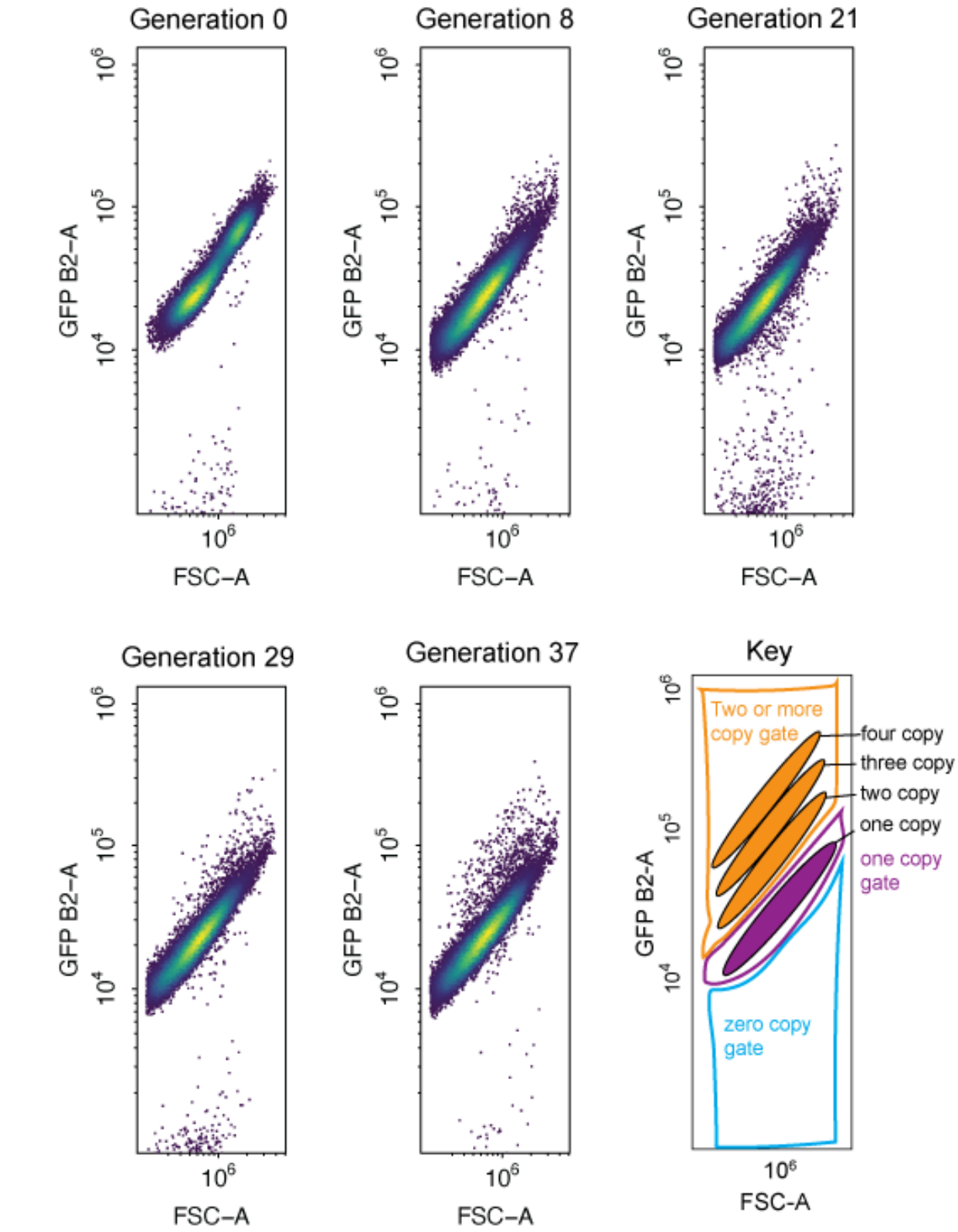


LTRΔ population 4

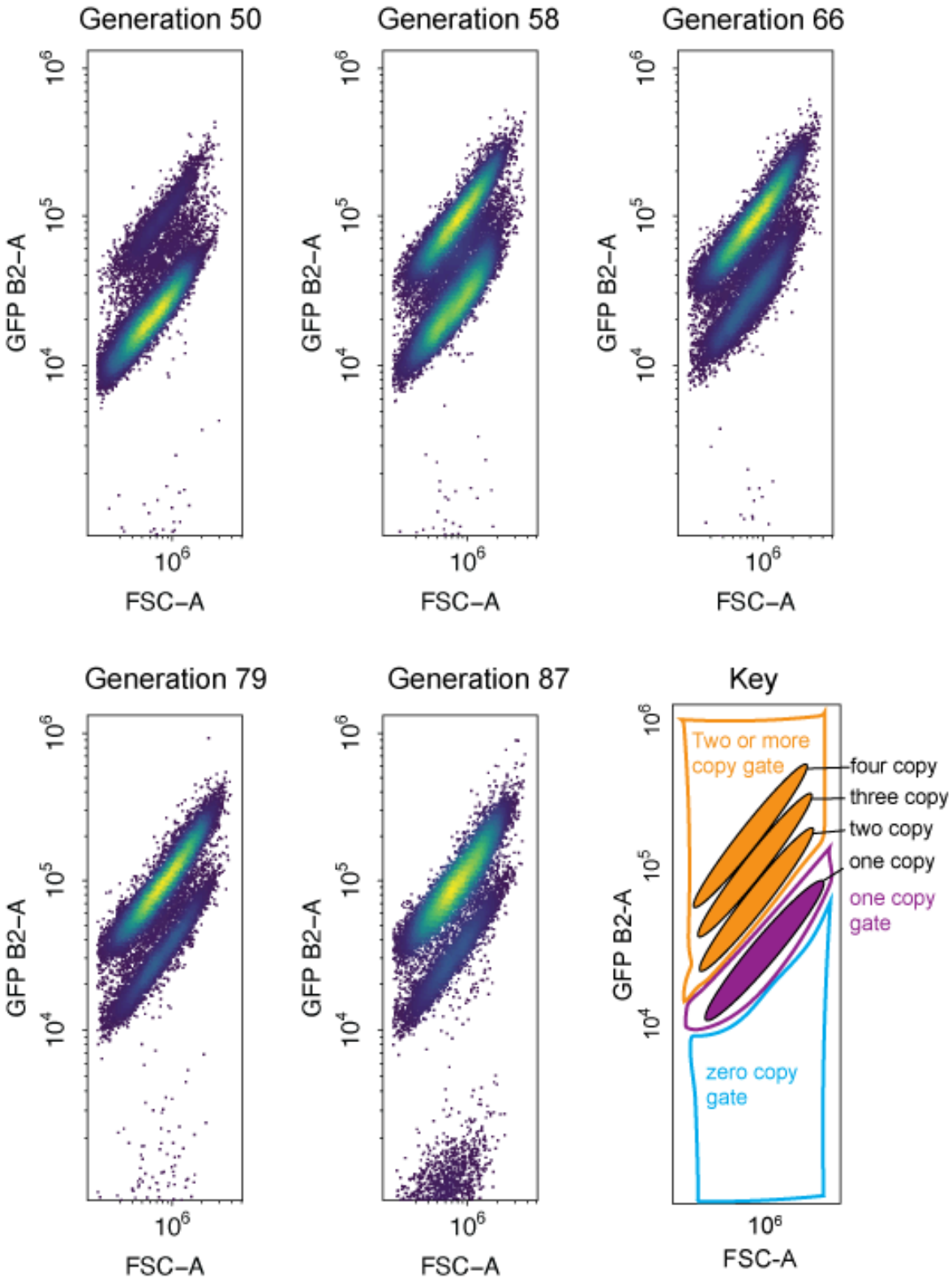


LTRΔ population 4

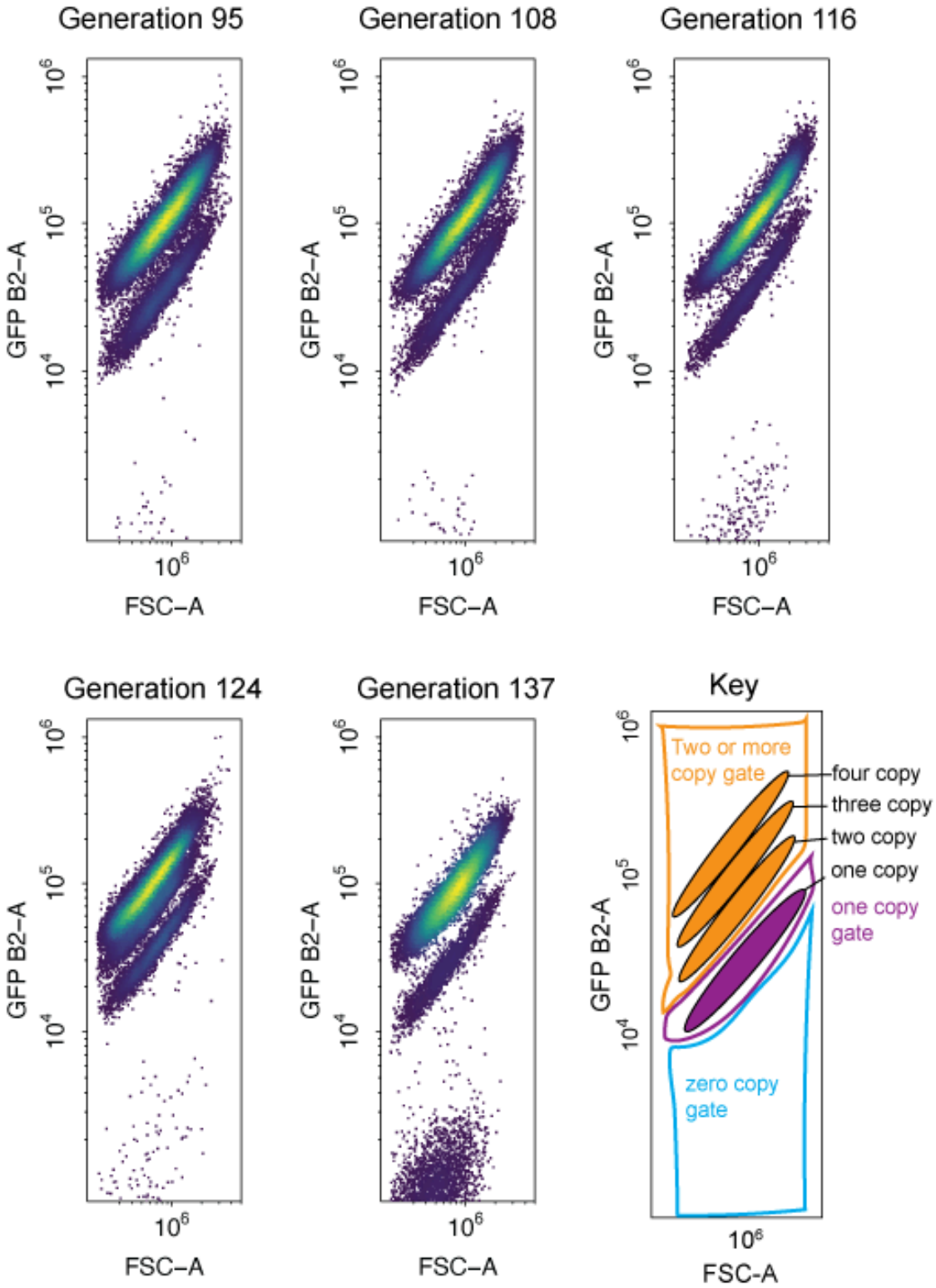




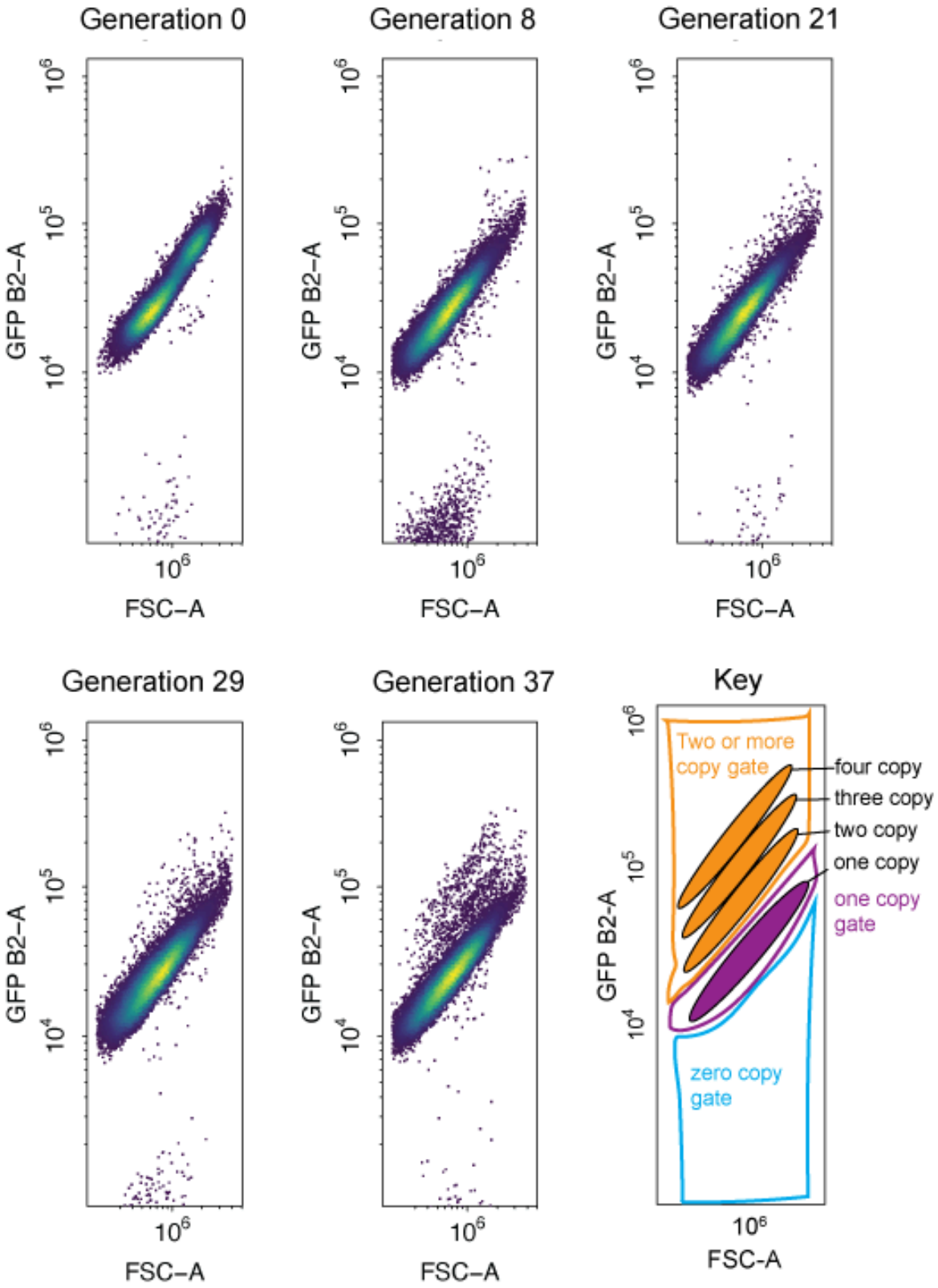
LTRΔ population 5



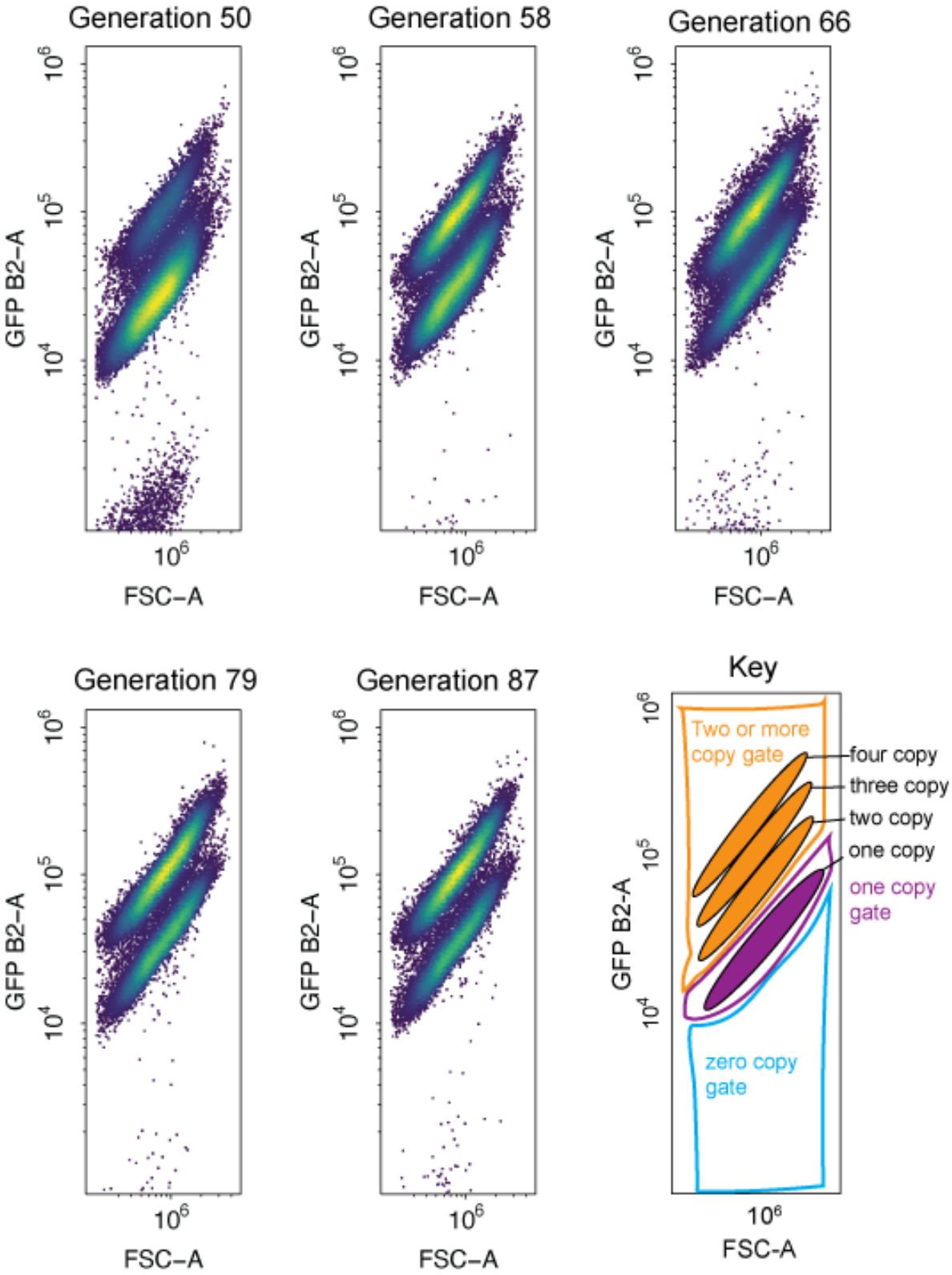
LTR△ population 5



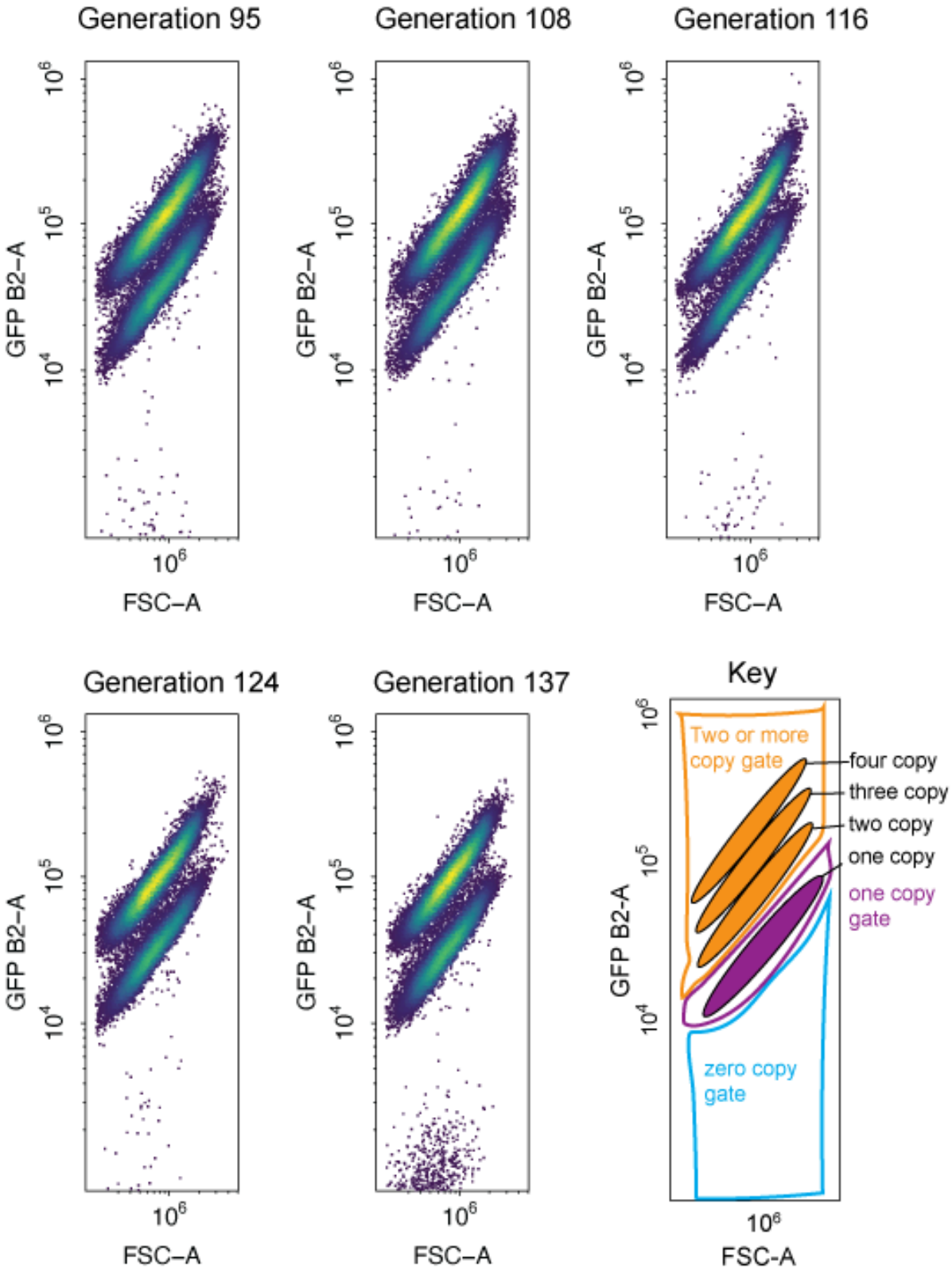
LTRΔ population 6



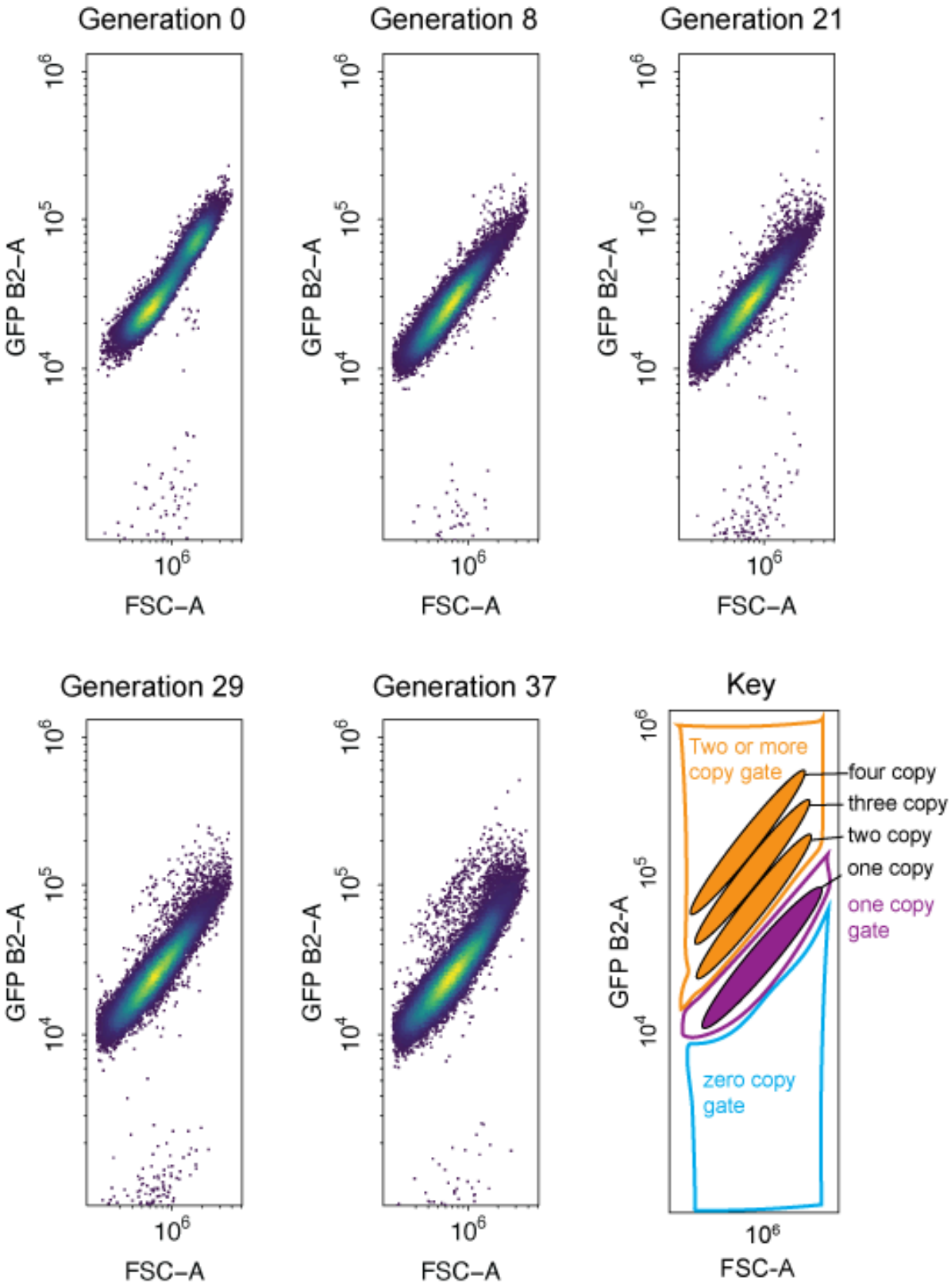
LTRΔ population 6



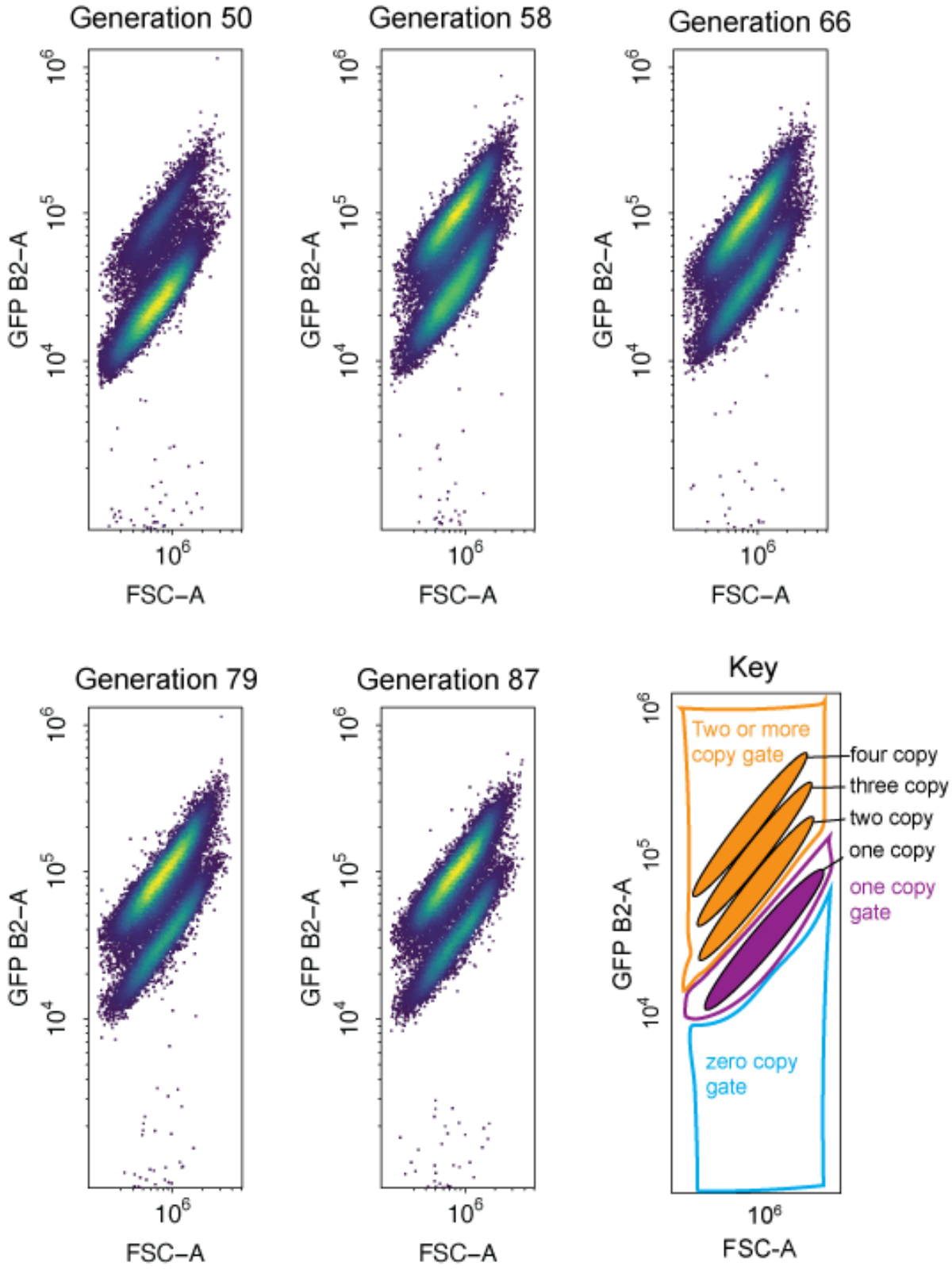
LTRΔ population 6



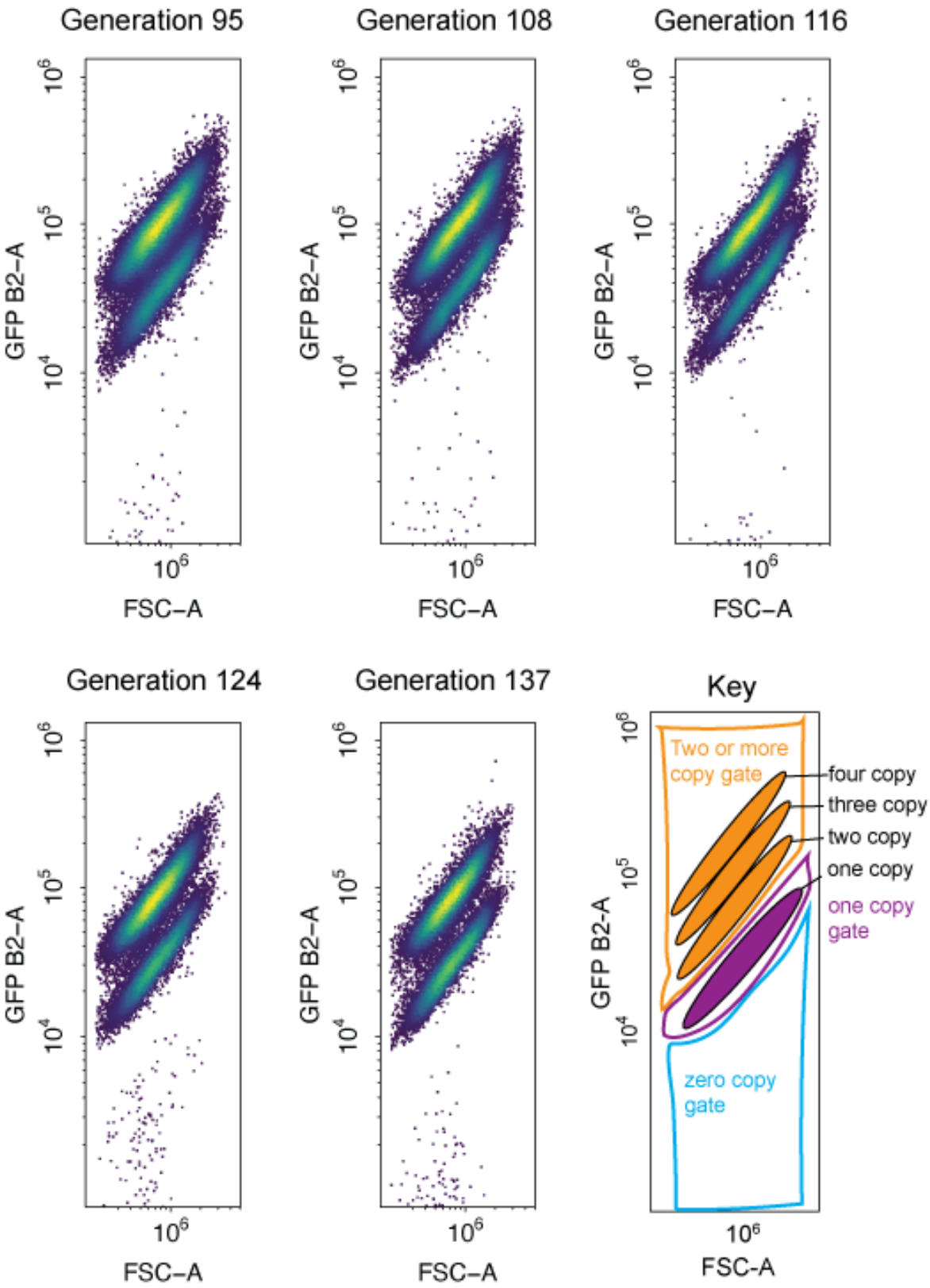
LTR∆ population 7



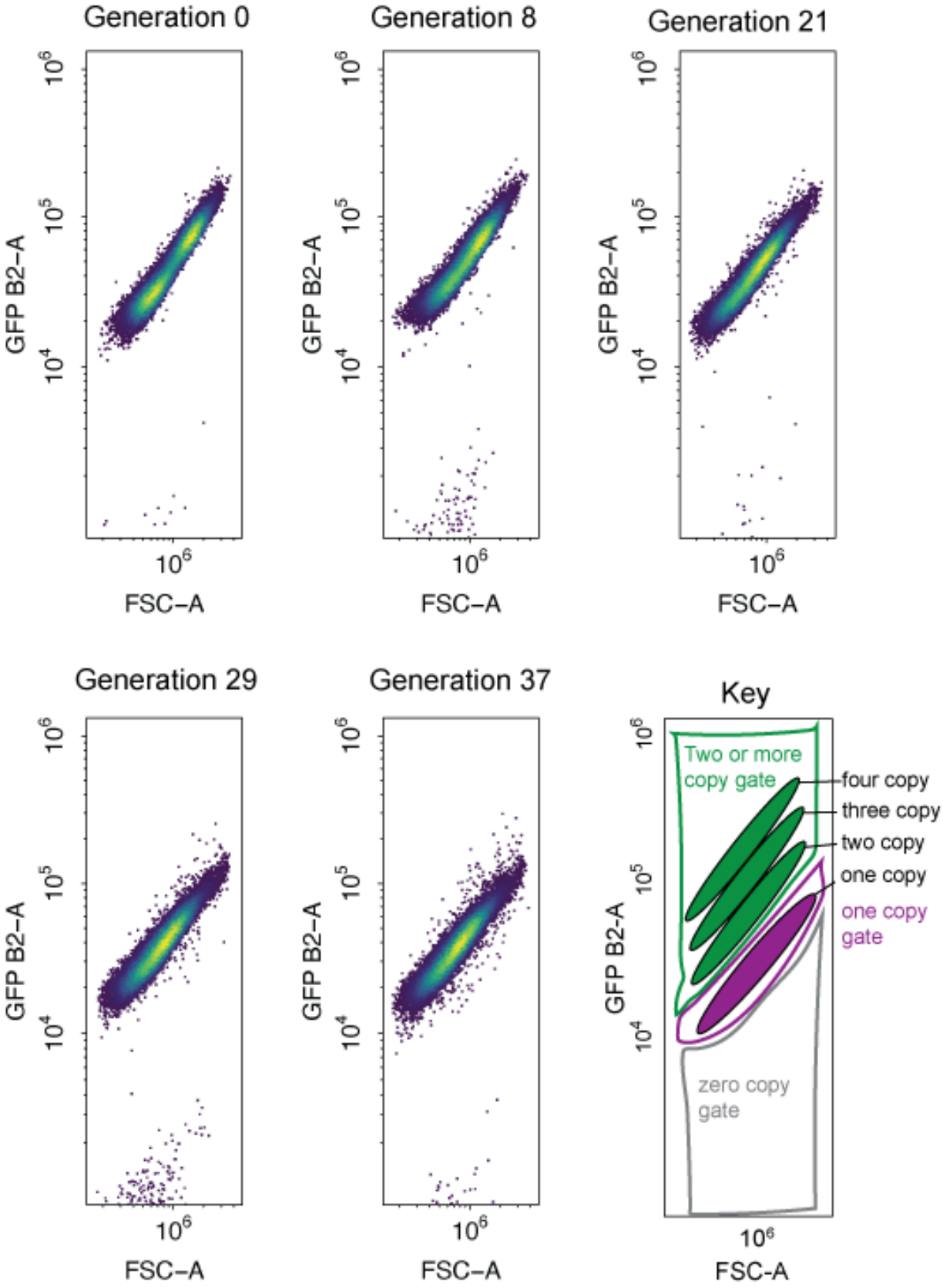
LTR∆ population 7



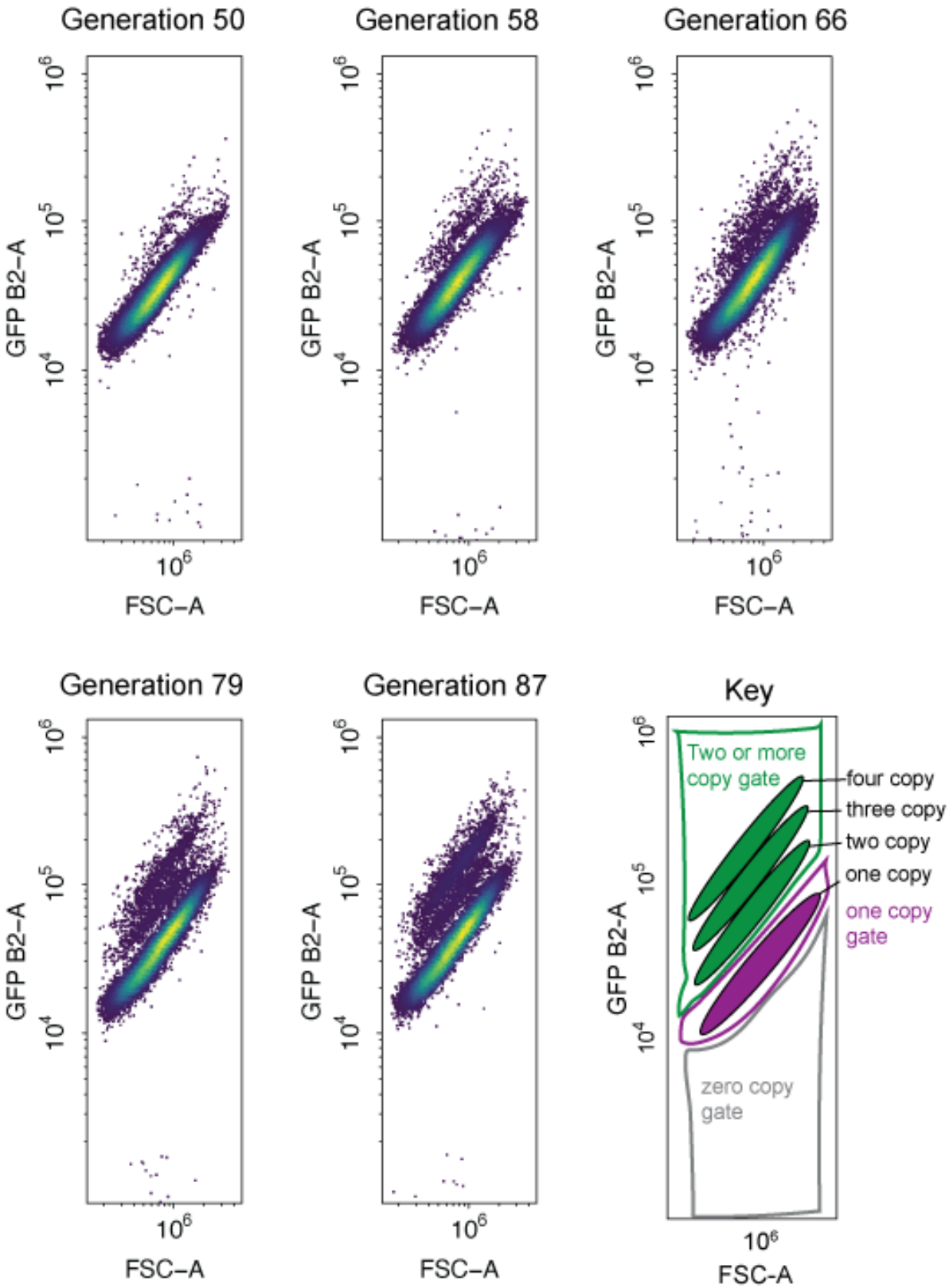
LTR∆ population 7



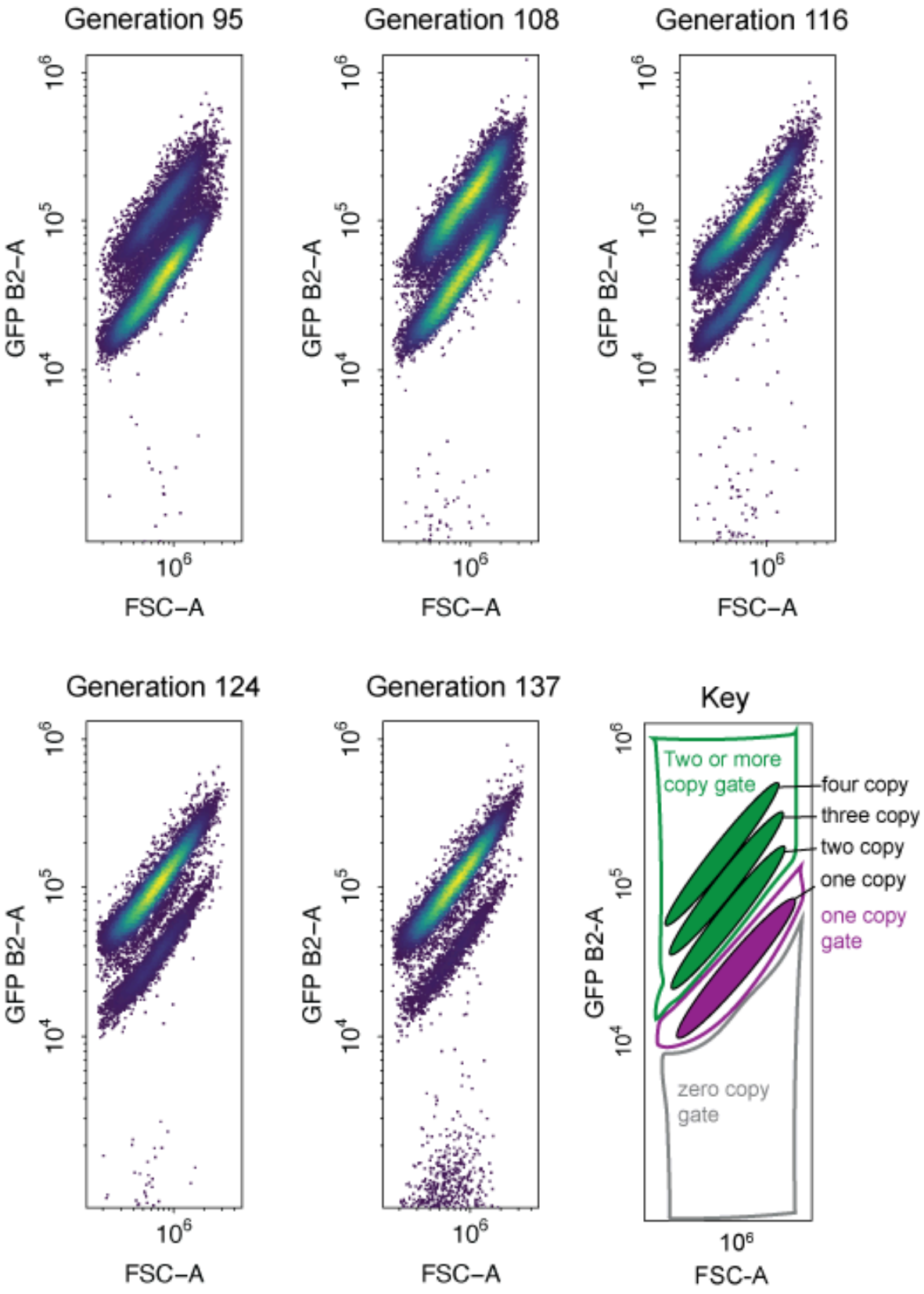
ARSΔ population 1

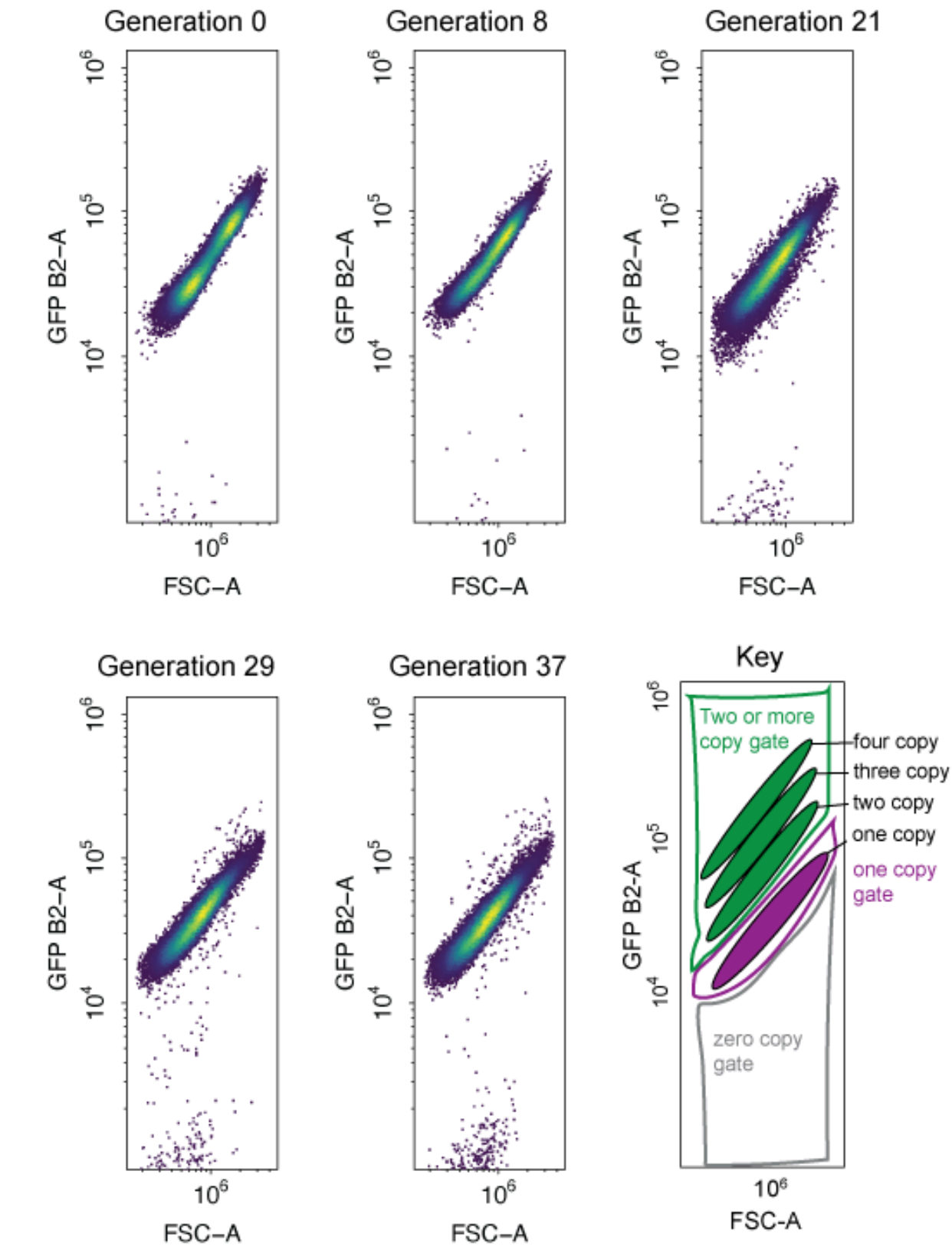


ARSΔ population 1

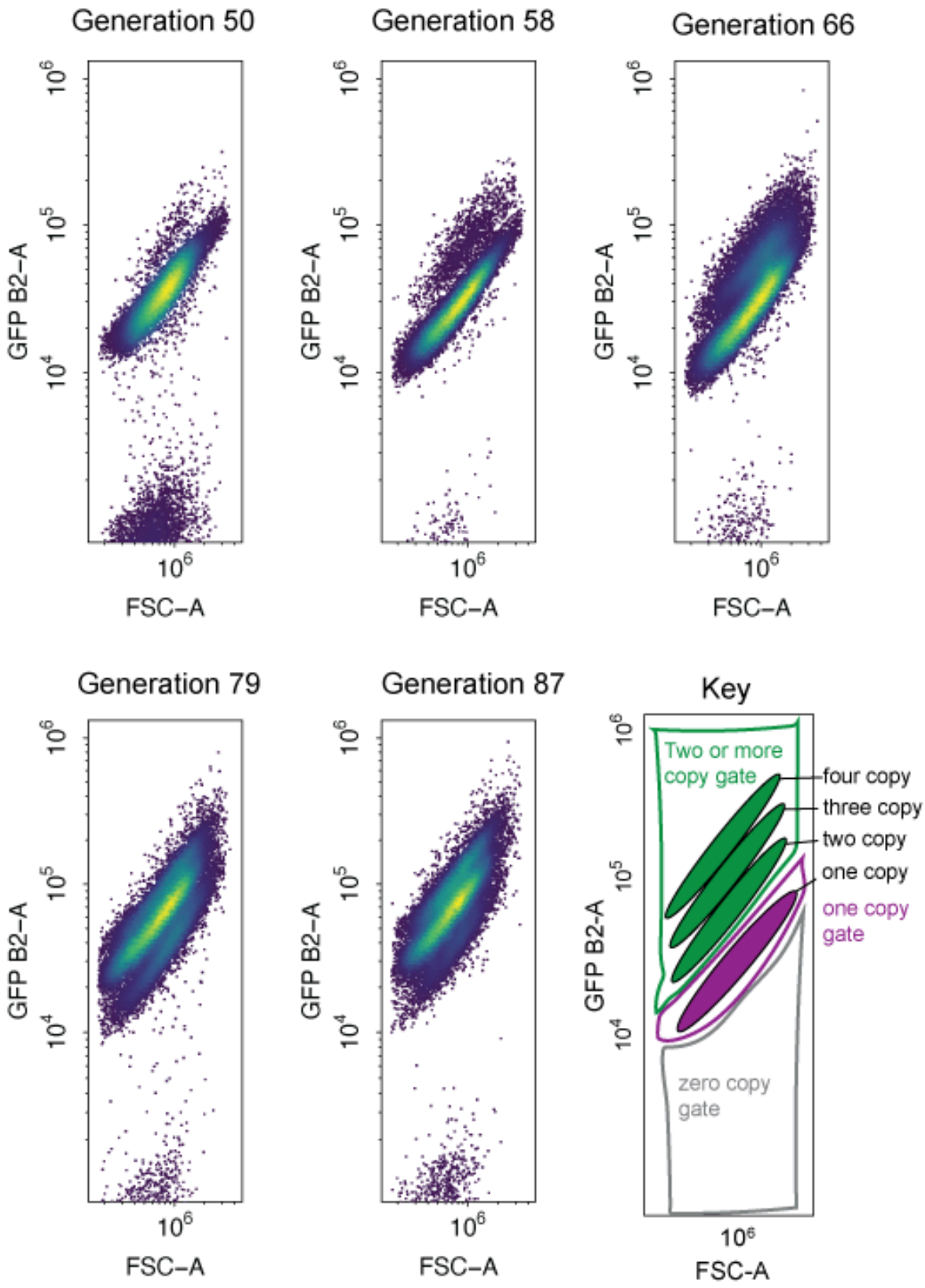


ARSΔ population 1

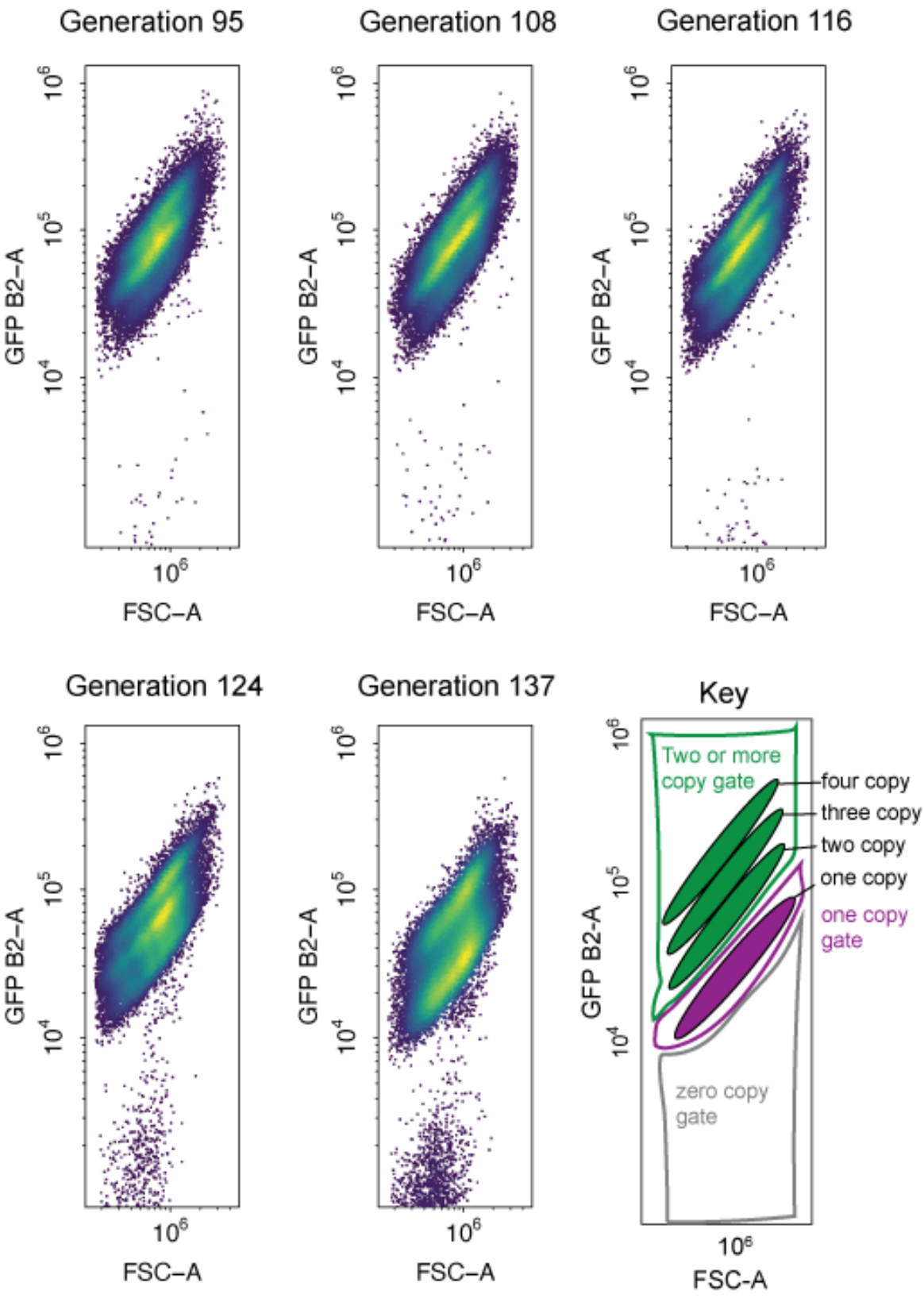




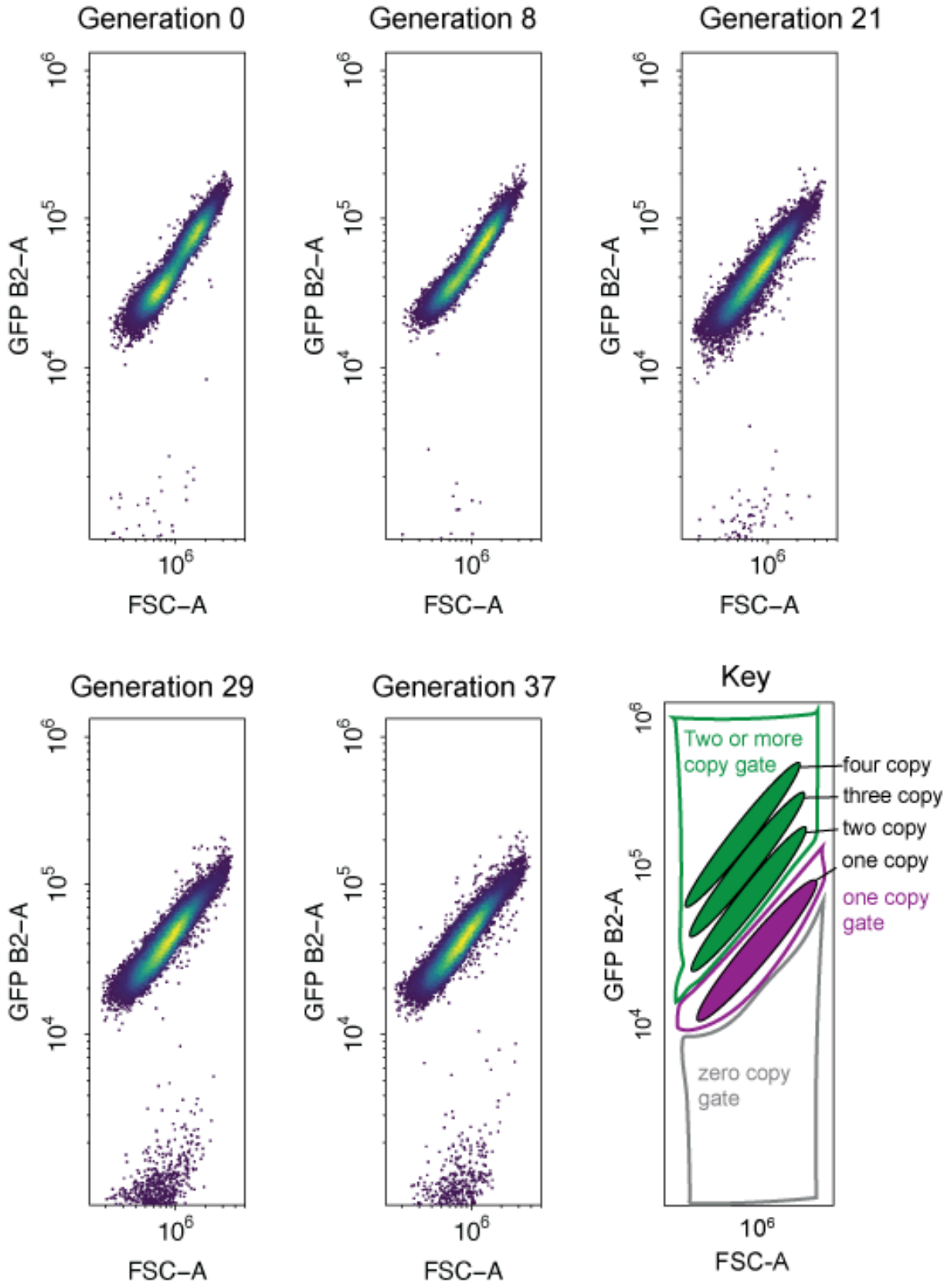
ARSΔ population 3



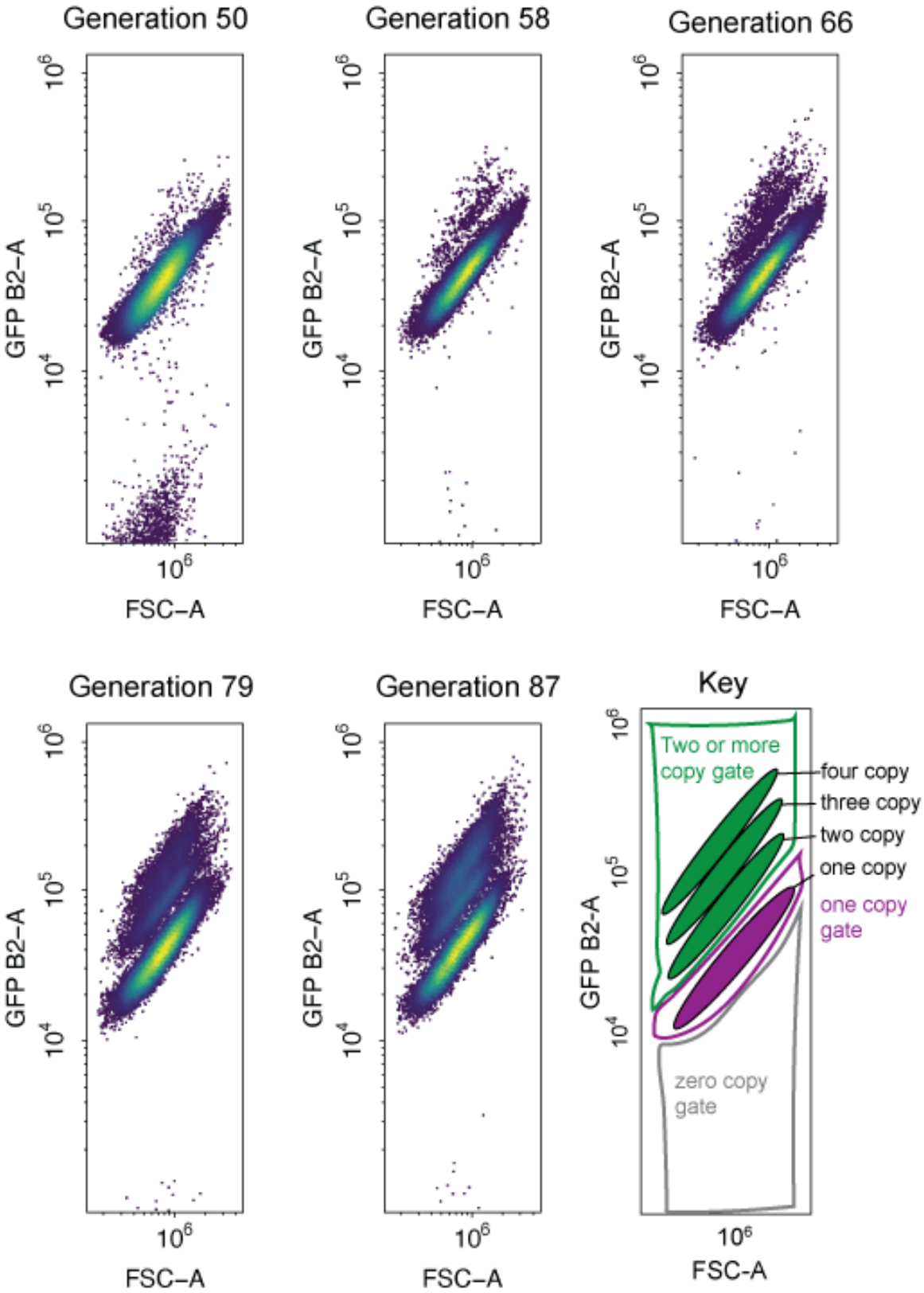
ARSΔ population 3



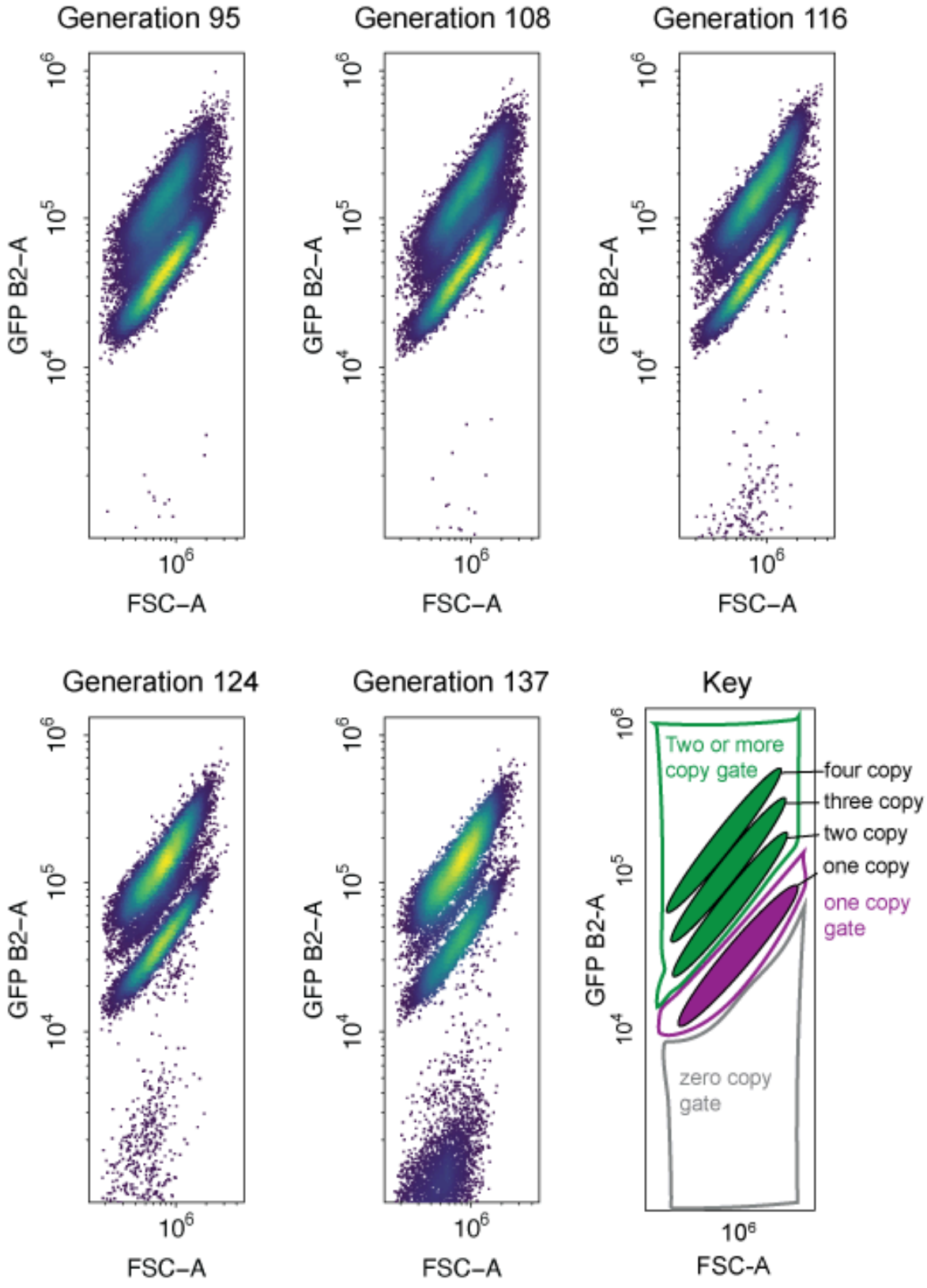
ARSΔ population 4



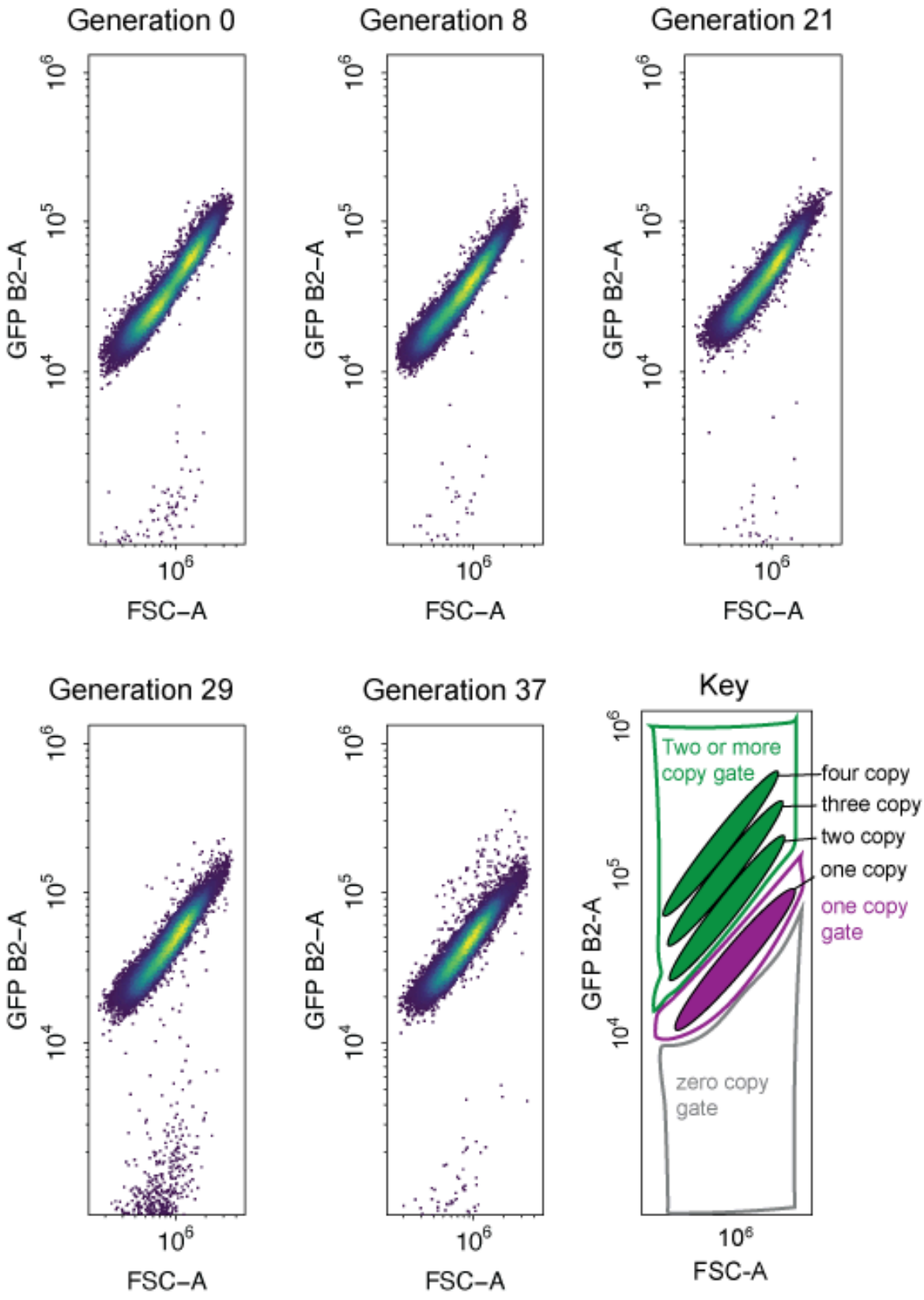
ARSΔ population 4



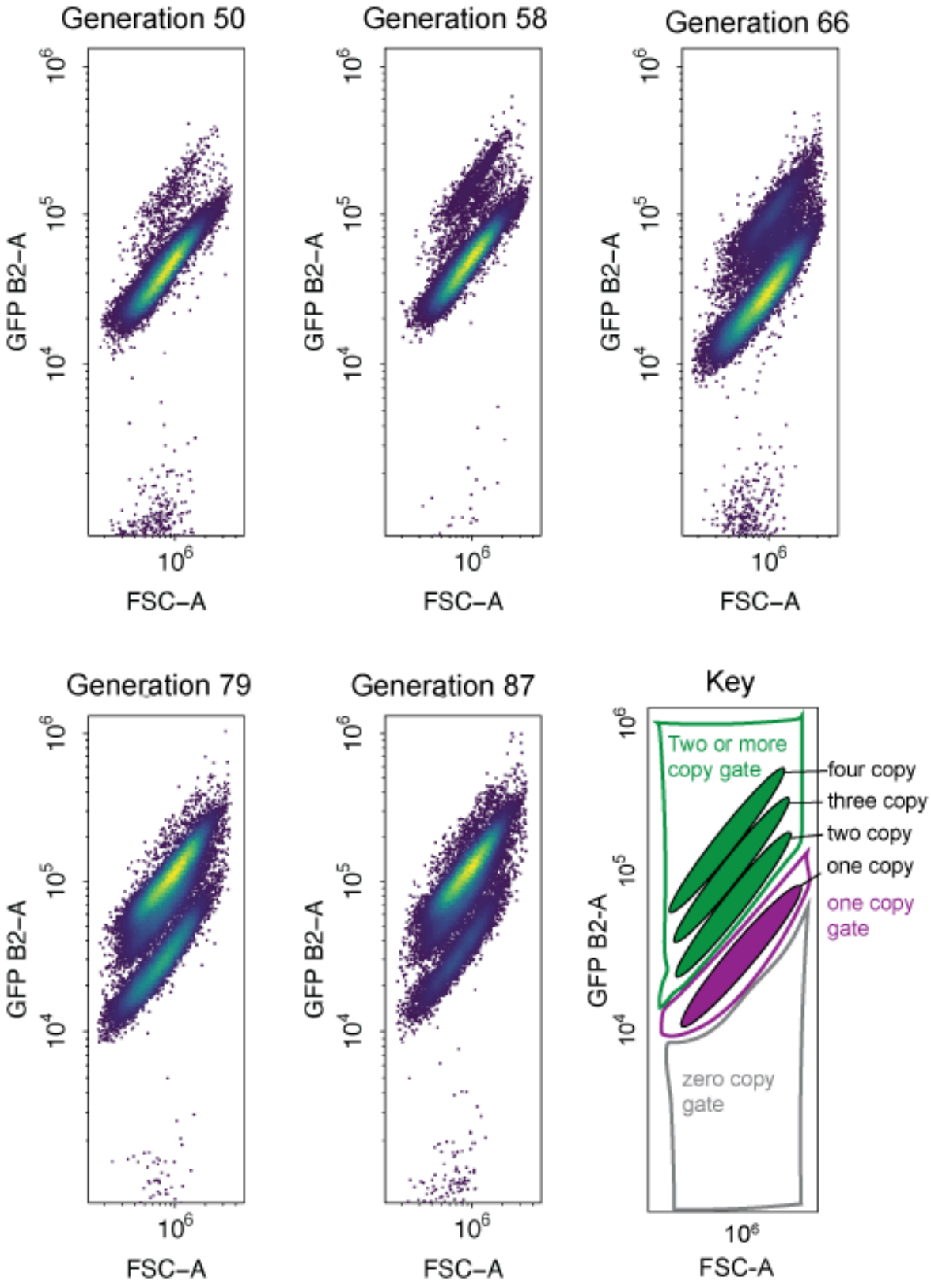
ARSΔ population 4



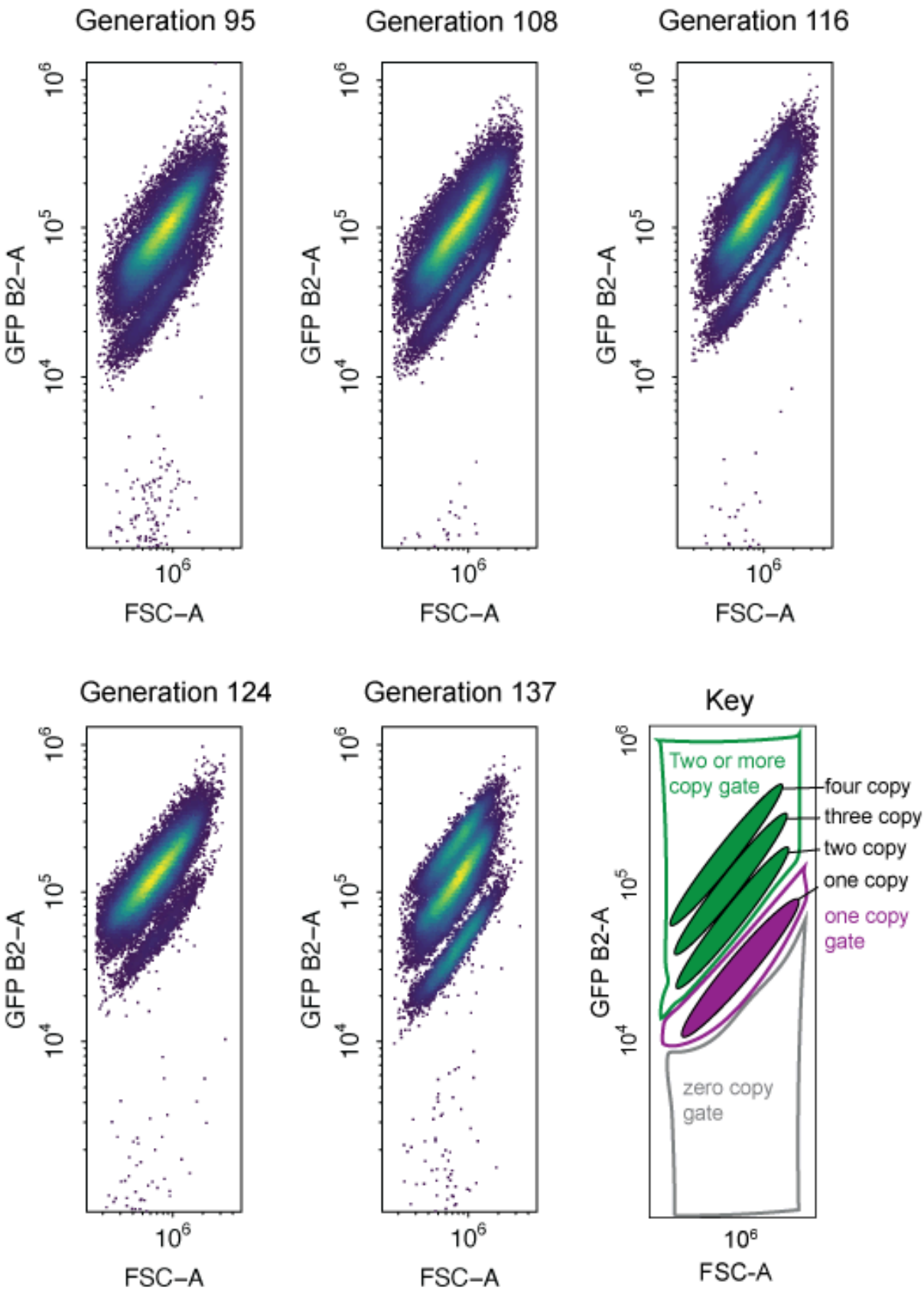
ARSΔ population 5



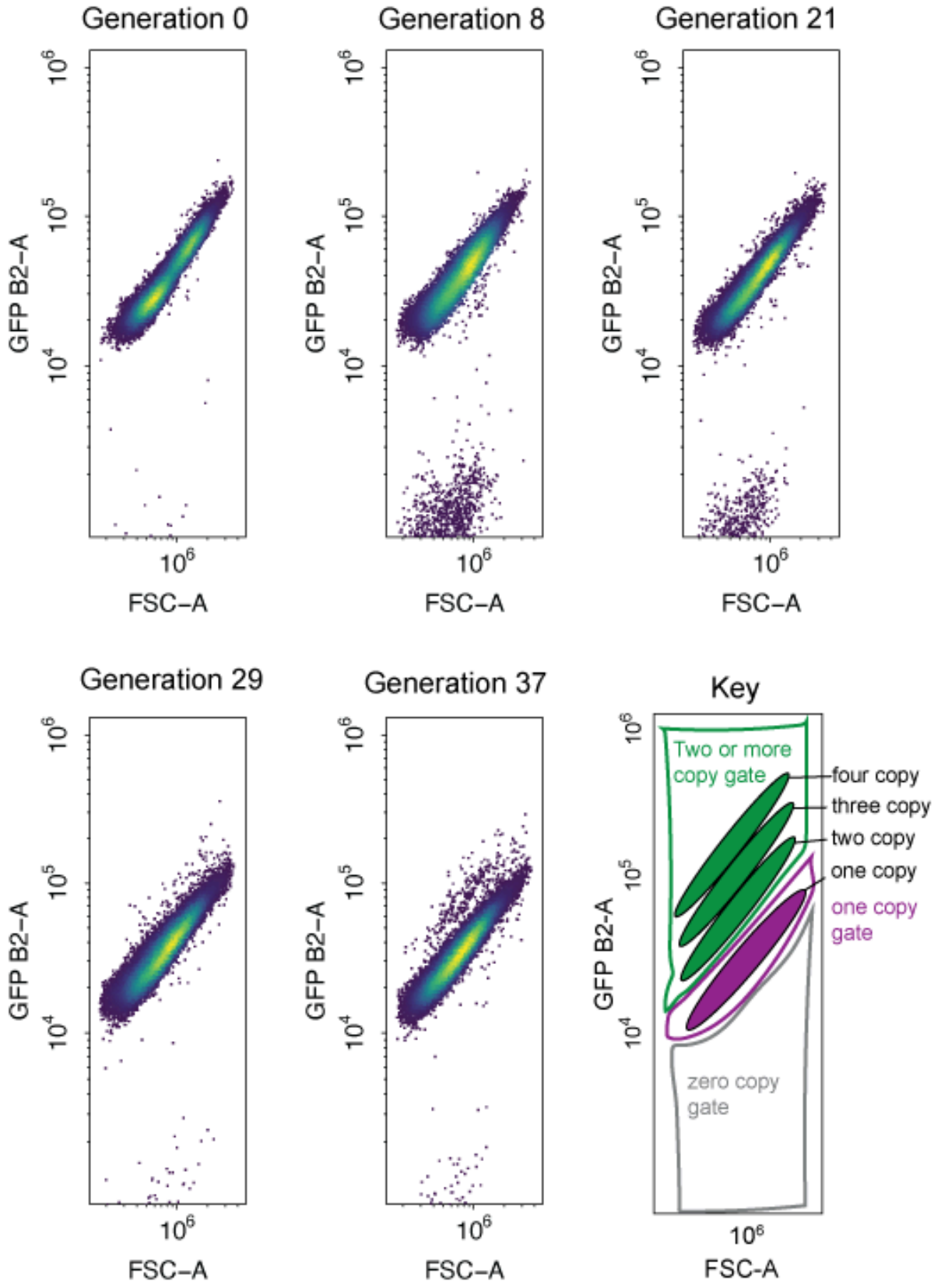
ARS∆ population 5



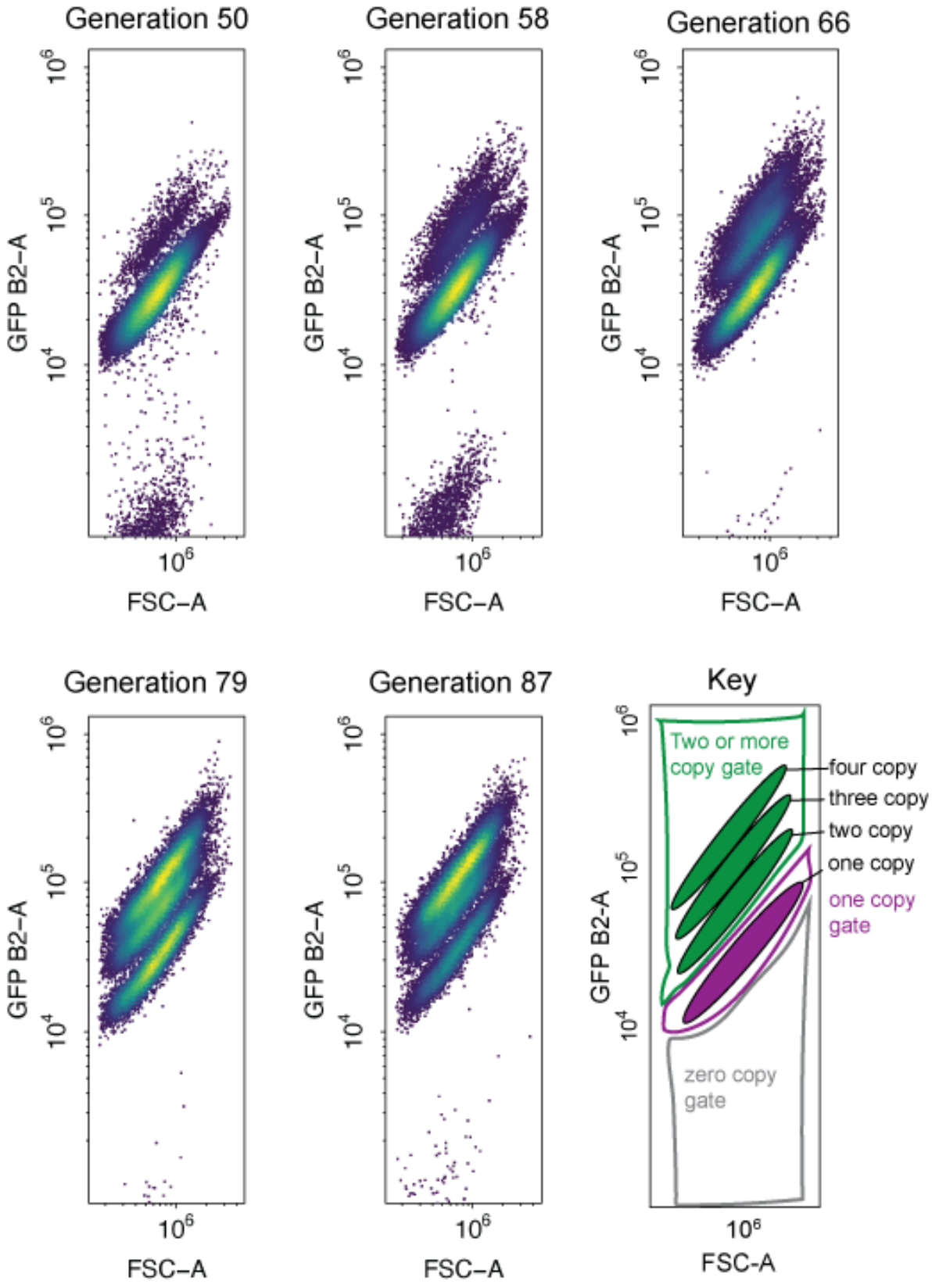
ARSΔ population 5

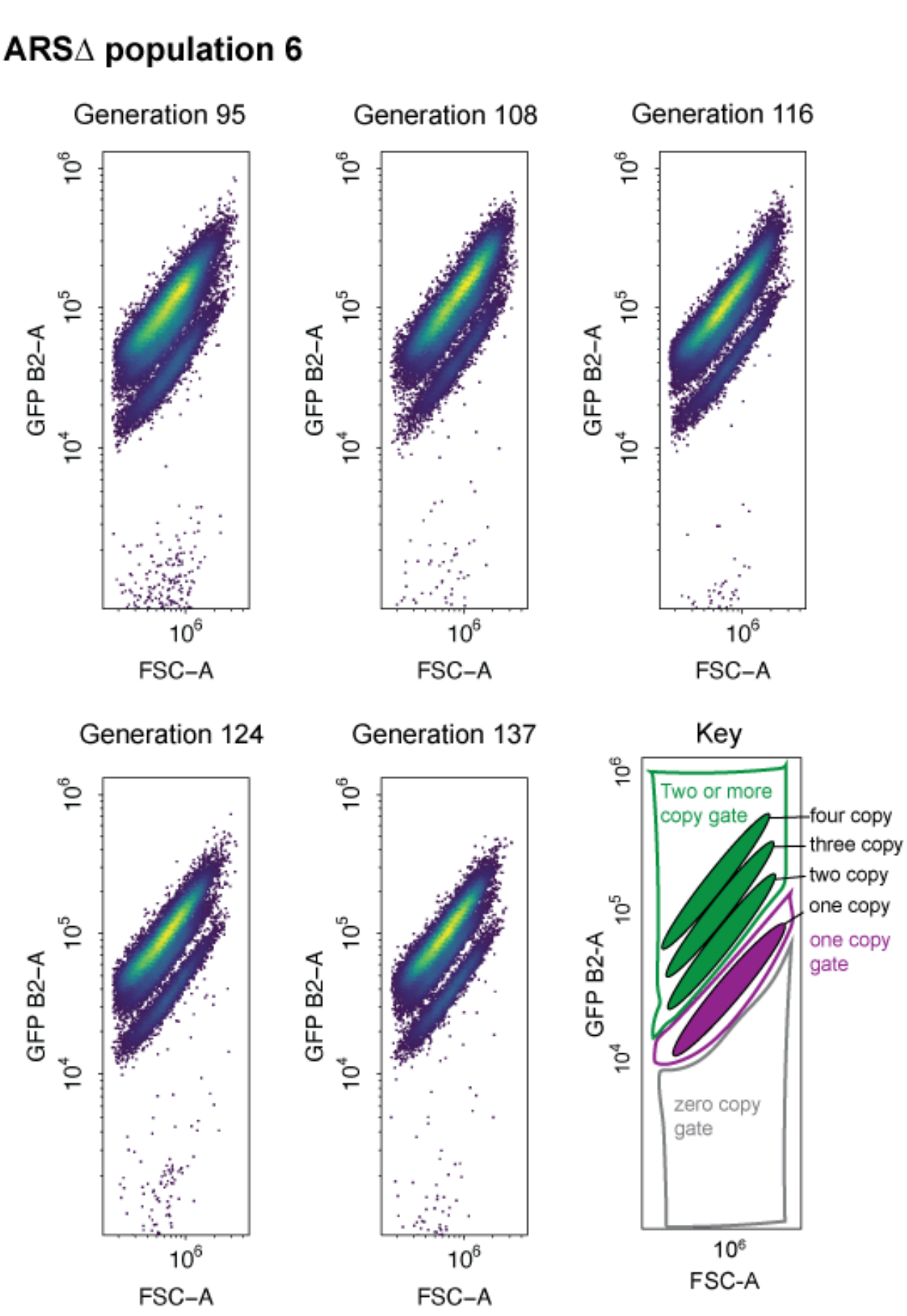


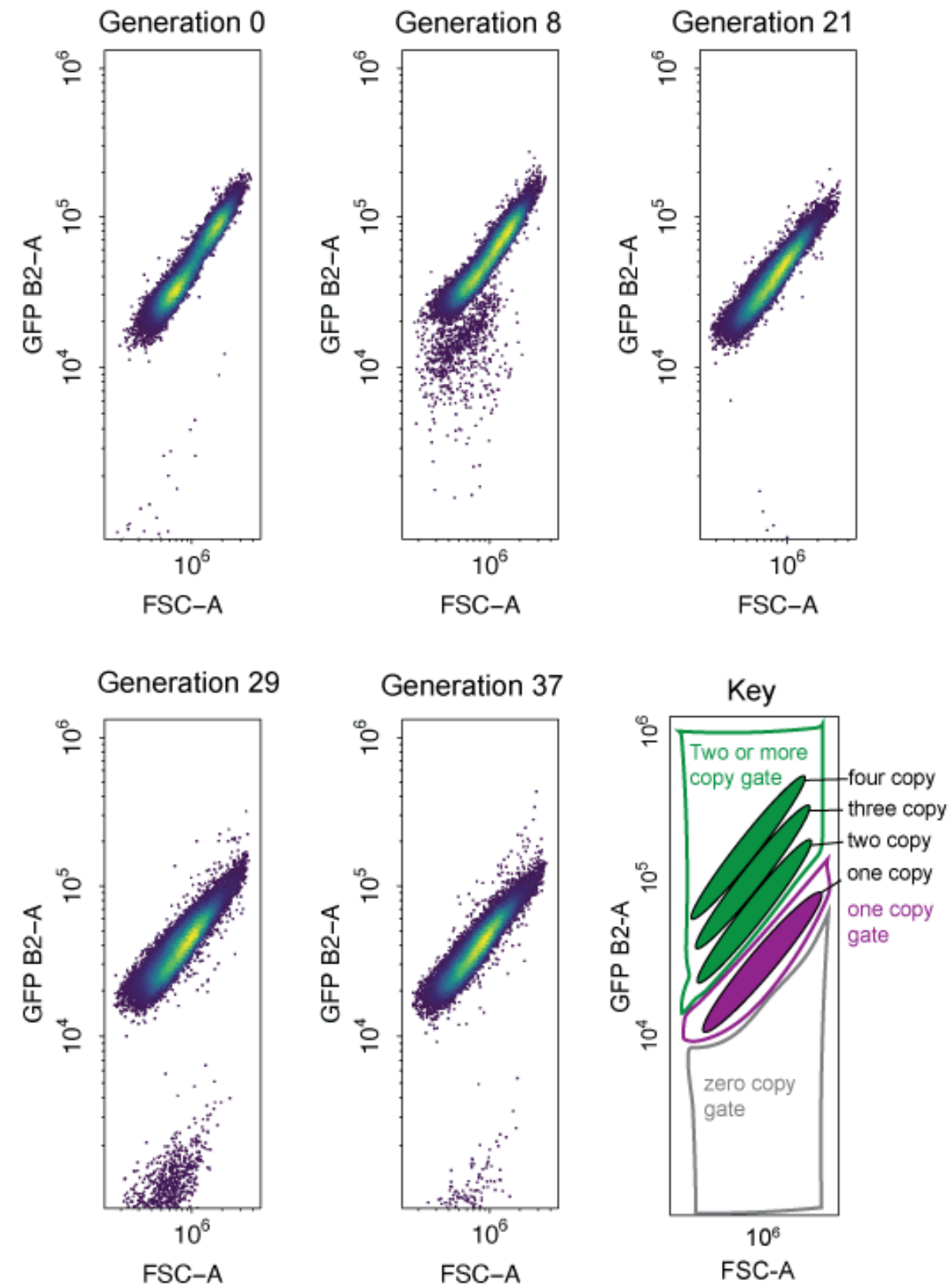
ARSΔ population 6



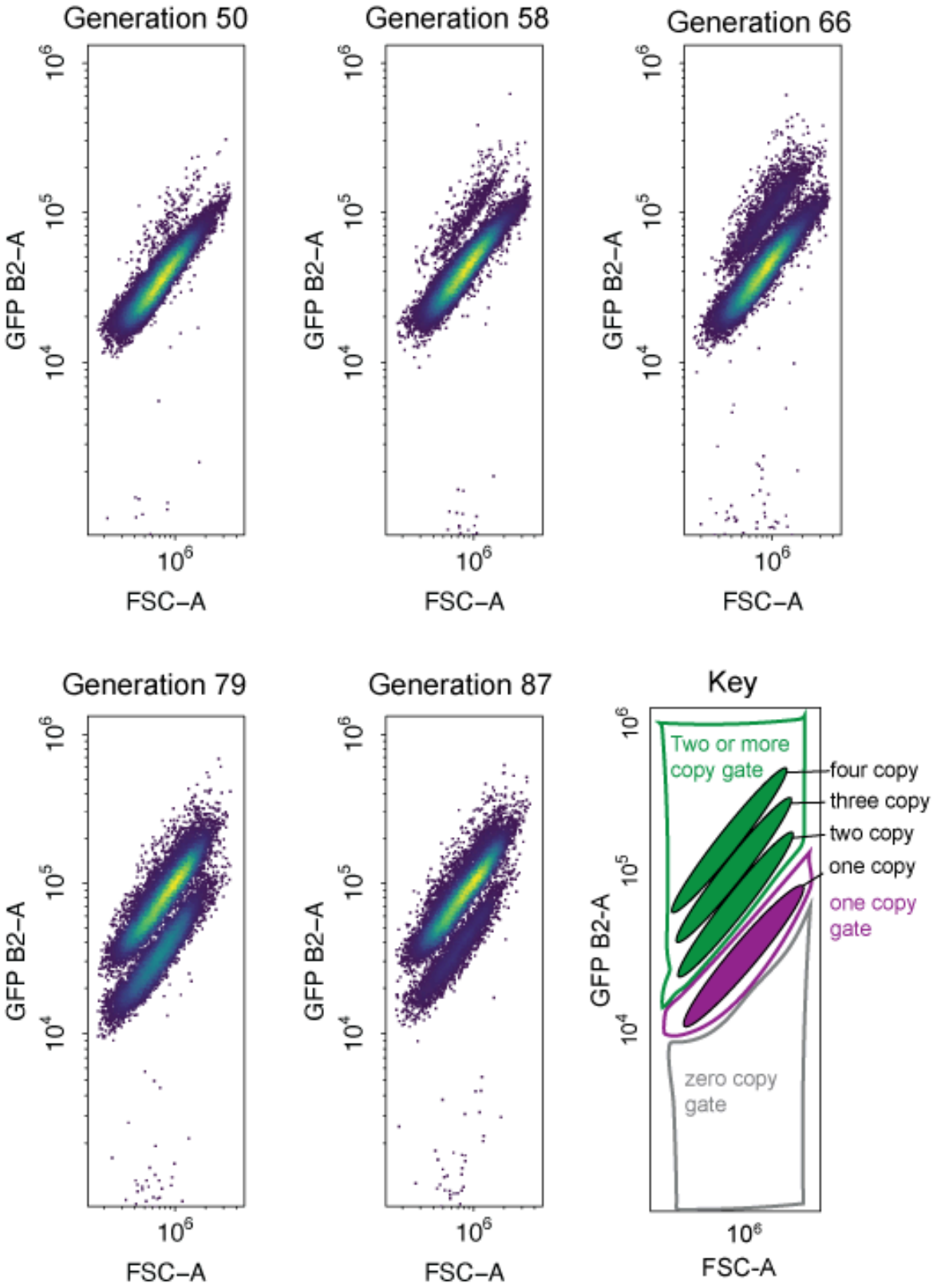
ARSΔ population 6



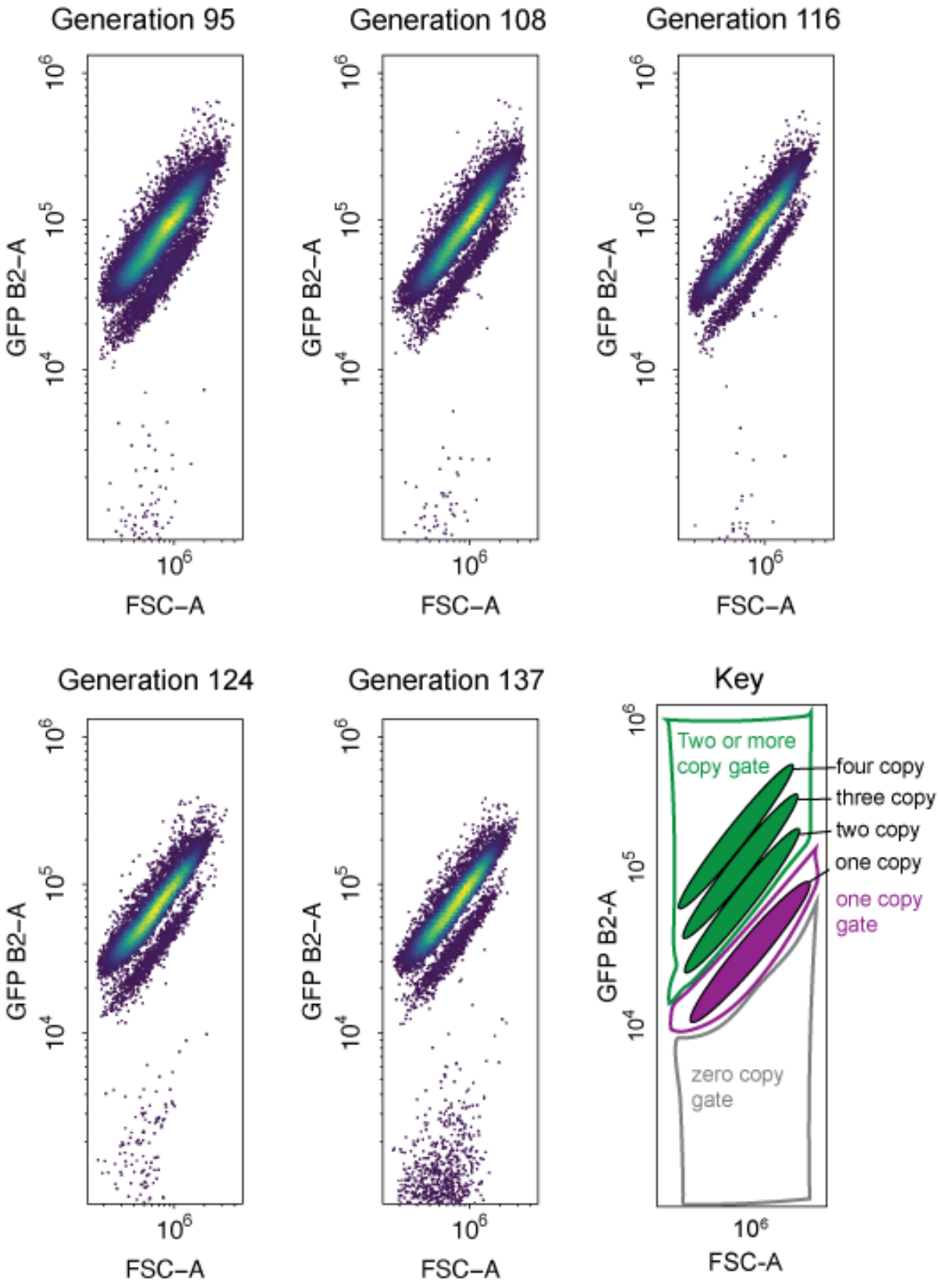


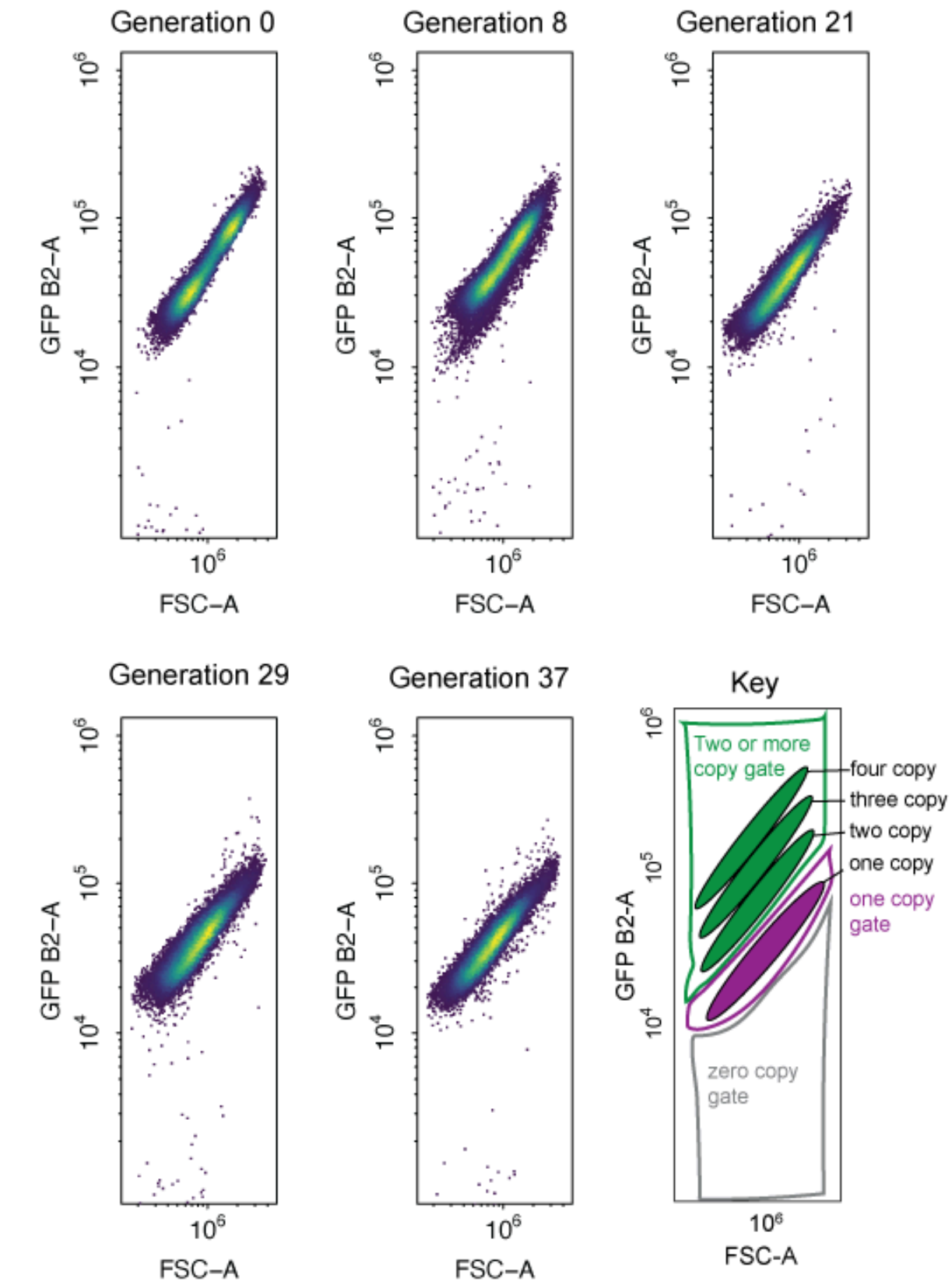


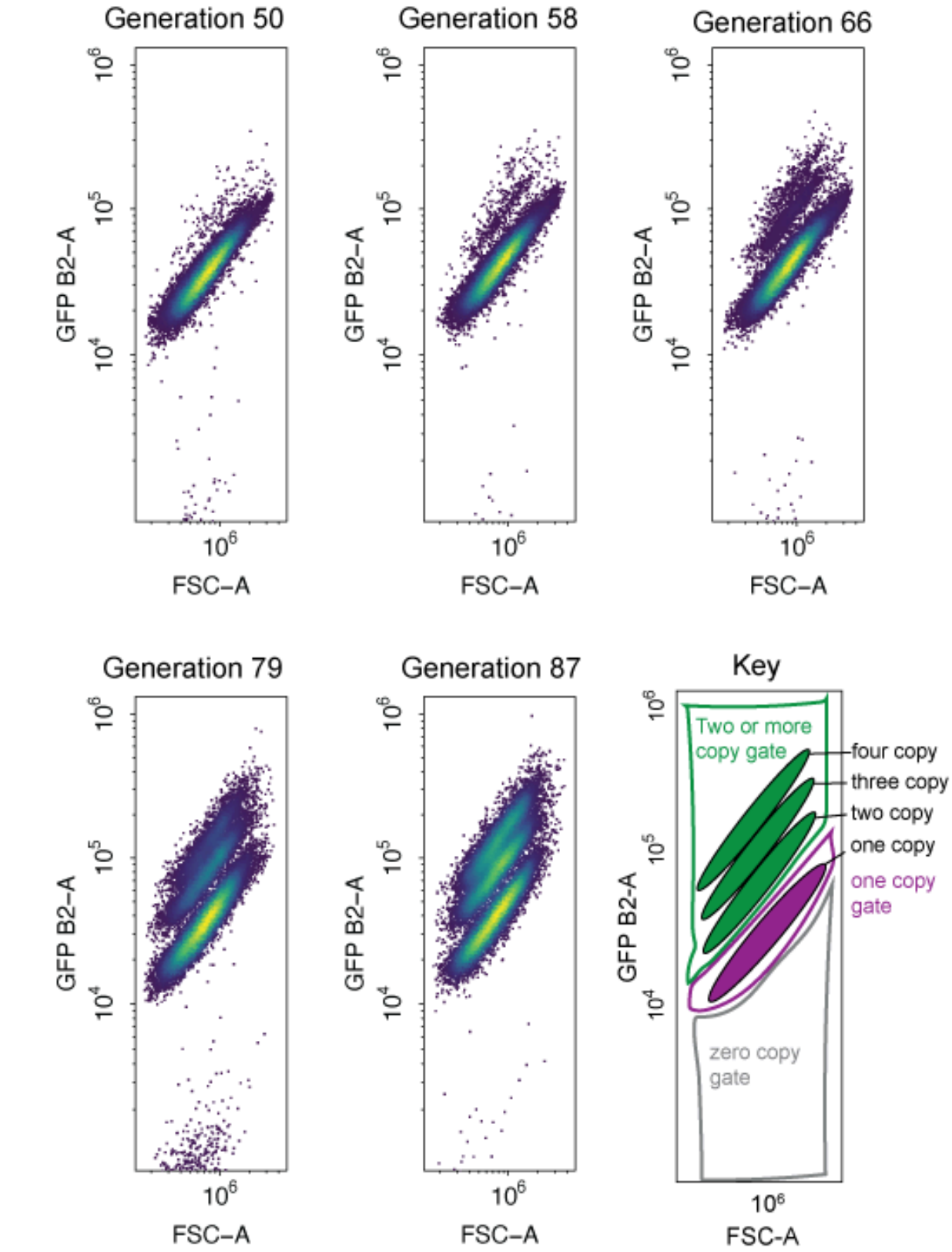
ARSΔ population 7



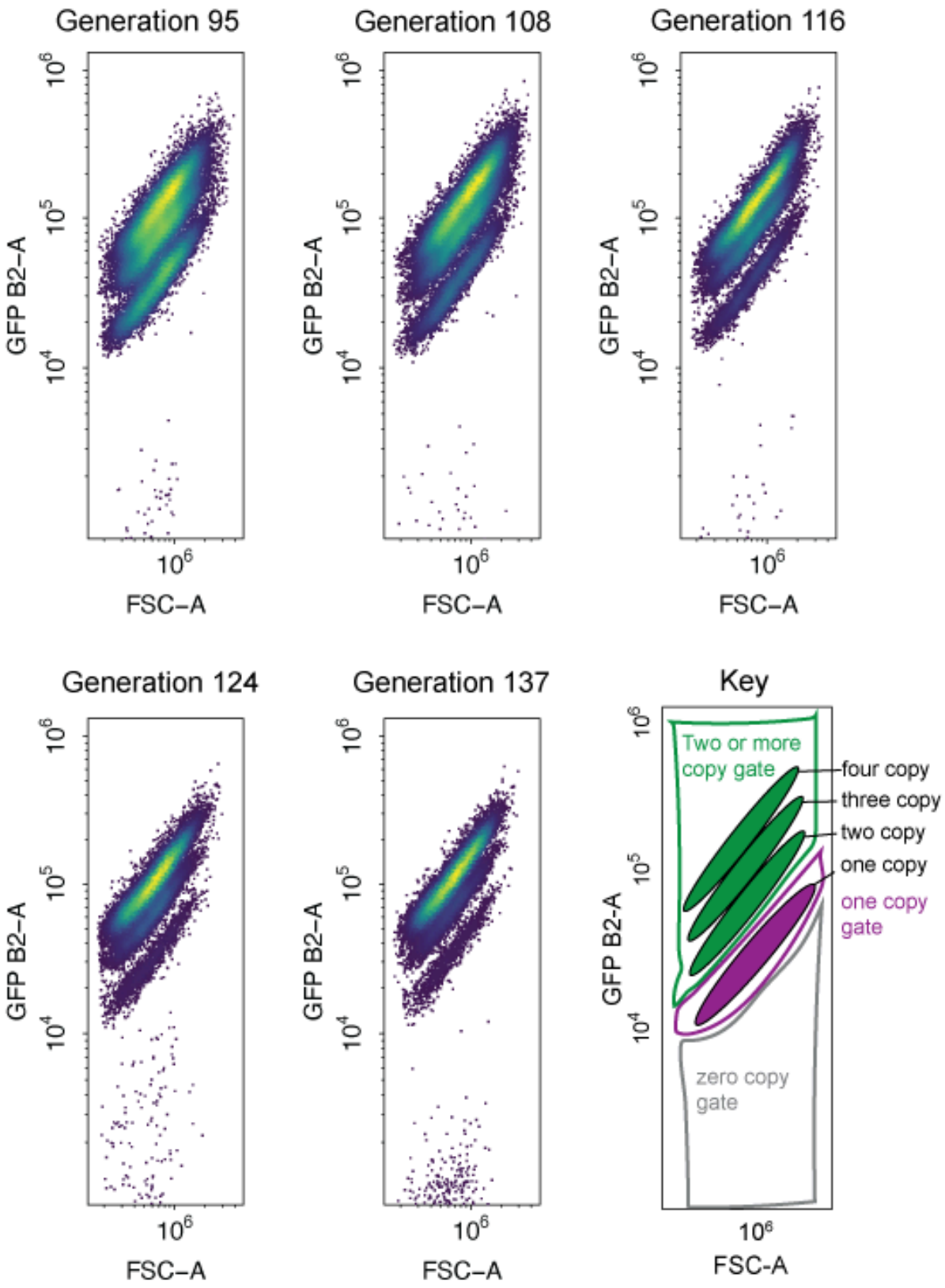
ARSΔ population 7



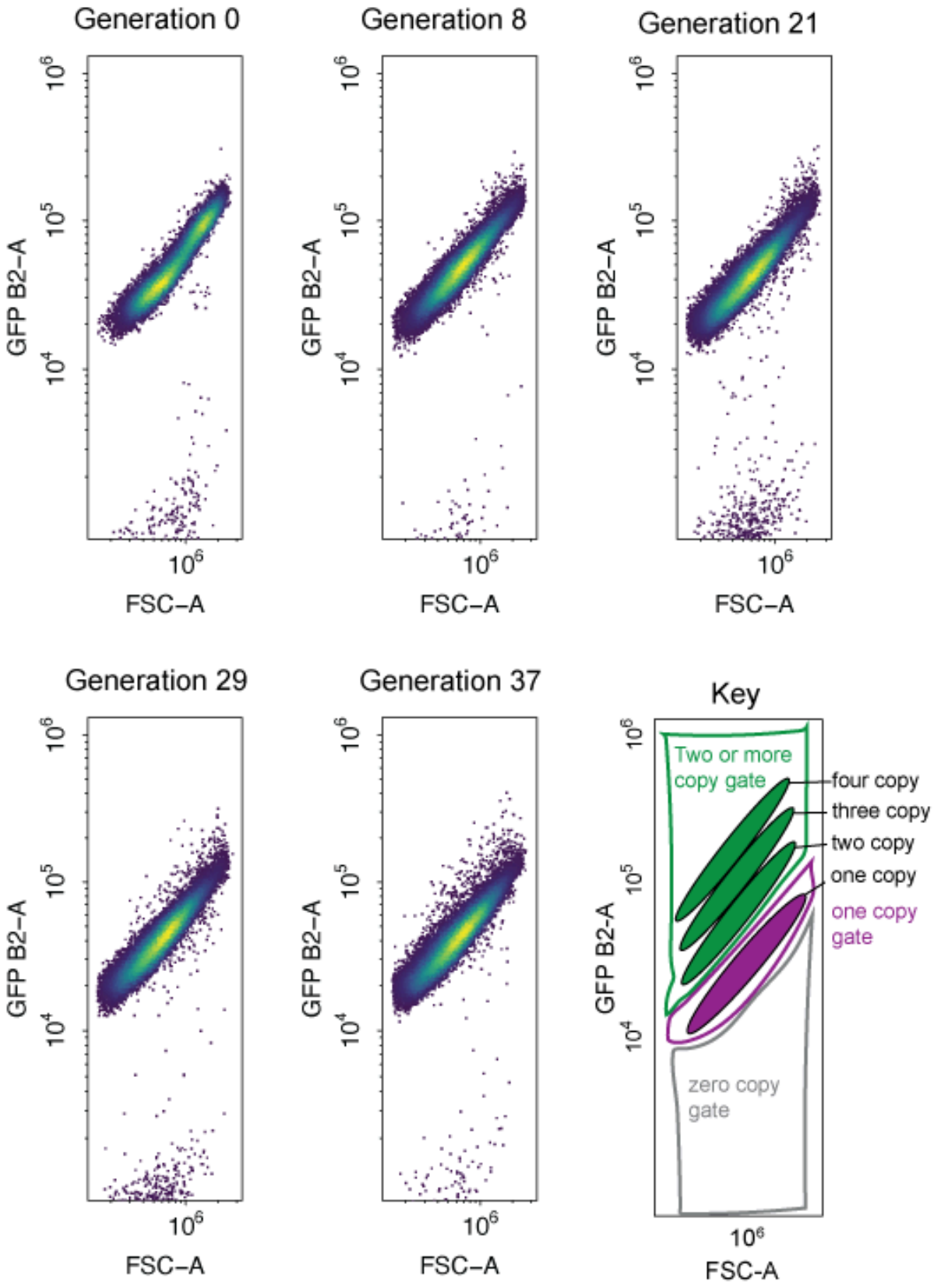




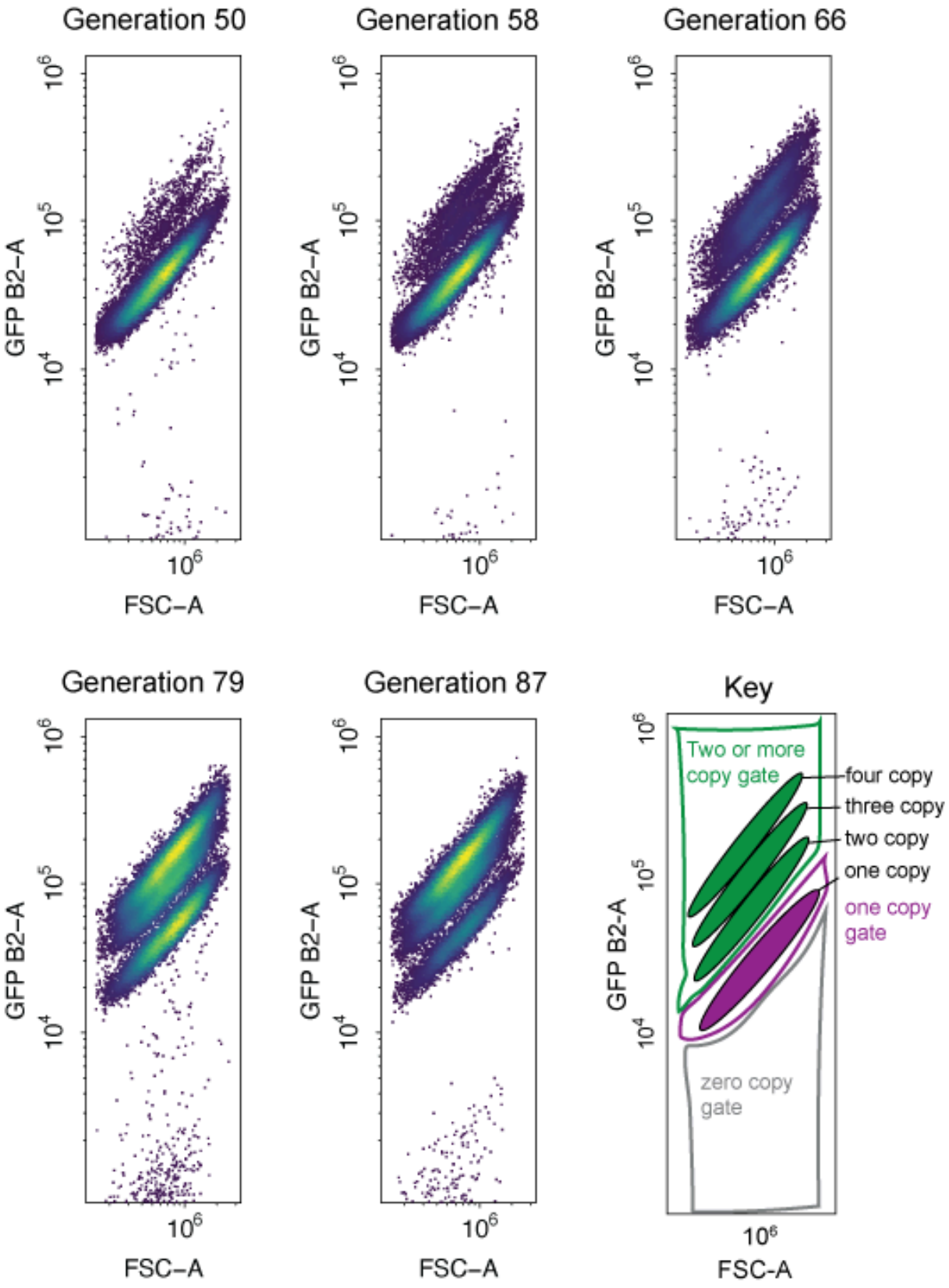
ARSΔ population 8



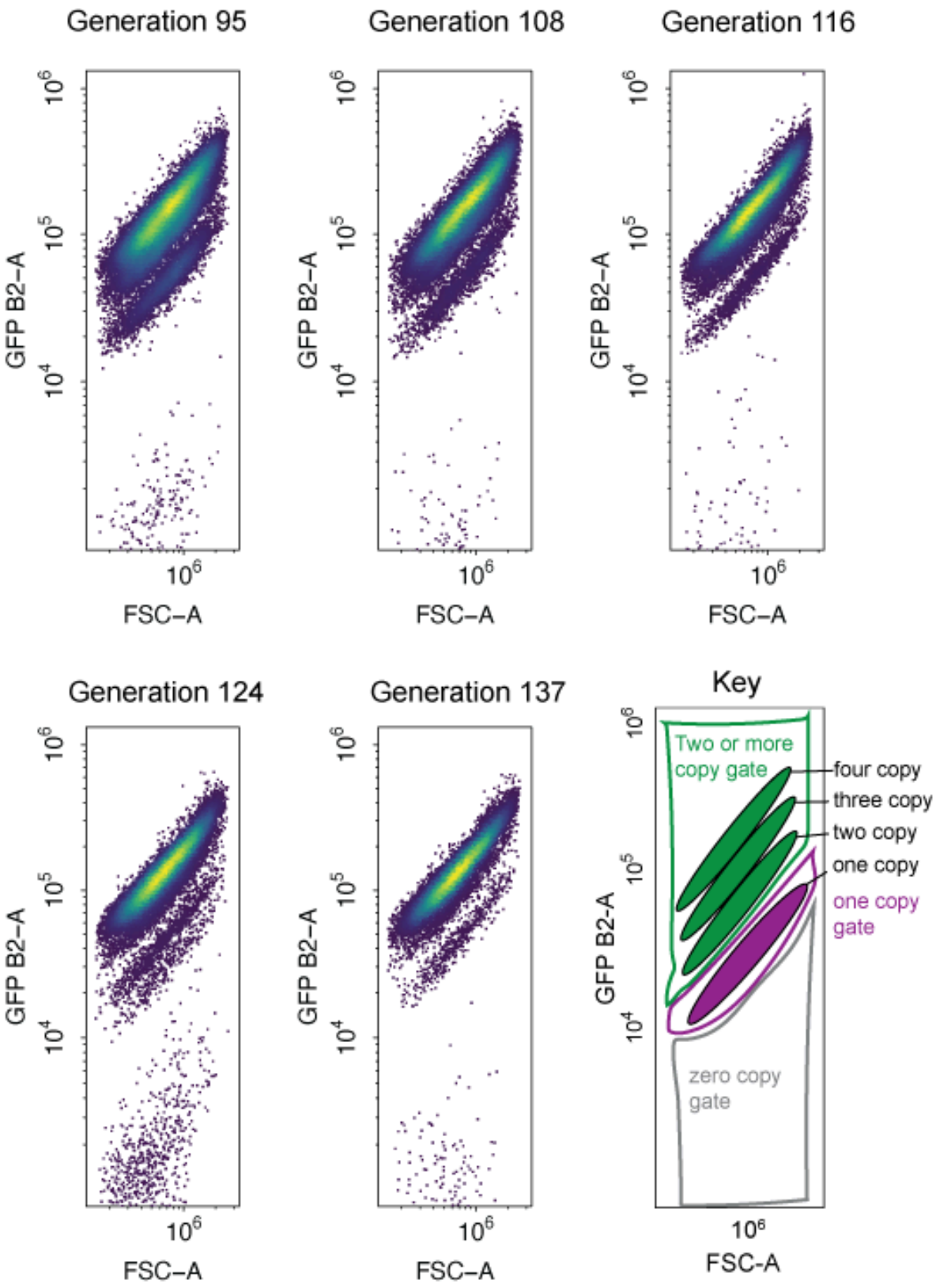
ALLΔ population 1



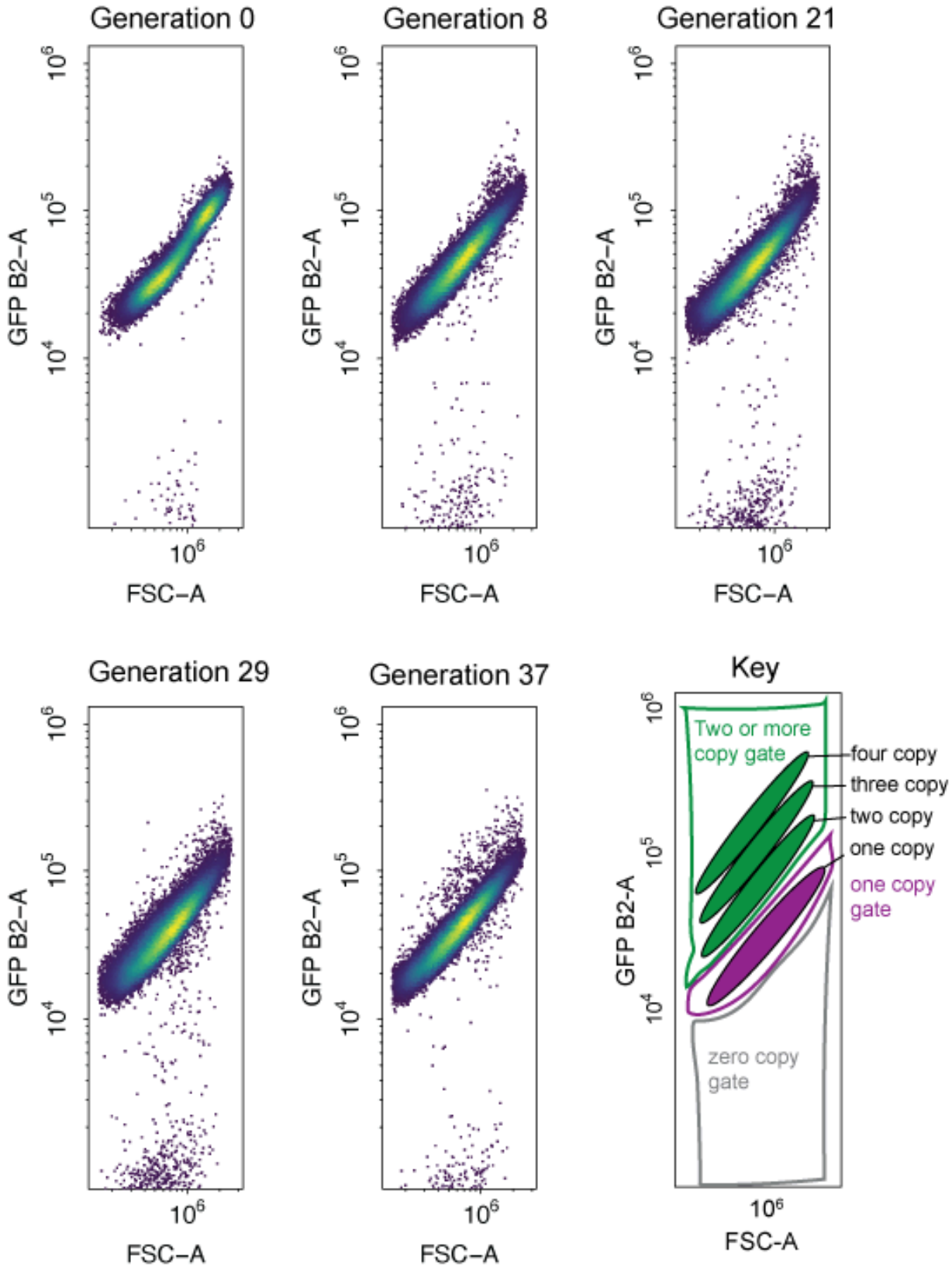
ALLΔ population 1



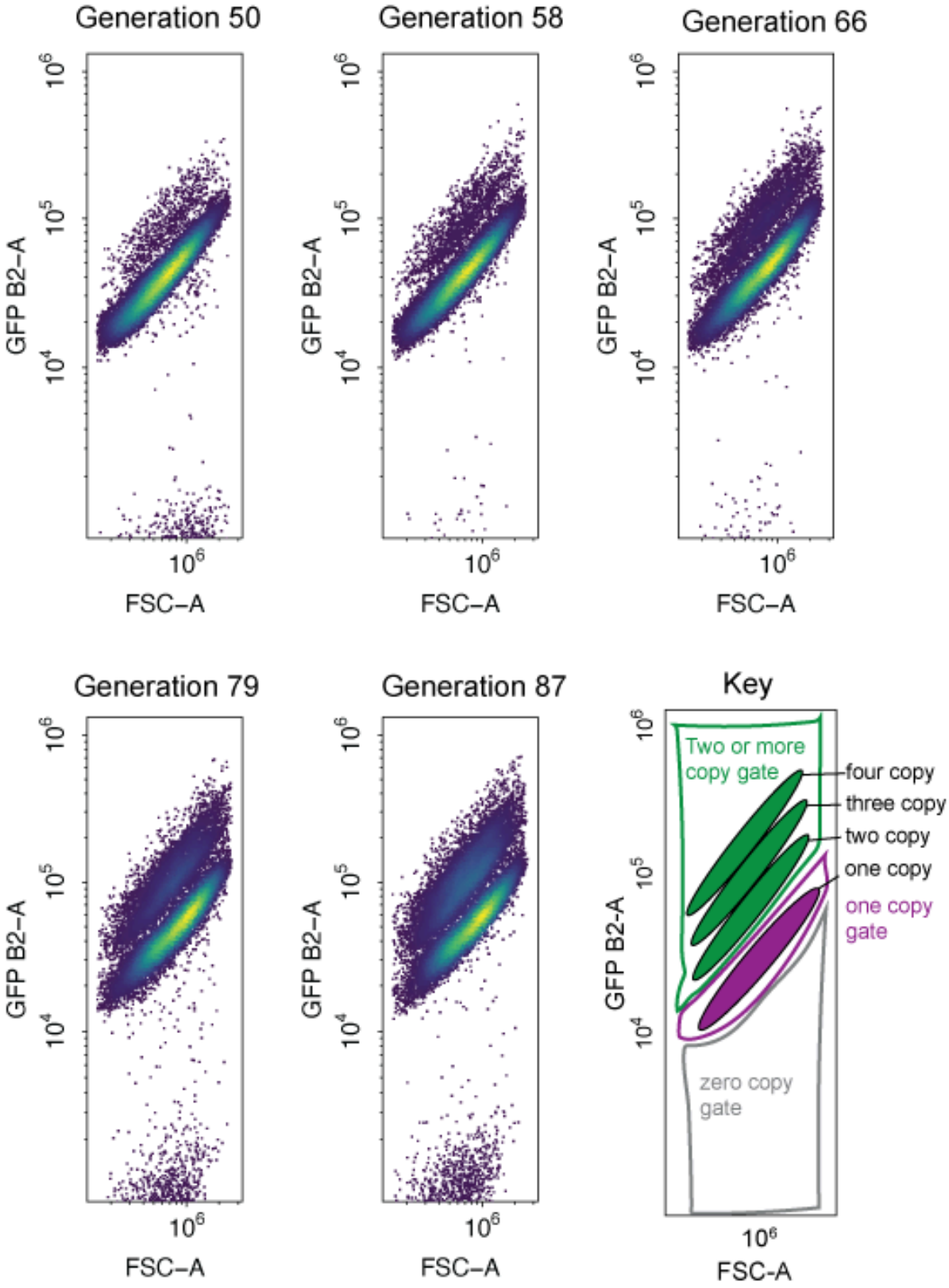
ALLΔ population 1



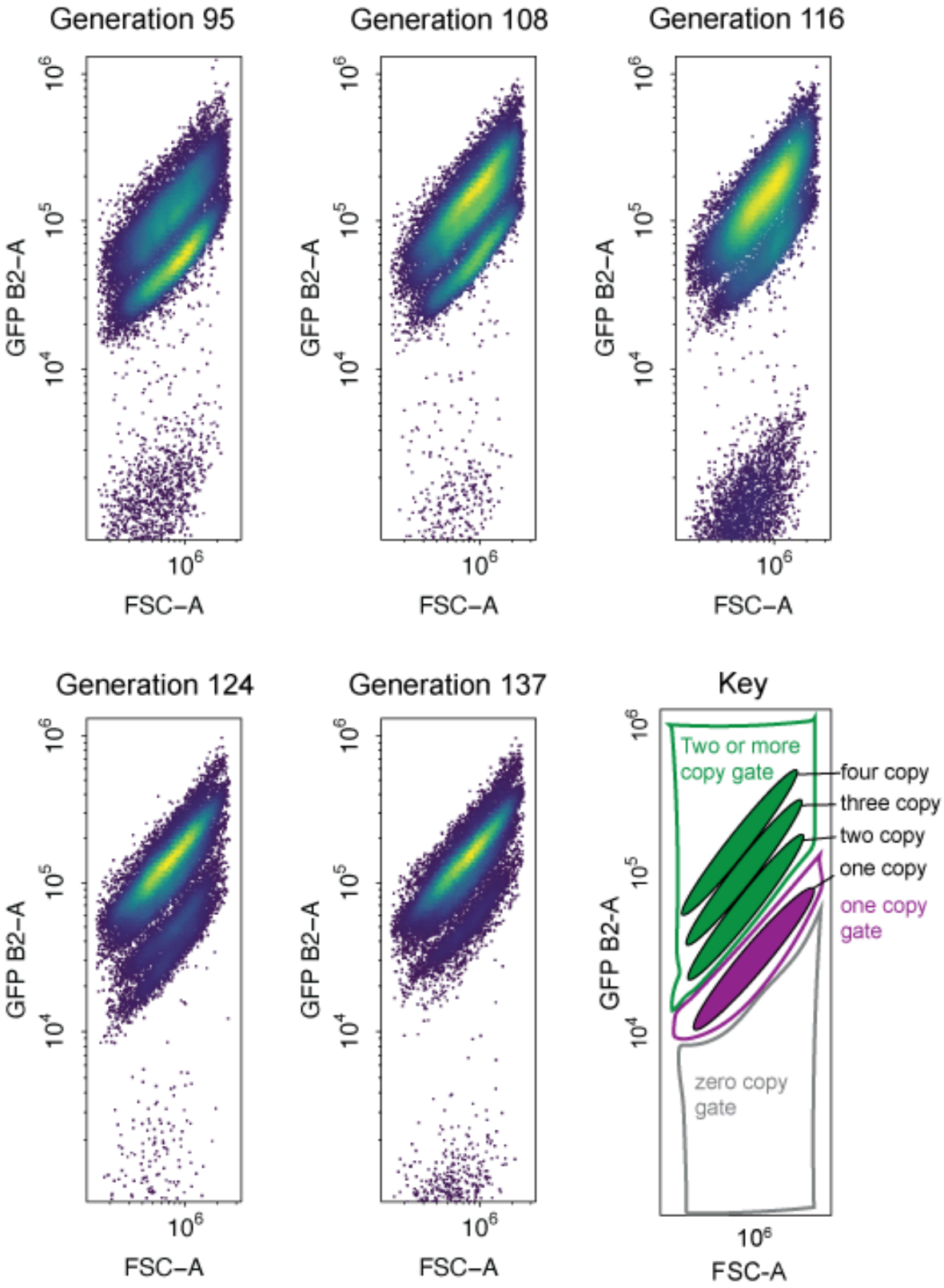
ALLΔ population 2

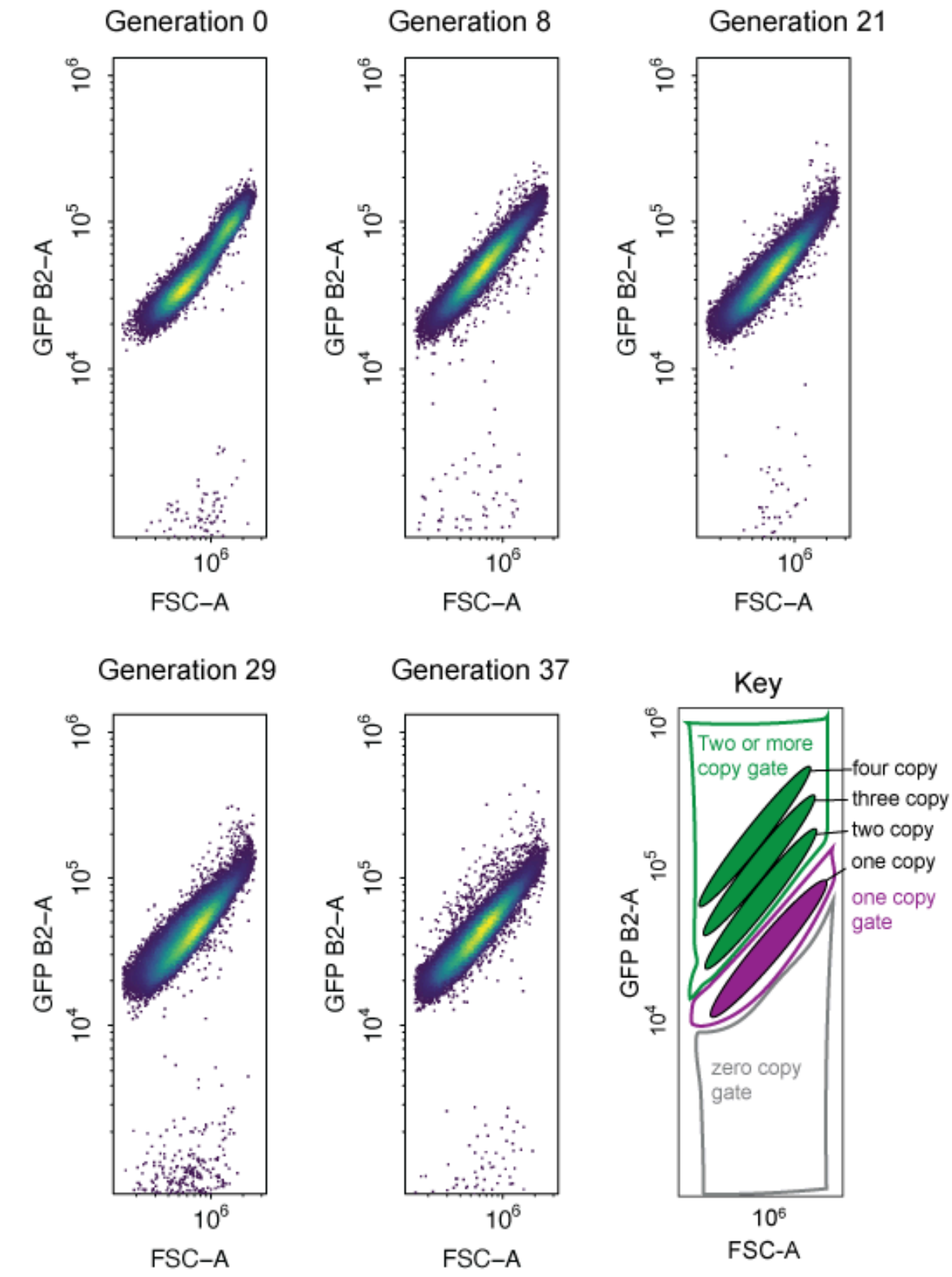


ALLΔ population 2

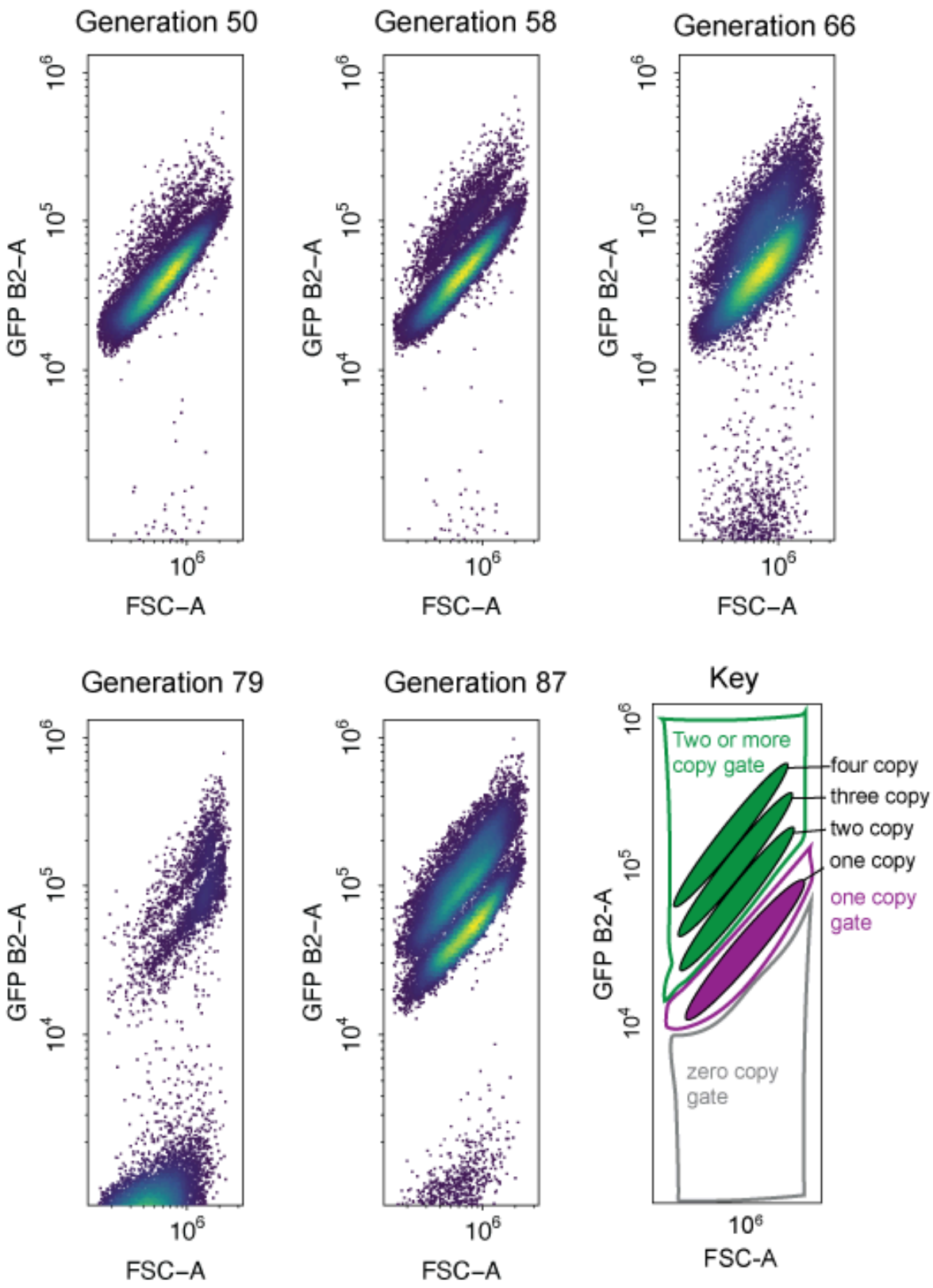


ALLΔ population 2

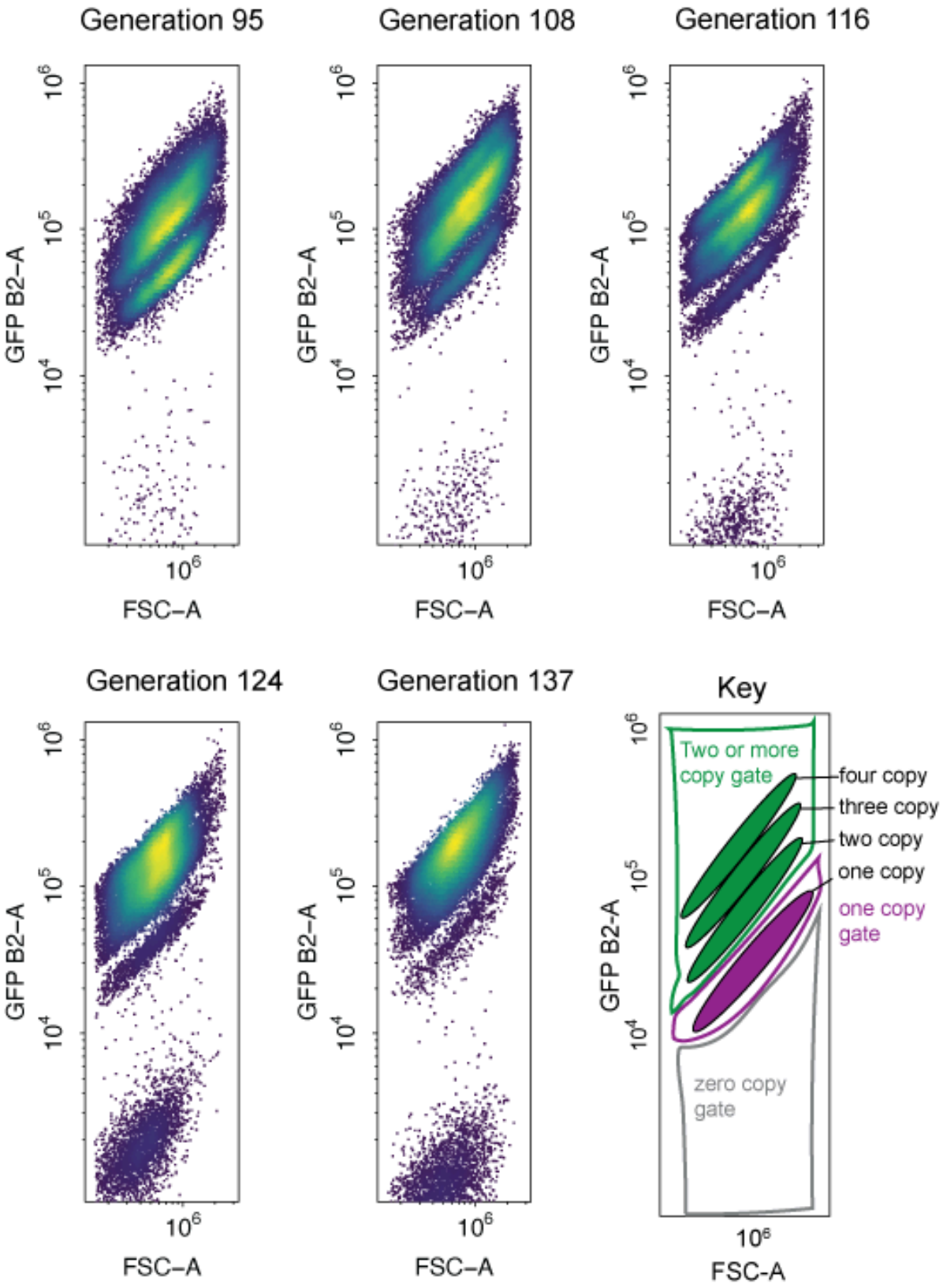




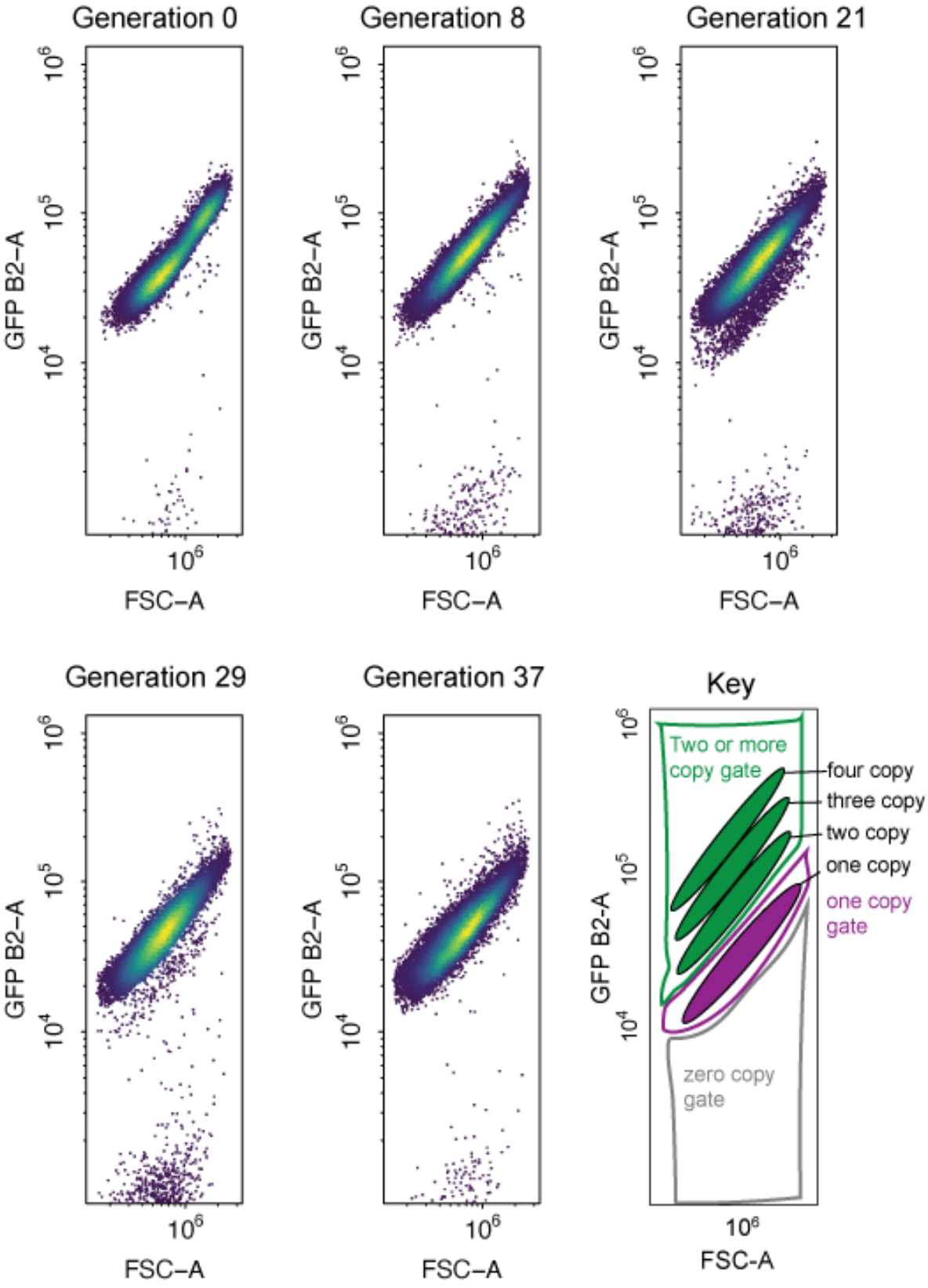
ALLΔ population 3



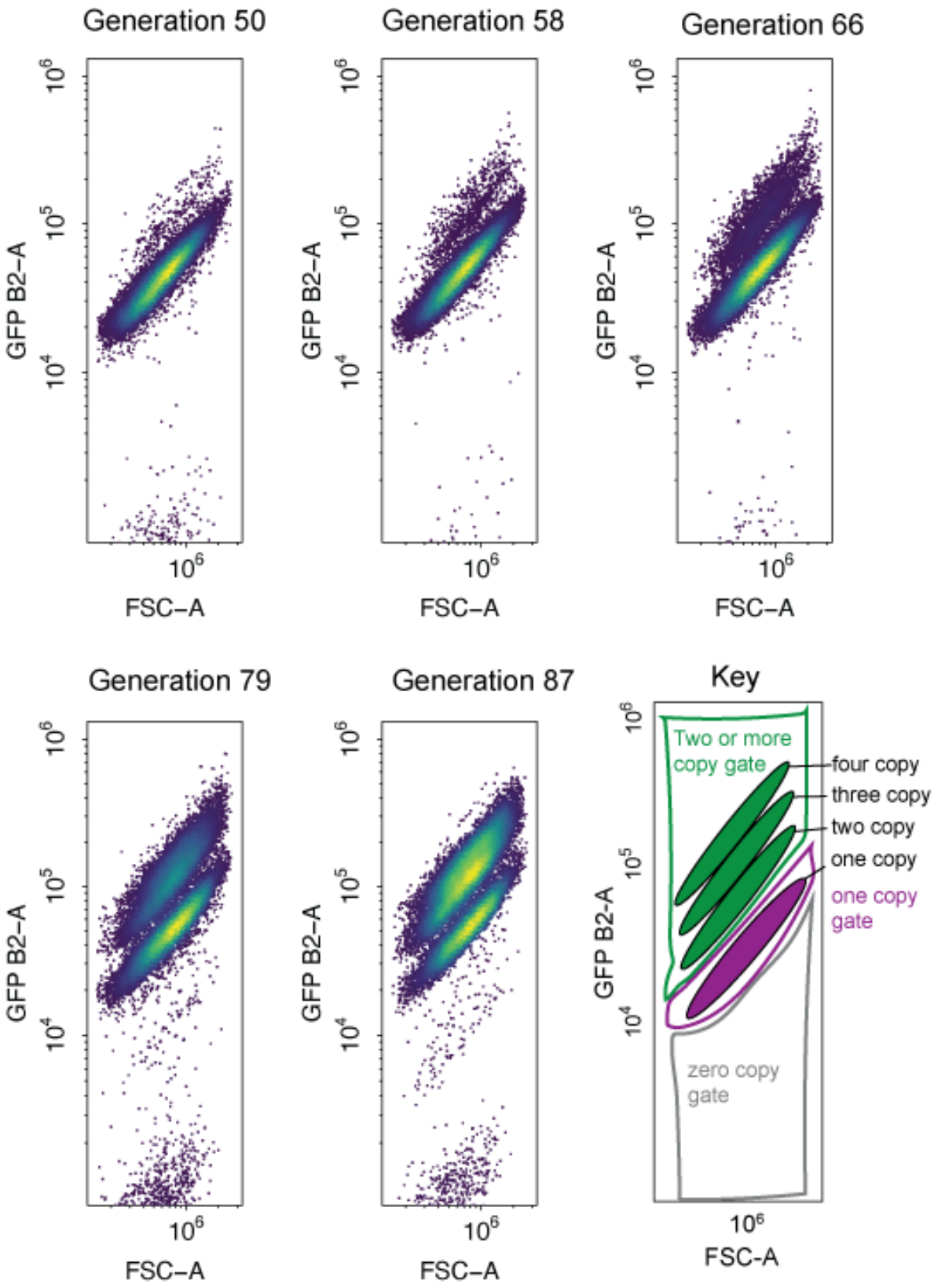
ALLΔ population 3



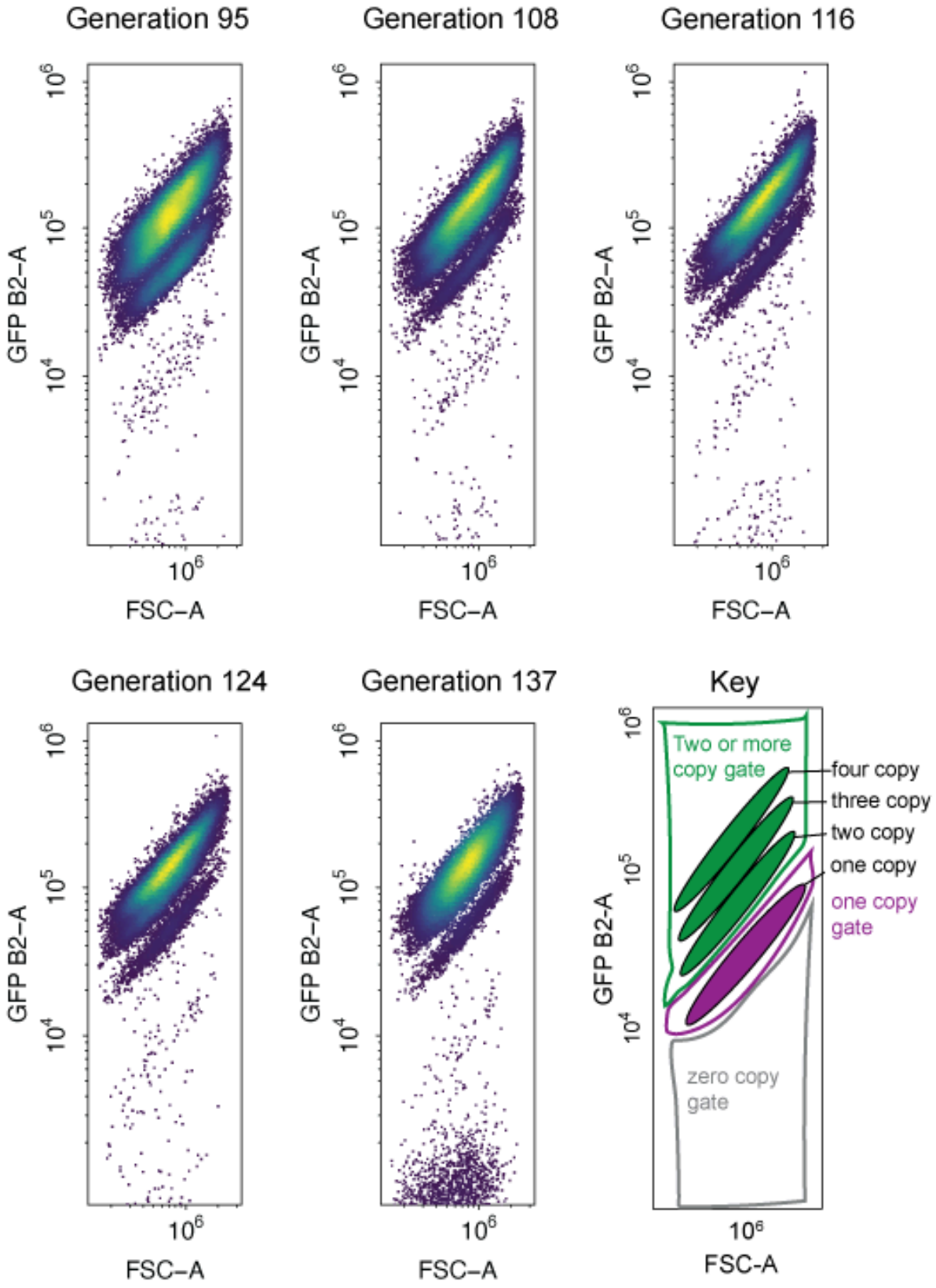
ALLΔ population 4



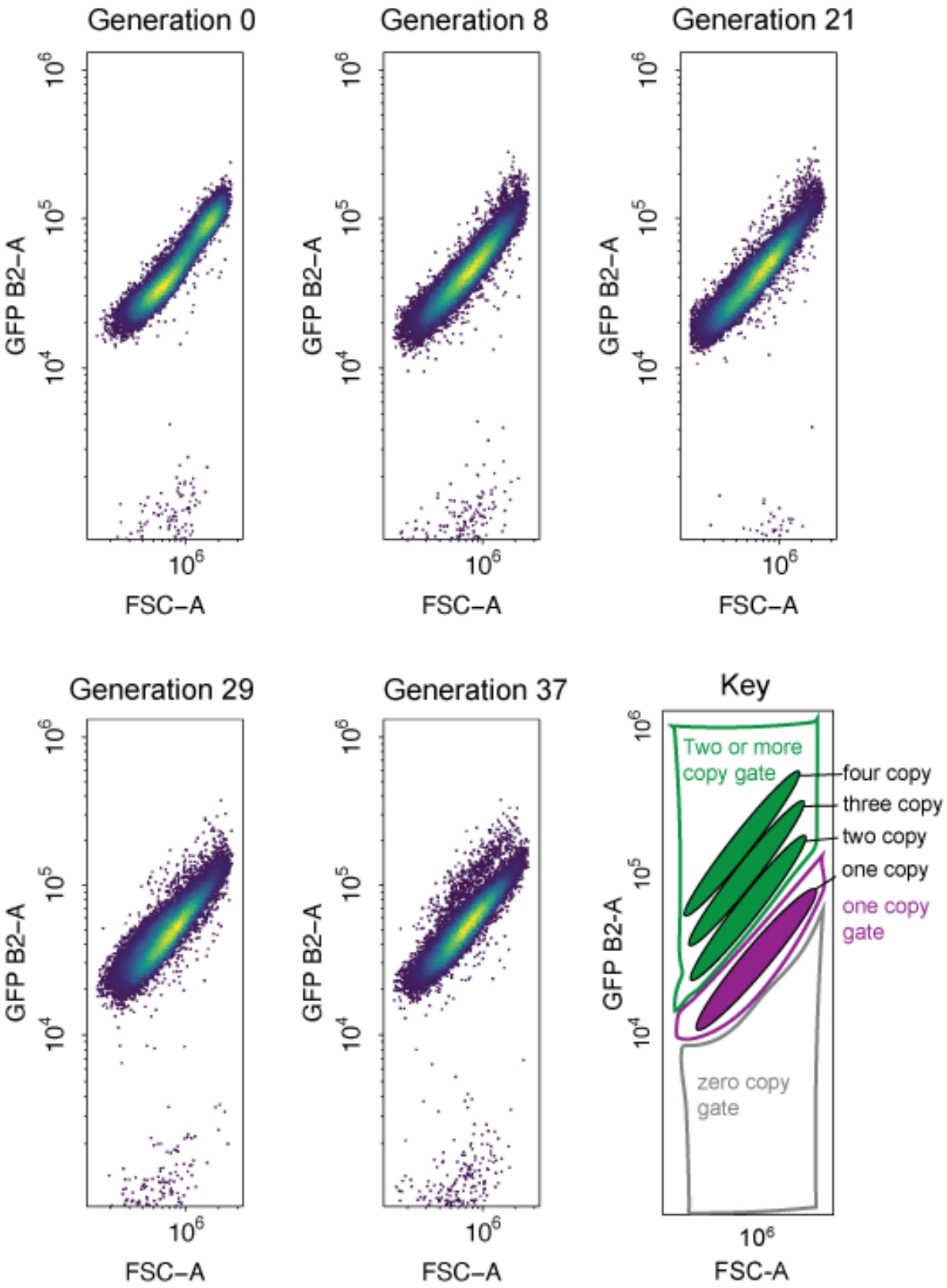
ALLΔ population 4



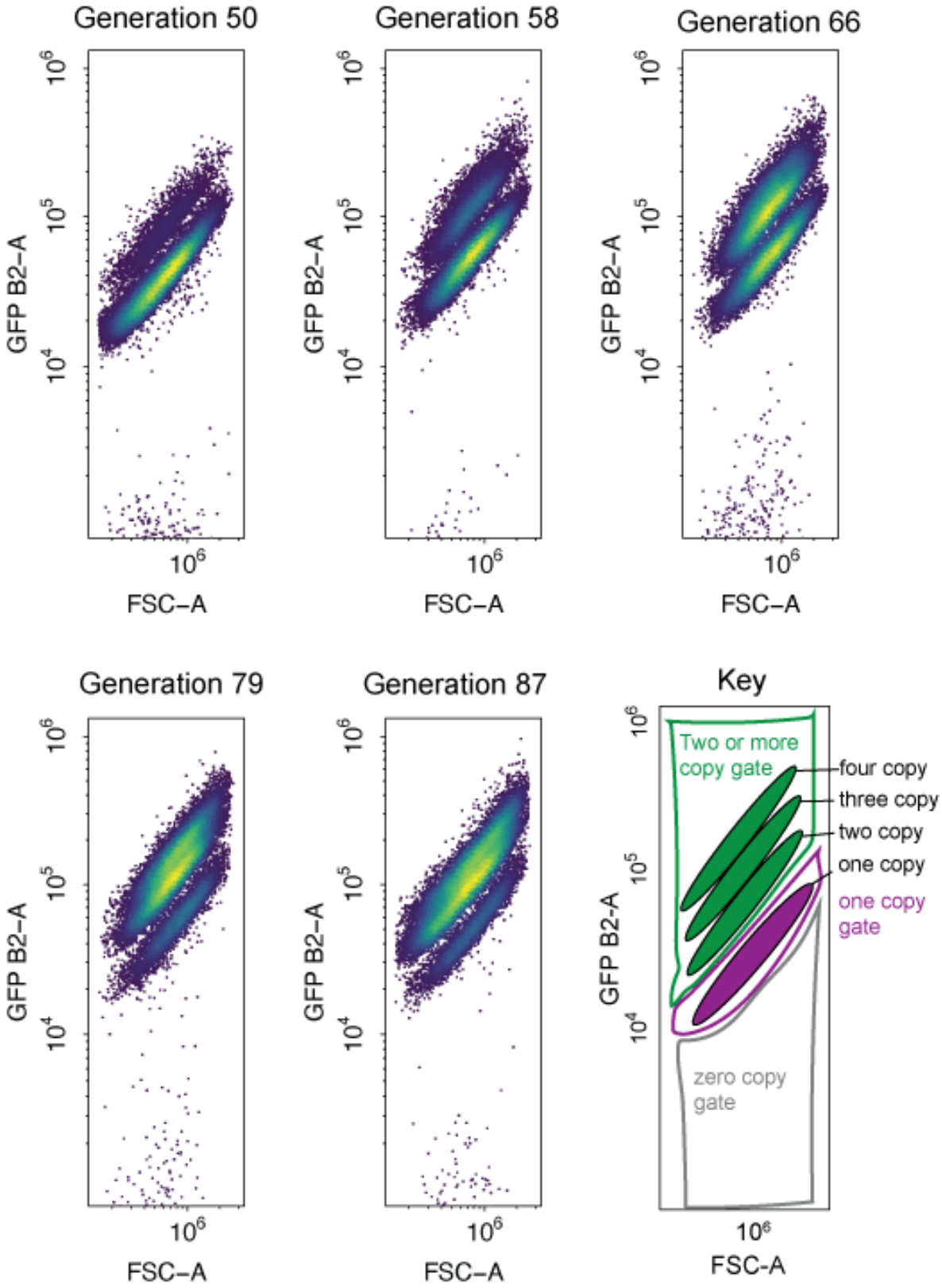
ALLΔ population 4



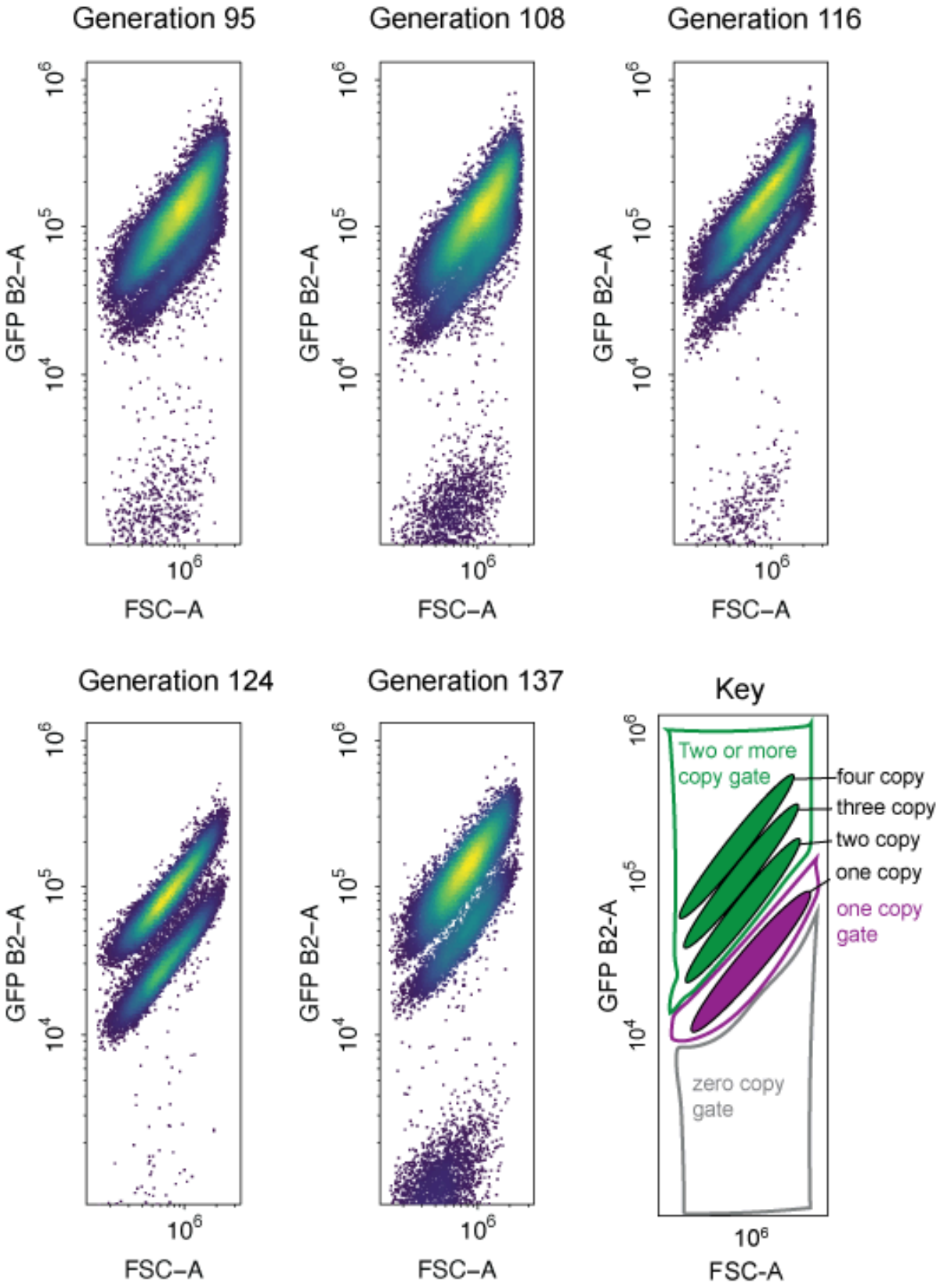
ALLΔ population 5



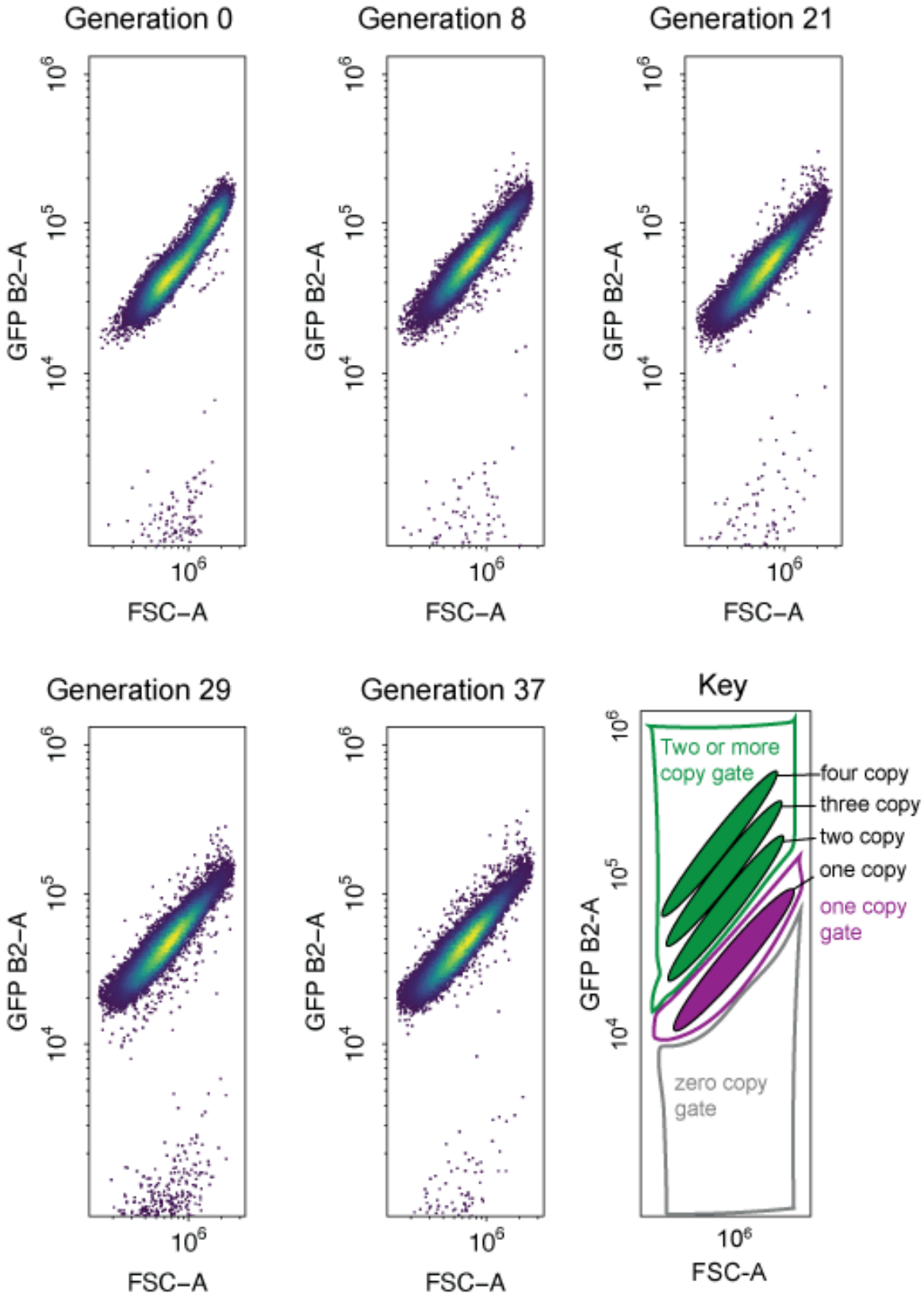
ALLΔ population 5



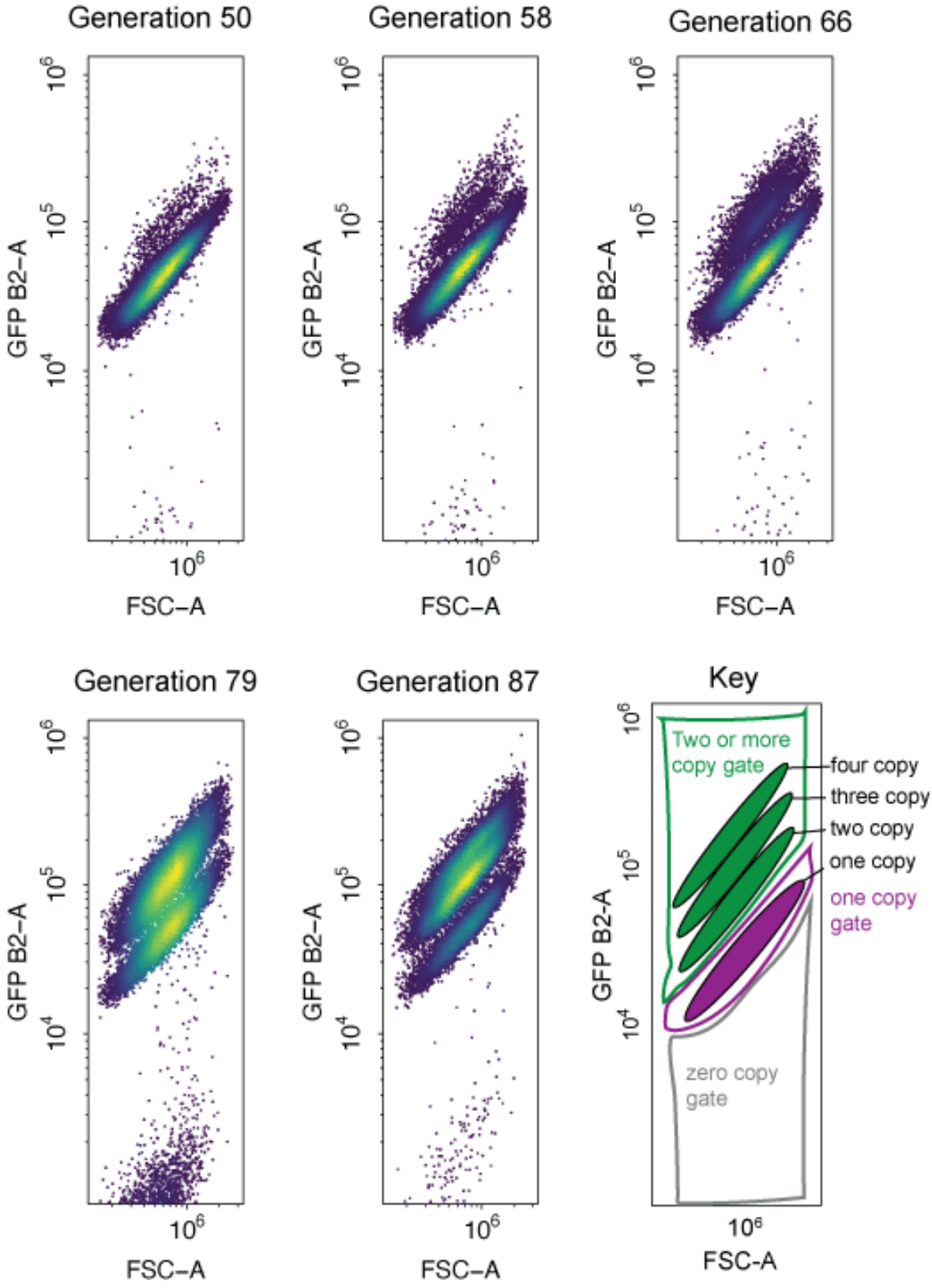
ALLΔ population 5



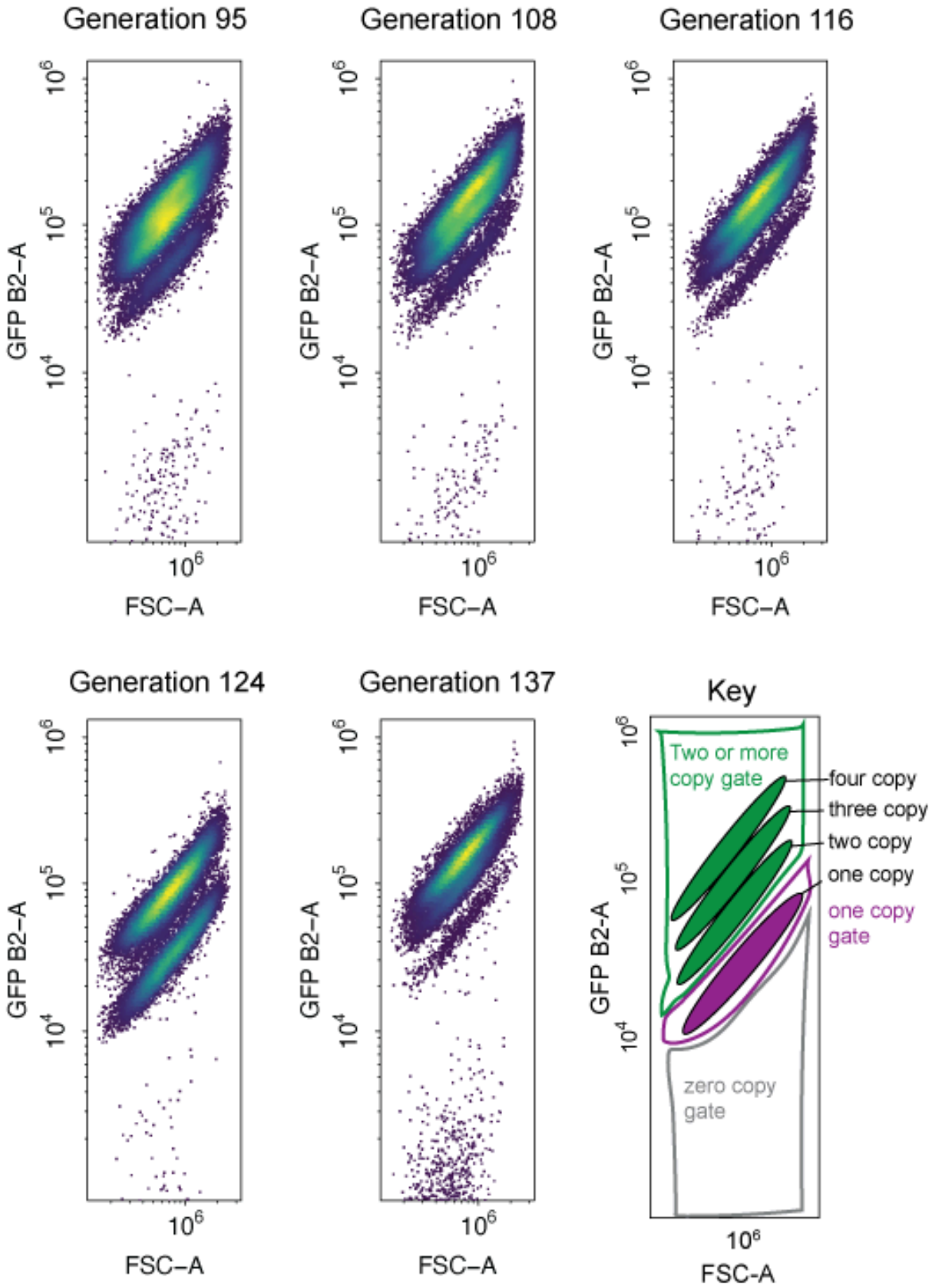
ALLΔ population 6



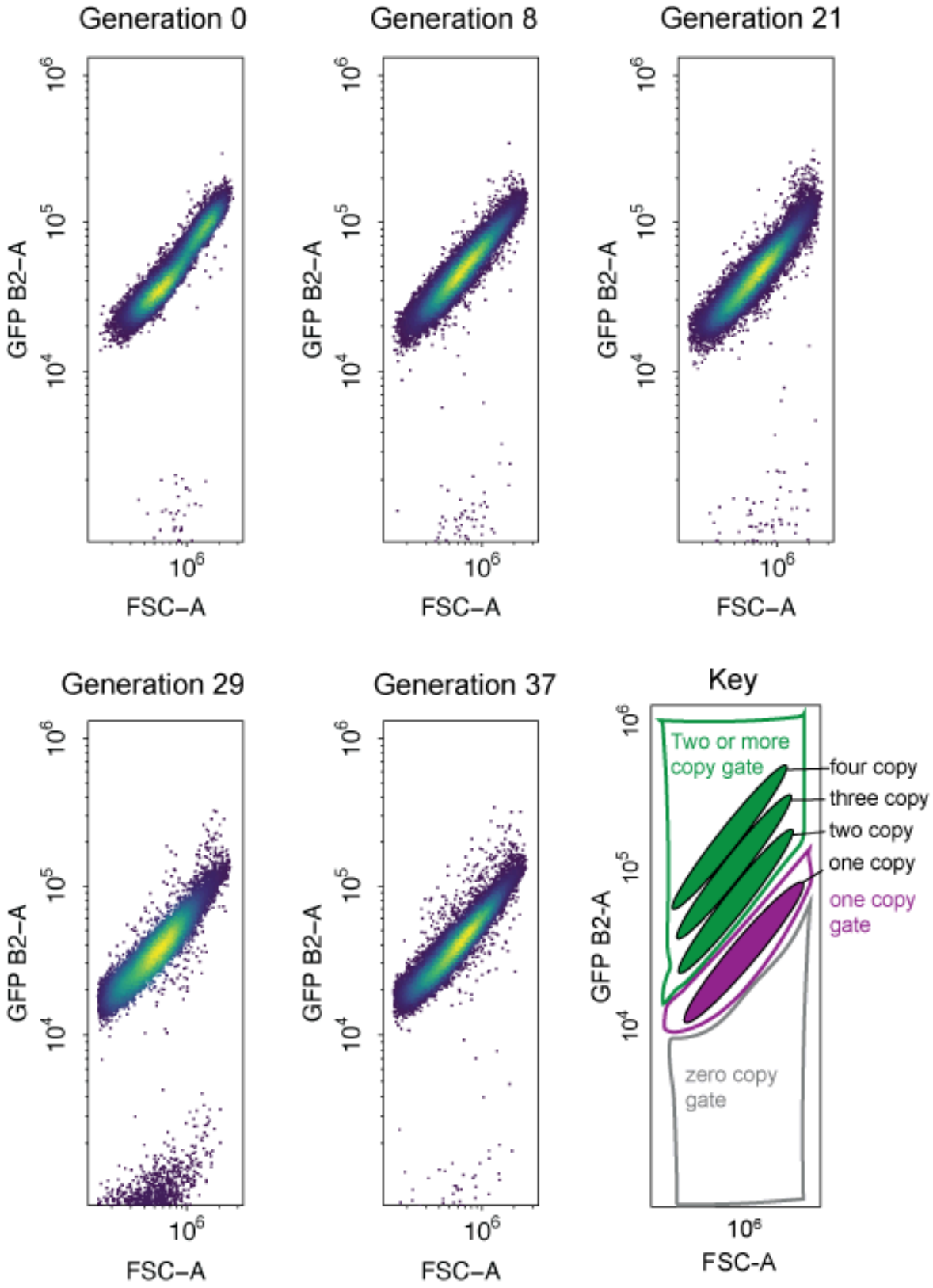
ALLΔ population 6



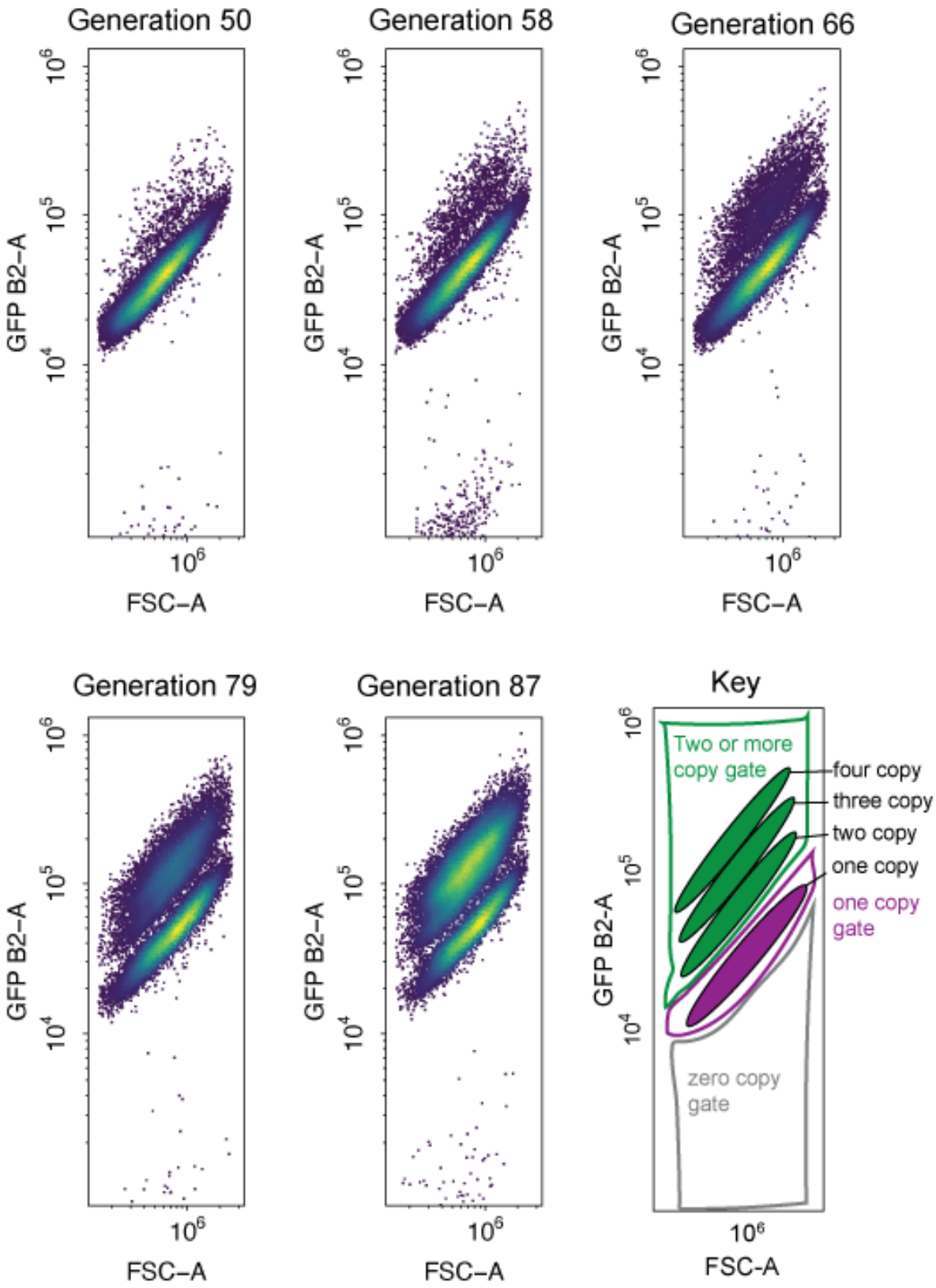
ALLΔ population 6



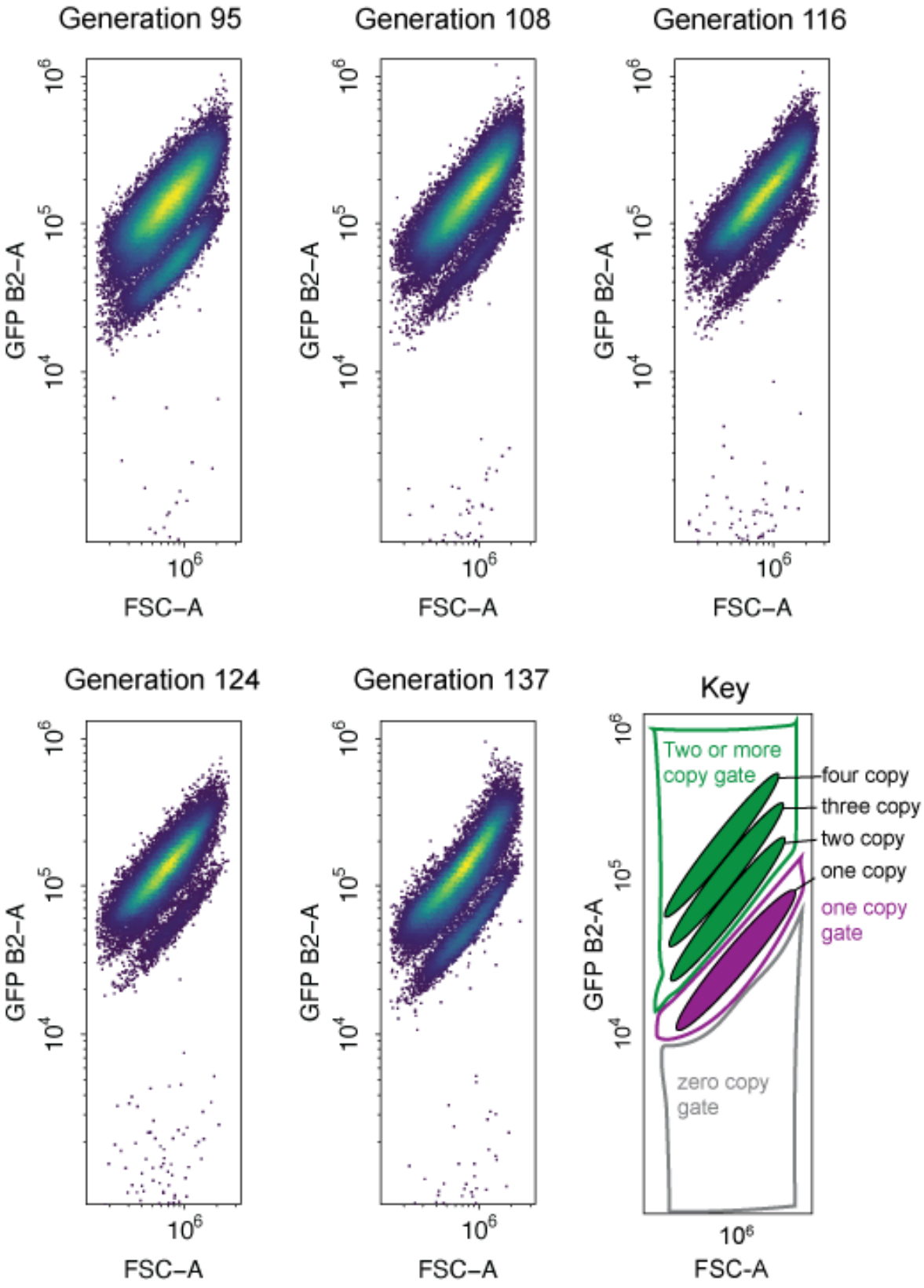
ALLΔ population 7



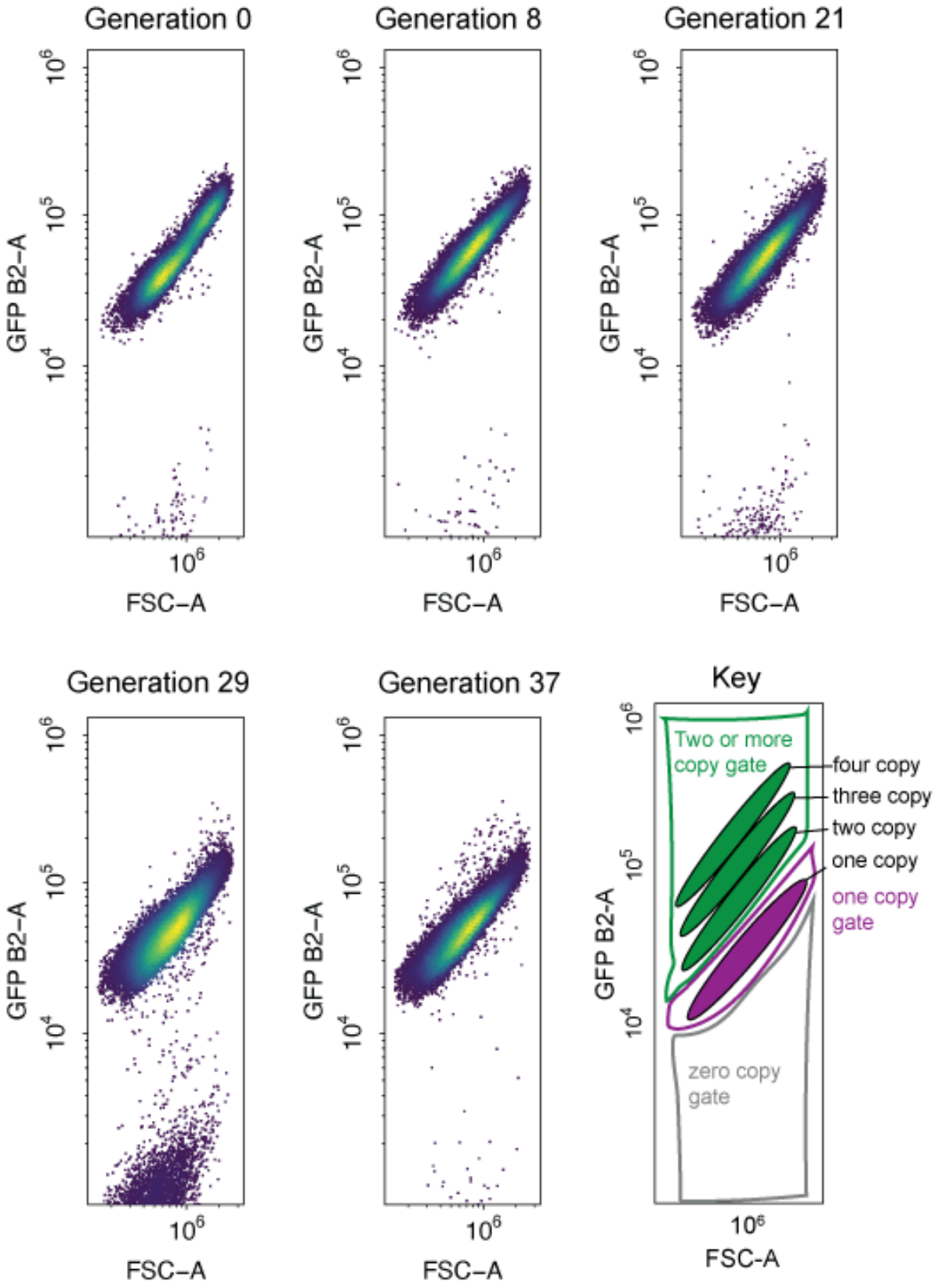
ALLΔ population 7



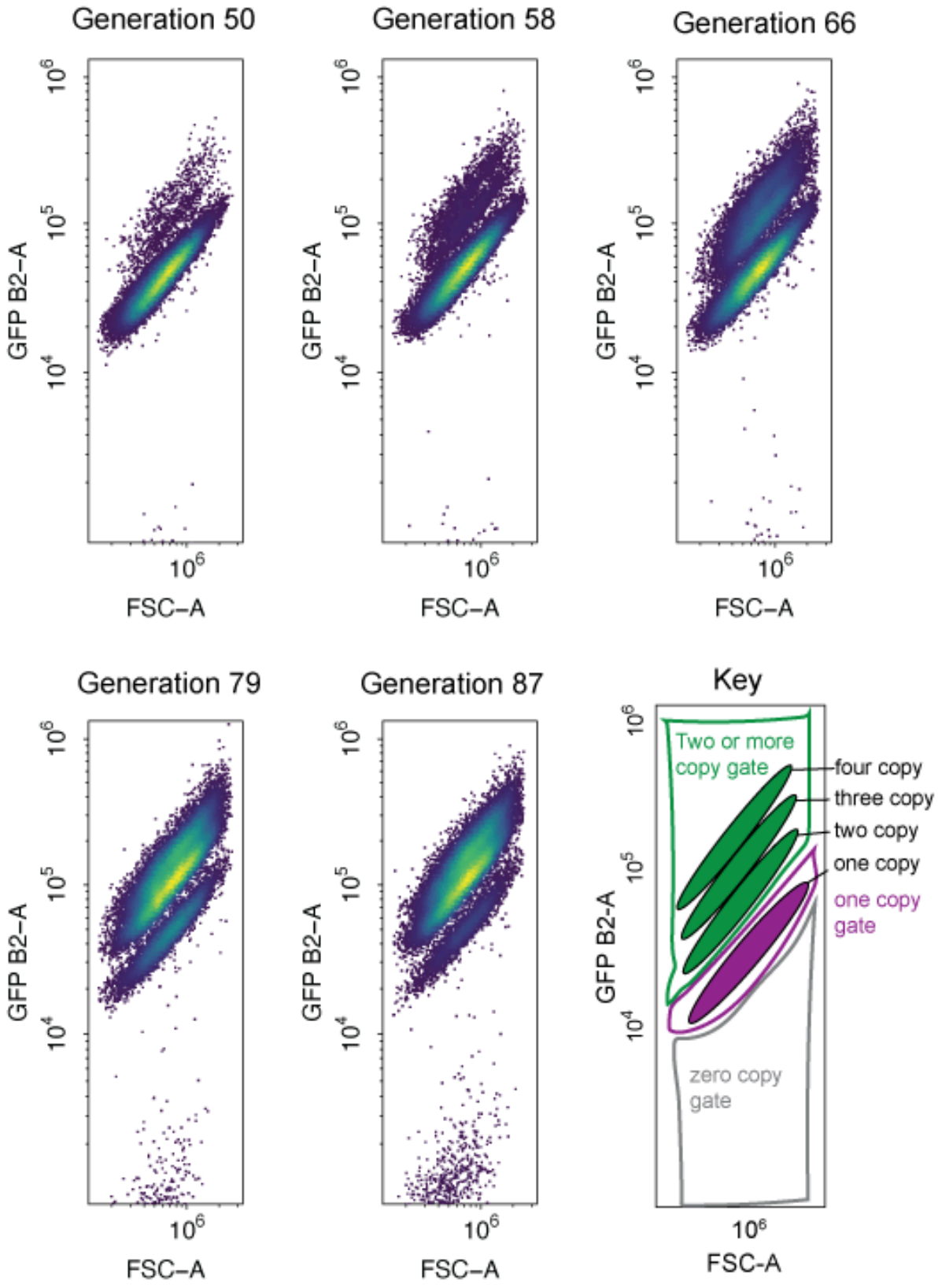
ALLΔ population 7



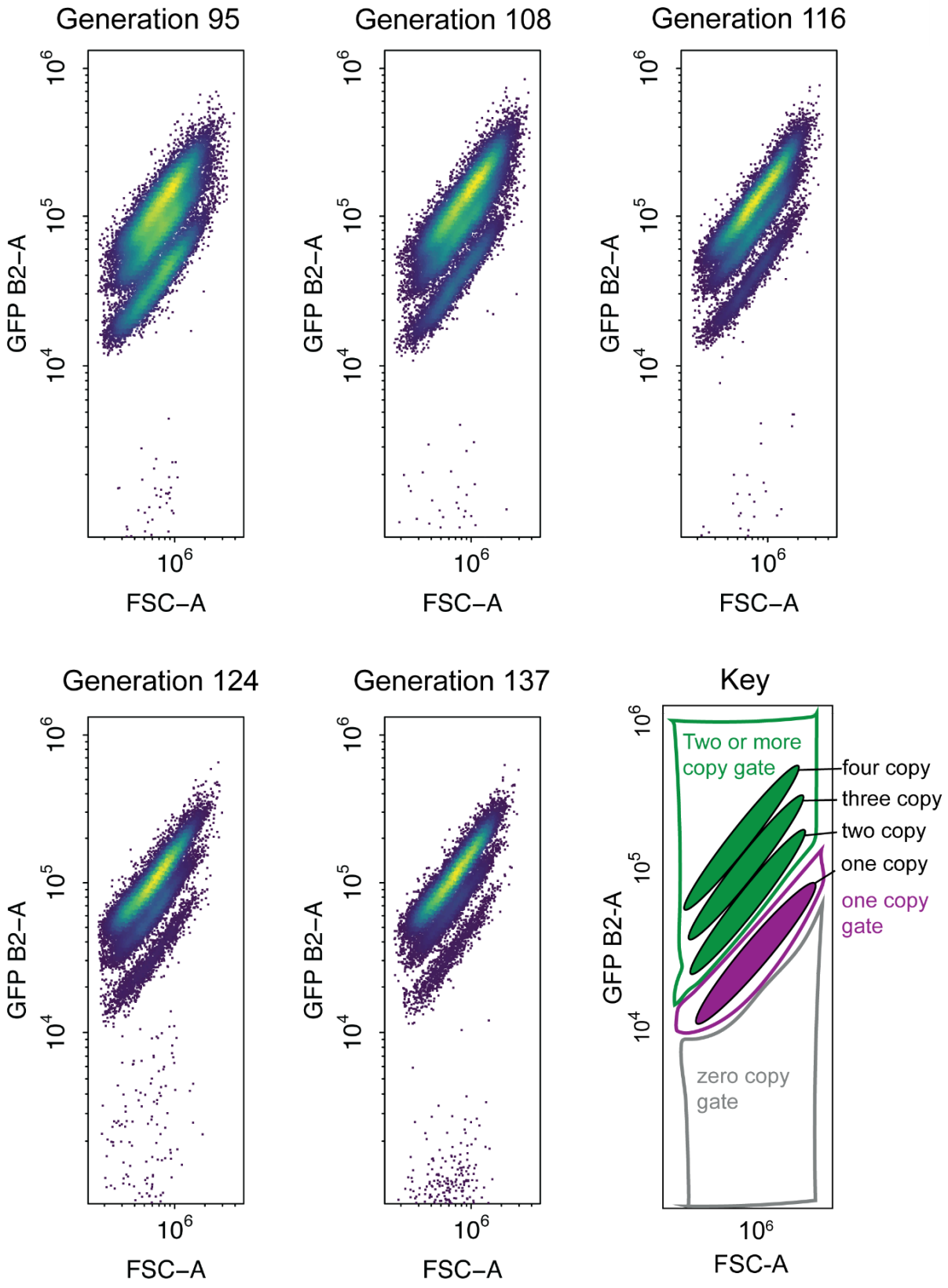
ALLΔ population 8



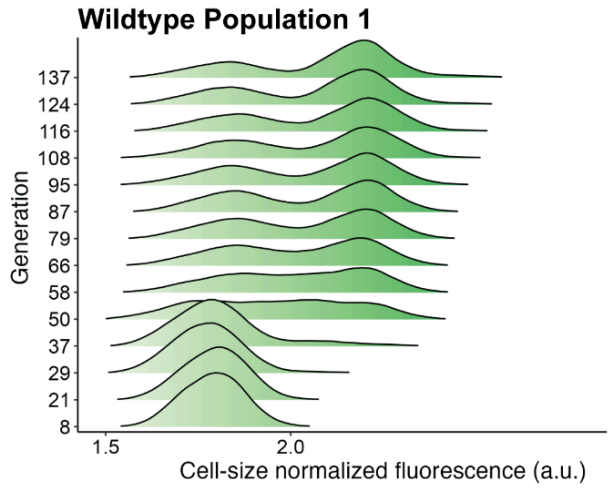
ALLΔ population 8

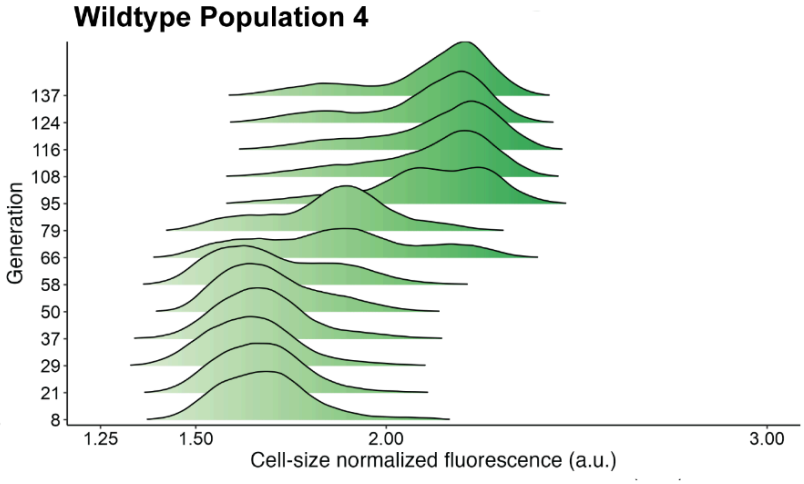


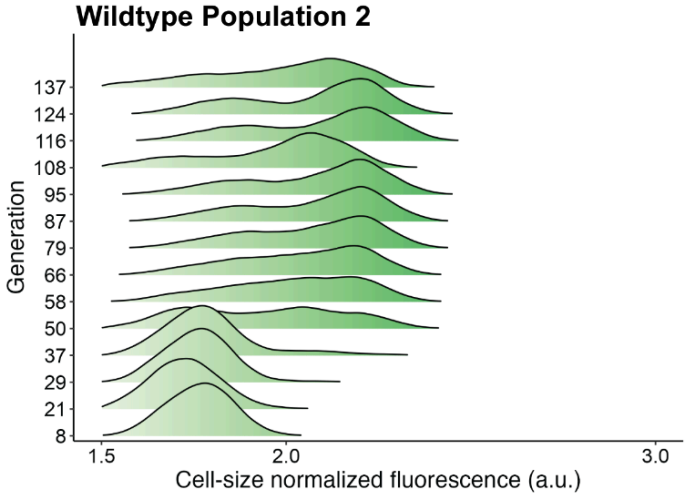
ARSΔ population 8

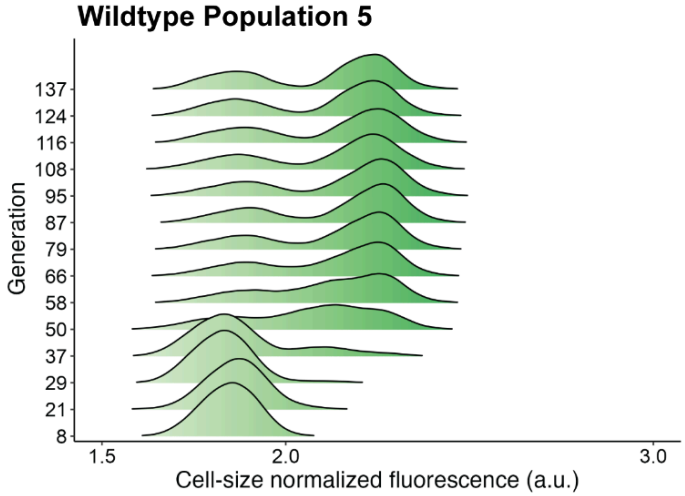


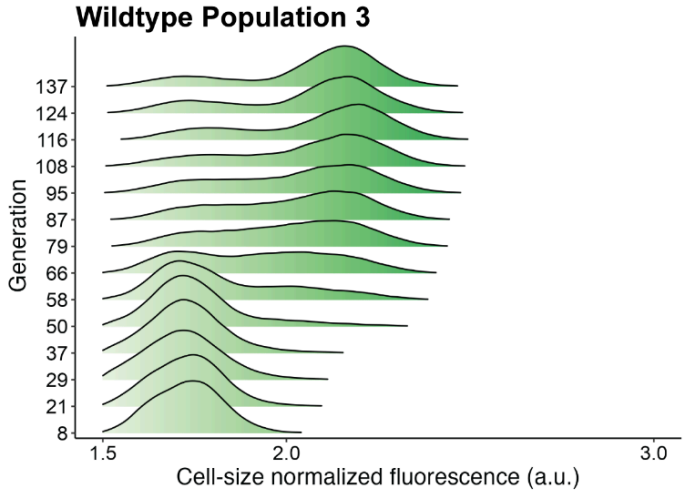
Supplementary Figure 2. Raw flow cytometry plots over the long term experimental evolution FSC-A is forward scatter-area which is a proxy for cell size. GFP fluorescence was measured using the B2-A channel in arbitrary units. Hierarchical gating was performed to identify zero-, one-, and two-or-more-copy populations. Within the two-or-more copy gate, distinct subpopulations formed consistent with having a two-, three-, four- copies of GFP.

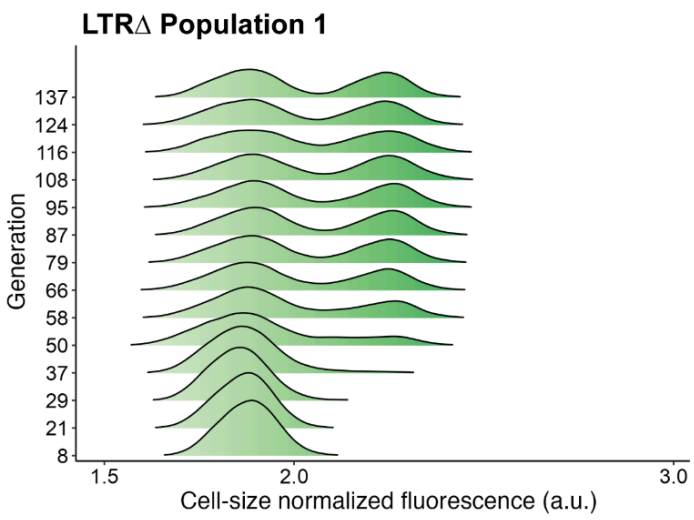


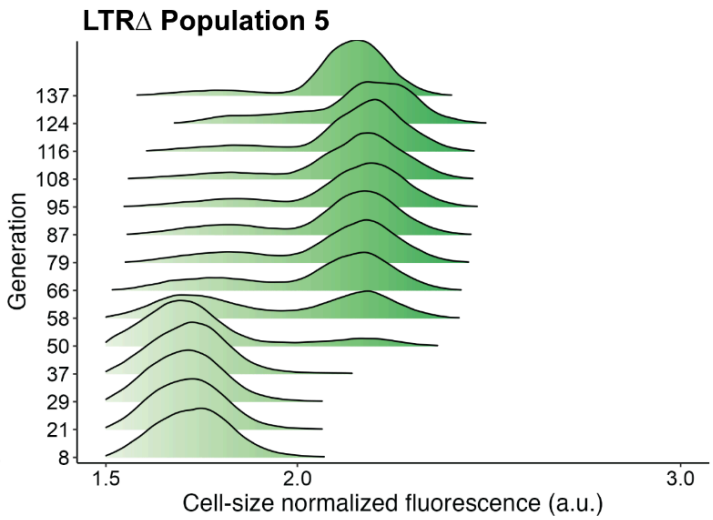


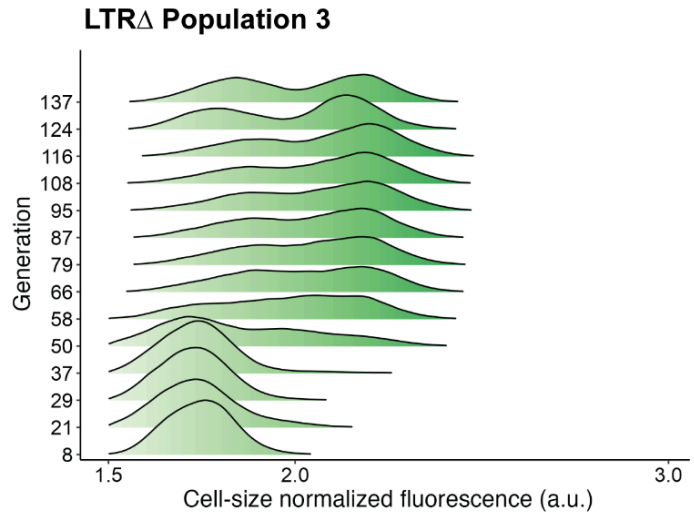


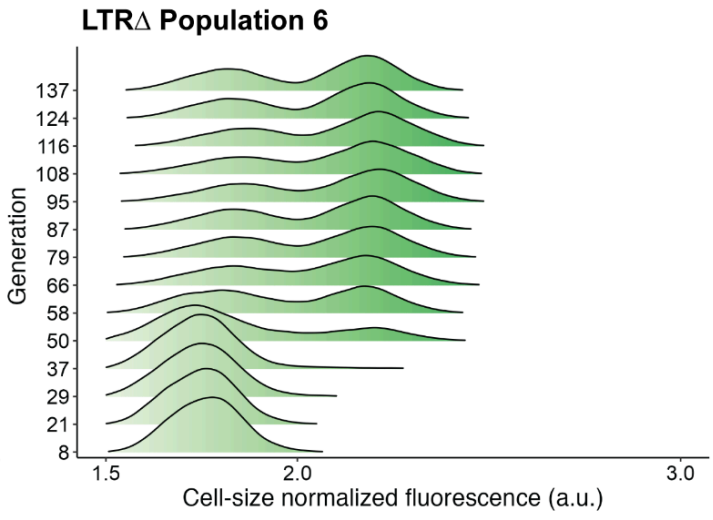


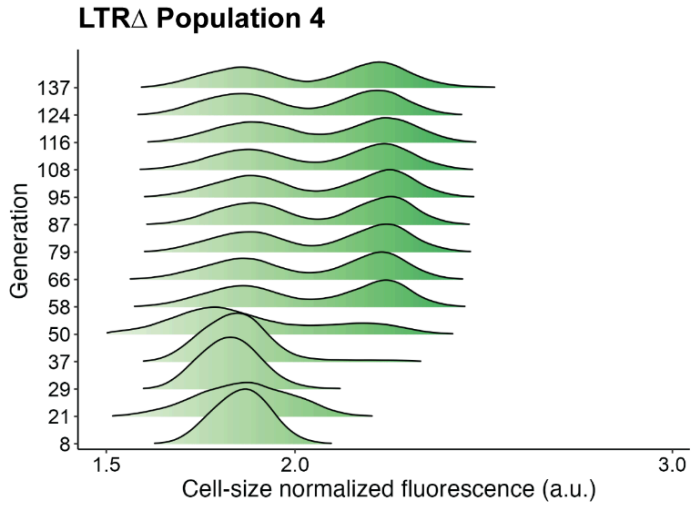


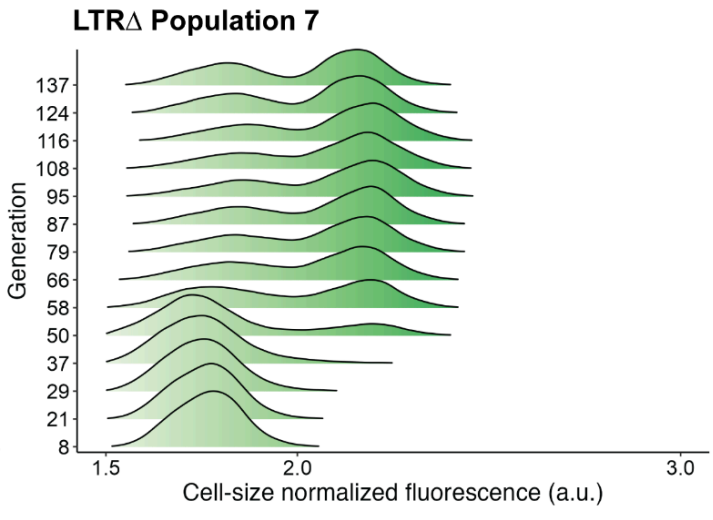


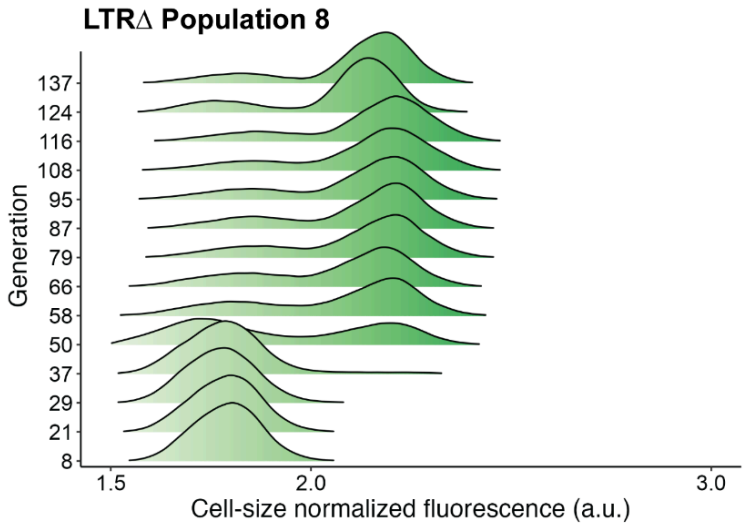


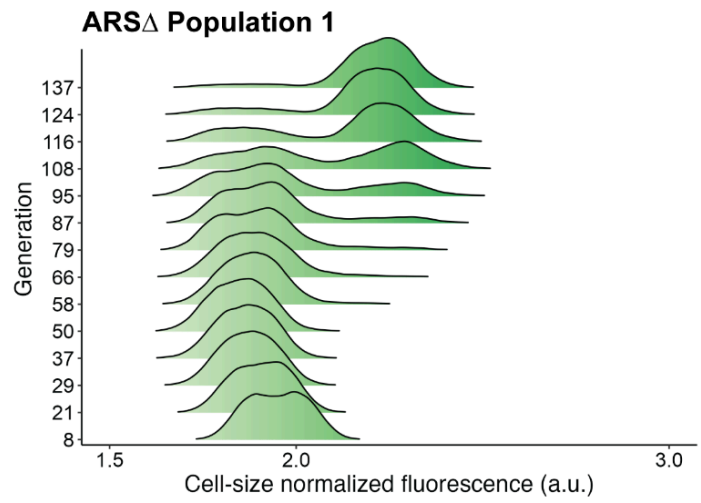


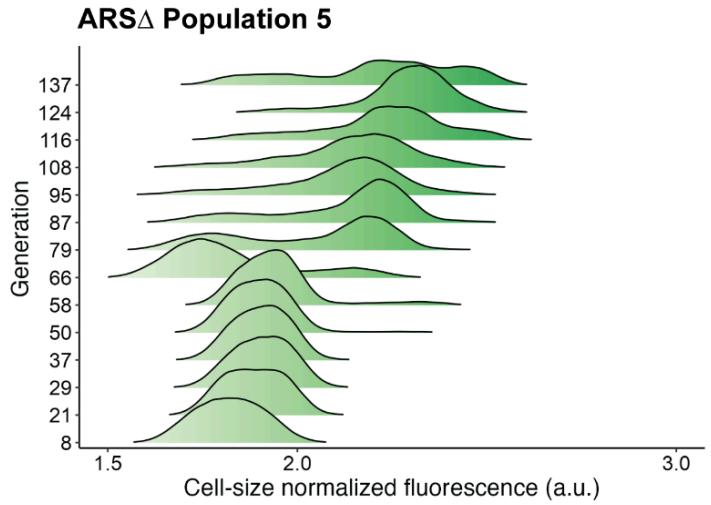


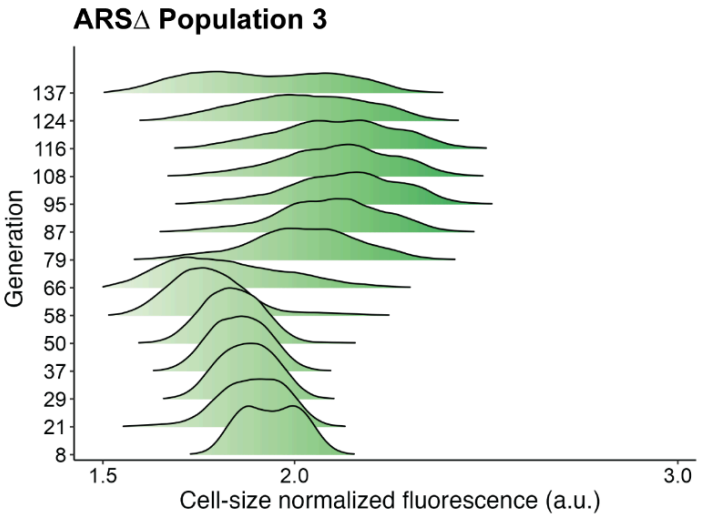


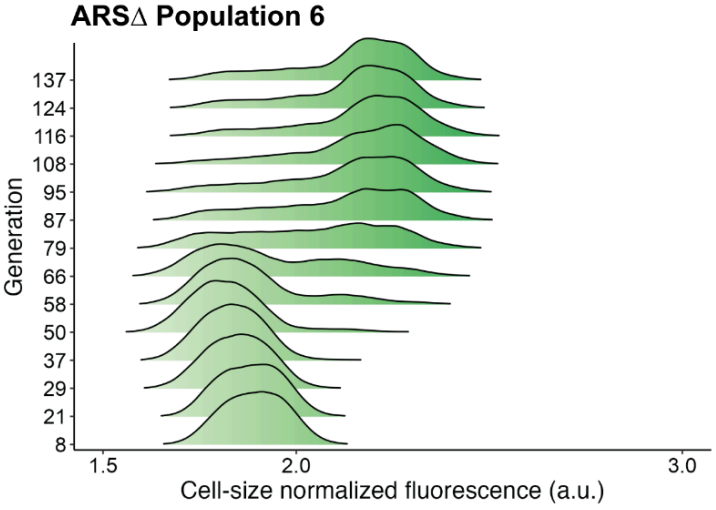


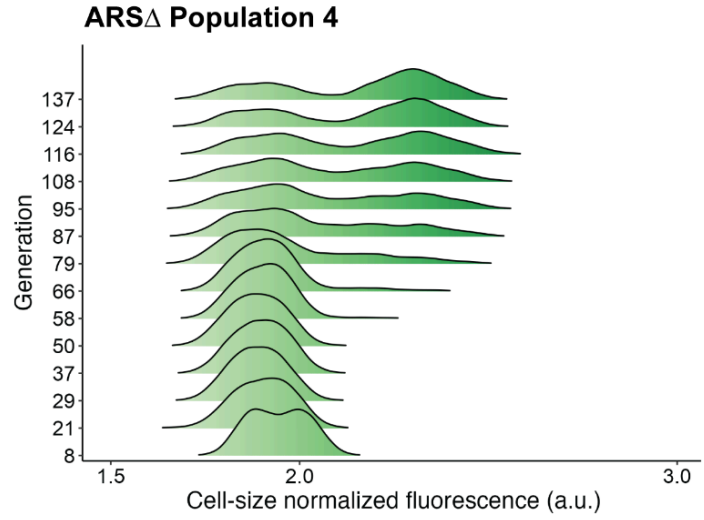


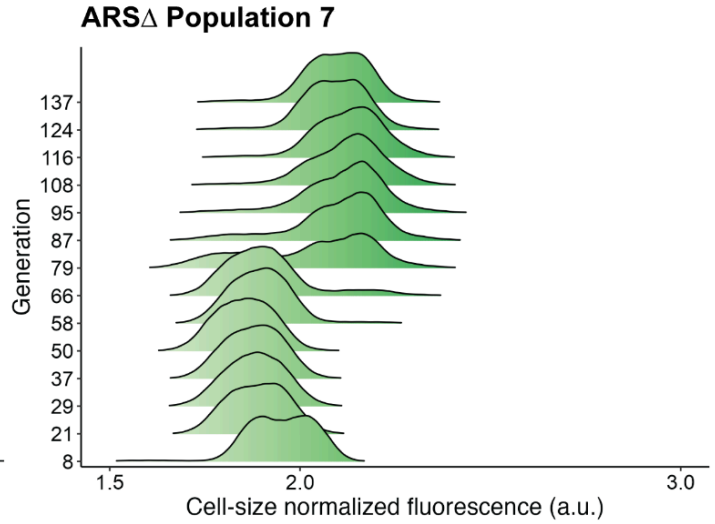


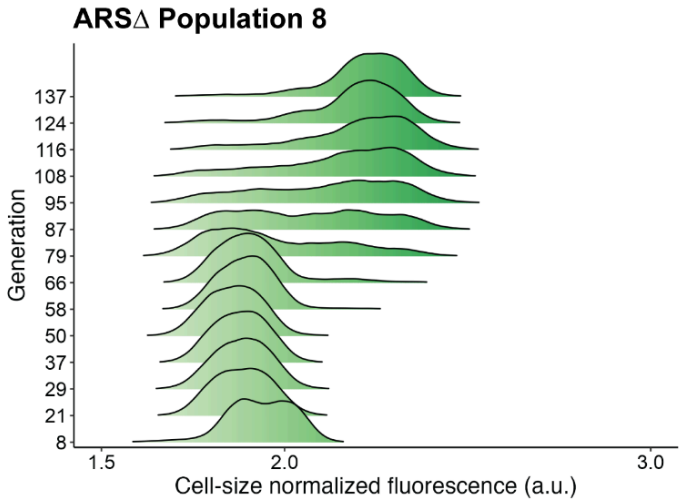


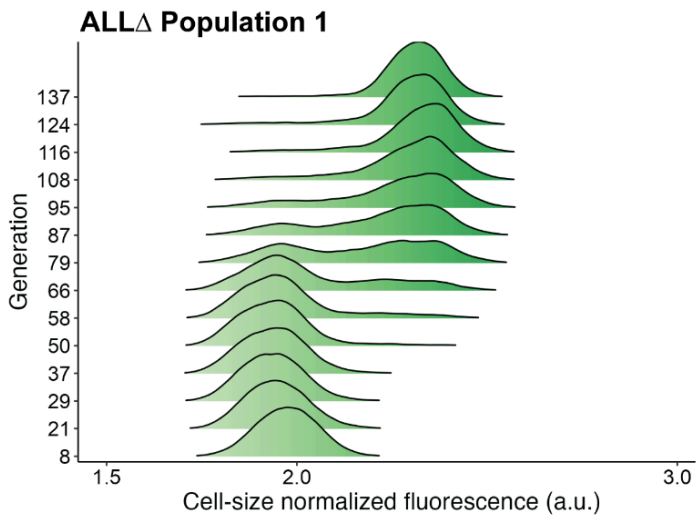


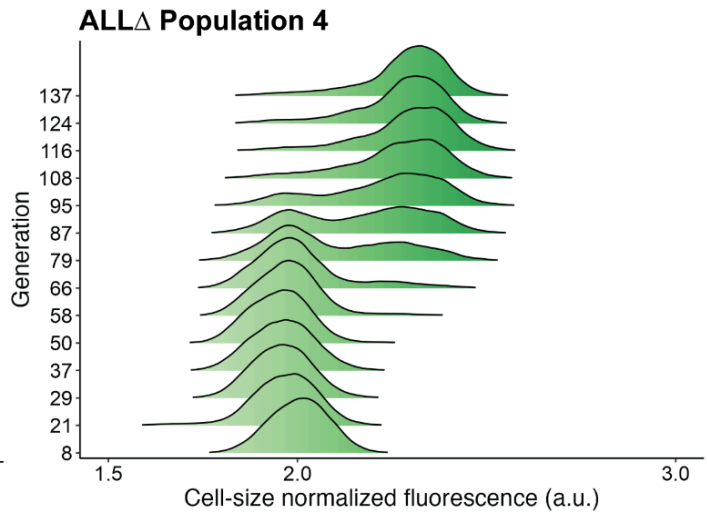


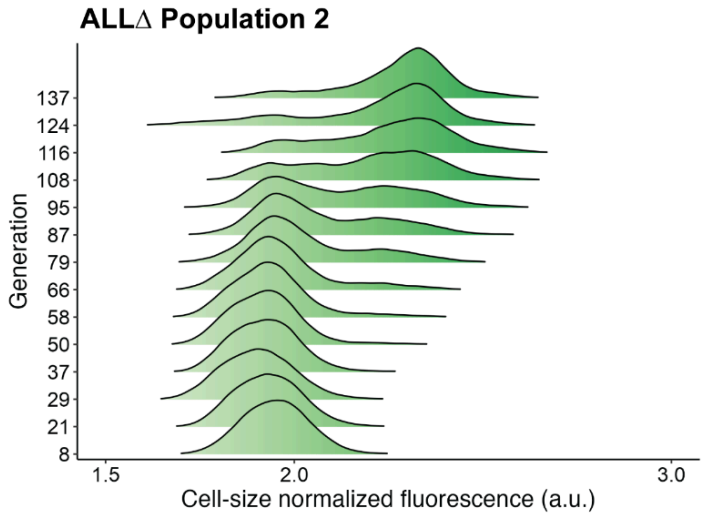


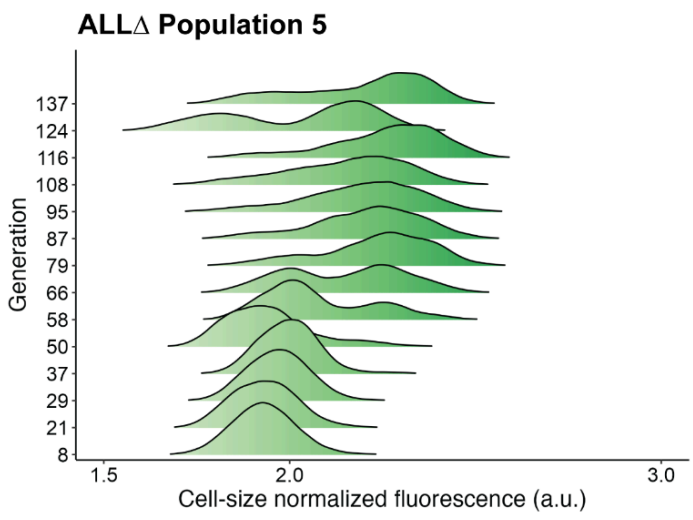


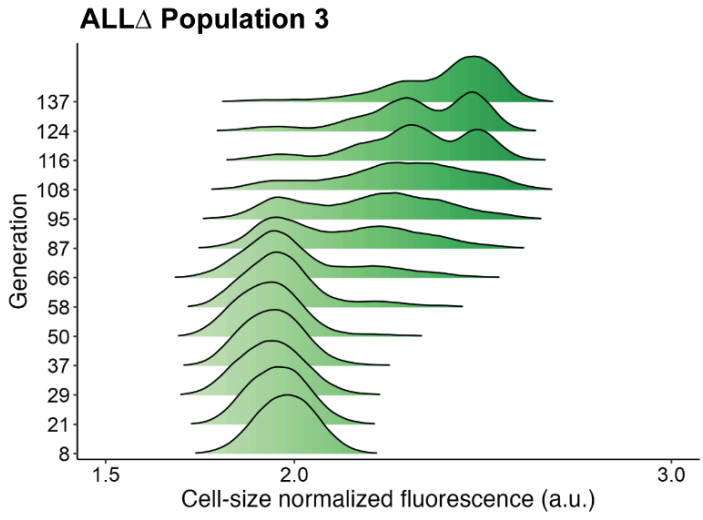


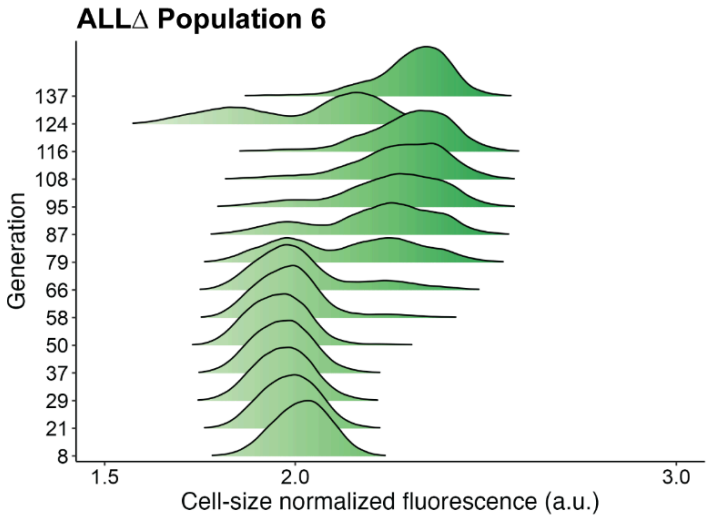


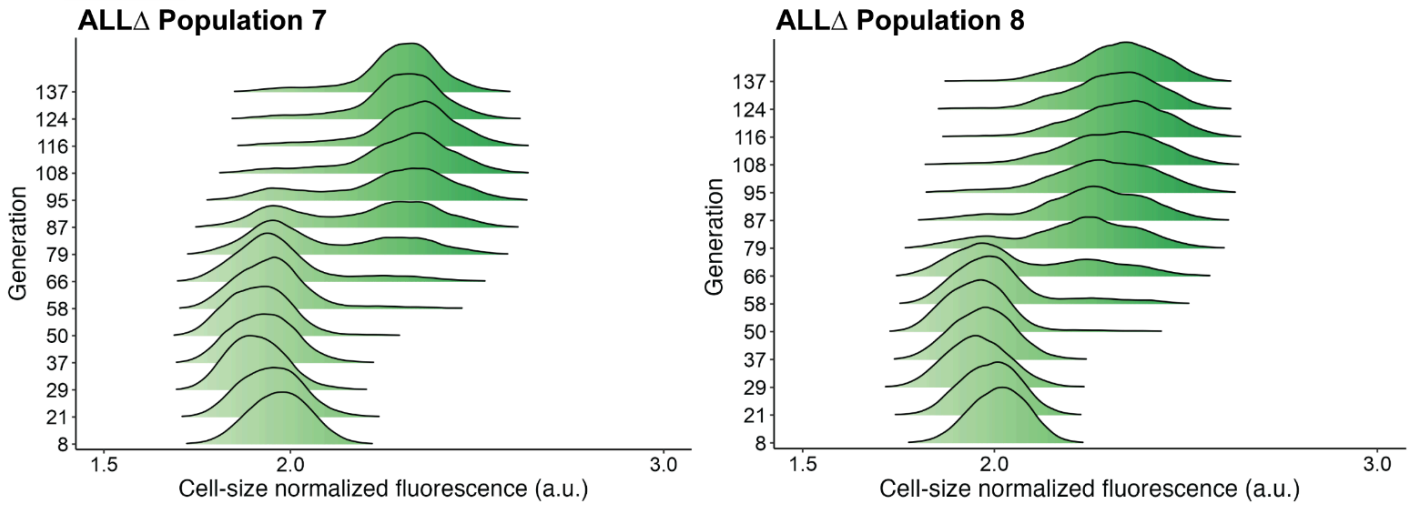




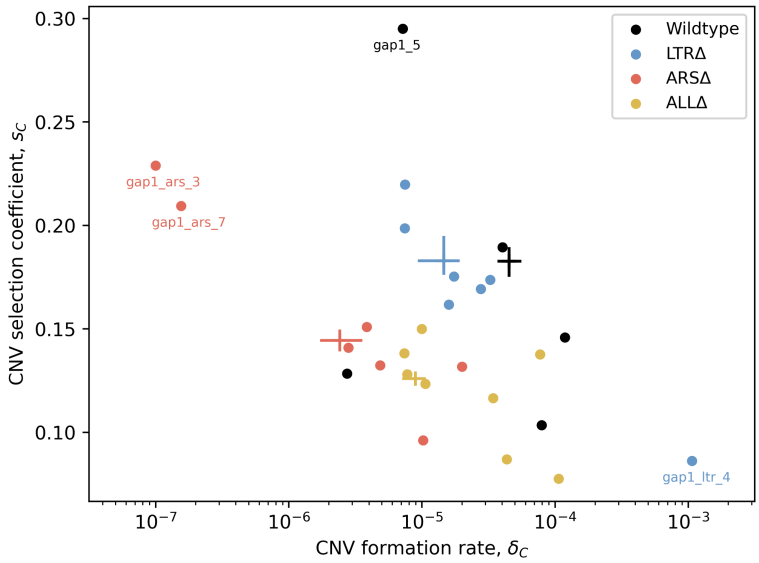




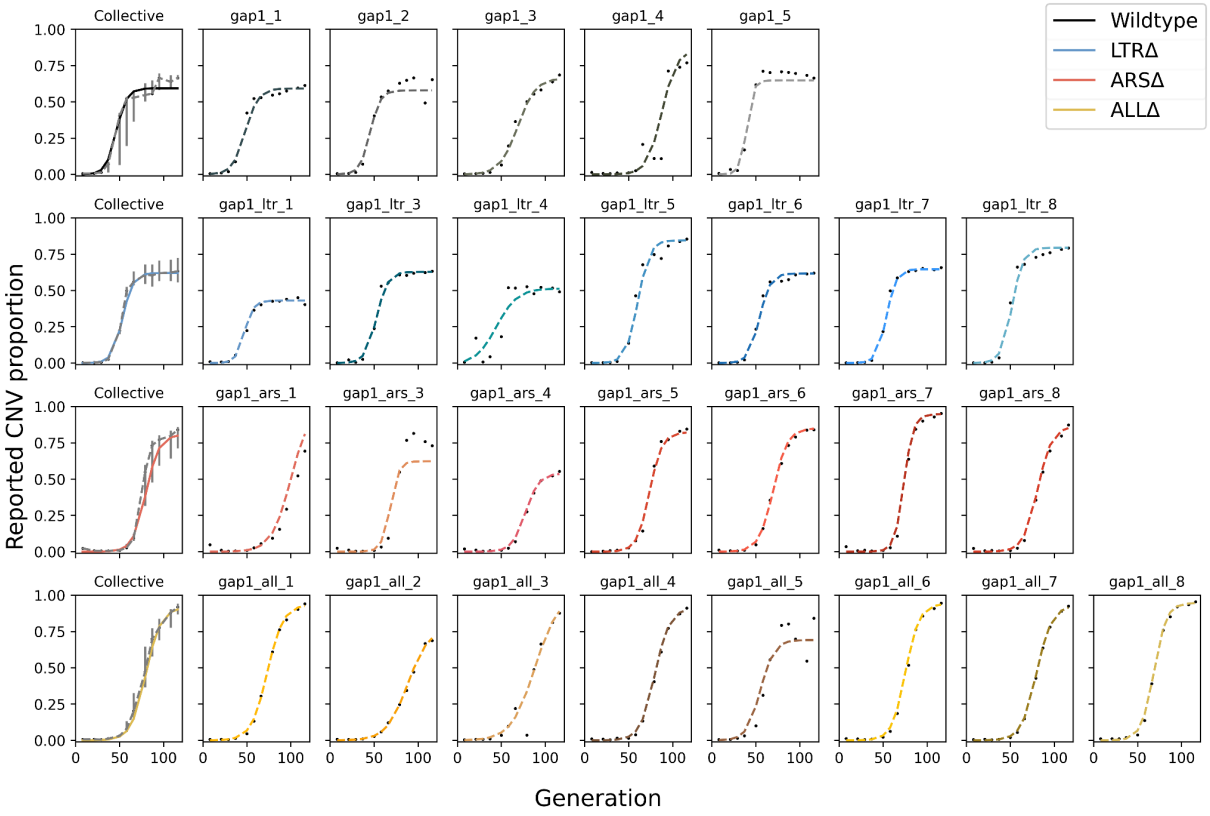




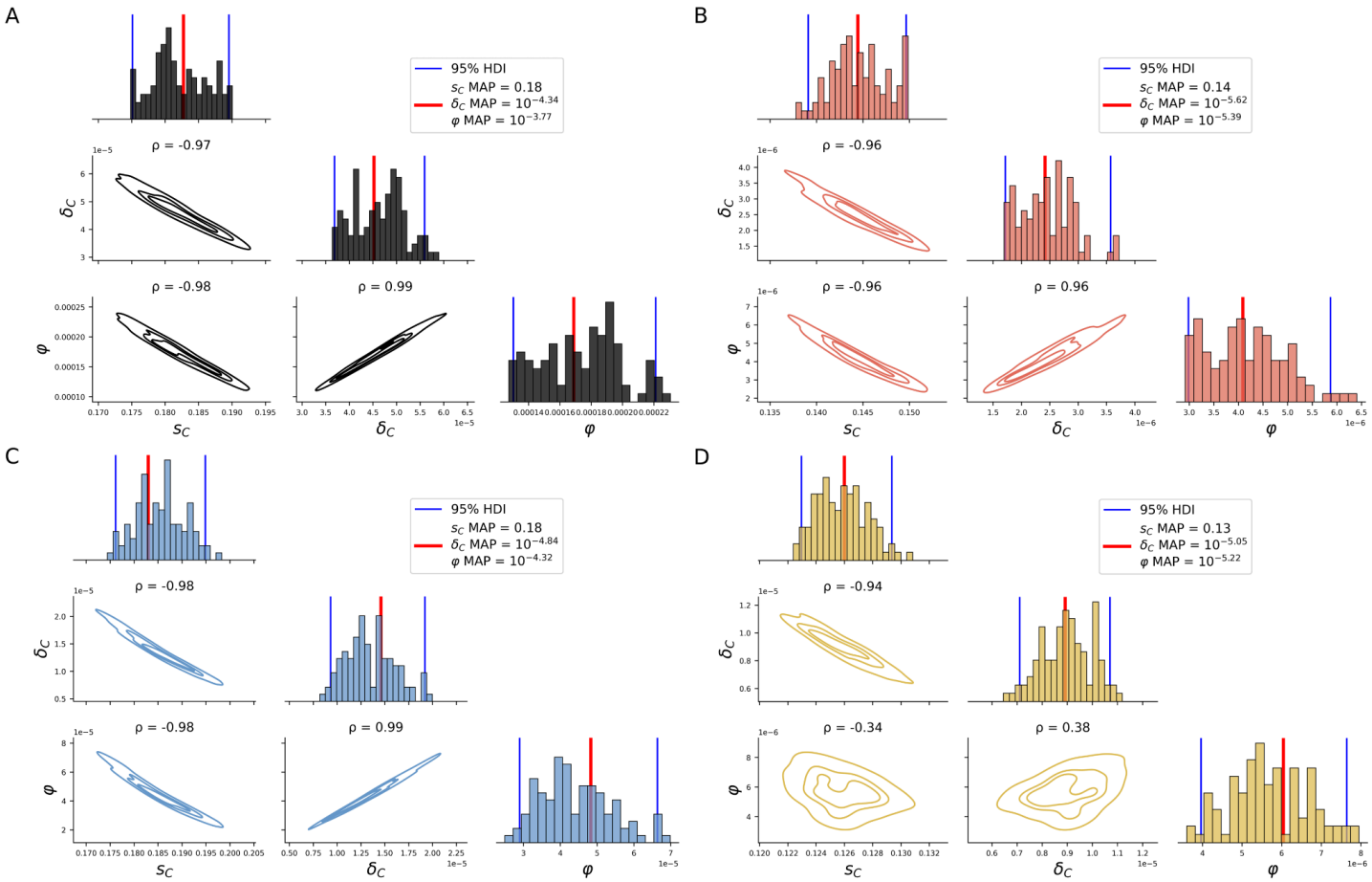
Supplementary Figure 3. Population GFP Ridgeplots. Density plots of cell-size normalized GFP fluorescence in arbitrary units (a.u.) for every population and timepoint over the course of long-term experimental evolution in glutamine-limited chemostats.



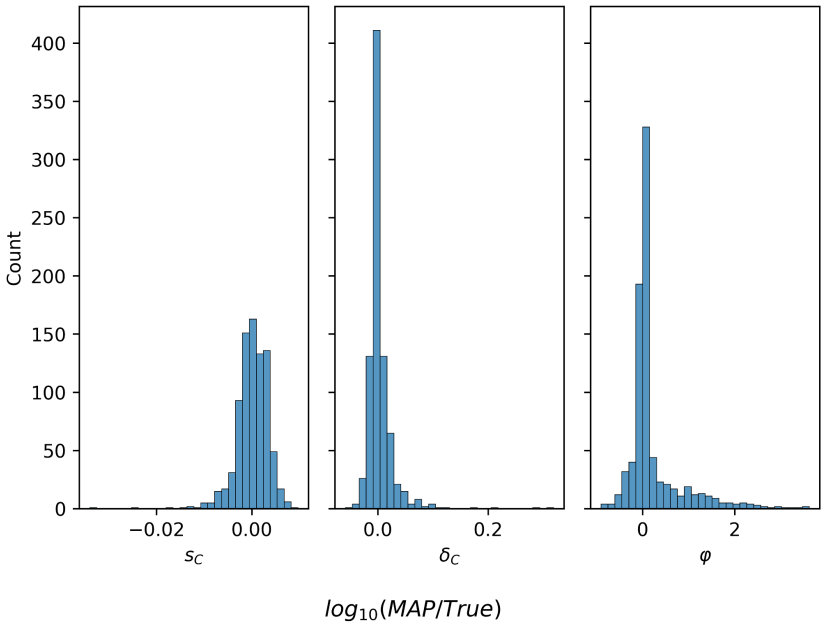
Supplementary Figure 4. MAP estimates of *GAP1* CNV formation rates (δ_C) and selection coefficients (s_C) for all replicate populations. Markers show MAP estimates from individual replicates, crosses show 50% HDI of collective posteriors. Extreme points are marked for comparison to data and posterior prediction, see Supplementary Figure 5 for posterior predictive checks.



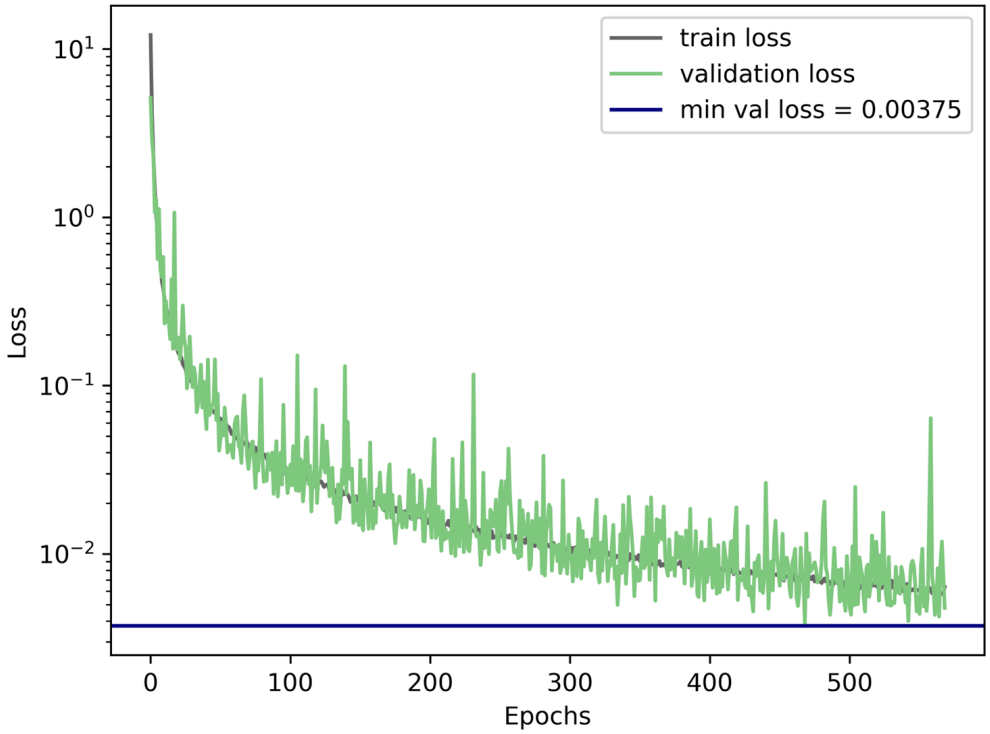
Supplementary Figure 5. Posterior predictive checks for all replicates. Black markers are the empirical observations, dashed line shows MAP prediction. The leftmost plot of each row shows the collective MAP prediction with empirical data's interquartile range (gray bars).



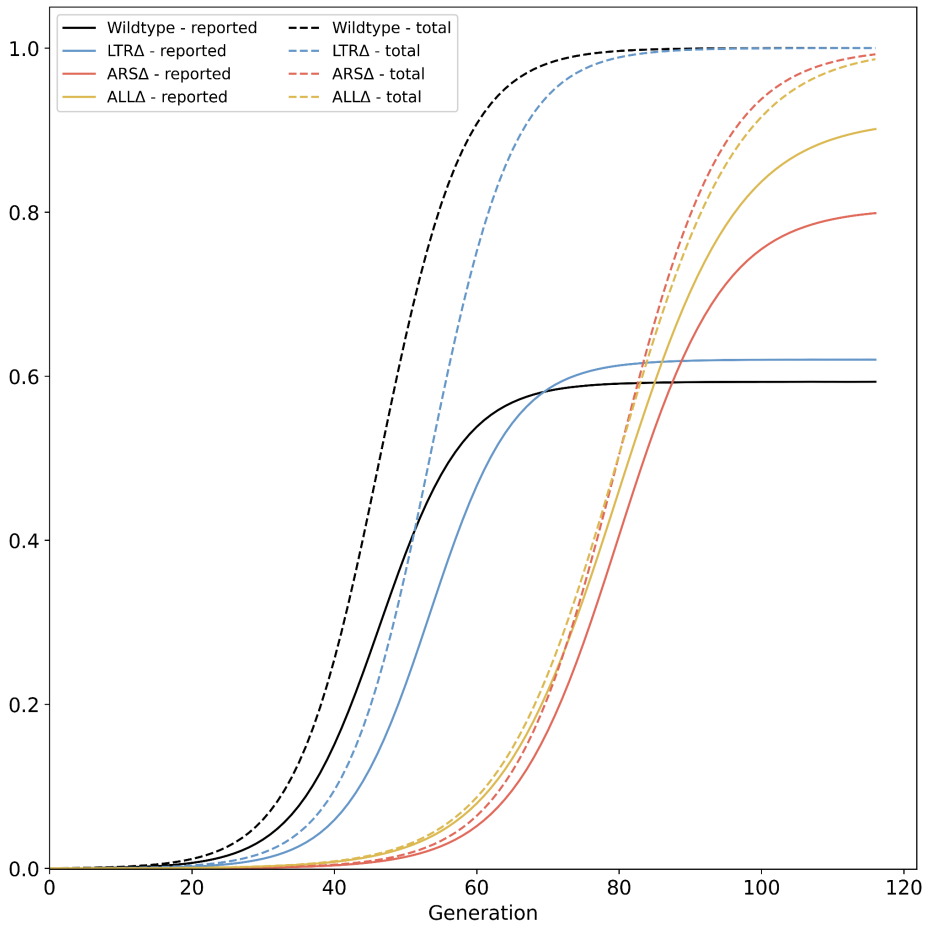
Supplementary Figure 6. Pairwise and marginal collective posteriors for all estimated model parameters. Diagonals show marginal collective posteriors per parameter per strain. Below-diagonal plots show pairwise KDEs for all pairs of model parameters. Collective joint MAPs (which may differ from collective marginal MAPs, as the marginal distribution integrates over all other parameters), are marked by a red vertical line. Panels are separated by strain: **(A)** WT, **(B)** ARSΔ, **(C)** LTRΔ, **(D)** ALLΔ.



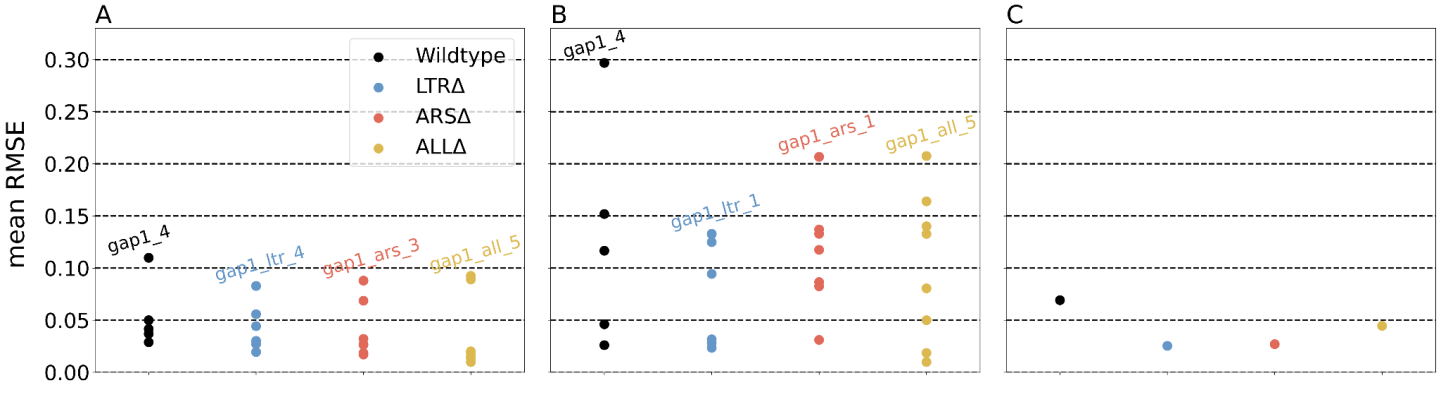
Supplementary Figure 7. Parameter estimation accuracy on synthetic data. Log-ratio of MAP estimate and true parameter value for 829 synthetic simulations in which the final reported *GAP1* CNV proportion is at least 0.3.



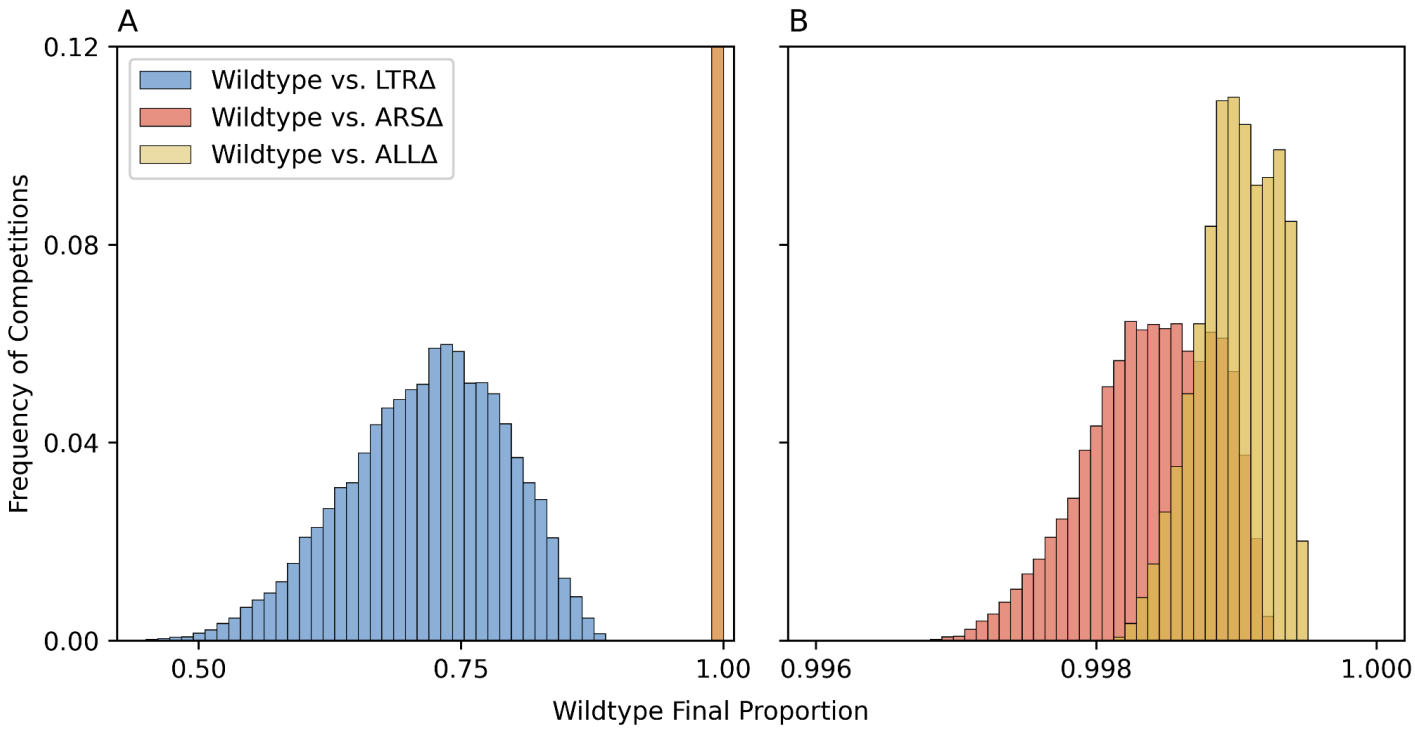
Supplementary Figure 8. Neural density estimator training and validation loss during training. Convergence threshold of 100 unimproved epochs (no decrease in minimal validation loss) was reached after 569 epochs.



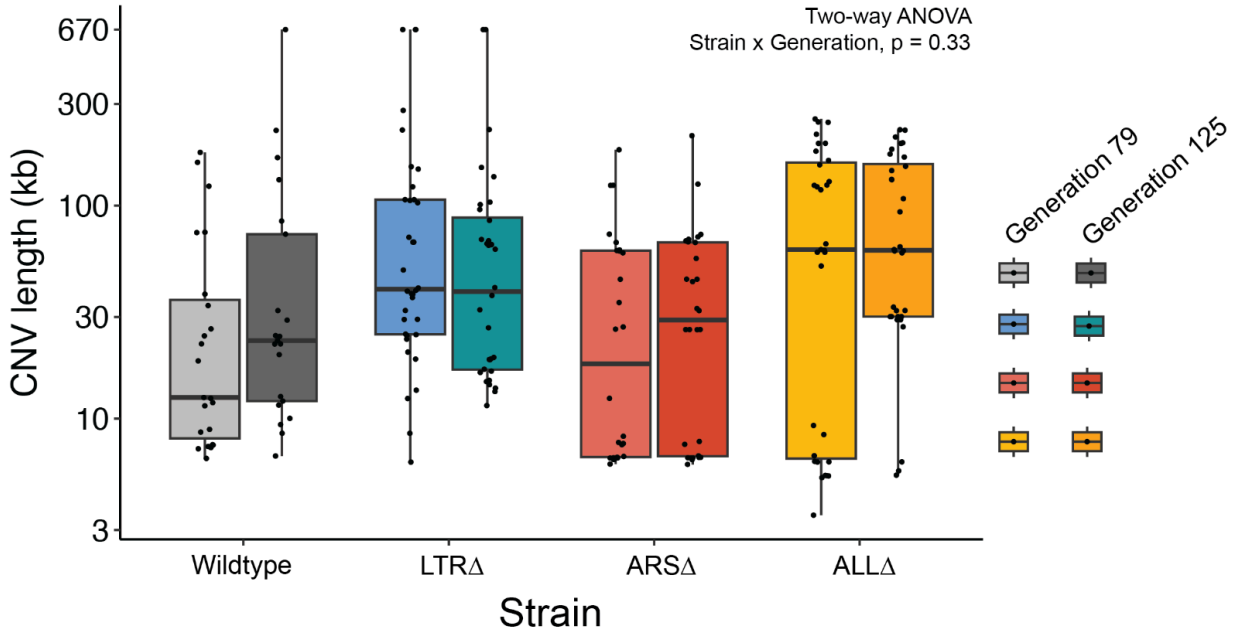
Supplementary Figure 9. Total *GAP1* CNV frequency. Solid lines show collective MAP predictions, dashed lines show the total proportion of *GAP1* CNVs, comprising unreported CNVs and reported CNVs generated during the experiment, as predicted by the evolutionary model.



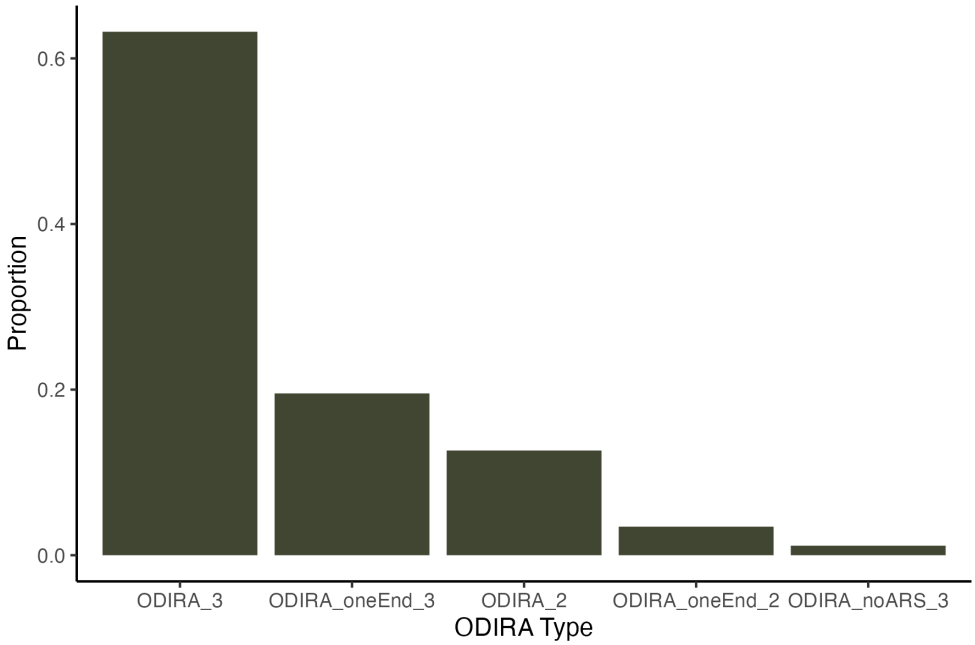
Supplementary Figure 10. Error estimation of parameter inference. Average root mean square errors (RMSE) of 50 posterior samples against the observed data. **(A)** Individual posteriors and individual replicates. **(B)** Collective posterior and individual replicates. **(C)** Collective posterior and empirical mean.



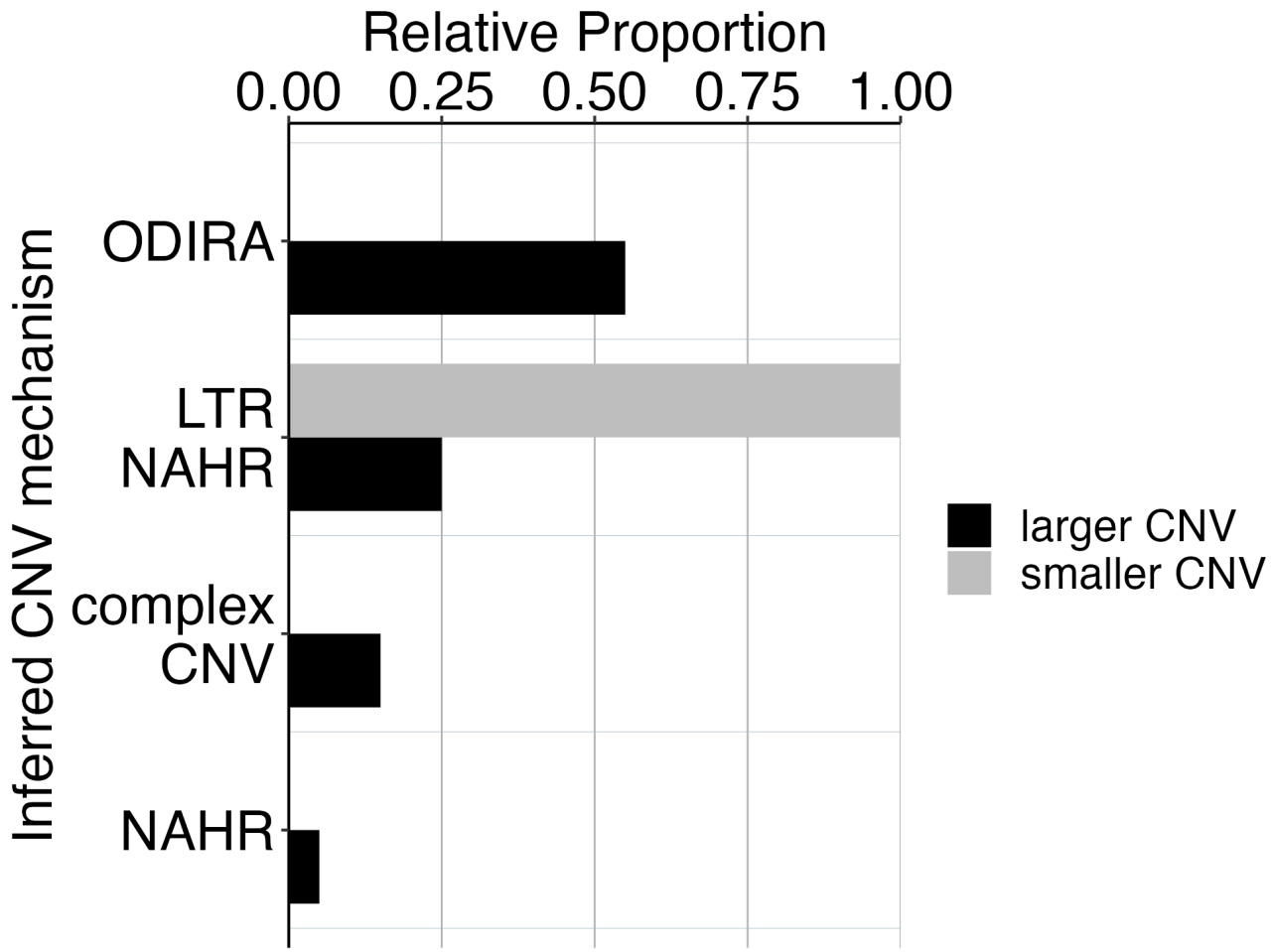
Supplementary Figure 11. Pairwise evolutionary competition predictions. We simulated evolutionary competitions in the experimental conditions of WT vs. genomic architecture mutants, starting from equal frequencies. The proportion at generation 116 of WT was predicted using 10,000 combinations of collective posterior samples for each pairwise competition. Overall, WT outcompetes all mutants because it adapts faster (due to faster CNV formation rate), but its advantage over ARSΔ and ALLΔ is much higher than its advantage over LTRΔ. (**A**) Histograms for three pairwise competitions. Note that ARSΔ and ALLΔ values overlap at this scale and are all in the rightmost bar. (**B)** High-resolution histograms for ARSΔ and ALLΔ.



Supplementary Figure 12. No significant interaction between strain and generation on CNV length. Boxplot of CNV length of clones by strain and generation of isolation. There is no significant interaction between strain and generation of isolated clone, and no significant effect of generation on CNV length (Two-way ANOVA, Strain x Generation, p = 0.33)



Supplementary Figure 13. **Types of ODIRA detected**. We found 87 ODIRA clones total regardless of strain. The majority of ODIRA clones fit the canonical definition of having two inverted junctions and 3 copies, 55/87 clones (63%) (ODIRA_3). We found four non-canonical types. We found 17 clones (20%) with only one inverted junction detected and 3 copies (ODIRA_oneEnd_3). We found 11 clones (13%) with two inverted junctions but only 2 copies (ODIRA_2) which may result from hairpin-capped double strand break repair. We found 3 clones (3.4%) with only one inverted junction detected and 2 copies (ODIRA_oneEnd_2). We found 1 clone (1.1%) with two inverted junctions but the amplified region did not contain an ARS.



Supplementary Figure 14. **CNV mechanisms in ARSΔ clones**. Two CNV sizes in ARSΔ clones correspond to different CNV mechanisms. We found two different groups of CNV lengths in the ARSΔ clones. 100% of smaller CNVs (6-8kb) correspond with a mechanism of NAHR between LTRs flanking the *GAP1* gene. Larger CNVs (8kb-200kb) correspond with other mechanisms that tend to produce larger CNVs, including ODIRA and NAHR between distal LTR elements. The smaller CNVs are indeed focal amplifications of *GAP1* that are 8kb or less.

Supplementary Files

Supplementary File 1. Ty-associated clones and locations of novel Ty insertions.

Supplementary File 2. CNV Clone Sequencing Analysis