bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Title: Vpr shapes the proviral landscape and polyclonal HIV-1 reactivation patterns in cultured cells

Authors: Edmond Atindaana[1], Sarah Emery[2], Cleo Burnett[1], Jake Pitcher[1], Jeffrey M. Kidd[2], Alice Telesnitsky[1*].

Affiliations:

[1]Departments of Microbiology and Immunology, and [2]Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109, USA.

*To whom correspondence should be addressed: ateles@umich.edu

One Sentence Summary: Expression properties and responsiveness to latency reactivation agents of individual HIV-1 proviral clones within polyclonal populations are masked by dominant clones and influenced by proviral proximity to certain epigenetic marks and by Vpr, a viral factor not previously known to affect latency and reactivation.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract:

Cell culture models suggest that the HIV-1 viral protein R (Vpr) is dispensable for latency establishment. However, whether Vpr affects the persistent proviral landscape and responsiveness to latency reversing agents (LRAs) is unclear. Here, integration site landscape, clonal dynamics, and latency reversal effects of Vpr were studied by comparing barcoded *vpr*+ and *vpr*- populations arising after infection of Jurkat cells in vitro. The results showed that individual integrant clones differed in fractions of LTR-active daughter cells: some clones gave rise to few to no LTR-active cells while for others almost all daughter cells were LTR-active. Integrant clones with at least 60% LTR-active cells (high LTR-active clones) contained proviruses positioned closer to preexisting enhancers (H3K27ac) and promoters (H3K4me3) than clones with <30% LTR-active cells (low LTR-active clones). Comparing *vpr*+ and *vpr*- populations revealed that the *vpr*+ population was depleted of high LTR-active clones. Complementing *vpr*-defective proviruses by transduction with *vpr* 16 days after infection led to rapid loss of high LTR-active clones, indicating that the effect of Vpr on proviral populations occurs post-integration. Comparing *vpr*+ and *vpr*- integration sites revealed that predominant *vpr*+ proviruses were farther from enhancers and promoters. Correspondingly, distances to these marks among previously reported intact HIV proviruses in ART-suppressed patients were more similar to those in the *vpr*+ pool than to *vpr*- integrants. To compare latency reactivation agent (LRA) responsiveness, the LRAs prostratin and JQ1 were applied separately or in combination. *vpr*+ and *vpr*- population-wide trends were similar, but combination treatment reduced virion release in a subset of *vpr*- clones relative to when JQ1 was applied separately, an effect not observed in *vpr*+ pools. Together, these observations highlight the importance of Vpr to proviral population dynamics, integration site landscapes, and responsiveness to latency reversing agents.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

HIV-1 establishes stable, long-lived reservoirs in patients treated with highly active antiretroviral therapy (HAART). Within this reservoir, cells that contain transcriptionally silent replication competent proviruses are inert to immune recognition and clearance as well as to current antiretroviral therapies; thus, even optimally treated patients remain infected (1). A cure for HIV-1 will require either ridding an individual of all infected cells (sterilizing cure) or reducing the reservoir to levels that enable the patient's immune control (functional cure).

Previous studies have demonstrated that latency establishment and maintenance is multifaceted and can involve the depletion of cellular factors such as the transcription factor NFkB, reduction in the level of the positive transcription elongation factor b (pTEFb), sequestration of pTEFb by the HEXIM/7SK RNA complex, the lack of CDK9 phosphorylation by CDK7 to promote pTEFb assembly, and promoter occlusion due to transcriptional interference from upstream host genes (2-7). The remodeling of the Nuc-1 nucleosome within the long terminal repeat (LTR) promoter, modulation of chromatin accessibility by histone modifying enzymes, and/or methylation of CpG sequences can also play roles in establishing and maintaining HIV-1 latency (8-10). Viral proteins and genetic elements may also contribute to latency, including deficiency of the viral transactivating protein Tat, which recruits pTEFb to the HIV promoter, deficiency of Rev, which mediates the export of unspliced HIV-1 RNA to the cytosol, or mutations in TAR or RRE *cis* acting RNA elements, which regulate these functions (11, 12).

Cure strategies that are under active investigation include "shock and kill" and "block and lock". In contrast to HAART, which prevents new infection, the shock and kill method involves inducing

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

69 virus expression, with the intention that this will lead to either cytopathic death of reactivated cells
70 or host immune recognition and clearance (13). Some proposed approaches involve perturbing
71 cellular pathways with small molecules or stimuli that complement intracellular deficiencies and
72 can reverse HIV-1 latency in experimental models. For example, prostratin and phorbol 12-
73 myristate 13-acetate (PMA) can reactivate latent HIV-1 through activation of protein kinase C
74 (PKC), leading to an increase in the level of NFkB. Other latency-reversing agents (LRAs) include
75 those which act to increase the level of pTEFb, including the BET bromodomains inhibitor JQ1
76 and a class of drugs known as ingenols. JQ1 notably inhibits the bromodomain of BRD4 and
77 prevents BRD4 from binding acelylated histones at promoters and specific enhancers, leading to
78 a decrease in lineage specific genes (14). Additional classes of LRAs act to modify the chromatin
79 environment: suberoyl anilide hydroxamic acid (SAHA) and entinostat being members of the class
80 that promotes a euchromatin state by inhibiting histone deacetylases (15-19).
81
82 The size of a patient's latent reservoir is estimated to be between one in a million and three in ten
83 thousand resting memory CD4+ T cells (20). The rarity of latently infected cells in virally
84 suppressed individuals makes the study of *in vitro* latency models necessary. Tissue culture models
85 developed to study HIV-1 latency and reactivation differ from one another in that some utilize
86 infectious virus while others employ HIV-1-based reporters. Prominent among this later class are
87 systems like the widely used J-Lat clones, which assess LTR promoter activity by reporter gene
88 expression but lack HIV-1 genetic elements believed to be dispensable in HIV-1 latency (21-23).
89 One such element is the *vpr* gene, whose product has been shown to play roles in efficient viral
90 infectivity, enhances CCR5-tropic HIV-1 replication in humanized mice, acts as a coactivator of
91 the LTR and enhances HIV-1 transcription by Vpr-mediated ubiquitination and degradation of

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Phosphorylated CTD Interacting Factor I (PCIFI) (24-27). Vpr also causes cell-cycle arrest and induces cell death through the depletion of the chromosome periphery protein CCDC137, and induces widespread changes in host gene expression in CD4+ T cells (28-30). For these reasons, some latency models use *vpr*-defective proviruses to allow the survival of cells harboring transcriptionally active proviruses. Cultured over time to allow proviral transcriptional silencing, these cells can then be used for reactivation studies (31-34).

Results from LRA reactivation studies using these models have been inconsistent, as demonstrated by Spina et al.(35). This discordance may be due to differences in the models of latency used (22). Some approaches used to produce silenced proviruses exclude latent clones that are established shortly after infection (23, 36). Moreover, HIV-1's semi-random integration site preference makes it likely that in each model, the integrants in a particular clonal cell line being studied have unique host chromosomal loci that may be subject to locus-specific regulation. The extent to which integration sites influence the reactivation of latent proviruses has long been recognized but not extensively explored (32, 37).

Molecules that are effective at reversing latency in various tissue culture models have been identified, but evidence thus far suggests that these are either too toxic to be therapeutically useful or that they fail to reduce the size of the latent reservoir in patients (19, 38-41). Fundamental aspects of latency in patients, such as the size of the latent reservoir, remain poorly defined. Recent studies have demonstrated that quantitative viral outgrowth assays (QVOA) underestimate the latent reservoir while PCR-based quantitation can overestimate it due to the predominance of defective proviruses (20), leading to ongoing efforts to improve reservoir quantification (42, 43).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Since accurately measuring the latent reservoir is challenging in and of itself, it seems likely that measuring changes in the latent reservoir before and after LRA treatment will be equally challenging. Because HIV-1 latency can be established and maintained by multiple mechanisms (4), an effective therapy will likely be one that disrupts multiple pathways.

To date, there is limited evidence addressing whether or not mechanistically distinct classes of LRAs affect all proviruses uniformly or have differential effects on individual proviral clones. In this study, the clonal structure and influence of hundreds to thousands of individual proviruses' host genomic loci on reactivation dynamics were investigated, with responses to distinct classes of LRAs in latent *vpr*- and *vpr*+ proviral pools compared. The results revealed differences in the clonal structures of populations that influenced total levels of reactivation and showed that the aggregate properties of polyclonal populations misrepresented the behavior of individual proviruses. Whether the presence of *vpr* affected integration at the time of infection or if its subsequent expression altered the structure of polyclonal populations or influenced activities of residual cells was examined. In response to dual treatment with JQ1, which affects pTEFb availability, and prostratin, which activates PKC, individual cell clones within polyclonal *vpr*- provirus populations exhibited either enhanced, additive or repressed reactivation, while the spectra of responses were more limited in type and magnitude in cell clones containing *vpr*+ proviruses. Together these findings suggest that Vpr plays a critical role in shaping transcriptionally inactive proviral populations and in defining their LRA responsiveness.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Results

vpr- and vpr+ proviral pools differed in numbers of LTR-active cells

To study the effects of Vpr on the responsiveness of individual proviruses in the absence of virus spread, Jurkat cells were infected with distinguishable *vpr*+ or *vpr*- versions of the NL4-3 derivative shown in Fig. 1A (44). Using the EF1α promoter to drive constitutive expression of the puromycin resistance gene allowed selection of infected cells independent of LTR expression and thus the retention of proviruses silenced shortly after infection. Each genomic RNA in the infecting virus contained a unique 20b randomized "barcode" inserted in U3, which was duplicated in both LTRs after reverse transcription and enabled tracking individual proviruses. We refer to the barcodes as "zip codes" because in the context of proviruses, they report the genetic neighborhoods of each integrant. Infected cell populations were passaged without cell cloning, thus generating polyclonal integrant populations, within which transcriptionally active cells were identified using LTR-driven GFP expression or by progeny genomic RNA in released Env- virions.

Previous work using zip coded *vpr*- derivatives of this vector has shown that for each integrant, the clone gives rise to a mixture of cellular progeny that includes some GFP+ cells and some GFP- cells (44), and that sibling cells switch between LTR-active (GFP+) and -inactive (GFP-) expression phenotypes in a phenomenon we called "flickering". To confirm that individual integrant clones in the current study also contained clone-specific mixtures of GFP+ and GFP- cells, single clones were isolated by limiting dilution from the *vpr*- pool and expanded, and cells from each clone were then subjected to flow cytometry. Consistent with previously reported differences in bimodal expression patterns among proviral clones, the results showed that each clonal pool was comprised of both GFP+ and GFP- cells, with GFP+ proportions that differed

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

among the clones: mostly GFP+ cells for some clones, and distinctly different GFP+ proportions for others (S1B Fig.).

In the current study, four independent polyclonal integrant pools were established by infecting roughly $5x10^6$ Jurkat T cells at a multiplicity of infection of <0.0005 to ensure infected cells contained one provirus per cell. Two of the pools contained *vpr*+ proviruses and two had integrants lacking *vpr*. Flow cytometry analysis after fourteen days of expansion showed significantly fewer GFP+ cells in the two *vpr*+ infected cell pools than in the two *vpr*- provirus pools ($p=0.042$, paired t-test; <5% vs. 73% GFP+, respectively Fig. 1B, S1A Fig.). These fourteen days post-infection cell pools were sorted into GFP+ and GFP- subpools, cellular DNA was harvested from an aliquot of the sorted cells immediately after sorting, and the zip codes of integrants were amplified from the cellular DNA and catalogued by high throughput sequencing.

After analyzing roughly three million sequencing reads per pool, zip codes were ordered by read abundance. Determining how many unique barcodes were present in each GFP+ pool revealed that similar *numbers* of zip codes (approximately 2000)—each indicative of a single integration event—were detected in the GFP+ sorted cells from both *vpr*+ and *vpr*- pools, even though the GFP+ fraction of total cells in the *vpr*+ pool was very low. This finding suggesting that the number of GFP+ integrants initially established did not differ markedly between pools is consistent with previous work in dendritic cells that indicated that the extent of proviral integration does not differ depending on the presence or absence of Vpr (45).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Combining population-wide GFP+ fractions in the unsorted cells with zip code read counts within sorted sub-pool libraries allowed calculating the proportion of LTR-active cells within each cell clone (Fig. 1C). Notably, for the top 500 most abundant clones in the unfractionated *vpr*+ and *vpr*- pools, median GFP+ proportions were significantly lower in the *vpr*+ pool (p=0.0001, Mann-Whitney U two-tailed test), indicating that most cells in the *vpr*+ pool were members of low LTR-active clones (Fig. 1C). When GFP+ sorted cells from *vpr*+ and *vpr*- pools were cultured further, no viable *vpr*+ GFP+ cells were detected after three days, and thus their clonal structures could not be analyzed further. These results indicated that integration events that resulted in GFP expression were equally likely in *vpr*+ and *vpr*- proviral pools when the cells were examined early post-infection. However, the depletion of the Vpr+ GFP+ cell sub-populations suggested that whenever cells flickered "on", the resulting GFP+ cells died upon further cell culturing.

Vpr shapes the clonal structure of infected pools

Observing fewer GFP+ cells in the *vpr*+ pool than in cells with *vpr*- proviruses was not surprising due to Vpr's well-known cytotoxic effects. However, a small fraction of *vpr*+ GFP+ cells were observed among unsorted *vpr*+ cells, even though sorted *vpr*+ GFP+ cells did not survive three days of culture.

One plausible reason unsorted *vpr*+ populations contained rare GFP+ cells was that these cells arose via recent phenotypic switches from clones that were largely inactive. To test this possibility, zip codes in GFP+ and GFP- sorted subpools were compared for both *vpr*+ and *vpr*- integrants (Fig. 1D). Ordering zip codes by their abundance and analyzing those that comprised the top 500 revealed that among *vpr*- cells, about 73.2% of all the zip codes were found in both subpools, while

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

4% were observed only in GFP- cells and 22.8% only in the GFP+ subpool. In contrast, 86.2% of the *vpr*+ cells' zip codes were found in both subpools, with 13.8% observed only in the GFP- subpool, and none of the zip codes found exclusively in the GFP+ subpool. This suggested that the small fraction of GFP+ cells in the *vpr*+ pool (Fig. 1B; left panel) resulted from the recent acquisition of LTR expression by cells from the larger GFP- cell pool. If the *vpr*- pool is assumed to be relatively representative of a population that results when all clones are equally viable, this suggests that when initially integrated, proviruses whose daughter cells were largely or always GFP+ were the dominant class of clones.

High LTR-active clones' proviruses are in close proximity to genome marks associated with active gene expression

The finding that LTR-active *vpr*+ cells were rapidly lost suggested that proviruses that had integrated into more active regions of the host genome might be selected against within polyclonal *vpr*+ populations. Therefore, we compared the proximities of *vpr*+ and *vpr*- integration sites to genomic features associated with active gene expression. First, zip code integration sites were determined using cellular DNA harvested ten days after establishment of each pool and mapped to 1171 and 1121unique sites in the human genome for *vpr*+ and *vpr*- respectively (see Methods). Then, the proportion of zip codes located in genes versus those not in genes (as defined by ENCODE for Jurkat cells (46) were compared (Fig. 2A, S2 Fig.). The results indicated that similar majorities of integrants were established within genes regardless of whether or not *vpr* was present. Next, *vpr*+ and *vpr*- integrants were compared for their proximity to specific genome marks associated with active gene expression that have been reported to pre-exist in Jurkat cells (47). No differences were found in distances to closest DNase I sensitivity sites, which are associated with

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

227 open chromatin (Mann-Whitney U test, p=0.1854) (Fig. 2B). However, proximities to H3K27ac
228 (associated with enhancers) (Mann-Whitney U two-tailed test, p<0.0001) and H3K4me3
229 (associated with active promoters) (Mann-Whitney U two-tailed test, p<0.0001) marks differed
230 significantly, with proviruses from *vpr*+ pools farther from these marks (Fig. 2C-D).
231
232 Next, integrants from *vpr*- pool were grouped into high LTR-active clones (GFP+ proportion >=
233 60%) and low LTR-active clones (GFP+ proportion < 30%) and compared for their proximities to
234 H3K27ac and H3K4me3 (Fig. 3A). The results showed that high LTR-active clones were closer
235 to both marks than low LTR-active clones and suggested that the difference in integration site
236 proximity observed between *vpr*+ and *vpr*- proviruses is due to the survival of high LTR-active
237 clones in the absence of Vpr.
238
239 Because the *vpr*+ pool was devoid of high LTR-active clones, its high- and low LTR-active
240 integrants could not readily be compared. However, with the reasoning that abundances of residual
241 LTR-active clones would gradually diminish over time, the dynamics of the v*pr*+ pool were
242 examined by comparing samples harvested two weeks apart (Fig. 3B). Examination of the 100
243 most abundant zip codes in the *vpr*+ cell pool showed that only 68 of the top 100 zip codes on day
244 10 post-infection (time point 1, or TP1) were observed among the 100 most abundant zip codes on
245 day 24. When integration site proximities to H3K4me3 and H3K27ac marks of these zip codes
246 were compared, those that were absent from the top-100 on day-24 (time point 2, TP2) (Fig. 3C)
247 were significantly closer to H3K4me3 (p=0.0193, Mann-Whitney U two-tailed test), and tended
248 to be closer to H3K27ac marks (p=0.1581, Mann-Whitney U two-tailed test) than the TP 1 clones
249 that remained within the top-100 most abundant clones at TP2.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

250

251 ***The spectra of LRA responses among vpr+ and vpr- clones differed***

252 Next, the reactivation properties of *vpr*- and *vpr*+ pools were compared and the behaviors of
253 individual proviral clones within the populations were determined. The latency reactivation agents
254 (LRAs) prostratin, a protein kinase C agonist, and JQ1, a bromodomain inhibitor, were applied
255 separately or in combination.

256

257 To examine reactivation in the total LTR-inactive cell sub-population of each pool, FACS-sorted
258 GFP- cells from both *vpr*+ and *vpr*- pools were subjected to prostratin (10 μM), JQ1 (2 μM), or
259 both drugs in combination for 24 hours. Reactivation was measured both by changes in the
260 frequency of GFP+ cells using flow cytometry and by the amount of virus release (Fig. 4A-B;
261 S3A-B Fig.). The results indicated that compared to single prostratin and JQ1 treatments, dual
262 treatment resulted in additive levels of reactivation in both *vpr*+ and *vpr*- populations by both
263 criteria compared to single prostratin or JQ1 treatment (Fig 4A shows reactivation monitored by
264 GFP+ cells and Fig. 4B shows virus release; left panel indicates reactivation for *vpr*+ and right
265 panel for *vpr*-). In dual LRA treatments for both pools, there was an approximately 4-fold increase
266 in GFP+ cells, while virus release increased by approximately 30-fold relative to DMSO control.
267 The most significant difference between *vpr*- and *vpr*+ pools was that the absolute amount of virus
268 release was 3-fold higher in *vpr*- pools, and the responsiveness to JQ1 was lower in the *vpr*+ pools.
269 The differences between pools in their extents of reactivation were not due to differences in cell
270 viability (S3C Fig.). The observation that reactivation was enhanced by dual prostratin and JQ1
271 treatment is consistent with previous works by Boehm *et al.,* and Darcis *et al.,* using the same

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

drugs in cell culture models of latency and ex vivo treated cells from HIV patients respectively (21, 48).

To investigate the effects of the treatments on individual GFP- clones, total virus release was quantified by p24 equivalent and cDNA was generated using viral genomic RNA upon LRA treatment. Zip codes were amplified from the viral cDNA and also from an untreated aliquot of the GFP- cellular DNA and high through-put sequenced. The results were normalized to calculate average virus release per treated cell for each clone (Fig. 4C; S3D Fig.). Analysis revealed that when cells were treated with both drugs, some clones in both the *vpr*+ and *vpr*- pools displayed enhanced virus release per treated cell compared to single treatment conditions. Interestingly, both proviral pools included a subset of clones that were not detectably reactivated by either prostratin or JQ1 when the LRAs were applied alone, but that were reactivated upon dual LRA treatment. Surprisingly, and only in the *vpr*- pool, reactivation levels observed under dual LRA conditions were lower than those observed with single LRA use for a subset of clones (Fig 4C: right panel. Compare, for example, *vpr*- JQ1 and Pros +JQ1 columns). These same patterns of reactivation were observed when experiments were repeated using a second set of independently established *vpr*- and *vpr*+ pools (S3D Fig.).

The finding that some clones were detectably reactivated by dual LRA treatment but not by either prostratin or JQ1 alone was interesting. A possible explanation for this was that these clones were in low abundance. If rare clones released little virus when prostratin and JQ1 were applied separately, their release might only rise to a detectable level in the cDNA library when these drugs were combined. However, the results showed that these were not rare clones. For example, the

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

most abundant clone in the GFP- *vpr*- library (pool 2; clone number 3), which accounted for about 15% of the treated GFP- pool cells (Supplementary Table), was one of the clones for which detectable virus was released only upon dual LRA treatment.

Complementation of the vpr-defective pool led to depletion of high LTR-active clones

As an additional test of the effects of Vpr on polyclonal population composition, a functional copy of the *vpr* gene was added to cells harboring *vpr*- proviruses 16 days post-infection. This was achieved by transducing the *vpr*- pool with a *tat*-deficient lentiviral vector containing LTR driven *vpr,* which also contained the fluorescent marker *mKO* expressed from the constitutive SV40 promoter. This resulted in expression of Vpr only when a pre-existing *vpr*- provirus was transcriptionally active. mKO expression was used as a proxy for *vpr* vector transduction, and enabled cells that were successfully transduced to be sorted in the phycoerythrin (PE) channel (Fig. 5A). 48 hours post-transduction, an unsorted aliquot was saved while the remaining cells were sorted to identify GFP- PE+ cells (S4 Fig.A-B) for work described below.

Comparing the proportions of GFP+ cells in the parental *vpr*- pool to the proportions in the same pool "reverted" *to vpr*+ by introduction of the *vpr* vector (hereafter referred to as the vpr^{rev} pool) showed a marked decrease in GFP+ cells between day-2 and day-14 post transduction (Fig. 5B). This selective loss of GFP+ transduced cells corroborated observations with the *vpr*+ pool that LTR-active cells were depleted in the presence of Vpr.

Proviral zip code abundances in unsorted *vpr*- and vpr^{rev} pools were then compared. Of the 500 most abundant clones in the *vpr*- pool, zip codes were split into two groups based on observed

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

fold-change in relative abundance: those that were reduced by 10-fold or more in *vpr*^{rev} (reduced clones), and those with no observed reduction in their relative abundance upon *vpr* addition (not reduced). A comparison between the "reduced" and "not reduced" groups revealed that nearly all *vpr*- pool clones that displayed high GFP+ proportions--that is, clones in which most member cells displayed LTR activity-- were reduced in the *vpr*^{rev} pool (Fig. 5C).

The *vpr*^{rev} pool was then used to address whether *vpr*- pools rendered *vpr*+ by complementation displayed reactivation patterns similar to those of the original *vpr*+ pools. Transduced GFP- *vpr*^{rev} subpools were subjected to prostratin, JQ1, or dual prostratin-JQ1 treatment and analyzed by flow cytometry and virus release (Fig. 5D). In general, the magnitude of reactivation in the transduced cells was diminished across all treatment conditions relative to parental *vpr*- GFP- cells, with cells responding less to JQ1 treatment than they did before Vpr addition. Consistent with expectations if the phenotypes above reflected Vpr, transduced cells showed enhanced reactivation upon dual LRA treatment (Fig. 5E; S4C Fig.). Notably, clones (such as clones numbered 53, 133, 161, 293, 420, etc.) that were highly responsive to JQ1 treatment in the parental untransduced *vpr*- pool, and that differentiated the LRA responsiveness of the *vpr*- pool from the *vpr*+ pool, were highly represented among those lost upon Vpr addition.

Genomic features of persistent intact proviruses in patients on ART are more similar to vpr+ than vpr- pool members

The proviruses in HIV patients on ART contain intact Vpr and have survived its selective pressures. Having observed differences between *vpr*+ and *vpr*- viruses in cultured cells in terms

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

of their reactivation patterns and proximity to genome features, we sought to determine which class of these proviruses more closely resembled those in persistent clinical isolates.

To do this, previously published data on patients' intact provirus integration sites (49) were analyzed for their proximity to H3K4me3, H3K27ac, and DNase I hypersensitivity sites that have been reported for Jurkat T cells and primary CD4+ T cells (see Methods) (47). When these proximities were compared to those in the *vpr*- and *vpr*+ pools established here, prominent proviruses in the *vpr*- pool, and not those in the *vpr*+ pool, were found to significantly differ from the patient proviruses in their proximities to H3K4me3 and H3K27ac marks (Fig 6. A-B; S5 Fig.), but not DNase I hypersensitivity sites (Fig. 6C). These results suggest that the integration site distribution of persistent proviruses in patients across the tested genome marks was similar to those observed in *in vitro* established *vpr*+ proviral populations but significantly different from what was observed within *vpr*- proviral pools.

Discussion

This work describes how Vpr shapes the overall provirus population structure, dominant host integration site landscape, and responsiveness to latency reversing agents in cultured cells. Cell-based models are critical to latency reactivation agent drug discovery and other HIV persistence work, but these models differ by cell type, form of HIV-1 used, whether they use isolated clones or polyclonal populations, and other parameters. Studies using polyclonal pools may implicitly assume that the pools contain an equivalent proportion of each proviral clone. However, as demonstrated here and previously, even when clonal target cell lines are infected, pools rapidly become dominated by a subset of proviral clones (44). Thus, population-wide trends may not

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

apply to all clones. Similarly, because proviral clones differ in expression patterns, clonal latency models' HIV-1 expression properties may represent the biases of the investigators who chose the clones more than selective pressures that shape the latent reservoir.

Primary CD4+ T cells are the closest tissue culture system to natural infection since CD4+ T cells are the primary target of HIV infection. However, these cells only survive 2-4 weeks *ex vivo* unless treated with antiapoptotic agents (38, 50) and do not proliferate unless stimulated. We previously compared expression patterns of zip coded proviruses in primary CD4+ T cells to those in Jurkat T cells, and observed indistinguishable spectra of expression patterns (44). However, extended cell passaging was required in the current work. Thus, we chose to use Jurkat cells, which limits the physiologic relevance of our findings, although Vpr has been noted to exert similar effects in primary CD4+ and Jurkat T cells (51).

Because identifying latent cells is challenging, many experimental latency systems include reporter genes and/or delete viral genes believed to be unnecessary for silencing and reactivation (33, 34, 52). For example, two prominent studies that used barcoded proviruses to track virus dissemination in animals used *vpr*- proviruses (33, 52). The impact of this variation is not well understood.

The zip coded provirus approach used here enabled addressing Vpr effects by comparing *vpr*+ and *vpr*- pools. *vpr*+ pools were found to contain significantly fewer GFP+ cells (indicative of transcriptionally active proviruses) than did *vpr*- pools, but similar numbers of proviral clones were detectable in both pools soon after their establishment. Consistent with previous findings (44), for

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

some *vpr*- integrant cell clones, essentially all member cells displayed HIV gene expression (which we designated high LTR-active clones); for other clones most cells did not express HIV (low LTR-active clones), and a smaller number of clones displayed similar numbers of active and inactive member cells. In contrast, within the most abundant *vpr*+ clones, most clonal member cells did not express HIV genes (low LTR-active clones). These findings may help explain why estimates of the transcriptionally active fraction of HIV-1 infected cells vary among previous studies. Using polyclonal *vpr*+ proviruses *in vitro*, Dahabieh *et al*. suggested that most integrated proviruses are latent, even though it has been estimated that latent cells represent only a small fraction of total infected cells in patients (29, 36, 53). Consistent with Dahabieh *et al.,* we found that HIV is transcriptionally inactive in most infected cells that persisted in *vpr*+ populations, but our additional work suggested this reflected the selective survival of low LTR-active clones.

Previous work has shown that within proviral clones, individual cells can shift from being silenced to expressing their proviruses, while maintaining the overall proportions of their daughter cells that possess active proviruses (37, 44, 54). For example, when barcodes are used to examine *vpr*- proviral clones within polyclonal infected cell populations, individual member cells within each clone either express HIV genes, as monitored by GFP reporter, p24 staining, or virion release, or they do not express HIV. Although some diminution of LTR expression is observed over time, for the most part each clone adopts a stable, heritable pattern of bimodal gene expression (44). The current work showed *vpr*+ and *vpr*- polyclonal populations did not differ upon integration but differed significantly in proportion of LTR-active cells after passaging. Together, these observations suggest that the low proportion of transcriptionally active cells in *vpr*+ integrant pools might reflect the selective depletion of high LTR-active clones (Fig.7); a notion that is further

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

409 supported by observations in *vpr*^{rev} pools, which were *vpr*- pools that were complemented with a
410 Vpr expression vector. According to this model, although the proportion of total cells that were
411 GFP+ was significantly lower in polyclonal *vpr*+ integrant populations than in *vpr*- proviral pools
412 at 14 days, the spectrum of *vpr*+ integration sites was likely indistinguishable from *vpr*- clones
413 upon initial establishment.
414
415 A little-discussed problem with transcriptionally silenced provirus models is that they tend to
416 "spontaneously revert" during passage and generate small proportions of cells with
417 transcriptionally active proviruses (37, 55). Our previous work and findings here suggest an
418 explanation slightly different from "spontaneous reversion" for the appearance of LTR-active cells
419 in passaged latency model cells. As evident from quantifying phenotypic switching within
420 expression-sorted polyclonal subpopulations, the phenotypes of active- and inactive-
421 subpopulations of each clone are not static, but instead equilibrate in a phenomenon we call
422 flickering. This suggests most LTR-inactive cells are not stably silenced but rather represent the
423 transient off-state of an on-off continuum. If so, "spontaneous reversion" by a small subset of
424 latency model cells, like the low proportion of GFP+ cells in the current study's *vpr*+ pool, might
425 more accurately be viewed as the programmed equilibration of a low LTR-active clone than the
426 outlier behavior in a subset of silenced cells.
427
428 Comparing integrants that dominated passaged *vpr*+ and *vpr*- pools suggested Vpr selects against
429 proviruses proximal to genome marks associated with active gene expression. Consistent with
430 previous work, *vpr*+ and *vpr*- pools showed no difference in preference for integration into actively
431 transcribed genes (56-58). However, predominant *vpr*+ proviruses, which represented clones with

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the largest numbers of daughter cells, were farther from the closest enhancer or promotor than were predominant *vpr*- proviruses. Importantly, zip codes of high LTR-active clones were closer to enhancers and promoters than those that had low GFP+ proportions. Furthermore, these clones diminished over time in the *vpr*+ pool and were disproportionately depleted in vpr^{rev} pools, upon the addition of *vpr* to the *vpr*- pools.

We also compared the effects of LRAs on *vpr*+ and *vpr*- proviral populations as well as individual clones. LRA responses vary among latency models (35), thus suggesting that combining mechanistically distinct LRAs might broaden treatment efficacy (10, 40, 59, 60). Here, when treatment with prostratin, which increases NFkB, was compared to JQ1, which enhances pTEFb availability, individual clones were affected to differing extents. Contrary to the expectation that dual treatment would enhance reactivation, prostratin plus JQ1 dual treatment suppressed some proviruses in the *vpr*- pool but did not display this phenotype in the *vpr*+ pool. The suppressed proviruses tended to be relatively close to promoters and enhancers, possibly suggesting a transcriptional interference role in this effect (3, 61). The adoption by the vpr^{rev} pools of the *vpr*+ pool's patterns of responsiveness to LRAs further corroborated the role of Vpr: even in this very limited analysis of two drugs' effects.

Analysis here suggested a correlation between LTR-active clones and integration positions relative to certain epigenetic marks. To address the relationship of these observations to persistent populations *in vivo*, where survival during cell proliferation shapes the reservoir (62), we mapped distances to pre-existing marks for previously published integration sites of intact, inferred to be intact, or large proviruses in three patients on antiretroviral therapy (ART) (49), and compared this

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

to integration site distances in our *vpr*+ and *vpr*- pools. This analysis indicated that proximities of *vpr*+ proviruses to the tested features were significantly more similar to those observed in patients than were *vpr*- pool members. This suggests that the observed effect of *vpr* in our *in vitro* system may also contribute to shaping the persistent latent reservoir in patients. If so, existing animal studies that have studied latency and reactivation using *vpr*- proviruses may largely be tracking the properties of proviruses that are eliminated during natural infection (52).

Materials and Methods

Tissue culture of cell lines

Two immortalized mammalian cell lines were used in this study. Human Embryonic Kidney (HEK) 293T and Jurkat cells were purchased from ATCC (Cat No. CRL3216 and TIB-152 respectively). Both cell lines were preserved as frozen stocks. 293T cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% Fetal Bovine Serum (FBS) (Gemini) and 125 µM gentamycin. Jurkat cells were cultured in RPMI media supplemented with 10% FBS, 100 U/ml penicillin, 100 µg/ml streptomycin, 2 mM glutamine and 55 µM β-mercaptoethanol. To propagate cells, frozen 1 ml aliquots of each cell line were thawed rapidly in 37° C water bath and added to 9 ml of their respective pre-warmed media. After mixing gently, cells were centrifuged at 2500 rpm for 5 minutes. Supernatant was discarded and the cell pellet was resuspended gently in pre-warmed media and plated in an appropriately sized tissue

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

478 culture plate. HEK 293T cells were sub-cultured when confluency was between 75-100%. Jurkat
479 cells were passaged such that the cell concentration was maintained between 8 x 10^5 and 1 x 10^6
480 cells/ml.
481
482 ***Zip coded vector and virus production***

483 Zip coded vector DNA templates were generated by digestion of previously described modified
484 "inside-out" forms of GKO, which here are called GPV+ and GPV$^-$ (44, 63), with ClaI and MluI.
485 The resulting 11.4 kb DNA fragment, which was devoid of plasmid backbone, was gel purified
486 using QIAquick Gel Extraction Kit (Cat No./ID: 28706 Qiagen, Germantown, MD). A 304 bp
487 insert fragment was generated by PCR with with GPV$^+$ or GPV$^-$ as template for *vpr*+ and *vpr*-
488 respectively, Phusion High-Fidelity DNA Polymerase (New England Biolabs, Inc., Ipswich, MA),
489 and primers 5'-GACAAGATATCCTTGATCTGNNN
490 NNNNNNNNNNNNNNNNGCCATCGATGTGGATCTACCACACACAAGGC-3' and 5'-
491 CGGTGCCTGATTAATTAAACGCGTGCTCGAGACCTGGAAAAAC-3'. The 11.4 kb and
492 304 bp degenerate barcode containing fragments were joined by Gibson assembly using HiFi DNA
493 assembly mix (New England Biolabs) in a molar ratio of 1:5. 3 µg of the resulting covalently
494 closed circular DNA was directly co-transfected with 330 ng of pHEF-VSV-G (Addgene Plasmid
495 #22501) into 70% confluent monolayer of 293T cells in a 10 cm dish using polyethylenimine
496 (Polysciences Inc, Warrington, PA.) at 4X total transfected DNA in 800 µl 150 mM NaCl. DMEM
497 was replaced with 4 ml RPMI1640 medium with 10% FBS and 1% Pen/strep 24 hours post
498 transfection and culture media was harvested by filtering through a 0.22 µm filter (Fisher
499 Scientific. Cat. No. 09-720-511) 48 hours post-transfection.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Generation of zip coded proviral pools

1000 µl virus-containing media was mixed with polybrene at a final concentration of 0.5 µg/ml and brought to a total volume of 2000 µl by addition of RPMI. This infection mixture was added to 5 x 10^6 Jurkat cells and incubated in two wells of a 12-well plate at 37 C with 5% CO_2 for 5 hours. Infected cells were then transferred to 10 ml falcon tubes and centrifuged for 5 minutes at 2500 rpm at 4℃. Following centrifugation, supernatants were replaced with fresh media and cell pellets were resuspended and cultured in two wells of a 6-well plate at 37℃ with 5% CO_2. At 24 hours post-infection, puromycin was added to a final concentration of 0.5 µg/ml. The infected cells were expanded into 6 cm culture plates without puromycin on day 5. Ten days post-infection, the culture supernatant was replaced with fresh media and the cultures were divided into aliquots, to be either frozen, prepared for integration-site sequencing or further expanded for four additional days and sorted into GFP+ and GFP- subpools for subsequent experiments.

Construction of vpr+ lentiviral vector and its use

A 2933 bp in vitro DNA synthesized *IRES-vpr mKO* fragment was ordered from Genewiz (South Plainfield, NJ) using an IRES sequence from pTRIPZ-hDDX5/7 (Addgene Plasmid #71307) (64) and *vpr* and *mKO* from GPV^+. An HIV-1 lentiviral vector fragment was generated by XbaI and MfeI digestion of pWA18puro (65) to remove its puromycin resistance cassette, and was gel purified using QIAquick Gel Extraction Kit (Cat No./ID: 28706 Qiagen, Germantown, MD). The resulting DNA fragments were joined by Gibson assembly using HiFi DNA assembly mix (New England Biolabs) following the manufacturer's protocol to generate pEA216-1. 10 µg of pEA216-

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1 was then co-transfected with 5 µg of the pCMVΔR8.2 helper plasmid (Addgene Plasmid #122263) and 1 µg of pHEF-VSV-G into 70% confluent monolayers of 293T cells in a 10 cm dish using polyethylenimine (Polysciences Inc, Warrington, PA.) at 4X total transfected DNA in 800 µl 150 mM NaCl. DMEM was replaced with 4 ml RPMI1640 medium with 10% FBS and 1% Pen/strep 24 hours post transfection and culture media was harvested by filtering through a 0.22 µm filter (Fisher Scientific. Cat. No. 09-720-511) 48 hours post-transfection. Parental *vpr*- pools were infected with this filtered media and sorted 48 hours post-transduction to generate *vpr*-^{rev} pools.

In determining the effect of Vpr addition on parental *vpr*- pools, zip code abundances in sequences amplified from genomic DNA of parental *vpr*- GFP+ cells and from Vpr-transduced *vpr*^{rev} GFP+ cells were compared. Two categories of zip codes were defined: reduced zip codes were those that were reduced 10-fold or more after Vpr transduction, whereas those with a change of 1.5-fold or less in their relative abundances were regarded as not affected by *vpr* addition (Not reduced).

Flow cytometry and cell sorting

For Fluorescence Activated Cell Sorting (FACS) analysis by flow cytometry, Jurkat cells were suspended in phosphate buffered saline (PBS) containing 1% FBS (FACS buffer). Dead cells were excluded from all analyses and sorting experiments using propidium iodide (PI). Acquisition was carried out on the FITC channel for GFP and PE channel for PI. Cell fluorescence was assessed using BD LRS Fortessa (BD Biosciences) and data were analyzed using FlowJo software, version 10.6 (FlowJo, LLC., Ashland, Oregon). Infected cells were sorted into GFP+ and GFP- sub-populations by flow cytometry using FACS Aria II (BD Biosciences, Franklin Lakes, NJ) or iCyt

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

544 Synergy SY3200 (Sony Biotechnology, San Jose, CA) cell sorters at the flow cytometry core of
545 the University of Michigan.

546 ***Latency reversing agents and reactivation***

547 JQ1 and prostratin were purchased from Sigma Aldrich. Each LRA was dissolved in DMSO
548 (Thermofisher) to produce stocks. For each experiment, stocks were added to culture medium to
549 achieve 2 μM JQ1 and 10 μM prostratin final concentrations. Dual LRA treatment was done by
550 adding the two LRAs to the same culture medium to achieve their respective single LRA
551 concentrations. For reactivation experiments, GFP- cells sorted on day 14 post-infection were co-
552 cultured with appropriate LRA for 24 hours. Cells were then centrifuged at 2000rpm, 4^{o}C for 5
553 mins, cells pellets were washed twice with ice cold FACS buffer after being stained with PI for 5
554 mins at room temperature. The resulting cells were then washed and assessed by flow cytometry
555 and the p24 concentration of culture supernatants was measured by a reverse transcriptase assay.
556

557 ***Zip code sequencing libraries***

558 Zip codes were amplified from the genomic DNA of infected cells as well as from the RNA of
559 virus released into cell media. Generation of zip coded sequencing libraries from infected cells
560 was initiated by harvesting DNA from an aliquot of 2×10^{6} infected cells. Genomic DNA extraction
561 was carried out using the Qiagen DNeasy Blood & Tissue Kit (Qiagen, Germantown, MD). Zip
562 codes were then amplified by polymerase chain reaction (PCR) from 200 ng of DNA template
563 using Phusion ® High-Fidelity (HF) DNA Polymerase (New England Biolabs) in HF Buffer.
564 Primers were designed to flank the zip code region (primer sequences: 5'-
565 NNACGAAGACAAGATATCCTTGATC-3' and 5'-NNTGTGTGGTAGATCCACATCG-3').
566 Multiple copies of these primers were created, each with a unique pair of known, randomized

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

nucleotides at the 5' end to confirm that no cross-contamination had occurred between samples. Reactions were cycled for 29 times with 30 second extension at 72°C and a 59°C annealing temperature. Zip code amplicons were purified using the DNA Clean and Concentrator-5 kit (Zymo Research, CA. Cat. No. D4013) and eluted in 15 µL of Milli-Q ® H_2O. To amplify zip codes from virus, a tissue culture plate of infected cells was decanted into a conical tube and centrifuged at 2500 rpm for 5 minutes. The virus-containing media was separated from the cell pellet and passed through a 0.22 µm filter. To concentrate virus, viral media was subjected to ultra-centrifugation (25,000 rpm) for 120 minutes through a 20% sucrose cushion. Viral pellets were then resuspended in 200 µl PBS and viral RNA was extracted using Quick-RNA Viral Kit (Zymo Research, CA. Cat. No. R1034 & R1035) following the manufacturer's protocol and eluted in 10 µl RNase-free water. cDNA was synthesized using 5 µl of the eluent as template using the U3 antisense primer 5' TGTGTGGTAGATCCACATCG-3' and M-MLV RT RNase [H-] (Promega, WI. Cat. No. MR3681) following the manufacturer's protocol. Zip codes were amplified from this cDNA using conditions above. The zip code amplicons were then used to generate MiSeq libraries for sequencing as described previously (44).

Integration site sequencing

Genomic DNA was extracted from *vpr*+ and *vpr*- cells 10 days post-infection using Qiagen DNeasy Blood & Tissue kit (Qiagen) and 200 ng of DNA was sheared to 1 kb fragments using Covaris M220 and micro- TUBE according to the manufacturer's recommended settings (Covaris, Woburn, MA). HIV-1 insertion site libraries were prepared and sequenced using methods described previously (44). All sequence data has been deposited to the Sequence Read Archive (SRA) under project accession PRJNA76782.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

590

591 ***Quantification of virus release***

592 Virions production was quantified using a real-time reverse-transcription PCR assay previously
593 described by Kharytonchyk et al (65). Briefly, viral lysates were prepared by adding 5 µl of culture
594 supernatant to 5 µl of lysis buffer. Using MS2 RNA as template, MS2 cDNA was synthesized with
595 viral lysates and quantified by real-time PCR in one reaction. Released virus was quantified and
596 normalized for p24 level based on values determined in parallel for reference samples.

597

598 ***Zip code quantification and analysis***

599 Zip codes were identified and quantified from Illumina sequencing reads using a previously
600 described custom suite of tools implemented in Python (https://github.com/KiddLab/hiv-zipcode-
601 tools). Briefly, 2x75 bp paired reads were merged together using *flash* v1.2.11 [73]. Zip codes
602 were identified by searching for known flanking sequence (with up to 1 mismatch). Only candidate
603 zip codes with a length of 17–23 nucleotides were considered and the read count for each unique
604 zip code was tabulated. To identify sets of zip codes for further analysis, zip code families, which
605 account for PCR and sequencing errors, were determined by clustering together the observed
606 unique zip codes. Abundances for the zip codes were then determined by assigning unique zip
607 codes to the most abundant family whose sequence was within 2 mismatches and summing their
608 associated read counts. Only zip code families with corresponding data in the integration site data
609 were selected for further analysis.

610

611 For each latency reversal treatment condition, clonal virus release was determined by multiplying
612 the fractional abundances of zip codes from the cDNA sequencing libraries of each treatment by

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

613 their corresponding sample's pool p24 concentration, as measured 24 hours post LRA treatment.
614 The resulting clonal p24 values for each zip code in prostratin, JQ1, and combination prostratin-
615 JQ1 treated samples were divided by clonal p24 values defined for cell samples exposed only to
616 DMSO to determine fold change after LRA treatment.
617
618 ***Determination of chromatin marks***
619 H3K27ac marks annotated for the Jurkat E-6-1 clone were sourced from http://chip-
620 atlas.org/view?id=SRX1041803, H3K4me3 and DNase I sensitivity sites datasets were
621 downloaded from the encode project (https://www.encodeproject.org/) with the identifiers
622 SRX1041803, and ENCFF304GVP for genome marks pre-existing in Jurkat cells, and
623 iENCFF341XUX, ENCFF053LHH (47). Bedtools was then used to map the distance to the closest
624 known annotated marks. Analysis of proximity to these genome marks was done using matplotlib
625 and scipy.stats packages in python and the results were exported into Graphpad Prism version
626 9.1.2 to plot graphs.
627
628 Acknowledgements
629 We thank Stephen Goff, Akira Ono and Joel Swanson for helpful comments on the manuscript.
630 This work was supported by NIH/NIAID grant number R33 AI116190 to AT and JMK and by a
631 pilot award to AT. from the Rogel Cancer Center NIH/NCI grant number P30CA046592.
632
633 **References and Notes**
634 1. Bailey J, Blankson JN, Wind-Rotolo M, Siliciano RF. Mechanisms of HIV-1 escape from
635 immune responses and antiretroviral drugs. Current opinion in immunology. 2004;16(4):470-6.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

636 2. Lenasi T, Contreras X, Peterlin BM. Transcriptional interference antagonizes proviral gene
637 expression to promote HIV latency. Cell host & microbe. 2008;4(2):123-33.
638 3. Han Y, Lin YB, An W, Xu J, Yang H-C, O'Connell K, Dordai D, Boeke JD, Siliciano JD,
639 Siliciano RF. Orientation-dependent regulation of integrated HIV-1 expression by host gene
640 transcriptional readthrough. Cell host & microbe. 2008;4(2):134-46.
641 4. Lassen K, Han Y, Zhou Y, Siliciano J, Siliciano RF. The multifactorial nature of HIV-1
642 latency. Trends in molecular medicine. 2004;10(11):525-31.
643 5. Williams SA, Greene WC. Regulation of HIV-1 latency by T-cell activation. Cytokine.
644 2007;39(1):63-74.
645 6. Kulkosky J, Sullivan J, Xu Y, Souder E, Hamer DH, Pomerantz RJ. Expression of latent
646 HAART-persistent HIV type 1 induced by novel cellular activating agents. AIDS research and
647 human retroviruses. 2004;20(5):497-505.
648 7. Mbonye U, Wang B, Gokulrangan G, Shi W, Yang S, Karn J. Cyclin-dependent kinase 7
649 (CDK7)-mediated phosphorylation of the CDK9 activation loop promotes P-TEFb assembly with
650 Tat and proviral HIV reactivation. Journal of Biological Chemistry. 2018;293(26):10009-25.
651 8. Archin NM, Margolis DM. Attacking latent HIV provirus: from mechanism to therapeutic
652 strategies. Current Opinion in HIV and AIDS. 2006;1(2):134-40.
653 9. Verdin E, Paras P, Van Lint C. Chromatin disruption in the promoter of human
654 immunodeficiency virus type 1 during transcriptional activation. The EMBO journal.
655 1993;12(8):3249-59.
656 10. Blazkova J, Trejbalova K, Gondois-Rey F, Halfon P, Philibert P, Guiguen A, Verdin E,
657 Olive D, Van Lint C, Hejnar J. CpG methylation controls reactivation of HIV from latency. PLoS
658 pathogens. 2009;5(8):e1000554.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

659 11. Emiliani S, Fischle W, Ott M, Van Lint C, Amella CA, Verdin E. Mutations in the tat gene
660 are responsible for human immunodeficiency virus type 1 postintegration latency in the U1 cell
661 line. Journal of virology. 1998;72(2):1666-70.
662 12. Emiliani S, Van Lint C, Fischle W, Paras P, Ott M, Brady J, Verdin E. A point mutation in
663 the HIV-1 Tat responsive element is associated with postintegration latency. Proceedings of the
664 National Academy of Sciences. 1996;93(13):6377-81.
665 13. Deeks SG. HIV: Shock and kill. Nature. 2012;487(7408):439.
666 14. Filippakopoulos P, Qi J, Picaud S, Shen Y, Smith WB, Fedorov O, Morse EM, Keates T,
667 Hickman TT, Felletar I. Selective inhibition of BET bromodomains. Nature.
668 2010;468(7327):1067-73.
669 15. Zhu J, Gaiha GD, John SP, Pertel T, Chin CR, Gao G, Qu H, Walker BD, Elledge SJ, Brass
670 AL. Reactivation of latent HIV-1 by inhibition of BRD4. Cell reports. 2012;2(4):807-16.
671 16. Contreras X, Schweneker M, Chen C-S, McCune JM, Deeks SG, Martin J, Peterlin BM.
672 Suberoylanilide hydroxamic acid reactivates HIV from latently infected cells. Journal of
673 Biological Chemistry. 2009;284(11):6782-9.
674 17. Edelstein LC, Micheva-Viteva S, Phelan BD, Dougherty JP. Activation of latent HIV type
675 1 gene expression by suberoylanilide hydroxamic acid (SAHA), an HDAC inhibitor approved for
676 use to treat cutaneous T cell lymphoma. AIDS research and human retroviruses. 2009;25(9):883-
677 7.
678 18. Jiang G, Mendes EA, Kaiser P, Sankaran-Walters S, Tang Y, Weber MG, Melcher GP,
679 Thompson III GR, Tanuri A, Pianowski LF. Reactivation of HIV latency by a newly modified
680 Ingenol derivative via protein kinase Cδ–NF-κB signaling. AIDS (London, England).
681 2014;28(11):1555.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

682 19. Archin NM, Espeseth A, Parker D, Cheema M, Hazuda D, Margolis DM. Expression of
683 latent HIV induced by the potent HDAC inhibitor suberoylanilide hydroxamic acid. AIDS research
684 and human retroviruses. 2009;25(2):207-12.
685 20. Eriksson S, Graf EH, Dahl V, Strain MC, Yukl SA, Lysenko ES, Bosch RJ, Lai J, Chioma
686 S, Emad F. Comparative analysis of measures of viral reservoirs in HIV-1 eradication studies.
687 PLoS pathogens. 2013;9(2):e1003174.
688 21. Darcis G, Kula A, Bouchat S, Fujinaga K, Corazza F, Ait-Ammar A, Delacourt N, Melard
689 A, Kabeya K, Vanhulle C. An in-depth comparison of latency-reversing agent combinations in
690 various in vitro and ex vivo HIV-1 latency models identified bryostatin-1+ JQ1 and ingenol-B+
691 JQ1 to potently reactivate viral gene expression. PLoS pathogens. 2015;11(7):e1005063.
692 22. Pace MJ, Agosto L, Graf EH, O'Doherty U. HIV reservoirs and latency models. Virology.
693 2011;411(2):344-54.
694 23. Jordan A, Bisgrove D, Verdin E. HIV reproducibly establishes a latent infection after acute
695 infection of T cells in vitro. The EMBO journal. 2003;22(8):1868-77.
696 24. Eckstein DA, Sherman MP, Penn ML, Chin PS, De Noronha CM, Greene WC, Goldsmith
697 MA. HIV-1 Vpr enhances viral burden by facilitating infection of tissue macrophages but not
698 nondividing CD4+ T cells. The Journal of experimental medicine. 2001;194(10):1407-19.
699 25. Kino T, Gragerov A, Slobodskaya O, Tsopanomichalou M, Chrousos GP, Pavlakis GN.
700 Human immunodeficiency virus type 1 (HIV-1) accessory protein Vpr induces transcription of the
701 HIV-1 and glucocorticoid-responsive promoters by binding directly to p300/CBP coactivators.
702 Journal of virology. 2002;76(19):9724-34.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

703 26. Sato K, Misawa N, Iwami S, Satou Y, Matsuoka M, Ishizaka Y, Ito M, Aihara K, An DS,
704 Koyanagi Y. HIV-1 Vpr accelerates viral replication during acute infection by exploitation of
705 proliferating CD4+ T cells in vivo. PLoS pathogens. 2013;9(12):e1003812.
706 27. Zhang Q, Kang Y, Wang S, Gonzalez GM, Li W, Hui H, Wang Y, Rana TM. HIV
707 reprograms host m6Am RNA methylome by viral Vpr protein-mediated degradation of PCIF1.
708 Nature Communications. 2021;12(1):1-13.
709 28. Zhang F, Bieniasz PD. HIV-1 Vpr induces cell cycle arrest and enhances viral gene
710 expression by depleting CCDC137. Elife. 2020;9:e55806.
711 29. Yao X-J, Mouland AJ, Subbramanian RA, Forget J, Rougeau N, Bergeron D, Cohen EA.
712 Vpr stimulates viral expression and induces cell killing in human immunodeficiency virus type 1-
713 infected dividing Jurkat T cells. Journal of virology. 1998;72(6):4686-93.
714 30. Bauby H, Ward CC, Hugh-White R, Swanson CM, Schulz R, Goujon C, Malim MH. HIV-
715 1 Vpr Induces Widespread Transcriptomic Changes in CD4+ T Cells Early Postinfection. mBio.
716 2021;12(3):e01369-21.
717 31. Chen H-C, Martinez JP, Zorita E, Meyerhans A, Filion GJ. Position effects influence HIV
718 latency reversal. Nature Structural and Molecular Biology. 2017;24(1):47.
719 32. Jordan A, Defechereux P, Verdin E. The site of HIV-1 integration in the human genome
720 determines basal transcriptional activity and response to Tat transactivation. The EMBO journal.
721 2001;20(7):1726-38.
722 33. Fennessey CM, Pinkevych M, Immonen TT, Reynaldi A, Venturi V, Nadella P, Reid C,
723 Newman L, Lipkey L, Oswald K. Genetically-barcoded SIV facilitates enumeration of rebound
724 variants and estimation of reactivation rates in nonhuman primates following interruption of
725 suppressive antiretroviral therapy. PLoS pathogens. 2017;13(5):e1006359.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

34. Jefferys SR, Burgos SD, Peterson JJ, Selitsky SR, Turner A-MW, James LI, Tsai Y-H, Coffey AR, Margolis DM, Parker J. Epigenomic characterization of latent HIV infection identifies latency regulating transcription factors. PLoS pathogens. 2021;17(2):e1009346.

35. Spina CA, Anderson J, Archin NM, Bosque A, Chan J, Famiglietti M, Greene WC, Kashuba A, Lewin SR, Margolis DM. An in-depth comparison of latent HIV-1 reactivation in multiple cell model systems and resting CD4+ T cells from aviremic patients. PLoS pathogens. 2013;9(12):e1003834.

36. Dahabieh MS, Ooms M, Simon V, Sadowski I. A doubly fluorescent HIV-1 reporter shows that the majority of integrated HIV-1 is latent shortly after infection. Journal of virology. 2013;87(8):4716-27.

37. Battivelli E, Dahabieh MS, Abdel-Mohsen M, Svensson JP, Da Silva IT, Cohn LB, Gramatica A, Deeks S, Greene WC, Pillai SK. Distinct chromatin functional states correlate with HIV latency reactivation in infected primary CD4+ T cells. Elife. 2018;7:e34655.

38. Xing S, Bullen CK, Shroff NS, Shan L, Yang H-C, Manucci JL, Bhat S, Zhang H, Margolick JB, Quinn TC. Disulfiram reactivates latent HIV-1 in a Bcl-2-transduced primary CD4+ T cell model without inducing global T cell activation. Journal of virology. 2011;85(12):6060-4.

39. Yang H-C, Xing S, Shan L, O'Connell K, Dinoso J, Shen A, Zhou Y, Shrum CK, Han Y, Liu JO. Small-molecule screening using a human primary cell model of HIV latency identifies compounds that reverse latency without cellular activation. The Journal of clinical investigation. 2009;119(11):3473-86.

40. Archin NM, Liberty A, Kashuba AD, Choudhary SK, Kuruc J, Crooks A, Parker D, Anderson E, Kearney M, Strain M. Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy. Nature. 2012;487(7408):482.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

749 41. Spivak AM, Andrade A, Eisele E, Hoh R, Bacchetti P, Bumpus NN, Emad F, Buckheit III
750 R, McCance-Katz EF, Lai J. A pilot study assessing the safety and latency-reversing activity of
751 disulfiram in HIV-1–infected adults on antiretroviral therapy. Clinical infectious diseases.
752 2013;58(6):883-90.
753 42. Bruner KM, Wang Z, Simonetti FR, Bender AM, Kwon KJ, Sengupta S, Fray EJ, Beg SA,
754 Antar AA, Jenike KM. A quantitative approach for measuring the reservoir of latent HIV-1
755 proviruses. Nature. 2019;566(7742):120-5.
756 43. Wonderlich ER, Subramanian K, Cox B, Wiegand A, Lackman-Smith C, Bale MJ, Stone
757 M, Hoh R, Kearney MF, Maldarelli F, Deeks SG, Busch MP, Ptak RG, Kulpa DA. Effector
758 memory differentiation increases detection of replication-competent HIV-l in resting CD4+ T cells
759 from virally suppressed individuals. PLoS pathogens. 2019;15(10):e1008074.
760 44. Read DF, Atindaana E, Pyaram K, Yang F, Emery S, Cheong A, Nakama KR, Burnett C,
761 Larragoite ET, Battivelli E, Verdin E, Planelles V, Chang C-H, Telesnitsky A, Kidd JM. Stable
762 integrant-specific differences in bimodal HIV-1 expression patterns revealed by high-throughput
763 analysis. PLOS Pathogens. 2019;15(10):e1007903.
764 45. Miller CM, Akiyama H, Agosto LM, Emery A, Ettinger CR, Swanstrom RI, Henderson
765 AJ, Gummuluru S. Virion-associated Vpr alleviates a postintegration block to HIV-1 infection of
766 dendritic cells. Journal of virology. 2017;91(13):e00051-17.
767 46. ENCODE. [Available from:
768 https://www.encodeproject.org/experiments/ENCSR000BXX/.
769 47. Consortium EP. An integrated encyclopedia of DNA elements in the human genome.
770 Nature. 2012;489(7414):57-74.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

48. Boehm D, Calvanese V, Dar RD, Xing S, Schroeder S, Martins L, Aull K, Li P-C, Planelles V, Bradner JE. BET bromodomain-targeting compounds reactivate HIV from latency via a Tat-independent mechanism. Cell cycle. 2013;12(3):452-62.

49. Einkauf KB, Lee GQ, Gao C, Sharaf R, Sun X, Hua S, Chen SM, Jiang C, Lian X, Chowdhury FZ. Intact HIV-1 proviruses accumulate at distinct chromosomal positions during prolonged antiretroviral therapy. The Journal of clinical investigation. 2019;129(3):988-98.

50. Kim M, Hosmane NN, Bullen CK, Capoferri A, Yang H-C, Siliciano JD, Siliciano RF. A primary CD4+ T cell model of HIV-1 latency established after activation through the T cell receptor and subsequent return to quiescence. Nature protocols. 2014;9(12):2755-70.

51. Gummuluru S, Emerman M. Cell cycle-and Vpr-mediated regulation of human immunodeficiency virus type 1 expression in primary and transformed T-cell lines. Journal of virology. 1999;73(7):5422-30.

52. Marsden MD, Zhang T-h, Du Y, Dimapasoc M, Soliman MS, Wu X, Kim JT, Shimizu A, Schrier A, Wender PA. Tracking HIV Rebound following Latency Reversal Using Barcoded HIV. Cell Reports Medicine. 2020;1(9):100162.

53. Stewart SA, Poon B, Jowett J, Chen I. Human immunodeficiency virus type 1 Vpr induces apoptosis following cell cycle arrest. Journal of Virology. 1997;71(7):5579-92.

54. Weinberger AD, Weinberger LS. Stochastic fate selection in HIV-infected patients. Cell. 2013;155(3):497-9.

55. McNamara LA, Ganesh JA, Collins KL. Latent HIV-1 infection occurs in multiple subsets of hematopoietic progenitor cells and is reversed by NF-κB activation. Journal of virology. 2012;86(17):9337-50.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

793 56. Sherrill-Mix S, Lewinski MK, Famiglietti M, Bosque A, Malani N, Ocwieja KE, Berry
794 CC, Looney D, Shan L, Agosto LM. HIV latency and integration site placement in five cell-based
795 models. Retrovirology. 2013;10(1):90.
796 57. Poon B, Chen IS. Human immunodeficiency virus type 1 (HIV-1) Vpr enhances expression
797 from unintegrated HIV-1 DNA. Journal of virology. 2003;77(7):3962-72.
798 58. Vansant G, Chen H-C, Zorita E, Trejbalová K, Miklík D, Filion G, Debyser Z. The
799 chromatin landscape at the HIV-1 provirus integration site determines viral expression. Nucleic
800 acids research. 2020;48(14):7801-17.
801 59. Rasmussen TA, Tolstrup M, Søgaard OS. Reversal of latency as part of a cure for HIV-1.
802 Trends in microbiology. 2016;24(2):90-7.
803 60. Chun T-W, Engel D, Mizell SB, Ehler LA, Fauci AS. Induction of HIV-1 replication in
804 latently infected CD4+ T cells using a combination of cytokines. Journal of Experimental
805 Medicine. 1998;188(1):83-91.
806 61. Greger IH, Proudfoot NJ, Demarchi F, Giacca M. Transcriptional interference perturbs the
807 binding of Sp1 to the HIV-1 promoter. Nucleic acids research. 1998;26(5):1294-300.
808 62. Reeves DB, Duke ER, Wagner TA, Palmer SE, Spivak AM, Schiffer JT. A majority of
809 HIV persistence during antiretroviral therapy is due to infected cell proliferation. Nature
810 communications. 2018;9(1):1-16.
811 63. Battivelli E, Dahabieh MS, Abdel-Mohsen M, Svensson JP, Da Silva IT, Cohn LB,
812 Gramatica A, Deeks S, Greene WC, Pillai SK. Chromatin Functional States Correlate with HIV
813 Latency Reversal in Infected Primary CD4+ T Cells. bioRxiv. 2018:242958.
814 64. Yang J, Zhao Y, Kalita M, Li X, Jamaluddin M, Tian B, Edeh CB, Wiktorowicz JE,
815 Kudlicki A, Brasier AR. Systematic determination of human cyclin dependent kinase (CDK)-9

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

816 interactome identifies novel functions in RNA splicing mediated by the DEAD Box (DDX)-5/17
817 RNA helicases. Molecular & Cellular Proteomics. 2015;14(10):2701-21.
818 65. Kharytonchyk S, King SR, Ndongmo CB, Stilger KL, An W, Telesnitsky A. Resolution of
819 Specific Nucleotide Mismatches by Wild-Type and AZT-Resistant Reverse Transcriptases during
820 HIV-1 Replication. Journal of molecular biology. 2016;428(11):2275-88.
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure Legends

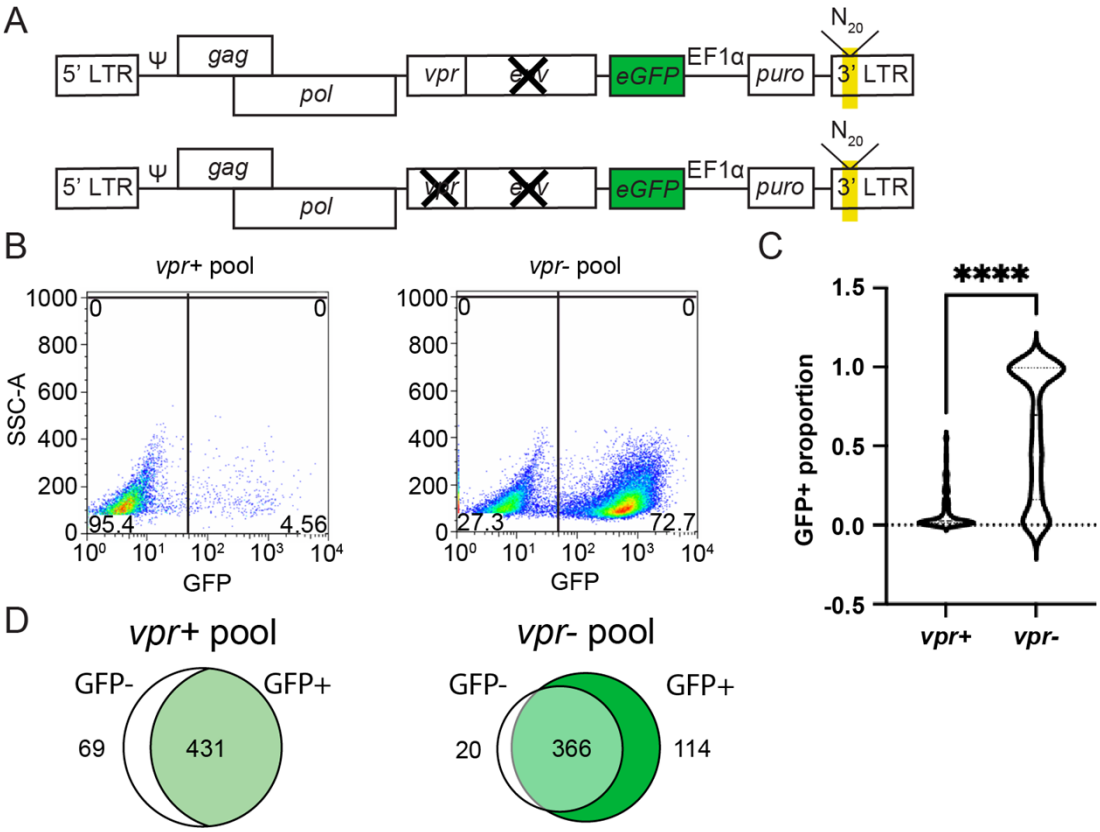


Figure 1. Generation of zip coded HIV-1 proviral pools. Polyclonal cell pools containing *vpr*+ and *vpr*- proviruses were established in Jurkat T cells and compared. **A.** Schematic of vectors used to produce zip coded virus (not to scale). Vectors were pseudotyped by co-transfection with pHEF-VSV-G into HEK293T cells. The yellow shade in the 3' LTR indicates position of 20bp randomized barcode in U3. **B.** Representative flow cytometry plots of polyclonal Jurkat T cells generated by infection with barcoded *vpr*+ (left panel) or *vpr*- (right panel) viruses, that were puromycin-selected for four days and expanded for an additional ten days. X axis indicates GFP signal detected in the FITC channel; numbers in each gate indicate percent of cells gated GFP- or GFP+. **C.** A comparison of the numbers of high and low LTR-active clones in *vpr*+ and *vpr*- pools. The fractions of total cells in each clone that sorted GFP+ (LTR-active) were calculated for the most abundant 500 clones in each pool by quantifying the fraction of each zip code (integrated

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

barcode) in the GFP+ sorted cells' genomic DNA to its quantity in the unsorted pool by high throughput sequencing and normalizing for the percent of the unsorted pool that was GFP+. Violin plots compare the GFP+ proportions (indicative of clones' fraction of LTR-active daughter cells) for the top 500 most abundant zip codes each pool. (p= **** indicated shows p<0.0001 Mann-Whitney U two-tailed test.). **D.** Venn diagrams of the number of zip codes in GFP+ and GFP- sorted sub pools for *vpr*+ and *vpr*- pools. The number in the white region indicates the number of zip codes observed only in GFP- sorted cells, that in the light-green region of intersection represents the number of zip codes that were present in both GFP+ and GFP- sorted cells, and the number in the dark green region represents zip codes present in only the green sorted cells. Reported zip codes represent those present in the top 78% and 88% of reads for *vpr*+ and *vpr*- respectively when clones were ordered by read abundance.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

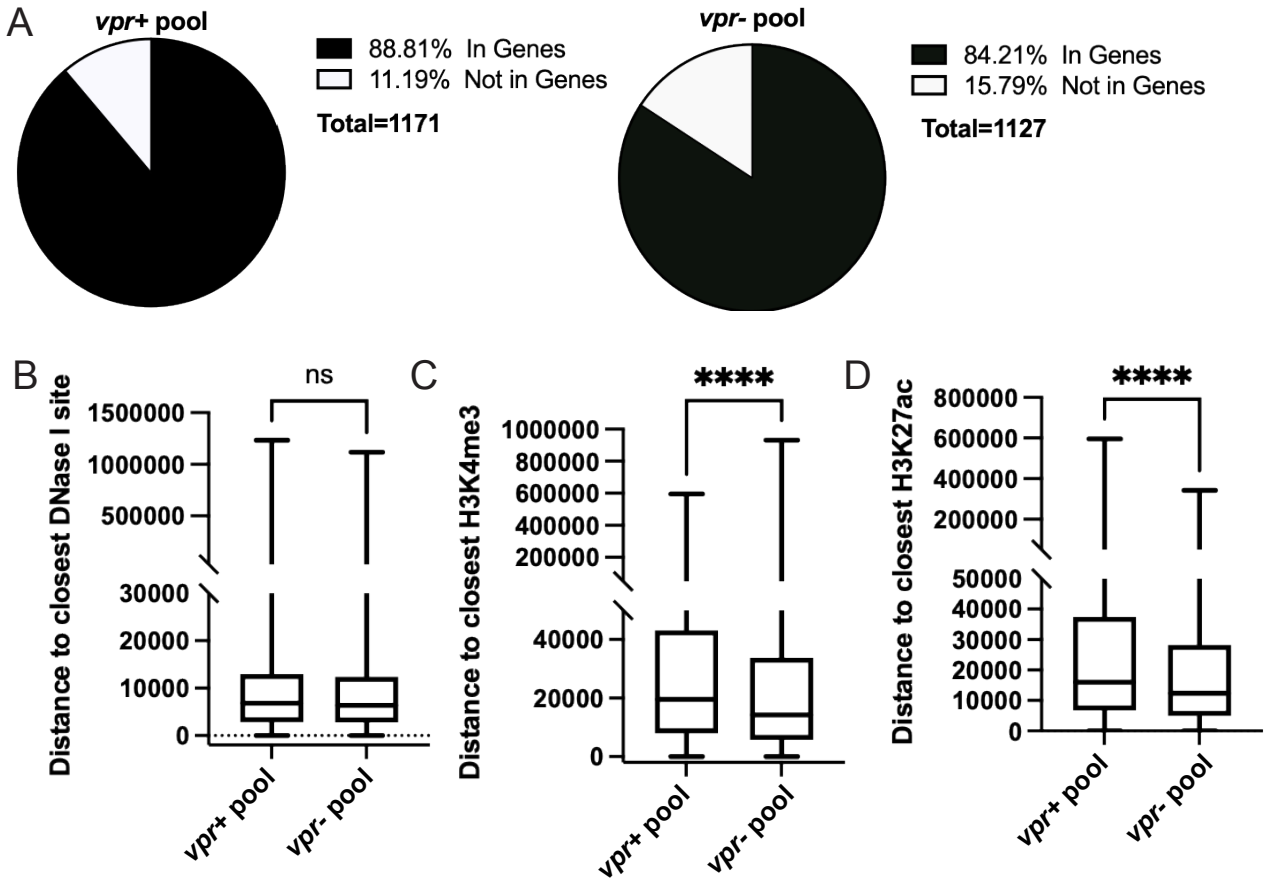
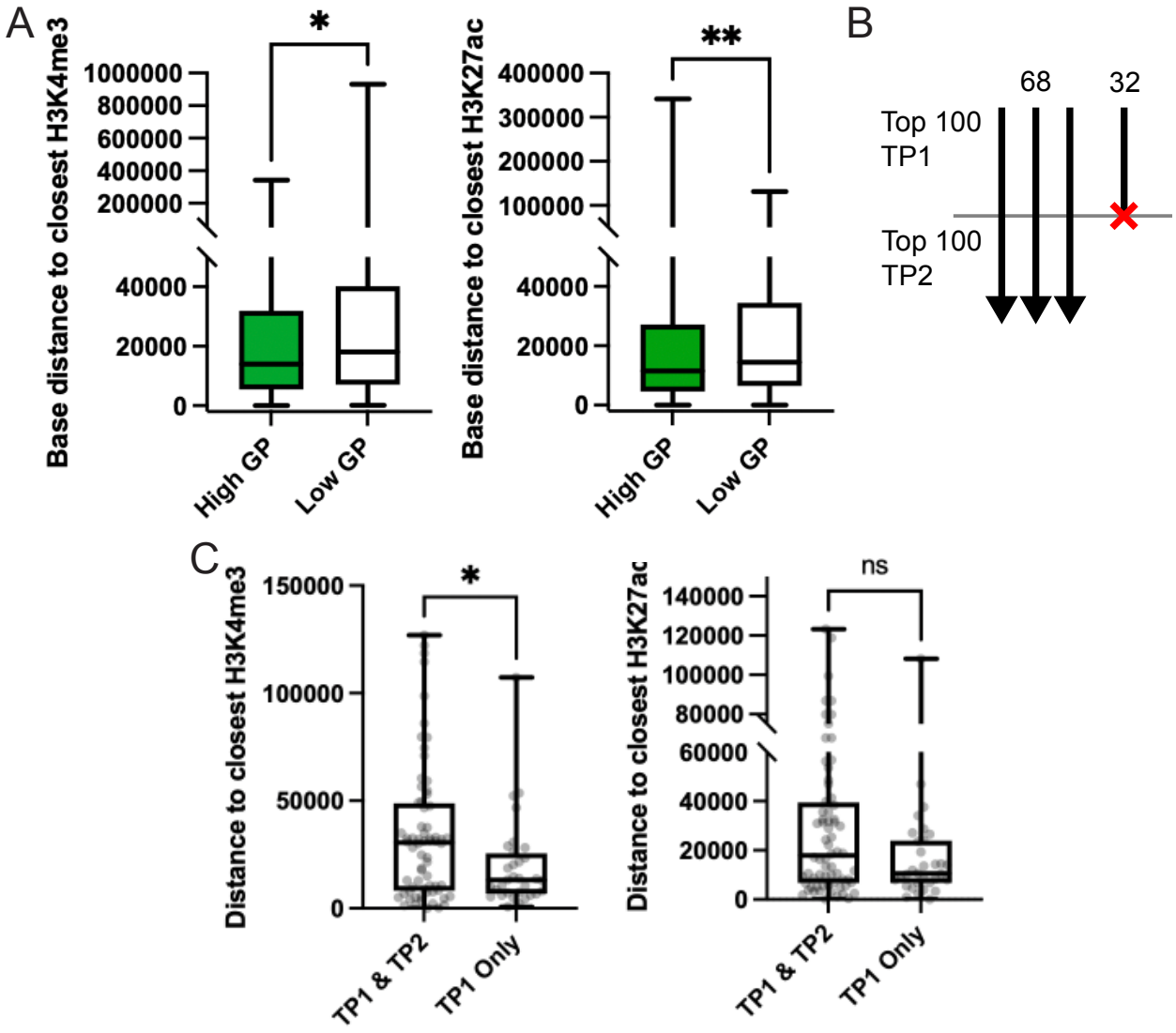


Figure 2. Comparison of integration sites in *vpr+* and *vpr-* pools. Integration sites for both *vpr+* and *vpr-* pool cells harvested 10 days post-infection were determined and distances to the nearest genome features indicated were mapped and compared. A. Pie charts showing the percentage of integration sites found within genes and those not found in genes for *vpr+* and *vpr-* pools. B.-D. Box plots showing pairwise comparisons between *vpr+* and *vpr-* integration sites' distances to the closest: B. DNase I hypersensitivity site, C. H3K4me3, and D. H3K27ac. Mann-Whitney U two-tailed test was conducted for pairwise comparisons for B-D (p= ns, and **** indicated shows $p>0.05$ and $p<0.0001$ respectively).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



881

882 **Figure 3. Integration site comparison within *vpr*+ and *vpr*- pools**. A. The GFP+ proportion of
883 each zip code from the *vpr*- pool was determined by quantification of the abundance in GFP+ cells'
884 genomic DNA to the total abundance in the genomic DNA from unsorted cells. Zip codes with
885 GFP+ proportion >= 0.6 were binned as "High" and zip codes (595 zip codes) with less than 0.3
886 GFP+ proportions were binned as "Low" (222 zip codes). Pairwise comparison of the distance to
887 the closest H3K4me3 (left) and H3K27ac (right) marks. B. Schematic diagram showing changes
888 in *vpr*+ pools' clonal composition over time. The number of zip codes in the top 100 most abundant
889 clones of the unsorted *vpr*+ pool sampled at time point one (TP1; 10 days post-infection) and the

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

890 number of zip codes that remained in the top 100 clones at time point two (TP2: 24 days post-
891 infection) are indicated. Arrows from TP1 to TP2 indicates zip codes present at both time points.
892 Arrows with red crosses indicate zip codes that were no longer present in the top 100 clones.
893 C. Box plots comparing integration site distances to active chromatin marks for the *vpr*+ pools
894 over time. The closest H3K4me3 or H3K27ac mark for zip codes that were among the top 100
895 clones at both TP1 &TP2 and for the clones that were lost from the top 100 at TP2 are indicated
896 (p=ns, indicated shows $p>0.05$; Mann-Whitney U two-tailed test).
897

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

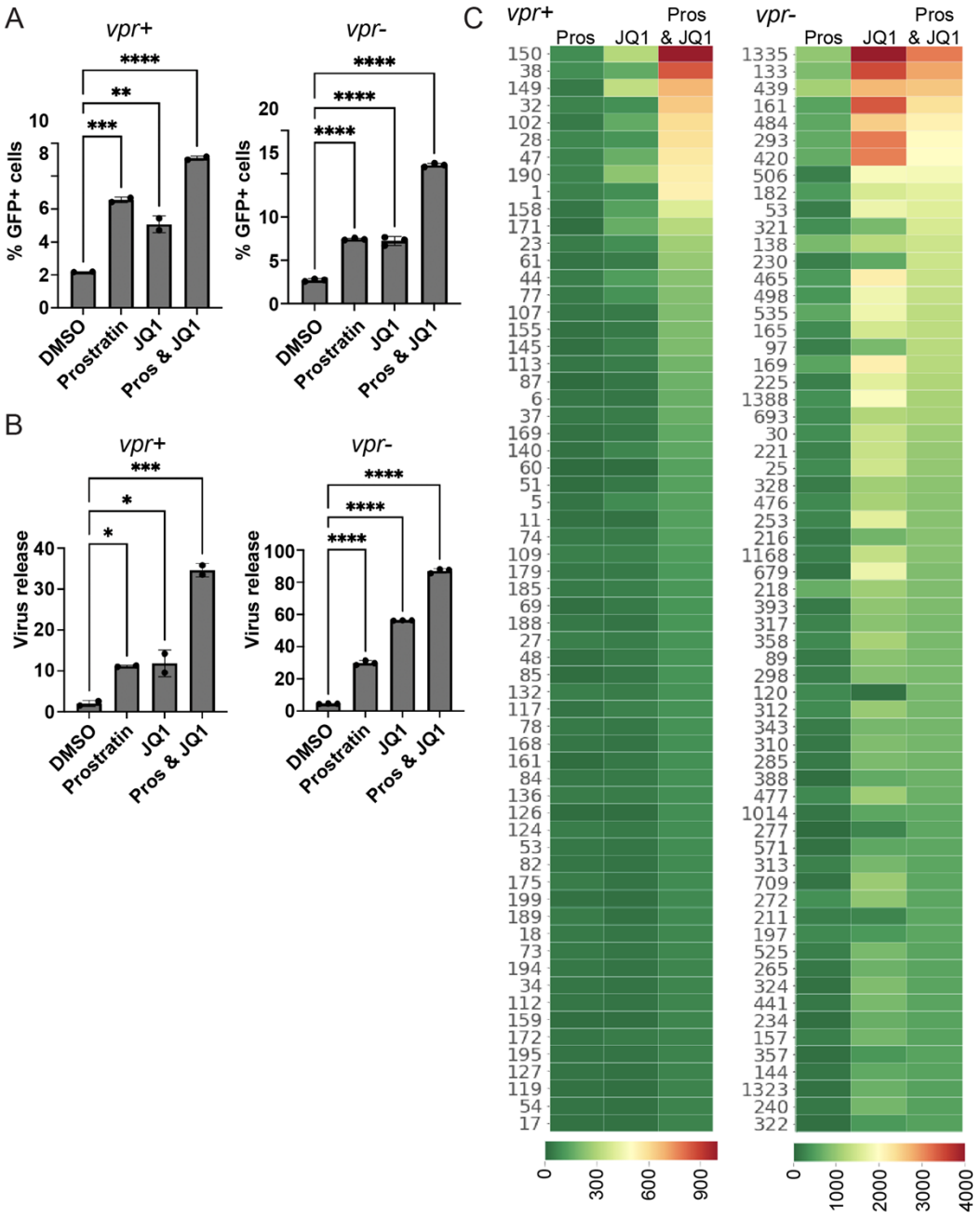
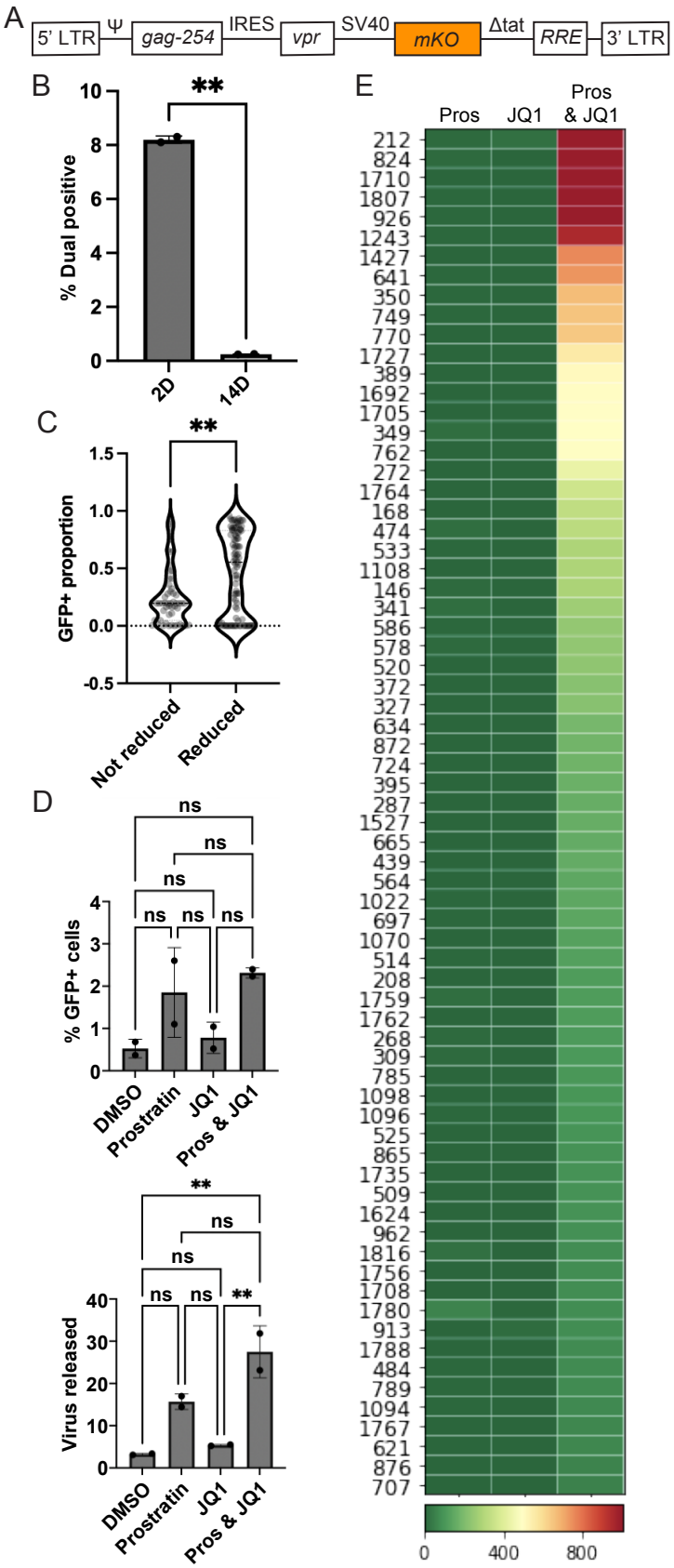


Figure 4. Effects of LRAs on zip coded pools and clones. *vpr*+ and *vpr*- GFP- cell fractions were exposed to 0.1% DMSO, 2μM JQ1, 2μM prostratin, or to a combination of prostratin and JQ1. Reactivation to GFP+ was measured at 24 hours post-infection by flow cytometry, and virus release was quantified using a reverse transcriptase assay and normalized to define p24 levels. A. Bar graphs showing the frequency of GFP+ cells post LRA treatment (left panel for each pool)

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and B. the amount of virus released into culture supernatant (right panel for each pool) for the indicated polyclonal pools. The error bars show the mean and standard deviation from two experimental treatment repetitions. C. Heatmaps of the clonal (zip code) virus release per treated GFP- cell (left: *vpr*+ pool and right: *vpr*- pool). Numbers at the left of each panel are clone identifiers generated by ordering proviral zip codes in decreasing relative abundance, as determined for the unsorted pools. Every row represents a unique cell clone's response. The clones were ordered from top to bottom by diminishing virus release per treated cell upon dual LRA treatment. The color bar indicates the extent of release per treated cell in arbitrary units.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



925

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

926 **Figure 5. Complementation of *vpr*- pools with Vpr expression vector.** *vpr* was added to *vpr*-
927 pools and effects on clone sizes were determined by high throughput analysis. A. Schematic
928 diagram of the Tat-deficient Vpr expression vector (not to scale), which contains *IRES-vpr* under
929 the HIV-1 LTR promoter and an SV40 promoter-driven mKO fluorescent reporter. B. Bar graph
930 of the percent GFP+ cells for two vpr^{rev} pools established in parallel at day-2 (2d) and day-14 (14d)
931 post-transduction with the Tat-deficient Vpr expression vector (p=0.0076, significant pairwise
932 comparison by paired t-test). C. Pairwise comparisons of GFP+ proportions between the clones
933 that were reduced (10-fold or more reduction in relative abundance) and those that were not
934 reduced (less than 1.5-fold reduction in relative abundance) after *vpr* addition (significant pairwise
935 comparison by Mann-Whitney U two-tailed test). D. Bar graphs showing the percent of GFP+ cells
936 24 hours post reactivation with JQ1, prostratin, and dual prostratin-JQ1 for the vpr^{rev} pool (left)
937 and *vpr*- pool (right). E. Heatmap of the clonal (zip code) virus release per treated GFP- cell of
938 vpr^{rev} pool. Numbers at the left of each panel are clone identifiers generated by ordering proviral
939 zip codes in decreasing relative abundance, as determined for the unsorted pools. Every row
940 represents a unique cell clone's response. The clones were ordered from top to bottom by
941 diminishing virus release per treated cell upon dual LRA treatment. The color bar indicates the
942 extent of release per treated cell in arbitrary units.
943
944
945
946
947
948

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

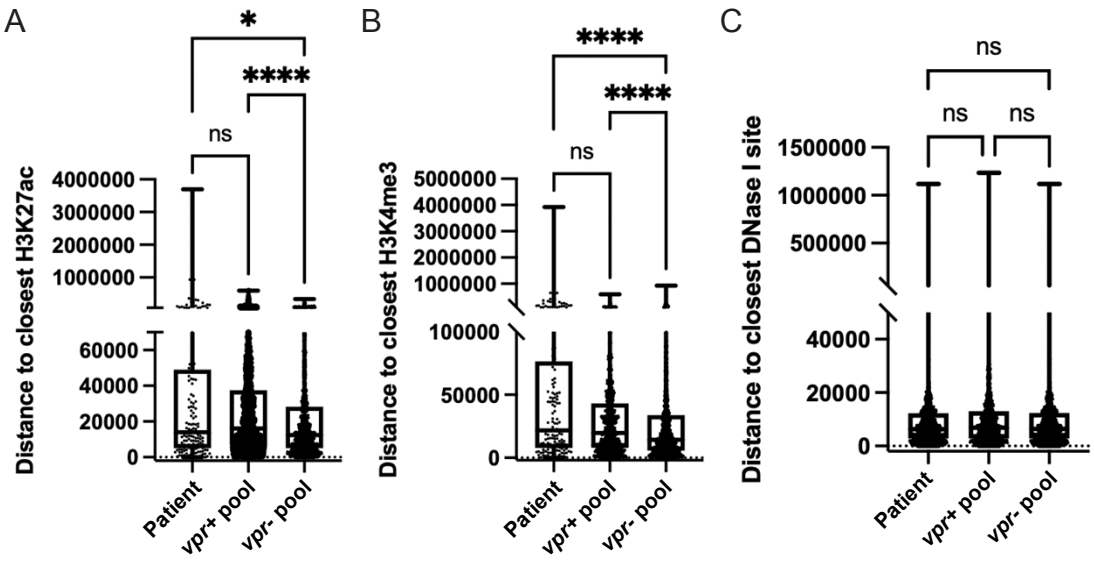
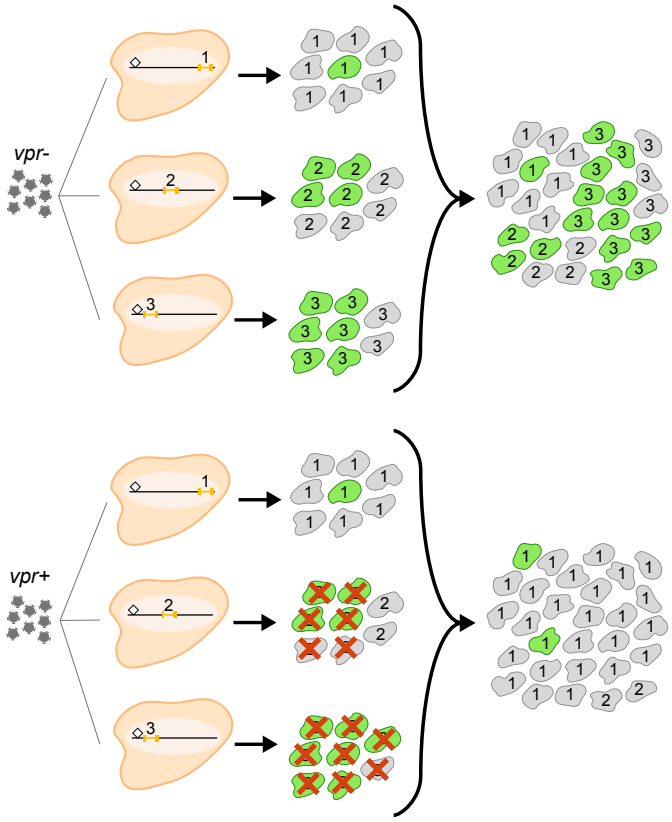


Figure 6. Comparison of patients' integration site features with *vpr*+ and *vpr*- pools. Previously published HIV integration sites from three patients were used to determine base distances to closest H3K27ac, H3K4me3, and DNase I sensitivity sites. Distances were compared to those of *vpr*+ and *vpr*- pools. Box plots comparing base distances of *vpr*+, *vpr*-, and patients' proviral integration sites to: A. H3K27ac, B. H3K4me3, and C. DNase I sensitivity sites. (Kruskal-Wallis test. p=ns, *, **, ***, **** indicate p>0.05, < 0.05, <0.01, 0.001, and 0.0001 respectively).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

965



966

967

968 **Figure 7. Schematic model of the effect of Vpr on integrant populations' proviral landscapes.**

969 From left: proviruses are integrated at the same genomic positions (designated 1, 2, and 3 above

970 the small orange bars, which indicate proviruses) regardless of whether the infecting virus was

971 Vpr+ or Vpr-. Note that integrants 1, 2, and 3 differ in their distances to the closest chromatin

972 mark, which is indicated by a small diamond, which may affect their expression characteristics

973 and responses to LRAs. Next, infected cells divide to form cell clones. For clone 1, cells that

974 express HIV genes (indicated by green cell) are rare (low LTR-active clone) and for other clones,

975 more cellular members of each clone express HIV (eg: clone 3: high LTR-active clone). This

976 pattern of bimodal gene expression is an intrinsic property of each clone. The red Xs indicate that

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

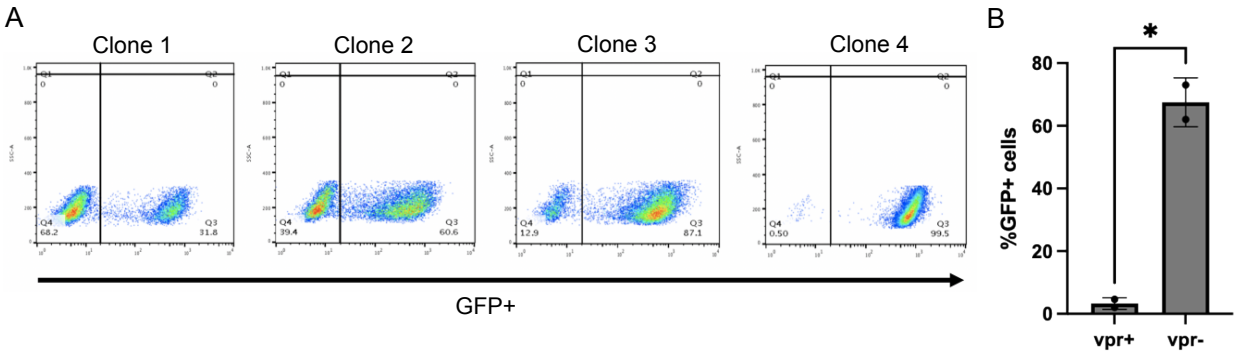
977 high LTR-active clone members are selectively depleted by Vpr. As a result, in the polyclonal
978 populations shown at right, low LTR-active clones are enriched in the *vpr*+ pool.

979

980

981

982



983

984 **S1 Fig. Comparison of the percent GFP+ of established cellular pools**. *vpr*+ and *vpr*- cell pools
985 that were generated by infection of jurkat cells with zip coded *vpr*+ and *vpr*- viruses, puromycin-
986 selected for five days, and expanded for an additional ten days. A. Four clones were isolated by
987 limiting dilution of *vpr*- pool, expanded for 14 days and subjected to flow cytometry. The flow
988 plots show clonal pools 1-4 have different percent of GFP+ cells. B. Bargraph showing the percent
989 GFP+ cells of two *vpr*+ pools and two *vpr*- pools (p=* indicates $p<0.05$; paired t-test).

990

991

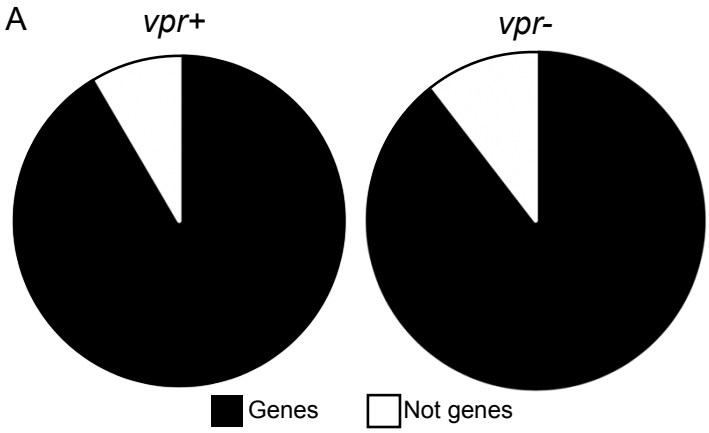
992

993

994

995

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

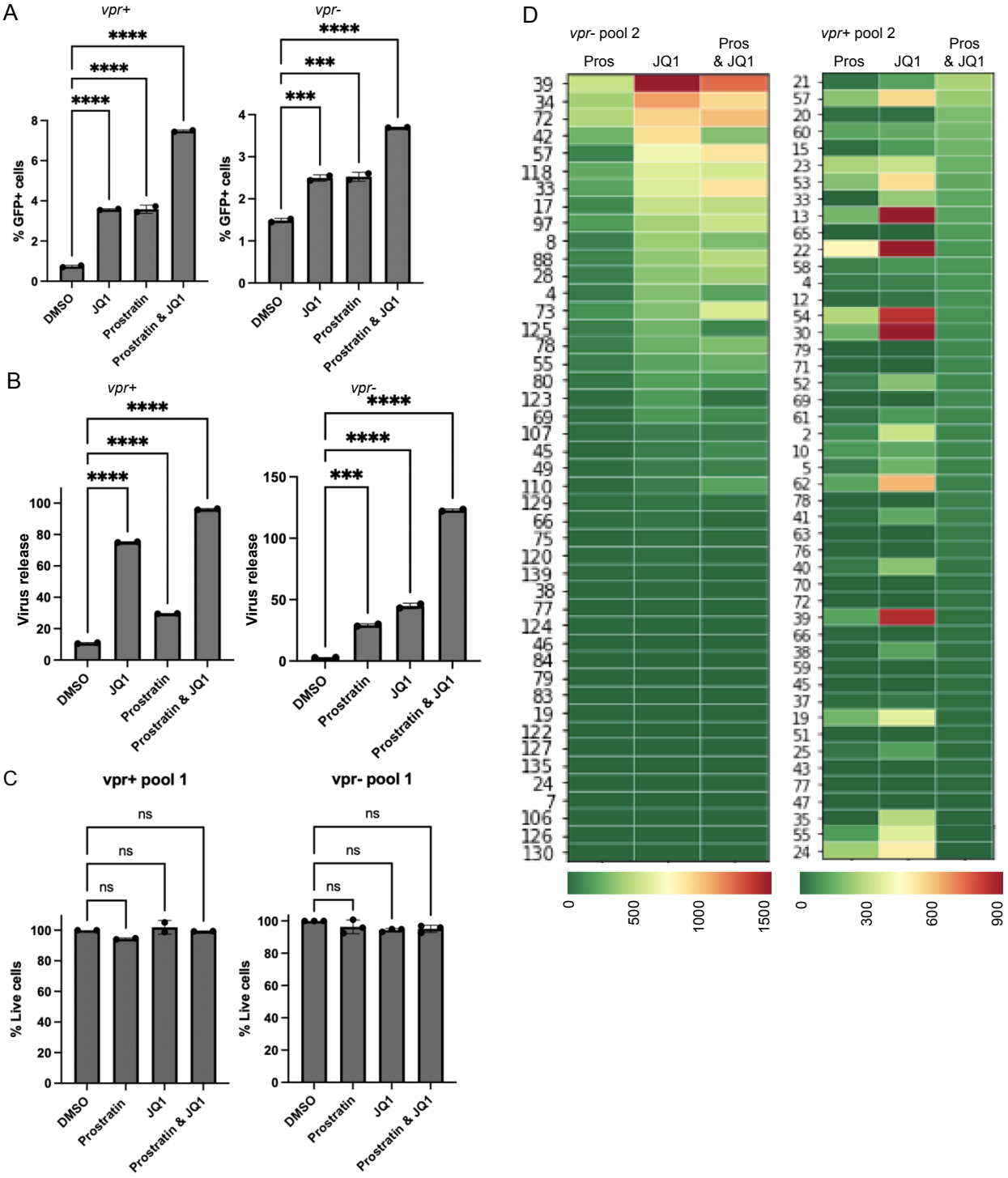


S2 Fig. Comparison of integration sites and GFP+ proportions between *vpr-* and *vpr+* pools. A. Integration sites were sequenced and mapped to genes in Jurkat T cells. 92% of integration sites were found in genes and 8% were not in genes for *vpr+* pool 2 (n=166) compared to 90% within genes and 10% not in genes for *vpr-* pool 2 (n=67).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1014

1015



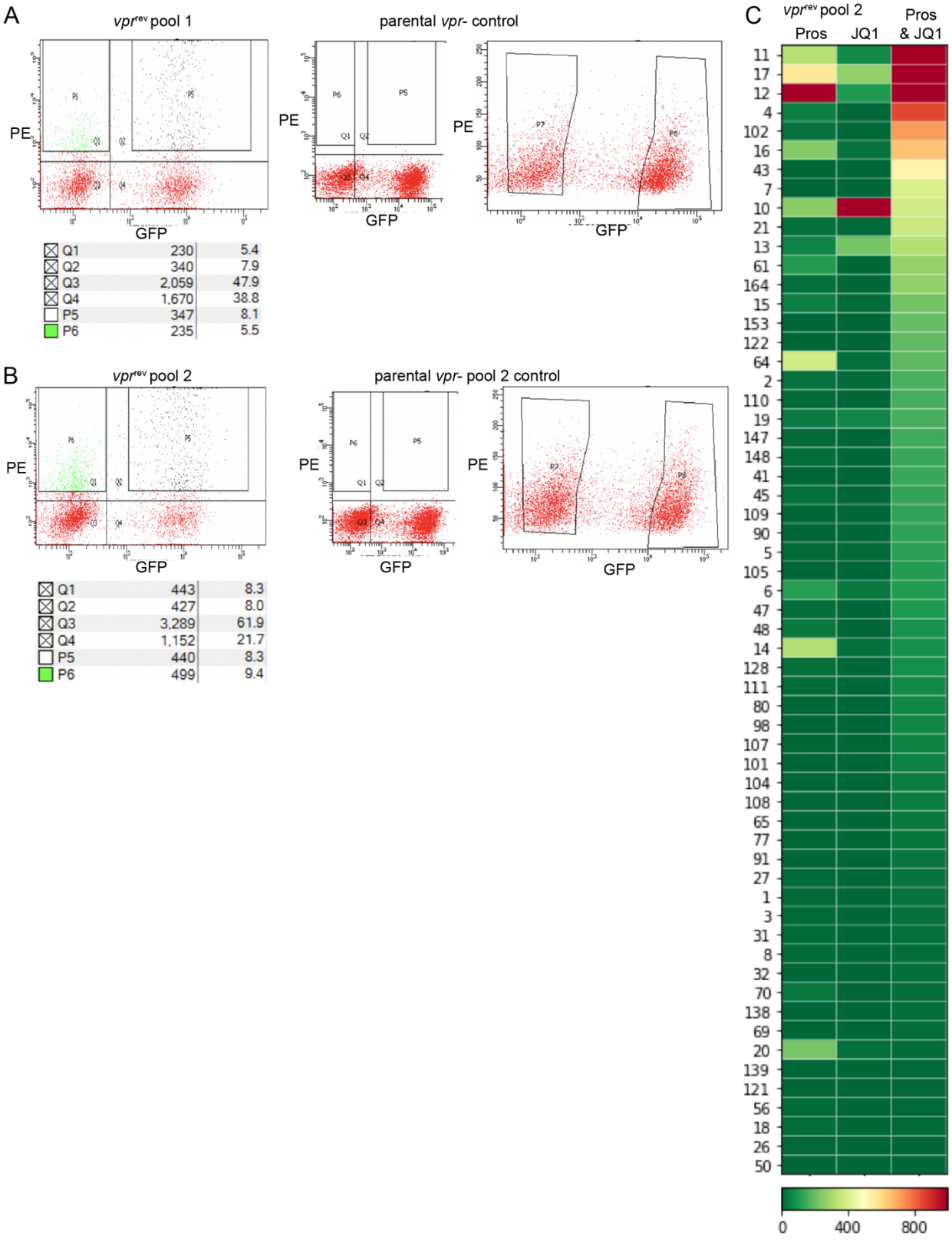
1016

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

S3 Fig. Cell viability and reactivation of additional *vpr*- and *vpr*+ pools. *vpr*+ and *vpr*- GFP- cells were exposed to 0.1% DMSO, 2μM JQ1, 2μM prostratin, and a combination of prostratin and JQ1. Reactivation was measured at 24 hours post-infection by flow cytometry, and virus release was quantified by reverse transcriptase assay. A. Bar graphs showing the frequency of GFP+ cells post LRA treatment (left panel for each pool) and the amount of virus released into culture supernatant (right panel for each pool) for the indicated polyclonal pools. The error bars show the mean and standard deviation from two experimental replicates. (Ordinary One-Way ANOVA; p=ns, ***, and **** indicate $p>0.05$, $p< 0.001$, and 0.0001 respectively). The error bars show mean and standard deviation. B. Viability was determined by propidium iodide staining and the percent viability was measured relative to the viability of cells in DMSO control. C. A heatmap of virus release per cell for the indicated clones (left: *vpr*- pool 2 and right: *vpr*+ pool 2). Every column represents a unique cell clone's response. The color bar indicates the extent of release in arbitrary units.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1039

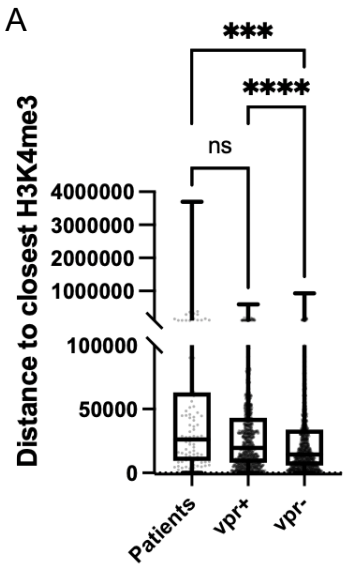


1040

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

S4 Fig. Sort gates for *vpr*-tranduced *vpr*- pools. Parental vpr- pools were transduced with Vpr-expression vector. 48 hours post-transduction, cells were sorted into PE+ GFP- and PE+ GFP+ cells. Flow plots for show the gates for sorting of A. Vpr-transduced *vpr*- pool 1 (left), parental *vpr*- pool 1 (right), B. Vpr- transduced *vpr*- pool 2 (left), and parental *vpr*- pool 2 (right). Quadrants P6 and P5 were collected for further analysis. B. Pairwise comparison of the distance to the closest H3K4me3 (left) and H3K27ac (right) marks of the zip codes that were similarly abundant at both time points (TP1 &TP2) and those that were only abundant at the first time point (TP1). Mann-Whitney two-tailed test p=ns and * indicate $p>0.05$ and $p<0.05$ respectively. C. A heatmap of virus release per treated cell for vpr^{rev}. The indicated clones in the first column are arranged from left to right by diminishing virus release per dual treatment, and Pros means prostratin.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



S5 Fig. Comparison of patient integration site distance to the nearest H3K4me3 mark to *vpr*+ and *vpr*- pools. Previously reported integration sites of intact proviruses from three HIV patients were mapped to the nearest pre-existing H3K4me3to in memory CD4+ T cells. Significant comparisons by Kruskal-Wallis test. p=ns, *, **, ***, **** indicates p>0.05, < 0.05, <0.01, 0.001, and 0.0001 respectively.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1076

1077 S Table. Table of zip codes, these clones' fractions of abundance, and integration sites of *vpr*- and

1078 *vpr*+ pools

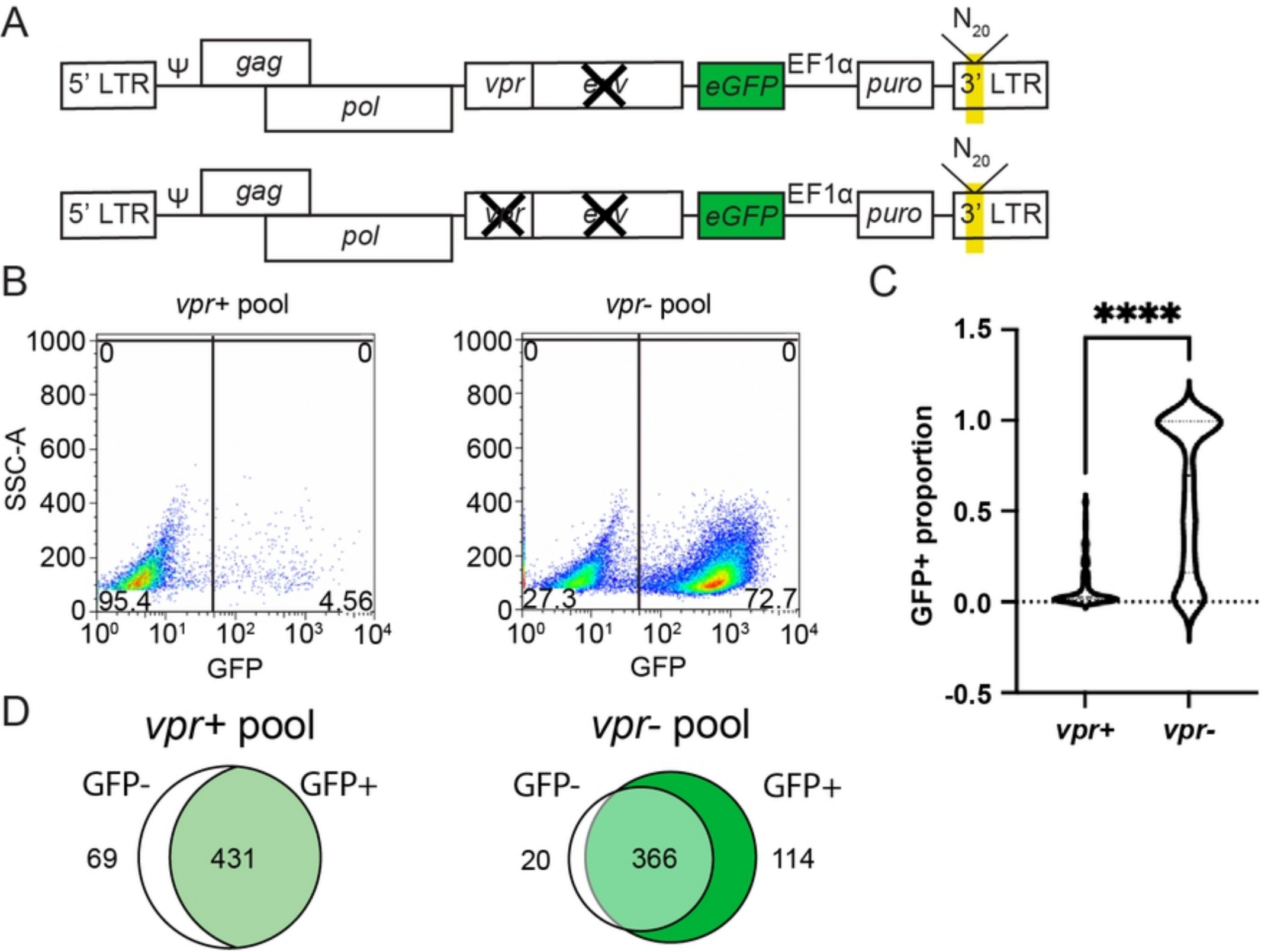

A
5' LTR
Ψ
gag
pol
vpr
env
eGFP
EF1α
puro
3' LTR
N20
B
vpr+ pool
vpr- pool
SSC-A
GFP
95.4
4.56
27.3
72.7
C
****
GFP+ proportion
vpr+
vpr-
D
vpr+ pool
vpr- pool
GFP-
GFP+
69
431
20
366
114

Figure 1

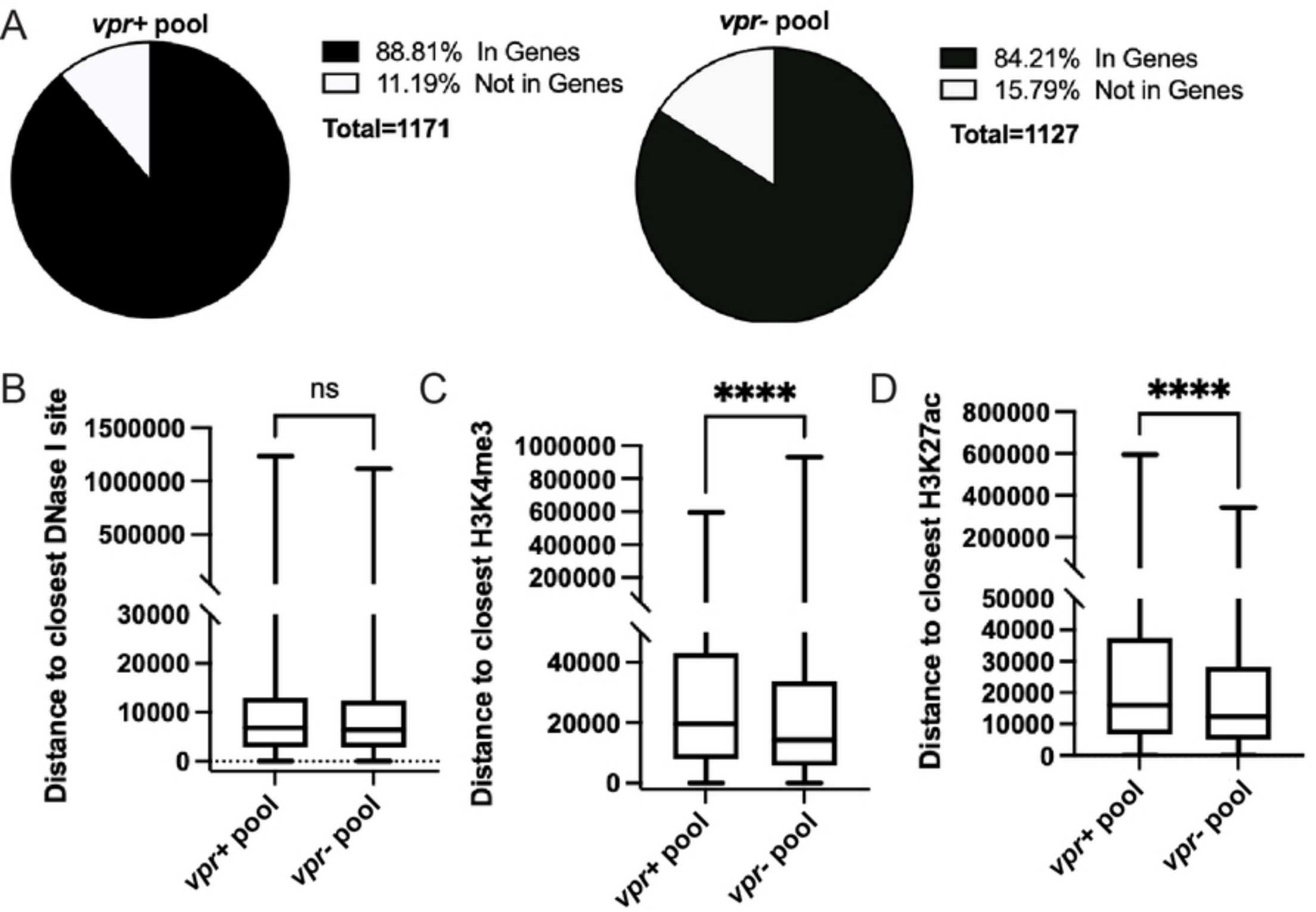

A
vpr+ pool
88.81% In Genes
11.19% Not in Genes
Total=1171
vpr- pool
84.21% In Genes
15.79% Not in Genes
Total=1127
B
ns
Distance to closest DNase I site
1500000
1000000
500000
30000
20000
10000
0
vpr+ pool
vpr- pool
C
****
Distance to closest H3K4me3
1000000
800000
600000
400000
200000
40000
20000
0
vpr+ pool
vpr- pool
D
****
Distance to closest H3K27ac
800000
600000
400000
200000
50000
40000
30000
20000
10000
0
vpr+ pool
vpr- pool

Figure 2

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

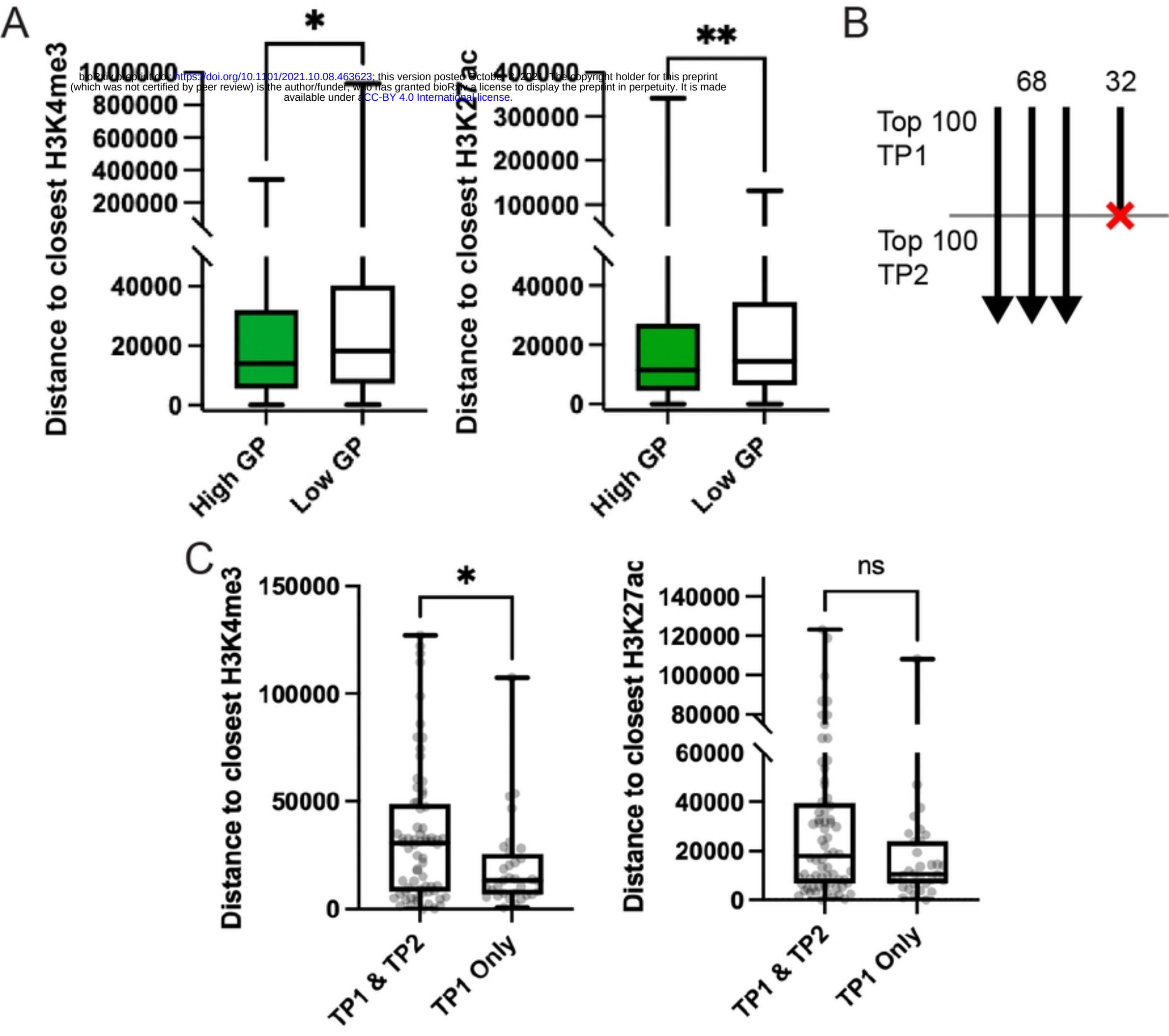


Figure 3

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

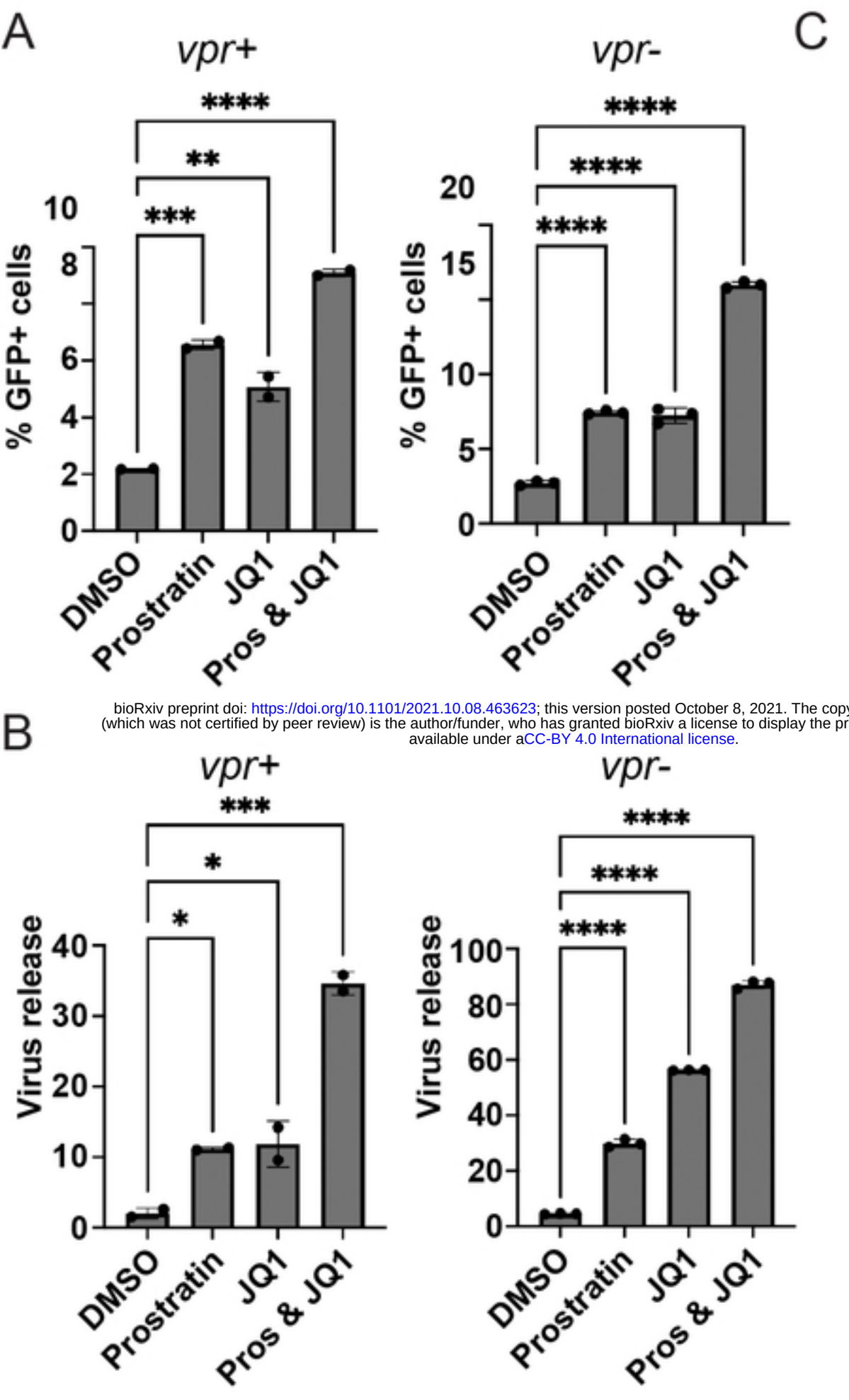
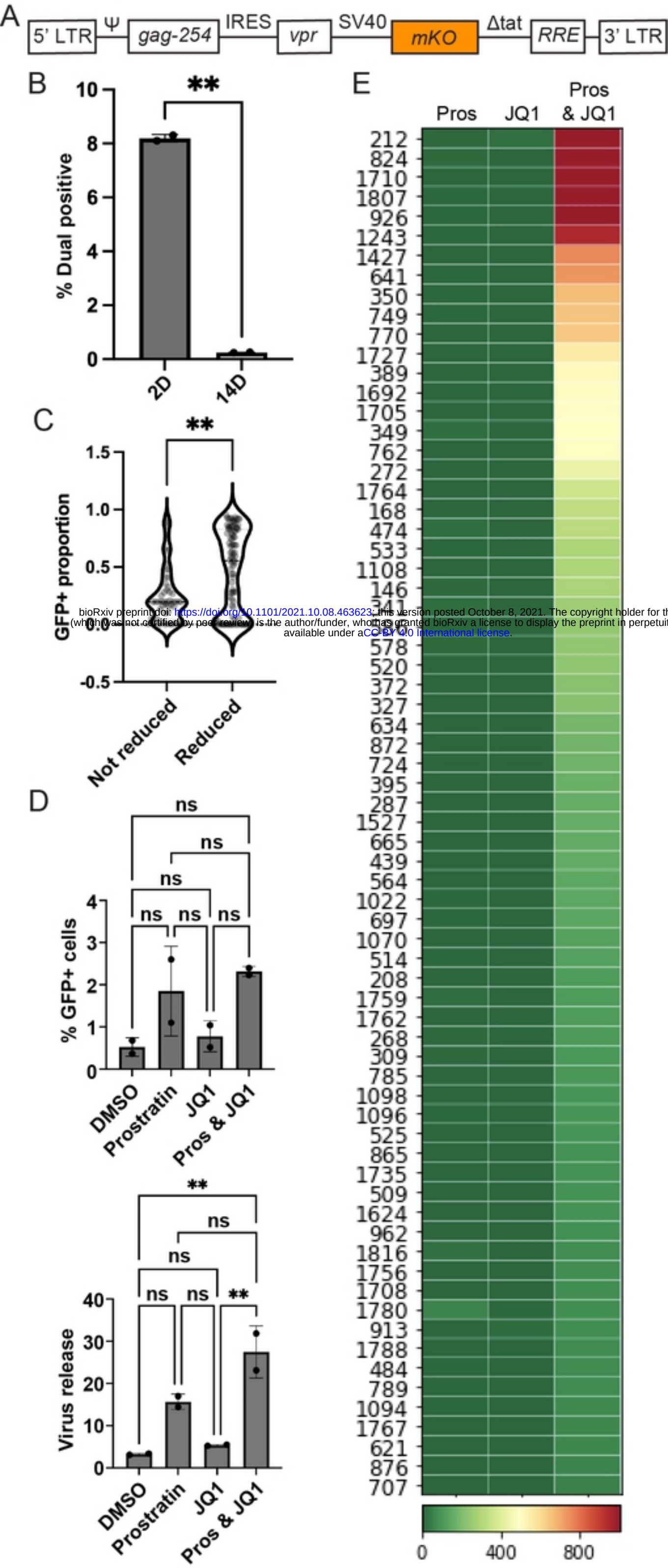


Figure 4



bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for th
(which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuit
available under aCC-BY 4.0 International license.

Figure 5

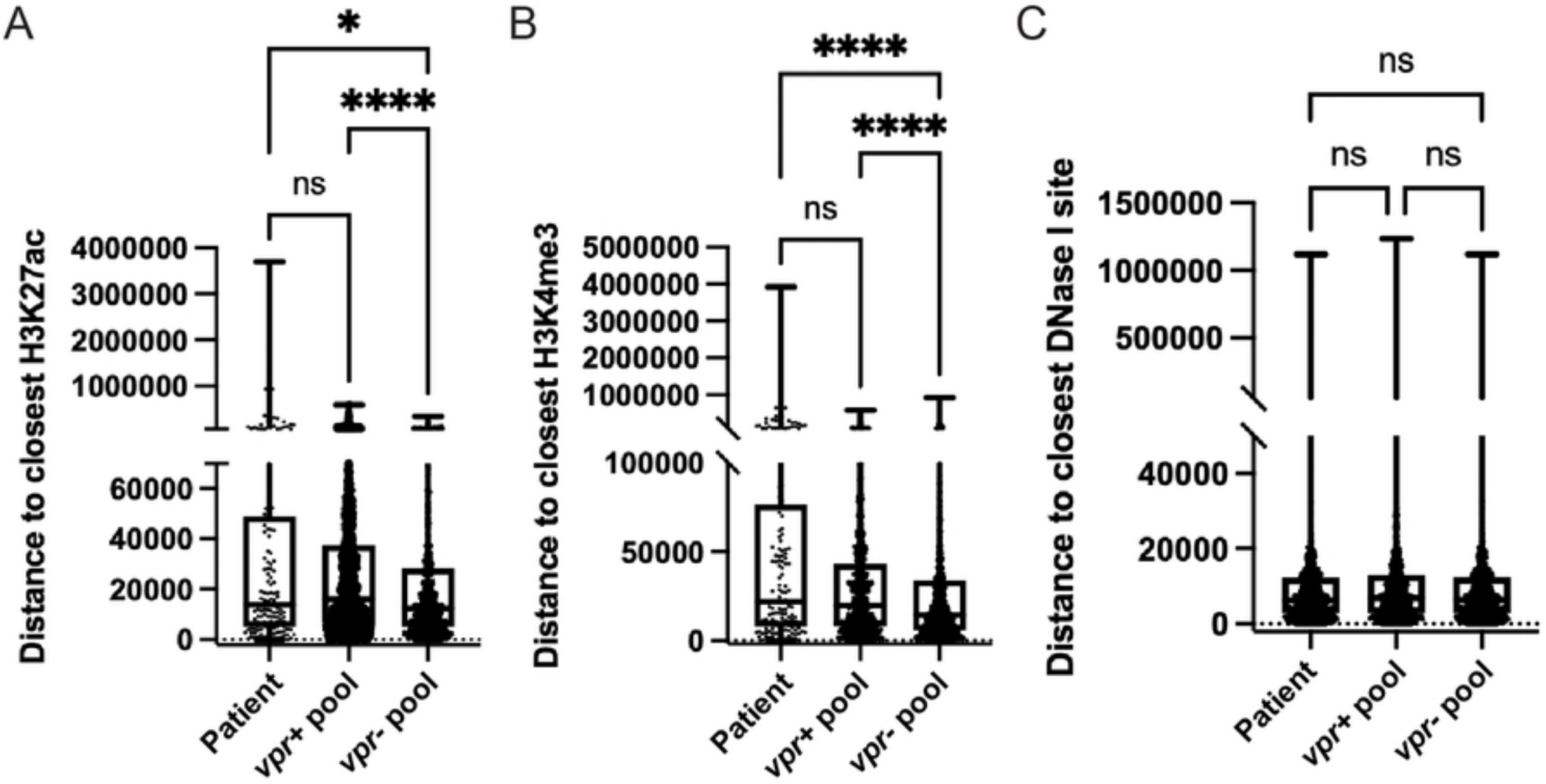

A
B
C
*
****
ns
Distance to closest H3K27ac
4000000
3000000
2000000
1000000
60000
40000
20000
0
Patient
vpr+ pool
vpr- pool
Distance to closest H3K4me3
5000000
4000000
3000000
2000000
1000000
100000
50000
0
Distance to closest DNase I site
1500000
1000000
500000
40000
20000
0

Figure 6

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

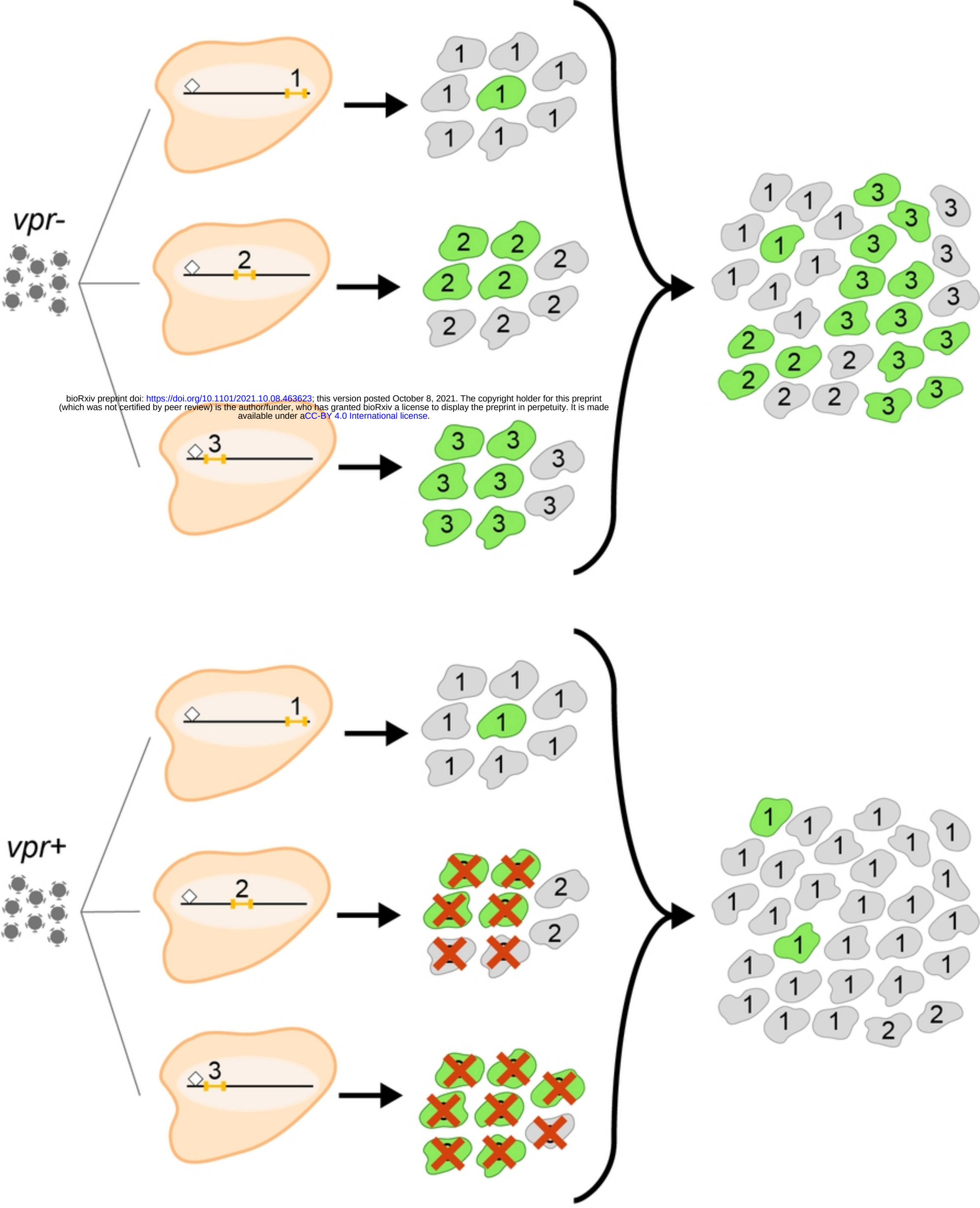
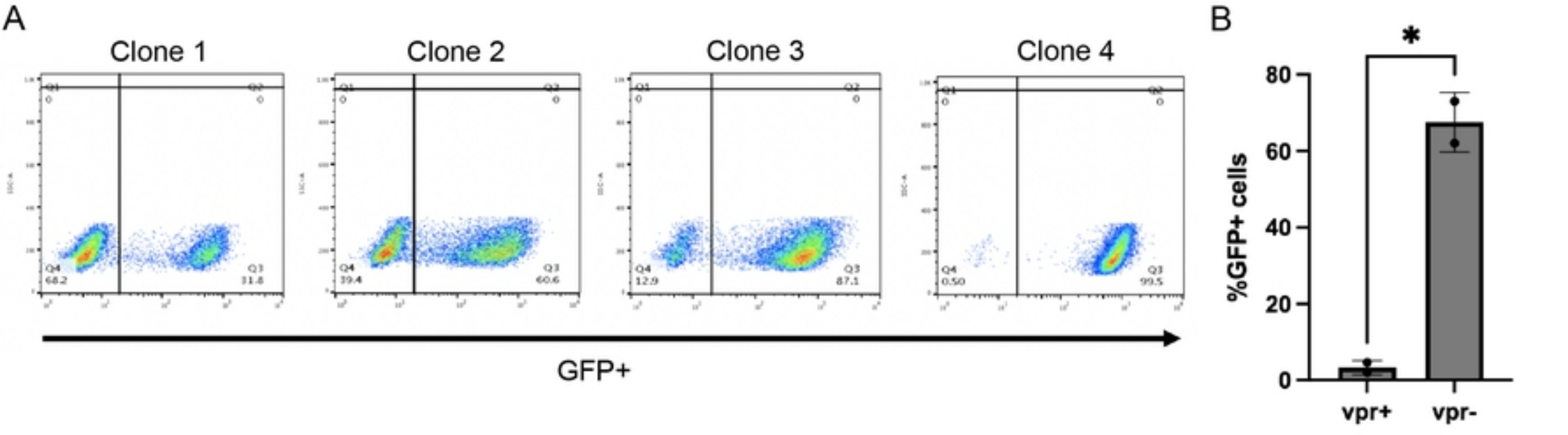
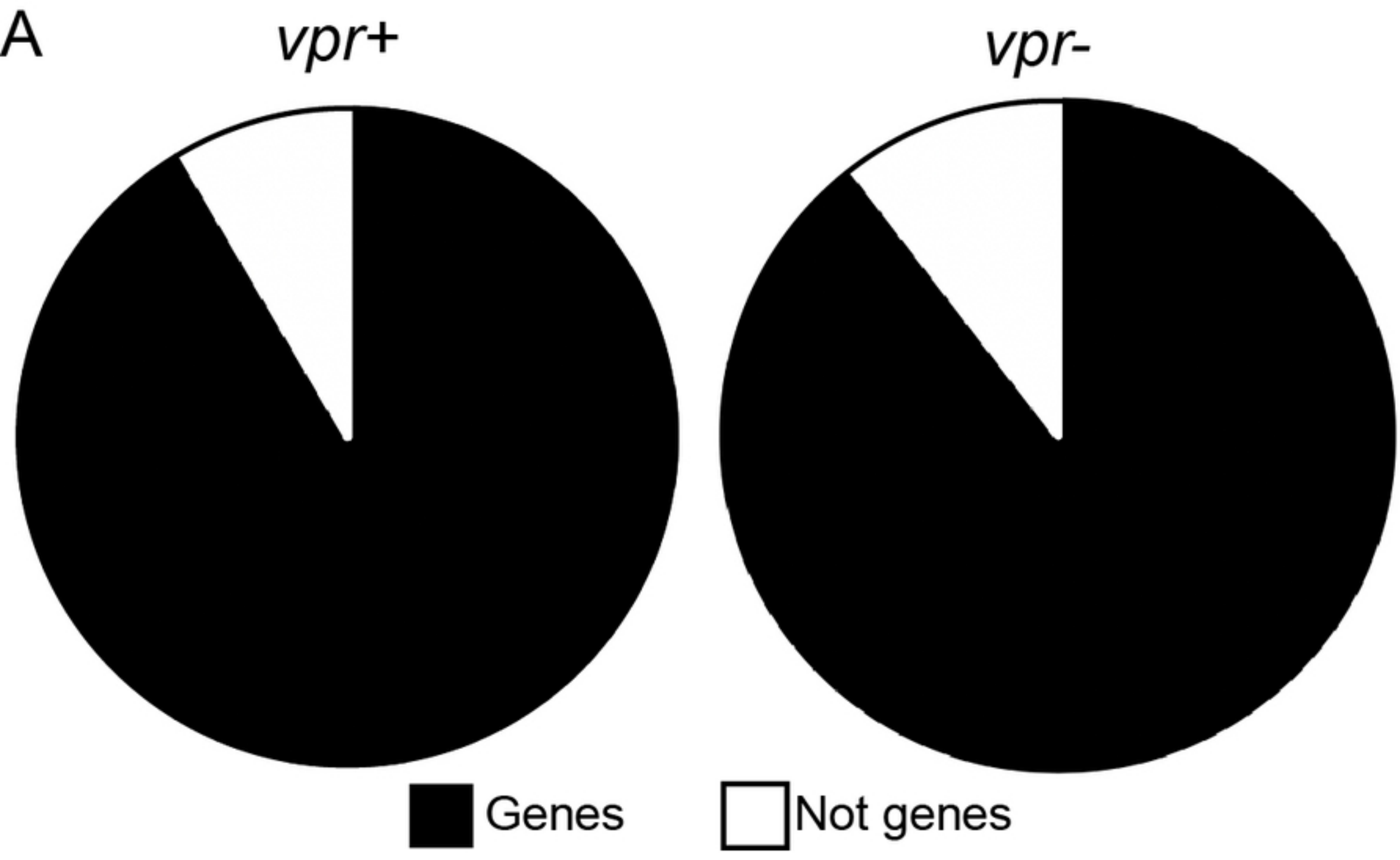


Figure 7

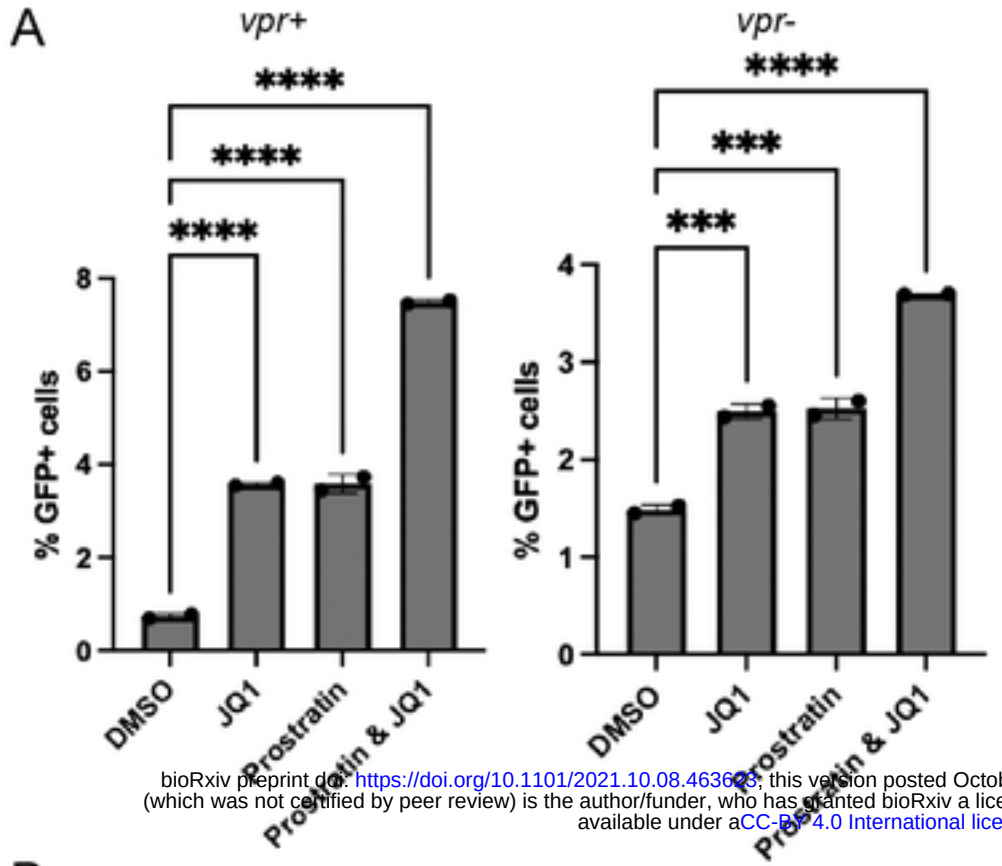

A
Clone 1
Clone 2
Clone 3
Clone 4
Q4
68.2
Q3
31.8
Q4
39.4
Q3
60.6
Q4
12.9
Q3
87.1
Q4
0.50
Q3
99.5
GFP+
B
*
%GFP+ cells
80
60
40
20
0
vpr+
vpr-

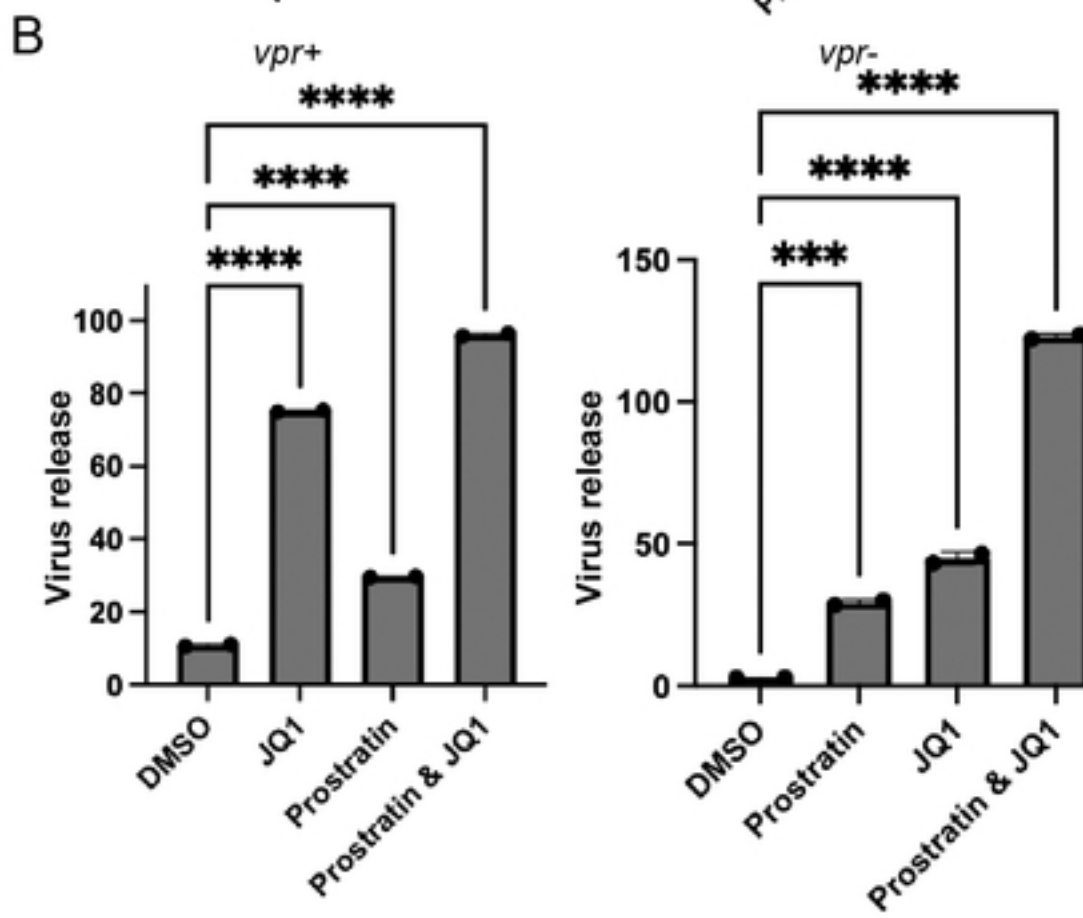
S1

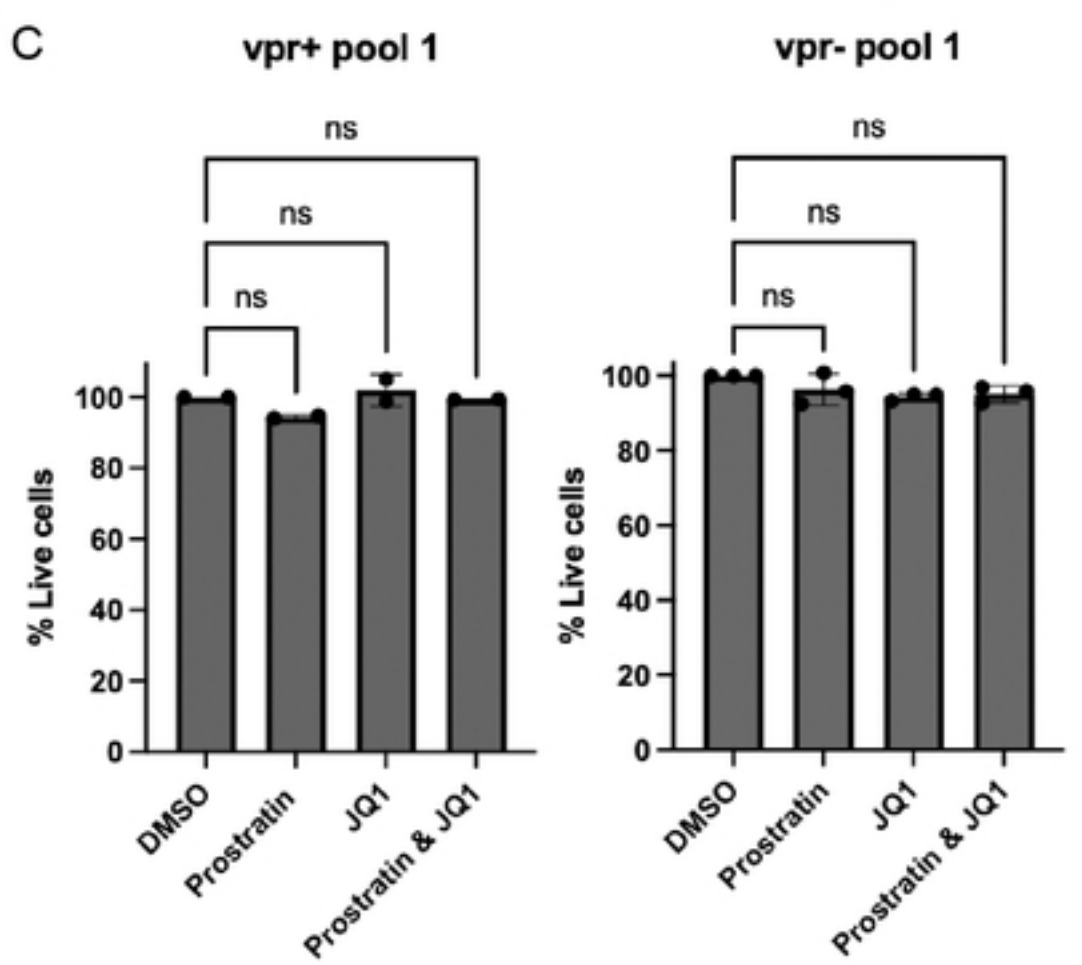


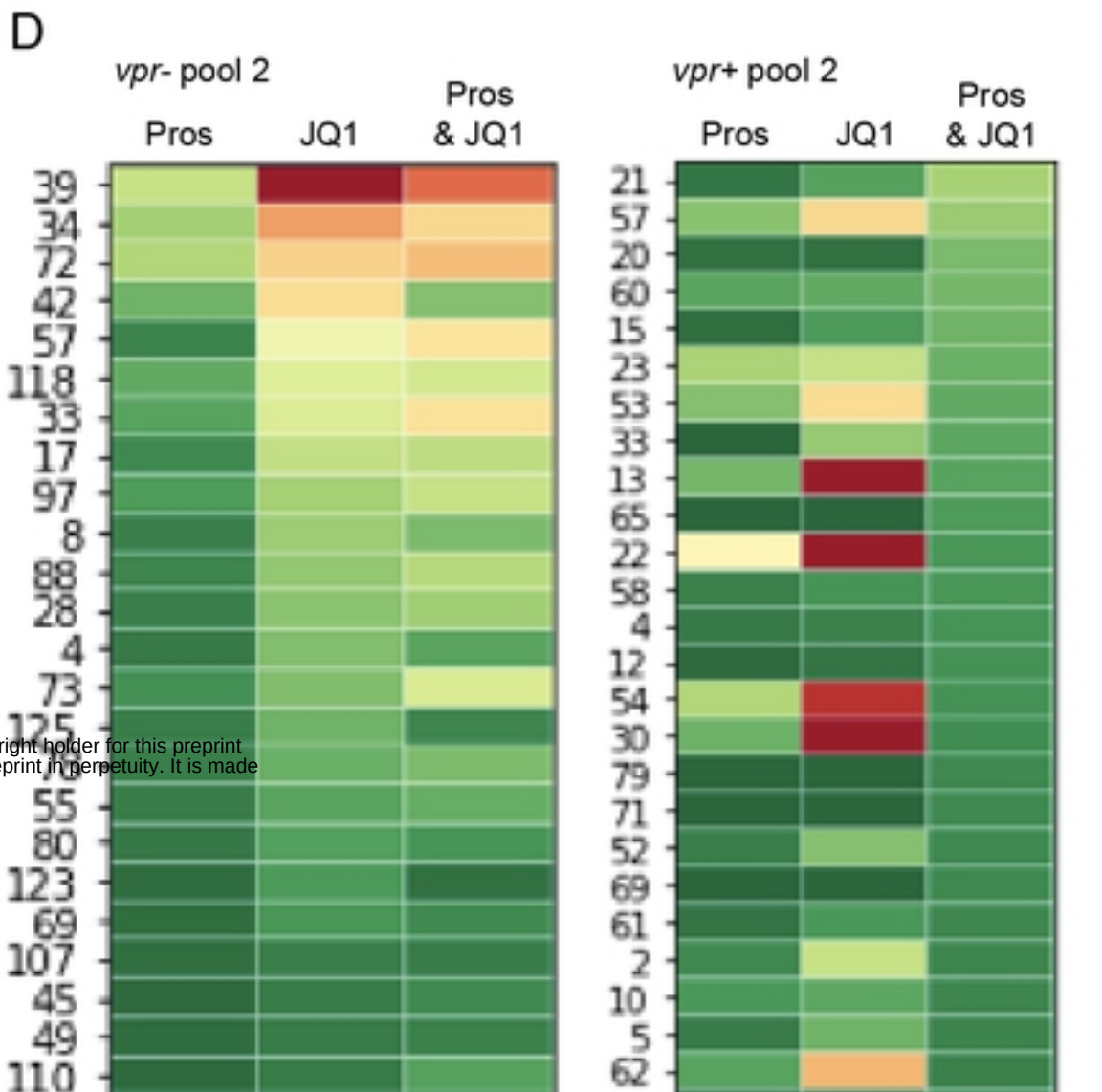
S2

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



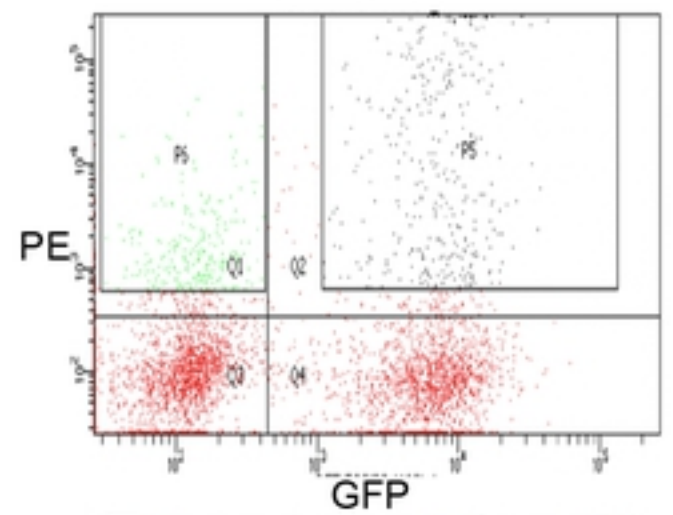




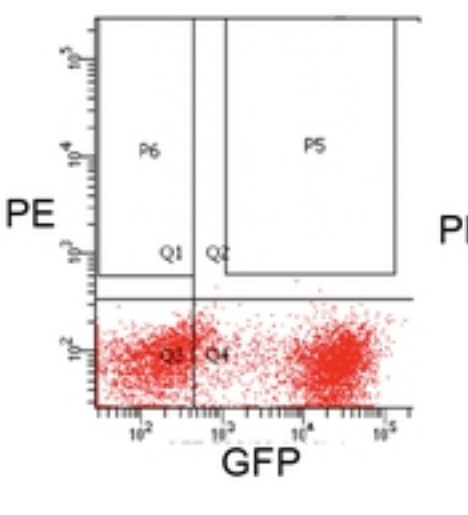


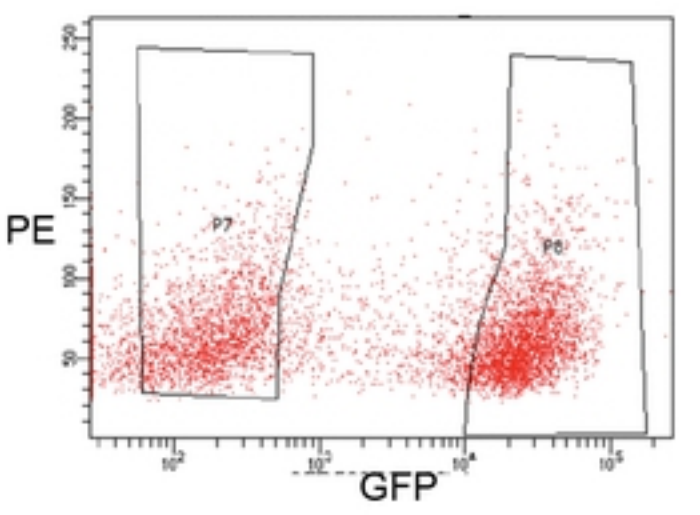
S3

A

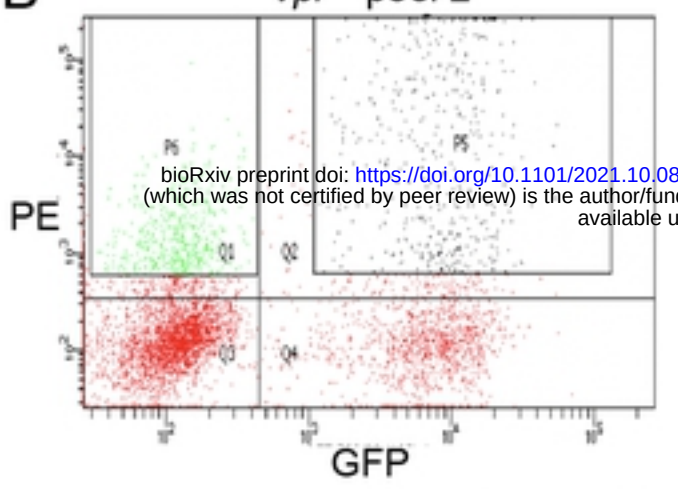


Q1	230	5.4
Q2	340	7.9
Q3	2,059	47.9
Q4	1,670	38.8
P5	347	8.1
P6	235	5.5

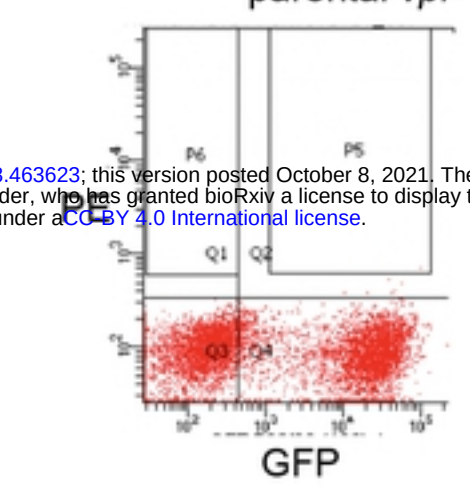


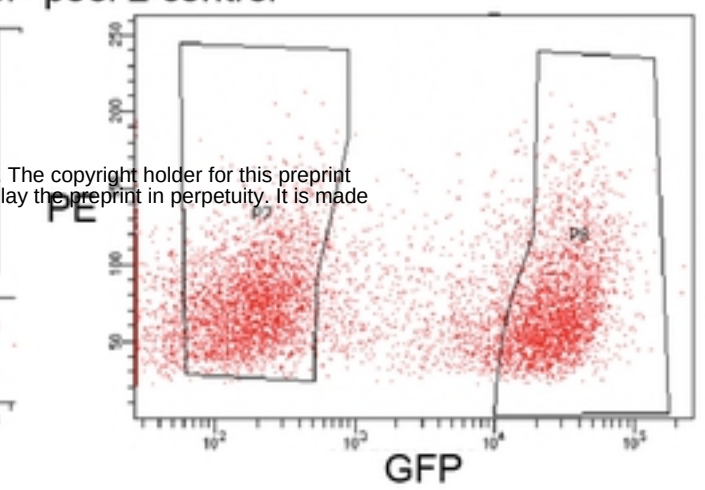


B



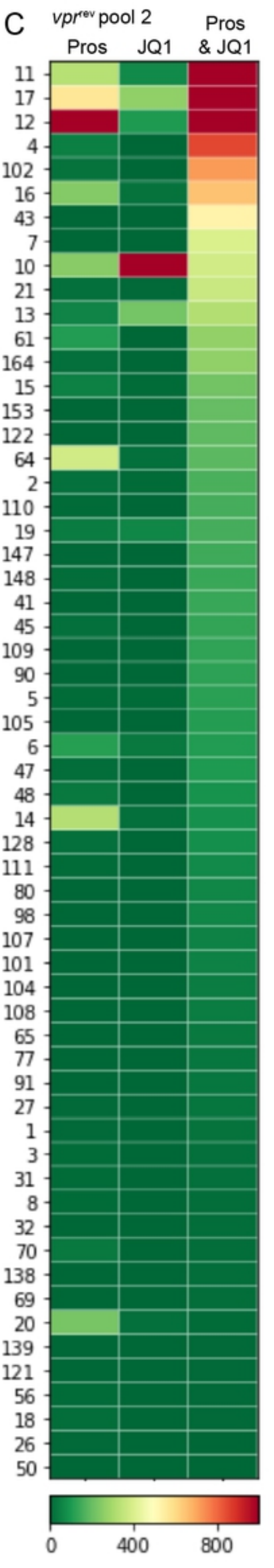
Q1	443	8.3
Q2	427	8.0
Q3	3,289	61.9
Q4	1,152	21.7
P5	440	8.3
P6	499	9.4



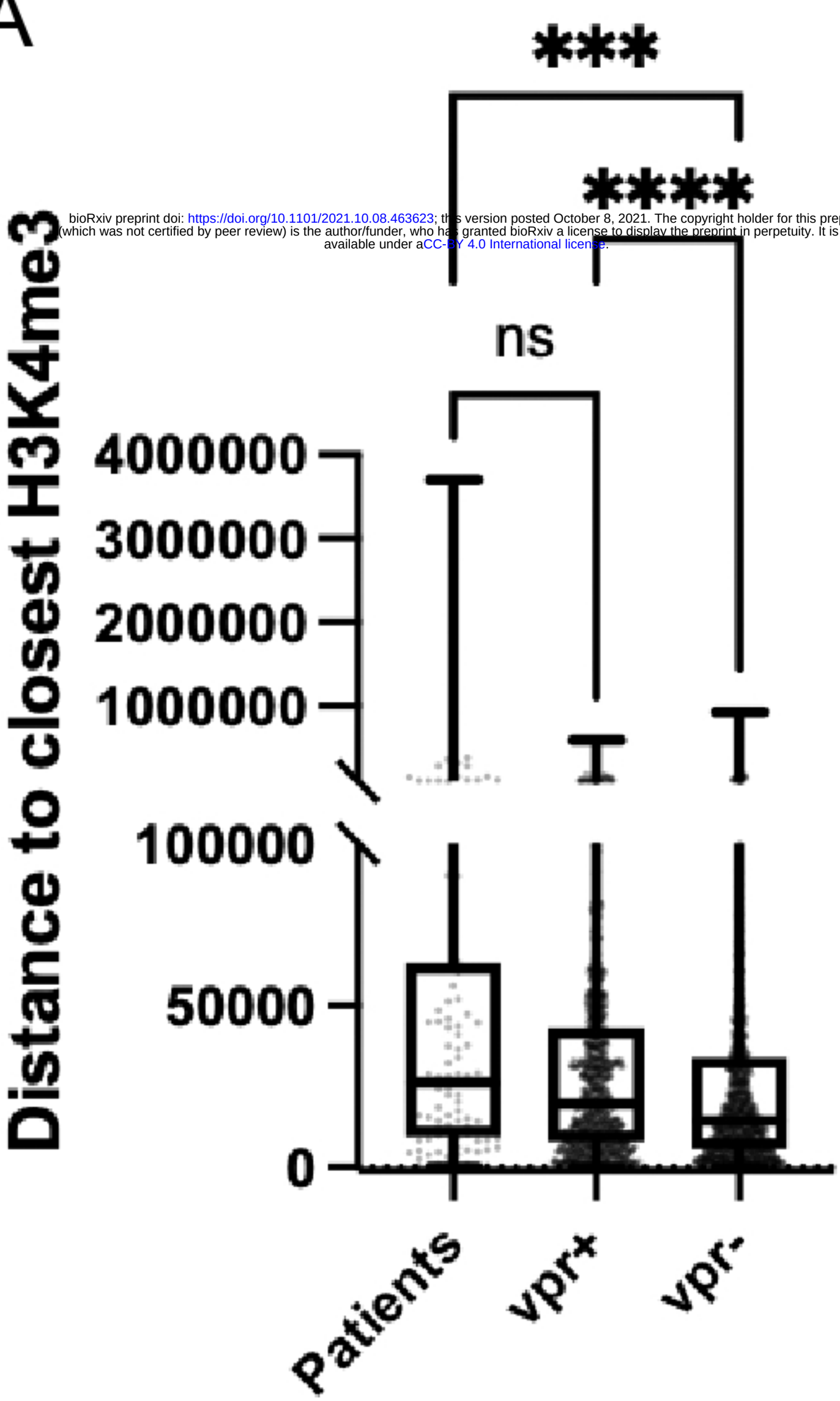


bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

C



S4

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.08.463623; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

A