Targeting evolutionary conserved sequences circumvents the evolution of resistance in a viral gene drive against human cytomegalovirus

Authors

Marius Walter[1]*, Rosalba Perrone[1], Eric Verdin[1]*.

[1] Buck Institute for Research on Aging, Novato, CA, United States.

*Correspondence to mwalter@buckinstitute.org, everdin@buckinstitute.org

Abstract

Gene drives are genetic systems designed to efficiently spread a modification through a population. Most engineered gene drives rely on CRISPR-Cas9 and were designed in insects or other eukaryotic species. We recently developed a viral gene drive in herpesviruses that efficiently spread into a population of wildtype viruses. A common consequence of gene drives is the appearance and selection of drive-resistant sequences that are no longer recognized by CRISPR-Cas9. Here, we analyze in cell culture experiments the evolution of resistance in a gene drive against human cytomegalovirus. We report that after an initial invasion of the wildtype population, a drive-resistant population is positively selected over time and outcompetes gene drive viruses. However, we show that targeting evolutionary conserved regions ensures that drive-resistant viruses have a replication defect, leading to a long-term reduction of viral levels. This marks an important step toward developing effective gene drives in viruses, especially for therapeutic applications.

Introduction

Herpesviruses are nuclear-replicating viruses with large dsDNA genomes (100-200 kb) that encode hundreds of genes (1). They establish life-long persistent infections and are implicated directly or indirectly in numerous human diseases (2). In particular, human cytomegalovirus (hCMV) is an important threat to immunocompromised patients, such as HIV-infected individuals, receivers of organ transplants and unborn infants.

Gene drives are genetic modifications designed to spread efficiently in a target population (3–6). They have been engineered principally in insect species, such as mosquitoes, and are seen as a potential strategy to eradicate vector-borne diseases, such as malaria and dengue. Most gene drives rely on CRISPR-mediated homologous recombination and have been restricted to sexually reproducing organisms. Using hCMV as a model, we recently developed a gene drive system in herpesviruses that rely on co-infection of cells by a wildtype and an engineered virus (7). Cleavage by Cas9 and repair by homologous recombination leads to the conversion of wildtype viruses into new recombinant viruses (Figure 1A). We demonstrated that gene drive viruses can replace their wildtype counterparts and spread in the viral population in cell culture experiments. Moreover, we showed that a gene drive virus presenting severe replicative defects could spread into the viral population and ultimately reduce viral levels. This development represents a novel therapeutic strategy against herpesviruses.

An important challenge lies in designing gene drive viruses with replicative defects that still spread efficiently in the viral population and ultimately reduce viral levels. In the ideal scenario, gene drive viruses are non-infectious but are complemented by wildtype factors upon co-infection and can spread in the population only as long as wildtype viruses are present. Our initial study targeted *UL23*, a viral gene that encodes a tegument protein involved in immune evasion. UL23 is dispensable in normal cell culture conditions, but necessary in presence of interferon γ (7–9). The tegument of herpesvirus is a layer of protein that lies between the genome-containing capsid and the viral envelope (1, 10). Tegument proteins are released in the cell upon viral entry and often involved in transcriptional activation and immune evasion. Tegument proteins are attractive targets for a viral gene drive because many of them function early in the viral cycle – where they could be complemented by a co-infecting wildtype virus – but only accumulate during later stages of the infection. In the present study, we designed gene drives against several tegument genes and showed that they indeed represent effective targets.

An important aspect of gene drive in sexually reproducing organisms, such as mosquitos, is the appearance and selection of drive-resistant alleles that are no longer recognized by the CRISPR guide RNA (gRNA) (11–13). Such alleles can already exist in the wild population or appear when the target site is repaired and mutated by non-homologous end-joining instead of homologous recombination (Figure 1A). These sequences are “immune” to future conversion into the drive allele, are often positively selected over time, and limit the ability to permanently modify a wildtype population (13–16).

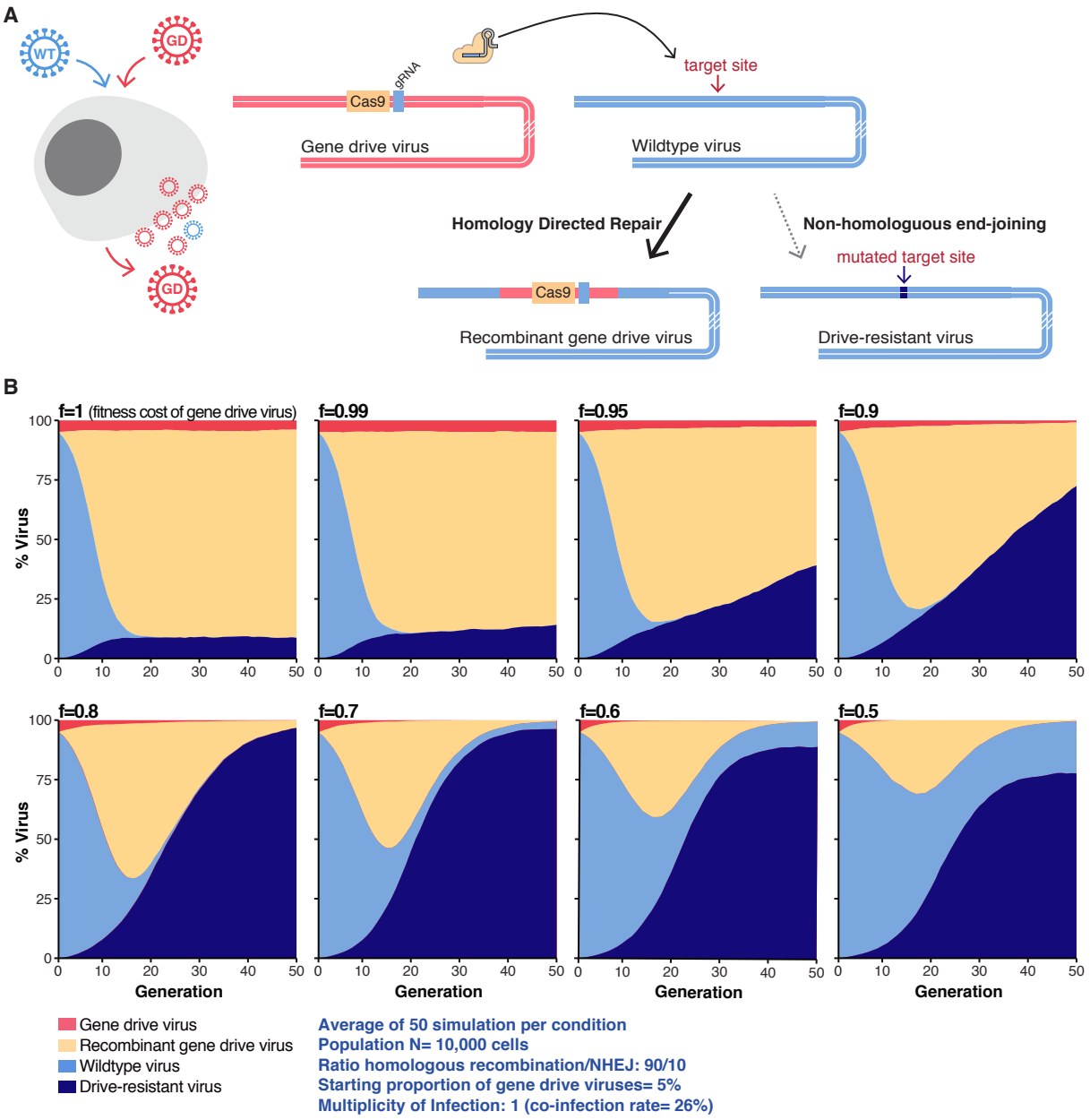


Figure 1: Evolution of gene drive resistance in numerical simulations
A. CRISPR-based gene drive sequences are, at a minimum, composed of Cas9 and a gRNA targeting the complementary wildtype locus. After co-infection of a cell by a wildtype and a gene drive virus, Cas9 targets and cleaves the wildtype sequence. Homology-directed repair of the damaged wildtype locus using the gene drive sequence as a repair template causes the conversion of the wildtype locus into a new gene drive sequence and the formation of new recombinant virus. In 5–10% of cases, repair of the damaged genome by non-homologous end-joining (NHEJ) creates drive-resistant viruses with a mutated target site. **B**. Numerical simulation showing that the appearance and positive selection of drive-resistant viruses depend on the replicative fitness cost f of gene-drive viruses. In these simulations, at each viral generation, N virtual cells were randomly infected and co-infected by N viruses, producing a new generation of viruses. When a cell is co-infected by wildtype and gene-drive viruses, wildtype viruses are converted to new gene drive viruses or resistant viruses in a 90/10 ratio.

In this report, we sought to investigate the appearance and evolution of resistance in a viral gene drive. Through numerical simulations and cell culture experiments with hCMV, we show that, after the initial propagation, a drive-resistant viral population is positively selected and outcompetes gene drive viruses over time. By designing and testing multiple gene drives that disable critical viral genes, we show that targeting evolutionary conserved sequences ensures that drive-resistant viruses have a replication defect that leads to a long-term reduction of viral levels.

Results

Drive-resistant viruses are positively selected over time.

We first attempted to model the evolution of gene drive resistance using numerical simulations. Our initial study indicated that after co-infection with a wildtype and a gene drive viruses, around 5–10% of viruses remained unconverted and had a mutated target site that rendered them resistant to the drive (7). Using a 10% estimate and considering in this scenario that resistant viruses would replicate at the same level as wildtype, we modelled the evolution of the viral population over time (Figure 1B, Supplementary Figure S1). These simulations showed that, depending on the replicative fitness cost of gene drive viruses, a population of drive-resistant viruses is often positively selected over time. At low fitness cost, the gene drive virus first spread efficiently in the wildtype population, but is outcompeted in the long run. At high fitness cost, a resistant population appears quickly and gene drive viruses disappear rapidly. Overall, these simulations predict that the appearance of drive-resistant viruses would prevent the gene drive from permanently remaining in the viral population.

We next evaluated experimentally the evolution of resistance in cell culture. Our initial experimental system involved an unmodified hCMV virus expressing GFP (strain Towne, referred hereafter as Towne-GFP) and an mCherry-expressing gene drive virus (GD-UL23) targeting *UL23* (7) (Figure 2A). *UL23* is dispensable in normal cell culture conditions (8, 9), but the GD-UL23 virus was built in a different viral strain (TB40/E) and replicated significantly slower than Towne-GFP (t-test: p=0.0065, Figure 2B). Human foreskin fibroblasts were co-infected at low multiplicity of infection (MOI=0.1) with the two viruses in different starting proportions, and the infection was subsequently propagated for several weeks. The proportion of the different viruses was evaluated over time by plaque assay (Figures 2C-D). The mCherry reporter enabled us to follow the spread of the gene drive in the viral population: viruses expressing mCherry represent gene drive viruses, and viruses expressing only GFP are unconverted – either unmodified or drive-resistant – viruses. In each of 12 biological replicates and independently of the starting proportion of GD-UL23, wildtype viruses were first converted to new gene drive viruses, and the population of GFP-only viruses reached a minimum level of 5–40% after 10–40 days. In a second phase, however, GFP-only viruses appeared to be positively selected, and their proportion rebounded until they represented the majority of the viral population. This showed that in the long term, drive-resistant viruses are positively selected.

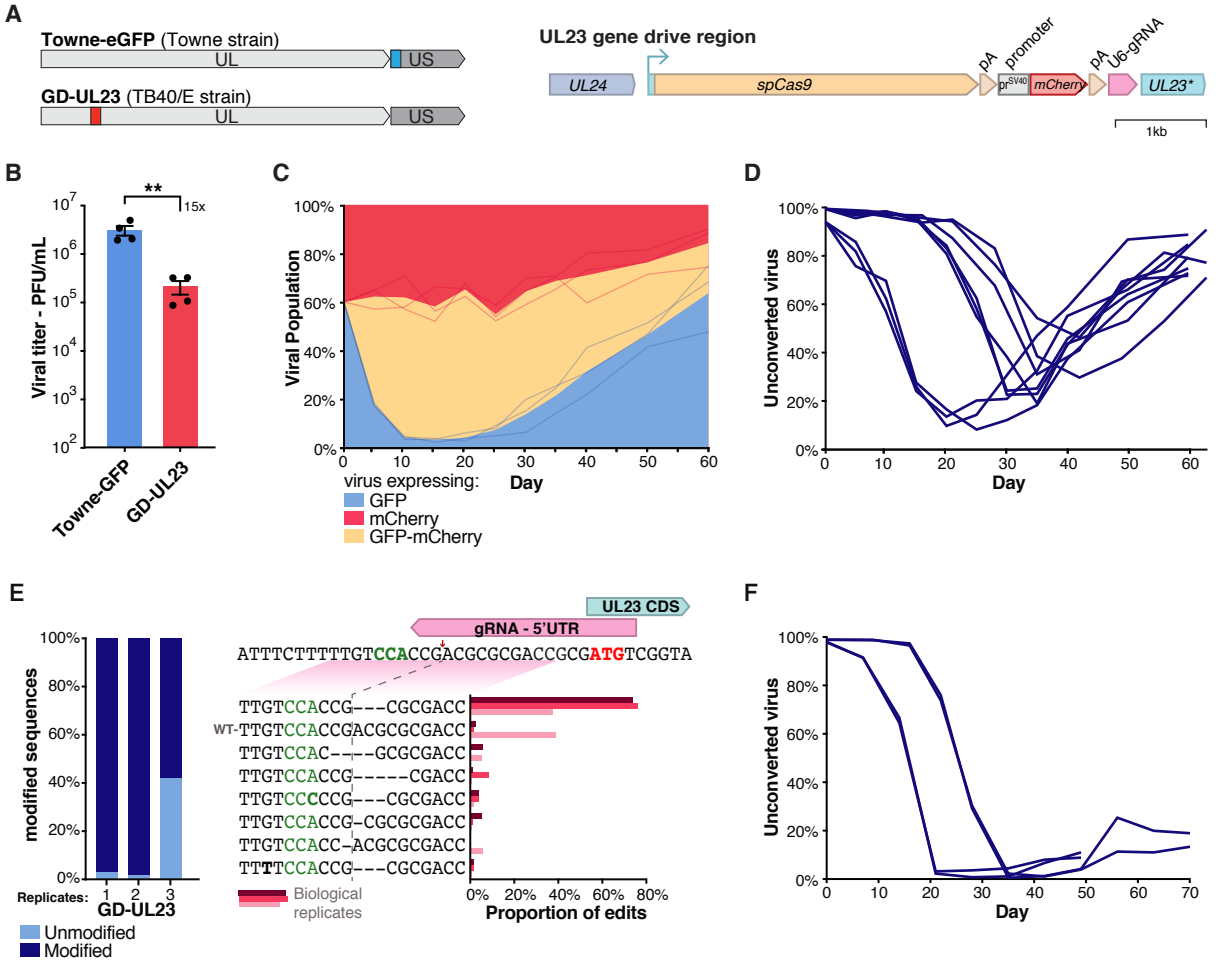


Figure 2: Evolution of resistance in a gene drive against UL23
A. Left: Localizations of gene drive and GFP cassettes on hCMV genomes. UL/US indicates Unique Long/Short genome segments. Right: the gene drive cassette comprises *spCas9* under the control of *UL23* endogenous promoter, followed by an SV40 polyA signal, an SV40 promoter driving an mCherry reporter, a beta-globin polyA signal and a U6-driven gRNA. **B**. Viral titer in fibroblasts infected (MOI=1) with Towne-GFP or GD-UL23 viruses after 7 days, as measured by plaque assay. Titers are expressed in PFU (plaque forming unit) per mL of supernatant. Error bars represent standard error of the mean (SEM) between biological replicates. n=4. **: $p < 0.01$, unpaired t-test. **C**. Proportion over time of viruses expressing GFP alone, mCherry alone, or both as measured by plaque assay, after co-infection with Towne-GFP and GD-UL23. Data show both the mean and the individual trajectory of biological replicates. n=3. **D**. Same as C, except that only the proportion of GFP-only viruses is represented. Every line corresponds to a biological replicate. n=9. **E**. Amplicon sequencing of the CRISPR target site at the end of the drive. Left: proportion of edited genomes, right: Relative contribution of each edits. CRISPR cleavage site is indicated by a red arrow, the protospacer adjacent motif (PAM) is highlighted in green, and *UL23* start codon in red. n=3. **F**. Proportion over time of viruses expressing only GFP after co-infection with Towne-GFP and GD^{Towne}-UL23. n=4.

These first experiments were carried out with two viruses from different strains, which complicate the interpretation of the results. To alleviate the influence of the viral strain, we analyzed the evolution of resistance in a strict Towne background. A gene drive plasmid (with a gRNA targeting *UL23* 5'UTR) was transfected into fibroblasts, and cells were subsequently infected with Towne-eGFP. The population of mCherry-expressing recombinant viruses and of GFP-only (unmodified or resistant) viruses was followed over time. As observed previously, the gene drive cassette spread efficiently until the population of GFP-only viruses reached a minimum after 70 days (Supplementary Figure S2A). The population of drive-resistant viruses at 70 days was characterized by amplicon sequencing of the target site, in three biological replicates (Figure 2E, Supplementary Figure S2B). As expected, the target site was heavily mutated: up to 98% of sequences were modified in some of the replicates. Interestingly, around 80% of sequences had the same 3-bp deletion. A parallel experiment performed with a gRNA targeting *UL23* start codon showed similar results, with a more diverse distribution of edits (Supplementary Figure S2). These results confirmed that drive-resistant viruses accumulate over time, limiting the impact of the drive.

Finally, we isolated and purified a gene drive virus against *UL23* in a Towne strain (GD^{Towne}-UL23). Fibroblasts were co-infected with Towne-GFP and GD^{Towne}-UL23, and the population of unconverted viruses expressing only GFP was analyzed over time. In this experiment, the drive achieved more than 99% penetrance, and the population of unconverted viruses only rebounded slightly and plateaued at around 10–20% (Figure 2F). UL23 is dispensable in normal cell culture conditions. This result showed that a gene drive without fitness cost can permanently and stably invade the wildtype population, as predicted by numerical simulations.

These numerical and experimental results indicate that, after the initial invasion of the wildtype population, a drive-resistant viral population is positively selected over time and outcompetes gene drive viruses. Importantly, however, we showed that, in the absence of associated fitness costs, gene drive sequences can reach almost full penetrance and be maintained indefinitely in the viral population.

Gene drives against conserved regions of *UL26* and *UL35* lead to permanent reduction of viral titers.

Drive-resistant viruses are created by imperfect repair of the CRISPR cleavage site. We reasoned that the appearance and selection of drive-resistant viruses could be circumvented if the mutation rendered viruses nonfunctional, for example if it knocked-out a critical viral gene. In this case, numerical simulation predicted that drive-resistant viruses would also be counter-selected and would not accumulate (Supplementary Figure S3), leading to a long-term reduction of viral titers. We, therefore, designed several gene drives targeting hCMV genes that are necessary for efficient viral replication (summary in Table 1). The gene drive cassette was inserted in the coding sequence of the viral gene. In addition, CRISPR gRNAs were designed in evolutionary conserved sequences, so that any mutation would potentially affect viral fitness. In this situation, both gene drive and drive-resistant viruses would have a replication defect, leading to a long-term reduction of viral levels.

We first built gene drive viruses against *UL122* (IE2), *UL79*, *UL99* and *UL55* (gB). These viral genes are absolutely required for viral replication, and mutant viruses are non-infectious. (17–20). These initial attempts were unsuccessful, and we observed no recombinant virus in co-infection experiments (Table 1). As explained in the introduction, we hypothesized that tegument proteins represent attractive targets. We then designed gene drive plasmids against the tegument genes *UL26*, *UL35*, *UL69* and *UL82,* whose absences lead to more moderate growth defects (21–24). Fibroblasts were independently transfected with the four plasmids, infected with Towne-GFP virus (MOI=1), and the population of recombinant viruses was followed over time (Table 1). We observed that the constructs against *UL26* and *UL35* could spread in the wildtype population, and these genes were selected for subsequent experiments.

Table 1: Summary of gene drive experiments against several viral genes

Gene	Function	Phenotype	Ability to drive
UL23	Tegument gene, involved in evasion of innate immune response (8)	Dispensable (8)	Drive efficiently, up to 99% of conversion
UL122	Encode Immediate early protein IE2, responsible for initiation of viral replication (17)	Essential (17)	Do not drive, no recombinant viruses could be observed
UL79	Regulation of transcription of late viral transcript (25)	Essential (18)	Do not drive, very few recombinant viruses
UL99	Tegument gene, essential for viral envelopment (19)	Essential (19)	Do not drive, no recombinant virus
UL55	Encode fusion protein gB	Essential (20)	Do not drive, no recombinant virus
UL26	Tegument gene, involved in immune evasion, transcriptional activation and virion stability (21, 26–28)	Severe growth defect (21)	Drive efficiently
UL35	Tegument gene, important for viral replication and virion formation (29)	Moderate growth defect (23)	Drive efficiently
UL69	Tegument gene, involved in unspliced mRNA nuclear export and cellular arrest (22, 30)	Severe growth defect (22)	Very limited drive at high MOI
UL82	Tegument gene, stimulate immediate early transcription and inhibit host innate response (24, 31)	Moderate growth defect (24)	Very limited drive at high MOI

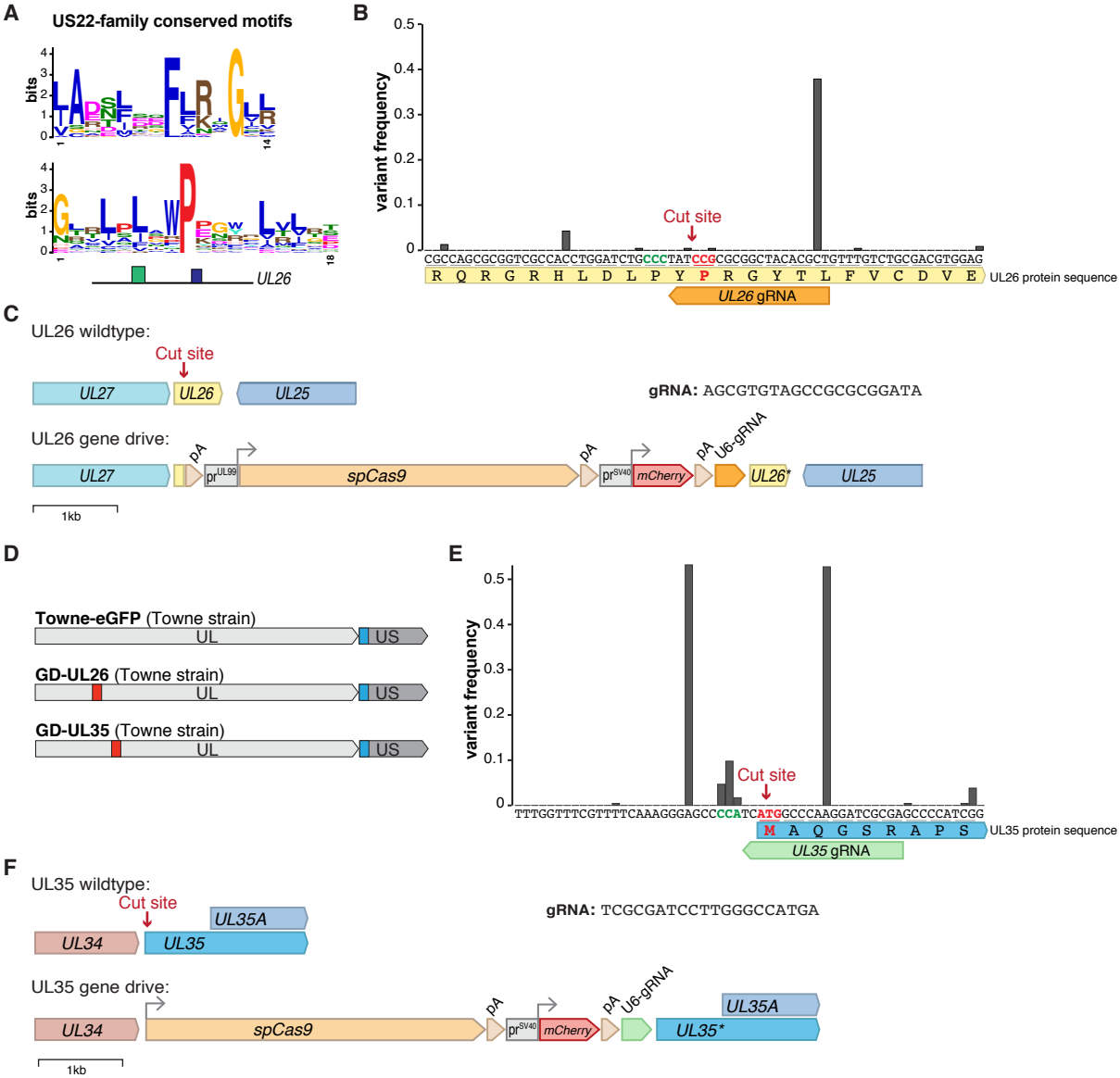


Figure 3: Design of gene drives against *UL26* and *UL35*
A. Logo sequences of the two most conserved motifs in the US22 family of herpesviral proteins. Bottom: localization of the two motifs on UL26 protein (blue: first motif, green: second motif). **B**. Frequency of genetic variants around *UL26* target site, from 235 hCMV strains. CRISPR protospacer adjacent motif (PAM) is highlighted in green, and the targeted proline codon in red. **C**. Wildtype and gene drive *UL26* regions. The gene drive cassette comprises an HSV1-TK polyA signal, *spCas9* under the control of an *UL99* late viral promoter, followed by a SV40 polyA signal, an SV40 promoter driving a mCherry reporter, a beta-globin polyA signal and a U6-driven gRNA. **D**. Localizations of gene drive and GFP cassette on hCMV genomes. UL/US indicates Unique Long/Short genome segments. **E**. Frequency of genetic variants around UL35 target site, from 235 hCMV strains. CRISPR PAM is highlighted in green, and the targeted start codon in red. **F**. Wildtype and gene drive *UL35* region. The gene drive cassette is the same as in C, except that *spCas9* expression is controlled by UL35 endogenous promoter.

CRISPR gRNAs against *UL26* and *UL35* were chosen in sequences evolutionarily conserved at the DNA and amino-acid levels. UL26 is a tegument protein involved in immune evasion, transcriptional activation and virion stability (21, 26–28). It is a member of the US22 family of herpesviral proteins (32), a family of proteins with conserved motifs across several herpesviruses. We aligned 31 protein sequences of US22 family members and screened for conserved motifs (Figure 3A, Supplementary Figure S4). Two highly conserved motifs were found, and we chose to design a gRNA that would disrupt a very conserved proline in the second motif. In parallel, available DNA sequences for 235 hCMV clinical and laboratory viruses were aligned, and the frequency of variants compared to the Towne reference was calculated (Figure 3B). The gRNA against *UL26* was chosen in a region of low variation, with less than 1% of sequenced hCMV viruses having a polymorphism in the first 18 bp of the gRNA sequence. A highly variable site was found in the gRNA sequence, but at a position (the 19th base) that presumably doesn't efficiently prevent DNA cleavage (33). Most repairs of the predicted cut site would disrupt the evolutionary conserved proline. The gene drive cassette comprised *Cas9*, an mCherry reporter and a U6-driven gRNA (Figures 3C-D). Of note, because the CRISPR target site against *UL26* was located far from the start codon (around 200 bp), a polyadenylation signal and a promoter with late kinetics (from hCMV *UL99*) were added to the gene drive cassette upstream of *Cas9*. The gRNA against *UL35* was designed in a similar manner. *UL35* encodes two isoforms with a common C-terminal domain but different N-termini (23, 34). Mutant viruses lacking the longer isoform have a modest replicative defect, and the gRNA against *UL35* targeted this first start codon, so that any mutations at the predicted cut site would abrogate translation of the long isoform (Figures 3D-F). This gRNA could not be designed in a region of low genetic variation (Figure 3E). This finding has little consequence for cell culture experiments, but suggests that this particular gene drive virus could be less efficient at targeting clinical hCMV strains.

Infectious gene drive viruses against *UL26* and *UL35* (GD-UL26 and GD-UL35, in Towne strain) were isolated and purified in fibroblasts. Of note, mutants lacking UL26 have severe growth defects (21) and the isolation and plaque assay of GD-UL26 viruses was carried on in cells stably expressing UL26. In wildtype fibroblast, GD-UL26 replication was almost completely abrogated compared to Towne-GFP (t-test: p=0.0005), while GD-UL35 replicated with a moderate but significant ninefold growth defect (t-test: p=0.0083), with viral plaques 50% smaller in average (t-test: p=0.0006) (Figures 4A-B).

To analyze the long-term evolution of a gene drive against *UL26* and *UL35,* fibroblasts were co-infected with Towne-GFP and either GD-UL26 or GD-UL35 in a 99/1% ratio, and compared to cells infected with only Towne-GFP (MOI=1). Viral titers and the population of unconverted GFP-only viruses representing unmodified or drive-resistant viruses was followed over time in four independent replicates (Figures 4C-E). In both situations and similarly to previous experiments, the drive first spread efficiently and the proportion of GFP-only viruses reached a minimum around 20% (for GD-UL35) or 50% (for GD-UL26). In a second time that corresponded to the positive selection of drive-resistant viruses, the population of GFP-only viruses rebounded, and gene drive viruses almost disappeared. Importantly, in both cases, viral titers dropped importantly when the drive reached its

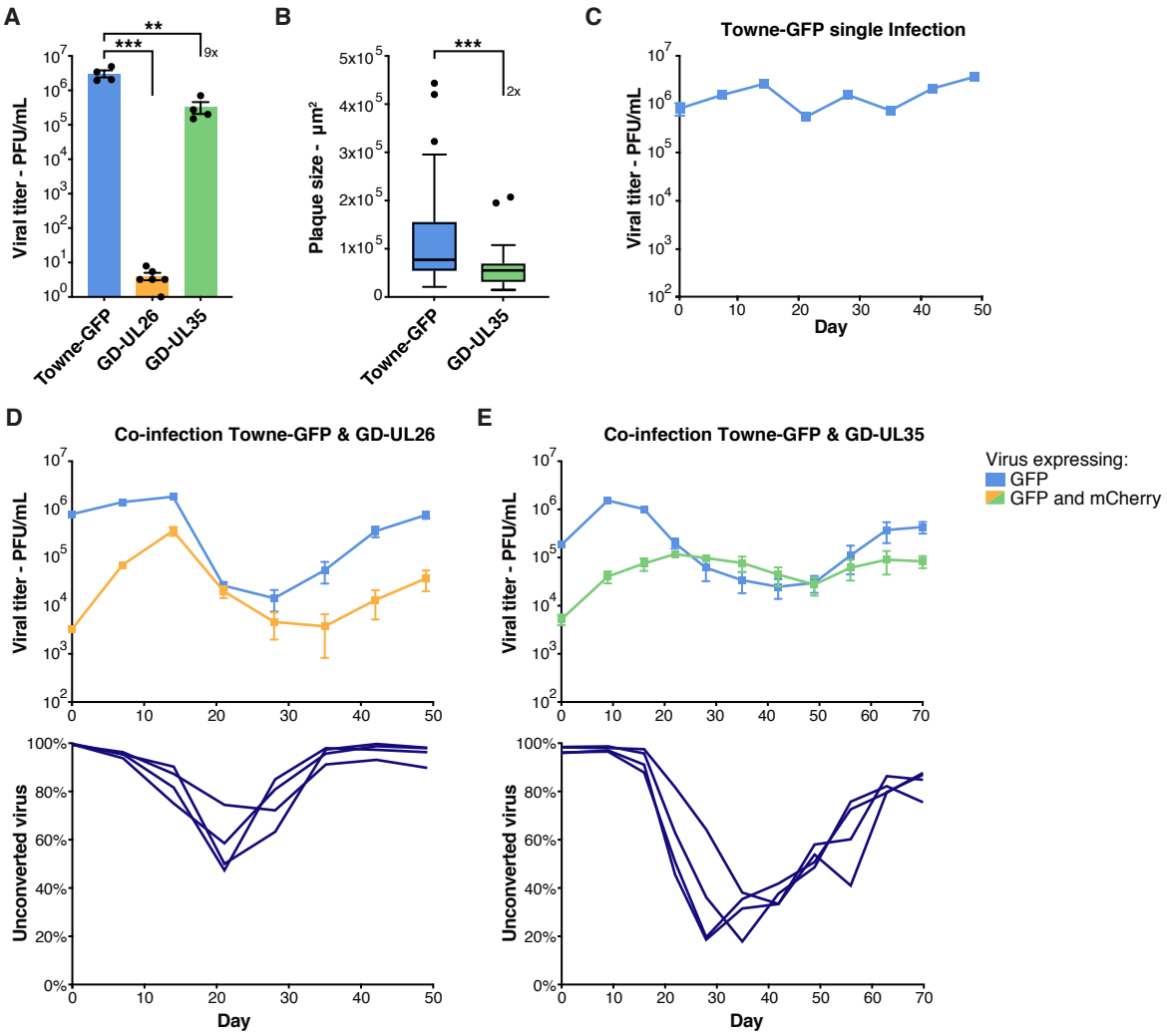


Figure 4: Gene drives against *UL26* and *UL35* spread in the viral population
A. Viral titer in fibroblasts infected (MOI=1) with Towne-GFP, GD-UL26 or GD-UL35 viruses after 7 days. n=4. **B**. Plaque size in fibroblasts infected with Towne-GFP or GD-UL35, after 8 days. Tukey box plot, n=40. **C**. Viral titer over time in fibroblasts infected with Towne-GFP. At each time point, supernatant was used to infect fresh cells to propagate the infection. n=4. **D-E**. Upper panels: Viral titer over time in fibroblasts co-infected with Towne-GFP and GD-UL26 (left) or GD-UL35 (right). Viruses either express GFP only, or both GFP and mCherry. Viruses expressing mCherry represent gene drive viruses. Lower panels: proportion of the viral population expressing only GFP, representing unconverted (unmodified or drive-resistant) viruses. n=4.
Titers were measured by plaque assay and are expressed in PFU (plaque forming unit) per mL of supernatant. Error bars represent SEM between biological replicates. Asterisks summarize results of unpaired t-tests. **: $p < 0.01$, ***: $p < 0.001$.

maximum, and remained lower until the end of the experiment when compared to cells infected with only Towne-GFP (Figure 4C). This finding suggested that drive-resistant viruses had a permanent replicative defect.

To investigate the population of drive-resistant viruses, viral clones resistant to either GD-UL26 or GD-UL35 and originating from three independent co-infection experiments were isolated and purified. PCR and Sanger sequencing of 11 viral clones for each condition revealed that the target site was mutated as expected (Figures 5A-B). For *UL26*, six clones had an out-of-frame mutation that would disrupt translation, and the five others had an in-frame mutation that nonetheless mutated the conserved proline. On the other hand, every clone resistant to GD-UL35 had a mutated start codon that would prevent translation of the long UL35 isoform. Interestingly, nine of 11 *UL35*-resistant clones had an identical 26 bp deletion that can probably be explained by the presence of microhomology segments on both sides of the cleavage site. The viral titer of these drive-resistant viruses was then compared to Towne-GFP. Viruses resistant to GD-UL26 had a significant 10-fold reduction of viral titers on average (t-test, p=0.0007, Figure 5C). Clones with out-of-frame mutations were in average more severely impaired (30-fold reduction, p=0.0057) than in-frame mutants (fivefold, p=0.0319), but the significant defect of these in-frame mutants confirmed that the targeted proline is important for viral fitness. Viruses resistant to GD-UL35 appeared to have a slight twofold replicative defect, compared to Towne-GFP, but this difference didn't reach statistical significance (p=0.1354, Figure 5D). These results indicated that drive-resistant viruses had acquired long-term replicative defects.

Our results show that a strategy of designing gene drives against conserved and critical viral sequences can overcome the evolution of resistance and lead to long-term reduction of viral levels.

Discussion

In this report, we analyzed the evolution of resistance in a viral gene drive against hCMV. Using multiple examples, we showed that, after the successful invasion of the wildtype population, drive-resistant viruses with a mutated target site are positively selected and outcompete gene drive viruses. These numerical and experimental results mimic perfectly what is observed in insect experiments (11–13): the rapid evolution of resistance prevents the fixation of the modified sequence in the population.

Gene drives in mosquitoes generally follow one of two basic strategies: population-suppression drives aim to eliminate the targeted insect population, whereas population-replacement drives aim to replace wild populations with engineered, often pathogen-resistant, animals (35). Population-replacement drives could be imagined in viruses, where for example a drug-responsive gene would be introduced in the viral population, but here we attempted to design population-suppression viral drives that lead to significant reduction of viral levels. The difficulty lies in designing gene drive viruses with replicative defects that spread efficiently. Our previous study focused on a gene drive against *UL23*, a viral gene that is dispensable in normal cell culture conditions (7, 8). UL23 is involved in immunity

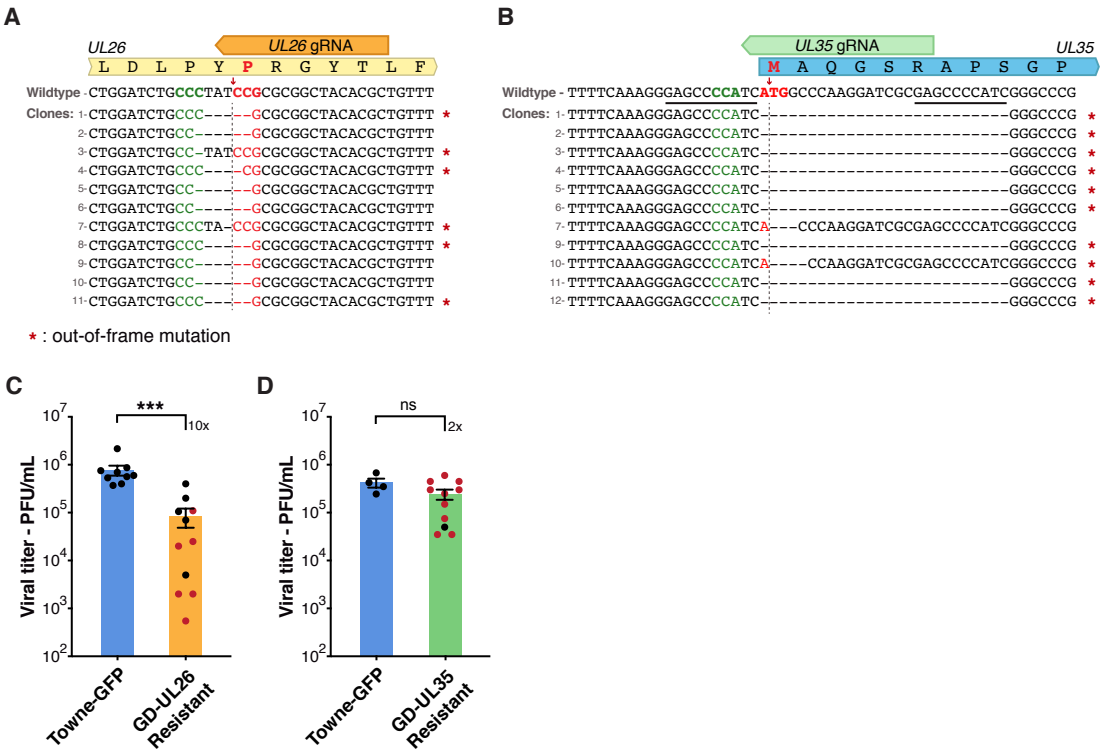


Figure 5: Drive-resistant viruses have long-term replicative defects
A-B. Sanger sequencing of the target site of 11 viral clones resistant to either GD-UL26 (left) or GD-UL35 (right). CRISPR cleavage sites are shown by red arrows, PAMs are highlighted in green and targeted codons in red. Red asterisks (*) indicate out-of-frame mutations. Microhomology sequences around *UL35* cleavage site are underlined in black. **C**. Viral titer in fibroblasts infected (MOI=0.1) with Towne-GFP or viruses resistant to GD-UL26 after 10 days. n=9 (Towne-GFP) or n=11 (GD-UL26 resistant). **D**. Viral titer in fibroblasts infected (MOI=1) with Towne-GFP or viruses resistant to GD-UL35 after 7 days. n=4 (Towne-GFP) or n=11 (GD-UL35 resistant).
Replicates with out-of-frames mutations are highlighted in red. Titers were measured by plaque assay and are expressed in PFU (plaque forming unit) per mL of supernatant. Error bars represent SEM between biological replicates. Asterisks summarize results of unpaired t-tests. ***: $p < 0.001$.

evasion, and the replication of *UL23* mutant viruses is severely impaired by interferon γ. We showed that a gene drive against *UL23* could still spread in the presence of interferon γ, which gave us the first indication that a gene drive with a replicative defect could spread. Here, we designed and tested gene drives against eight additional viral genes that are necessary for efficient viral infection independently of the culture condition (Table 1). We showed that gene drives against the tegument genes *UL26* or *UL35* could spread efficiently in the viral population. GD-UL35 only had a moderate replicative defect, but the GD-UL26 virus was almost non-infectious. However, co-infection with unmodified Towne-GFP virus efficiently complemented the defective viruses and enabled their replication. The modification spread efficiently into the viral population as long as wildtype viruses were present, but it was unable to propagate further afterward. This represents an important example for the design of suppression drives against herpesviruses.

In insects, targeting evolutionarily conserved sequences, such as the *Doublesex* gene, mitigates the appearance of drive-resistant sequences (16, 36, 37), and our approach followed similar principles. Some regions of the hCMV genome are highly variable, but others show a high degree of sequence conservation, and both *UL26* and *UL35* are reported to be among the most conserved hCMV genes (38). We designed gRNAs in regions of low genetic variation, and targeted sequences encoding conserved amino acids. GD-UL26 and GD-UL35 are examples of gene drives with high and low fitness costs, respectively. As predicted by numerical simulations (Figures 1B, 4D-E), drive-resistant viruses were selected rapidly in the gene drive against *UL26*, and GD-UL26 didn't reach more than 50% of the population. By contrast, the drive against UL35 had a higher penetrance and reached up to 80% of the population. In both cases, drive-resistant viruses had mutations that can be predicted to seriously affect the fitness of the virus. Viruses resistant to GD-UL26 had a significant growth defect compared to wildtype viruses. Absence of UL35 is reported to cause a modest growth defect (23), and GD-UL35-resistant viruses lacking UL35 start codon were slightly attenuated. This result was not statistically significant in our cell culture experiments, but absence of UL35 would likely have important consequences *in vivo.* In summary, this work presents important proofs of principle for the design of viral gene drives. It demonstrates that both gene drive and drive-resistant viruses can have replication defects, leading to a long-term reduction of viral levels.

Our hope is to ultimately design similar gene drive systems that could be used to treat herpesvirus diseases. Patients infected with a herpesvirus and unable to control it could be superinfected with a gene drive virus that would reduce the infection. hCMV reactivation in immunosuppressed patients after organ transplant could be one use (39, 40). How a viral gene drive spreads *in vivo* and how the immune system reacts to superinfection will have to be studied in animal models. Our work nonetheless brings important considerations about viral dynamics. In particular, in our gene drives against *UL26* or *UL35*, viral levels dropped importantly when drive penetrance reached its maximum, with a 10–100-fold reduction of viral titers (Figures 4D-E). This transient drop of viral levels could have important implications *in vivo,* as it could give the immune system a transient window to control the infection. Similarly, even a small decrease of viral fitness caused by drive-resistant mutations could have huge benefits for the infected patient. Indeed, *in vivo*, a successful gene drive

wouldn't require to reduce viral levels significantly by itself, but to do so just enough for the immune system to take control of the infection.

Existing genetic variation in hCMV or other herpesviruses would hamper the capacity of a gene drive to efficiently target wild viruses in infected patients. We designed CRISPR gRNAs in regions of low genetic diversity, but the number of available genomes (235) is small compared to the size of the hCMV population. A better assessment of the genetic diversity of herpesviruses will be necessary before a gene drive can be used against wild viruses. Nonetheless, it is easy to imagine that patients could be treated with an array of gene drive viruses, each targeting different variants or different locations. Such a strategy would increase redundancy and limit the probability that variants could escape the drive.

As a final note, the development of gene drives raises important ecological and biosafety concerns, and our approach follows the guidelines established by the NIH and the National Academy of Science (41, 42). In particular, our work was conducted using laboratory viral strains unable to infect human hosts (43), and thus, there were no risks of inadvertent release of gene drive viruses into the wild.

Material and Methods

Cells and viruses

Human foreskin fibroblast cells were obtained from the ATTC (#SCRC-1041) and cultured in DMEM (10-013-CV, Corning, Corning, NY, USA), supplemented with 10% FBS (Sigma-Aldrich, USA) and 100 µm/L penicillin/streptomycin (Corning, USA). Cells were regularly tested negative for mycoplasma and used between passages 3 and 15.

hCMV TB40/E-Bac4 (44) and Towne-GFP (T-BACwt)(17) were kindly provided by Edward Mocarski (Emory University, USA). Viral stocks were prepared and plaque assays performed exactly as reported (7).

Co-infection experiments were performed by co-infecting with wildtype Towne-GFP and gene drive viruses for 1 h, with a total MOI of 0.1–1. Experiments were conducted using 12-well plates with 1 mL of medium per well. For time-course experiments over multiple weeks, 100–200 µL of supernatant was used to inoculate fresh cells for 1 h before changing media.

For plaque size analysis, images of around 40 fluorescent viral plaques were acquired with a Nikon Eclipse Ti2 inverted microscope and Nikon acquisition software (NIS-Element AR 3.0). Plaque size in pixels was measured using ImageJ (v2.1.0) and then converted to μm^2.

Cloning and generation of gene drive viruses

The gene drive construct against *UL23* was as described (7). The core gene drive cassette comprises a codon-optimized *SpCas9* (from *Streptococus pyogenes*), followed by an mCherry fluorescent reporter and a U6-driven gRNA (Figure 2A). This cassette is surrounded by homology arms specifics to the site of integration. Gene drive plasmids against UL122, UL79, UL99, UL55, UL26, UL35, UL82 and UL69 were built by serial modifications of the UL23 gene drive plasmid. Briefly, homology arms and the gRNA for UL23 locus were removed by restriction enzyme digestion, and replaced by new homology arms and gRNAs by Gibson cloning (NEB, USA), using PCR products or synthesized DNA fragments (GeneArt™ String™ fragments, ThermoFisher, USA).

Cell lines stably expressing *UL26* or other viral genes were generated using lentiviral constructs. Lentivirus expression plasmids were cloned by serial modifications of Addgene plasmid #84832 (45), using digestion/ligation and gibson cloning. The final constructs expressed the viral gene of interest, followed by in-frame puromycin and BFP reporters interleaved with self-cleavable 2A peptides under a EF1a promoter. Lentiviruses were produced in HEK293 cells using standard protocols as reported (46). Fibroblasts were then transduced with lentiviruses and selected with 1 µg/mL puromycin for 1 week before being used (ant-pr-1, Invivogen, USA).

Purification of gene drive virus was performed as reported (7). Briefly, fibroblasts were transfected by nucleofection (Kit V4XP-2024, Lonza, Switzerland) with the gene drive plasmid, and infected 48 h later with Towne-GFP virus. Recombinant viruses expressing mCherry were isolated and purified by several rounds of serial dilutions and plaque purification.

Viral clones resistant to either *UL26* or *UL35* gene drives were isolated by plaque purification of GFP-only viruses at the end of co-infection experiments. Mini-stocks were titrated by plaque assays, and the sequence of the target site was analyzed by PCR and Sanger sequencing. (UL26 primers: F: GGCGCGTTATAAGCACCGTGG, R: GCCGATGACGCGCAACTGA; UL35 primers: F: ACGTCACTGGAGAACAATAAAGCGT, R: GGCACGCCAAAGTTGAGCAG). 12 resistant clones originating from 3 independent experiments were first isolated, but in both cases only 11 could be successfully purified and sequenced.

Amplicon sequencing

Total DNA was extracted from infected cells with Qiagen DNeasy kit. A 470 bp PCR product surrounding the *UL23* cut site was amplified using Phusion high-fidelity polymerase (NEB, USA) and column purified (Macherey-Nagel, Germany). Primers contained a 5'-overhang compatible with Illumina NGS library preparation. Amplicons were pooled and sequenced on an Illumina Miseq (2x300 paired-end). Library preparation and sequencing was performed by SeqMatic (Fremont, CA, USA). Analysis of genome editing outcomes from sequencing data was generated using CRISPResso2 software pipeline (34). Forward primer: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTTGGGGCATAAAACACCG; Reverse primer: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCCAGGTACAGTTCAGACGG.

Sequence alignment and motif analysis

Protein sequences of the herpesviridae US22 protein family were downloaded from Uniprot and curated to remove duplicates. Motif discovery among the 31 protein sequences was performed using the Meme Suite 5.3.0 (http://meme-suite.org/), using OOPS parameter and looking for motifs 10–30 in lengths (47).

Full-length genome sequences of 235 hCMV clinical and laboratory strains were downloaded from the NIAID Virus Pathogen Database and Analysis Resource (ViPR: https://www.viprbrc.org/)(48). The frequencies of variants around the CRISPR target site were calculated using simple Python and R scripts. Briefly, we recovered the sequences around the target site, using BLAST locally with the sequence of the target site as the query and the 235 hCMV sequence as the search database. Base frequencies on the aligned sequences were calculated and plotted using R packages APE v5.3 and ggplot2 v3.3.0 (49, 50).

Statistics and reproducibility

Analyses were run using GraphPad Prism version 8.1.1 for macOS (GraphPad Software, USA, www.graphpad.com). The only statistical test used in this study was unpaired two-tailed t-test to compare wildtype and modified viruses. Exact p-values and summary are reported in the text and in the source data.

Numerical simulations

Numerical simulations of viral gene drive were computed using a simplified viral replication model. Briefly, in each viral generation, N virtual cells were randomly infected and co-infected by N viruses (MOI =1), producing a new generation of viruses. In this new generation, wildtype viruses co-infected with drive viruses were converted to new gene-drive viruses or resistant viruses with a ratio of 90/10. Gene drive viruses replicate with a fitness

cost f, and the coinfection rate is calculated from the MOI assuming a Poisson distribution. The code and a more thorough description are available at https://github.com/mariuswalter/ViralDrive.

Data and code availability

The data supporting the findings of this study are available within the paper and its supplementary files. Amplicon Sequencing data have been deposited in the Short Read Archive with BioProject accession no. PRJNA556897. Lentivirus expression plasmids for *UL23, UL122, UL79, UL99, UL55, UL26* and *UL23* will be deposited in Addgene. Viruses and other reagents developed in this study are available upon request and subject to standard material transfer agreements with the Buck Institute. Source data are provided with this paper. Any other relevant data are available upon reasonable request. Code developed for numerical simulations is available on GitHub (https://github.com/mariuswalter/ViralDrive).

Acknowledgement

We thank Edward Mocarski (Emory University) for providing viral stocks. We thank members of the Verdin laboratory for technical and conceptual help. This study was funded through institutional support from the Buck Institute for Research on Aging.

Author contributions

M.W. designed the study. M.W. and R.P. conducted experiments. E.V. supervised and funded the project. M.W. and E.V. wrote the manuscript.

Competing interests

A patent application describing the use of a gene drive in DNA viruses has been filed by the Buck Institute for Research on Aging (Application number PCT/US2019/034205, pending, inventor: M.W.). E.V. and R.P. declare no competing interests.

References

1. P. E. Pellett, B. Roizman, “Herpesviridae” in *Fields Virology*, D. M. Knipe, P. M. Howley, Eds. (Wolters Kluwer Health, 2013), pp. 1802–1822.
2. C. J. Burrell, C. R. Howard, F. A. Murphy, “Chapter 17 - Herpesviruses” in *Fenner and White’s Medical Virology (Fifth Edition)*, C. J. Burrell, C. R. Howard, F. A. Murphy, Eds. (Academic Press, 2017), pp. 237–261.
3. A. Burt, Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proceedings. Biol. Sci* **270** (2003).
4. V. M. Gantz, *et al.*, Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. *Proceedings of the National Academy of Sciences* **112**, E6736–E6743 (2015).
5. A. Hammond, *et al.*, A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nat. Biotechnol.* **34**, 78–83 (2016).
6. J. Champer, A. Buchman, O. S. Akbari, Cheating evolution: engineering gene drives to manipulate the fate of wild populations. *Nat. Rev. Genet.* **17**, 146–159 (2016).
7. M. Walter, E. Verdin, Viral gene drive in herpesviruses. *Nat. Commun.* **11**, 4884 (2020).
8. L. Feng, *et al.*, Human cytomegalovirus UL23 inhibits transcription of interferon-γ stimulated genes and blocks antiviral interferon-γ responses by interacting with human N-myc interactor protein. *PLoS Pathog.* **14**, e1006867 (2018).
9. W. Dunn, *et al.*, Functional profiling of a human cytomegalovirus genome. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 14223–14228 (2003).
10. R. F. Kalejta, Tegument proteins of human cytomegalovirus. *Microbiol. Mol. Biol. Rev.* **72**, 249–65, table of contents (2008).
11. J. Champer, *et al.*, Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. *PLoS Genet.* **13**, e1006796 (2017).
12. A. M. Hammond, *et al.*, The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. *PLoS Genet.* **13**, e1007039 (2017).
13. R. L. Unckless, A. G. Clark, P. W. Messer, Evolution of Resistance Against CRISPR/Cas9 Gene Drive. *Genetics* **205**, 827–841 (2017).
14. C. Noble, B. Adlam, G. M. Church, K. M. Esvelt, M. A. Nowak, Current CRISPR gene drive systems are likely to be highly invasive in wild populations. *Elife* **7** (2018).
15. D. W. Drury, A. L. Dapper, D. J. Siniard, G. E. Zentner, M. J. Wade, CRISPR/Cas9 gene drives in genetically variable and nonrandomly mating wild populations. *Science Advances* **3**, e1601910 (2017).
16. J. M. Marshall, A. Buchman, H. M. Sánchez C., O. S. Akbari, Overcoming evolved resistance to population-suppressing homing-based gene drives. *Sci. Rep.* **7**, 3776 (2017).
17. A. Marchini, H. Liu, H. Zhu, Human cytomegalovirus with IE-2 (UL122) deleted fails to express early lytic genes. *J. Virol.* **75**, 1870–1878 (2001).
18. Y.-C. Perng, Z. Qian, A. R. Fehr, B. Xuan, D. Yu, The human cytomegalovirus gene UL79 is required for the accumulation of late viral transcripts. *J. Virol.* **85**, 4841–4852 (2011).
19. M. C. Silva, Q.-C. Yu, L. Enquist, T. Shenk, Human cytomegalovirus UL99-encoded pp28 is required for the cytoplasmic envelopment of tegument-associated capsids. *J. Virol.* **77**, 10594–10605 (2003).
20. M. K. Isaacson, T. Compton, Human cytomegalovirus glycoprotein B is required for virus entry and cell-to-cell spread but not for virion attachment, assembly, or egress. *J. Virol.* **83**, 3891–3903 (2009).
21. K. Lorz, *et al.*, Deletion of open reading frame UL26 from the human cytomegalovirus genome results in reduced viral growth, which involves impaired stability of viral particles. *J. Virol.* **80**, 5423–5434 (2006).

22. M. L. Hayashi, C. Blankenship, T. Shenk, Human cytomegalovirus UL69 protein is required for efficient accumulation of infected cells in the G1 phase of the cell cycle. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 2692–2696 (2000).

23. G. Maschkowitz, S. Gärtner, H. Hofmann-Winkler, H. Fickenscher, M. Winkler, Interaction of Human Cytomegalovirus Tegument Proteins ppUL35 and ppUL35A with Sorting Nexin 5 Regulates Glycoprotein B (gpUL55) Localization. *J. Virol.* **92** (2018).

24. W. A. Bresnahan, T. E. Shenk, UL82 virion protein activates expression of immediate early viral genes in human cytomegalovirus-infected cells. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 14506–14511 (2000).

25. Y.-C. Perng, *et al.*, Human Cytomegalovirus pUL79 Is an Elongation Factor of RNA Polymerase II for Viral Gene Transcription. *PLoS Pathog.* **10**, e1004350 (2014).

26. J. Munger, D. Yu, T. Shenk, UL26-deficient human cytomegalovirus produces virions with hypophosphorylated pp28 tegument protein that is unstable within newly infected cells. *J. Virol.* **80**, 3541–3548 (2006).

27. T. Stamminger, *et al.*, Open reading frame UL26 of human cytomegalovirus encodes a novel tegument protein that contains a strong transcriptional activation domain. *J. Virol.* **76**, 4836–4847 (2002).

28. Y. J. Kim, *et al.*, Consecutive Inhibition of ISG15 Expression and ISGylation by Cytomegalovirus Regulators. *PLoS Pathog.* **12**, e1005850 (2016).

29. K. Schierling, C. Buser, T. Mertens, M. Winkler, Human cytomegalovirus tegument protein ppUL35 is important for viral replication and particle formation. *J. Virol.* **79**, 3084–3096 (2005).

30. P. Lischka, Z. Toth, M. Thomas, R. Mueller, T. Stamminger, The UL69 transactivator protein of human cytomegalovirus interacts with DEXD/H-Box RNA helicase UAP56 to promote cytoplasmic accumulation of unspliced RNA. *Mol. Cell. Biol.* **26**, 1631–1643 (2006).

31. Y.-Z. Fu, *et al.*, Human Cytomegalovirus Tegument Protein UL82 Inhibits STING-Mediated Signaling to Evade Antiviral Immunity. *Cell Host Microbe* **21**, 231–243 (2017).

32. M. S. Chee, *et al.*, Analysis of the Protein-Coding Content of the Sequence of Human Cytomegalovirus Strain AD169 in *Cytomegaloviruses*, (Springer Berlin Heidelberg, 1990), pp. 125–169.

33. T. Zheng, *et al.*, Profiling single-guide RNA specificity reveals a mismatch sensitive core sequence. *Sci. Rep.* **7**, 40638 (2017).

34. Y. Liu, B. J. Biegalke, The human cytomegalovirus UL35 gene encodes two proteins with different functions. *J. Virol.* **76**, 2460–2468 (2002).

35. A. E. Williams, A. W. E. Franz, W. R. Reid, K. E. Olson, Antiviral Effectors and Gene Drive Strategies for Mosquito Population Suppression or Replacement to Mitigate Arbovirus Transmission by Aedes aegypti. *Insects* **11** (2020).

36. A. Adolfi, *et al.*, Efficient population modification gene-drive rescue system in the malaria mosquito Anopheles stephensi. *Nat. Commun.* **11**, 5553 (2020).

37. K. Kyrou, *et al.*, A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes. *Nat. Biotechnol.* **36**, 1062–1066 (2018).

38. S. Sijmons, *et al.*, High-throughput analysis of human cytomegalovirus genome diversity highlights the widespread occurrence of gene-disrupting mutations and pervasive recombination. *J. Virol.* **89**, 7673–7695 (2015).

39. L. K. Dropulic, J. I. Cohen, Severe viral infections and primary immunodeficiencies. *Clin. Infect. Dis.* **53**, 897–909 (2011).

40. G. Haidar, N. Singh, Viral infections in solid organ transplant recipients: novel updates and a review of the classics. *Curr. Opin. Infect. Dis.* **30**, 579–588 (2017).

41. E. A. M. National Academies of Sciences, *Gene Drives on the Horizon* (National Academies Press, 2016).

42. C. Emerson, S. James, K. Littler, F. (fil) Randazzo, Principles for gene drive research. *Science* **358**, 1135–1136 (2017).

43. G. W. G. Wilkinson, *et al.*, Human cytomegalovirus: taking the strain. *Med. Microbiol. Immunol.* **204**, 273–284 (2015).

44. C. Sinzger, *et al.*, Cloning and sequencing of a highly productive, endotheliotropic virus strain derived from human cytomegalovirus TB40/E. *J. Gen. Virol.* **89**, 359–368 (2008).

45. M. A. Horlbeck, *et al.*, Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *Elife* **5**, e19760 (2016).

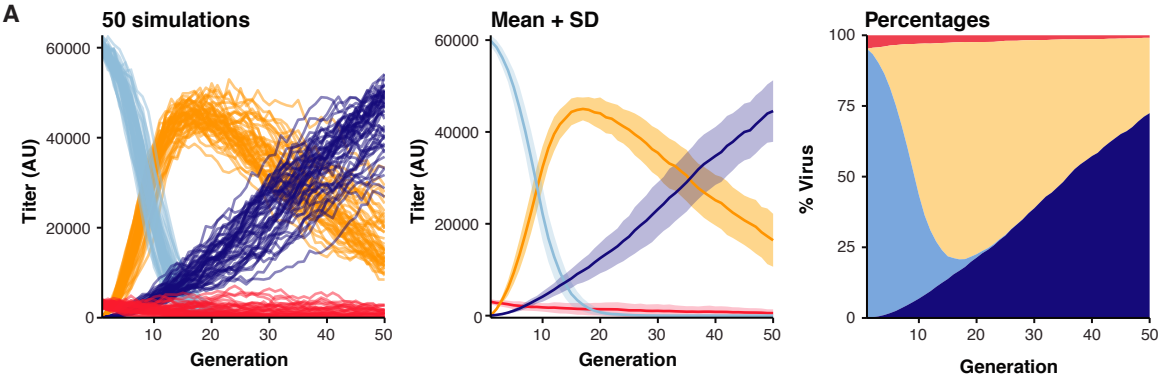
46. A. Vallejo-Gracia, *et al.*, FOXO1 promotes HIV latency by suppressing ER stress in T cells. *Nat Microbiol* (2020) https:/doi.org/10.1038/s41564-020-0742-9.

47. T. L. Bailey, *et al.*, MEME Suite: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).

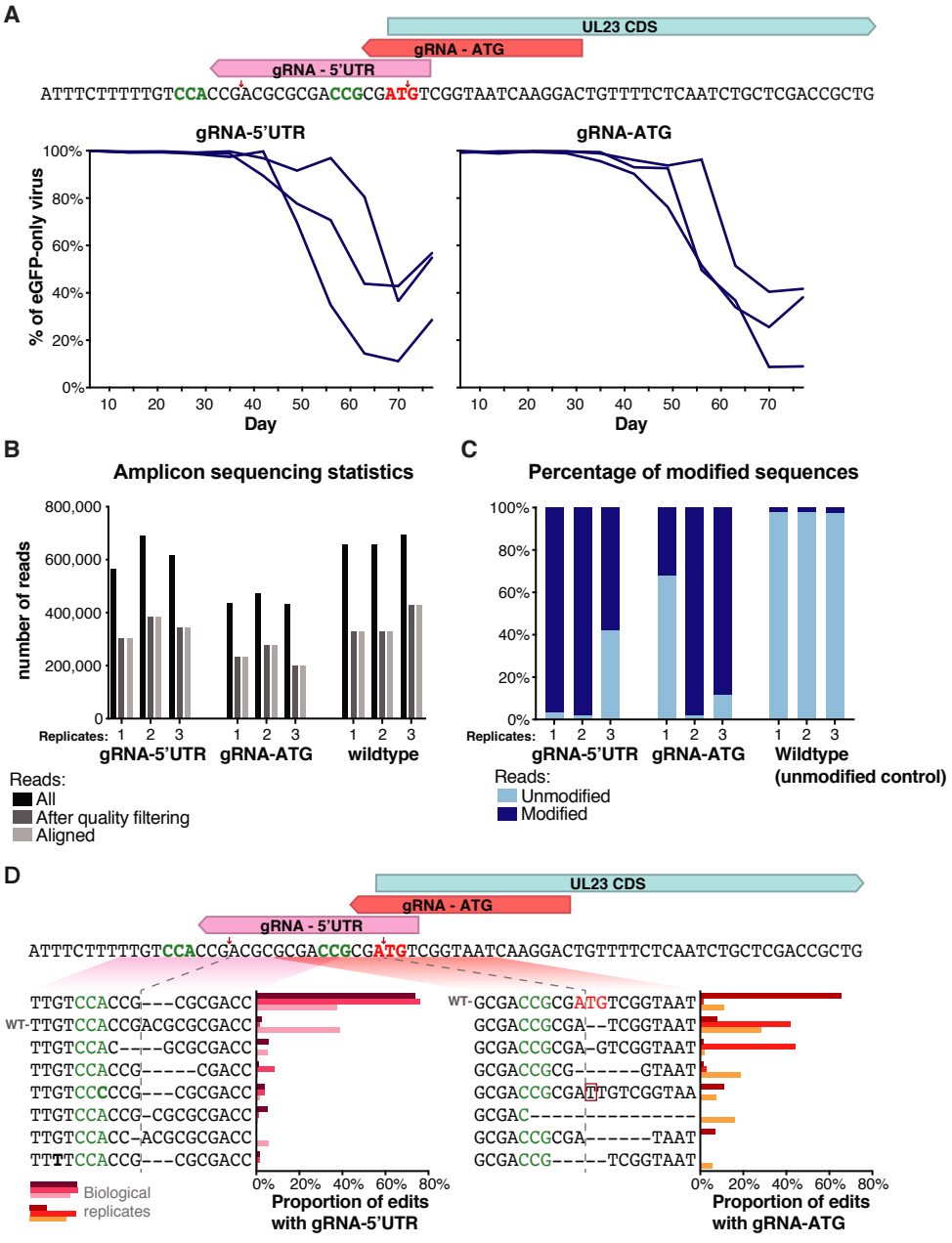
48. B. E. Pickett, *et al.*, ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res.* **40**, D593–8 (2012).

49. E. Paradis, K. Schliep, ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).

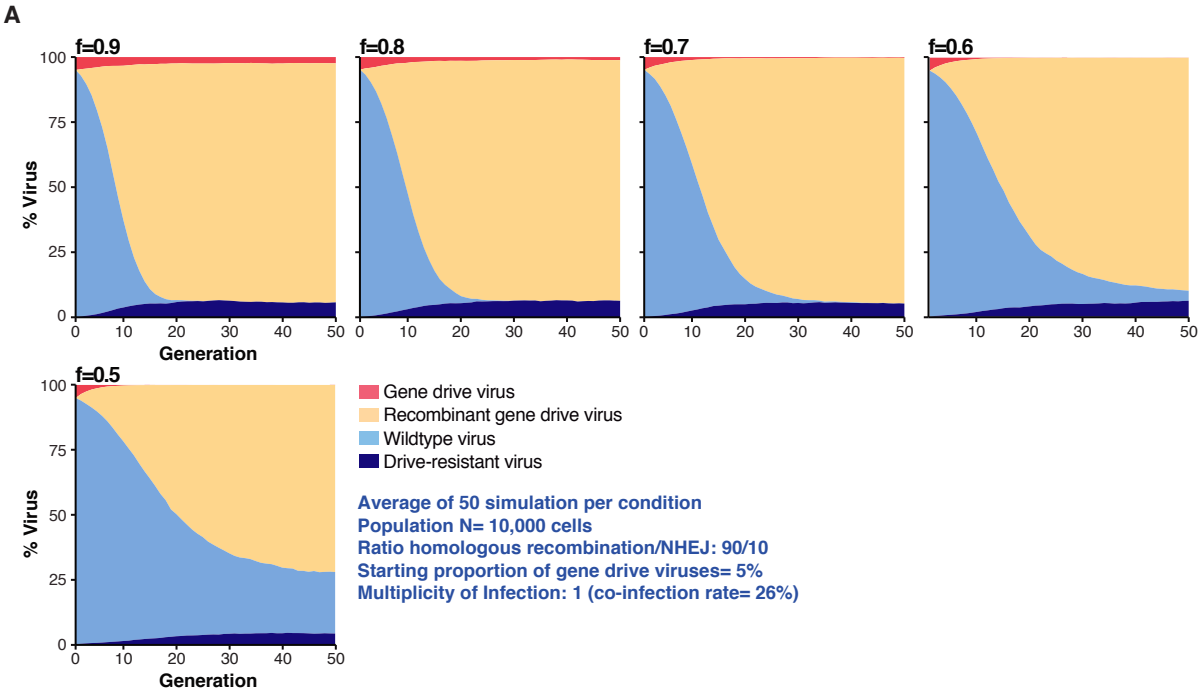
50. H. Wickham, ggplot2: Elegant Graphics for Data Analysis (2016).



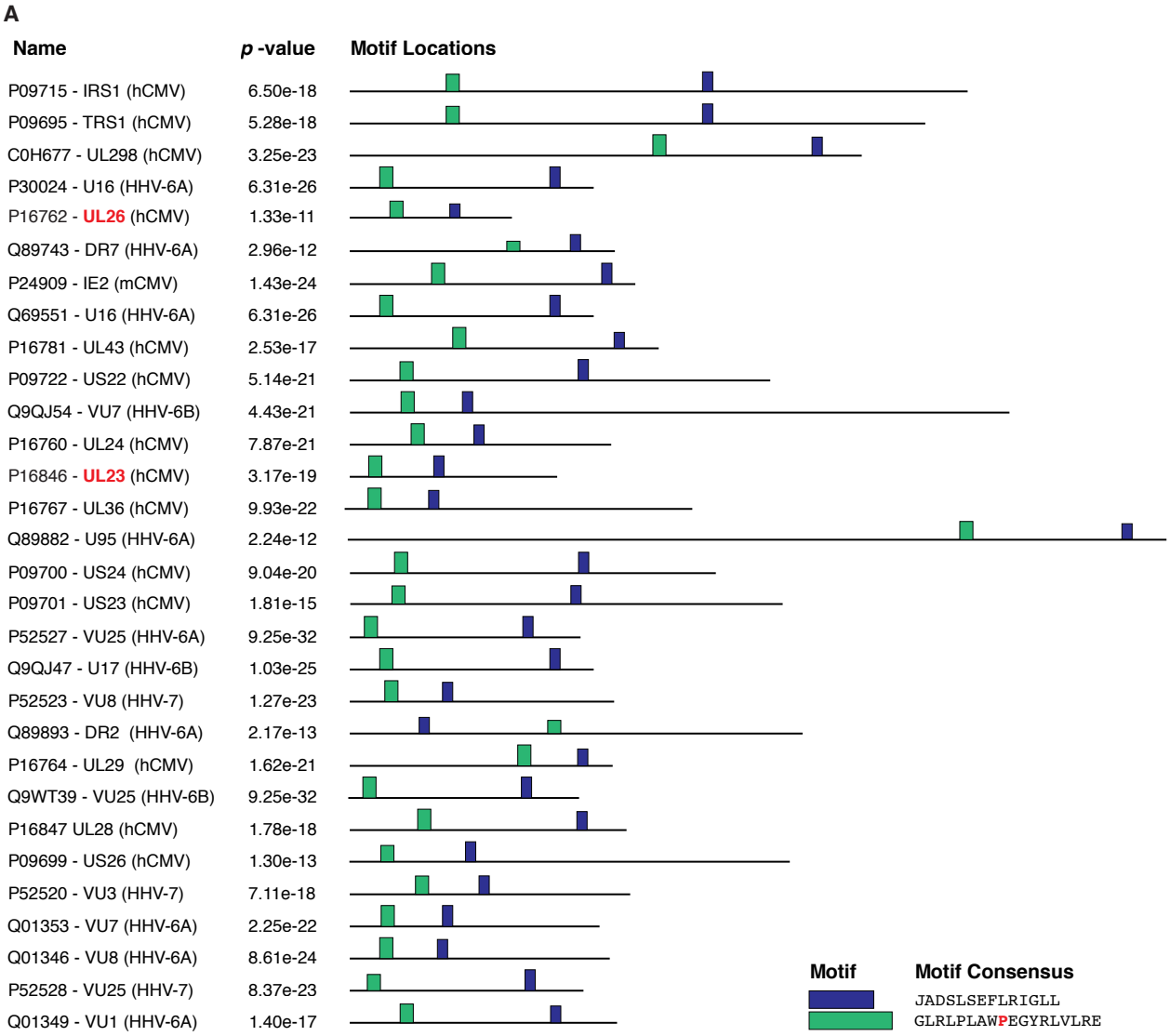
Supplementary Figure S1: Example of numerical simulation
Example of a simulation with a replicative fitness cost f=0.9 for gene-drive viruses. Left panel shows 50 independent simulations. Middle: mean and standard deviation of the 50 simulations. Right: mean proportion of the different viruses.



Supplementary Figure S2: Amplicon sequencing of drive-resistant target sites
A. Proportion over time of viruses expressing only GFP in a gene drive against *UL23* 5'UTR or ATG. Fibroblasts were initially transfected with a gene drive plasmid against *UL23* 5'UTR or *UL23* start codon, and subsequently infected with Towne-GFP. **B**. Amplicon-sequencing statistics of three biological replicates, at day 70. Towne-GFP-infected cells were also sequenced as an unmodified control. **C**. Proportion of edited genomes after amplicon sequencing. **D**. Relative contribution of each edits.
CRISPR cleavage sites are shown by red arrows, PAMs are highlighted in green and UL23 start codon in red. Results for the gRNA against UL23 5'UTR are the same as in Figure 2, but are duplicated here for easier comparison.



Supplementary Figure S3: Numerical simulations with defective drive-resistant viruses
Numerical simulation with gene-drive and drive-resistant viruses having the same fitness cost f. In these simulations, at each viral generation, N virtual cells were randomly infected and coinfected by N viruses, producing a new generation of viruses. When a cell was coinfected by wildtype and gene-drive viruses, wildtype viruses are converted to new gene drive viruses or resistant viruses in a 90/10 ratio.



Supplementary Figure S4: Localization of conserved motifs in the US22 protein family
Localization of conserved motifs in 31 proteins of the US22 family of herpesviral proteins. First column gives Uniprot IDs, gene names and viruses of origin.