An Adaptable Platform for Directed Evolution in Human Cells

Chet M. Berman,[1‡] Louis J. Papa III,[1‡] Samuel J. Hendel,[1‡] Christopher L. Moore,[1] Patreece H. Suen,[1] Alexander F. Weickhardt,[1] Ngoc-Duc Doan,[1] Caiden M. Kumar,[1] Taco G. Uil,[2†] Vincent L. Butty,[3] Robert C. Hoeben,[2] Matthew D. Shoulders[1,*]

[1]Department of Chemistry and [3]BioMicroCenter, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Massachusetts 02139, United States; [2]Department of Cell and Chemical Biology, Leiden University Medical Center, 2300 RC, Leiden, The Netherlands

ABSTRACT: The discovery and optimization of biomolecules that reliably function in metazoan cells is imperative for both the study of basic biology and the treatment of disease. We describe the development, characterization, and proof-of-concept application of a platform for directed evolution of diverse biomolecules of interest (BOIs) directly in human cells. The platform relies on a custom-designed adenovirus variant lacking multiple genes, including the essential DNA polymerase and protease genes, features that allow us to evolve BOIs encoded by genes as large as 7 kb while attaining the mutation rates and enforcing the selection pressure required for successful directed evolution. High mutagenesis rates are continuously attained by trans-complementation of a newly engineered, highly error-prone form of the adenoviral polymerase. Selection pressure that couples desired BOI functions to adenoviral propagation is achieved by linking the functionality of the encoded BOI to the production of adenoviral protease activity by the human cell. The dynamic range for directed evolution can be enhanced to several orders of magnitude via application of a small molecule-based adenoviral protease inhibitor to modulate selection pressure during directed evolution experiments. This platform makes it possible, in principle, to evolve any biomolecule activity that can be coupled to protease expression or activation by simply serially passaging adenoviral populations carrying the BOI. As proof-of-concept, we use the platform to evolve, directly in the human cell environment, several transcription factor variants that maintain high levels of function while gaining resistance to a small molecule inhibitor. We anticipate that this platform will substantially expand the repertoire of biomolecules that can be reliably and robustly engineered for both research and therapeutic applications in metazoan systems.

INTRODUCTION

Directed evolution methodologies have transformed our ability to generate biomolecules with improved or novel functionalities.[1-6] The vast majority of directed evolution experiments are performed in either acellular environments, bacteria, or yeast. While these strategies have yielded many successes, they also frequently lead to products that fail to function optimally when later introduced into complex metazoan systems. The evolved functions can be derailed by off-target interactions, poor protein folding or stability, pleiotropic outputs, or other serious problems that arise because the biomolecules were discovered and optimized in overly simplistic environments.[7-9] This frontier challenge could be most directly addressed by leveraging the human cell itself as the design, engineering, and quality control factory for directed evolution-mediated biomolecule discovery and optimization.

Extant strategies for directed evolution in human cells rely almost entirely on fluorescent screens to identify active biomolecule variants. The most common technique is *in vitro* plasmid mutagenesis followed by transfection and screening.[10] This approach is slow, labor-intensive, and significantly constrains library sizes. Other methods include *in vivo* mutagenesis through somatic hypermutation in immune cells followed by screening or selection.[11,12] More recently, robotic cell-picking techniques have been used to more comprehensively screen for desired phenotypes across multiple dimensions (e.g., both extent and localization of a fluorescent signal).[9] These methods, while valuable, are still slow, inefficient, and have limited library sizes (~10^5 variants for the most recent robotic platform). Another recent development has been the use of cytidine deaminase fused to Cas9 variants to introduce mutations into endogenous genes in human cells and selecting or screening for desired phenotypes.[13-15] However, these methods require the design and synthesis of many guide RNAs to tile along regions of interest, which requires repeated rounds of sequencing and guide RNA redesign as mutations accumulate. Moreover, directed evolution achieved via *in vivo* mutagenesis of the human genome is limited by the slow growth rate of human cells and the high potential for false positives ('cheaters') associated with any strategy that relies on cell selection or screening.

A broadly useful human cell-based directed evolution platform requires several critical features: (1) Large mutational libraries expressed in the human cell; (2) Selection schemes providing a broad dynamic range for selection and minimal opportunities for cheating; (3) Capacity to evolve multiple biomolecule functions; (4) Applicability across multiple cell types; and (5) Ideally, a minimal need for experimenter intervention during evolution experiments.

Inspiration for such a platform can be drawn from prior efforts coupling biomolecule function to viral replication using HIV[16] or bacteriophage.[17] However, HIV-based strategies suffer from an inability of the virus to propagate under strong selection pressure or in most cell types, and raise safety concerns surrounding large-scale HIV culture. The M13 bacteriophage used in phage-assisted continuous evolution provides large mutational libraries and enables rapid rounds of selection and mutagenesis for biomolecules carrying out diverse functions, but only permits directed evolution in bacterial cells.

With these parameters and challenges in mind, we aimed to devise a broadly useful human cell-based directed evolution platform. We rationalized that adenovirus type-5 would be a practical vector for directed evolution of biomolecules in hu-

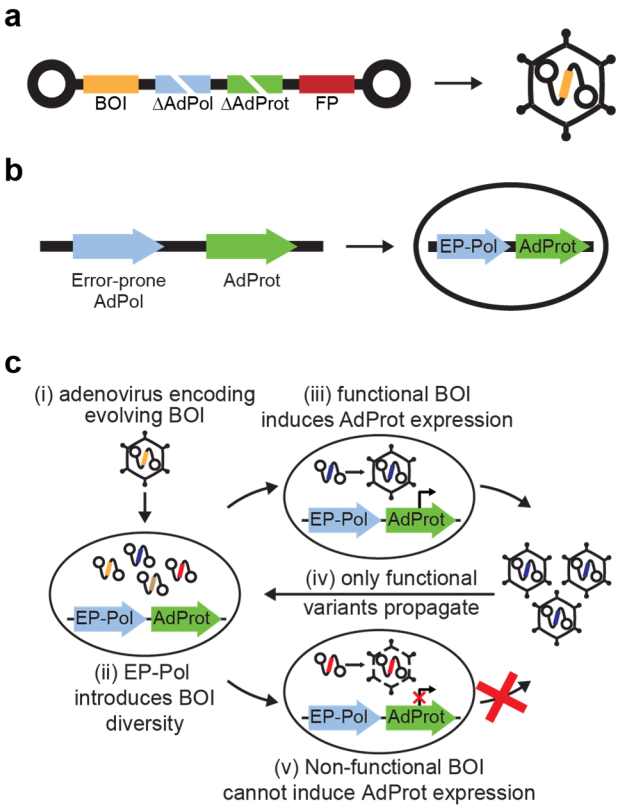


Figure 1. Human cell-based directed evolution platform overview. (a) Schematic of an engineered adenovirus type-5 vector in which genes for adenoviral polymerase (AdPol) and protease (AdProt) are removed and a gene encoding the biomolecule of interest (BOI) for directed evolution is introduced, as well as a fluorescent protein (FP) for visualization during infection. (b) Schematic of engineered human cells constitutively expressing a highly error-prone AdPol (termed EP-Pol) and conditionally expressing AdProt at levels directly dependent on BOI activity. (c) Schematic for adenoviral-based directed evolution of BOIs in human cells: (i) The BOI is delivered into the human cell via adenoviral infection. (ii) EP-Pol introduces mutations into the BOI gene, generating a mutational library. (iii) The desired BOI function is coupled to the expression or activity of AdProt such that (iv) only functional BOI variants result in viral propagation. (v) If the BOI variant is non-functional, AdProt is not expressed or active and the adenovirus encoding that variant is outcompeted.

man cells, owing to its genetic tractability and broadly infectious nature in many human cell types.[18,19] Conceptually, if the replication of a highly mutagenic adenovirus somehow depended on the activity of a biomolecule of interest (BOI) encoded in the adenoviral genome, then a simple directed evolution scheme for evolving diverse BOI functions in human cells could be feasible.

To achieve this concept, we first deleted the essential adenoviral DNA polymerase (AdPol) and protease (AdProt) genes from an adenoviral genome that also encoded the BOI for evolution **(Figure 1a)**. The resulting adenovirus deletion variant is incapable of replication outside engineered human cells. We trans-complemented the missing AdPol by constitutive expression, within human cells, of a newly engineered and highly mutagenic AdPol variant to enable the generation of large mutational libraries during viral replication. AdProt expression in the human cells was then engineered to depend conditionally upon BOI function (**Figure 1b**). Directed evolution experiments in this system rely on simply serially passaging the BOI-encoding adenovirus while mutagenesis and selection continuously occur (**Figure 1c**).

Here, we present the key features of this new platform, including mutagenesis, selection, and enrichment parameters. We further demonstrate the platform's utility via proof-of-concept directed evolution experiments in which we evolved, directly in the human cell environment, multiple transcription factor variants that maintained high levels of function while gaining resistance to a small molecule inhibitor. Altogether, we believe that this platform holds significant potential to not only enable the development of new research tools, but also to enhance our understanding of metazoan evolutionary biology and our ability to rapidly generate and optimize biomolecular therapeutics.

RESULTS

Mutagenesis. Adenovirus type-5 relies on its own DNA polymerase, AdPol, for replication of its double-stranded DNA genome.[20] The high fidelity AdPol has an estimated mutation rate of ~1.3×10^{-7} mutations per base per viral passage, based on high fidelity deep sequencing experiments performed by Sanjúan and co-workers.[21] Such a low mutation rate is insufficient to generate the large library sizes necessary for laboratory time-scale directed evolution. We therefore sought to increase the mutation rate of adenovirus by engineering a highly mutagenic variant of AdPol.

Previous studies identified two amino acid substitutions in AdPol, F421Y and D827A, that separately increase the mutation rate of AdPol, likely through distinct mechanisms (**Figure 2a**).[22] In the ϕ29 bacteriophage polymerase,[23] an AdPol homolog, the amino acid analogous to F421 occurs in the proofreading exonuclease domain, suggesting that the F421Y AdPol variant may have weakened proofreading capacity. The amino acid analogous to D827 occurs in the fingers domain involved in selection of incoming nucleotides, again suggesting a possible mechanism for the reduced fidelity of D827A AdPol. We reasoned that combining these two substitutions to create the F421Y/D827A AdPol double-mutant, which we termed error-prone AdPol (or EP-Pol), would allow us to further increase the mutation rate while still supporting robust adenovirus propagation.

To test this hypothesis, we first used recombineering to inactivate the AdPol gene encoded by the adenovirus type-5 genome via an internal deletion (see **Table S1** for a list of adenoviral constructs employed). Next, we stably transduced HEK293A cells with an HA-tagged version of either wild-type AdPol or EP-Pol (see **Table S2** for a list of cell lines employed). We observed that ΔAdPol adenoviruses (CFP.ΔAdPol.GFP where CFP and GFP correspond to cyan and green fluorescent protein, respectively) propagated only on cells that expressed either AdPol (**Figure S1**) or EP-Pol *in trans* (**Figure 2b**). Further, we observed that EP-Pol and wild-type AdPol both supported robust ΔAdPol-adenovirus replication.

We next assessed the mutation rate endowed by EP-Pol. After passaging ΔAdPol-adenovirus (AdGLΔPol) on EP-Pol trans-complementing human cells for 10 serial passages, we deep sequenced a 6.5 kb region of the genome obtained from a pool of about 30 viral clones (**Figure 2c**; see also **Table S3**). This sequencing revealed a mutation rate of 3.7×10^{-5} mutations per base per passage. As the adenoviral genome is ~35 kb, this mutation rate indicates that EP-Pol introduced ~1.3 mutations into the genome per infected cell per passage. Moreover, EP-Pol displayed a broad mutational spectrum, including both transitions and transversions (**Figure 2d**).

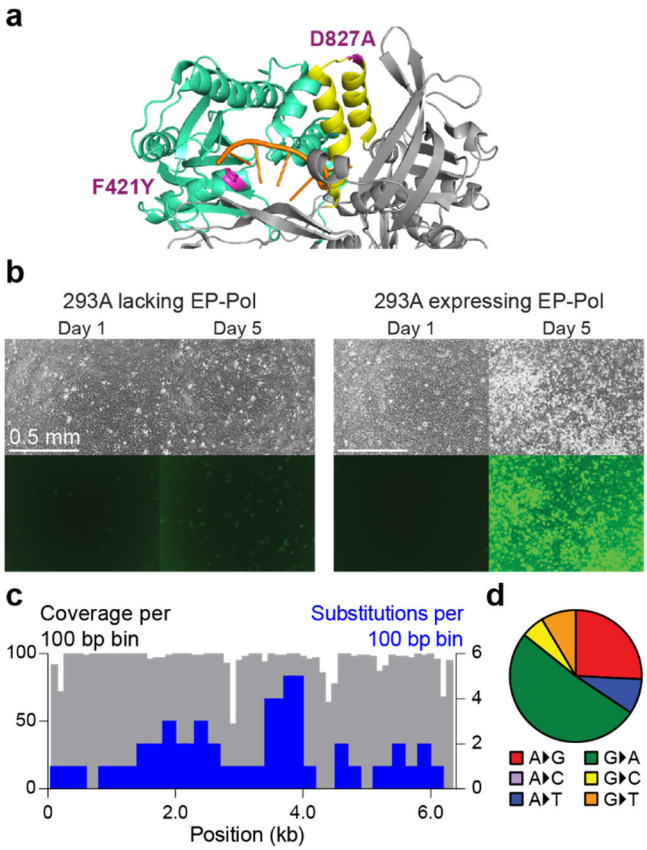


Figure 2. (a) Crystal structure of the ϕ29 DNA polymerase (PDBID 1XHZ),[23] an AdPol homolog, with the locations of homologous mutations used to create EP-Pol shown in magenta. (b) Either parental HEK293A cells or cells constitutively expressing EP-Pol were infected with a GFP-encoding ΔAdPol-adenovirus (CFP.ΔAdPol.GFP). The virus propagated only on EP-Pol trans-complementing cells. Similar results were obtained for wild-type AdPol (**Figure S1**). (c) ΔAdPol-adenovirus (AdGLΔPol) was serially passaged on EP-Pol expressing cells for ten passages, after which a 6.5 kb genomic fragment was amplified from an ~30 clone pool. Illumina sequencing identified mutations throughout the amplified region. For substitution values, see **Table S3**. (d) Mutational spectrum of EP-Pol evaluated by next-generation sequencing.

Previously, the same sequencing procedure was carried out for wild-type AdPol.[22] Because only one mutation introduced by wild-type AdPol was detected across two separate trials in that experiment, it was not possible to define an actual mutation rate for wild-type AdPol. In contrast, 60 mutations and 13 indels were observed for EP-Pol and, compared to the previously reported mutation rate of wild-type AdPol determined by another method,[21] the mutation rate of EP-Pol is enhanced ~280-fold. Thus, EP-Pol greatly increases the number of mutations introduced per viral passage. Based on these comparisons, the EP-Pol mutation rate is similar to highly mutagenic RNA viruses that can readily evolve on laboratory timescales.[24-26]

We next estimated the lower limit of the library size in a given passage (or 'round') of directed evolution using EP-Pol. A typical round of directed evolution might reasonably involve infecting 3.0×10^8 human cells at a low MOI. Each round of directed evolution ends once 100% of cells (~3.0×10^8 cells) are infected. Because ~1.3 mutations are introduced per cell per replication, and because there is at least one replication in each round of evolution since the infection occurs at low MOI, we estimate that there are ~4×10^8 adenoviral variants after one passage. Assuming a typical 1 kb gene encoding the BOI comprises ~1/30 of the engineered adenoviral genome, there would be ~1.3×10^7 variants of the BOI in the population after one round of evolution. This calculation is a lower limit because it does not account for any genetic diversity at the beginning of each round. Additionally, there is likely to be more than a single replication in each round of evolution, which would further increase library complexity. Regardless, even this conservative estimate indicates that we can generate virtually all single, many double, and some triple mutants in a single round of evolution. Notably, the mutations are continuously introduced instead of requiring *in vitro* mutagenesis physically separated from selection and propagation steps.

Selection. Our next objective was to design an appropriate selection scheme capable of coupling BOI activity to adenoviral propagation. After extensive testing of assorted adenoviral genes, we developed such a scheme based on deleting the gene for adenoviral protease (AdProt) from the viral genome and then providing AdProt *in trans* from the human host cell.[27] AdProt has vital functions in viral uncoating, DNA replication, and viral maturation.[28,29] Importantly, AdProt is a 'late gene' expressed mainly after DNA replication of the adenoviral genome.[29] Because AdProt is not required in the early stages of infection, BOI variants can be generated by mutagenesis before selection pressure is applied during a given infection.

We began by testing whether AdProt trans-complementation could be achieved in the context of an adenovirus already requiring AdPol trans-complementation. We stably expressed AdProt in an AdPol-expressing cell line, termed "producer" cells (see **Table S2**). Next, we monitored the progress of an adenovirus infection of ΔAdProtΔAdPol-adenovirus on AdPol-expressing versus AdPol- and AdProt-expressing cells. We observed that only the cell line constitutively expressing both AdProt and AdPol supported robust replication of ΔAdProtΔAdPol-adenovirus (**Figure S2**). Thus, host cell expression of AdPol and AdProt can successfully support the replication of an AdPol- and AdProt-deleted adenovirus, permitting both the facile production of ΔAdProtΔAdPol-adenoviruses and providing a potential mechanism to impart selection pressure in a directed evolution experiment.

We next evaluated the capacity of this AdProt-complementation strategy to confer sufficient selection pressure to drive a directed evolution workflow. For this purpose, we performed a competition experiment on a model BOI, the tetracycline (tet)-transactivator (tTA).[30,31] Wild-type tTA (tTA_{wt}) binds its endogenous operator, with a consensus sequence of 5′-CCTATCAGTGATAGA-3′, to induce downstream gene transcription. A tTA variant (tTA_{mut}) that is incapable of binding to the endogenous operators has also been reported.[32] tTA_{mut} instead possesses enhanced affinity for the mutant 5′-CCcgTCAGTGAcgGA-3′ operator. We engineered ΔAdProtΔAdPol-adenoviruses that expressed either tTA_{wt} and mCherry (tTA_{wt}.mCherry) or tTA_{mut} and GFP (tTA_{mut}.GFP). We then stably transduced AdPol-expressing HEK293A cells with a lentiviral vector that provided AdProt under control of the endogenous tTA operator (termed "selector" cells, see **Table S2**). In this cell line, tTA_{wt}.mCherry adenovirus should be able to strongly induce AdProt and propagate, whereas tTA_{mut}.GFP should not induce AdProt and therefore should not form infectious virions. Because these viruses express different fluorescent markers, relative viral populations can be assessed using flow cytometry upon infection of human cells that do not

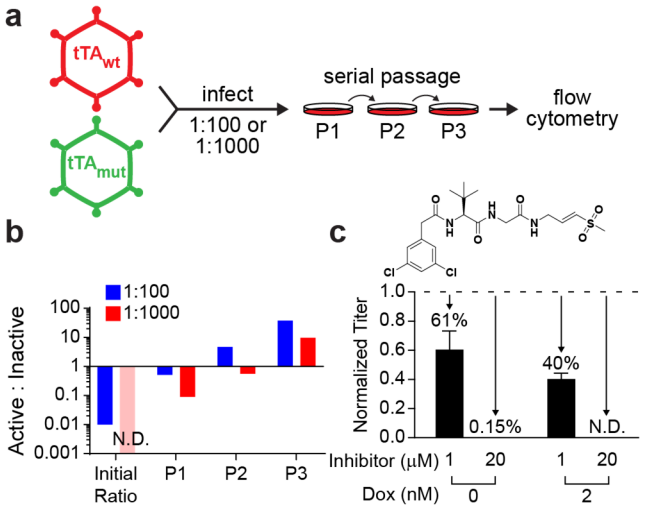


Figure 3. (a) Schematic of the competition experiment between adenoviruses that carry the gene for wild-type tetracycline transactivator (tTA$_{wt}$.mCherry) versus viruses that carry inactive tTA (tTA$_{mut}$.GFP). HEK293A cells stably encoding the gene for adenoviral protease (AdProt) under control of the endogenous tTA operator are infected by an initial ratio of 1:100 or 1:1,000 tTA$_{wt}$.mCherry to tTA$_{mut}$.GFP viruses. Viral media was serially passaged onto a new plate of cells for three rounds. The viral populations were then determined via flow cytometry. (b) Quantification of flow cytometry data from the competition experiment. The proportion of tTA$_{wt}$.mCherry adenoviruses relative to tTA$_{mut}$.GFP adenoviruses rapidly increased with each passage. The initial ratio of the 1:1,000 sample (labeled N.D.; not detectable) was not experimentally quantifiable owing to the low amount of tTA$_{wt}$.mCherry adenovirus present, and was therefore derived by dilution of the 1:100 initial ratio. For raw flow cytometry data, see **Figures S3** and **S4**. (c) AdProt-based selection pressure in combination with administration of a small molecule AdProt inhibitor (structure shown) provides access to an orders of magnitude-wide dynamic range of selection pressure. tTA-inducible AdProt cells were infected with tTA$_{wt}$.mCherry adenovirus, and treated with a combination of doxycycline (dox) and the AdProt inhibitor. The resulting viral supernatant was titered by flow cytometry. Titers were normalized to infections performed in the absence of the AdProt inhibitor. The titer of the adenovirus treated with 20 μM AdProt inhibitor and 2 nM dox was too low to be accurately detected (N.D.; not detectable).

express AdProt in order to prevent propagation and therefore more accurately quantify the resulting viral populations.

To test our hypothesis that AdProt induction could enable enrichment of active over inactive BOI variants, we co-infected tTA$_{wt}$.mCherry and tTA$_{mut}$.GFP at an MOI of ~0.25 in selector cells (see **Table S2**) at initial ratios of 1:100 or 1:1,000 (**Figure 3a**). We then performed three serial passages on selector cells, and analyzed the resulting viral populations via infection of AdPol-expressing but AdProt-lacking HEK293A cells followed by flow cytometry. In the initial passage, the tTA$_{wt}$.mCherry adenovirus enriched at least 40–50-fold over the tTA$_{mut}$.GFP adenovirus (**Figure 3b**). Furthermore, across three rounds of passaging, the tTA$_{wt}$.mCherry adenoviruses were consistently enriched to > 90% of the adenoviral population regardless of the starting ratios. Thus, our AdProt-based selection strategy can rapidly enrich active BOIs that are initially present at low frequency in a viral population.

We next applied this tTA-based genetic circuit to evaluate the dynamic range of AdProt selection. Our approach was to employ an allosteric inhibitor of tTA, doxycycline (dox), to tune AdProt expression levels. In the presence of dox, tTA is unable to bind its target operator and induce AdProt expression. Using this approach, based on AdProt transcript levels we were able to access up to a 14-fold change in AdProt expression (**Figure S5a**). Notably, we observed a strong correlation between dox concentration and viral titer over this entire range (**Figure S5b**).

We note that because AdProt can perform multiple catalytic turnovers, there is likely an upper bound to the number of active AdProt molecules required for replication, at which point additional AdProt induction will not result in greater viral replication. As a result, selection pressure would be reduced very low for any evolved BOIs that are able to induce AdProt above the upper bound. A small molecule inhibitor of AdProt could provide a way to dynamically tune selection pressure to reduce AdProt activity below the upper limit as a given directed evolution experiment proceeds. Indeed, when we challenged tTA$_{wt}$.mCherry-expressing adenoviruses with various concentrations of the vinyl sulfone AdProt inhibitor shown in **Figure 3c**,[33] we found that the inhibitor could reduce the infectious titer of the tTA$_{wt}$.mCherry virus up to 650-fold, providing ready access to a dynamic range of selection pressure between 2–3 orders of magnitude in size. Moreover, we observed that the AdProt inhibitor even further reduced infectious titer in the presence of dox (**Figure 3c**), highlighting the capacity of AdProt inhibition to expand the dynamic range for selection at a variety of baseline AdProt expression levels. Notably, the vinyl sulfone AdProt inhibitor was not toxic at the concentrations used (**Figure S6**).

Directed evolution of functional, drug-resistant tTA variants in human cells. We next sought to test the feasibility of actually evolving BOI function in human cells using this platform. For proof-of-concept, we specifically aimed to evolve tTA variants that retained transcription-inducing activity but gained resistance to their small molecule inhibitor, dox. Specifically, we serially passaged our tTA$_{wt}$.mCherry virus in the presence of dox in a "selector" cell line (see **Table S2**) that inducibly expressed AdProt under control of the endogenous tTA operator. We maintained a low multiplicity of infection (~0.05) to minimize the probability that viruses encoding distinct tTA variants would co-infect the same cell, at least at an early stage of each passage. Such double-infections could result in "hitchhiking," in which low fitness variants can be temporarily maintained in the population by infecting the same cell as high fitness variants. Such hitchhikers could slow the pace of selection. We transferred viral supernatant to fresh cell plates upon the appearance of spreading infection, with the goal of selecting for viruses that encode functional, but dox-resistant, tTA variants.

We ran two evolution experiments in parallel (Trials 1 and 2) with different selection pressure strategies (**Figure 4a**). In Trial 1, we tuned the selection pressure over time, increasing the dox concentration from 2 nM up to 20 μM. In Trial 2, we kept selection pressure constant and high by maintaining the dox concentration at 200 nM. In order to test whether dox-resistant tTA variant enriched in the population, we used the viral media from each passage in Trial 1 to infect a "phenotyping" cell line (see **Table S2**) containing GFP under control of the endogenous tTA operator in the presence of dox. This phenotyping cell line lacked AdProt, allowing the virus to infect the cells and induce GFP expression, but not to proliferate. We measured GFP induction by the viral population harvested after each serial passage in the presence of 20 μM dox in these phenotyping cells using flow cytometry (**Figure 4b**). Substantial dox-resistant

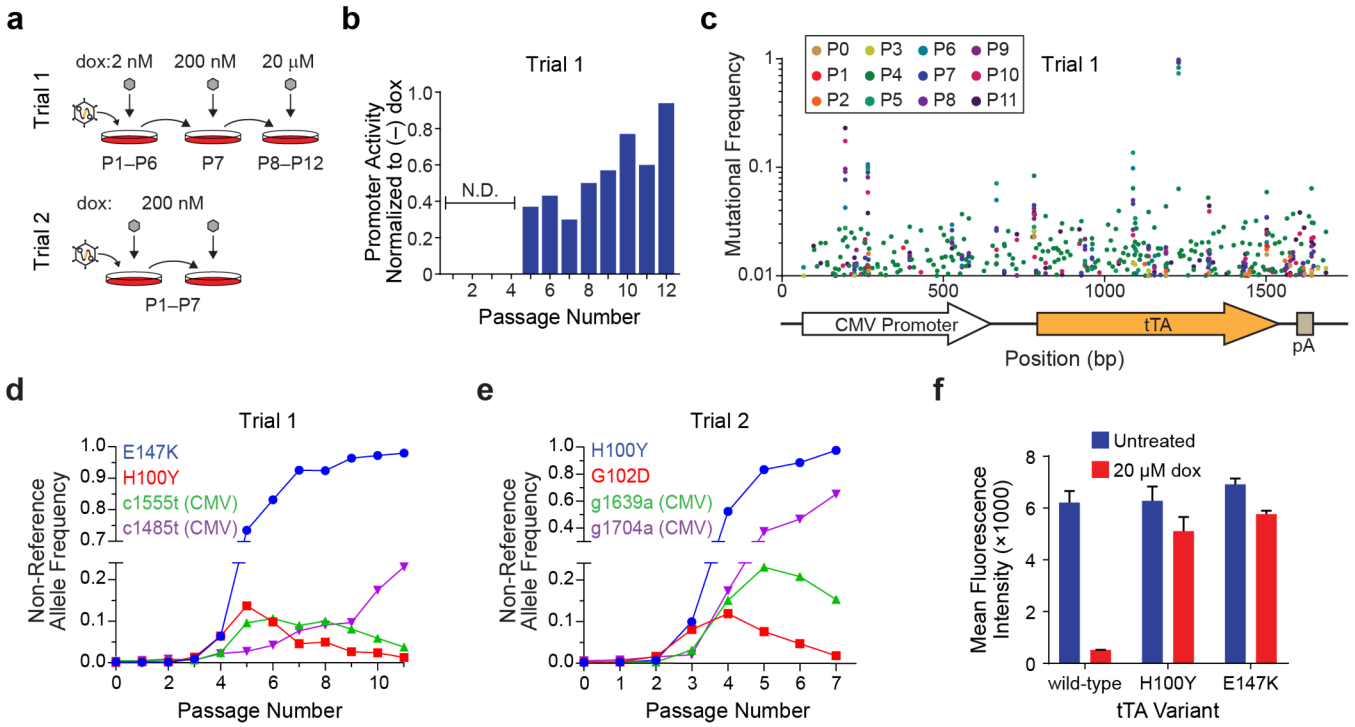


Figure 4. (a) Serial-passaging schemes for evolving functional tTA variants that gain dox resistance in human cells. Two approaches to selection pressure were used, either with increasing dox concentrations (Trial 1) or a constant, moderate dox concentration (Trial 2). (b) tTA-induced GFP expression in the presence of dox after each round of evolution for Trial 1. Phenotyping cells were infected with passaged viral populations and analyzed by flow cytometry. The percentage of infected GFP-positive cells at each passage in the presence of dox was normalized to the percentage of infected GFP-positive cells at each passage in the absence of dox. N.D. = not detectable owing to low viral titer. (c) Non-reference allele frequencies for all mutations observed at ≥ 1% frequency over the course of the directed evolution experiment for Trial 1 (see **Figure S7** for Trial 2). A schematic of the sequenced amplicon is shown below the x-axis for reference. (d) Mutational trajectories of four mutations identified in Trial 1, including two non-coding mutations in the CMV promoter upstream of the tTA gene. (e) Mutational trajectories of four abundant mutations identified in Trial 2, including two non-coding mutations in the CMV promoter upstream of the tTA gene. (f) Plasmids encoding the tTA variants that fixed in Trials 1 and 2 were transfected, along with the pLVX-TRE3G.eGFP reporter plasmid, into HEK293A cells with or without dox (N = 3). Two days later, flow cytometry was performed to examine tTA variant activity in the presence versus the absence of 20 μM dox.

tTA activity emerged by passage 5, suggesting that dox-resistant variant(s) of tTA may have arisen and enriched in the viral population.

We next examined whether mutations in the tTA gene contributed to this decreased dox sensitivity. We amplified and sequenced a 1.75-kb region of the adenoviral genome containing the tTA open reading frame from virus harvested at each passage during both Trials. Using this approach, we detected > 200 unique mutations that attained ≥ 1% frequency by passage 4 in Trial 1, even though promoter activity at passage 4 was still undetectable (**Figure 4c**). In Trial 2, 43 mutations attained ≥ 1% by passage 4 **(Figure S7)**. By passage 5, a single amino acid substitution in tTA attained > 70% frequency in the viral population in both trials (E147K in Trial 1 and H100Y in Trial 2), rapidly becoming fully fixed in the population thereafter (**Figures 4d** and **4e**). Both mutations observed were previously reported to confer dox-resistance in tTA,[34] which we further confirmed through transient co-transfection of a plasmid encoding GFP under control of the endogenous tTA operator along with wild-type, E147K, or H100Y tTA-encoding plasmids into HEK293A cells in the presence or absence of dox (**Figure 4f**). Additional mutations that were also previously reported to confer dox-resistance were also observed at > 10% frequency early in the directed evolution experiment (H100Y in Trial 1 and G102D in Trial 2).

In Trial 2, we also analyzed the possible effects of hitchhikers on the enrichment of active variants. Our approach was to harvest the adenovirus at two different timepoints: (*i*) either early, when ~75% of cells were infected and co-infection was minimized or (*ii*) very late, after full cytopathic effect was achieved and most cells were co-infected. We found that even under high co-infection conditions (late harvest) dox-resistant variants continued to enrich, possibly even more than under low co-infection conditions (early harvest; **Figure S8**). Thus, co-infection did not hinder the enrichment of active variants.

These results highlight both the different outcomes that can result from repeated evolution experiments and the capacity of our platform to explore sequence space in human cells. Additionally, we were able to evolve biomolecules using two different selection pressure protocols (gradually increasing pressure or constant, high pressure). In summary, our directed evolution protocol can successfully generate and rapidly enrich functional BOI variants in human cells, merely by serial passaging of a BOI-encoding adenovirus.

Design of alternative selection circuits. In the interest of extending the utility of our platform beyond the directed evolution of transcription factors, we sought to demonstrate how alternative selection circuits could be used to evolve different types of functions. We created two new selection circuits for a user-defined recombinase activity and aminoacyl-tRNA synthetase activity (**Figure S9a** and **b**).[35,36] We transfected both the

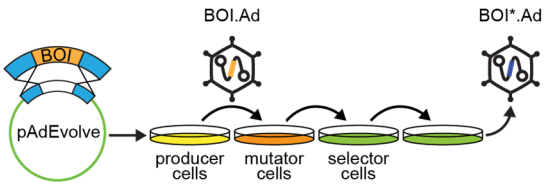


Figure 5. The gene encoding a biomolecule of interest (BOI) is first inserted into pAdEvolve. "Producer" cells (see cell lines listed in **Table S2**) are used to generate ΔAdProtΔAdPol-adenoviruses carrying the BOI gene. If desired, the BOI gene can be mutated prior to selection by first passaging the adenovirus on a "mutator" cell line constitutively expressing EP-Pol. A "selector" cell line tailored to the activity of interest is generated by the researcher, followed by serial passaging of viral supernatants on the selector cells.

Cre-recombinase (Cre, **Figure S9a**) and leucyl-tRNA synthetase (LeuRS, **Figure S9b**) AdProt selection circuits into HEK293A cells expressing AdPol and then monitored the replication of AdProt-deleted adenoviruses expressing Cre, LeuRS, or a control, inactive BOI (tTA). For the recombinase circuit, we found that the Cre-containing adenovirus replicated > 20-fold better than a control adenovirus (**Figure S9c**). For the aminoacyl-tRNA synthetase circuit, we observed the LeuRS-containing adenovirus was able to replicate while the control adenovirus could not replicate to detectable levels. All adenoviruses replicated robustly on a control circuit that constitutively expressed protease. These data indicate that our platform can be easily adapted to select for desired recombinase and amino-acyl tRNA synthetase activities.

DISCUSSION

We report here the development, characterization, and proof-of-principle application of a highly adaptable platform for directed evolution of diverse BOI functions in human cells. In this platform, human cells are infected by a BOI-encoding adenovirus lacking the essential AdProt and AdPol genes (**Figure 1c**). A newly engineered, highly error-prone variant of AdPol, EP-Pol, constitutively expressed by the human cells, replicates the adenoviral genome. The resulting error-prone DNA replication introduces mutations into the BOI gene at a high rate, thereby continuously generating mutant libraries for selection. BOI variants are then expressed during viral infection of the human cell, and continuously tested for activity via a selection couple in which functional BOI variants induce higher levels of AdProt activity stemming from an AdProt gene cassette installed in the human cells. Because AdProt activity is linked to the virus' capacity to propagate, functional BOI variants are continuously enriched in the evolving viral population, whereas non-functional BOI variants result in non-viable virions that cannot propagate.

Application of the platform is straightforward, such that genes encoding a BOI can be integrated into the adenoviral genome using Gateway cloning,[37] followed by plasmid transfection into a producer cell line that constitutively expresses both AdPol and AdProt to generate a starter adenovirus population (**Figure 5**). Directed evolution then simply involves serial passaging of the adenovirus on user-defined 'selector cells'.

In developing this platform, we chose to use adenovirus rather than a natively mutagenic RNA virus owing to adenovirus' relative safety, broad tropism, ease of manipulation, and capacity to propagate even under strong selection pressure. The adenoviruses used for directed evolution experiments were E1-, E3-, AdPol- and AdProt-deleted. All of these genes are required for adenoviral replication in the wild. Thus, the safety of working with these adenovirus deletion variants is maximized as they can only replicate in human cells that provide these essential genes *in trans*, and cannot replicate in unmodified human cells.[22,27,38] Moreover, the removal of this large portion of the adenoviral genome means that genes as large as ~7 kb can potentially be introduced and evolved in our platform. The broad tropism of adenovirus[18] is beneficial because it means that directed evolution experiments can, in principle, be performed in many different human cell types depending on the objective of a particular experiment. Finally, from a genome engineering perspective, our optimized recombineering protocols (see **Supporting Information**) allow the necessary facile manipulation of the adenoviral genome.[39]

Despite the manifold benefits of the choice to use adenovirus, we faced a significant challenge because both wild-type and even the previously reported error-prone AdPol variants[22] are relatively high fidelity, and therefore unlikely to enable the creation of mutational libraries at a sufficiently high rate to support continuous directed evolution of novel BOIs. To address this issue, we engineered EP-Pol, a highly mutagenic AdPol variant that pushes the adenoviral mutation rate into the regime of RNA viruses such as HIV and influenza that are well-known to rapidly evolve on laboratory timescales.[26,40,41] We used trans-complementation of EP-Pol via constitutive expression in the host cell to prevent reversion to wild-type AdPol that could occur if we modified an adenovirally encoded AdPol gene, thereby ensuring that mutagenic activity remains at a constant, high level throughout directed evolution experiments. We note that the optimized EP-Pol mutagenesis system may have applications beyond our directed evolution system. For instance, EP-Pol could be used to more rapidly assess resistance pathways to treatment of adenovirus infections or to improve the properties of adenovirus for therapeutic purposes.[22,42]

We note that this mutagenesis approach does introduce mutations into the adenoviral genome outside the gene for the BOI that can potentially be negatively selected and consequently reduce library size. The 6.5 kb genomic region we sequenced (**Figure 2**) was chosen because it contained both protein coding regions necessary for adenoviral replication and non-coding regions that should not face severe selection pressure. Comparing these domains across the sequenced region, we observed only a two-fold difference between the mutation rate in the inactivated AdPol gene, which should not be under any selection pressure in our trans-complementing system, and the neighboring pIX, IVa2, and pTP genes, suggesting that such selection only impacts our mutation rate at most two-fold.

Because AdPol selectively replicates only adenoviral DNA, EP-Pol can only introduce mutations into the adenoviral genome. This mutagenesis technique thus represents an improvement over other strategies that evolve genes directly in the human genome. In such strategies, off-target mutations can arise through basal or through the enhanced mutagenesis rates, which can subvert selection pressure and generate false positives. Furthermore, even recent mutagenesis methods that target specific genes within the human genome, by using somatic hypermutation[11,12] or Cas9-fusion proteins,[13-15] still display significant off-target genetic modification.[43-45] Especially given the large size of the human genome, many pathways to cheating selection may be available. Our use of an orthogonal replication system means that the human host cells are discarded and replaced with each passage, preventing mutation accumulation in the human cell that could potentially cheat selection pressure. As a result,

false positives are restricted to the ~30 kb viral genome, providing much more limited escape options than might be found in the entire human genome. This advantage, combined with the much more rapid expansion of adenovirus relative to human cells allowing a larger number of directed evolution rounds in a given time period, highlights the ability of our platform to quickly scan mutational space with minimal risk of selection subversion.

We found that AdProt can serve as a robust selectable marker for adenovirus-mediated directed evolution in human cells. As an enzyme with catalytic activity, we might not expect AdProt to exhibit a dynamic range of selection. However, we observed that AdProt was able to modulate viral titers ~10-fold in response to protease levels. Importantly, we discovered that a small molecule inhibitor of protease could be easily used to further enhance this dynamic range to several orders of magnitude. It is noteworthy that the AdProt inhibitor may also be employed to actively fine-tune selection stringency over the course of a directed evolution experiment, simply by modulating the compound's concentration in cell culture media.

We used this AdProt-based selection to evolve transcriptionally active variants of tTA that gained dox-resistance. Across two replicates of the experiment, two different tTA variants ultimately fixed in the population, both of which were indeed dox-resistant. We also observed a large number of lower frequency mutations at various passages above our 1% threshold for detection. The observation of these variants suggests that our platform is effectively screening sequence space for a selective advantage, particularly as the vast majority of mutations are unlikely to ever attain a frequency of 1% in the evolving viral population.

While this proof-of-concept experiment specifically highlights how AdProt-based selection could be used to evolve transcription factors, the platform should be readily generalizable to evolve a variety of other biological functions. Here, we demonstrated how our system can enable directed evolution of DNA recombinases and amino-acyl tRNA synthetases. Beyond just these selection circuits, examples of the necessary selection couples already exist for an assortment of other protein classes, including TALENs,[46] proteases,[47] protein-protein interactions,[48] RNA polymerases,[17,49] Cas9,[50] and beyond.

Looking forward, we envision a number of improvements that would further enhance this platform's practicability and applicability. The current system relies on serial passaging of adenovirus on adherent cells. Transitioning to suspension cells would enable variant libraries several orders of magnitude larger than we can currently explore. The integration of emerging targeted mutagenesis techniques, such as MutaT7[51] or CRISPR-X,[14] could further focus mutations only to the BOI gene and also increase mutation library size. Additionally, the present system is only capable of positive selection. Implementation of a negative selection strategy would enable our platform to evolve biomolecules that are more selective and specific for a given activity. We note that phage-assisted continuous evolution in bacteria can afford larger library sizes with more tunable mutation rates, in addition to dynamic selections that occur on the order of hours, not days.[17] Critically, while adenovirus-mediated directed evolution explores mutational space more slowly than phage-assisted continuous evolution, it makes possible similar experiments in the metazoan cell environment for the first time.

CONCLUDING REMARKS

Our platform offers several advantages relative to extant strategies for human cell-based directed evolution that rely on time-intensive screens and extensive *in vitro* manipulations. The use of adenovirus allows researchers to continuously mutate, select, and amplify genes of interest by simply transferring viral supernatant from one cell plate to the next. Owing to this simple viral passaging protocol, library sizes are restricted only by a researcher's tissue culture capacity. Cheating is minimized because mutations are specifically directed to the viral genome. Safety is maximized because the adenoviruses used lack multiple genes required for replication in the wild. Moreover, the user-defined nature of the selector cell and the broad tropism of adenovirus type 5 enable directed evolution to be performed in a diverse array of human cell types.

By making it possible for researchers to evolve diverse BOI functions in the same environment in which the BOIs are intended to function, we believe this human cell-based directed evolution platform holds significant potential to enable researchers to rapidly evolve a wide variety of biomolecules in human cells. Thus, this method should impact not just the development of new tools for research, but also our understanding of metazoan evolutionary biology and our ability to rapidly generate effective biomolecular therapeutics.

MATERIALS AND METHODS

Cloning methods: All PCR reactions for cloning and assembling recombineering targeting cassettes were performed using Q5 High Fidelity DNA Polymerase (New England BioLabs). Restriction cloning was performed using restriction endonucleases and Quick Ligase from New England BioLabs (see **Supporting Information**). Adenoviral constructs were engineered using *ccdB* recombineering, as previously described[39] and further optimized by us (see **Supporting Information**). Primers were obtained from Life Technologies and Sigma-Aldrich (**Table S4**). The TPL Gene block was obtained from Integrated DNA Technologies (**Table S4**). Sequences for all plasmids developed here can be obtained from GenBank using the accession numbers provided in **Table S5**.

Cell culture: Cells were cultured at 37 °C and 5% CO_2(g). New cell lines were derived from a parent HEK293A cell line (ATCC) and cultured in Dulbecco's modified Eagle's medium (DMEM; Cellgro) supplemented with 10% fetal bovine serum (FBS; Cellgro), 1% penicillin-streptomycin (Cellgro), and 1% L-glutamine (Cellgro). For assays involving the tetracycline (Tet)-dependent transcriptional activation system (directed evolution of dox insensitivity, promoter activity assays, and reverse genetics), Tet-approved FBS (Takara Bio) was used. The producer and mutator cell lines (**Table S2**) were cultured in 50 μg/mL hygromycin (Thermo Fisher) to stably maintain transgenes, while the selector and phenotyping cell lines (**Table S2**) were cultured in 1 μg/mL puromycin (Corning) for the same purpose.

Generation of cell lines by lentiviral transduction: In a typical protocol, ~9×10^6 HEK293FT cells (Thermo Fisher) were plated on a poly-D-lysine-coated 10 cm plate. The next day, the cells were co-transfected with plasmids from a third-generation lentiviral packaging system:[52] 15 μg RRE, 6 μg REV, 3 μg VSVG, and 15 μg transfer vector using 60 μL Lipofectamine 2000 (Thermo Fisher). Cultures were maintained in 5 mL total volume of OPTI-MEM (Gibco) during the transfection. After 8 h, the media was exchanged for fresh DMEM. After 48 h, media was harvested and centrifuged for 5 min at 3,200 × *g* to clear the cell debris. The supernatant was used to transduce

HEK293A cells supplemented with 4 μg/mL polybrene (Sigma-Aldrich). After 24 h, the media was exchanged for fresh DMEM. 48 h later, media was exchanged again for DMEM containing appropriate antibiotics to select stable cell lines.

Adenovirus production: Adenoviruses were produced by transfecting a PacI (New England BioLabs)-linearized vector into appropriate trans-complementing HEK293A cells (ΔAdPol adenoviruses on wild-type AdPol cells, ΔAdProtΔAdPol adenoviruses on producer cells; see **Table S2**). 24 μg of PacI-linearized adenovirus vectors mixed with 144 μL polyethyleneimine (Sigma-Aldrich) in 1 mL OptiMEM (Gibco) was added to a 15 cm plate of producer cells (**Table S2**; ~3×10^7 cells). Media was replaced 8 h post-transfection, and then intermittently replaced every 2–3 days until plaques were observed (typically ~3 weeks). Once plaques were detected, cytopathic effect was observed in all cells within 5 days. Upon complete cytopathic effect, the cells and media were harvested and subjected to three freeze/thaw cycles. The cell debris was removed by centrifugation at 3,200 × *g* for 15 min and the supernatant stored at –80 °C.

Mutagenesis rate determination: The mutagenic potential of AdPol variants was evaluated following a previously reported protocol.[22] Briefly, a polymerase-deleted Ad5, AdGLΔPol, was subjected to 10 serial passages on cultures of 911 cells[53] expressing EP-Pol in order to accumulate mutations. After 10 serial passages, 911 cells expressing wild-type AdPol were infected in duplicate 6-well plates at ~50 plaque-forming units/well in order to amplify pools of 50 viral clones for sequencing. Based on a plaque assay of one of the duplicates (which was overlaid with agarose), the actual number of plaque-forming viral clones in the pool obtained from the other duplicate (which was not overlaid with agarose) was estimated to be ~27. Using pools of 50 or fewer clonal viruses ensured that mutations present in only one clone will be present at a frequency above the threshold of detection. From the 27-clone viral pool, a 6.5-kb fragment was amplified and prepared for deep sequencing. Libraries were subjected to 32 cycles of single-read sequencing by an Illumina Genome Analyzer II. Using the short read analysis pipeline SHORE,[54] these reads were mapped against the reference sequence allowing up to two mismatches or gaps, after which low quality base calls within the obtained mappings were individually masked. Mutations were subsequently scored using a minimal variant frequency requirement of 0.25% and a minimal local sequencing depth requirement of 1200 for both the forward and the reverse read mappings. Previous experiments showed that these settings were able to account for sequencing errors and accurately score mutations.[22]

AdPol and AdProt trans-complementation assays: The day before beginning the assay, a 6-well plate was seeded with ~1 × 10^6 of the indicated cells. The next day, individual wells were infected with the indicated adenoviruses at a low MOI (< 0.5) in order to permit observation of the presence or absence of a spreading infection. AdPol and EP-Pol trans-complementation (see **Figure S1** for AdPol and **Figure 2b** for EP-Pol) was tested by monitoring CFP.ΔAdPol.GFP adenovirus infection on either AdPol- or EP-Pol-expressing HEK293A cells. Pictures were taken with an Olympus U-TB190 microscope. AdProt and AdPol double trans-complementation (see **Figure S2**) was tested by monitoring ΔAdProtΔAdPol-adenovirus (**Table S1**) infection on producer cells. Pictures were taken with a Nikon Eclipse TE200 microscope.

Determining adenoviral titer by flow cytometry: Adenoviral titers were determined through flow cytometry. Known volumes of AdPol- and AdProt-deleted viral supernatants were added to AdPol-expressing HEK293A cells. 2–3 days post-infection, cells were washed once with media, stained with 0.2 μg/mL DAPI, and then analyzed on a BD LSR II Analyzer for fluorescent protein expression. Infectious titers were determined by measuring the percentage of cells infected by a known volume of virus. To minimize counting cells that were infected by more than one virus and to minimize any background fluorescence, data were only considered if they fell within the linear range, which typically encompassed samples where 1–10% of cells were infected.

Competition experiments: A confluent dish of selector cells (**Table S2**; ~15 million cells) was infected with either a 1:100 or 1:1,000 mixture of tTA_{wt}:tTA_{mut} adenovirus (MOI ~ 0.25; **Table S1**). Plates were monitored for the appearance of spreading infection, defined by fluorescent "comets" or plaques, every 24 h. One day after the observation of spreading infection, 1 mL of media was transferred to a new semi-confluent dish (~1×10^7 cells) of selector cells for the next passage (see **Table S2**), and 2 mL of media was stored at –80 °C for later analysis. To analyze the relative amounts of each virus present after each passage, we measured the relative adenoviral titers by flow cytometry (see above). The ratio of tTA_{wt} and tTA_{mut} viruses was determined by taking the ratio of cells expressing only mCherry and only GFP.

AdProt inhibitor experiments: A confluent 12-well plate of selector cells (**Table S2**) (~4×10^5 cells/well) was infected with tTA_{wt}.mCherry adenovirus (MOI ~ 5). After 4 h, the cells were washed with PBS (Corning), and the AdProt inhibitor was added at the indicated concentrations (0 μM, 1 μM, 20 μM) in the absence or presence of 2 nM doxycycline (dox; Sigma-Aldrich). After 6 days, media and cells were harvested and subjected to three freeze/thaw cycles, and analyzed by flow cytometry (see above).

AdProt inhibitor toxicity assay: A 96-well plate of HEK293A cells were treated with the AdProt inhibitor at concentrations up to 20 μM for 5 days (**Figure S6**). A CellTiter-Glo Luminescent Cell Viability Assay (Promega) was performed according to the manufacturer's instructions. Readings were normalized to the 0 μM AdProt inhibitor samples.

RT-qPCR on selector cells: A confluent plate of selector cells (**Table S2**; ~4×10^5 cells/well) was transfected with 1.25 μg of pTet-Off Advanced (Takara Bio). 2 days later, cells were harvested and the RNA was extracted using an E.Z.N.A Total RNA Kit (Omega Bio-Tek). cDNA was prepared from 1 μg of purified RNA using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). qPCR analysis for *AdProt* (primers: AdProt.Forward and AdProt.Reverse) and the housekeeping gene *RPLP2* (primers: RPLP2.Forward and RPLP2.Reverse; **Table S4**) on a LightCycler 480 II (Roche). *AdProt* transcript levels were normalized to untransfected selector cells (**Table S2**).

Dox dose-response experiment: A confluent 24-well plate of selector cells (**Table S2**; ~1.5×10^5 cells/well) was infected with tTA_{wt}.mCherry adenovirus (MOI ~5). After 4 h, the cells were washed with DMEM (Corning), and dox was added at the indicated concentrations (0 nM, 0.02 nM, 0.1 nM, 0.2 nM, 1 nM, or 2 nM). After 5 days, media and cells were harvested and subjected to three freeze/thaw cycles, followed by analysis of titers using flow cytometry.

Continuous evolution workflow: Before initiating directed evolution, 500 μL of a tTA_{wt}.mCherry adenovirus was amplified on mutator cells (see **Table S2**) to create a diverse viral population. After 5 days, cytopathic effect was observed in all cells. This amplified virus was harvested with three freeze/thaw cycles. Three 15 cm, semi-confluent dishes of selector cells (**Table S2**) (~1×10^7 cells/plate) were infected with either 250, 500, or 1,000 μL of the amplified virus in the presence of dox. Plates were monitored for plaques every day. If more than one plate displayed a plaque on the same day, the plate with the lowest volume of virus added was used for the next round of evolution. The day after a plaque was observed, typically every 4–8 days, three 15 cm semi-confluent dishes of selector cells were again infected in the presence of dox. The three dishes were infected with 250, 500, or 1,000 μL of media from the previous round by direct transfer without a freeze/thaw step. 2 mL of media were saved in Eppendorf tubes and stored at –80 °C for future analysis. In Trial 2, an additional media harvest was performed after full cytopathic effect was observed. In Trial 1, the concentration of dox was increased to 200 nM at passage 7 and then to 20 μM in passages 8–12. In Trial 2, the concentration of dox was held constant at 200 nM for all seven passages.

Measuring promoter activity of viral populations: To follow changes in promoter activity developing during Trial 1, phenotyping cells (**Table S2**) were plated in a 96-well plate at ~40,000 cells/well. The next day, 30 μL of media from passages 1–12 was used to infect two rows of the 96-well plate. Media was removed 5 h post-infection and replaced with media containing 0 μM or 20 μM dox. The cells were then analyzed by flow cytometry (see above for sample preparation) for simultaneous expression of mCherry, indicating that the cell was infected, and GFP, indicating that the promoter was activated by the tTA protein.

Viral genome isolation for next-generation sequencing: Using a viral DNA isolation kit (NucleoSpin Virus; Macherey-Nagel), DNA was harvested from 200 μL of the media that was saved after each round of evolution. A 1.75 kb region of DNA encompassing the CMV promoter and the tTA gene was PCR-amplified from 1 μL of the harvested DNA for 20 rounds of amplification using 5′-ctacataagacccccaccttatatattctttcc-3′ and 5′-agcgggaaaactgaataagaggaagtgaaatc-3′ forward and reverse primers, respectively. The resulting PCR product was purified and prepared for Illumina sequencing via the Nextera DNA Library Prep protocol (Illumina). 250 bp paired-end sequencing was run on a MiSeq (Illumina). Sequencing reads were aligned to the amplicon sequence, which was derived from the tTA_{wt}.mCherry adenovirus sequence using bwa mem 0.7.12-r1039 [RRID:SCR_010910]. Allele pileups were generated using samtools v1.5 mpileup [RRID:SCR_002105] with flags -d 10000000 --excl-flags 2052, and allele counts/frequencies were extracted.[55,56] Each position within the tTA gene and CMV promoter had at least 1,000-fold coverage.

Reverse genetics of tTA variants: HEK-293A cells were seeded in a 12-well plate at ~4×10^5 cells/well. The next day, 0.2 μg of the pBud.tTA.mCherry vector was co-transfected with 1 μg of the pLVX-TRE3G.eGFP vector using 7.2 μL of polyethyleneimine (Polysciences) and 100 μL OPTI-MEM. 8 h post-transfection, media was exchanged and 20 μM dox was added. 48 h post-transfection, cells were analyzed by flow cytometry (see above for sample preparation). Promoter activity was calculated based on the mean fluorescence intensity of GFP fluorescence, backgated for only mCherry-expressing cells.

Testing of recombinase and synthetase selection circuits: HEK-293A cells expressing wt-AdPol were plated at 3.5×10^5 cells/well in a 12-well plate. The next day, 1 μg of the plasmid for each circuit ($(LoxP)_2$Term.AdProt, AdProt(STOP), or AdProt.FLAG as a positive control) was transfected into six wells of a 12-well plate using 6 μL of polyethyleneimine in 100 μL of OPTI-MEM. For the AdProt(STOP) circuit, 0.5 μg was co-transfected with 0.5 μg pLeu-tRNA.GFP(STOP). Media was changed 4 h post-transfection. The next day, transfected wells were infected with either the relevant BOI virus (Table S1; Cre.Ad for $(LoxP)_2$Term.AdProt, and LeuRS.Ad for AdProt(STOP)) or TTA_{wt}.mCherry as a negative control at an MOI of 5. Cells were washed 3× with media 3 h post-infection. After 4 days, media and cells were harvested and subject to three freeze/thaw cycles, followed by analysis of titers using flow cytometry.

ASSOCIATED CONTENT

Supporting Information. This material is available free of charge via the Internet at http://pubs.acs.org.

Nine figures including trans-complementation data, representative flow cytometry data, and additional evolutionary data are included, along with text and tables describing plasmid construction and viral generation.

AUTHOR INFORMATION

Corresponding Author

*mshoulde@mit.edu

Present Addresses

† Janssen Infectious Diseases and Vaccines, Pharmaceutical Companies of Johnson and Johnson, Leiden, The Netherlands

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. ‡These authors contributed equally.

ACKNOWLEDGMENTS

This work was supported by an NIH Director's New Innovator Award (1DP2GM119162), the MISTI Global Seeds Fund, and MIT (all to M.D.S.) and by the European Union through the 6th Framework Program GIANT (contract no. 512087; to R.C.H.). C.M.B. was supported by an NIGMS/NIH Inter-Departmental Biotechnology Training Program (T32-GM008334). C.L.M., S.J.H., and L.J.P. were supported by National Science Foundation Graduate Research Fellowships under Grant No. 1122374. This work was also supported in part by the NIH/NIEHS under award P30-ES002109 and by Cancer Center Support (core) Grant P30-CA14051 from the NIH/NCI.

REFERENCES

1 Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat Rev Genet* **16**, 379-394, doi:10.1038/nrg3927 (2015).

2 Gai, S. A. & Wittrup, K. D. Yeast surface display for protein engineering and characterization. *Curr. Opin. Struct. Biol.* **17**, 467-473, doi:10.1016/j.sbi.2007.08.012 (2007).

3 Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10**, 866-876, doi:10.1038/nrm2805 (2009).

4 Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat. Biotechnol.* **22**, 1567-1572, doi:10.1038/nbt1037 (2004).

5 Branon, T. C. *et al.* Efficient proximity labeling in living cells and organisms with TurboID. *Nat Biotechnol*, doi:10.1038/nbt.4201 (2018).

6 Arzumanyan, G. A., Gabriel, K. N., Ravikumar, A., Javanpour, A. A. & Liu, C. C. Mutually Orthogonal DNA Replication Systems In Vivo. *ACS Synth Biol* **7**, 1722-1729, doi:10.1021/acssynbio.8b00195 (2018).

7 Zetsche, B. *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759-771, doi:10.1016/j.cell.2015.09.038 (2015).

8 Peck, S. H., Chen, I. & Liu, D. R. Directed evolution of a small-molecule-triggered intein with improved splicing properties in mammalian cells. *Chem Biol* **18**, 619-630, doi:10.1016/j.chembiol.2011.02.014 (2011).

9 Piatkevich, K. D. *et al.* A robotic multidimensional directed evolution approach applied to fluorescent voltage reporters. *Nature chemical biology* **14**, 352-360, doi:10.1038/s41589-018-0004-9 (2018).

10 Banaszynski, L. A., Chen, L. C., Maynard-Smith, L. A., Ooi, A. G. & Wandless, T. J. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* **126**, 995-1004, doi:10.1016/j.cell.2006.07.025 (2006).

11 Wang, C. L., Yang, D. C. & Wabl, M. Directed molecular evolution by somatic hypermutation. *Protein Eng. Des. Sel.* **17**, 659-664, doi:10.1093/protein/gzh080 (2004).

12 Wang, L., Jackson, W. C., Steinbach, P. A. & Tsien, R. Y. Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16745-16749, doi:10.1073/pnas.0407752101 (2004).

13 Ma, Y. *et al.* Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat Methods* **13**, 1029-1035, doi:10.1038/nmeth.4027 (2016).

14 Hess, G. T. *et al.* Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nature methods* **13**, 1036-1042, doi:10.1038/nmeth.4038 (2016).

15 Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-424, doi:10.1038/nature17946 (2016).

16 Das, A. T. *et al.* Viral evolution as a tool to improve the tetracycline-regulated gene expression system. *J. Biol. Chem.* **279**, 18776-18782, doi:10.1074/jbc.M313895200 (2004).

17 Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499-503, doi:10.1038/nature09929 (2011).

18 Lucher, L. A. Abortive adenovirus infection and host range determinants. *Curr. Top. Microbiol. Immunol.* **199 (Pt 1)**, 119-152 (1995).

19 Amalfitano, A. & Chamberlain, J. S. Isolation and characterization of packaging cell lines that coexpress the adenovirus E1, DNA polymerase, and preterminal proteins: implications for gene therapy. *Gene therapy* **4**, 258-263, doi:10.1038/sj.gt.3300378 (1997).

20 Hoeben, R. C. & Uil, T. G. Adenovirus DNA replication. *Cold Spring Harb Perspect Biol* **5**, a013003, doi:10.1101/cshperspect.a013003 (2013).

21 Risso-Ballester, J., Cuevas, J. M. & Sanjuán, R. Genome-wide estimation of the spontaneous mutation rate of human adenovirus 5 by high-fidelity deep sequencing. *PLOS Pathog.* **12**, e1006013, doi:10.1371/journal.ppat.1006013 (2016).

22 Uil, T. G. *et al.* Directed adenovirus evolution using engineered mutator viral polymerases. *Nucleic Acids Res.* **39**, e30, doi:10.1093/nar/gkq1258 (2011).

23 Kamtekar, S. *et al.* Insights into strand displacement and processivity from the crystal structure of the protein-primed DNA polymerase of bacteriophage phi29. *Mol. Cell* **16**, 609-618, doi:10.1016/j.molcel.2004.10.019 (2004).

24 Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M. & Belshaw, R. Viral mutation rates. *J. Virol.* **84**, 9733-9748, doi:10.1128/JVI.00694-10 (2010).

25 Davis, J. N. & van den Pol, A. N. Viral mutagenesis as a means for generating novel proteins. *J Virol* **84**, 1625-1630, doi:10.1128/JVI.01747-09 (2010).

26 Phillips, A. M. *et al.* Host proteostasis modulates influenza evolution. *eLife* **6**, e28652, doi:10.7554/eLife.28652 (2017).

27 Elahi, S. M., Oualikene, W., Naghdi, L., O'Connor-McCourt, M. & Massie, B. Adenovirus-based libraries: efficient generation of recombinant adenoviruses by positive selection with the adenovirus protease. *Gene therapy* **9**, 1238-1246, doi:10.1038/sj.gt.3301793 (2002).

28 Greber, U. F., Webster, P., Weber, J. & Helenius, A. The role of the adenovirus protease in virus entry into cells. *EMBO J.* **15**, 1766-1777 (1996).

29 Webster, A., Leith, I. R. & Hay, R. T. Activation of adenovirus-coded protease and processing of preterminal protein. *J. Virol.* **68**, 7292-7300 (1994).

30 Gossen, M. & Bujard, H. Tight Control of gene expression in mammalian cells by tetracycline responsive promoters. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5547-5551 (1992).

31 Loew, R., Heinz, N., Hampf, M., Bujard, H. & Gossen, M. Improved Tet-responsive promoters with minimized background expression. *BMC Biotechnol* **10**, 81, doi:10.1186/1472-6750-10-81 (2010).

32 Krueger, M., Scholz, O., Wisshak, S. & Hillen, W. Engineered Tet repressors with recognition specificity for the tetO-4C5G operator variant. *Gene* **404**, 93-100, doi:10.1016/j.gene.2007.09.002 (2007).

33 Grosche, P. *et al.* Structure-based design and optimization of potent inhibitors of the adenoviral protease. *Bioorg. Med. Chem. Lett.* **25**, 438-443, doi:10.1016/j.bmcl.2014.12.057 (2015).

34 Hecht, B., Muller, G. & Hillen, W. Noninducible Tet repressor mutations map from the operator binding motif to the C terminus. *J. Bacteriol.* **175**, 1206-1210, doi:10.1128/jb.175.4.1206-1210.1993 (1993).

35 Meinke, G., Bohm, A., Hauber, J., Pisabarro, M. T. & Buchholz, F. Cre Recombinase and Other Tyrosine Recombinases. *Chemical Reviews* **116**, 12785-12820, doi:10.1021/acs.chemrev.6b00077 (2016).

36 Italia, J. S. *et al.* Expanding the genetic code of mammalian cells. *Biochemical Society Transactions* **45**, 555-562, doi:10.1042/bst20160336 (2017).

37 Hartley, J. L., Temple, G. F. & Brasch, M. A. DNA cloning using in vitro site-specific recombination. *Genome Res* **10**, 1788-1795, doi:DOI 10.1101/gr.143000 (2000).

38 Russell, W. C. Update on adenovirus and its vectors. *J Gen Virol* **81**, 2573-2604, doi:10.1099/0022-1317-81-11-2573 (2000).

39 Wang, H. *et al.* Improved seamless mutagenesis by recombineering using ccdB for counterselection. *Nucleic Acids Res.* **42**, e37, doi:10.1093/nar/gkt1339 (2014).

40 Meyerhans, A. *et al.* Temporal fluctuations in HIV quasispecies in vivo are not reflected by sequential HIV isolations. *Cell* **58**, 901-910, doi:10.1016/0092-8674(89)90942-2 (1989).

41 O'Loughlin, T. L., Greene, D. N. & Matsumura, I. Diversification and specialization of HIV protease function during in vitro evolution. *Mol Biol Evol* **23**, 764-772, doi:10.1093/molbev/msj098 (2006).

42 Myers, N. D., Skorohodova, K. V., Gounder, A. P. & Smith, J. G. Directed evolution of mutator adenoviruses resistant to antibody neutralization. *J. Virol.* **87**, 6047-6050, doi:10.1128/JVI.00473-13 (2013).

43 Meng, F. L. *et al.* Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell* **159**, 1538-1548, doi:10.1016/j.cell.2014.11.014 (2014).

44 Kim, D. *et al.* Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat. Biotechnol.* **35**, 475-480, doi:10.1038/nbt.3852 (2017).

45 Wang, C. L., Harper, R. A. & Wabl, M. Genome-wide somatic hypermutation. *Proc Natl Acad Sci U S A* **101**, 7352-7356, doi:10.1073/pnas.0402009101 (2004).

46 Hubbard, B. P. *et al.* Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nat. Methods.* **12**, 939-942, doi:10.1038/nmeth.3515 (2015).

47 Dickinson, B. C., Packer, M. S., Badran, A. H. & Liu, D. R. A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. *Nat. Commun.* **5**, 5352, doi:10.1038/ncomms6352 (2014).

48 Badran, A. H. *et al.* Continuous evolution of Bacillus thuringiensis toxins overcomes insect resistance. *Nature* **533**, 58-63, doi:10.1038/nature17938 (2016).

49 Carlson, J. C., Badran, A. H., Guggiana-Nilo, D. A. & Liu, D. R. Negative selection and stringency modulation in phage-assisted continuous evolution. *Nature chemical biology* **10**, 216-222, doi:10.1038/nchembio.1453 (2014).

50 Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57-63, doi:10.1038/nature26155 (2018).

51 Moore, C. L., Papa, L. J., III & Shoulders, M. D. A Processive Protein Chimera Introduces Mutations across Defined DNA Regions In Vivo. *J Am Chem Soc*, doi:10.1021/jacs.8b04001 (2018).

52 Dull, T. *et al.* A third-generation lentivirus vector with a conditional packaging system. *J. Virol.* **72**, 8463-8471 (1998).

53 Fallaux, F. J. *et al.* Characterization of 911: a new helper cell line for the titration and propagation of early region 1-deleted adenoviral vectors. *Hum Gene Ther* **7**, 215-222, doi:10.1089/hum.1996.7.2-215 (1996).

54 Ossowski, S. *et al.* Sequencing of natural strains of Arabidopsis thaliana with short reads. *Genome Res* **18**, 2024-2033, doi:10.1101/gr.080200.108 (2008).

55 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

56 Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987-2993, doi:10.1093/bioinformatics/btr509 (2011).

An Adaptable Platform for Directed Evolution in Human Cells

Chet M. Berman, Louis J. Papa III, Samuel J. Hendel, Christopher L. Moore, Patreece H. Suen, Alexander F. Weickhardt, Ngoc-Duc Doan, Caiden M. Kumar, Taco G. Uil, Vincent L. Butty, Robert C. Hoeben, Matthew D. Shoulders

Page	Contents
S1	Table of Contents
S2–S5	Supporting Methods
S6–S14	Supporting Figures
S15–S17	Supporting Tables
S18	Supporting References

SUPPORTING METHODS

New vectors and cell lines reported here are available from the Principal Investigator upon reasonable request.

General: LV-Cre pLKO.1 was a gift from Elaine Fuchs (Addgene plasmid #25997)[1] and pANAP was a gift from Peter Schultz (Addgene plasmid #48696).[2] Plasmid sequences can be obtained from GenBank using the accession numbers provided in **Table S5**.

Wild-type AdPol and EP-Pol vectors: The lentiviral vector encoding HA-tagged wild-type AdPol was previously described.[3] Mutations were introduced by site-directed mutagenesis.

CMV.AdProt vector: A 641 bp fragment containing adenoviral protease (AdProt) was amplified from the Ad5 genome using the primers BamHI.AdProt Forward and SalI.AdProt Reverse (**Table S4**) and ligated into pTRE-Tight (Clontech) using BamHI and SalI to make the pTRE-Tight.AdProt vector. The Ad5 Tripartite leader sequence (TPL) was amplified from the TPL gene block using the primers TPL.GA.Forward and TPL.GA.Reverse (**Table S4**) and the pTRE-Tight.AdProt vector was amplified using the primers Tight.AdProt.GA.Forward and Tight.AdProt.GA.Reverse (**Table S4**). The TPL and pTRE-Tight.AdProt amplicons were assembled using the HiFi DNA assembly kit (New England Biolabs) to create the pTRE-Tight.TPL.AdProt vector. From this vector, an 852 bp fragment containing TPL.AdProt was amplified using the primers NotI.TPL.AdProt.Forward and XbaI.TPL.AdProt.Reverse (**Table S4**) and inserted into the pENTR1A vector (Thermo Fisher) using NotI and XbaI. The LR clonase II enzyme mixture (Thermo Fisher) was used to recombine the TPL.AdProt fragment from pENTR1A.TPL.AdProt into pLenti.CMV.Hygro (w117-1) (Thermo Fisher).

TRE3G.AdProt vector: TPL-AdProt was amplified from pTRE-Tight.TPL.AdProt using the primers TPL.AdProt.GA.Forward and TPL.AdProt.GA.Reverse (**Table S4**) and assembled with NotI-digested pLVX.Tight.Puro (Takara Biosciences) using the HiFi DNA assembly kit to form pLVX.Tight.TPL.AdProt.Puro. A fragment containing TPL.AdProt was obtained from pLVX.Tight.TPL.AdProt.Puro by digestion with EcoRI and BamHI and ligated into the pLVX.TRE3G vector (Takara Bio) to create the pLVX.TRE3G.AdProt vector.

TRE3G.eGFP vector: A 762 bp fragment containing eGFP was amplified from the eGFP-N3 vector (Takara Bio) using the primers NotI.eGFP.Forward and EcoRI.eGFP.Reverse (**Table S4**) and ligated into the pLVX-TRE3G vector (Takara Bio) using NotI and EcoRI to create the pLVX-TRE3G.eGFP vector.

tTA variant vectors: A 743 bp fragment containing mCherry was amplified from a pcDNA3.1-mCherry template plasmid using the primers NotI.mCherry.Forward and XhoI.mCherry.Reverse (**Table S4**) and inserted into the pBudCE4.1 vector (Thermo Fisher) using NotI and XhoI to create the pBud.mCherry vector. A 771 bp fragment containing tTA was amplified from a tTA.mCherry adenoviral vector using the primers SalI.TTA.Forward and BamHI.TTA.Reverse (**Table S4**) and inserted into the pBud.mCherry vector using BamHI and SalI to create the pBud.tTA.mCherry vector. Site-directed mutagenesis was then performed on pBud.tTA.mCherry using a QuickChange II XL Site-Directed Mutagenesis Kit (Agilent) to generate the indicated point mutations in tTA (**Figure 4f**).

AdProt.flag vector: From the pTRE-Tight.TPL.AdProt vector, an 852 bp fragment containing TPL.AdProt was amplified using primers NotI.TPL.Forward and XbaI.AdProt.Reverse (**Table S4**) and inserted into the pENTR1A vector using NotI and XbaI to form pENTR1A.TPL.AdProt. A FLAG epitope tag was inserted into pENTR1A.TPL.AdProt using primers pENTR1A.AdProt.FLAG.Forward and pENTR1A.AdProt.FLAG.Reverse (**Table S4**) and using the QuickChange II XL Site-Directed Mutagenesis Kit to form pENTR1A.TPL.AdProt.FLAG. This vector was then recombined with pcDNA-DEST40 using LR Clonase II Enzyme Master Mix to form pcDNA.TPL.AdProt.FLAG.

$(LoxP)_2$Term.AdProt vector: A vector containing an SV40-polyA terminator flanked by two loxP sites was purchased from GeneArt (ThermoFisher). From this vector, a 370 bp fragment containing the floxed SV40-terminator signal was amplified using primers LoxP2Term.GA.Forward and LoxP2Term.GA.Reverse (**Table S4**). The

pENTR1A.TPL.AdProt.FLAG vector was linearized using pENT.AdProt.GA.Forward and pENT.AdProt.GA.Reverse (**Table S4**). The two amplicons were assembled to form pENTR1A.$(LoxP)_2$Term.TPL.AdProt.FLAG using the NEB HiFi DNA assembly kit. This vector was then recombined with pcDNA-DEST40 using LR Clonase II Enzyme Master Mix to form pcDNA.$(LoxP)_2$Term.TPL.AdProt.FLAG.

AdProt(STOP) vector: pENTR1A.TPL.AdProt.FLAG was mutagenized using primers L8.STOP.Forward and L8.STOP.Reverse (**Table S4**) to form pENTR1A.TPL.AdProt(STOP).FLAG using the QuickChange II XL Site-Directed Mutagenesis Kit. This vector was then recombined with pcDNA-DEST40 using LR Clonase II Enzyme Master Mix to form pcDNA.TPL.AdProt(STOP).FLAG.

pLeu-tRNA.LeuRS vector: A 2607 bp fragment containing LeuRS, the *E. coli* leucyl-tRNA synthetase, was amplified from DH10B *E. coli* genomic DNA using the primers HindIII.LeuRS.Forward and XhoI.LeuRS.Reverse (**Table S4**) and inserted into pANAP[2] using HindIII and XhoI to create the pLeu-tRNA.LeuRS vector.

pLeu-tRNA.GFP(STOP) vector: Site-directed mutagenesis was performed on the pcDNA3.1-CMV.GFP plasmid using a QuickChange II XL Site-Directed Mutagenesis Kit (Agilent) and the primers Tyr40TAG.Forward and Tyr40TAG.Reverse to introduce a premature stop codon at postion 40 in eGFP. Then a 750 bp fragment containing eGFP(STOP) was amplified from the site-directed mutagenesis product using the primers HindIII.eGFP.Forward and XhoI.eGFP.Reverse (**Table S4**) and inserted into pANAP[2] using HindIII and XhoI to create the pLeu-tRNA.GFP(STOP) vector.

Adenoviral constructs: Adenoviral constructs were engineered using *ccdB* recombineering, as previously described,[4] in DH10B *Escherichia coli* carrying the adenovirus type 5 genome in a chloramphenicol-resistant bacterial artificial chromosome (AdBAC). Cells carrying an AdBAC were transformed with the temperature-sensitive psc101-gbaA recombineering plasmid,[4] plated on LB (Difco) agar (Alfa Aesar) with 10 μg/mL tetracycline (CalBioChem) and 10 μg/mL chloramphenicol (Alfa Aesar), and incubated for 24 h at 30 °C. Colonies were selected and grown in LB containing 10 μg/mL tetracycline and 10 μg/mL chloramphenicol overnight at 30 °C (18–21 h). Overnight cultures were diluted 25-fold in LB with 10 μg/mL tetracycline and 10 μg/mL chloramphenicol and grown at 30 °C for ~2 h until attaining an OD_{600} of 0.3–0.4. The ccdA antitoxin and recombineering machinery were then induced by adding L-arabinose (Chem-Impex) and L-rhamnose (Sigma Aldrich) to a final concentration of 2 mg/mL each and then growing the cultures at 37 °C for 40 min to an OD_{600} of ~0.6. The cultures were then placed on ice, washed twice with ice-cold, sterile ddH_2O, resuspended in ~25 μL of ice-cold, sterile ddH_2O, and electroporated with ~200 ng of the appropriate kan-ccdB targeting cassette (1.8 kV, 5.8 ms, 0.1 cm cuvette, BioRad Micropulser). The cells were then recovered in super optimal broth with catabolite repression (SOC; Teknova) with 2 mg/mL L-arabinose at 30 °C for 2 h, then plated on LB agar plates with 50 μg/mL kanamycin (Alfa Aesar) and 2 mg/mL L-arabinose and incubated for 24 h at 30 °C. Colonies that grew under these conditions had incorporated the kan-ccdB targeting cassette and were picked in triplicate and grown in LB with 50 μg/mL kanamycin and 2 mg/mL L-arabinose at 30 °C for 18–21 h. Note that the colonies were picked in triplicate because multimers of the AdBAC formed at a high rate (~30–50% of colonies) during the first recombineering step. Such multimers cannot be successfully recombineered in the next step. Picking three colonies and recombineering them separately in parallel increases the chances of picking a monomer that can be successfully recombineered. The cultures were then diluted 25-fold in LB with 50 μg/mL kanamycin and 2 mg/mL L-arabinose and grown at 30 °C for ~2 h until they reached an OD_{600} of 0.3–0.4. The recombineering machinery was then induced by adding L-rhamnose to a final concentration of 2 mg/mL and then growing the cultures at 37 °C for 40 min to an OD_{600} of ~0.6. The cultures were then placed on ice, washed twice with ice-cold, sterile ddH_2O, resuspended in ~25 μL of ice-cold, sterile ddH_2O, and electroporated with ~200 ng of the final targeting cassette intended to replace the kan-ccdB cassette currently integrated in the genome (1.8 kV, 5.8 ms, 0.1 cm cuvette, BioRad Micropulser). The cells were then recovered in SOC with 2 mg/mL L-arabinose at 30 °C for 2 h, and then were washed once with LB to remove the L-arabinose and prevent continued production of the ccdA antitoxin. The cultures were then plated on LB agar plates at various dilutions with 10 μg/mL tetracycline and 10 μg/mL chloramphenicol and incubated for 24 h at 37 °C. Without the ccdA antitoxin, the ccdB toxin will kill cells that have not replaced the integrated kan-ccdB cassette with the final targeting cassette. The colonies that grow should,

in principle, have the desired final targeting cassette integrated, but were always screened by PCR or sequencing to confirm cassette integration as some colonies may simply inactivate the ccdB toxin.

The following modifications were made using the primers in **Table S4** to obtain the adenoviruses (**Table S1**) used in this work:

Modification	**Genotype**	**KanccdB cassette primers used with R6K-kan-ccdB template plasmid (unless stated otherwise)**	**Final targeting cassette oligos or primers and template (if applicable)**	**Purpose of modification**
AdPol Deletion	ΔAdPol	Pol.kanccdB.Forward and Pol.kanccdB.Reverse	delPol.Forward and delPol.Reverse (annealed oligos)	To prevent evolution of the adenoviral polymerase. The error-prone version was expressed in trans.
Insertion of mCherry	E4R-mCherry	E4.kanccdB.Forward and E4.kanccdB.Reverse	E4.SV40.Promoter.Forward and E4.SV40.Reverse were used to amplify from pcDNA3.1-mCherry template plasmid	mCherry was inserted to enable the visualization of infected cells. The E4 position with the rightward facing orientation was previously shown to allow for optimal expression and viral titer.[5]
Insertion of eGFP	E4R-eGFP	E4.kanccdB.Forward and E4.kanccdB.Reverse	E4.SV40.Promoter.Forward and E4.SV40.Reverse were used to amplify from pcDNA3.1-eGFP template plasmid	eGFP was inserted to enable the visualization of infected cells. The E4 position with the rightward facing orientation was previously shown to allow for optimal expression and viral titer.[5]
AdProt Deletion	ΔAdProt	AdProt.kanccdB.Forward and AdProt.kanccdB.Reverse	delAdProt.Forward and delAdProt.Reverse (annealed oligos)	This essential viral gene was deleted so that viral replication could become dependent on the conditional expression of the adenoviral protease in trans.
Insertion of active tTA	E1L-tTA	E1.kanccdB.Forward and E1.kanccdB.Reverse	E1.CMV.Promoter Forward and E1.bGH.polyA.Reverse used to amplify from pcDNA3.1-tTA template plasmid	tTA was inserted as the evolution target that must evolve to express adenoviral protease from the host genome for efficient viral propagation. The E1 position with the leftward facing orientation was previously shown to allow for optimal expression and viral titer.[5]
Insertion of Cre Recombinase	E1L-Cre	TetR.kanccdB.Forward and TetR.kanccdB.Reverse	Cre.Forward and Cre.Reverse used to amplify from LV-Cre pLKO.1 (Addgene #25997)	Cre recombinase was7 inserted as a model BOI for selection circuit experiments.
Insertion of LeuRS	E1L-LeuRS	E1L.KanccdB.Forward and E1.kanccdB.Reverse used to amplify from pcDNA3.1-KanFDEST template plasmid	E1.CMV.Promoter.Forward and E1.bGH.polyA.Reverse used to amplify from pLeu-tRNA.LeuRS	LeuRS aminoacyl-tRNA synthetase was inserted as a model BOI for selection circuit experiments.
Insertion of DEST cassette	E1L-DEST	E1L.KanccdB.Forward and E1.kanccdB.Reverse used to amplify from pcDNA3.1-	Not applicable, only the first step is required	Insertion of a DEST cassette into the E1 position with the leftward facing orientation. The DEST cassette has *att*R sites that allow

		KanFDEST template plasmid		users to insert genes via Gateway cloning.
Replacement of the low copy BAC origin with the high copy pUC origin	Not applicable	N/A, the replacement is a one step recombineering since the origin switches from chloramphenicol to ampicillin resistant	BAC2pUC.Forward and BAC2pUC.Reverse used to amplify the pUC origin cassette from pAd/CMV/V5-DEST (Thermo Fisher).	Switching to a high copy pUC origin allowed for the preparation of concentrated, purified DNA, which was necessary for transfection and successful adenovirus production.

Once a clone with all of the desired genetic changes was found and confirmed by Sanger sequencing, the AdBAC single-copy replication origin was replaced with the high copy pUC origin. The cells with the correct clone were grown in LB containing 10 μg/mL tetracycline and 10 μg/mL chloramphenicol overnight at 30 °C (18–21 h). Overnight cultures were diluted 25-fold in LB with 10 μg/mL tetracycline and 10 μg/mL chloramphenicol and grown at 30 °C for ~2 h until attaining an OD_{600} of 0.3–0.4. The recombineering machinery was then induced by adding L-rhamnose to a final concentration of 2 mg/mL each and then growing the cultures at 37 °C for 40 min to an OD_{600} of ~0.6. The cultures were then placed on ice, washed twice with ice-cold, sterile ddH_2O, resuspended in ~25 μL of ice-cold, sterile ddH_2O, and electroporated with ~200 ng of the pUC origin cassette (1.8 kV, 5.8 ms, 0.1 cm cuvette; BioRad Micropulser). The cells were recovered in SOC at 30 °C for 2 h, then plated on LB agar plates with 100 μg/mL ampicillin and incubated for 24 h at 37 °C. The resulting amp-resistant colonies should have the pUC origin inserted and were checked by verifying expected restriction digestion patterns. The colonies were grown in 25 mL LB containing 100 μg/mL ampicillin and the DNA was purified using the Zymo-PURE II plasmid midiprep kit (Zymo Research) according to the manufacturer's instructions. The DNA was digested with PacI overnight at 37 °C in order to liberate and linearize the adenoviral genome. The linearized DNA was purified using the E.Z.N.A. cycle pure kit (Omega Bio-tek) according to the manufacturer's instructions.

SUPPORTING FIGURES

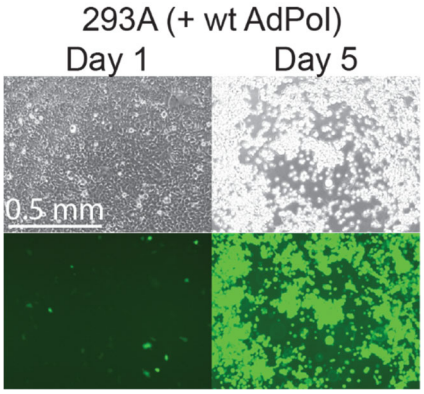


Figure S1. Trans-complementation of wild-type adenoviral polymerase (AdPol). Parental HEK293A cells stably expressing wt AdPol were infected with a GFP-encoding ΔAdPol-adenovirus (CFP.ΔAdPol.GFP). The virus propagated robustly on these AdPol expressing cells.

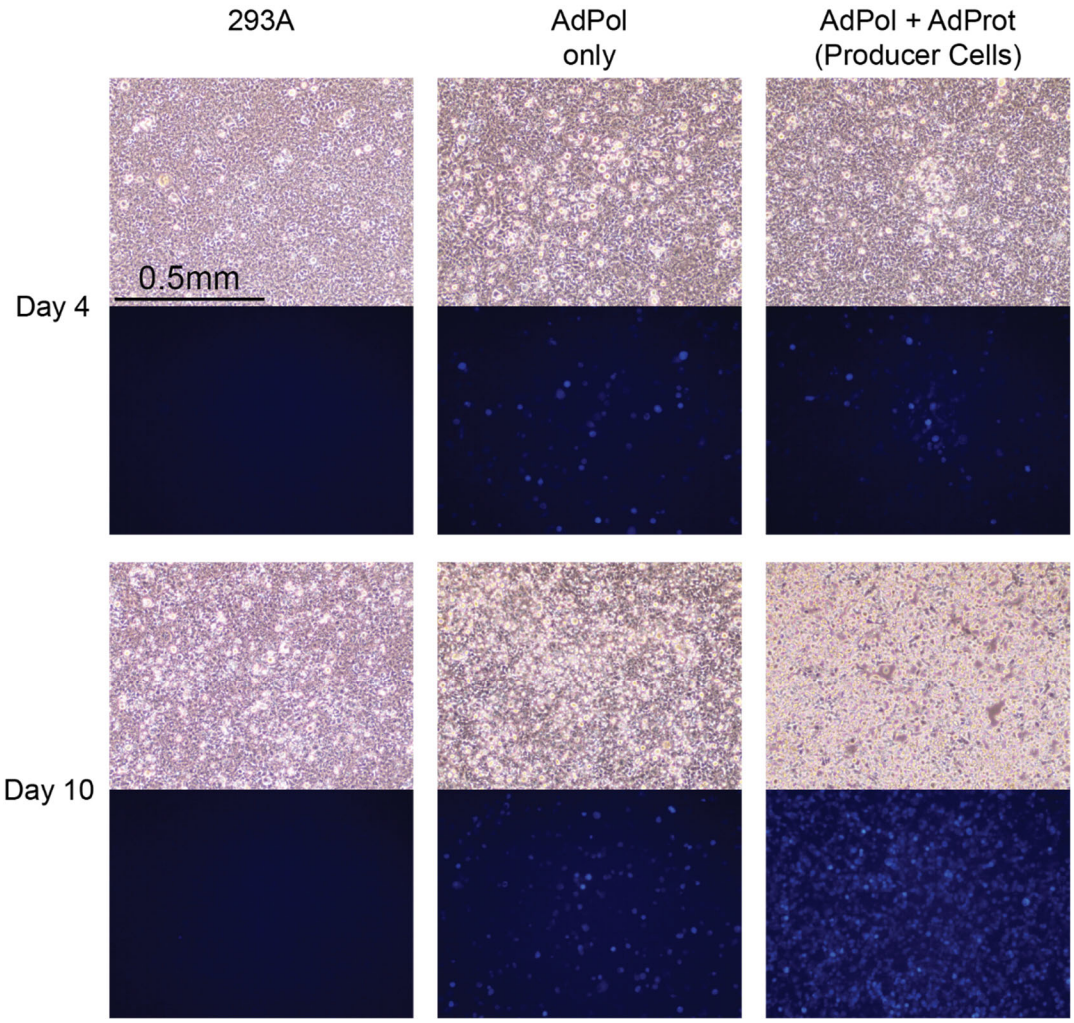


Figure S2. ΔAdProtΔAdPol-adenovirus was used to infect HEK293A cells, AdPol-expressing cells, or producer cells (**Table S2**) at a low multiplicity of infection (< 0.5). The infection was monitored over 10 days. The parental HEK293A cells showed no visible sign of infection, likely because without AdPol expression the copy number of the CFP gene was too low in the cell to easily visualize fluorescence with laboratory microscopes. The AdPol-expressing cells showed a strong CFP signal indicating a robust infection, however the infection did not spread owing to a lack of AdProt. In contrast, the producer cells trans-complementing both AdPol and AdProt were able to support a spreading infection, with every cell in the plate infected by day 10.

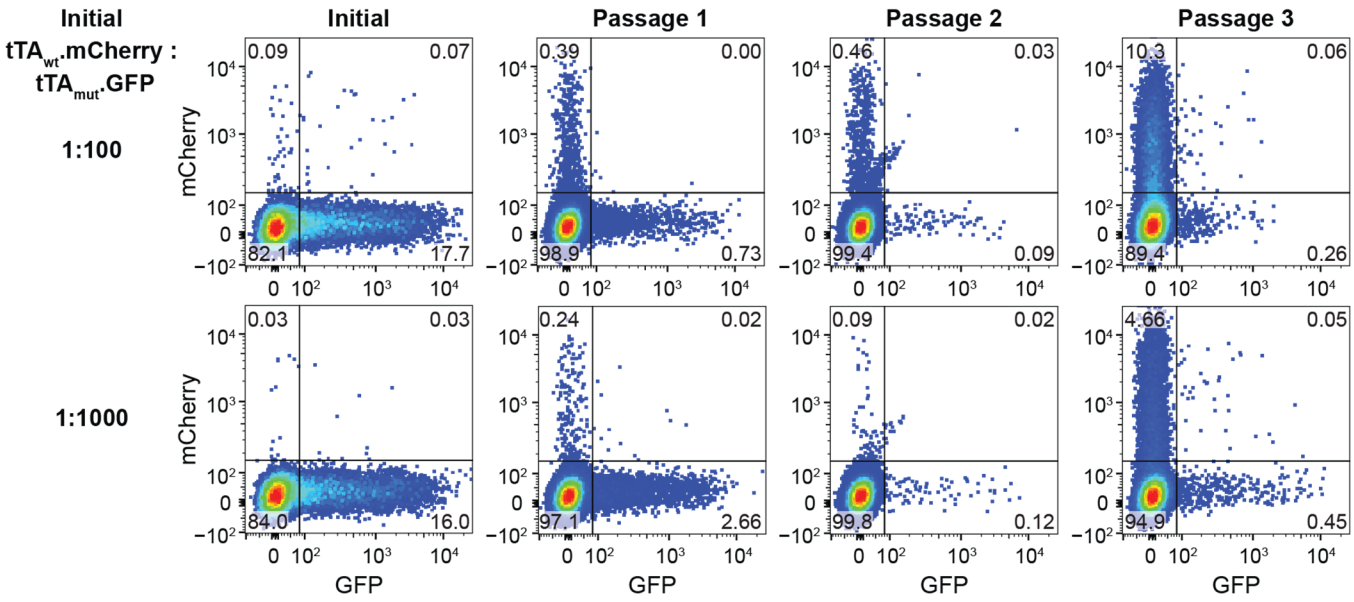
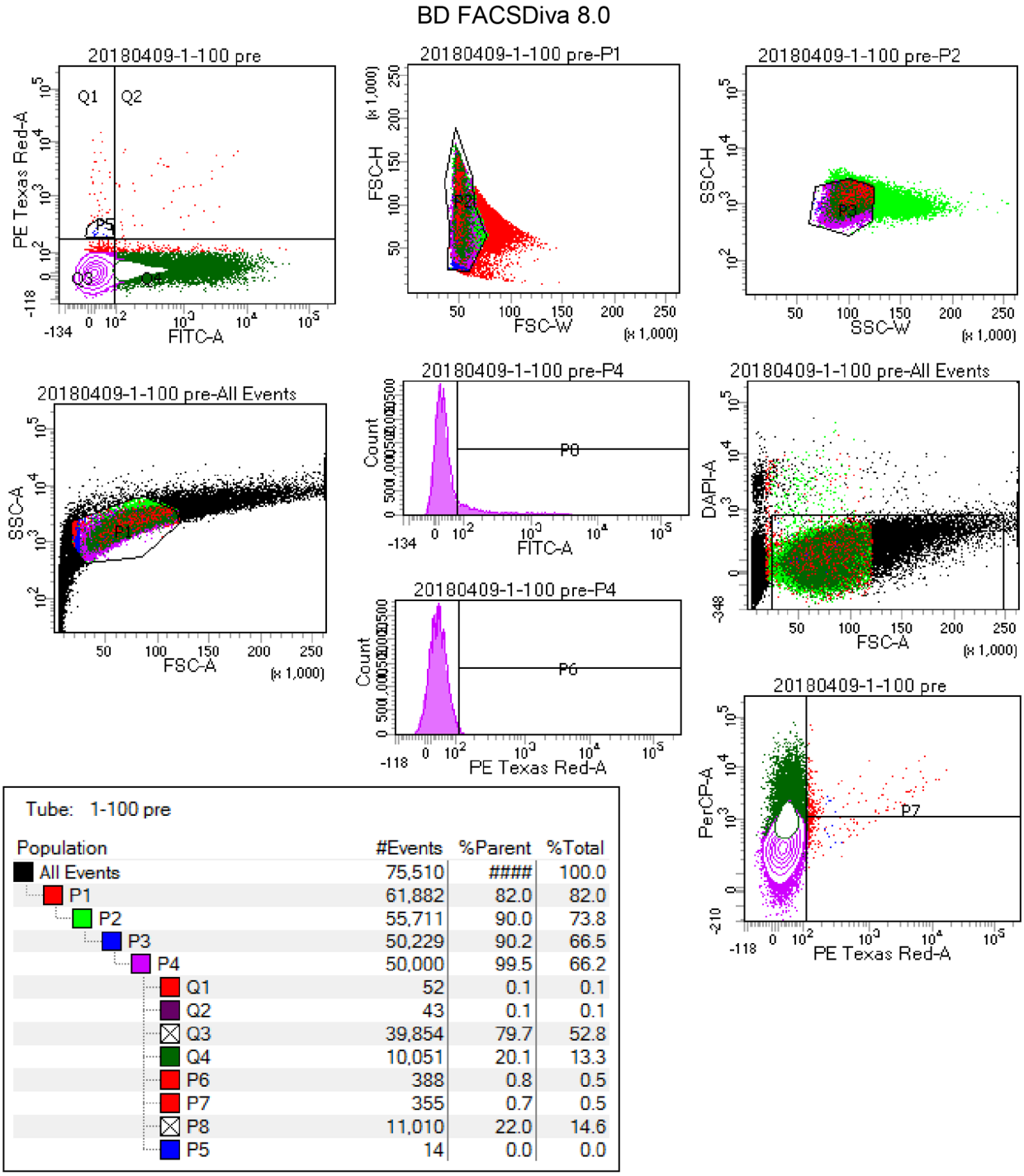


Figure S3. Flow cytometry data showing infection with mixed samples of tTA$_{wt}$.mCherry adenoviruses and tTA$_{mut}$.GFP adenoviruses (**Table S1**) over three serial passages. Density plots show cells infected with tTA$_{wt}$.mCherry adenoviruses (Q1), tTA$_{mut}$.GFP adenoviruses (Q4), or both (Q2). Quantifications of each quadrant as a percentage of the total population are shown.



Tube: 1-100 pre

Population	#Events	%Parent	%Total
All Events	75,510	####	100.0
P1	61,882	82.0	82.0
P2	55,711	90.0	73.8
P3	50,229	90.2	66.5
P4	50,000	99.5	66.2
Q1	52	0.1	0.1
Q2	43	0.1	0.1
Q3	39,854	79.7	52.8
Q4	10,051	20.1	13.3
P6	388	0.8	0.5
P7	355	0.7	0.5
P8	11,010	22.0	14.6
P5	14	0.0	0.0

Figure S4. Example raw flow cytometry data. Gates P1–P3 were used to eliminate cell debris and cell aggregates. Gate P4 was used to remove dead cells by only gating for DAPI-negative cells. Gates Q1–Q4 were used to gate for GFP-positive and mCherry-positive cells. This specific data set was used to calculate the initial ratio of tTA_{wt}.mCherry virus to tTA_{mut}.GFP virus in the competition experiment (**Figure 2a** and **Figure S3**).

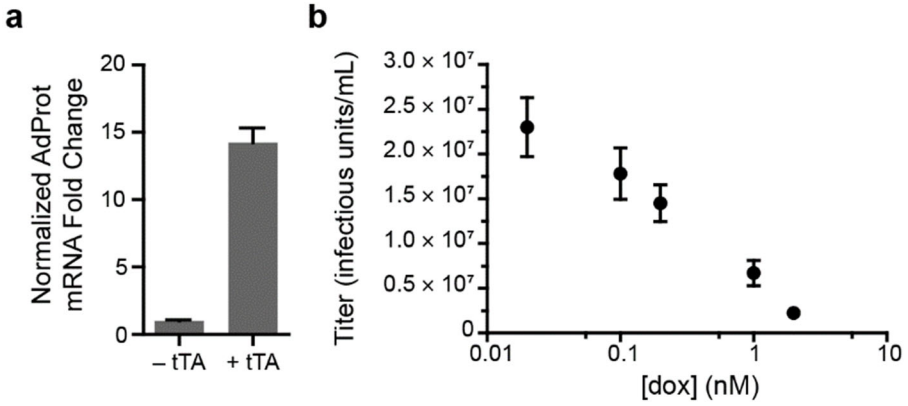


Figure S5. (a) AdProt transcript levels were analyzed by qPCR in selector cells 2 days after transfection with pTet-Off Advanced. AdProt transcript levels normalized to untransfected selector cells (**Table S2**). (b) Titer of tTA_{wt}.mCherry adenovirus after infection of selector cells (**Table S2**) treated with varying amounts of doxycycline (dox). Titers are reported as "infectious units/mL" defined as the number of fluorescent cells per mL of viral supernatant used during the infection.

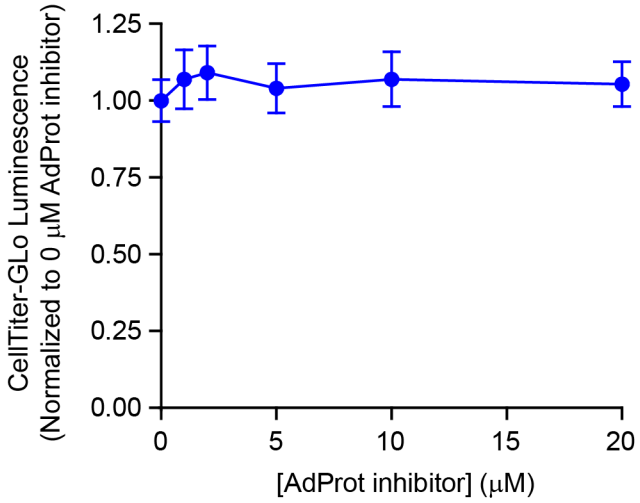


Figure S6. CellTiter-GLo assay to assess any impact of the AdProt inhibitor on cell viability at concentrations up to 20 μM (*N* = 5, 5-day treatment).

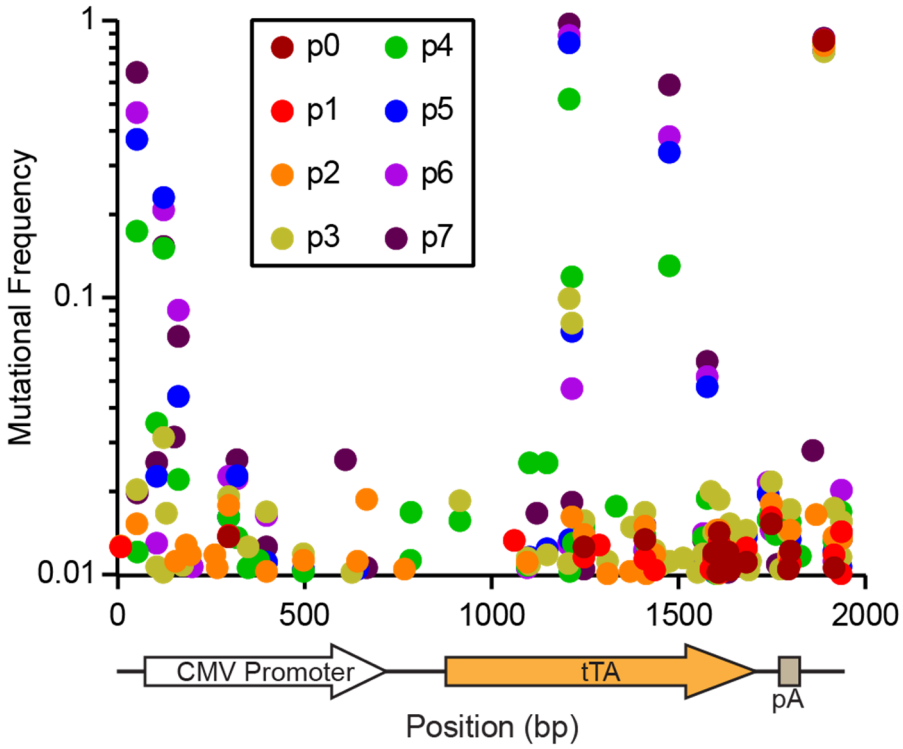


Figure S7. Non-reference allele frequencies at ≥ 1% frequency over the course of the directed evolution experiment for Trial 2.

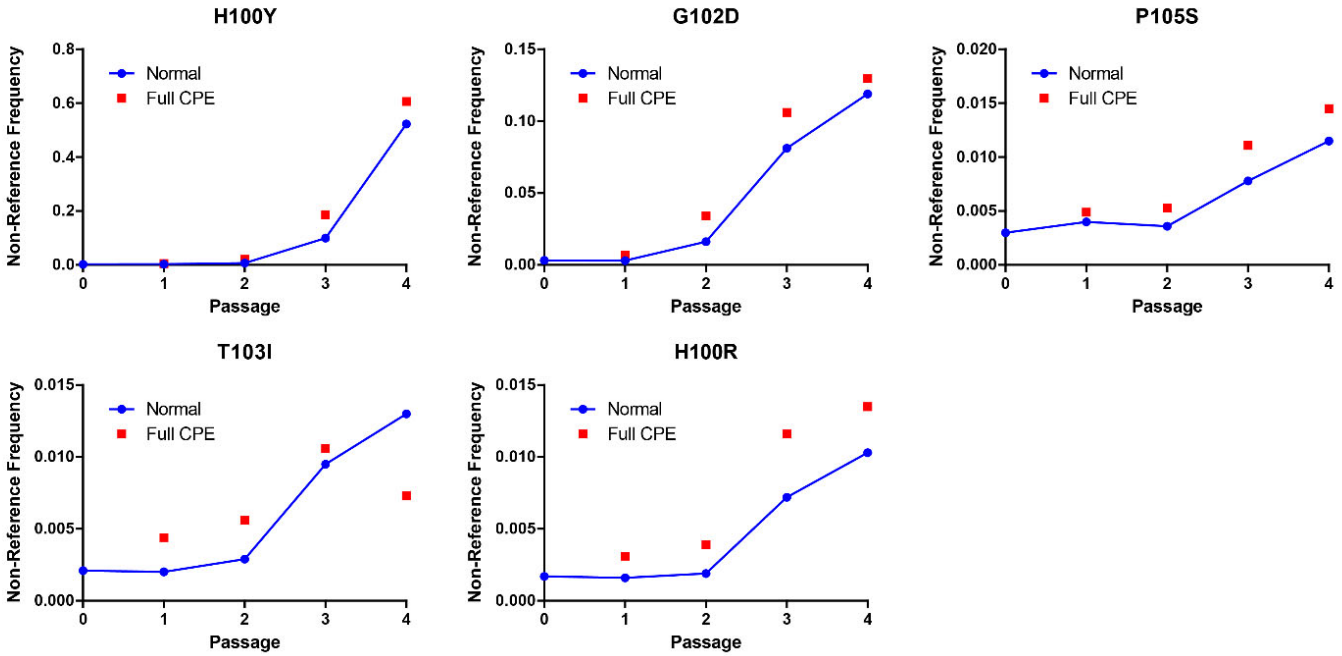


Figure S8. Comparison of mutation frequencies in Trial 2 using the early harvesting protocol (75% of cells infected) and late harvesting for each passage after full CPE was attained. Five previously reported[6] doxycycline-resistant variants that reached a frequency of 1% by passage 4 are shown.

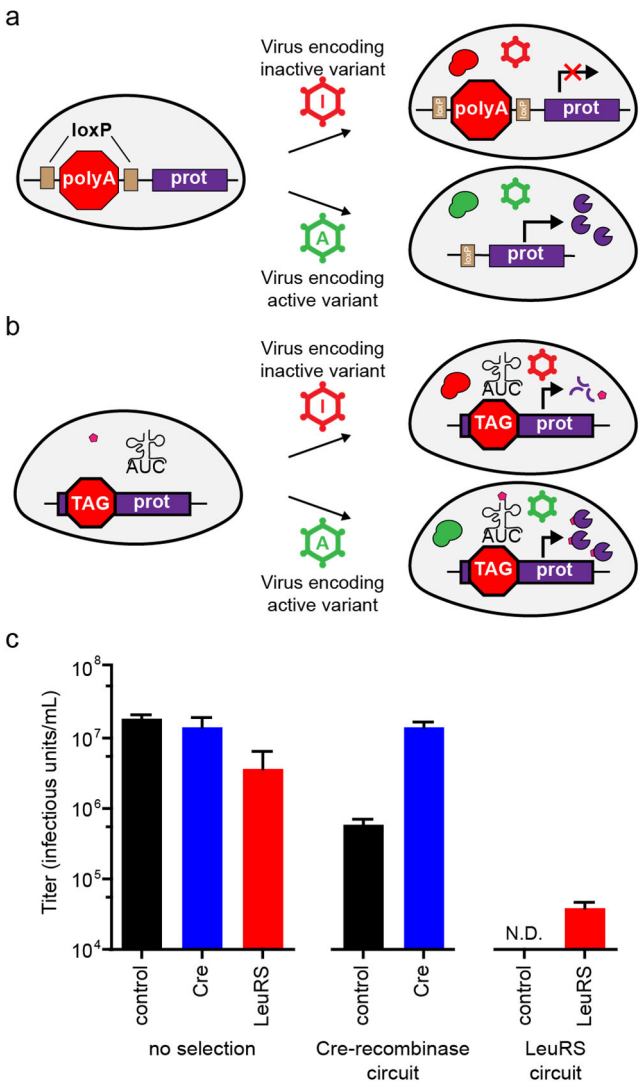


Figure S9. (a) Selection circuit designed for AdProt-based selection of Cre recombinase activity. A floxed SV40-polyA terminator signal prevents transcription of AdProt unless Cre deletes the terminator by recombination at target loxP sites. (b) Selection circuit designed for AdProt-based selection of leucyl-tRNA synthetase (LeuRS) activity. A premature amber stop codon (TAG) prevents translation of full-length AdProt unless LeuRS charges an amber-anticodon tRNA with leucine (pink) and suppresses the amber stop codon. (c) Cells were transfected with either a constitutive protease control (no selection, AdProt.FLAG), the Cre-recombinase circuit ($(LoxP)_2$Term.AdProt), or the LeuRS circuit (AdProt(STOP)) with the relevant tRNAs (pLeu-tRNA.GFP(STOP)). Transfected cells were then infected with ΔAdProt.adenovirus carrying tTA (control, TTA_{wt}.mCherry), Cre (Cre.Ad), or LeuRS (LeuRS.Ad). The infections were allowed to progress for four days before they were harvested and titered by flow cytometery. Titers are provided in infectious units per milliliter. N.D. indicates that the control virus was not detected after passaging on the synthetase selection circuit.

SUPPORTING TABLES

Table S1. Adenoviruses constructed and used in this study.

Name	Modifications relative to wild-type Ad5
AdCFP	E1R-CFP ΔE1 ΔE3
CFP.ΔAdPol.GFP	E1R-CFP ΔE1 ΔE3 ΔAdPol E4R-GFP
tTA_{wt}.mCherry	E1L-tTA ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry
tTA_{mut}.GFP	E1L-tTA_{aak} ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-GFP
Cre.Ad	E1L-Cre ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry
LeuRS.Ad	E1L-LeuRS ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry
AdEvolve-DEST	E1L-DEST ΔE1 ΔE3 ΔAdProt ΔAdPol E4R-mCherry
ΔAdProtΔAdPol-adenovirus	E1R-CFP ΔE1 ΔE3 ΔAdProt ΔAdPol
AdGLΔPol[3]	E1L-Luciferase-GFP ΔE1 ΔE3

Note: All viruses used in this work were derived from AdCFP except for AdGLΔPol, which was previously reported.[3]

Table S2. Cell lines used in this study.

Cell line	Polymerase	Transgene cassette
Producer	AdPol	CMV.AdProt
Mutator	EP-Pol	CMV.AdProt
Selector	EP-Pol	TRE3G.AdProt
Phenotyping	AdPol	TRE3G.eGFP

Note: All cell lines were derived from HEK293A cells.

Table S3. Tabulation of next-generation sequencing results and experimental parameters used to estimate the EP-Pol mutation rate.[3]

Estimated number of clones sequenced	Size of the region sequenced and analyzed (bp)	Substitution load per million bp	Substitutions per Ad genome per viral generation
27.3*	6020	365	1.31**

* Viral pool size was estimated based on intra-experiment titrations during pool preparations
**Assuming a genome size of 36 kb and that 27.3 genomes were sequenced. Each of the 10 passages was defined as a generation.

Table S4. Primers used to construct lentiviral and adenoviral plasmids through cloning and recombineering.

Name	Sequence
BamHI.AdProt.Forward	5′-aaaaaaggatccaccatgggctccagtgag-3′
SalI.AdProt.Reverse	5′-aaaaagtcgacttacatgtttttcaagtgacaaaaagaag-3′
EcoRI.TPL.Forward	5′-aaaaaagcggccgcactctcttccgcatcg-3′
BamHI.TPL.Reverse	5′-aaaaaatctagattacatgtttttcaagtgacaaaaagaag-3′
TPL.GA.Forward	5′-atcgcctggagaattcactctcttccgcatcgct-3′
TPL.GA.Reverse	5′-ctcactggagcccattgcgactgtgactggttag-3′
TPL Gene Block	5′-aaaaaagaattcactctcttccgcatcgctgtctgcgagggccagctgttgggctcgcggttgaggacaa actcttcgcggtctttccagtactcttggatcggaaacccgtcggcctccgaacaggtactccgccgccgagg gacctgagcgagtccgcatcgaccggatcggaaaacctctcgagaaaggcgtctaaccagtcacagtcgc aggatccttttt-3′
Tight.AdProt.GA.Forward	5′-atgggctccagtgagcag-3′
Tight.AdProt.GA.Reverse	5′-gaattctccaggcgatctg-3′
NotI.TPL.AdProt.Forward	5′-aaaaaagcggccgcactctcttccgcatcg-3′
XbaI.TPL.AdProt.Reverse	5′-aaaaaatctagattacatgtttttcaagtgacaaaaagaag-3′
TPL.AdProt GA.Forward	5′-tggagaaggatccgcactctcttccgcatcgct-3′
TPL.AdProt GA.Reverse	5′-atctagagccggcgcttacatgtttttcaagtgacaaaaagaag-3′
NotI.eGFP.Forward	5′-aaaaaaagcggccgccgccaccatggtgag-3′
EcoRI.eGFP.Reverse	5′-aaaaaagaattccggccgctttacttgtac-3′
NotI.mCherry.Forward	5′-aaaaaagcggccgcgcaccatggtgagcaag-3′
XhoI.mCherry.Reverse	5′-aaaaaactcgagactacttgtacagctcgtccatg-3′
SalI.TTA.Forward	5′-aaaaaagtcgacatgtctagactggacaagagcaaag-3′
BamHI.TTA.Reverse	5′-aaaaaaggatccttacccggggagcatgtcaagg-3′

NotI.TPL.Forward	5′-aaaaaagcggccgcactctcttccgcatcg-3′
XbaI.AdProt.Reverse	5′-aaaaaatctagattacatgtttttcaagtgacaaaaagaag-3′
pENTR1A.AdProt.FLAG.Forward	5′-taatctagacccagctttcttgtacaaagttggcattataag-3′
pENTR1A.AdProt.FLAG.Reverse	5′-agaaagctgggtctagattacttatcgtcgtcatccttgtaatccatgtttttcaagtgacaaaaagaagtggcg-3′
LoxP2Term.GA.Forward	5′-agtcgactggatccggtaccgccgcatcaacgagctc-3′
LoxP2Term.GA.Reverse	5′-gagagtgcggccgcgaattcgaggcccagagggtacc-3′
pENT.AdProt.GA.Forward	5′-gaattcgcggccgcac-3′
pENT.AdProt.GA.Reverse	5′-ggtaccggatccagtcgac-3′
L8.STOP.Forward	5′-cagtgagcaggaatagaaagccattgtcaaagatcttggttgtgg-3′
L8.STOP.Reverse	5′-ctttgacaatggctttctattcctgctcactggagcccattg-3′
E1.kanccdB.Forward	5′-atacaaaactacataagacccccaccttatatattctttcccaccctaaccctcatcagtgccaacatagtaag-3′
E1.kanccdB.Reverse	5′-aataagaggaagtgaaatctgaataattttgtgttactcatagcgcgtaaccgctcattaggcgggc-3′
TetR.kanccdB.Forward	5′-tggaactaatcatatgtggcctggagaaacagctaaagtgcgaaagcggcccgctcattaggcgggc-3′
TetR.kanccdB.Reverse	5′-cgcgaacaaatgtggtatggctgattatgatcctctagagataattctagccctcatcagtgccaacatagtaag-3′
E1.CMV.Promoter.Forward	5′-atacaaaactacataagacccccaccttatatattctttcccaccctaagccacgcccacagatatacgcgttgacattg-3′
E1.bGH.polyA.Reverse	5′-aataagaggaagtgaaatctgaataattttgtgttactcatagcgcgtaatagaagccatagagcccac-3′
E4.kanccdB.Forward	5′-caaaaaacccacaacttcctcaaatcgtcacttccgttttcccacgttacccctcatcagtgccaacatagtaag-3′
E4.kanccdB.Reverse	5′-agtaacttgtatgtgttgggaattgtagttttcttaaaatgggaagtgacccgctcattaggcgggc-3′
E4.SV40.Promoter.Forward	5′-caaaaaacccacaacttcctcaaatcgtcacttccgttttcccacgttacttctgtggaatgtgtgtcagttaggg-3′
E4.SV40.polyA.Reverse	5′-agtaacttgtatgtgttgggaattgtagttttcttaaaatgggaagtgacctctagctagaggtcgacggtatac-3′
Pol.kanccdB.Forward	5′-tcccgcgcttcttggaactttacattgtgggccacaacatcaacggccctccctcatcagtgccaacatagtaag-3′
Pol.kanccdB.Reverse	5′-ggcacctcggaacggttgttaattacctgggcggcgagcacgatctcgtcccgctcattaggcgggc-3′
delPol.Forward	5′-gcgcggccttccggagcgaggtgtgggtgagcgcaaaggtgtccctgaccatgaccagcatgaaggcacgagctgcttcccaaaggcccccatccaag-3′
delPol.Reverse	5′-cttggatgggggcctttgggaagcagctcgtgcccttcatgctggtcatggtcagggacacctttgcgctcacccacacctcgctccggaaggccgcgc-3′
AdProt.kanccdB.Forward	5′-ggcaacgccacaacataaagaagcaagcaacatcaacaacagctgccgccccctcatcagtgccaacatagtaag-3′
AdProt.kanccdB.Reverse	5′-tacaaataaaagcatttgcctttattgaaagtgtctctagtacattatttccgctcattaggcgggc-3′
delAdProt.Forward	5′-ggcaacgccacaacataaagaagcaagcaacatcaacaacagctgccgccaaataatgtactagagacactttcaataaaggcaaatgcttttatttgta-3′
delAdProt.Reverse	5′-tacaaataaaagcatttgcctttattgaaagtgtctctagtacattatttggcggcagctgttgttgatgttgcttgCttctttatgttgtggcgttgcc-3′
HindIII.LeuRS.Forward	5′-aaaaaaaagcttatgcaagagcaataccgccc-3′
XhoI.LeuRS.Reverse	5′-aaaaaactcgagttagccaacgaccagattgaggag-3′
Cre.Forward	5′-tggctagcgtttaaacttaagcttggtacccctccgcggggatcctctaggccaccatgcccaagaagaagaggaag-3′
Cre.Reverse	5′-cgcgaacaaatgtggtatggctgattatgatcctctagagtaattctagctaatcgccatcttccagcagg-3′
BAC2pUC.Forward	5′-cccgggaattcggatctgc-3′
BAC2pUC.Reverse	5′-ccgggaattcggatccttgaagac-3′
Tyr40TAG.Forward	5′-agggcgatgccacctagggcaagctg-3′
Tyr40TAG.Reverse	5′-cagcttgccctaggtggcatcgccct-3′
HindII.eGFP.Forward	5′-aaaaaaaagcttgccaccatggtgagcaagg-3′
XhoI.eGFP.Reverse	5′-aaaaaactcgagttacttgtacagctcgtccatgcc-3′
AdProt.Forward	5′-gggtacccaactccatgctc-3′
AdProt.Reverse	5′-aagtggcgctcctaatctgc-3′
RPLP2.Forward	5′-ccattcagctcactgataaccttg-3′
RPLP2.Reverse	5′-cgtcgcctcctacctgct-3′

Table S5. Plasmid sequence accession numbers.

Vector Name	GenBank Accession Number
Wild-type AdPol vector	MH325099
EP-Pol vector	MH325100
pTRE-Tight.TPL.AdProt	MH325101
CMV.AdProt	MH325102
pLVX.TRE3G.AdProt	MH325103
pLVX.TRE3G.eGFP	MH325104
pBud.tTA.mCherry	MH325105
R6K-kan-ccdB	MH325106
pcDNA3.1-mCherry template plasmid	MH325107
pcDNA3.1-GFP template plasmid	MH325108
pcDNA3.1-tTA template plasmid	MH325109
pcDNA3.1-tTA_{aak} template plasmid	MH325110
pcDNA3.1-KanFDEST template plasmid	MH325111
AdCFP	MH325112
tTA_{wt}.mCherry	MH325113
tTA_{mut}.GFP	MH325114
AdEvolve-DEST	MH325115
ΔAdProtΔAdPol-adenovirus	MH325116
pLeu-tRNA.LeuRS	MH777597
pcDNA3.1-CMV.GFP	MH777595
pLeu-tRNA.GFP(STOP)	MH777596

SUPPORTING REFERENCES

[1] Beronja, S., Livshits, G., Williams, S., and Fuchs, E. (2010) Rapid functional dissection of genetic networks via tissue-specific transduction and RNAi in mouse embryos, *Nat Med 16*, 821-827.
[2] Chatterjee, A., Guo, J., Lee, H. S., and Schultz, P. G. (2013) A genetically encoded fluorescent probe in mammalian cells, *J Am Chem Soc 135*, 12540-12543.
[3] Uil, T. G., Vellinga, J., de Vrij, J., van den Hengel, S. K., Rabelink, M. J., Cramer, S. J., Eekels, J. J., Ariyurek, Y., van Galen, M., and Hoeben, R. C. (2011) Directed adenovirus evolution using engineered mutator viral polymerases, *Nucleic Acids Res 39*, e30.
[4] Wang, H., Bian, X., Xia, L., Ding, X., Muller, R., Zhang, Y., Fu, J., and Stewart, A. F. (2014) Improved seamless mutagenesis by recombineering using ccdB for counterselection, *Nucleic Acids Res 42*, e37.
[5] Suzuki, M., Kondo, S., Pei, Z., Maekawa, A., Saito, I., and Kanegae, Y. (2015) Preferable sites and orientations of transgene inserted in the adenovirus vector genome: The E3 site may be unfavorable for transgene position, *Gene Therapy 22*, 421-429.
[6] Hecht, B., Muller, G., and Hillen, W. (1993) Noninducible Tet repressor mutations map from the operator binding motif to the C terminus, *J. Bacteriol. 175*, 1206-1210.