

MIT Open Access Articles

An Adaptable Platform for Directed Evolution in Human Cells

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Berman, Chet M. et al. "An Adaptable Platform for Directed Evolution in Human Cells." Journal of the American Chemical Society 140 (2019): 18093-18103 © 2019 The Author(s)

As Published: <https://dx.doi.org/10.1021/JACS.8B10937>

Publisher: American Chemical Society (ACS)

Persistent URL: <https://hdl.handle.net/1721.1/125551>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.





Published in final edited form as:

J Am Chem Soc. 2018 December 26; 140(51): 18093–18103. doi:10.1021/jacs.8b10937.

An Adaptable Platform for Directed Evolution in Human Cells

Chet M. Berman^{1,‡,†}, Louis J. Papa III^{1,‡}, Samuel J. Hendel^{1,‡}, Christopher L. Moore¹, Patreece H. Suen¹, Alexander F. Weickhardt¹, Ngoc-Duc Doan, Caiden M. Kumar¹, Taco G. Uil^{2,§}, Vincent L. Butty³, Robert C. Hoeben², and Matthew D. Shoulders^{1,*}

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

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 adenoviral 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.

*Corresponding Author mshoulde@mit.edu.

†Thayer School of Engineering, Dartmouth, Hanover, New Hampshire, USA

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

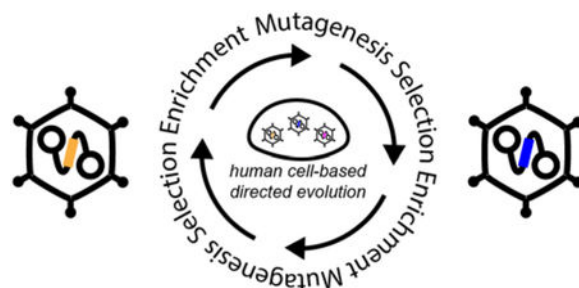
Author Contributions

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.

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.

Graphical Abstract



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 screens (often fluorescent) 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 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 ($\sim 10^5$ variants for the most recent robotic platform).⁹ Another recent development has been the use of cytidine deaminase fused to Cas9 variants to introduce mutations into endogenous genes in human cells, followed by 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 HIV¹⁶ 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 human cells, owing to its genetic tractability and broadly infectious nature in many human cell types.^{18, 19} Further, decades of research have shown that adenovirus tolerates an extremely wide range of transgenes, ensuring broad applicability of an adenovirus-based platform to diverse directed evolution targets. 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 designed, 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 $\sim 1.3 \times 10^{-7}$ 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 $\phi 29$ 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 AdPol adenoviruses (CFP-AdPol.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 AdPol-adenovirus replication.

We next assessed the mutation rate endowed by EP-Pol. After passaging AdPol-adenovirus (AdGL-Pol) 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 27 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 ~ 36 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).

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 insertions were observed for EP-Pol. Compared to the previously reported mutation rate of wild-type AdPol determined by another method,²¹ however, 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 analyses, 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 could conclude once 100% of cells ($\sim 3.0 \times 10^8$ 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 $\sim 4 \times 10^8$ adenoviral variants after one passage. Assuming a typical 1 kb gene encoding the BOI comprises $\sim 1/30$ of the engineered adenoviral genome, there would be $\sim 1.3 \times 10^7$ 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. Regardless, even this conservative estimate indicates that we can generate virtually all single, many double, and some triple mutants in a typical BOI 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, viral maturation, and cell entry.^{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 AdProt AdPol-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 AdProt AdPol-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 AdProt AdPol-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 AdProt AdPol-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 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 using a total 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 order of magnitude range (Figure S5b).

We note that there is likely to be 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 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 strengthen selection

pressure 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 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 initial 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. Co-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 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 likely to be co-infected. We found that even under very high co-infection conditions (late harvest), dox-resistant variants continued to enrich, possibly even more than under low co-infection conditions (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 highlighting 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 (Figures S9a and b).³⁵³⁶ We transfected both the 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 (Figure S9c). 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 essential genes *in trans*, and cannot replicate in unmodified human cells.²²²⁷³⁸ 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 adenovirus¹⁸ 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 variants²² 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.²⁶⁴⁰⁴¹ 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.²²⁴²

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 hypermutation¹¹¹² or Cas9-fusion proteins¹³⁻¹⁵ 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 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 note that one theoretical cheating pathway could be recombination of the AdProt gene from the human cell genome into the adenovirus genome. However, we designed the adenovirus genome to lack any significant homology with the AdProt gene, greatly minimizing the risk of AdProt recombination. We did not observe AdProt re-introduction in any of our experiments.

We used the AdProt-based selection strategy 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. 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,¹⁷⁴⁹ Cas9,⁵⁰ GPCRs,⁵¹ and beyond. Indeed, a context where this platform will prove particularly valuable will be the evolution of complex activities like these, requiring multiple adaptive mutations not readily accessible via traditional transformation of plasmid libraries into human cells.

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 MutaT7⁵² or CRISPR-X,¹⁴ could further focus mutations only to the BOI gene, increase library size, and ensure sustained high mutation rates. 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, 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. Moreover, our system will allow for the continuous evolution of complex proteins that require significant time to reach their active state and therefore may not be possible to evolve using viruses, such as some RNA viruses, that replicate extremely rapidly. Thus, the platform provides a compelling option in any situation where the evolution of optimal BOI variants is unlikely to succeed in simpler systems.

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 described³⁹ 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% CO₂(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, $\sim 9 \times 10^6$ 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:⁵³ 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 $3,200 \times g$ 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 (AdPol adenoviruses on wild-type AdPol cells, AdProt AdPol 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; $\sim 3 \times 10^7$ 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 $3,200 \times g$ 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, AdGL Pol, was subjected to 10 serial passages on cultures of 911 cells⁵⁴ 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,⁵⁵ 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 $\sim 1 \times 10^6$ 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. AdPol.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 AdProt AdPol-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 $0.2 \mu\text{g/mL}$ 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 × 10⁷ 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 × 10⁵ 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 12-well plate of selector cells (Table S2; ~4 × 10⁵ cells/well) was transfected with 1.25 μg of pTet-Off Advanced (Takara Bio) using 7.5 μL of polyethyleneimine (Polysciences) and 100 μL OPTI-MEM. 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 × 10⁵ 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 μ L 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) ($\sim 1 \times 10^7$ cells/plate) were infected with either 250, 500, or 1,000 μ L 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 μ L of adenovirus-containing 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 μ M 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 $\sim 40,000$ cells/well. The next day, 30 μ L 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 μ M or 20 μ M 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 μ L 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 μ L of the harvested DNA for 20 rounds of amplification using 5' - ctacataagacccccaccttatatattctttcc-3' and 5' - agcgggaaaactgaataagaggaagtgaatc-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.⁵⁶⁵⁷ 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 $\sim 4 \times 10^5$ 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)₂Term.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)₂Term.AdProt, and LeuRS.Ad for AdProt(STOP)) or TTA_{wt}.mCherry as a negative control at an MOI of 5. Cells were washed 3 \times 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

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 MS, and Liu DR (2015) Methods for the directed evolution of proteins, *Nat. Rev. Genet* 16, 379–394. [PubMed: 26055155]
- [2]. Gai SA, and Witttrup KD (2007) Yeast surface display for protein engineering and characterization, *Curr. Opin. Struct. Biol* 17, 467–473. [PubMed: 17870469]
- [3]. Romero PA, and Arnold FH (2009) Exploring protein fitness landscapes by directed evolution, *Nat. Rev. Mol. Cell Biol* 10, 866–876. [PubMed: 19935669]
- [4]. Shaner NC, Campbell RE, Steinbach PA, Giepmans BN, Palmer AE, and Tsien RY (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein, *Nat. Biotechnol* 22, 1567–1572. [PubMed: 15558047]
- [5]. Branon TC, Bosch JA, Sanchez AD, Udeshi ND, Svinkina T, Carr SA, Feldman JL, Perrimon N, and Ting AY (2018) Efficient proximity labeling in living cells and organisms with TurboID, *Nat. Biotechnol* 36, 880–887. [PubMed: 30125270]

- [6]. Arzumanyan GA, Gabriel KN, Ravikumar A, Javanpour AA, and Liu CC (2018) Mutually Orthogonal DNA Replication Systems In Vivo, *ACS Synth. Biol* 7, 1722–1729. [PubMed: 29969238]
- [7]. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A, Koonin EV, and Zhang F (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system, *Cell* 163, 759–771. [PubMed: 26422227]
- [8]. Peck SH, Chen I, and Liu DR (2011) Directed evolution of a small-molecule-triggered intein with improved splicing properties in mammalian cells, *Chem. Biol* 18, 619–630. [PubMed: 21609843]
- [9]. Piatkevich KD, Jung EE, Straub C, Linghu C, Park D, Suk HJ, Hochbaum DR, Goodwin D, Pneumatikakis E, Pak N, Kawashima T, Yang CT, Rhoades JL, Shemesh O, Asano S, Yoon YG, Freifeld L, Saulnier JL, Riegler C, Engert F, Hughes T, Drobizhev M, Szabo B, Ahrens MB, Flavell SW, Sabatini BL, and Boyden ES (2018) A robotic multidimensional directed evolution approach applied to fluorescent voltage reporters, *Nat. Chem. Biol* 14, 352–360. [PubMed: 29483642]
- [10]. Banaszynski LA, Chen LC, Maynard-Smith LA, Ooi AG, and Wandless TJ (2006) A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules, *Cell* 126, 995–1004. [PubMed: 16959577]
- [11]. Wang CL, Yang DC, and Wabl M (2004) Directed molecular evolution by somatic hypermutation, *Protein Eng. Des. Sel* 17, 659–664. [PubMed: 15516361]
- [12]. Wang L, Jackson WC, Steinbach PA, and Tsien RY (2004) Evolution of new nonantibody proteins via iterative somatic hypermutation, *Proc. Natl. Acad. Sci. U.S.A* 101, 16745–16749. [PubMed: 15556995]
- [13]. Ma Y, Zhang J, Yin W, Zhang Z, Song Y, and Chang X (2016) Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells, *Nat. Methods* 13, 1029–1035. [PubMed: 27723754]
- [14]. Hess GT, Fresard L, Han K, Lee CH, Li A, Cimprich KA, Montgomery SB, and Bassik MC (2016) Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells, *Nat. Methods* 13, 1036–1042. [PubMed: 27798611]
- [15]. Komor AC, Kim YB, Packer MS, Zuris JA, and Liu DR (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage, *Nature* 533, 420–424. [PubMed: 27096365]
- [16]. Das AT, Zhou X, Vink M, Klaver B, Verhoef K, Marzio G, and Berkhout B (2004) Viral evolution as a tool to improve the tetracycline-regulated gene expression system, *J. Biol. Chem* 279, 18776–18782. [PubMed: 14761948]
- [17]. Esvelt KM, Carlson JC, and Liu DR (2011) A system for the continuous directed evolution of biomolecules, *Nature* 472, 499–503. [PubMed: 21478873]
- [18]. Lucher LA (1995) Abortive adenovirus infection and host range determinants, *Curr. Top. Microbiol. Immunol* 199 (Pt 1), 119–152. [PubMed: 7555052]
- [19]. Amalfitano A, and Chamberlain JS (1997) Isolation and characterization of packaging cell lines that coexpress the adenovirus E1, DNA polymerase, and preterminal proteins: implications for gene therapy, *Gene Ther.* 4, 258–263. [PubMed: 9135740]
- [20]. Hoeben RC, and Uil TG (2013) Adenovirus DNA replication, *Cold Spring Harb. Perspect. Biol* 5, a013003. [PubMed: 23388625]
- [21]. Risso-Ballester J, Cuevas JM, and Sanjuán R (2016) Genome-wide estimation of the spontaneous mutation rate of human adenovirus 5 by high-fidelity deep sequencing, *PLOS Pathog.* 12, e1006013. [PubMed: 27824949]
- [22]. Uil TG, Vellinga J, de Vrij J, van den Hengel SK, Rabelink MJ, Cramer SJ, Eekels JJ, Ariyurek Y, van Galen M, and Hoeben RC (2011) Directed adenovirus evolution using engineered mutator viral polymerases, *Nucleic Acids Res.* 39, e30. [PubMed: 21138963]
- [23]. Kamtekar S, Berman AJ, Wang J, Lazaro JM, de Vega M, Blanco L, Salas M, and Steitz TA (2004) Insights into strand displacement and processivity from the crystal structure of the protein-primed DNA polymerase of bacteriophage phi29, *Mol. Cell* 16, 609–618. [PubMed: 15546620]

- [24]. Sanjuán R, Nebot MR, Chirico N, Mansky LM, and Belshaw R (2010) Viral mutation rates, *J. Virol* 84, 9733–9748. [PubMed: 20660197]
- [25]. Davis JN, and van den Pol AN (2010) Viral mutagenesis as a means for generating novel proteins, *J. Virol* 84, 1625–1630. [PubMed: 19906913]
- [26]. Phillips AM, Gonzalez LO, Nekongo EE, Ponomarenko AI, McHugh SM, Butty VL, Levine SS, Lin YS, Mirny LA, and Shoulders MD (2017) Host proteostasis modulates influenza evolution, *eLife* 6, e28652. [PubMed: 28949290]
- [27]. Elahi SM, Oualikene W, Naghdi L, O'Connor-McCourt M, and Massie B (2002) Adenovirus-based libraries: efficient generation of recombinant adenoviruses by positive selection with the adenovirus protease, *Gene Ther.* 9, 1238–1246. [PubMed: 12215891]
- [28]. Greber UF, Webster P, Weber J, and Helenius A (1996) The role of the adenovirus protease in virus entry into cells, *EMBO J.* 15, 1766–1777. [PubMed: 8617221]
- [29]. Webster A, Leith IR, and Hay RT (1994) Activation of adenovirus-coded protease and processing of preterminal protein, *J. Virol* 68, 7292–7300. [PubMed: 7933113]
- [30]. Gossen M, and Bujard H (1992) Tight Control of gene expression in mammalian cells by tetracycline responsive promoters, *Proc. Natl. Acad. Sci. U.S.A* 89, 5547–5551. [PubMed: 1319065]
- [31]. Loew R, Heinz N, Hampf M, Bujard H, and Gossen M (2010) Improved Tet-responsive promoters with minimized background expression, *BMC Biotechnol.* 10, 81. [PubMed: 21106052]
- [32]. Krueger M, Scholz O, Wisshak S, and Hillen W (2007) Engineered Tet repressors with recognition specificity for the tetO-4C5G operator variant, *Gene* 404, 93–100. [PubMed: 17928170]
- [33]. Grosche P, Sirockin F, Mac Sweeney A, Ramage P, Erbel P, Melkko S, Bernardi A, Hughes N, Ellis D, Combrink KD, Jarousse N, and Altmann E (2015) Structure-based design and optimization of potent inhibitors of the adenoviral protease, *Bioorg. Med. Chem. Lett* 25, 438–443. [PubMed: 25571794]
- [34]. 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. [PubMed: 8432715]
- [35]. Meinke G, Bohm A, Hauber J, Pisabarro MT, and Buchholz F (2016) Cre recombinase and other tyrosine recombinases, *Chem. Rev* 116, 12785–12820. [PubMed: 27163859]
- [36]. Italia JS, Zheng Y, Kelemen RE, Erickson SB, Addy PS, and Chatterjee A (2017) Expanding the genetic code of mammalian cells, *Biochem. Soc. Trans* 45, 555–562. [PubMed: 28408495]
- [37]. Hartley JL, Temple GF, and Brasch MA (2000) DNA cloning using in vitro site-specific recombination, *Genome Res.* 10, 1788–1795. [PubMed: 11076863]
- [38]. Russell WC (2000) Update on adenovirus and its vectors, *J. Gen. Virol* 81, 2573–2604. [PubMed: 11038369]
- [39]. Wang H, Bian X, Xia L, Ding X, Muller R, Zhang Y, Fu J, and Stewart AF (2014) Improved seamless mutagenesis by recombineering using ccdB for counterselection, *Nucleic Acids Res.* 42, e37. [PubMed: 24369425]
- [40]. Meyerhans A, Cheynier R, Albert J, Seth M, Kwok S, Sninsky J, Morfeldt-Månson L, Asjö B, and Wain-Hobson S (1989) Temporal fluctuations in HIV quasispecies in vivo are not reflected by sequential HIV isolations, *Cell* 58, 901–910. [PubMed: 2550139]
- [41]. O'Loughlin TL, Greene DN, and Matsumura I (2006) Diversification and specialization of HIV protease function during in vitro evolution, *Mol. Biol. Evol* 23, 764–772. [PubMed: 16423863]
- [42]. Myers ND, Skorohodova KV, Gounder AP, and Smith JG (2013) Directed evolution of mutator adenoviruses resistant to antibody neutralization, *J. Virol* 87, 6047–6050. [PubMed: 23487468]
- [43]. Meng FL, Du Z, Federation A, Hu J, Wang Q, Kieffer-Kwon KR, Meyers RM, Amor C, Wasserman CR, Neuberg D, Casellas R, Nussenzweig MC, Bradner JE, Liu XS, and Alt FW (2014) Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability, *Cell* 159, 1538–1548. [PubMed: 25483776]
- [44]. Kim D, Lim K, Kim ST, Yoon SH, Kim K, Ryu SM, and Kim JS (2017) Genome-wide target specificities of CRISPR RNA-guided programmable deaminases, *Nat. Biotechnol* 35, 475–480. [PubMed: 28398345]

- [45]. Wang CL, Harper RA, and Wabl M (2004) Genome-wide somatic hypermutation, *Proc. Natl. Acad. Sci. U.S.A* 101, 7352–7356. [PubMed: 15123833]
- [46]. Hubbard BP, Badran AH, Zuris JA, Guilinger JP, Davis KM, Chen L, Tsai SQ, Sander JD, Joung JK, and Liu DR (2015) Continuous directed evolution of DNA-binding proteins to improve TALEN specificity, *Nat. Methods* 12, 939–942. [PubMed: 26258293]
- [47]. Dickinson BC, Packer MS, Badran AH, and Liu DR (2014) A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations, *Nat. Commun* 5, 5352. [PubMed: 25355134]
- [48]. Badran AH, Guzov VM, Huai Q, Kemp MM, Vishwanath P, Kain W, Nance AM, Evdokimov A, Moshiri F, Turner KH, Wang P, Malvar T, and Liu DR (2016) Continuous evolution of *Bacillus thuringiensis* toxins overcomes insect resistance, *Nature* 533, 58–63. [PubMed: 27120167]
- [49]. Carlson JC, Badran AH, Guggiana-Nilo DA, and Liu DR (2014) Negative selection and stringency modulation in phage-assisted continuous evolution, *Nat. Chem. Biol* 10, 216–222. [PubMed: 24487694]
- [50]. Hu JH, Miller SM, Geurts MH, Tang W, Chen L, Sun N, Zeina CM, Gao X, Rees HA, Lin Z, and Liu DR (2018) Evolved Cas9 variants with broad PAM compatibility and high DNA specificity, *Nature* 556, 57–63. [PubMed: 29512652]
- [51]. Armbruster BN, Li X, Pausch MH, Herlitze S, and Roth BL (2007) Evolving the lock to fit the key to create a family of G protein-coupled receptors potently activated by an inert ligand, *Proc. Natl. Acad. Sci. U.S.A* 104, 5163–5168. [PubMed: 17360345]
- [52]. Moore CL, Papa LJ 3rd, and Shoulders MD (2018) A processive protein chimera introduces mutations across defined DNA regions in vivo, *J. Am. Chem. Soc* 140, 11560–11564. [PubMed: 29991261]
- [53]. Dull T, Zufferey R, Kelly M, Mandel RJ, Nguyen M, Trono D, and Naldini L (1998) A third-generation lentivirus vector with a conditional packaging system, *J. Virol* 72, 8463–8471. [PubMed: 9765382]
- [54]. Fallaux FJ, Kranenburg O, Cramer SJ, Houweling A, Van Ormondt H, Hoeben RC, and Van Der Eb AJ (1996) 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. [PubMed: 8788172]
- [55]. Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, and Weigel D (2008) Sequencing of natural strains of *Arabidopsis thaliana* with short reads, *Genome Res.* 18, 2024–2033. [PubMed: 18818371]
- [56]. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, and 1000 Genome Project Data Processing Subgroup. (2009) The Sequence Alignment/Map format and SAMtools, *Bioinformatics* 25, 2078–2079. [PubMed: 19505943]
- [57]. Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data, *Bioinformatics* 27, 2987–2993. [PubMed: 21903627]

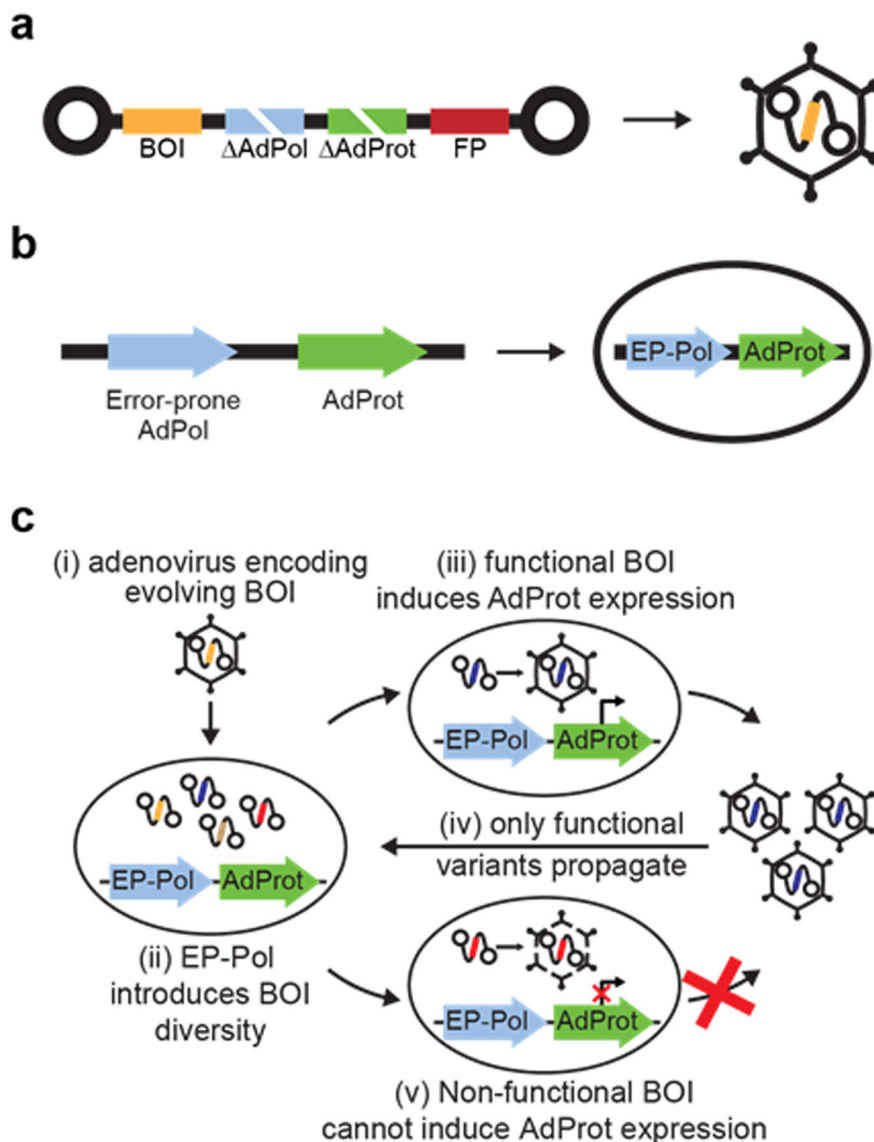


Figure 1.

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.

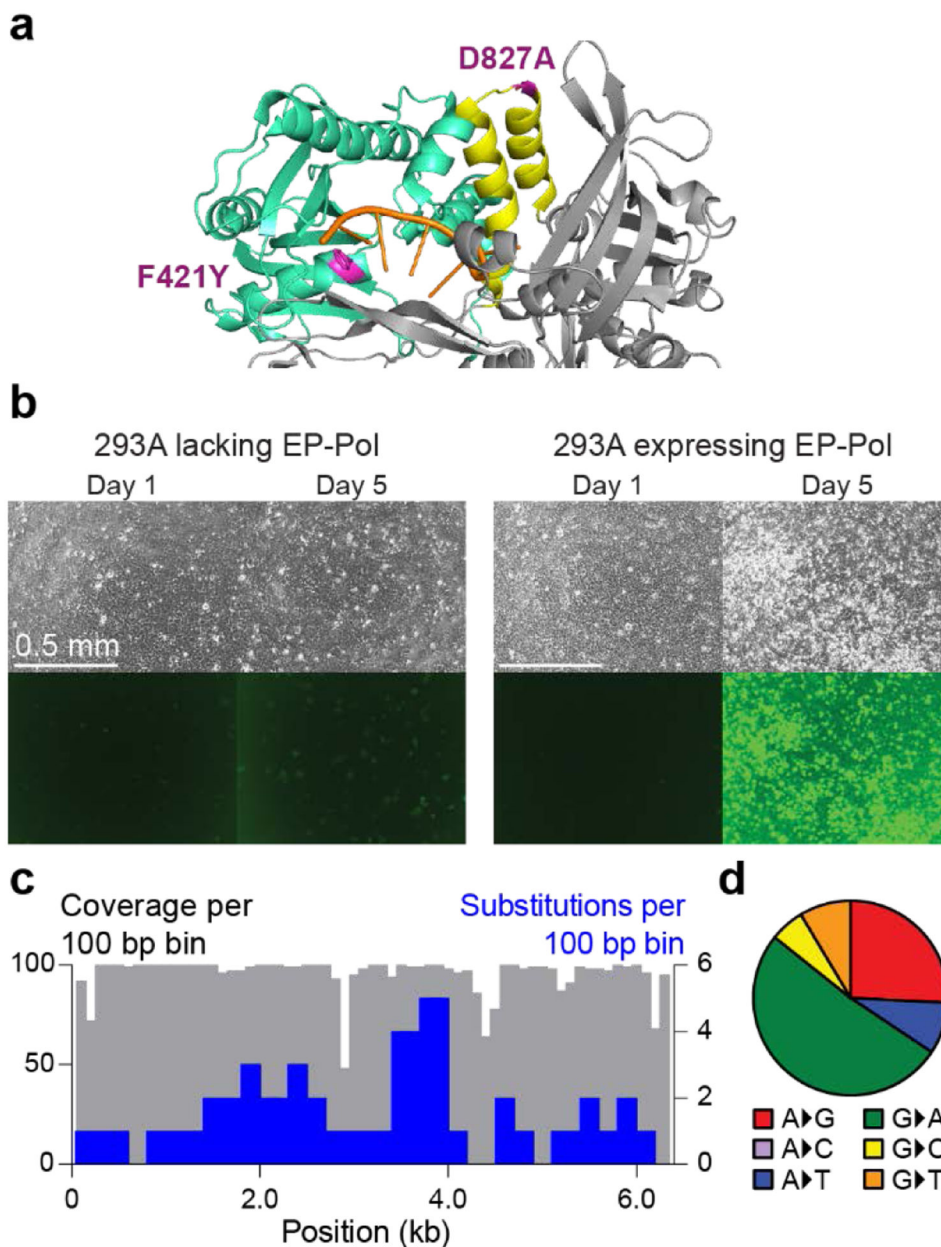
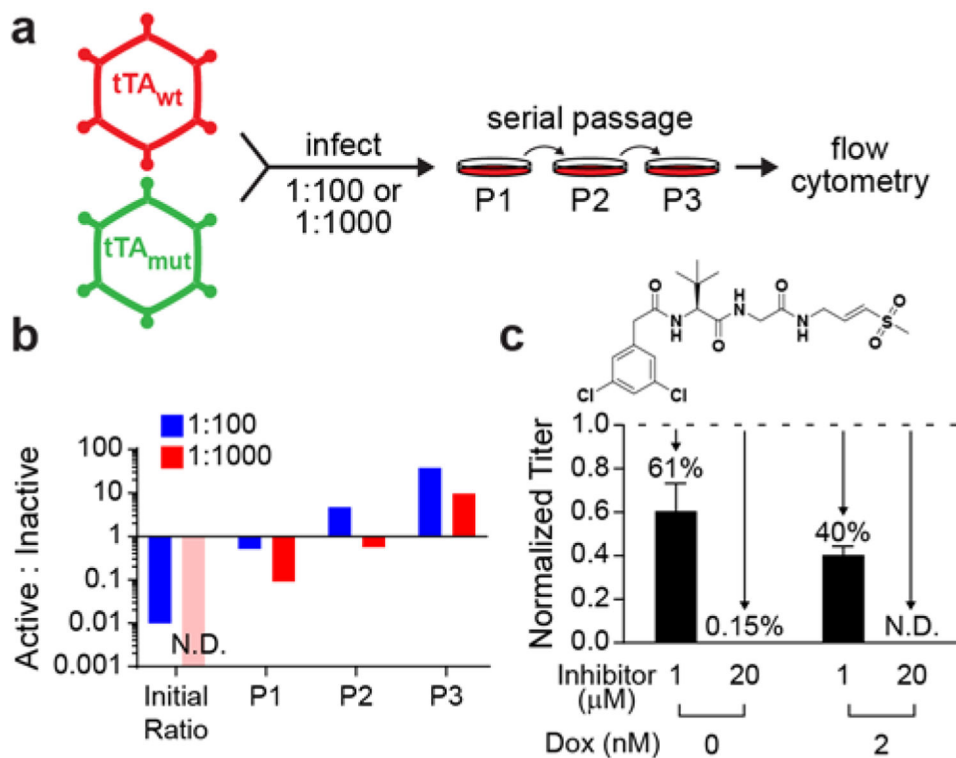
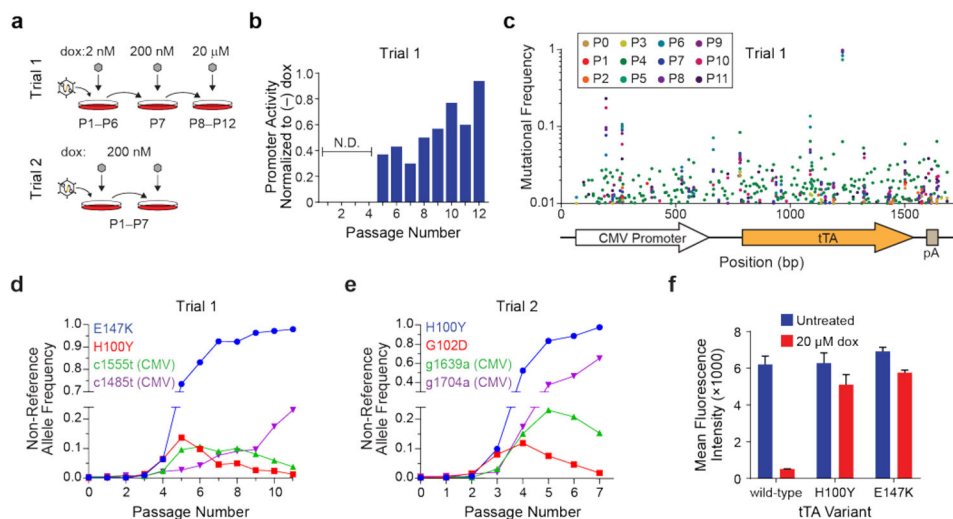


Figure 2. (a) Crystal structure of the ϕ 29 DNA polymerase (PDBID 1XHZ),²³ an AdPol homolog, with the locations of homologous mutations used to create EP-Pol shown in magenta. (b) Either parental HEK293A cells or cells constitutively expressing EP-Pol were infected with a GFP-encoding AdPol-adenovirus (CFP AdPol.GFP). The virus propagated only on EP-Pol trans-complementing cells. Similar results were obtained for wild-type AdPol (Figure S1). (c) AdPol-adenovirus (AdGL Pol) was serially passaged on EP-Pol expressing cells for ten passages, after which a 6.5 kb genomic fragment was amplified from an ~27 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.

**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).

**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.

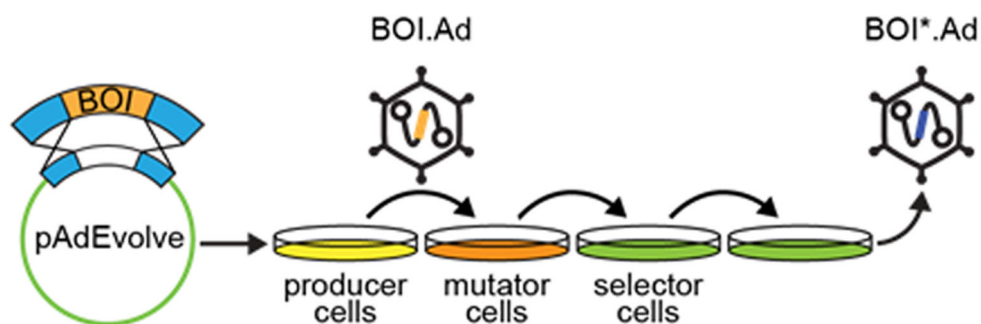


Figure 5.

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 AdProt AdPol-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.