Continuous Directed Evolution in Mammalian Cells

by

Samuel Joseph Hendel

B.A., Amherst College (2015)

Submitted to the Department of Chemistry in Partial Fulfillment of the Requirements for the

Degree of Doctor of Science in Chemistry

at the

Massachusetts Institute of Technology

May 2022

Signature of the author:

_____

Department of Chemistry

April 19, 2022

Certified by:

_____

Matthew Shoulders

Associate Professor of Chemistry

Thesis Supervisor

Accepted by:

_____

Adam Willard

Associate Professor

Graduate Officer

This doctoral thesis has been examined by a committee of the Department of Chemistry as follows:

_____

Alex K. Shalek

Associate Professor

Department of Chemistry, MIT

Thesis Committee Chair

_____

Matthew D. Shoulders

Associate Professor

Department of Chemistry, MIT

Thesis Supervisor

_____

Laura L. Kiessling

Novartis Professor of Chemistry

Department of Chemistry, MIT

Thesis Committee Member

Continuous Directed Evolution in Mammalian Cells

by

Samuel Joseph Hendel

Submitted to the Department of Chemistry in Partial Fulfillment of the Requirements for the
Degree of Doctor of Science in Chemistry

Abstract

Directed evolution is a powerful methodology for the creation of new biomolecules with user-desired functions. Most directed evolution experiments are performed *in vitro*, in bacteria, or in yeast, even when the evolved biomolecule is intended to function in mammalian cells. As a result, the functions of biomolecules evolved in these environments are often derailed in the complex mammalian cellular environment. The development of highly efficacious methods for directed evolution in mammalian cells has severely lagged behind similar methods in single-celled organisms, owing to the relative difficulties of both mammalian cell culture and genomic engineering in mammalian cells. In this thesis, I describe the development and subsequent application of a high-throughput, adaptable, virus-based continuous directed evolution method that uses the mammalian cell for simultaneously mutagenizing, expressing, and selecting an evolving gene of interest. This platform functions by making adenoviral propagation in mammalian cells dependent upon the activity of a virally encoded gene of interest, which is continuously mutagenized by a highly error-prone engineered adenoviral polymerase. We demonstrated the platform's efficacy in proof-of-principle evolution experiments by evolving a transcription factor to be insensitive to a small molecule inhibitor. We then engineered selection circuits for evolving endogenous human G-protein coupled receptors, wherein viral replication is coupled to an endogenous signaling pathway. We also engineered selection circuits for evolving exogenous CRISPR systems, wherein viral replication is coupled to an exogenous transcriptional couple. For both selection circuits, we demonstrated selection pressure sufficient to drive a directed evolution campaign through viral replication assays. Finally, we highlight a wide range of biomolecules for which directed evolution mammalian cells would be impactful, but was previously out of reach before the development of virus-based continuous evolution methods.

Thesis Supervisor: Matthew D. Shoulders
Title: Associate Professor of Chemistry

Continuous Directed Evolution in Mammalian Cells

"It is those who know little, and not those who know much, who so positively assert that this or that problem will never be solved by science." – Charles Darwin, *The Descent of Man*

"That is the essence of science: ask an impertinent question, and you are on the way to the pertinent answer." – Jacob Bronowski, *The Ascent of Man*

Table of contents

List of figures

List of tables

Abbreviations

aaRS	aminoacyl tRNA synthetase
AAV	adeno-associated virus
Acr	anti-CRISPR protein
Ad	adenovirus
ADCY	adenylyl cyclase
AdPol	adenoviral polymerase
AdProt	adenoviral protease
ADRB3	β3-adrenoceptor
AID	activation-induced cytidine deaminase
AMP	adenosine monophosphate
Arch	archrhodopsin
ASAP	accelerated sensor of action potentials
ASCL1	Achaete-scute homolog 1
ATP	adenosine triphosphate
AVPR2	vasopressin V2 receptor
BCR-ABL	breakpoint cluster region-Abelson fusion protein
BOI	biomolecule of interest
CALCRb	calcitonin receptor-like receptor
cAMP	cyclic AMP
Cas9	CRISPR associated protein 9
cDNA	complementary DNA
CFP	cerulean fluorescent protein
CHRM3	muscarinic acetylcholine receptor M3
CMV	cytomegalovirus promoter
CNO	clozapine-*N*-oxide
CPE	cytopathic effect
CRE	cAMP response element
CREB	CRE binding protein
CRISPR	clustered regularly interspaced short palindromic repeats
CRISPRa	CRISPR activation
CSD	C-terminal superdegron
DMEM	Delbecco's modified Eagle medium
DNA	deoxyribonucleic acid
dox	doxycycline
DREADD	designer receptor exclusively activated by designer drugs
E1	adenoviral early region 1
E3	adenoviral early region 3
EP-Pol	error-prone adenoviral polymerase
ER	endoplasmic reticulum
FBS	fetal bovine serum
FDA	Food and Drug Administration
FKBP	FK506 binding protein
FP	fluorescent protein

FSHR	follicle stimulating hormone receptor
G protein	guanine nucleotide-binding protein
GCGR	glucagon receptor
GDP	guanosine diphosphate
GFP	green fluorescent protein
GOI	gene of interest
GPCR	G protein-coupled receptor
gRNA	guide RNA
GTP	guanosine triphosphate
H^+	hydrogen ion
HEK	human embryonic kidney cells
HIV	human immunodeficiency virus
hM_3R	human M3 receptor
HTR6	5-hydroxytryptamine receptor 6
kb	kilobase pairs
Lenti	lentivirus
MC1R	melanocortin 1 receptor
MC2R	melanocortin 2 receptor
MC3R	melanocortin 3 receptor
MEK	mitogen-activated protein kinase kinase
MOI	multiplicity of infection
MP6	mutagenesis plasmid 6
mPACE	mammalian PACE
mRNA	messenger RNA
NGS	next-generation sequencing
nsP4	nonstructural protein 4
p65	TAD of NF-κB
PACE	phage-assisted continuous evolution
PAM	protospacer adjacent motif
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	Protein Data Bank
PGK	3-phosphoglycerate kinase
pH	potential of hydrogen
PKA	protein kinase A
pTP	pre-terminal protein
qPCR	quantitative PCR
RASSL	receptor activated solely by synthetic ligands
REV	HIV Rev protein
RNA	ribonucleic acid
RNA Pol II	RNA polymerase II
RRE	Rev response element
Rta	Epstein-Barr virus R transactivator
sgRNA	single guide RNA

SLF*   synthetic ligand for FKBP
SSG   Sindbis structural genome
T7   T7 bacteriophage RNA polymerase
TAD   transcriptional activation domain
TAM   targeted AID-mediated mutagenesis
TRACE   T7 polymerase-driven continuous editing
TRE   tetracycline response element
TREC   therapeutic receptor-effector complex
tRNA   transfer RNA
tTA   tetracycline transactivator
TTN   titin
UTR   untranslated region
VEGAS   viral evolution of genetically actuating sequences
VPR   VP64–p65 TAD–Rta TAD fusion
VSVG   vesicular stomatitis virus glycoprotein
Xfect   transfected
YFP   yellow fluorescent protein

Chapter 1: Methods for directed evolution in mammalian cells

1.1 Summary

Directed evolution experiments are typically carried out using *in vitro* systems, bacteria, or yeast—even when the goal is to probe or modulate mammalian biology. Performing directed evolution in systems that do not match the intended mammalian environment severely constrains the scope and functionality of targets that can be evolved. This chapter reviews recent developments that enable researchers to use the mammalian cell itself as the setting for directed evolution, and presents an overview of frontier challenges and high-impact targets for this approach.

1.2 Introduction

Directed evolution is a powerful methodology for creating biomolecules with new and improved properties. In a typical directed evolution experiment, a library of genetic variants is generated through mutagenesis of an initial sequence. The expressed biomolecules are then selected or screened for a desired activity. Iterative cycles of this process produce increasingly optimized biomolecules. This approach has yielded molecules useful to industry and research and numerous critical biotechnologies, and was the subject of the 2018 Nobel Prize in Chemistry.[1] Aside from the many practical applications, valuable fundamental information about biomolecular structure, evolutionary biology, and organismal fitness can often be obtained using directed evolution.

Directed evolution can be used to create biomolecular tools to perturb or interrogate many mammalian systems, with both fundamental research and therapeutic applications. Directed evolution could even optimize for function in a specific cell type or genetic background, providing a platform for developing targeted or personalized medicine. From a basic science perspective, researchers could gain new insights into the structure and function of understudied mammalian proteins or those with unknown structures, as well as explore principles of mammalian evolutionary biology.

Yet, despite all this potential, the promised impact of directed evolution on mammalian biology and medicine has not yet been fully realized. Instead, many impactful targets and pathways have never been pursued using directed evolution, and many attempted directed evolution campaigns have yielded products that fail to function properly in mammalian cells. Published examples of the latter include monobodies,[2] fluorescent proteins,[3] inteins,[4-5] aminoacyl-tRNA synthetases,[6] and membrane proteins.[7]
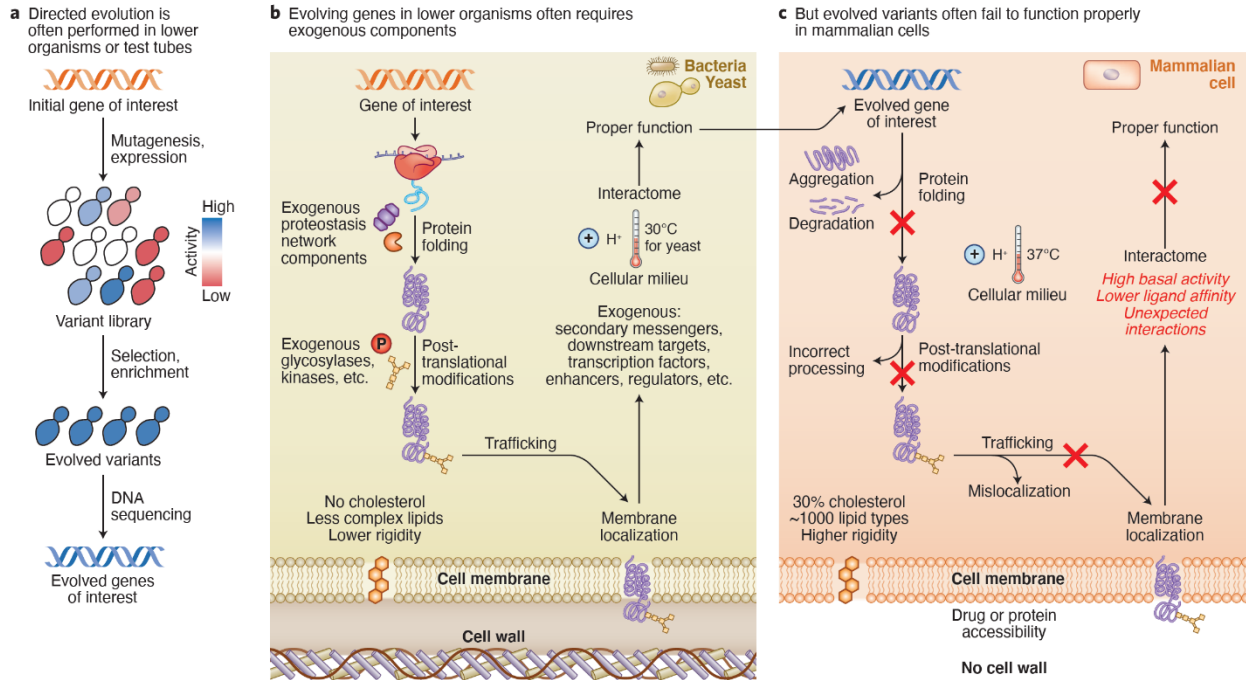
Figure 1.1 Directed evolution for function in mammalian cells. (a) In most directed evolution experiments, protein variants are expressed in test tubes, bacteria, or yeast. Genes of interest are mutagenized and expressed, creating a mutant library with varying activities. Successful variants (blue) are selected and enriched in the population more than unsuccessful variants (red). Evolved variants are then identified through DNA sequencing. (b–c) Illustration of some of the challenges associated with performing directed evolution in lower organisms—in this example, directed evolution of a mammalian membrane protein. (b) To evolve mammalian proteins in lower organisms, exogenous components from mammalian cells frequently need to be introduced for proper folding, modification, trafficking, and more. Membrane proteins also often require extensive interactions with other downstream exogenous components for proper function, requiring additional exogenous components. (c) Proteins evolved *in vitro*, in bacteria, or in yeast often fail to function properly once expressed in mammalian cells. Function can be derailed in mammalian cells because the evolved protein may aggregate, be degraded, be incorrectly processed or modified, mislocalize, interact with inappropriate partners, or lack catalytic or other activity for any number of additional reasons.

What is responsible for these failures? Directed evolution experiments are most commonly performed *in vitro*, in bacteria, or in yeast (**Figure 1.1a**), even when the product is intended to be used in mammalian cells or when the goal is to study mammalian biology.[8] Unfortunately, most mammalian proteins do not retain their function when expressed in lower organisms, including in single-celled eukaryotes like yeast,[9] commonly owing to misfolding or aggregation, the absence of complex signaling pathways unique to mammalian cells, or other differences (**Figure 1.1b**). Even when these obstacles can be overcome, the lower organism is still a highly artificial selection system. As a consequence, the desired functions of evolved products can be derailed in mammalian cells by unintended intermolecular interactions, poor folding, unexpected modifications or cellular localization, and many other serious problems (**Figure 1.1c**). Importantly, this limitation applies not only to mammalian proteins themselves but also to those derived from other sources, including bacteria and yeast, that could be evolved to perturb or probe mammalian biology.

The apparent solution to all these issues is to use the mammalian cell itself as the design, engineering, and quality control setting for directed evolution. Although simple on its face, this solution immediately runs up against numerous obstacles. Many practical reasons, both technical (e.g., slow growth rate of mammalian cells, engineering challenges) and economic (e.g., labor-intensive mammalian cell culture, expensive media), have deterred researchers from using mammalian cells in directed evolution campaigns. These challenges, while real, have led the directed evolution field down what can prove to be an unproductive path—attempting to evolve function in poorly matched model systems.

Recent technological advances have comprehensively reshaped the mammalian cell directed evolution landscape. New mammalian cell-specific platforms, including bespoke strategies for library creation and variant selection, are democratizing mammalian cell-based directed evolution. In this chapter, I present these emerging methods and some frontier challenges in the field. After first discussing key elements and challenges of mammalian cell-based directed evolution campaigns, I focus attention on *ex mammalia* and *in mammalia* mutagenesis methods. Finally, I introduce virus-aided continuous evolution methods that integrate *in mammalia* mutagenesis with efficient expression and selection, a concept which will be explored more deeply in the remaining chapters.
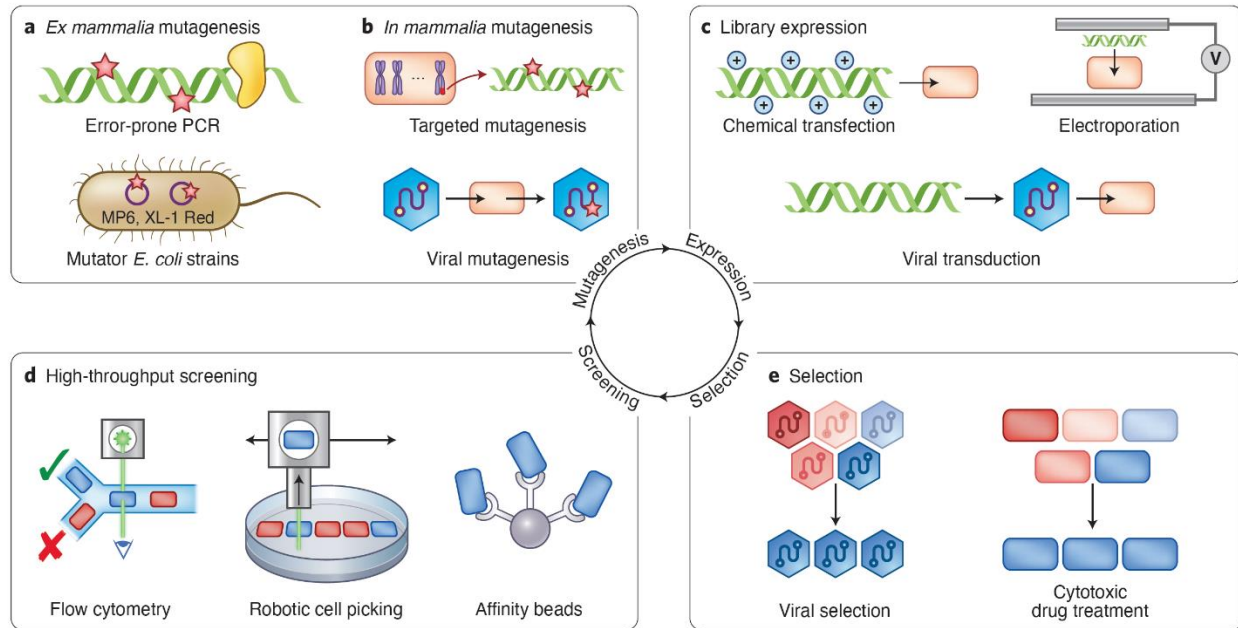
Figure 1.2 Directed evolution in mammalian cells. (a–b) First, genes of interest are mutagenized. (a) *Ex mammalia* mutagenesis strategies, such as error-prone PCR[10] or mutator strains of *Escherichia coli* like XL-1 Red[11] and bacteria transformed with the mutagenic MP6 plasmid,[12] mutate genes of interest outside of the mammalian cell. (b) *In mammalia* mutagenesis strategies, such as targeted mutagenesis or viral mutagenesis, mutate genes directly in the mammalian cell. (c) Once genes of interest are mutagenized, the variant library must be expressed in mammalian cells, commonly through chemical transfection, electroporation, or viral transduction. (d–e) Cells must then be assessed for appropriate activity, either by screening or selection, and genes encoding active variants enriched in the population. (d) High-throughput screening methods include flow cytometry, cellular imaging followed by robotic cell picking,[30] and cellular binding to affinity beads. (e) Selection methods may include cytotoxic drug treatment and viral selection.

1.3 General principles and challenges of mammalian cell-based directed evolution

Individual cycles of a directed evolution experiment have three main components: mutagenesis, expression, and screening or selection (**Figure 1.2**). For each component, mammalian cells present unique challenges.

1.3.1 Mutagenesis. The first step in each cycle of a directed evolution experiment is mutagenesis, in which a large library of genetic variants of a gene of interest is created. Global mutagenesis approaches introduce off-target mutations throughout the genome. Considering the immense size and complexity of the mammalian genome, there is a high probability that such off-target mutations can lead to the creation of 'cheaters' that bypass screening or selection in unproductive ways, completely derailing directed evolution experiments.

Mammalian directed evolution strategies therefore rely upon targeted mutagenesis, meaning mutagenesis directed to a specific genetic region. Targeted mutagenesis methods can be categorized as either *ex mammalia* or *in mammalia*. *Ex mammalia* mutagenesis, or mutagenesis that occurs outside of mammalian cells, includes random techniques, such as error-prone PCR[10] and mutator strains of bacteria,[11-12] as well as non-random techniques, such as site-directed mutagenesis and synthetic library generation (**Figure 1.2a**). Non-random mutagenesis provides precise control over which genetic variants are created, but these libraries are typically much smaller in size than those produced using random methods.[13] Therefore, most *ex mammalia* mutagenesis strategies rely on random mutagenesis.

In mammalia mutagenesis, or mutagenesis that occurs inside mammalian cells, currently relies on somatic hypermutation,[14] CRISPR-based DNA targeting,[15] highly processive RNA polymerases,[16-17] or viral replication (**Figure 1.2b**). As opposed to *ex mammalia* mutagenesis, *in mammalia* mutagenesis enables multiple rounds of mutagenesis without needing to excise genetic libraries from the mammalian cell population to allow for further rounds of mutagenesis. Many of these strategies currently have a limited mutational spectrum, owing to their reliance on DNA-damaging enzymes to drive mutagenesis. Non-random *in mammalia* mutagenesis methods, such as Cas9-induced homology-directed repair using designed single-stranded DNA repair templates, can promote the introduction of a greater spectrum of point mutations[18-19] but generate smaller libraries.

1.3.2 Expression. Once a mutagenized genetic library is generated, the protein variants must be expressed either transiently or stably within mammalian cells. In transient expression methods, including chemical transfection and electroporation, cells express protein variants without genomically encoding or replicating the variant's genetic information (**Figure 1.2c**). Challenges associated with transient expression include transfection efficiency, inconsistent

expression levels, expression of multiple variants in a single cell, and compatibility with a user-desired cell line. Further, if the genes encoding transiently expressed biomolecule variants are not somehow replicated *in mammalia* over multiple generations, they cannot be mutagenized *in mammalia*, hampering iterability. If, however, viruses are used to encode, propagate, and transiently express genetic variants, then *in mammalia* mutagenesis can be used alongside transient expression.

Stable expression, including retroviral transduction, can address many of the problems inherent to transient expression. Stably encoded genes can be mutated *in mammalia*, and protein expression levels may be more consistent. Nonetheless, cells can be infected by multiple retroviruses, and some commonly used retroviruses such as lentivirus are packaged with two genomes.

1.3.3 Selection or screening. Once a library of protein variants is expressed in mammalian cells, the expressed variants must somehow be assessed for appropriate activity and then enriched in the population. This step can be categorized as either screening or selection, and the details depend entirely on the particular biomolecular activity being evolved. Screening involves physically separating cells on the basis of a given phenotype, commonly fluorescence-based flow cytometry or cellular display (**Figure 1.2d**). Screening-based methods can be quite labor-intensive since cells usually must be manually manipulated before, during, and after each round. Another challenge is determining appropriate activity cutoffs for each round, which must be modified to become more stringent over the course of a directed evolution experiment.

Selection-based strategies usually rely on the survival and replication of cells or viruses to enrich genetic variants (**Figure 1.2e**). Commonly, selection involves treating cells that encode a diverse library of genetic variants with a cytotoxic drug. Only those cells encoding genetic variants that provide resistance to the cytotoxic drug can survive selection to replicate and enrich in the population. Selection is in many ways much preferable to screening, as it requires minimal researcher intervention, is easy to scale and iterate, and requires little specialized equipment. Nonetheless, the slow growth rate of mammalian cells, the relatively narrow scope of biomolecules that negate the impact of cytotoxic perturbations, and the tendency for cells to spontaneously attain resistance to many such perturbations via undesired 'cheating' pathways, mean that selection-based approaches have historically been mostly limited to studies on oncoproteins.[20-24] Recently developed virus-based selection methods[25-26], described below, overcome many of the limitations of these past cell-based methods, leading to a vastly expanded scope of biomolecules evolvable through selection in mammalian cells.
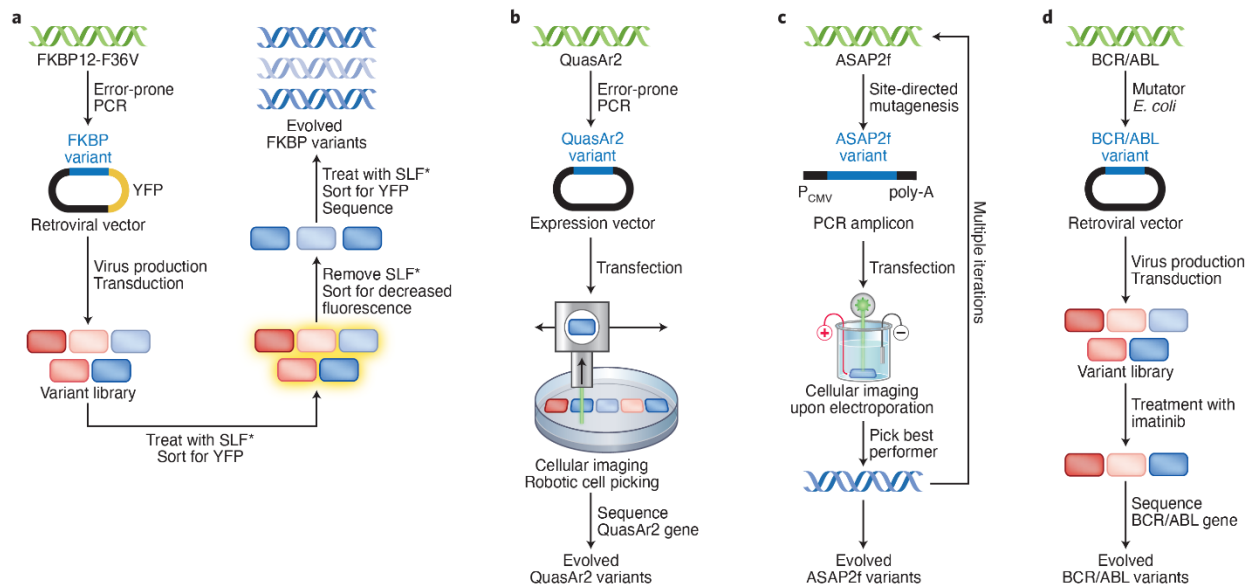
Figure 1.3 Selected examples of directed evolution experiments that use ex mammalia mutagenesis. (a) Directed evolution of a small-molecule regulated destabilizing domain by Wandless and co-workers.[27] (b) Robotic cell sreening-based directed evolution of a voltage reporter by Boyden and co-workers.[30] (c) Iterative directed evolution of a voltage reporter by Dieudonné, Lin, and co-workers.[13] (d) Directed evolution of BCR/ABL by Daley and co-workers.[21]

1.4 Mammalian cell-based directed evolution enabled by *ex mammalia* mutagenesis

Ex mammalia mutagenesis combined with cellular screening is historically the most common approach for directed evolution in mammalian cells (**Figure 1.3**). *Ex mammalia* mutagenesis techniques were originally designed to support directed evolution in test tubes or lower organisms,[11-12] and typically do not account for the unique challenges associated with mutagenesis within mammalian cells. Nonetheless, despite the constraints that *ex mammalia* techniques place on library size and throughput, these strategies can produce valuable evolved biomolecules that interact with and depend on the unique mammalian cellular environment.[13, 20-21, 27-31]

1.4.1 Screening-based methods enabled by ex mammalia *mutagenesis.* An informative example of *ex mammalia* mutagenesis followed by cellular screening is the development of small molecule-controlled destabilizing domains by Wandless and co-workers.[27] Destabilizing domains are proteins that exist largely in an unfolded state. Proteins fused to destabilizing domains are rapidly degraded by the proteasome. However, in the presence of a stabilizing small molecule, the folded state is more prevalent. The result is small molecule-mediated control of the level of any destabilizing domain-fused protein in cells and animals. It is critical that destabilizing domains intended for use in mammalian systems are optimized directly in mammalian cells, as lower organisms can have vastly different proteostasis networks and degradation pathways. To evolve a destabilizing domain based on the human protein FKBP12, Wandless and co-workers used error-prone PCR and retroviral transduction to create and then stably express a library of ~30,000 FKBP12–YFP fusion protein variants in 3T3 mouse cells (**Figure 1.3a**).[27] From this library, 10,000 cells expressing FKBP12–YFP variants were treated with the FKBP12-stabilizing ligand SLF* and screened for YFP fluorescence using flow cytometry. Selected cells were then cultured in the absence of SLF* and subsequently screened again by flow cytometry for a decrease in YFP fluorescence. A final round of flow cytometry was performed upon addition of SLF* and, from the population expressing fluorescent YFP, 72 FKBP12 clones were sequenced, revealing several frequently occurring mutations and resulting in small molecule-regulated destabilizing domains for use in mammalian cells. A similar strategy has since been used to identify small molecule-regulated destabilizing domain versions of other proteins.[28-29] These destabilizing domains are now widely used to confer small molecule control of the levels of proteins ranging from Cas9[32] to membrane proteins[28, 33] in mammalian cells.

Another instructive example is the directed evolution of membrane-spanning ion channels with valuable new functions. Boyden and co-workers used *ex mammalia* mutagenesis, transient transfection, and an innovative robotic screening approach to evolve the membrane-bound

voltage reporter QuasAr2[34] for improved properties such as brightness and localization.[30] A library of Arch variants was generated through error-prone PCR (**Figure 1.3b**) and subsequently transfected into HEK293 cells. Detailed images of ~10,000 cells were computationally evaluated for desired biomolecular properties, such as membrane localization, brightness, and more. Cells expressing QuasAr2 variants with the desired properties were then physically detached from the dish using a robotic micropipette and sequenced. Two of the identified variants, Archon1 and Archon2, displayed improved brightness, photostability, and cellular localization, and allow detailed imaging of neuronal activity from many neurons simultaneously.

A second approach to ion channel directed evolution in mammalian cells was demonstrated by Dieudonné, Lin, and co-workers.[13] ASAP family voltage sensors are created via fusion of a circularly permuted GFP variant within the transmembrane region of a voltage-sensing phosphatase.[35] First-generation ASAP voltage sensors suffer from relatively slow kinetics and are, therefore, unable to temporally resolve rapid action potentials. Dieudonné, Lin, and co-workers developed a streamlined platform for directed evolution of ASAP voltage sensors by transfecting PCR-amplified, mutated constructs directly into mammalian cells (**Figure 1.3c**), avoiding the time-consuming and inefficient step of plasmid library generation. They then screened ~400 variants per round of evolution by inducing an electrical potential and observing, through imaging, cells with improved voltage sensing. After six iterative rounds of evolution, a sextuple-mutant of ASAP2f was identified, termed ASAP3, that showed more rapid activation kinetics similar to Archon1 and with improved molar brightness.

1.4.2 Selection-based methods enabled by **ex mammalia** *mutagenesis.* While less common than screening-based methods, selection-based methods combined with *ex mammalia* mutagenesis have also been used, especially in efforts to illuminate how oncoproteins evolve drug resistance during cancer treatment. The small molecule imatinib inhibits BCR-ABL, an oncogene encoding an aberrant tyrosine kinase that can cause chronic myeloid leukemia and other cancers.[36] Within just two years of FDA approval, 16 imatinib-resistant genetic mutations were observed in patients, creating a need to more comprehensively understand and predict the possible mutations that may permit imatinib resistance. Daley and co-workers used a single-round directed evolution technique to profile drug-resistant BCR-ABL variants in murine B cells.[21] They employed a mutator strain of bacteria to create a BCR-ABL variant library (**Figure 1.3d**), introduced the library into a retroviral vector, transduced ~500,000 murine B cells with the library, and then treated the cells with imatinib for 10 days. Upon picking surviving cell colonies and sequencing the BCR-ABL variants, they identified 13 of the 16 known drug-resistance mutations. More significantly, they also discovered 100 novel drug-resistant variants, some of which were

later observed in imatinib-treated cancer patients.[37] Similar approaches have since been deployed to illuminate drug resistance pathways in other oncoproteins, including farnesyl transferase[31] and mitogen-activated protein kinase.[20]

1.4.3 Summary. Proteins such as destabilizing domains, ion channels, and oncoproteins intimately rely on the unique mammalian cellular environment and complex mammalian pathways for proper function. These features cannot feasibly be moved into lower organisms or a test tube and, therefore, the success of these directed evolution campaigns depended on optimizing function within mammalian cells.

Unfortunately, while *ex mammalia* mutagenesis techniques are well-established and straightforward to employ, they have significant drawbacks. Most significantly, the labor-intensive nature of variant library expression in and extraction from mammalian cells mean that in most cases only a single round of evolution is performed on what is typically quite a small variant library. In cases where multiple rounds of evolution are performed, sometimes fewer than 10^3 variants are tested in each round. More complex directed evolution targets typically require multiple rounds of evolution to be performed iteratively on variant libraries that are orders of magnitude larger than in these studies.

1.5 Mammalian cell-based directed evolution enabled by *in mammalia* targeted mutagenesis

A key bottleneck in most early examples of mammalian cell-based directed evolution experiments is that only one round of mutagenesis is performed. Breakthrough techniques to enable *in mammalia* targeted mutagenesis (**Figures 1.2b** and **1.4a**) facilitate directed evolution experiments that require multiple rounds of evolution without the inefficient and cumbersome steps of library extraction, *ex mammalia* mutagenesis, and re-expression.
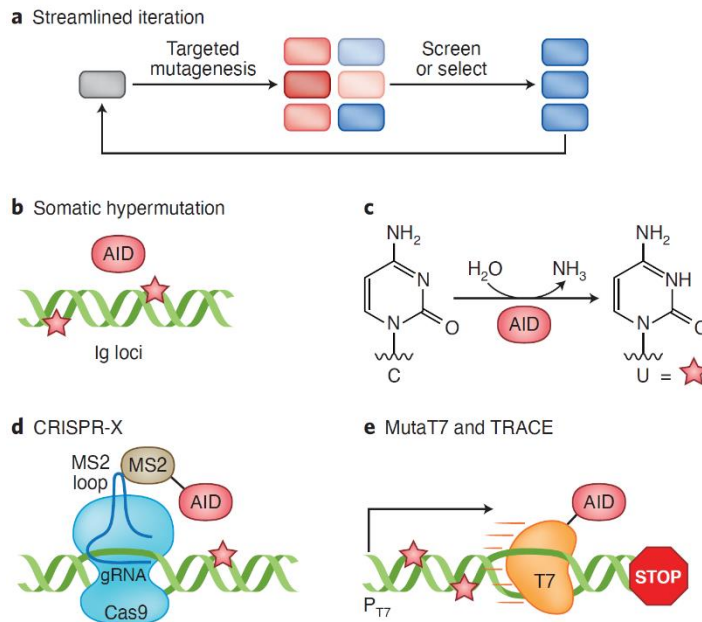
Figure 4. Targeted in mammalia mutagenesis. (a) Mutating genes of interest directly in mammalian cells streamlines iteration and increases the number of rounds of evolution that can be performed, but targeted mutagenesis of only the gene of interest for evolution is essential to avoid cheaters that escape selection pressure or subvert screening strategies. (b) Activation-induced cytidine deaminase (AID) mutagenizes hypervariable regions of immunoglobulin (Ig) loci,[14] although evidence suggests that significant off-target mutagenesis occurs.[44,45] (c) AID induces point mutations by deaminatingcytosine (C), creating uracil (U) bases that can be unfaithfully repaired or may base-pair with adenine upon replication, resulting in the introduction of C→T mutations. (d) CRISPR-X targets mutagenesis to user-defined DNA sequences through a targeted guide RNA (gRNA) with a MS2 binding loop that recruits an MS2-AID fusion.[50] This fusion protein mutagenizes DNA within a relatively narrow window of DNA (10–50 bp), but with much higher target specificity than AID alone. (e) Processive mutagenesis methods like TRACE[17] leverage the MutaT7[16] concept, in which AID is fused to a highly processive T7 polymerase. MutaT7 targets genes preceded by the T7 promoter (P_{T7}) and delimited by a terminator sequence.

Perhaps the most significant obstacle to diversifying genes *in mammalia* is the immense size of the mammalian genome, providing many opportunities for off-target mutations that can cheat screening or selection strategies. Unfortunately, many of the highly specific genomic editing strategies or commonly used genetic tools that function well in lower organisms do not exist for mammalian cells. For example, mammalian plasmids (e.g., latent viruses such as SV40[38]) and mammalian artificial chromosomes,[39] both of which facilitate *in vivo* mutagenesis in bacteria or yeast (e.g., in the OrthoRep platform for directed evolution in yeast[40]), do not nearly approach the utility of the corresponding systems in lower organisms. Despite these obstacles, the last five years have witnessed the development of a number of targeted mutagenesis strategies that use bespoke technologies to address problems presented by the large mammalian genome and other aspects of mammalian cell biology.

1.5.1 Somatic hypermutation. In mammalia targeted mutagenesis occurs naturally in B cells during the somatic hypermutation of antibody genes, mediated in part by the mutagenic activity of activation-induced cytidine deaminase (AID, **Figure 1.4b**).[41] Cytidine deaminases are a class of DNA-damaging enzymes that convert cytosine into uracil (**Figure 1.4c**). The resulting uracil bases may be unfaithfully repaired or may base pair with adenine instead of guanine upon replication, thereby introducing C→T mutations throughout targeted gene regions.

Several groups have taken advantage of AID-mediated mutagenesis to perform screening-based directed evoluti on of antibodies[42] and fluorescent proteins[43] in mammalian cells. Often, B cells that naturally express AID are used as a host cell to which genes of interest are introduced and continuously mutagenized. Unfortunately, although originally thought to be targeted to specific genetic loci, it is now clear that AID induces mutations genome-wide in cultured cells, especially at highly transcribed regions.[44-45] Supporting this observation, AID expression is cytotoxic when highly expressed, requiring researchers to artificially moderate expression levels during directed evolution experiments.[44] Thus, the need for truly targeted DNA-damaging enzymes has emerged.

1.5.2 Deactivated RNA-guided endonucleases paired with DNA-damaging enzymes. CRISPR enzymes such as Cas9 are bacteria-derived endonucleases that can be targeted to any genomic sequence containing an appropriate protospacer adjacent motif (PAM) using a single guide RNA (sgRNA). At an sgRNA-targeted genomic site, the endonuclease induces double-strand breaks that are commonly repaired by either non-homologous end joining or homology-directed repair.[15]

In principle, the resulting DNA damage can be used to diversify genes of interest for directed evolution experiments. Indeed, Cas9-induced in-frame insertions and deletions in

targeted oncogenes have been used to generate and enrich drug-resistant protein variants.[46-47] However, non-homologous end-joining is much less likely to generate functional variants than homology-directed repair, owing to the high potential for insertions and deletions to induce frameshifts and potentially cause early termination. A Cas9-based method that specifically promotes homology-directed repair has been used to evolve a red fluorescent protein to function in mammalian lysosomes.[19] Griesbeck and co-workers inserted a single copy of a mRuby-based fluorescent protein into mammalian genomes using the Flp-recombinase/FRT system. Single-stranded oligonucleotides encoding point mutants were then used as donor templates to promote the desired homology-directed repair at five amino acid codons with ~20% efficiency. Using this approach, a new variant of mRuby was identified, termed mCRISPRed, that showed stability at low pH and proved functional even in mammalian lysosomes.

Researchers have also sought to develop CRISPR-based mutagenesis methods that specifically promote point mutations. Inspired by base editors pioneered by Liu and co-workers,[48] two technologies were developed that combine the DNA-targeting ability of Cas9 with the point mutation-inducing activity of wild-type or engineered variants of cytidine deaminases such as AID and APOBEC (**Figure 1.4d**). In an approach termed targeted AID-mediated mutagenesis (TAM), Xing and co-workers fused AID to deactivated Cas9 variants (dCas9) that bind, but do not cut, specific DNA sequences.[49] In a proof-of-principle directed evolution experiment, this dCas9-AID fusion was then used to target point mutations to BCR-ABL in K562 cells. Upon passaging cells in the presence of imatinib, cells encoding drug-resistant BCR-ABL variants were readily enriched. In a related approach, termed CRISPR-X, a cytidine deaminase was fused to the RNA aptamer-binding protein MS2. The MS2-AID fusion can be recruited to a desired genetic locus by interaction with a dCas9-binding sgRNA that also contains an MS2-binding RNA aptamer domain.[50] CRISPR-X was used to target mutations to a proteasome subunit in K562 cells to generate and identify variants resistant to the proteasome inhibitor bortezomib. These methods have also been applied to evolve other endogenous genes, including antibodies.[51] Excitingly, the advent of additional DNA-damaging enzymes, such as adenosine deaminases,[52-53] may expand the accessible mutational spectrum using this type of approach for targeted mutagenesis.

Emerging CRISPR-based tools constitute a major improvement over natural somatic hypermutation as a strategy to target point mutations to DNA in mammalian cells. Nonetheless, there are still some significant limitations. TAM and CRISPR-X target mutations only to very small (~10–50 bp) windows of DNA that surround designed sgRNAs.[49-50] These small targeting windows mean that many different sgRNAs must be tiled along a gene to mutagenize the entire sequence. Critically, some regions may not be accessible owing to the absence of Cas9-specific PAMs.[15]

Further, since repeated rounds of directed evolution result in mutation accumulation, new sgRNAs are required as the guide recognition sequences mutate.

     1.5.3 Processive targeted mutagenesis. Bearing in mind the limitations of CRISPR-based technologies, a new type of strategy for targeted *in mammalia* mutagenesis was more recently developed. Like Cas9, RNA polymerases can be precisely targeted to specific regions of DNA. Unlike Cas9, RNA polymerases can be extremely processive. For example, the bacteriophage-derived T7 RNA polymerase is able to traverse >10,000 bp of DNA,[54] directed by the highly-specific T7 promoter. In the *Escherichia coli*-based MutaT7 system, a T7 polymerase was fused to a cytidine deaminase to create a processive, highly targeted mutagenic protein that introduces point mutations *in vivo* in any stretch of DNA downstream of a T7 promoter (**Figure 1.4e**).[16, 55] The TRACE platform built on this MutaT7 concept to mutagenize specific genes of interest directly in mammalian cells.[17] TRACE has been used to screen for GFP variants with blue-shifted emission spectra in HEK293T cells and to select for MEK variants resistant to small molecule inhibitors in A375 cells.

     The high processivity of RNA polymerases permits the mutation of kilobases of DNA without significant sequence limitations, such as the availability of PAM sequences or the need to synthesize and introduce multiple, custom sgRNAs for each gene of interest during each round of evolution. Multiple strategies to terminate T7 polymerase can be used to prevent unwanted mutagenesis downstream of the targeted region,[16, 56] and installation of a second T7 promoter facing the opposite direction of the protein coding sequence allows for the introduction of both C→T and G→A mutations.[16] Moreover, fusion to other DNA-damaging enzymes, such as improved variants of the adenosine deaminases[52-53] and a related approach to induce C→G transversions[57] are beginning to allow expansion of the mutational spectrum of these technologies,[56] as was also the case in the CRISPR systems. A key limitation of MutaT7-based systems, such as TRACE, is that they require installation of a T7 promoter upstream of the gene of interest. The significance of this limitation depends largely on whether the user seeks to evolve an endogenous gene encoded in its unmodified genomic context.

     1.5.4 Summary. By simplifying the process of mutant library generation and expression, the capacity of targeted *in mammalia* mutagenesis techniques to facilitate multiple rounds of directed evolution is potentially transformative. In theory, it is now possible to easily mutagenize any gene targeted by an RNA-guided endonuclease or by a T7 polymerase directly in mammalian cells. Ultimately, improvements in targeted *in mammalia* mutagenesis focus only on mutagenesis but not on further steps. Directed evolution experiments using these advanced genetic

diversification techniques still rely on traditional methods of selection or screening to enrich cells encoding desired variants.

1.6 Virus-based continuous directed evolution in mammalian cells

Despite improvements with *in mammalia* mutagenesis, a number of issues with typical cell-based platforms continue to restrict broad application. First, enrichment of functional variants can be limited by the slow rate of replication of mammalian cells, often requiring individual rounds of evolution to take a week or longer. Second, assessment techniques limit the types of biomolecules that can be evolved. Selection-based methods can only evolve functions that impact cell viability, while screening methods require an observable phenotype that can be screened. Many valuable targets of directed evolution do not fit easily in either category.

New implementations of virus-based continuous directed evolution in mammalian cells are primed to address these obstacles and have potential to revolutionize the field. The concept of virus-based directed evolution is simple: Introduce the gene of interest into the genome of a mammalian virus that replicates with low fidelity. Then, implement a selection couple that ties the ability of the virus to propagate in mammalian cells to the function of the directed evolution target. From there, simply propagating the virus results in the generation and enrichment of desired gene variants. In the field of bacteria-based directed evolution, PACE (phage-assisted continuous evolution)[58] has enabled the rapid and efficient evolution of many diverse protein functions by simply coupling the desired protein function to the ability of M13 phage to replicate in *E. coli*.[59] Analogous mammalian-cell based systems are poised to do the same.

A number of general principles must be considered when designing a broadly useful virus-based mammalian directed evolution platform. First, the rate of viral growth is important. Some mammalian viruses replicate faster than others or replicate with a larger burst size. However, viral replication must not outpace the rate of expression, maturation, and function of the protein of interest. Second, platforms should be able to evolve as many genes of interest in as many different mammalian cell types as possible. Viruses with large packaging capacities and broad tropisms are preferable.[60] Third, platforms should be designed with practicability in mind. Virus generation, packaging efficiency, and other design features such as visual aids to monitor viral propagation should be considered. Lastly, the most important consideration is safety. Many laboratory viral strains have built-in safety strategies to prevent their replication in wild-type hosts. These strategies include deleting genes non-essential to cell culture,[61] using extensive trans-complementation,[62] or developing laboratory-adapted strains that are incapable of producing a

replicative infection in organisms. These types of safety features should always be included and verified in any mammalian virus-based directed evolution platform.

     1.6.1 mPACE-style platforms for continuous directed evolution in mammalian cells. While there exist some intriguing early reports of viral evolution in mammalian cells,[63-65] early methods are ill-equipped for a broad range of substrates and often have serious safety liabilities. The landscape has changed with the reports of two adaptable virus-based directed evolution platforms for mammalian cells from our laboratory (see also Chapter 2)[25] and from Roth, English, and co-workers.[26] Both of these mammalian PACE (mPACE)-style platforms rely on custom engineering of a virus and a host cell to enable mutagenesis, selection, and amplification to occur simultaneously and continuously throughout an evolution experiment. First, essential viral genes are deleted from the viral genome (**Figure 1.5a**). Second, the gene of interest for evolution is introduced into the viral genome. Finally, the desired function of the gene of interest is then coupled to transcription, translation, or activation of the missing viral proteins in the host cell (**Figure 1.5b**). In this way, viruses that contain active genes of interest are rescued by trans-complementation and rapidly enrich in the population (**Figure 1.5c**). Viruses that do not contain active genes of interest are not rescued and cannot replicate. Critically, in an mPACE-style system, the mammalian cells are replaced after each round of evolution, avoiding the problem off-target mutation accumulation. Creativity is required for designing an appropriate selection couple, but the success of bacterial PACE highlights the tremendous breadth of functions that can be evolved using this type of system, ranging from protein stability to diverse enzymatic activities.[58, 66-69]

     The mPACE-style system introduced by our group, which is discussed in detail in Chapter 2, employs the double-stranded DNA virus adenovirus to achieve continuous directed evolution in ma mmalian cells (**Figures 1.5d** and **1.5e**), with mutations introduced during viral replication by a highly error-prone, engineered version of the adenoviral DNA polymerase.[25] The second mPACE-style system, introduced by Roth, English, and co-workers and termed VEGAS (viral evolution of genetically actuating sequences), operates on the same principles but uses the single-stranded RNA alphavirus Sindbis (**Figures 1.5f** and **1.5g**) with mutations introduced by the endogenous Sindbis replicase.[26] Both systems rely on conditional expression of essential viral genes to drive selection, either the adenoviral protease (AdProt) or the Sindbis structural genome (SSG), respectively, and both mPACE-style platforms demonstrate their potential with proof-of-principle experiments.
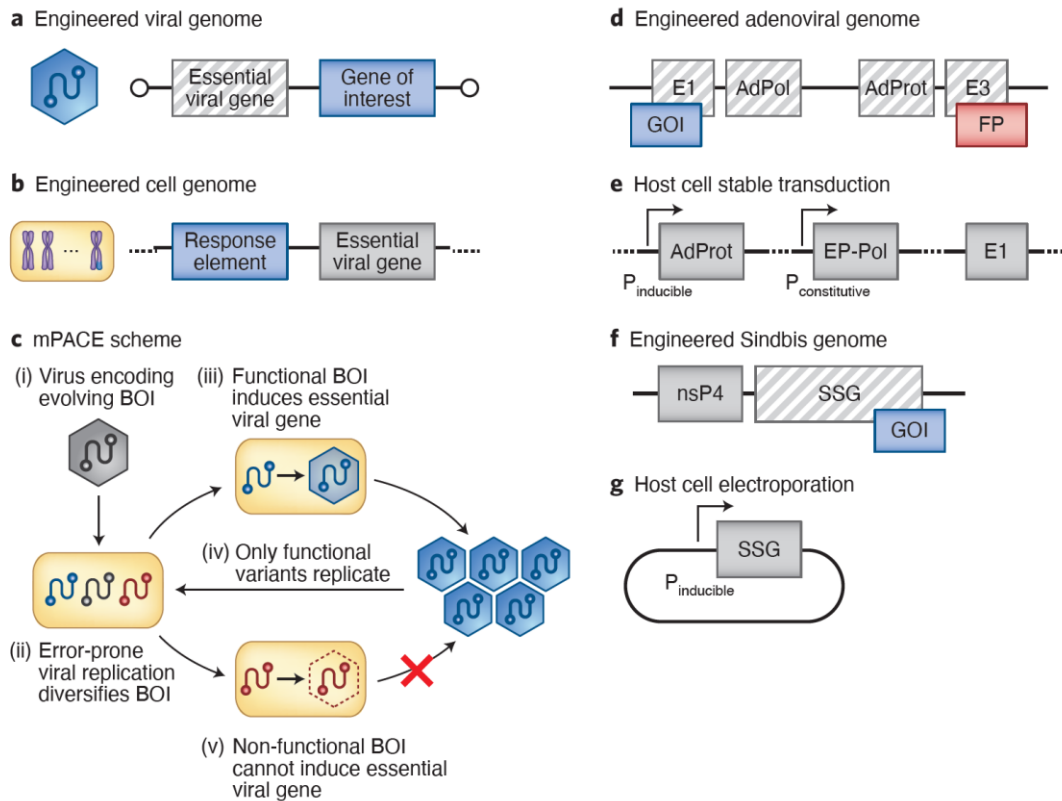
Figure 5 (a–b) Mammalian PACE, or mPACE, utilizes two engineered components, a virus and a host cell. (a) The engineered virus lacks essential viral genes required for replication, but contains an evolving gene of interest. (b) The engineered mammalian cell encodes the deleted essential viral genes. Transcription, translation, or the activity of the conditionally trans-complemented viral genes is coupled to the desired activity of the evolving gene of interest. (c) These components combine to create a directed evolution platform, where the activity of a biomolecule of interest (BOI) drives viral replication. Over time, viruses encoding active genes of interest (GOIs) enrich in the population while viruses with inactive genes of interest do not replicate and are depleted. (d–e) Overview of the adenovirus-based mPACE directed evolution platform.[25] (d) For safety and to clear space for insertion of a gene of interest (GOI) to be evolved, the engineered adenoviral genome has the E1 genes, E3 genes, the adenoviral polymerase (AdPol), and the adenoviral protease (AdProt) deleted from the wild-type genome. A gene encoding a fluorescent protein (FP) is also included to allow for easy visualization of directed evolution campaigns and viral titering. (e) E1, AdPol, and AdProt are stably trans-complemented by the host cell. The E3 region, which is dispensable in cell culture, is not trans-complemented at all to provide an additional safety feature. Mutagenesis of an encoded gene of interest is conferred by a constitutively expressed, highly error-prone variant of AdPol (EP-Pol). AdProt activity is inducibly coupled to the desired function of the GOI. (f–g) Overview of the Sindbis virus-based mPACE directed evolution platform (VEGAS).[26] (f) Multiple genes are deleted from the wild-type Sindbis genome, including the capsid protein, E1, E2, 6K, and E3, known collectively as the Sindbis structural genome (SSG). (g) The missing SSG genes are inducibly expressed in response to the desired activity of the GOI via a gene cassette introduced by electroporation before each round of evolution. Mutagenesis of an encoded gene of interest is conferred by the wild-type Sindbis replicase (nsP4; or nonstructural protein 4).

Both platforms present distinct advantages. The adenovirus-based system has a number of features that enhance ease-of-use, including a fluorescent marker to visualize viral infection and compatibility with Gateway[TM] cloning to introduce genes of interest for evolution. The approach is compatible with a small molecule inhibitor[70] of the selection marker, AdProt, which can be applied to tune the stringency of selection pressure over the course of a directed evolution campaign.[25] Because adenovirus contains a double-stranded DNA genome, this system is also compatible with the targeted mutagenesis techniques discussed earlier. The system includes multiple distinct safety features to prevent production of a replication-competent virus. Three separate genetic regions are trans-complemented, decreasing the chance that a viable virus can be created. In addition, the E3 region essential for replication in living organisms is deleted and not trans-complemented at all. Lastly, adenovirus has an unusually broad tropism, allowing application in a vast array of different mammalian cell types.[71]

VEGAS also has unique advantages. Sindbis virus can be generated more quickly than adenovirus and is readily compatible with *ex mammalia*-diversified libraries. Sindbis has a faster replication cycle than adenovirus, ~4 hours for Sindbis versus ~24 hours for adenovirus, accelerating iterative rounds of evolution. Roth and co-workers estimate a mutation rate of 4×10^{-4} mutations per base per replication, whereas the adenoviral system has a lower estimated rate of 4×10^{-5} mutations per base per replication. It is important to balance useful mutation rates with mutational error catastrophe and so it is difficult to know the best mutation rate *a priori*—a single mutation is sufficient to render tTA fully functional yet inhibitor-insensitive[25] whereas the consensus tTA variant from VEGAS had 47 mutations—but a higher mutation rate is often valuable. More significant than the mutation rate, Roth, English, and co-workers report a largely unbiased mutational spectrum for the Sindbis replicase. It is noteworthy in this regard, however, that evolved sequences produced by VEGAS show an extremely high percentage of A→G transitions, including 46 of 47 reported mutations in the consensus sequence from the tTA evolution experiment—an observation that appears inconsistent with unbiased mutagenesis.[26]

1.6.2 Summary. Amongst all mammalian cell-based directed evolution platforms, virus-based continuous directed evolution platforms provide the largest library sizes, fastest experimental timescales, and allow, in principle, for the broadest range of biomolecules to be evolved. These mPACE-style techniques would still greatly benefit from additional features to improve scope and efficiency. First, current mPACE selection schemes only allow for positive selection, which may lead to poly-specific, promiscuous, or even constitutive activity. Developing negative selection schemes would allow the evolution of specific protein activity or environment-dependent protein activity, such as in the presence of a small molecule, in a particular cell type,

or in a particular genetic background. Second, implementing fully continuous mPACE systems may increase library size, accelerate experimental timescales, and reduce researcher effort. Third, mutations are currently driven entirely by error-prone viral polymerases. The result is both some bias in mutational spectra as well as continual mutagenesis of the viral genome, which may eventually cause error catastrophe. Finally, published mPACE directed evolution campaigns all rely on direct transcriptional couples. Demonstration of adaptability to other types of couples that evolve non-transcription-related activities is essential.

Although there is more work to be done, virus-based directed evolution in mammalian cells presents a quantum leap over previous *in mammalia* evolution methods, providing simultaneous mutagenesis, selection, and enrichment on rapid timescales. As has proven readily achievable in bacterial PACE, mPACE-style approaches are likely to enable the evolution of targets previously unattainable by other screening or selection-based methods. The two platforms presented here therefore lay the groundwork for new discoveries and innovative applications in mammalian cells.

1.7 Conclusion

In conclusion, recent years have shown exciting progress in the development of new methods for directed evolution in mammalian cells. Initial studies that used ex mammalian mutagenesis and screening or selection methods have since been supplanted by higher throughput and more adaptable methods, particularly due to the innovations in *in mammalia* mutagenesis and the invention of new virus-based continuous methods. These newly developed methods for directed evolution in mammalian cells present a flexible and powerful solution to the problem of engineering new protein functionalities in mammalian cells.

In this thesis, I will discuss the development and application of our platform toward the evolution of new protein classes. In Chapter 2, I discuss the development of our virus-based continuous directed evolution platform, from conception to proof-of-principle evolution experiments. In Chapter 3, I discuss our efforts adapting our platform toward the directed evolution of G-protein coupled receptors. In Chapter 4, I discuss our efforts adapting our platform toward the directed evolution of CRISPR systems. In Chapter 5, I present a perspective on the future of mammalian cell-based directed evolution, including a wide range of impactful targets to pursue.

Just as directed evolution has revolutionized protein engineering in bacteria and yeast, the new methods presented here are poised to revolutionize protein engineering in mammalian cells. Both the challenges and the successes presented in this work can inform researchers

looking to use directed evolution in their own research, and may provide inspiration for a new generation of researchers who wish to look beyond the status quo to explore new frontiers of protein engineering.

1.8 References

1.	Drummond, D. A.; Bloom, J. D., A Nobel Prize for evolution. *Evolution* **2019,** *73* (3), 630-631.

2.	Visintin, M.; Tse, E.; Axelson, H.; Rabbitts, T. H.; Cattaneo, A., Selection of antibodies for intracellular function using a two-hybrid *in vivo* system. *Proc. Natl. Acad. Sci. U. S. A.* **1999,** *96* (21), 11723-8.

3.	Herwig, L.; Rice, A. J.; Bedbrook, C. N.; Zhang, R. K.; Lignell, A.; Cahn, J. K. B.; Renata, H.; Dodani, S. C.; Cho, I.; Cai, L.; Gradinaru, V.; Arnold, F. H., Directed evolution of a bright near-infrared fluorescent rhodopsin using a synthetic chromophore. *Cell Chem. Biol.* **2017,** *24* (3), 415-425.

4.	Buskirk, A. R.; Ong, Y. C.; Gartner, Z. J.; Liu, D. R., Directed evolution of ligand dependence: small-molecule-activated protein splicing. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (29), 10505-10.

5.	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.* **2011,** *18* (5), 619-30.

6.	Italia, J. S.; Peeler, J. C.; Hillenbrand, C. M.; Latour, C.; Weerapana, E.; Chatterjee, A., Genetically encoded protein sulfation in mammalian cells. *Nat. Chem. Biol.* **2020,** *16* (4), 379-382.

7.	Armbruster, B. N.; Li, X.; Pausch, M. H.; Herlitze, S.; Roth, B. L., 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.* **2007,** *104* (12), 5163-8.

8.	Pourmir, A.; Johannes, T. W., Directed evolution: selection of the host organism. *Comput. Struct. Biotechnol. J.* **2012,** *2*, e201209012.

9.	Kachroo, A. H.; Laurent, J. M.; Yellman, C. M.; Meyer, A. G.; Wilke, C. O.; Marcotte, E. M., Systematic humanization of yeast genes reveals conserved functions and genetic modularity. *Science* **2015,** *348* (6237), 921-5.

10.	Cirino, P. C.; Mayer, K. M.; Umeno, D., Generating mutant libraries using error-prone PCR. *Methods Mol. Biol.* **2003,** *231*, 3-9.

11.	Muteeb, G.; Sen, R., Random mutagenesis using a mutator strain. *Methods Mol. Biol.* **2010,** *634*, 411-9.

12.	Badran, A. H.; Liu, D. R., Development of potent *in vivo* mutagenesis plasmids with broad mutational spectra. *Nat. Commun.* **2015,** *6* (1), 8425.

13.	Villette, V.; Chavarha, M.; Dimov, I. K.; Bradley, J.; Pradhan, L.; Mathieu, B.; Evans, S. W.; Chamberland, S.; Shi, D.; Yang, R.; Kim, B. B.; Ayon, A.; Jalil, A.; St-Pierre, F.; Schnitzer, M. J.; Bi, G.; Toth, K.; Ding, J.; Dieudonné, S.; Lin, M. Z., Ultrafast two-photon imaging of a high-gain voltage indicator in awake behaving mice. *Cell* **2019,** *179* (7), 1590-1608.e23.

14.	Odegard, V. H.; Schatz, D. G., Targeting of somatic hypermutation. *Nat. Rev. Immunol.* **2006,** *6* (8), 573-83.

15.	Pickar-Oliver, A.; Gersbach, C. A., The next generation of CRISPR-Cas technologies and applications. *Nat. Rev. Mol. Cell Biol.* **2019,** *20* (8), 490-507.

16.	Moore, C. L.; Papa, L. J., 3rd; Shoulders, M. D., A processive protein chimera introduces mutations across defined DNA regions *in vivo. J. Am. Chem. Soc.* **2018,** *140* (37), 11560-11564.

17.	Chen, H.; Liu, S.; Padula, S.; Lesman, D.; Griswold, K.; Lin, A.; Zhao, T.; Marshall, J. L.; Chen, F., Efficient, continuous mutagenesis in human cells using a pseudo-random DNA editor. *Nat. Biotechnol.* **2020,** *38* (2), 165-168.

18.	Sansbury, B. M.; Hewes, A. M.; Kmiec, E. B., Understanding the diversity of genetic outcomes from CRISPR-Cas generated homology-directed repair. *Commun. Biol.* **2019,** *2* (1), 458.

19.   Erdogan, M.; Fabritius, A.; Basquin, J.; Griesbeck, O., Targeted *in situ* protein diversification and intra-organelle validation in mammalian cells. *Cell Chem. Biol.* **2020,** *27*, 610-621.

20.   Emery, C. M.; Vijayendran, K. G.; Zipser, M. C.; Sawyer, A. M.; Niu, L.; Kim, J. J.; Hatton, C.; Chopra, R.; Oberholzer, P. A.; Karpova, M. B.; MacConaill, L. E.; Zhang, J.; Gray, N. S.; Sellers, W. R.; Dummer, R.; Garraway, L. A., MEK1 mutations confer resistance to MEK and B-RAF inhibition. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106* (48), 20411-6.

21.   Azam, M.; Latek, R. R.; Daley, G. Q., Mechanisms of autoinhibition and STI-571/Imatinib resistance revealed by mutagenesis of BCR-ABL. *Cell* **2003,** *112* (6), 831-843.

22.   McDermott, M.; Eustace, A. J.; Busschots, S.; Breen, L.; Crown, J.; Clynes, M.; O'Donovan, N.; Stordal, B., *In vitro* development of chemotherapy and targeted therapy drug-resistant cancer cell lines: a practical guide with case studies. *Frontiers in oncology* **2014,** *4*, 40.

23.   Booth, L.; Roberts, J. L.; Tavallai, M.; Webb, T.; Leon, D.; Chen, J.; McGuire, W. P.; Poklepovic, A.; Dent, P., The afatinib resistance of *in vivo* generated H1975 lung cancer cell clones is mediated by SRC/ERBB3/c-KIT/c-MET compensatory survival signaling. *Oncotarget* **2016,** *7* (15), 19620-30.

24.   Forment, J. V.; Herzog, M.; Coates, J.; Konopka, T.; Gapp, B. V.; Nijman, S. M.; Adams, D. J.; Keane, T. M.; Jackson, S. P., Genome-wide genetic screening with chemically mutagenized haploid embryonic stem cells. *Nat. Chem. Biol.* **2017,** *13* (1), 12-14.

25.   Berman, C. M.; Papa, L. J., 3rd; Hendel, S. J.; Moore, C. L.; Suen, P. H.; Weickhardt, A. F.; Doan, N. D.; Kumar, C. M.; Uil, T. G.; Butty, V. L.; Hoeben, R. C.; Shoulders, M. D., An adaptable platform for directed evolution in human cells. *J. Am. Chem. Soc.* **2018,** *140* (51), 18093-18103.

26.   English, J. G.; Olsen, R. H. J.; Lansu, K.; Patel, M.; White, K.; Cockrell, A. S.; Singh, D.; Strachan, R. T.; Wacker, D.; Roth, B. L., VEGAS as a platform for facile directed evolution in mammalian cells. *Cell* **2019,** *178* (3), 748-761.

27.   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* **2006,** *126* (5), 995-1004.

28.   Iwamoto, M.; Bjorklund, T.; Lundberg, C.; Kirik, D.; Wandless, T. J., A general chemical method to regulate protein stability in the mammalian central nervous system. *Chem. Biol.* **2010,** *17* (9), 981-8.

29.   Miyazaki, Y.; Imoto, H.; Chen, L. C.; Wandless, T. J., Destabilizing domains derived from the human estrogen receptor. *J. Am. Chem. Soc.* **2012,** *134* (9), 3942-5.

30.   Piatkevich, K. D.; Jung, E. E.; Straub, C.; Linghu, C.; Park, D.; Suk, H. J.; Hochbaum, D. R.; Goodwin, D.; Pnevmatikakis, E.; Pak, N.; Kawashima, T.; Yang, C. T.; Rhoades, J. L.; Shemesh, O.; Asano, S.; Yoon, Y. G.; Freifeld, L.; Saulnier, J. L.; Riegler, C.; Engert, F.; Hughes, T.; Drobizhev, M.; Szabo, B.; Ahrens, M. B.; Flavell, S. W.; Sabatini, B. L.; Boyden, E. S., A robotic multidimensional directed evolution approach applied to fluorescent voltage reporters. *Nat. Chem. Biol.* **2018,** *14* (4), 352-360.

31.   Raz, T.; Nardi, V.; Azam, M.; Cortes, J.; Daley, G. Q., Farnesyl transferase inhibitor resistance probed by target mutagenesis. *Blood* **2007,** *110* (6), 2102-2109.

32.   Maji, B.; Moore, C. L.; Zetsche, B.; Volz, S. E.; Zhang, F.; Shoulders, M. D.; Choudhary, A., Multidimensional chemical control of CRISPR-Cas9. *Nat. Chem. Biol.* **2017,** *13* (1), 9-11.

33.   Dietz, O.; Rusch, S.; Brand, F.; Mundwiler-Pachlatko, E.; Gaida, A.; Voss, T.; Beck, H. P., Characterization of the small exported *Plasmodium falciparum* membrane protein SEMP1. *PLoS One* **2014,** *9* (7), e103272.

34.   Hochbaum, D. R.; Zhao, Y.; Farhi, S. L.; Klapoetke, N.; Werley, C. A.; Kapoor, V.; Zou, P.; Kralj, J. M.; Maclaurin, D.; Smedemark-Margulies, N.; Saulnier, J. L.; Boulting, G. L.; Straub, C.; Cho, Y. K.; Melkonian, M.; Wong, G. K.; Harrison, D. J.; Murthy, V. N.; Sabatini, B. L.;

Boyden, E. S.; Campbell, R. E.; Cohen, A. E., All-optical electrophysiology in mammalian neurons using engineered microbial rhodopsins. *Nat. Methods* **2014,** *11* (8), 825-33.

35.	St-Pierre, F.; Marshall, J. D.; Yang, Y.; Gong, Y.; Schnitzer, M. J.; Lin, M. Z., High-fidelity optical reporting of neuronal electrical activity with an ultrafast fluorescent voltage sensor. *Nat. Neurosci.* **2014,** *17* (6), 884-889.

36.	Capdeville, R.; Buchdunger, E.; Zimmermann, J.; Matter, A., Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. *Nat. Rev. Drug Discov.* **2002,** *1* (7), 493-502.

37.	Soverini, S.; Hochhaus, A.; Nicolini, F. E.; Gruber, F.; Lange, T.; Saglio, G.; Pane, F.; Muller, M. C.; Ernst, T.; Rosti, G.; Porkka, K.; Baccarani, M.; Cross, N. C.; Martinelli, G., BCR-ABL kinase domain mutation analysis in chronic myeloid leukemia patients treated with tyrosine kinase inhibitors: recommendations from an expert panel on behalf of European LeukemiaNet. *Blood* **2011,** *118* (5), 1208-15.

38.	Mahon, M. J., Vectors bicistronically linking a gene of interest to the SV40 large T antigen in combination with the SV40 origin of replication enhance transient protein expression and luciferase reporter activity. *Biotechniques* **2011,** *51* (2), 119-28.

39.	Brown, D. M.; Glass, J. I., Technology used to build and transfer mammalian chromosomes. *Exp. Cell Res.* **2020,** *388* (2), 111851.

40.	Ravikumar, A.; Arzumanyan, G. A.; Obadi, M. K. A.; Javanpour, A. A.; Liu, C. C., Scalable, continuous evolution of genes at mutation rates above genomic error thresholds. *Cell* **2018,** *175* (7), 1946-1957.

41.	Pavri, R.; Nussenzweig, M. C., AID targeting in antibody diversity. *Adv. Immunol.* **2011,** *110*, 1-26.

42.	Bowers, P. M.; Horlick, R. A.; Neben, T. Y.; Toobian, R. M.; Tomlinson, G. L.; Dalton, J. L.; Jones, H. A.; Chen, A.; Altobell, L., 3rd; Zhang, X.; Macomber, J. L.; Krapf, I. P.; Wu, B. F.; McConnell, A.; Chau, B.; Holland, T.; Berkebile, A. D.; Neben, S. S.; Boyle, W. J.; King, D. J., Coupling mammalian cell surface display with somatic hypermutation for the discovery and maturation of human antibodies. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108* (51), 20455-60.

43.	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.* **2004,** *101* (48), 16745-9.

44.	Al-Qaisi, T. S.; Su, Y. C.; Roffler, S. R., Transient AID expression for *in situ* mutagenesis with improved cellular fitness. *Sci. Rep.* **2018,** *8* (1), 9413.

45.	Wang, C. L.; Harper, R. A.; Wabl, M., Genome-wide somatic hypermutation. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (19), 7352-6.

46.	Donovan, K. F.; Hegde, M.; Sullender, M.; Vaimberg, E. W.; Johannessen, C. M.; Root, D. E.; Doench, J. G., Creation of novel protein variants with CRISPR/Cas9-mediated mutagenesis: turning a screening by-product into a discovery tool. *PLoS One* **2017,** *12* (1), e0170445.

47.	Ipsaro, J. J.; Shen, C.; Arai, E.; Xu, Y.; Kinney, J. B.; Joshua-Tor, L.; Vakoc, C. R.; Shi, J., Rapid generation of drug-resistance alleles at endogenous loci using CRISPR-Cas9 indel mutagenesis. *PLoS One* **2017,** *12* (2), e0172177.

48.	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* **2016,** *533* (7603), 420-4.

49.	Ma, Y.; Zhang, J.; Yin, W.; Zhang, Z.; Song, Y.; Chang, X., Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* **2016,** *13* (12), 1029-1035.

50.	Hess, G. T.; Fresard, L.; Han, K.; Lee, C. H.; Li, A.; Cimprich, K. A.; Montgomery, S. B.; Bassik, M. C., Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* **2016,** *13* (12), 1036-1042.

51.   Devilder, M. C.; Moyon, M.; Gautreau-Rolland, L.; Navet, B.; Perroteau, J.; Delbos, F.; Gesnel, M. C.; Breathnach, R.; Saulquin, X., *Ex vivo* evolution of human antibodies by CRISPR-X: from a naive B cell repertoire to affinity matured antibodies. *BMC Biotechnol.* **2019,** *19* (1), 14.

52.   Gaudelli, N. M.; Komor, A. C.; Rees, H. A.; Packer, M. S.; Badran, A. H.; Bryson, D. I.; Liu, D. R., Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* **2017,** *551* (7681), 464-471.

53.   Gaudelli, N. M.; Lam, D. K.; Rees, H. A.; Sola-Esteves, N. M.; Barrera, L. A.; Born, D. A.; Edwards, A.; Gehrke, J. M.; Lee, S. J.; Liquori, A. J.; Murray, R.; Packer, M. S.; Rinaldi, C.; Slaymaker, I. M.; Yen, J.; Young, L. E.; Ciaramella, G., Directed evolution of adenine base editors with increased activity and therapeutic application. *Nat. Biotechnol.* **2020**.

54.   Thiel, V.; Herold, J.; Schelle, B.; Siddell, S. G., Infectious RNA transcribed *in vitro* from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. *J. Gen. Virol.* **2001,** *82* (Pt 6), 1273-1281.

55.   Park, H.; Kim, S., Gene-specific mutagenesis enables rapid continuous evolution of enzymes *in vivo*. *Nucleic Acids Res.* **2021**.

56.   Alvarez, B.; Mencia, M.; de Lorenzo, V.; Fernandez, L. A., *In vivo* diversification of target genomic sites using processive base deaminase fusions blocked by dCas9. *Nat. Commun.* **2020,** *11* (1), e6436.

57.   Kurt, I. C.; Zhou, R.; Iyer, S.; Garcia, S. P.; Miller, B. R.; Langner, L. M.; Grunewald, J.; Joung, J. K., CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat. Biotechnol.* **2021,** *39* (1), 41-46.

58.   Esvelt, K. M.; Carlson, J. C.; Liu, D. R., A system for the continuous directed evolution of biomolecules. *Nature* **2011,** *472* (7344), 499-503.

59.   Morrison, M. S.; Podracky, C. J.; Liu, D. R., The developing toolkit of continuous directed evolution. *Nat. Chem. Biol.* **2020,** *16* (6), 610-619.

60.   Dai, X.; Zhang, X.; Ostrikov, K.; Abrahamyan, L., Host receptors: the key to establishing cells with broad viral tropism for vaccine production. *Crit. Rev. Microbiol.* **2020,** *46* (2), 147-168.

61.   Danthinne, X.; Imperiale, M. J., Production of first generation adenovirus vectors: a review. *Gene Ther.* **2000,** *7* (20), 1707-14.

62.   Chira, S.; Jackson, C. S.; Oprea, I.; Ozturk, F.; Pepper, M. S.; Diaconu, I.; Braicu, C.; Raduly, L. Z.; Calin, G. A.; Berindan-Neagoe, I., Progresses towards safe and efficient gene therapy vectors. *Oncotarget* **2015,** *6* (31), 30675-703.

63.   Mullick, A.; Xu, Y.; Warren, R.; Koutroumanis, M.; Guilbault, C.; Broussau, S.; Malenfant, F.; Bourget, L.; Lamoureux, L.; Lo, R.; Caron, A. W.; Pilotte, A.; Massie, B., The cumate gene-switch: a system for regulated expression in mammalian cells. *BMC Biotechnol.* **2006,** *6*, 43.

64.   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.; Hoeben, R. C., Directed adenovirus evolution using engineered mutator viral polymerases. *Nucleic Acids Res.* **2011,** *39* (5), e30.

65.   Das, A. T.; Zhou, X.; Vink, M.; Klaver, B.; Verhoef, K.; Marzio, G.; Berkhout, B., Viral evolution as a tool to improve the tetracycline-regulated gene expression system. *J. Biol. Chem.* **2004,** *279* (18), 18776-82.

66.   Richter, M. F.; Zhao, K. T.; Eton, E.; Lapinaite, A.; Newby, G. A.; Thuronyi, B. W.; Wilson, C.; Koblan, L. W.; Zeng, J.; Bauer, D. E.; Doudna, J. A.; Liu, D. R., Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* **2020**.

67.   Wang, T.; Badran, A.; Huang, T.; Liu, D., Continuous directed evolution of proteins with improved soluble expression. *Nat. Chem. Biol.* **2018,** *14*.

68.   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.* **2014,** *5* (1), 5352.

69.   Miller, S. M.; Wang, T.; Randolph, P. B.; Arbab, M.; Shen, M. W.; Huang, T. P.; Matuszek, Z.; Newby, G. A.; Rees, H. A.; Liu, D. R., Continuous evolution of SpCas9 variants compatible with non-G PAMs. *Nat. Biotechnol.* **2020,** *38* (4), 471-481.

70.   Mac Sweeney, A.; Grosche, P.; Ellis, D.; Combrink, K.; Erbel, P.; Hughes, N.; Sirockin, F.; Melkko, S.; Bernardi, A.; Ramage, P.; Jarousse, N.; Altmann, E., Discovery and structure-based optimization of adenain inhibitors. *ACS Med Chem Lett* **2014,** *5* (8), 937-941.

71.   Havenga, M. J.; Lemckert, A. A.; Ophorst, O. J.; van Meijer, M.; Germeraad, W. T.; Grimbergen, J.; van Den Doel, M. A.; Vogels, R.; van Deutekom, J.; Janson, A. A.; de Bruijn, J. D.; Uytdehaag, F.; Quax, P. H.; Logtenberg, T.; Mehtali, M.; Bout, A., Exploiting the natural diversity in adenovirus tropism for therapy and prevention of disease. *J. Virol.* **2002,** *76* (9), 4612-20.

Chapter 2: A method for continuous directed evolution in mammalian cells

*Denotes equal contributions

2.1 Author Contributions

L.J.P., C.M.B., and S.J.H. designed and performed experiments throughout the chapter. L.J.P. performed recombineering for the construction of the adenoviral vectors. C.M.B. and S.J.H. engineered cell lines and performed tissue culture experiments related to platform characterization. S.J.H. performed adenoviral protease trans-complementation experiments. S.J.H. performed the directed evolution of the tet-transactivator. S.J.H. and C.M.B. characterized tet-transactivator mutants. L.J.P. and C.M.B. constructed the selection circuits for Leucyl-tRNA synthetase and Cre recombinase. C.M.B. performed Leucyl-tRNA synthetase and Cre recombinase selection. L.J.P. and N.-D.D. synthesized the adenoviral protease inhibitor. C.L.M. contributed to project conception, data analysis, and early-stage experiments. P.H.S. and A.F.W. performed experiments and analyzed data. C.M.K. assisted in cloning and recombineering. T.G.U. and R.C.H. designed the error-prone adenoviral polymerase and characterized its mutagenesis activity. V.L.B. analyzed deep sequencing data. M.D.S. conceived the project, acquired funding, and supervised experiments and data analysis.

2.2 Summary

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 seven kilobases 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 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.

2.3 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 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.

Prior to the development of this platform, strategies for directed evolution in human cells relied almost entirely on screens or cytotoxic cell-based selections to identify active biomolecule variants (see Chapter 1).[10] These prior methods severely limit library sizes, the number of cycles of evolution, and the scope of biomolecules that can be evolved. To solve these problems, 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[11] or bacteriophage.[12] 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.
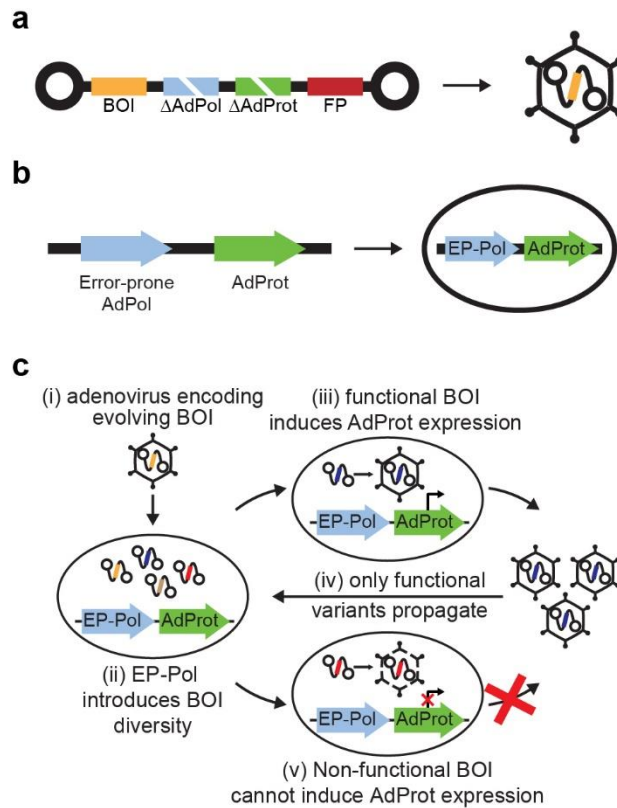
Figure 2.1 Human-cell-based directed evolution platform overview. (a) Schematic of an engineered adenovirus type 5 vector in which genes for adenoviral polymerase (AdPol) and protease (AdProt) are removed and a gene encoding the biomolecule of interest (BOI) for directed evolution is introduced, as well 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.

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.[13-14] Furthermore, 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 2.1a**). 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 2.1b**). Directed evolution experiments in this system rely on simply serially passaging the BOI-encoding adenovirus while mutagenesis and selection continuously occur (**Figure 2.1c**).

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 enhance our understanding of metazoan evolutionary biology and our ability to rapidly generate and optimize biomolecular therapeutics.

2.4 Results

2.4.1 Mutagenesis. Adenovirus type 5 relies on its own DNA polymerase, AdPol, for replication of its double-stranded DNA genome.[15] The high-fidelity AdPol has an estimated mutation rate of $\sim 1.3 \times 10^{-7}$ mutations per base per viral passage, based on deep sequencing experiments performed by Sanjuán and co-workers.[16] 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 2.2a**).[17] In the φ29 bacteriophage polymerase,[18] an AdPol homologue, 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[19] to inactivate the AdPol gene encoded by the adenovirus type 5 genome via an internal deletion. Next, we stably transduced HEK293A cells with an HA-tagged version of either wild-type AdPol or EP-Pol. 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 EP-Pol in trans (**Figure 2.2b**). Furthermore, we observed that EP-Pol and wild-type AdPol both supported robust ΔAdPol adenovirus replication.
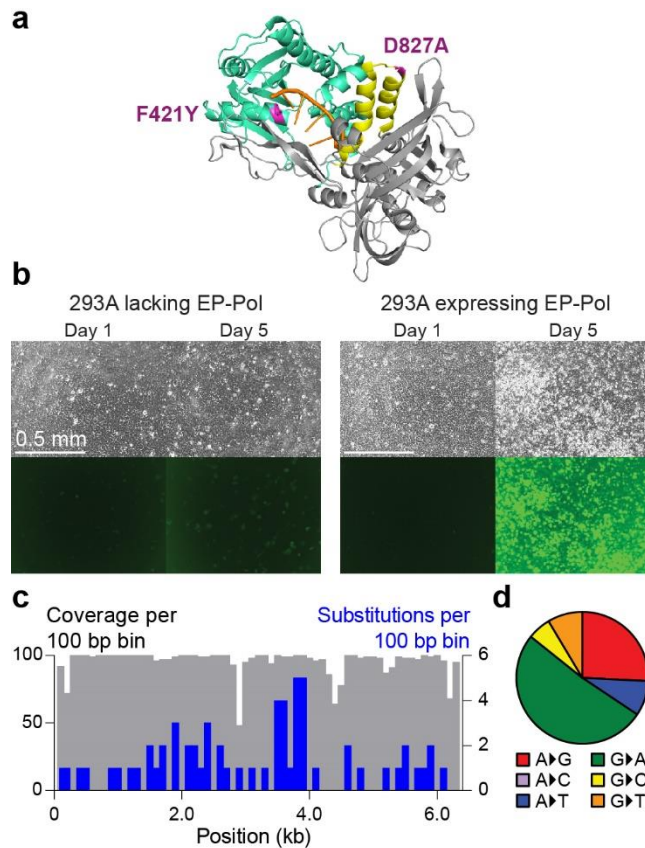
Figure 2.2 An error-prone adenoviral polymerase induces mutations throughout the adenoviral genome. (a) Crystal structure of the φ29 DNA polymerase (PDB ID 1XHZ),[18] an AdPol homologue, with the locations of homologous mutations used to create EP-Pol shown in magenta. (b) Either parental HEK293A cells 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. (c) ΔAdPol adenovirus (AdGLΔPol) was serially passaged on EP-Pol expressing cells for 10 passages, after which a 6.5 kb genomic fragment was amplified from a ~27 clone pool. Illumina sequencing identified mutations throughout the amplified region. For substitution values, see **Table 2.1**. (d) Mutational spectrum of EP-Pol evaluated by next-generation sequencing.

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 27 viral clones (**Figure 2.2c** and **Table 2.1**). 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 2.2d**).

Previously, the same sequencing procedure was carried out for wild-type AdPol.[17] 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,[20] however, the mutation rate of EP-Pol is enhanced ~280-fold. Thus, EP-Pol greatly increases the number of mutations introduced per viral passage. On the basis of these analyses, the mutation rate of adenovirus passaged in the presence of EP-Pol is similar to that of highly mutagenic RNA viruses that can readily evolve on laboratory time scales.[16, 21-22]

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 the 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. 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.

Table 2.1 Substitution values of an error-prone adenoviral polymerase.

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.

2.4.2 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.[23] AdProt has vital functions in viral uncoating, DNA replication, viral maturation, and cell entry.[24-25] Importantly, AdProt is a "late gene" expressed mainly after DNA replication of the adenoviral genome.[25] 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 (**Table 2.2**). 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 2.3**). 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).[26-28] Wild-type tTA (tTAwt) binds its endogenous operator, with a consensus sequence of 5′-CCTATCAGTGATAGA-3′, to induce downstream gene transcription. A tTA variant (tTAmut) that is incapable of binding to the endogenous operators has also been reported.[28] tTAmut instead possesses enhanced affinity for the mutant 5′-CC**CG**TCAGTGA**CG**GA-3′ operator. We engineered ΔAdProtΔAdPol adenoviruses that expressed either tTAwt and mCherry (tTAwt.mCherry) or tTAmut and GFP (tTAmut.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, **Table 2.2**). In this cell line, tTAwt.mCherry adenovirus should be able to strongly induce AdProt and propagate, whereas tTAmut.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 to prevent propagation and therefore more accurately quantify the resulting viral populations.

Table 2.2 Table of cell lines.

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.
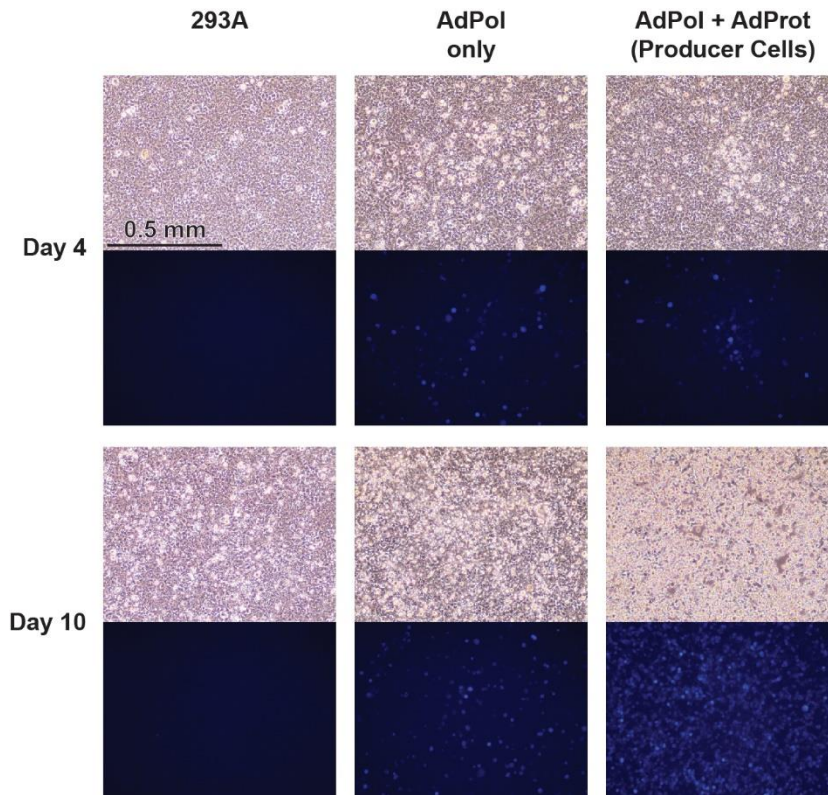
Figure 2.3 Double trans-complementation of adenoviral polymerase and adenoviral protease. ΔAdProtΔAdPol adenovirus was used to infect HEK293A cells, AdPol-expressing cells, or producer cells (**Table 2.2**) at a low multiplicity of infection (<0.5). The infection was monitored over 10 days. The parental 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.

To test our hypothesis that AdProt induction could enable enrichment of active over inactive BOI variants, we co-infected tTAwt.mCherry and tTAmut.GFP using a total MOI of ~0.25 in selector cells at initial ratios of 1:100 or 1:1000 (**Figure 2.4a**). 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 tTAwt.mCherry adenovirus was enriched at least 40–50-fold over the tTAmut.GFP adenovirus (**Figure 2.4b**). Furthermore, across three rounds of passaging, the tTAwt.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, on the basis of AdProt transcript levels, we were able to access up to a 14-fold change in AdProt expression (**Figure 2.4c**). Notably, we observed a strong correlation between dox concentration and viral titer over this entire order of magnitude range (**Figure 2.4d**).

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 tTAwt.mCherry-infected cells with various concentrations of the vinyl sulfone AdProt inhibitor shown in **FIGURE 2.4e**,[29] we found that the inhibitor could reduce the resulting infectious titer of the tTAwt.mCherry virus up to 650-fold, providing ready access to a dynamic range of selection pressure between 2 and 3 orders of magnitude in size. Moreover, we observed that the AdProt inhibitor even further reduced infectious titer in the presence of dox (**Figure 2.4e**), highlighting the capacity of AdProt inhibition to strengthen the selection pressure at a variety of baseline AdProt expression levels. Notably, the vinyl sulfone AdProt inhibitor was not toxic at the concentrations used (**Figure 2.4f**).
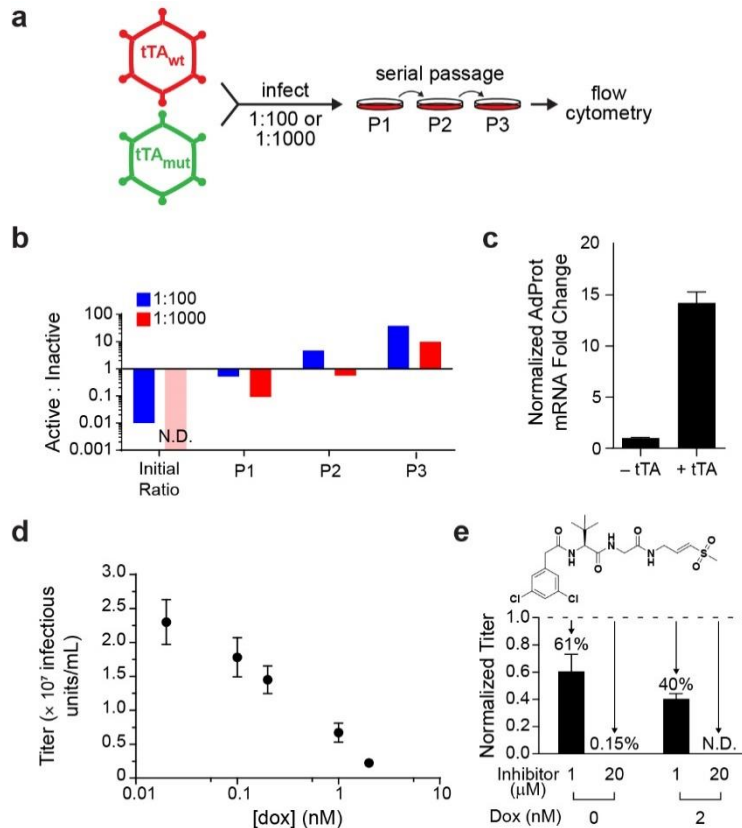
Figure 2.4 Competition experiment between virally encoded active and inactive genes of interest. (a) Schematic of the competition experiment between adenoviruses that carry the gene for wild-type tetracycline transactivator (tTAwt.mCherry) versus viruses that carry inactive tTA (tTAmut.GFP). HEK293A cells stably encoding the gene for adenoviral protease (AdProt) under control of the endogenous tTA operator (termed "selector" cells, **Table 2.2**) were infected at an initial ratio of 1:100 or 1:1000 tTAwt.mCherry to tTAmut.GFP viruses. Viral media were 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 tTAwt.mCherry adenoviruses relative to tTAmut.GFP adenoviruses rapidly increased with each passage. The initial ratio of the 1:1000 sample (labeled N.D., not detectable) was not experimentally quantifiable owing to the low amount of tTAwt.mCherry adenovirus present and was therefore derived via dilution of the 1:100 initial ratio. (c) 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. (d) Titer of tTAwt.mCherry adenovirus resulting from infection of selector cells 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 titering infections. (e) 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 tTAwt.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 resulting from treatment with 20 µM AdProt inhibitor and 2 nM dox was too low to be accurately detected (N.D., not detectable).

2.4.3 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 tTAwt.mCherry virus in the presence of dox in selector cells 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 2.5a**). In trial 1, we tuned the selection pressure over time, increasing the dox concentration from 2 nM to 20 µM. In trial 2, we kept the selection pressure constant and high by maintaining the dox concentration at 200 nM. 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 (**Table 2.2**) containing GFP under control of the endogenous tTA operator in the presence of dox. The 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 2.5b**). 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.
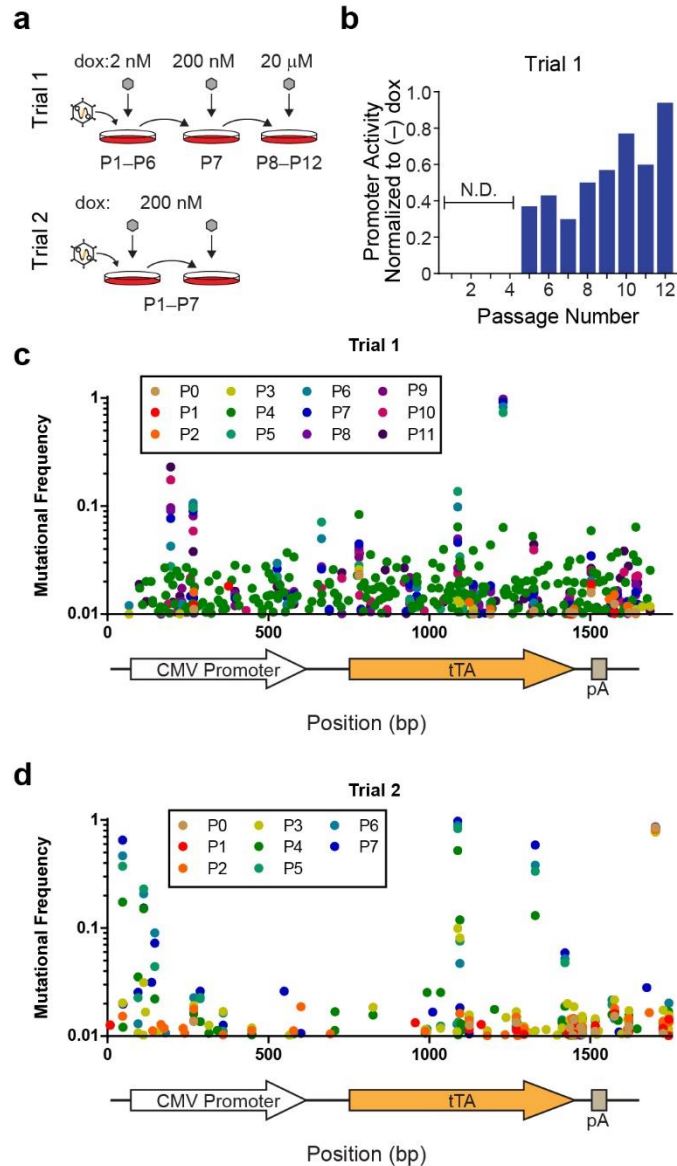
Figure 2.5 Proof-of-principle directed evolution experiment. (a) Serial-passaging schemes for evolving functional tTA variants that gain dox resistance in human cells. Two approaches to selection pressure were used, either with increasing dox concentrations (trial 1) or with a constant, moderate dox 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–d) Non-reference allele frequencies for all mutations observed at ≥1% frequency over the course of the directed evolution experiment for (c) trial 1 and (d) trial 2. A schematic of the sequenced amplicon is shown below the x-axis for reference.
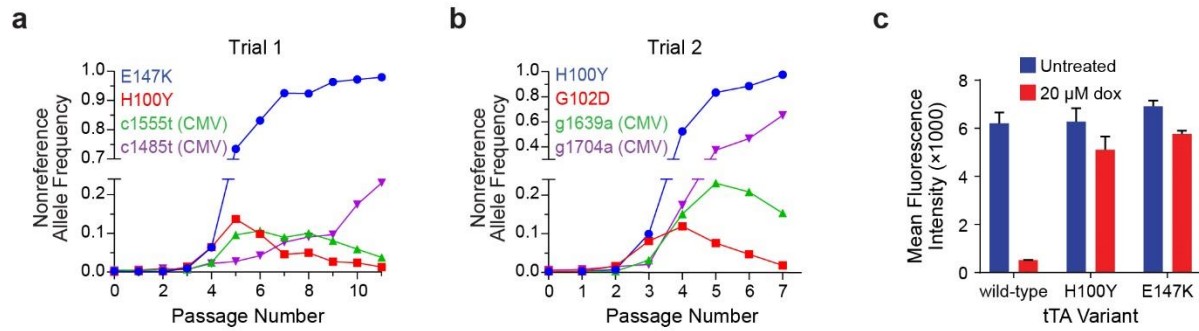
Figure 2.6 Mutational trajectories of proof-of-principle evolution experiments. (a) Mutational trajectories of four mutations identified in trial 1, including two non-coding mutations in the CMV promoter upstream of the tTA gene. (b) Mutational trajectories of four abundant mutations identified in trial 2, including two non-coding mutations in the CMV promoter upstream of the tTA gene. (c) Plasmids encoding the tTA variants fixed in trials 1 and 2 were transfected, along with the pLVX-TRE3G.eGFP reporter plasmid, into HEK293A cells with or without dox. Two days later, flow cytometry was performed to examine tTA variant activity in the presence versus the absence of 20 µM dox.

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 2.5c**). In trial 2, 43 mutations attained ≥1% frequency by passage 4 (**Figure 2.5d**). 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 (**Figure 2.6a–b**). Both mutations observed were previously reported to confer dox resistance in tTA,[30] 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 2.6c**). 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 time points: (i) either early, when ~75% of the 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 2.7**). 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.
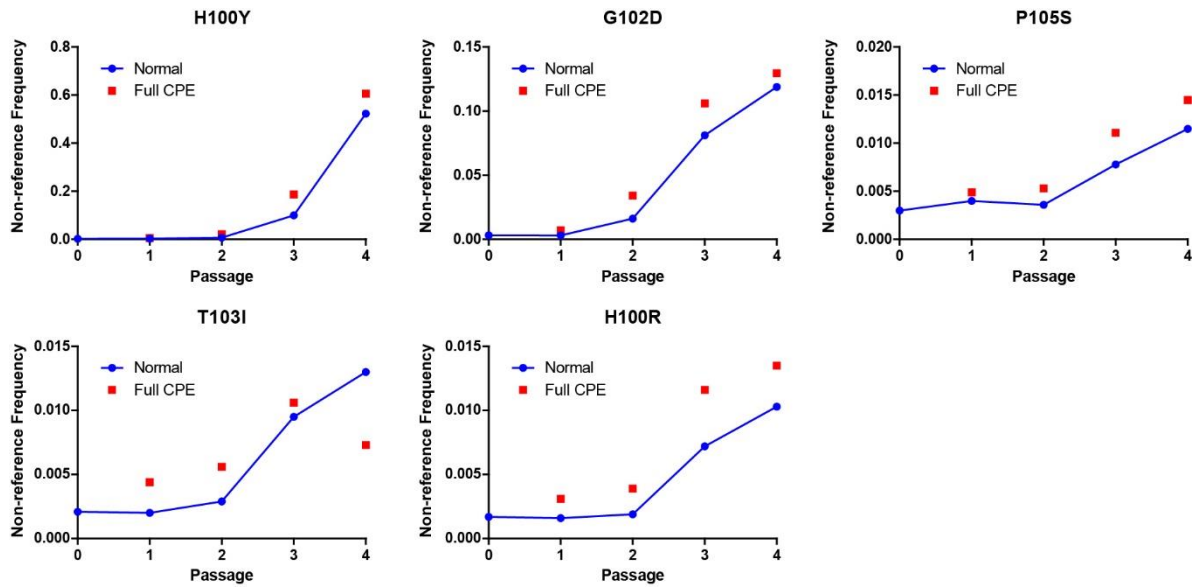
Figure 2.7 Comparison of early and late harvesting protocols. Comparison of mutation frequencies in trial 2 using the early harvesting protocol and late harvesting for each passage after full CPE was attained. Five doxycycline-resistant variants that reached a frequency of 1% by passage 4 are shown.
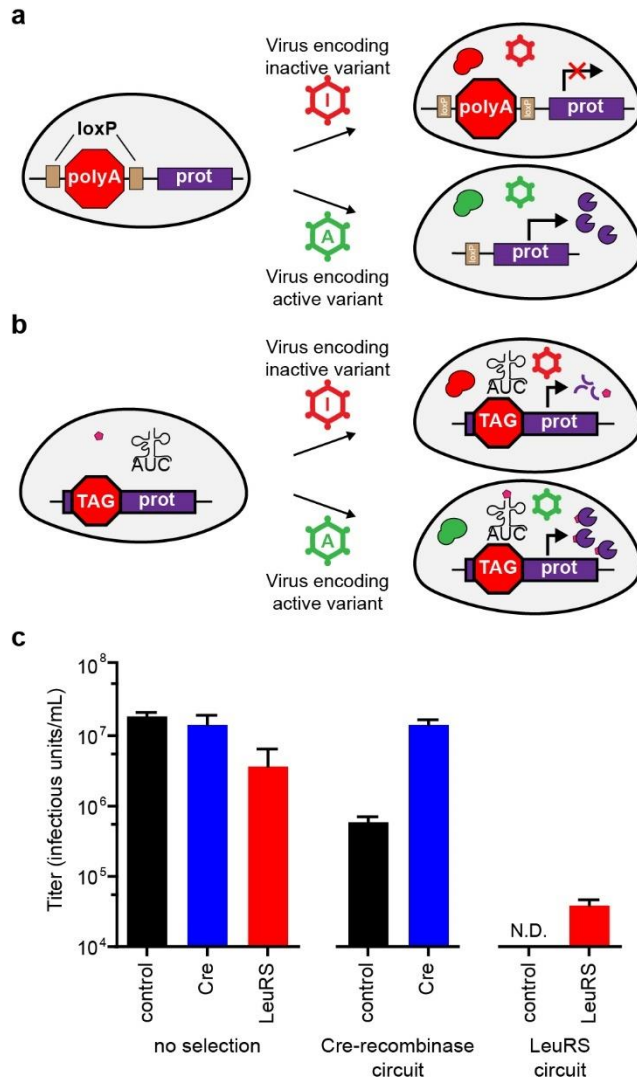
Figure 2.8 Demonstration of multiple additional selection circuits. (a) Selection circuit designed for AdProt-based selection of Cre recombinase activity. A floxed SV40-polyA terminator signal prevents transcription of AdProt unless Cre deletes the terminator by recombination at target loxP sites. (b) 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)2Term.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, tTAwt.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.

2.4.4 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 (**Figure 2.8a,b**).[31-32] We transfected both the Cre-recombinase (Cre; **Figure 2.8a**) and leucyl-tRNA synthetase (LeuRS; **Figure 2.8b**) 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 2.8c**). 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 2.8c**). 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 aminoacyl-tRNA synthetase activities.

2.5 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 2.1c). 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's 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,[33] followed by plasmid transfection into a producer cell line that constitutively expresses both AdPol and AdProt to generate a starter adenovirus population (**Figure 2.9**). Directed evolution then simply involves serial passaging of the adenovirus on user-defined "selector cells".
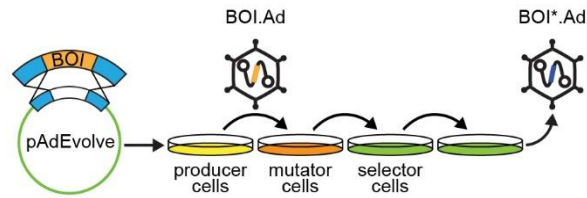
Figure 2.9 A generalized scheme for directed evolution in mammalian cells. A gene encoding a biomolecule of interest (BOI) is first inserted into pAdEvolve. "Producer" cells are used to generate ΔAdProtΔAdPol adenoviruses carrying the BOI gene. If desired, the BOI gene can be mutated prior to selection 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.

In developing this platform, we chose to use adenovirus rather than a natively mutagenic RNA virus owing to adenovirus's 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.[17, 23, 34] 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[13] 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 allow the necessary facile manipulation of the adenoviral genome.[19, 35]

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[17] 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 time-scales.[22, 36-37] 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.[17, 38]

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 the library size. The 6.5 kb genomic region we sequenced (**Figure 2.2c**) 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 2-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 that in the neighboring pIX, IVa2, and pTP genes, suggesting that such selection only impacts our mutation rate at most 2-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 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[39-40] or Cas9-fusion proteins[41-43] still display significant off-target genetic modification.[39-40, 44-45] 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 large dynamic range of selection. However, we observed that AdProt was able to modulate viral titers ~10-fold in response to protease levels. Importantly, we also showed 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 reintroduction 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 aminoacyl-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,[12, 49] Cas9,[50] GPCRs,[51] and beyond. Indeed, a context where this platform should 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 MutaT7[52] or CRISPR-X,[42] could further focus mutations only to the BOI gene, increase the 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.[12] 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.

2.6 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.

2.7 Materials and Methods

2.7.1 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. Adenoviral constructs were engineered using ccdB recombineering, as previously described[30] and further optimized by our laboratory.[19] Primers were obtained from Life Technologies and Sigma-Aldrich (**Table 2.3**). The tripartite leader sequence (TPL) gene block was obtained from Integrated DNA Technologies (**Table 2.3**).

2.7.2 Cell culture. Cells were cultured at 37 °C and 5% CO2(g). New cell lines were derived from a parent HEK293A cell line (Thermo Fisher) 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 were cultured in 50 µg/mL hygromycin (Thermo Fisher) to stably maintain transgenes, while the selector and phenotyping cell lines were cultured in 1 µg/mL puromycin (Corning) for the same purpose.

Table 2.3 Table of primers.

Name	Sequence
BamHI.AdProt.Forward	5'-aaaaaaggatccaccatgggctccagtgag-3'
SalI.AdProt.Reverse	5'-aaaaagtcgacttacatgttttttcaagtgacaaaaagaag-3'
TPL.GA.Forward	5'-atcgcctggagaattcactctcttccgcatcgct-3'
TPL.GA.Reverse	5'-ctcactggagcccattgcgactgtgactggttag-3'
TPL Gene Block	5'-aaaaaagaattcactctcttccgcatcgctgtctgcgagggccagctgttgggctcgcggttgaggacaa actcttcgcggtctttccagtactcttggatcggaaacccgtcggcctccgaacaggtactccgccgccgagggacctgagcgagtccgcatcgaccggatcgg aaaacctctcgagaaaggcgtctaaccagtcacagtcgcaggatccttttt-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'-aaaaaatctagattacatgttttttcaagtgacaaaaagaag-3'
TPL.AdProt.GA.Forward	5'-tggagaaggatccgcactctcttccgcatcgct-3'
TPL.AdProt.GA.Reverse	5'-atctagagccggcgcttacatgttttttcaagtgacaaaaagaag-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'-aaaaaatctagattacatgttttttcaagtgacaaaaagaag-3'
pENTR1A.AdProt.FLAG.Forward	5'-taatctagacccagctttcttgtacaaagttggcattataag-3'
pENTR1A.AdProt.FLAG.Reverse	5'-agaaagctgggtctagattacttatcgtcgtcatccttgtaatccatgtttttcaagtgacaaaaagaagt ggcg-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'-atacaaaactacataagaccccccaccttatatattctttcccacccttaaccctcatcagtgccaacatag taag-3'
E1.kanccdB.Reverse	5'-aataagaggaagtgaaatctgaataattttgtgttactcatagcgcgtaaccgctcattaggcgggc-3'
Cre.kanccdB.Forward	5'-tggaactaatcatatgtggcctggagaaacagctaaagtgcgaaagcggcccgctcattaggcgggc-3'
Cre.kanccdB.Reverse	5'-cgcgaacaaatgtggtatggctgattatgatcctctagagataattctagccctcatcagtgccaacatag taag-3'
E1.CMV.Promoter.Forward	5'-atacaaaactacataagaccccccaccttatatattctttcccacccttaagccacgcccacagatatacgc gttgacattg-3'
E1.bGH.polyA.Reverse	5'-aataagaggaagtgaaatctgaataattttgtgttactcatagcgcgtaatagaagccatagagcccac-3'
E4.kanccdB.Forward	5'-caaaaaacccacaacttcctcaaatcgtcacttccgttttcccacgttaccccctcatcagtgccaacatag taag-3'
E4.kanccdB.Reverse	5'-agtaacttgtatgtgttgggaattgtagttttcttaaaatgggaagtgacccgctcattaggcgggc-3'
E4.SV40.Promoter.Forward	5'-caaaaaacccacaacttcctcaaatcgtcacttccgttttcccacgttacttctgtggaatgtgtgtcagtta ggg-3'
E4.SV40.polyA.Reverse	5'-agtaacttgtatgtgttgggaattgtagttttcttaaaatgggaagtgacctctagctagaggtcgacggta tac-3'
Pol.kanccdB.Forward	5'-tcccgcgcttcttggaactttacattgtgggccacaacatcaacggccctccctcatcagtgccaacatag taag-3'
Pol.kanccdB.Reverse	5'-ggcacctcggaacggttgttaattacctgggcggcgagcacgatctcgtcccgctcattaggcgggc-3'
delPol.Forward	5'-gcgcggccttccggagcgaggtgtgggtgagcgcaaaggtgtccctgaccatgaccagcatgaagg gcacgagctgcttcccaaaggcccccatccaag-3'
delPol.Reverse	5'-cttggatggggcctttgggaagcagctcgtgccctttcatgctggtcatggtcagggacacctttgcgctc acccacacctcgctccggaaggccgcgc-3'
AdProt.kanccdB.Forward	5'-ggcaacgccacaacataaagaagcaagcaacatcaacaacagctgccgcccctcatcagtgcca acatagtaag-3'
AdProt.kanccdB.Reverse	5'-tacaaataaaagcatttgcctttattgaaagtgtctctagtacattatttccgctcattaggcgggc-3'
delAdProt.Forward	5'-ggcaacgccacaacataaagaagcaagcaacatcaacaacagctgccgccaaataatgtactaga gacactttcaataaaggcaaatgctttttatttgta-3'
delAdProt.Reverse	5'-tacaaataaaagcatttgcctttattgaaagtgtctctagtacattatttggcggcagctgttgttgatgttgcttg Cttctttatgttgtggcgttgcc-3'

2.7.3 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: 15 µg of RRE, 6 µg of REV, 3 µg of VSVG, and 15 µg of transfer vector using 60 µL of Lipofectamine 2000 (Thermo Fisher). Cultures were maintained in 5 mL total volume of Opti-MEM (Gibco) during the transfection. After 8 h, the medium was exchanged for fresh DMEM. After 48 h, the medium was harvested and centrifuged for 5 min at 3200 × 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 medium was exchanged for fresh DMEM. After an additional 48 h, the medium was exchanged again for DMEM containing appropriate antibiotics to select stable cell lines.

2.7.4 Adenovirus production. Adenoviruses were produced by transfecting a PacI (New England BioLabs)-linearized vector into appropriate trans-complementing HEK293A cells. A 24 µg sample of PacI-linearized adenovirus vectors mixed with 144 µL of polyethylenimine (Polysciences) in 1 mL of Opti-MEM (Gibco) was added to a 15 cm plate of producer cells (~3×10^7 cells). The medium 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, full 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 3200 × g for 15 min and the supernatant was stored at −80 °C.

2.7.5 Mutagenesis rate determination. The mutagenic potential of AdPol variants was evaluated following a previously reported protocol.[17] Briefly, a polymerase-deleted adenovirus type 5, AdGLΔPol, was subjected to 10 serial passages on cultures of 911 cells[53] expressing EP-Pol to accumulate mutations. After 10 serial passages, 911 cells expressing wild-type AdPol were infected in duplicate six-well plates at ~50 plaque-forming units/well to amplify pools of 50 viral clones for sequencing. On the basis of 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 would 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.[17]

2.7.6 AdPol and AdProt trans-complementation assays. The day before beginning the assay, a six-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) to permit observation of the presence or absence of a spreading infection. EP-Pol trans-complementation was tested by monitoring CFP.ΔAdPol.GFP adenovirus infection on EP-Pol-expressing HEK293A cells. Pictures were taken with an Olympus U-TB190 microscope. AdProt and AdPol double trans-complementation was tested by monitoring ΔAdProtΔAdPol adenovirus infection on producer cells. Pictures were taken with a Nikon Eclipse TE200 microscope.

2.7.7 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. Two to three days post-infection, the cells were washed once with medium, stained with 0.2 μg/mL DAPI (4′,6-diamidino-2-phenylindole, Thermo Fisher), 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 the cells were infected.

2.7.8 Competition experiments. A confluent dish of ~15 million selector cells was infected with either a 1:100 or 1:1000 mixture of tTAwt.mCherry:tTAmut.GFP adenovirus. The plates were monitored for the appearance of spreading infection, defined by fluorescent plaques, every 24 h. One day after the observation of spreading infection, 1 mL of medium was transferred to a new semi-confluent dish (~1 × 10^7 cells) of selector cells for the next passage, and 2 mL of medium 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 tTAwt.mCherry and tTAmut.GFP viruses was determined by taking the ratio of cells expressing only mCherry and only GFP.

2.7.9 AdProt inhibitor experiments. A confluent 12-well plate of selector cells (~4 × 10^5 cells/well) was infected with tTAwt.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, 1, and 20 μM) in the absence or presence of 2 nM doxycycline (dox; Sigma-Aldrich). After 6 days, the

medium and cells were harvested and subjected to three freeze/thaw cycles and analyzed by flow cytometry (see above).

2.7.10 AdProt inhibitor toxicity assay. A 96-well plate of HEK293A cells was treated with the AdProt inhibitor at concentrations up to 20 µM for 5 days. 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.

2.7.11 RT-qPCR on selector cells. A confluent 12-well plate of selector cells ($\sim 4 \times 10^5$ cells/well) was transfected with 1.25 µg of the pTet-Off Advanced vector (Takara Bio) using 7.5 µL of polyethylenimine (Polysciences) and 100 µL of Opti-MEM. Two days later, the 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 a high-capacity cDNA reverse transcription kit (Applied Biosystems). qPCR analysis for AdProt (**Table 2.3**; primers: AdProt.Forward and AdProt.Reverse) and the housekeeping gene RPLP2 (**Table 2.3**; primers: RPLP2.Forward and RPLP2.Reverse) was performed on a LightCycler 480 II (Roche). AdProt transcript levels were normalized to untransfected selector cells.

2.7.12 Dox dose–response experiment. A confluent 24-well plate of selector cells ($\sim 1.5 \times 10^5$ cells/well) was infected with tTAwt.mCherry adenovirus (MOI ≈ 5). After 4 h, the cells were washed with DMEM (Corning) and dox was added at the indicated concentrations (0, 0.02, 0.1, 0.2, 1, or 2 nM). After 5 days, the medium and cells were harvested and subjected to three freeze/thaw cycles, followed by analysis of titers using flow cytometry.

2.7.13 Continuous evolution workflow. Before initiation of directed evolution, 500 µL of a tTAwt.mCherry adenovirus was amplified on mutator cells to create a diverse viral population. After 5 days, full 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 ($\sim 1 \times 10^7$ cells/plate) were infected with 250, 500, or 1000 µL of the amplified virus in the presence of dox. The 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 1000 µL of adenovirus-containing medium from the previous round by direct transfer without a freeze/thaw step. A 2 mL volume of medium was saved in Eppendorf tubes and stored at −80 °C for future analysis. In trial 2, an additional medium harvest was performed after a 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.

2.7.14 Measuring promoter activity of viral populations. To follow changes in promoter activity developing during trial 1, phenotyping cells were plated in a 96-well plate at ~40000 cells/well. The next day, 30 µL of medium from passages 1–12 was used to infect two rows of the 96-well plate. Medium was removed 5 h post-infection and replaced with medium containing 0 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.

2.7.15 Next-generation sequencing of evolved tTA variants. Using a viral DNA isolation kit (NucleoSpin Virus; Macherey-Nagel), DNA was harvested from 200 µL of the medium 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). A 250 bp paired-end sequencing was performed on a MiSeq (Illumina). Sequencing reads were aligned to the amplicon sequence, which was derived from the tTAwt.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. (56,57) Each position within the tTA gene and CMV promoter had at least 1000-fold coverage.

2.7.16 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 polyethylenimine (Polysciences) and 100 µL of Opti-MEM. After 8 h post-transfection, medium was exchanged, and 20 µM dox was added. After 48 h post-transfection, the cells were analyzed by flow cytometry (see above for sample preparation). Promoter activity was calculated on the basis of the mean fluorescence intensity of GFP fluorescence, backgated for only mCherry-expressing cells.

2.7.17 Testing of recombinase and synthetase selection circuits. HEK-293A cells expressing 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 polyethylenimine in 100 µL of Opti-MEM. For the AdProt(STOP) circuit, 0.5 µg was co-transfected with 0.5 µg of pLeu-

tRNA.GFP(STOP). The medium was changed 4 h post-transfection. The next day, transfected wells were infected with either the relevant BOI virus (Cre.Ad for (LoxP)2Term.AdProt, and LeuRS.Ad for AdProt(STOP)) or TTAwt.mCherry as a negative control at an MOI of 5. The cells were washed three times with medium 3 h post-infection. After 4 days, the medium and cells were harvested and subjected to three freeze/thaw cycles, followed by analysis of titers using flow cytometry.

2.8 References

1.   Packer, M. S.; Liu, D. R., Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **2015,** *16* (7), 379-94.
2.   Gai, S. A.; Wittrup, K. D., Yeast surface display for protein engineering and characterization. *Curr. Opin. Struct. Biol.* **2007,** *17* (4), 467-73.
3.   Romero, P. A.; Arnold, F. H., Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **2009,** *10* (12), 866-76.
4.   Shaner, N. C.; Campbell, R. E.; Steinbach, P. A.; Giepmans, B. N.; Palmer, A. E.; Tsien, R. Y., Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat. Biotechnol.* **2004,** *22* (12), 1567-72.
5.   Branon, T. C.; Bosch, J. A.; Sanchez, A. D.; Udeshi, N. D.; Svinkina, T.; Carr, S. A.; Feldman, J. L.; Perrimon, N.; Ting, A. Y., Efficient proximity labeling in living cells and organisms with TurboID. *Nat. Biotechnol.* **2018,** *36* (9), 880-887.
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.* **2018,** *7* (7), 1722-1729.
7.   Zetsche, B.; Gootenberg, J. S.; Abudayyeh, O. O.; Slaymaker, I. M.; Makarova, K. S.; Essletzbichler, P.; Volz, S. E.; Joung, J.; van der Oost, J.; Regev, A.; Koonin, E. V.; Zhang, F., Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **2015,** *163* (3), 759-71.
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* **2011,** *18* (5), 619-30.
9.   Piatkevich, K. D.; Jung, E. E.; Straub, C.; Linghu, C.; Park, D.; Suk, H. J.; Hochbaum, D. R.; Goodwin, D.; Pnevmatikakis, E.; Pak, N.; Kawashima, T.; Yang, C. T.; Rhoades, J. L.; Shemesh, O.; Asano, S.; Yoon, Y. G.; Freifeld, L.; Saulnier, J. L.; Riegler, C.; Engert, F.; Hughes, T.; Drobizhev, M.; Szabo, B.; Ahrens, M. B.; Flavell, S. W.; Sabatini, B. L.; Boyden, E. S., A robotic multidimensional directed evolution approach applied to fluorescent voltage reporters. *Nat. Chem. Biol.* **2018,** *14* (4), 352-360.
10.   Hendel, S. J.; Shoulders, M. D., Directed evolution in mammalian cells. *Nat. Methods* **2021,** *18* (4), 346-357.
11.   Das, A. T.; Zhou, X.; Vink, M.; Klaver, B.; Verhoef, K.; Marzio, G.; Berkhout, B., Viral evolution as a tool to improve the tetracycline-regulated gene expression system. *J. Biol. Chem.* **2004,** *279* (18), 18776-82.
12.   Esvelt, K. M.; Carlson, J. C.; Liu, D. R., A system for the continuous directed evolution of biomolecules. *Nature* **2011,** *472* (7344), 499-503.
13.   Lucher, L. A., Abortive adenovirus infection and host range determinants. *Curr. Top. Microbiol. Immunol.* **1995,** *199 (Pt 1),* 119-52.
14.   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 Ther.* **1997,** *4* (3), 258-63.
15.   Hoeben, R. C.; Uil, T. G., Adenovirus DNA replication. *Cold Spring Harb. Perspect. Biol.* **2013,** *5* (3), a013003.
16.   Sanjuán, R.; Nebot, M. R.; Chirico, N.; Mansky, L. M.; Belshaw, R., Viral mutation rates. *J. Virol.* **2010,** *84* (19), 9733-9748.
17.   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.; Hoeben, R. C., Directed adenovirus evolution using engineered mutator viral polymerases. *Nucleic Acids Res.* **2011,** *39* (5), e30.
18.   Kamtekar, S.; Berman, A. J.; Wang, J.; Lazaro, J. M.; de Vega, M.; Blanco, L.; Salas, M.; Steitz, T. A., Insights into strand displacement and processivity from the crystal structure of the protein-primed DNA polymerase of bacteriophage phi29. *Mol. Cell* **2004,** *16* (4), 609-18.

19.   Papa, L. J., 3rd; Shoulders, M. D., Genetic engineering by DNA recombineering. *Curr Protoc Chem Biol* **2019,** *11* (3), e70-e70.

20.   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.* **2016,** *12* (11), e1006013.

21.   Davis, J. N.; van den Pol, A. N., Viral mutagenesis as a means for generating novel proteins. *J. Virol.* **2010,** *84* (3), 1625-30.

22.   Phillips, A. M.; Gonzalez, L. O.; Nekongo, E. E.; Ponomarenko, A. I.; McHugh, S. M.; Butty, V. L.; Levine, S. S.; Lin, Y. S.; Mirny, L. A.; Shoulders, M. D., Host proteostasis modulates influenza evolution. *eLife* **2017,** *6*, e28652.

23.   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 Ther.* **2002,** *9* (18), 1238-46.

24.   Greber, U. F.; Webster, P.; Weber, J.; Helenius, A., The role of the adenovirus protease in virus entry into cells. *EMBO J.* **1996,** *15* (8), 1766-1777.

25.   Webster, A.; Leith, I. R.; Hay, R. T., Activation of adenovirus-coded protease and processing of preterminal protein. *J. Virol.* **1994,** *68* (11), 7292-300.

26.   Gossen, M.; Bujard, H., Tight control of gene expression in mammalian cells by tetracycline responsive promoters. *Proc. Natl. Acad. Sci. U.S.A.* **1992,** *89*, 5547-5551.

27.   Loew, R.; Heinz, N.; Hampf, M.; Bujard, H.; Gossen, M., Improved Tet-responsive promoters with minimized background expression. *BMC Biotechnol.* **2010,** *10*, 81.

28.   Krueger, M.; Scholz, O.; Wisshak, S.; Hillen, W., Engineered Tet repressors with recognition specificity for the tetO-4C5G operator variant. *Gene* **2007,** *404* (1-2), 93-100.

29.   Grosche, P.; Sirockin, F.; Mac Sweeney, A.; Ramage, P.; Erbel, P.; Melkko, S.; Bernardi, A.; Hughes, N.; Ellis, D.; Combrink, K. D.; Jarousse, N.; Altmann, E., Structure-based design and optimization of potent inhibitors of the adenoviral protease. *Bioorg. Med. Chem. Lett.* **2015,** *25* (3), 438-43.

30.   Hecht, B.; Muller, G.; Hillen, W., Noninducible Tet repressor mutations map from the operator binding motif to the C terminus. *J. Bacteriol.* **1993,** *175* (4), 1206-10.

31.   Meinke, G.; Bohm, A.; Hauber, J.; Pisabarro, M. T.; Buchholz, F., Cre recombinase and other tyrosine recombinases. *Chem. Rev.* **2016,** *116* (20), 12785-12820.

32.   Italia, J. S.; Zheng, Y.; Kelemen, R. E.; Erickson, S. B.; Addy, P. S.; Chatterjee, A., Expanding the genetic code of mammalian cells. *Biochem. Soc. Trans.* **2017,** *45* (2), 555-562.

33.   Hartley, J. L.; Temple, G. F.; Brasch, M. A., DNA cloning using *in vitro* site-specific recombination. *Genome Res.* **2000,** *10* (11), 1788-1795.

34.   Russell, W. C., Update on adenovirus and its vectors. *J. Gen. Virol.* **2000,** *81* (Pt 11), 2573-604.

35.   Wang, H.; Bian, X.; Xia, L.; Ding, X.; Muller, R.; Zhang, Y.; Fu, J.; Stewart, A. F., Improved seamless mutagenesis by recombineering using ccdB for counterselection. *Nucleic Acids Res.* **2014,** *42* (5), e37.

36.   Meyerhans, A.; Cheynier, R.; Albert, J.; Seth, M.; Kwok, S.; Sninsky, J.; Morfeldt-Månson, L.; Asjö, B.; Wain-Hobson, S., Temporal fluctuations in HIV quasispecies *in vivo* are not reflected by sequential HIV isolations. *Cell* **1989,** *58* (5), 901-910.

37.   O'Loughlin, T. L.; Greene, D. N.; Matsumura, I., Diversification and specialization of HIV protease function during *in vitro* evolution. *Mol. Biol. Evol.* **2006,** *23* (4), 764-72.

38.   Myers, N. D.; Skorohodova, K. V.; Gounder, A. P.; Smith, J. G., Directed evolution of mutator adenoviruses resistant to antibody neutralization. *J. Virol.* **2013,** *87* (10), 6047-50.

39.   Wang, C. L.; Harper, R. A.; Wabl, M., Genome-wide somatic hypermutation. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (19), 7352-6.

40.   Wang, C. L.; Yang, D. C.; Wabl, M., Directed molecular evolution by somatic hypermutation. *Protein Eng. Des. Sel.* **2004,** *17* (9), 659-64.

41.   Ma, Y.; Zhang, J.; Yin, W.; Zhang, Z.; Song, Y.; Chang, X., Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* **2016,** *13* (12), 1029-1035.

42.   Hess, G. T.; Fresard, L.; Han, K.; Lee, C. H.; Li, A.; Cimprich, K. A.; Montgomery, S. B.; Bassik, M. C., Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* **2016,** *13* (12), 1036-1042.

43.   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* **2016,** *533* (7603), 420-4.

44.   Meng, F. L.; Du, Z.; Federation, A.; Hu, J.; Wang, Q.; Kieffer-Kwon, K. R.; Meyers, R. M.; Amor, C.; Wasserman, C. R.; Neuberg, D.; Casellas, R.; Nussenzweig, M. C.; Bradner, J. E.; Liu, X. S.; Alt, F. W., Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell* **2014,** *159* (7), 1538-48.

45.   Kim, D.; Lim, K.; Kim, S. T.; Yoon, S. H.; Kim, K.; Ryu, S. M.; Kim, J. S., Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat. Biotechnol.* **2017,** *35* (5), 475-480.

46.   Hubbard, B. P.; Badran, A. H.; Zuris, J. A.; Guilinger, J. P.; Davis, K. M.; Chen, L.; Tsai, S. Q.; Sander, J. D.; Joung, J. K.; Liu, D. R., Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nat. Methods* **2015,** *12* (10), 939-42.

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.* **2014,** *5*, 5352.

48.   Badran, A. H.; Guzov, V. M.; Huai, Q.; Kemp, M. M.; Vishwanath, P.; Kain, W.; Nance, A. M.; Evdokimov, A.; Moshiri, F.; Turner, K. H.; Wang, P.; Malvar, T.; Liu, D. R., Continuous evolution of *Bacillus thuringiensis* toxins overcomes insect resistance. *Nature* **2016,** *533* (7601), 58-63.

49.   Carlson, J. C.; Badran, A. H.; Guggiana-Nilo, D. A.; Liu, D. R., Negative selection and stringency modulation in phage-assisted continuous evolution. *Nat. Chem. Biol.* **2014,** *10* (3), 216-22.

50.   Hu, J. H.; Miller, S. M.; Geurts, M. H.; Tang, W.; Chen, L.; Sun, N.; Zeina, C. M.; Gao, X.; Rees, H. A.; Lin, Z.; Liu, D. R., Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **2018,** *556* (7699), 57-63.

51.   Armbruster, B. N.; Li, X.; Pausch, M. H.; Herlitze, S.; Roth, B. L., 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.* **2007,** *104* (12), 5163-8.

52.   Moore, C. L.; Papa, L. J., 3rd; Shoulders, M. D., A processive protein chimera introduces mutations across defined DNA regions *in vivo*. *J. Am. Chem. Soc.* **2018,** *140* (37), 11560-11564.

53.   Fallaux, F. J.; Kranenburg, O.; Cramer, S. J.; Houweling, A.; Van Ormondt, H.; Hoeben, R. C.; Van Der Eb, A. J., Characterization of 911: a new helper cell line for the titration and propagation of early region 1-deleted adenoviral vectors. *Hum. Gene Ther.* **1996,** *7* (2), 215-22.

54.   Ossowski, S.; Schneeberger, K.; Clark, R. M.; Lanz, C.; Warthmann, N.; Weigel, D., Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Res.* **2008,** *18* (12), 2024-33.

Chapter 3: Virus-based continuous directed evolution of GPCRs in mammalian cells

3.1 Author contributions

Samuel J. Hendel, Amanuela A. Mengiste, and Matthew D. Shoulders designed the experiments. S.J.H. and A.A.M. performed the cloning, virus and cell line engineering, characterization assays, directed evolution experiments. Sequencing data was analyzed by Vincent L. Butty at the BioMicroCenter at MIT.

3.2 Identifying targets to evolve with our directed evolution platform

3.2.1 A framework for choosing targets. In Chapter 2, I described the development of a virus-based continuous directed evolution platform for mammalian cells. Our next task was to move beyond proof-of-principle by evolving high valued, never-before-evolved targets directly in mammalian cells.

Adapting our virus-based directed evolution experiments toward new targets requires engineering and characterization steps. While there are limitless targets for which directed evolution in mammalian cells would be impactful, practicability limits the number of independent targets that can be meaningfully pursued in parallel. Furthermore, some target families are more impactful than others, often those that are important for human health or are the target for therapeutics. Lastly, our directed evolution platform evolution would be particularly useful for evolving targets which have been evolved in lower organisms and the evolved products have failed to function properly in mammalian cells (see Chapter 1). Careful analysis of which directed evolution targets to pursue is essential to maximizing the probability of success and the relative impact of any directed evolution campaign.

In assessing both the impact and practicality of evolving proteins in mammalian cells, three criteria can be useful to assess which directed evolution targets should be pursued:

1. Successful evolution would impact the study of mammalian biology or the treatment of disease.
2. The selection circuit should be applicable to a broad class of evolvable activities.
3. Directed evolution of the desired activities in lower organisms is impossible or insufficient.

Mammalian cells present unique challenges. Successfully evolving a particular target in our system requires careful engineering of a mammalian host cell with a robust selection circuit for continuous directed evolution, which can be notoriously difficult. Moreover, the large genomes and the relative inefficiency of mammalian cell-compatible genome editing technologies makes engineering selections in mammalian cells even more challenging. Therefore, creating a selection circuit that has the broadest applicability possible will maximize the potential impact of the engineering effort.

Mammalian cells also present unique opportunities. Mammalian cells have numerous sophisticated signaling pathways that are not present in lower organisms. Creating selection circuits based on endogenous pathways in mammalian cells can allow a target to be evolved in conjunction with crucial intracellular activities. Furthermore, endogenous signal transduction pathways in mammalian cells often overlap, involving the same secondary messengers upon activation of a different pathways.

3.2.2 GPCRs: An ideal target class for directed evolution in mammalian cells. G-protein coupled receptors (GPCRs) are an ideal target class for directed evolution in mammalian cells. From their diverse functions to their importance to human health, GPCRs present exciting and vast opportunities for a directed evolution campaign.

GPCRs are seven-transmembrane receptors that transmit myriad extracellular signals into physiological responses (**Figure 3.1a, b**).[1-2] The largest class of membrane proteins in the human genome and the third largest class overall, GPCRs are ubiquitous in the nervous and immune systems[3] and play vital roles in our senses of smell,[4] and taste,[5] and sight.[6] Although the roughly 800 GPCRs in the human genome[7] are predicted to share broadly similar structural features, they can be activated by a huge diversity of perturbations, from proteins and hormones to ions and photons.

Studying GPCR-based signaling has yielded crucial insights into the development of therapeutics.[3] GPCRs are the targets of ~34% of FDA approved pharamaceuticals (as of 2017),[8] including those that treat hypertension, allergies, and chronic pain. Furthermore, studying mutations in GPCRs can provide crucial insights into the function of GPCRs. Naturally occurring mutations[9] within GPCR genes can cause disease,[10] and pharmacogenomic testing of GPCRs can inform pharmaceutical treatment using a personalized medicine approach. Although structural studies of GPCRs are notoriously challenging, new advances in cryo-EM techniques are now elucidating complex GPCR functionalities and enabling improved GPCR engineering.[11]

Engineered GPCRs play an important role in biomedical research.[12] GPCRs engineered to respond to new ligands go by many names, including receptors activated solely by synthetic ligands (RASSLs), therapeutic receptor-effector complexes (TRECs), neoceptors, and designer receptors exclusively activated by designer drugs (DREADDs).[13] One of the most widely-used systems is a DREADD engineered from the human M3 receptor (hM$_3$R) to be activated by clozapine-*N*-oxide (CNO; **Figure 3.1c**).[14] This DREADD, termed hM3D, is activated by CNO, an otherwise pharmalogically inert derivative of clozapine, which is a ligand for hM$_3$R and other GPCRs. DREADDs that bind CNO are widely used, particularly to study complex neurological processes with spatiotemporal control.[15-16]
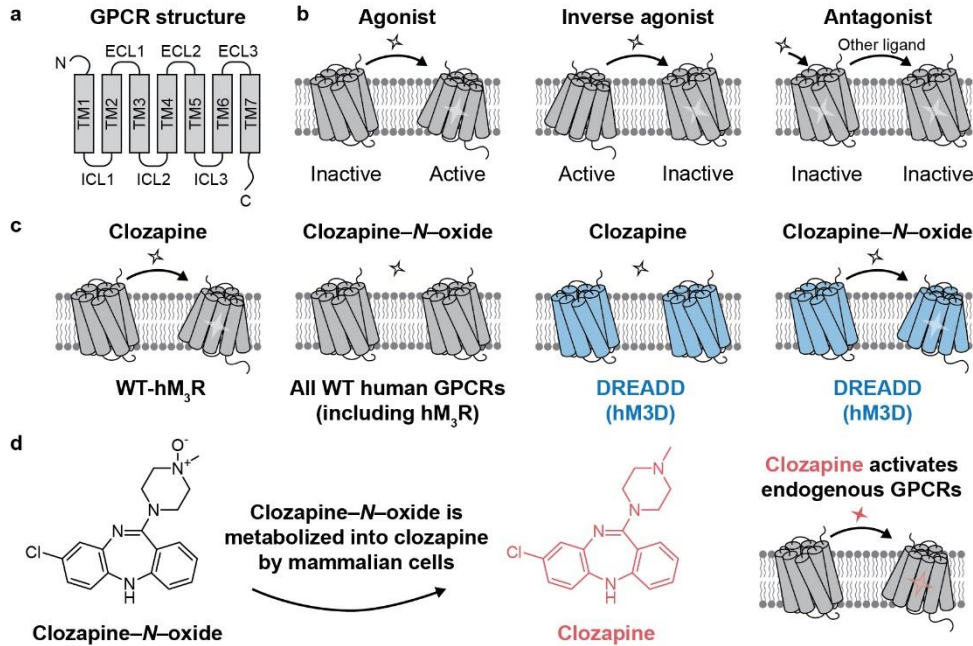
Figure 3.1 Overview of GPCR structure and function. (a) G protein-coupled receptors (GPCRs) are membrane proteins involved in numerous signal transduction pathways. GPCRs have an extracellular N-terminal domain, seven transmembrane domains (TM1–7), three extracellular loops (ECL1–3), three intracellular loops (ICL1–3), and an intracellular C-terminal domain. (b) Perturbagens such as drugs most commonly impact GPCR function through agonism, inverse agonism, and antagonism, although other modes of activation such as allosteric regulation may also occur. (c) GPCRs can be engineered to respond to synthetic ligands, such as designer receptors exclusively activated by designer drugs (DREADDs).[14] Clozapine is an agonist for wild-type human M3 receptor (hM$_3$R); however, clozapine-N-oxide (CNO) is not an agonist for hM$_3$R or any other human GPCR. The first DREADD, hM$_3$D, is variant of hM$_3$R that is activated upon treatment with CNO but responds to neither clozapine nor acetylcholine, the natural agonist of hM$_3$R. (d) One major problem with CNO is that it is converted to clozapine *in vivo*. Clozapine then can act throughout the body on wild-type targets such as hM$_3$R and other GPCRs.[19]

However, currently available engineered GPCRs have several undesirable properties (**Figure 3.1d**).[17] CNO has difficulty penetrating the blood-brain barrier,[18] requiring researchers to use high concentrations of the water-insoluble CNO *in vivo* to observe proper effects. The problems associated with the use of high concentrations of water-insoluble ligands are compounded by the fact that CNO has been shown to be metabolized into clozapine *in vivo*.[19] Clozapine, as opposed to CNO, does accumulate in the brain[20] and can activate multiple endogenous GPCRs.[21] Other engineered GPCRs show similar problems, such as low activity or poor expression in mammalian cells,[22] leading the field to search for new and improved technologies.[23]

Directed evolution of GPCRs had previously been performed in yeast, particularly for the development of hM3D. However, the directed evolution campaign to produce the first DREADDs created numerous evolved variants that functioned properly in yeast but failed to retain their functionality in mammalian cells.[14, 24] Furthermore, although yeast have two endogenous GPCRs in their genome,[25] the GPCR signaling network in yeast is much less complex than mammalian systems[26] and does not contain the numerous downstream signaling components necessary for function in mammalian cells.

Evolving GPCRs directly in mammalian cells would solve the problems of yeast-based directed evolution, and could tremendously benefit GPCR research. Receptors could be designed to be activated by truly pharmacologically inert ligands, to have higher affinity to a given designer ligand, to attain higher activity, or to activate GPCR pathways for which there are currently no engineered GPCRs. Furthermore, even the simplest evolution of evolving uninduced or constitutive activity would yield important biochemical information[9, 27-28] into GPCR structure and mechanisms of activation.[9]
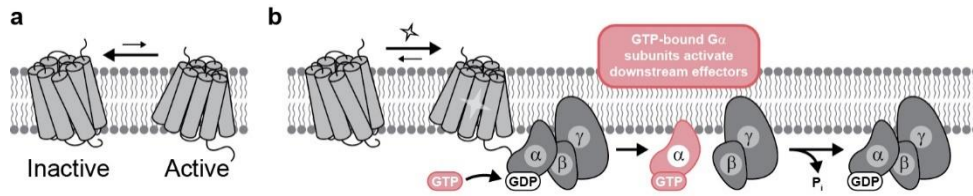
Figure 3.2 Molecular biology of GPCR signaling pathways. (a) GPCRs are in equilibrium between active and inactive conformations. (b) Canonically, when an agonist or perturbation pushes this equilibrium toward the GPCR's active state, a heterotrimeric $G\alpha\beta\gamma$ complex is enabled to interact with the active GPCR to promote nucleotide exchange in the $G\alpha$ subunit. The GTP-bound $G\alpha$ subunit will dissociate from the heterotrimeric complex and activate downstream effectors.

3.3 Directed evolution of GPCRs using adenovirus-based directed evolution

3.3.1 Molecular biology of GPCRs. Canonically,[2] GPCRs function as allosteric membrane proteins, such that they exist in an equilibrium of active and inactive conformations (**Figure 3.2a**). As opposed to inactive conformations, active conformations promote the binding of intracellular transducers, including heterotrimeric guanine nucleotide binding proteins (G proteins; **Figure 3.2b**).[29] Heterotrimeric G proteins consist of one α, one β, and one γ subunit, of which there are 18, 5, and 12 different genes in humans, respectively. When in the active conformation, a GPCR can bind the G protein heterotrimer and induce an allosteric change that increases the $G\alpha$ rate of exchange of GDP and GTP. Once bound to GTP, the $G\alpha$ subunit dissociates from the $G\beta\gamma$ subunit and interacts with downstream effector proteins. After an active-state lifetime ranging from milliseconds to minutes, the $G\alpha$ subunits hydrolyze the bound GTP to form GDP, which inactivates the $G\alpha$ subunit and promotes the reconstitution of the $G\alpha\beta\gamma$ heterotrimer. Notably, there are counterexamples to this generalized pathway—active GPCRs can also interact with arrestins and other intracellular factors[30] and agonists can affect GPCR allostery without affecting agonist binding or downstream effectors.[31] Furthermore, $G\beta\gamma$ heterodimers likely have important roles in signaling, although elucidating the roles of specific $G\beta$ and $G\gamma$ subunits has proven challenging.

Acknowledging these important considerations, GPCR signaling pathways are largely defined by the $G\alpha$ subunit that is activated by the GPCR and the effector protein that is associated with the active $G\alpha$ subunit. These effector proteins influence downstream pathways by changing the concentration of secondary messengers. For example, active $G\alpha_s$ and $G\alpha_{olf}$ subunits interact with adenylyl cyclases to promote the cyclization of ATP into the secondary messenger cyclic AMP (cAMP).[32] Knowledge of which $G\alpha$ subunits an active GPCR associates with can be important for determining which signaling pathways the GPCR activates.

3.3.2 Development of a GPCR-based directed evolution scheme. GPCR-based signaling pathways often connect extracellular signals that activate GPCRs to gene expression through sophisticated transcriptional activation networks.[33] These transcriptional activation networks are most commonly controlled by a regulatable transcription factor associated with a particular G-protein-coupled pathway. Importantly for our purposes, these transcription factors bind specific response elements in DNA that can induce expression of downstream genes.

Response elements can be engineered and adapted to couple GPCR activity to the expression of any desired protein. The expression of reporter proteins such as luciferase or fluorescent proteins has been linked to numerous signaling cascades, including multiple GPCR-

based signaling pathways. By encoding selection markers such as AdProt under the control of these response elements, we envisioned that we could create a selection circuit based on the activity of any protein involved in the signaling pathway.

3.3.3 The cyclic AMP signaling pathway. One of the best characterized and most widely used GPCR-based response elements in mammalian cells is the cAMP response element (CRE) involved in the cAMP signaling pathway (**Figure 3.3**). An important secondary messenger in many living organisms, cAMP can be generated in cells by adenylyl cyclases activated by GTP-bound $G\alpha_s$ and $G\alpha_{olf}$ subunits. Levels of cAMP can also be elevated upon $G\alpha_q/G\alpha_{11}$ activation.[34] Once generated, cAMP can impact diverse activities in the cell, including activating protein kinase A (PKA), nucleotide exchange factors, and cyclic-nucleotide-gated ion channels. When active, PKA phosphorylates substrates including the cAMP-response element binding protein (CREB),[35] a transcription factor that induces expression of genes under control of CRE after phosphorylation.[36]

Intracellular increases in cAMP can be monitored by CRE-based reporter systems, allowing researchers genetic tools to robustly record intracellular activation.[37] The diversity of protein activities regulated by cAMP and the existence of CRE-based reporter systems make this pathway an appealing avenue toward making a GPCR-based selection circuit. Simply encoding our system's[38] selection marker under control of CRE could enable continuous directed evolution of proteins involved in this pathway, allowing unprecedented throughout and flexibility to engineer diverse and important biological activities.
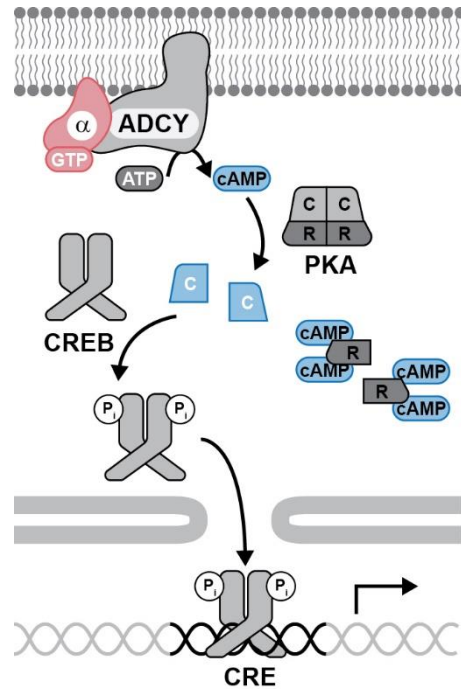
Figure 3.3 GPCR-based activation of the cAMP response element. GTP-bound active Gα subunits can interact with an adenylyl cyclase (ADCY) to induce the conversion of ATP to the secondary messenger cyclic AMP (cAMP). In the presence of elevated levels of cAMP, the two regulatory (R) subunits of the heterotetramer protein kinase A bind cAMP and dissociate from the catalytic (C) subunits. The catalytic subunits phosphorylate numerous protein targets including the cAMP response element binding protein (CREB), which translocates to the nucleus upon phosphorylation and induces transcription of genes downstream of genes controlled by the cAMP response element (CRE).

3.3.4 Developing a selection circuit for virus-based directed evolution of GPCRs. In order to ensure that a selection circuit can be based on $G\alpha_s$-coupled GPCRs, we needed to validate that the CRE can be used to induce reporter genes upon activation of a model GPCR.

CRE was identified as a 30 bp sequence common to 5′ untranslated regions of cAMP responsive genes.[39] CREB and other CREB-like transcription factors specifically bind a highly conserved 8 bp region of DNA (TGACGTCA),[40] a fact that later engineering efforts took advantage of in generating CRE-responsive reporter systems. A CRE-based luciferase reporter plasmid was previously developed to monitor cAMP production in HEK293A cells (**Figure 3.4a**).[37] This reporter plasmid encodes CRE preceding firefly luciferase fused to two human codon optimized protein degradation tags (termed hPEST and hCL1) to decrease protein expression.[41-42] The plasmid also encodes a hygromycin resistance marker, which can be used to select for cells that received the plasmid.

Our first step was to confirm that CRE-based reporter plasmid functions properly in our HEK293A cells—namely that the expression of luciferase increase upon activation of GPCRs that increase intracellular cAMP. We initially chose two GPCRs with potent small molecule activators: muscarinic acetylcholine receptor M3 (CHRM3),[43] which is activated by carbachol, and follicle stimulating hormone receptor (FSHR),[44] which is activated by follicle stimulating hormone (FSH). We observed a dose-dependent increase in luciferase expression for both GPCRs (**Figure 3.4 b–c**), confirming that these GPCRs increase intracellular cAMP upon activation. Furthermore, the lack of increase luciferase expression upon treatment with carbachol or FSH alone shows that the exogenously expressed GPCRs were responsible for the observed effect (**Figure 3.4d**).

In our system, the evolving gene of interest is encoded in the evolving adenoviral genome. Although GPCRs have been previously encoded and delivered in adenovirus, GPCR-encoding adenoviruses deleted for AdPol and AdProt have not previously been reported. Therefore, our next objective was to show that we could encode GPCRs in our engineered adenoviral vector and demonstrate functional GPCR expression. Using our previously developed AdEvolveDEST system (Chapter2), we generated AdPol- and AdProt-deleted viruses encoding human codon optimized CHRM3 and FSHR genes. Using cells stably expressing the CRE-based luciferase reporter plasmid (Xfect-CRE-Luc cells), we showed that viral delivery of CHRM3 and FSHR can activate CRE and that luciferase expression is dependent upon addition of the GPCRs' respective agonists (**Figure 3.4e–f**).
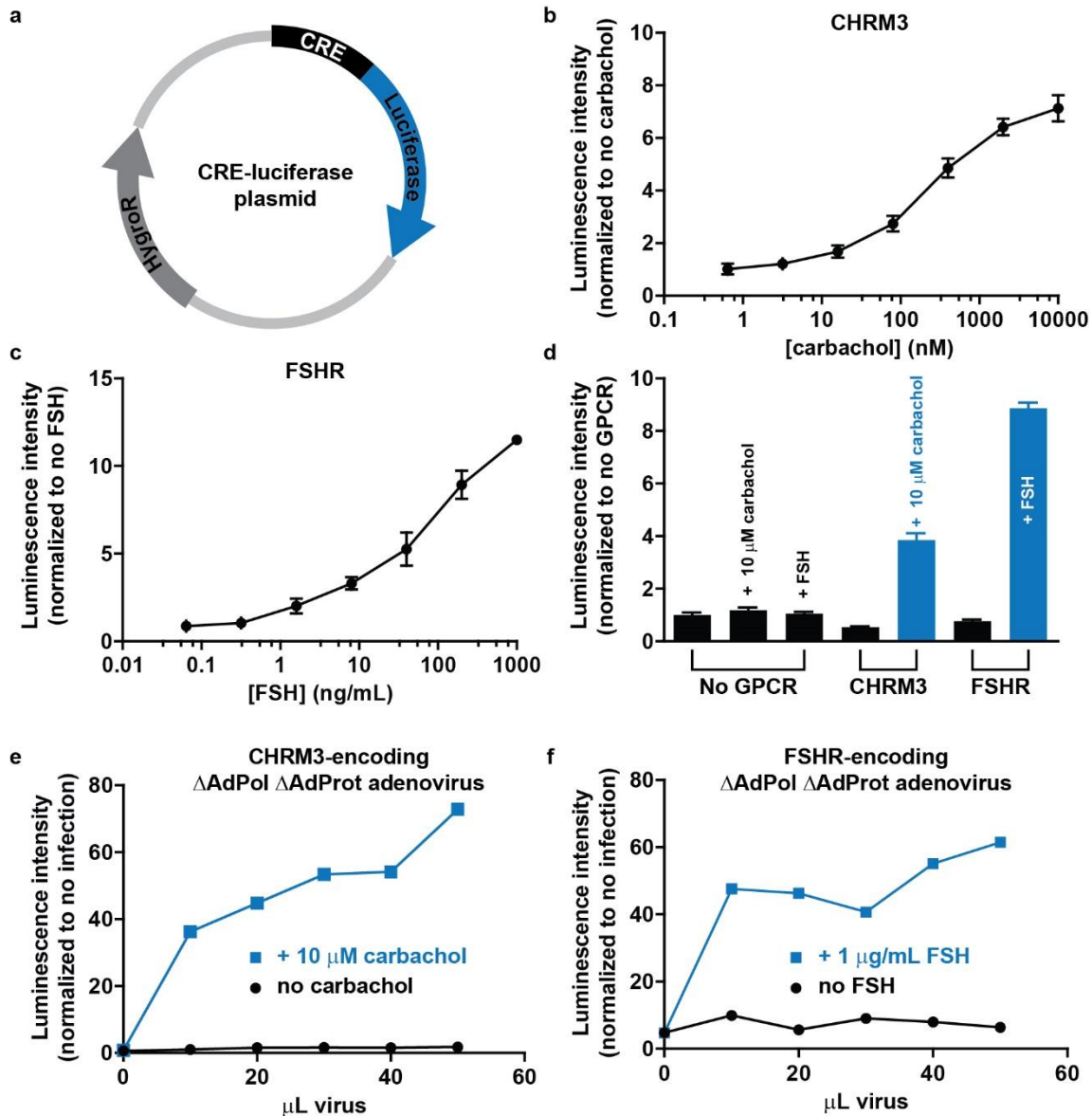
Figure 3.4 Testing a CRE-based luciferase reporter plasmid. (a) Overview of plasmid map for the CRE-Luciferase reporter plasmid, which encodes luciferase under control of the cAMP response element (CRE). The reporter plasmid also encodes a constitutively expressed hygromycin resistance marker (HygroR), which renders cells resistant to treatment with the cytotoxic antibiotic hygromycin. (b) Dose-dependent activation of CRE by carbachol-based agonism of the muscarinic acetylcholine receptor M3 (CHRM3), as evidenced by an increase in luciferase-based luminescence. (c) Dose-dependent activation of CRE by follicle stimulating hormone (FSH)-based agonism of the follicle stimulating hormone receptor (FSHR), as evidenced by an increase in luciferase-based luminescence. (d) Agonist-induced CRE activation depends on the exogenous expression of the target GPCR. For FSHR and CHRM3, both GPCR and the respective agonist are necessary to observe luciferase induction from the CRE-luciferase plasmid. (e–f) Viral delivery of GPCRs also induces CRE activation upon drug treatment. CHRM3 and FSHR-encoding adenovirus are transduced into cells expressing the CRE-Luciferase reporter plasmid with or without their respective agonists, after which luciferase levels are measured through luminescence.

In our system, expression of AdProt from a mammalian host cell drives the replication of our gene-of-interest-encoding adenovirus. Therefore, we next created a lentiviral vector that encodes AdProt under control of CRE (**Figure 3.5a**), and transduced EP-Pol expressing cells with the resulting lentivirus to create Lenti-CRE-AdProt cells. Upon CHRM3 transfection in Lenti-CRE-AdProt cells, AdProt transcript levels increased upon carbachol treatment (**Figure 3.5b**). Similar increases were observed when CHRM3 was delivered using AdPol- and AdProt-deleted adenovirus (**Figure 3.5b**). These increases are comparable to the levels observed with tTA in TRE-AdProt-based selector cells (**Figure 3.5c**), which had ample selection pressure to drive a directed evolution campaign. Finally, we tested whether CHRM3-encoding viruses would preferentially replicate on Lenti-CRE-AdProt cells in the presence of carbachol. Two days post-infection, we observed markedly increased viral replication (**Figure 3.5d**), suggesting the functionality of selection based on GPCR activity.

In summary, these data show that the activity of GPCRs can be coupled to adenoviral propagation through the cAMP-induced expression of AdProt. With a functional selection circuit in hand, we were well-positioned to perform directed evolution of GPCRs in mammalian cells.

3.3.5 Parallelized directed evolution of ten GPCRs. The simplest GPCR activity to evolve using this system would be high basal or constitutive activity, such that the evolved GPCR would induce the cAMP signaling pathway without any activating perturbation. Because GPCRs are in an equilibrium between active and inactive conformational states (**Figure 3.2a**), some level of basal activity is observed for most GPCRs.[2] Mutations to key structural regions of GPCRs can increase basal activity, and knowledge of which mutations constitutively activate a GPCR can be crucial for identifying disease-causing variants,[45] particularly for the hundreds of GPCRs for which there is little to no structural information.[1, 46] Since constitutively active variants have been implicated in human disease,[9, 28] variants that show constitutive GPCR activity could be used to study the impact of such aberrant GPCR signaling on disease. Furthermore, constitutively active GPCR variants could be extremely valuable chemical biology tools when fused to or combined with other functional subunits such as destabilizing domains.[47] Practically relevant, evolution toward higher constitutive activity is straightforward with our current selection circuit, and would be a valuable milestone towards demonstrating that our platform can evolve more complicated GPCR activities.
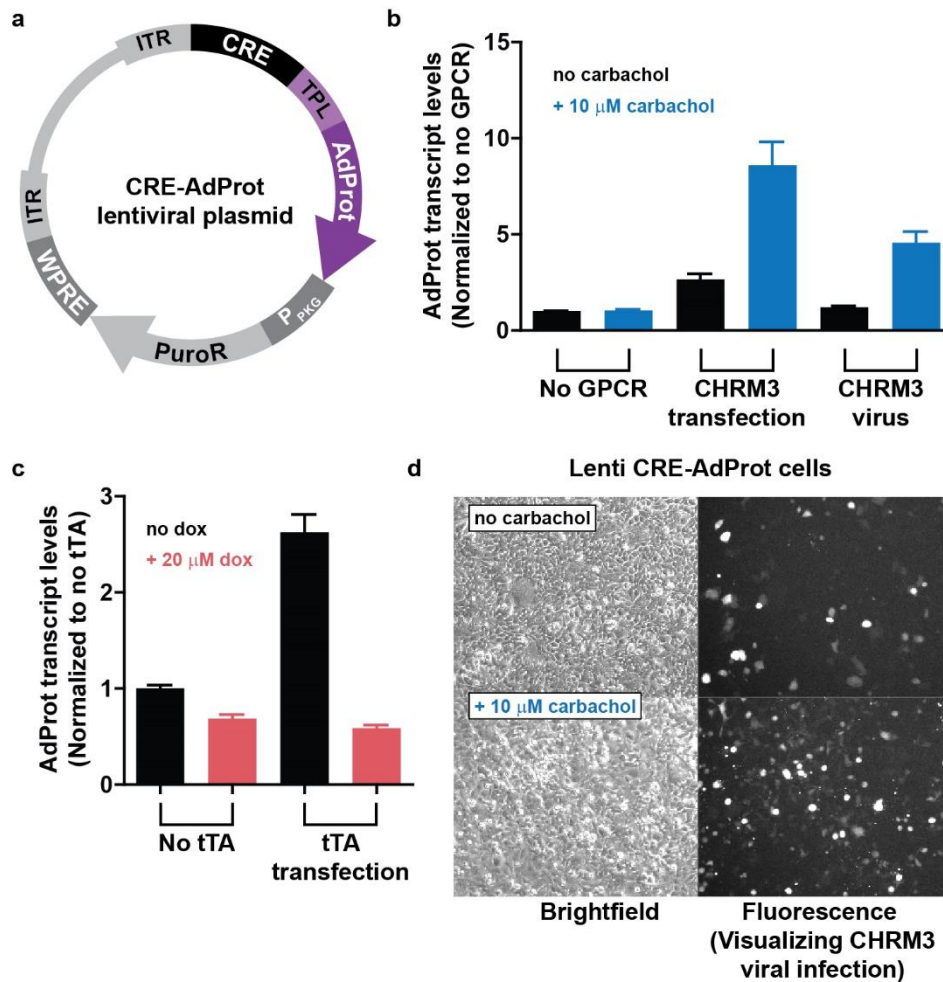
Figure 3.5 Induction of AdProt upon GPCR-activated cAMP signaling. (a) Lentiviral plasmid map for Lenti-CRE-AdProt cells. ITR, inverted terminal repeat; CRE, cAMP response element; TPL, adenoviral tripartite leader sequence; AdProt, adenoviral protease; P_{PGK}, 3-phosphoglycerate kinase promoter; PuroR, puromycin resistance marker; WPRE, woodchuck hepatitis virus posttranscriptional regulatory element. (b) Measuring AdProt transcript levels in Lenti-CRE-AdProt cells upon CHRM3 expression and activation. Cells were either transfected with a CHRM3-encoding expression plasmid or infected with a CHRM3-encoding AdPol- and AdProt-deleted adenovirus at an MOI ~ 1.0, with or without 10 µM carbachol. After 72 hours, RNA was harvested and AdProt transcript levels were measured through qPCR. (c) Measuring AdProt transcript levels in Lenti-TRE-AdProt cells (see "selection cells" in Chapter 2) upon tTA expression. Cells were transfected with a tTA-encoding expression plasmid and treated with or without 20 µM doxycycline. (d) Brightfield and fluorescence microscopy of our selection circuit. Lenti-CRE-AdProt cells were infected with a low MOI of CHRM3-encoding virus, with or without 10 µM carbachol.

One of the advantages of virus-based continuous directed evolution systems is the ease with which directed evolution experiments can be run in parallel. There are hundreds of GPCRs that activate the cAMP signaling pathway.[7, 32] A successful directed evolution campaign for even a single GPCR could be impactful—the evolution of ten in parallel would be an exciting and far-reaching demonstration of the importance and potential of our directed evolution platform.

Ten GPCRs associated with the cAMP signaling pathway were chosen for their importance to human health and mammalian biology (**Table 3.1**). We first tested the constitutive activity of the ten wild-type GPCRs (**Figure 3.6**) by infecting GPCR-encoding viruses into Xfect-CRE-Luc cells at a multiplicity of infection (MOI) of ~1.0. We observed a diversity of GPCR constitutive activity; most GPCRs showed nearly basal activity, but HTR6 showed relatively high constitutive activity. Since constitutively active GPCRs should be under relatively less selection pressure to evolve increased CRE activation, we may observe different evolutionary outcomes for HTR6, such as increased expression or improved protein folding efficiency.

Table 3.1 Ten human GPCRs associated with the cAMP signaling pathway.

GPCR	Abbreviation	Native ligand	Tissues	
β3-adrenoceptor	ADRB3	Noradrenaline	Brain	Ovaries
Vasopressin V2 receptor	AVPR2	Vasopressin	Adipose tissue	
Calcitonin receptor-like receptor	CALCRb	Adrenomedullin or CGRP	Lung	Adipose tissue
Muscarinic acetylcholine receptor M3	CHRM3	Acetylcholine	Brain	Salivary gland
Follicle stimulating hormone receptor	FSHR	Follicle stimulating hormone	Ovaries	Testes
Glucagon receptor	GCGR	Glucagon	Liver	Kidney
5-hydroxytryptamine receptor 6	HTR6	Serotonin	Brain	
Melanocortin 1 receptor	MC1R	Melanocyte stimulating hormones	Pituitary gland	Thyroid
Melanocortin 2 receptor	MC2R	Adrenocorticotropic hormone	Adrenal gland	
Melanocortin 3 receptor	MC3R	Adrenocorticotropic hormone	Brain	
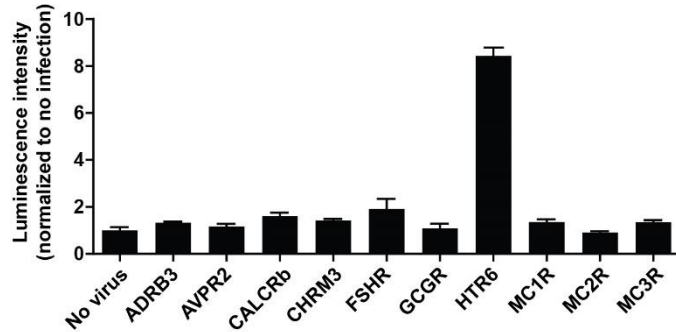
Figure 3.6 Constitutive activity of ten GPCRs used in directed evolution campaigns. Constitutive cAMP signaling of ten GPCR-encoding adenoviruses was tested via luciferase assay.
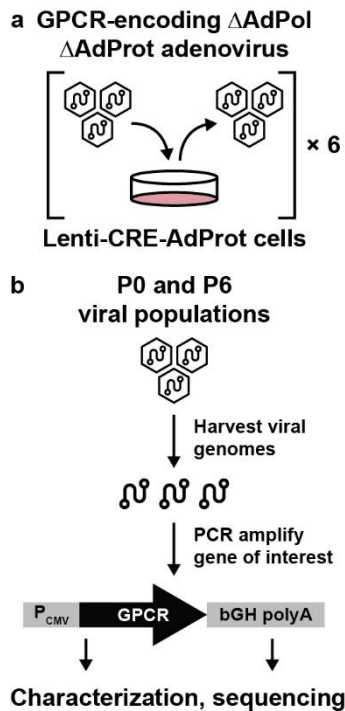
Figure 3.7 Parallelized directed evolution of ten GPCRs. (a) AdPol- and AdProt-deleted adenoviruses encoding one of ten GPCRs (Table 3.1) were passaged on Lenti-CRE-AdProt cells for six passages. (b) Viral media was collected before and after selection, from which viral genomes were harvested and populations of genes of interest were amplified through PCR for subsequent characterization and sequencing.

We then passaged these ten GPCR-encoding adenoviruses on Lenti-CRE-AdProt cells for six consecutive passages (**Figure 3.7a**). We harvested the viral DNA of all GPCRs before selection (P0) and after passage six (P6). From the prepared viral DNA, we subcloned a PCR-amplified region of the viral genome encoding the gene of interest into bacterial plasmids and sequenced eight individual colonies for all GPCRs except ADRB3, which is discussed further below (**Table 3.2**). We observed some missense mutations, but observed many wild-type sequences and, perhaps most discouragingly, numerous stop codons, insertions and deletions. These deactivating mutations were a first indication that our evolutionary circuit may be enriching for non-functional GPCRs.

We then PCR-amplified a region of the viral genome encoding the CMV promoter, the gene of interest, and the poly-adenylation tail (**Figure 3.7b**) and transfected this population into Xfect-CRE-Luc cells to observe the activity of the GPCR-expressing populations at P0 and P6 (**Figure 3.8**). Although the same pattern of activity emerged as observed in **Figure 3.6** (e.g., HTR6 showed a high level of constitutive activity), we did not observe any increase in activity from P0 to P6.

3.3.6 Next-generation sequencing data of evolved populations. To identify any enriched variants within our evolved populations, and to quantify their level of enrichment, we performed next-generation sequencing on the evolved viral genomes. Briefly, using the same region of the viral genome as in **Figure 3.7b**, we prepared amplicons for sequencing using the Nextera XT DNA Library Preparation Kit and performed next-generation sequencing on an Illumina MiSeq with 150 bp paired-end reads following standard Illumina protocols.
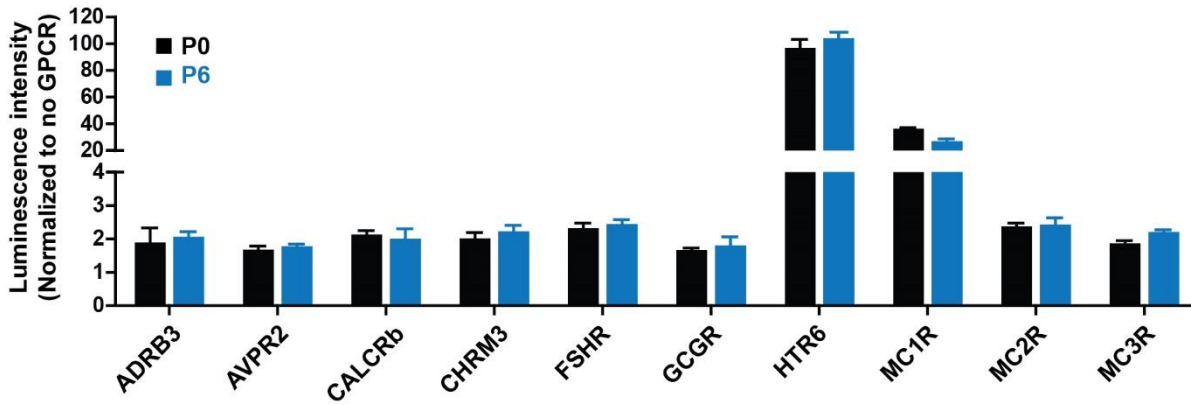
Figure 3.8 Comparing basal activity of GPCRs before and after selection. Gene of interest populations were PCR-amplified from P0 and P6 viral samples and transfected into Xfect-CRE-Luc cells to assess basal activity. Luciferase assays showed no substantial increase in cAMP signaling activity between P0 and P6.
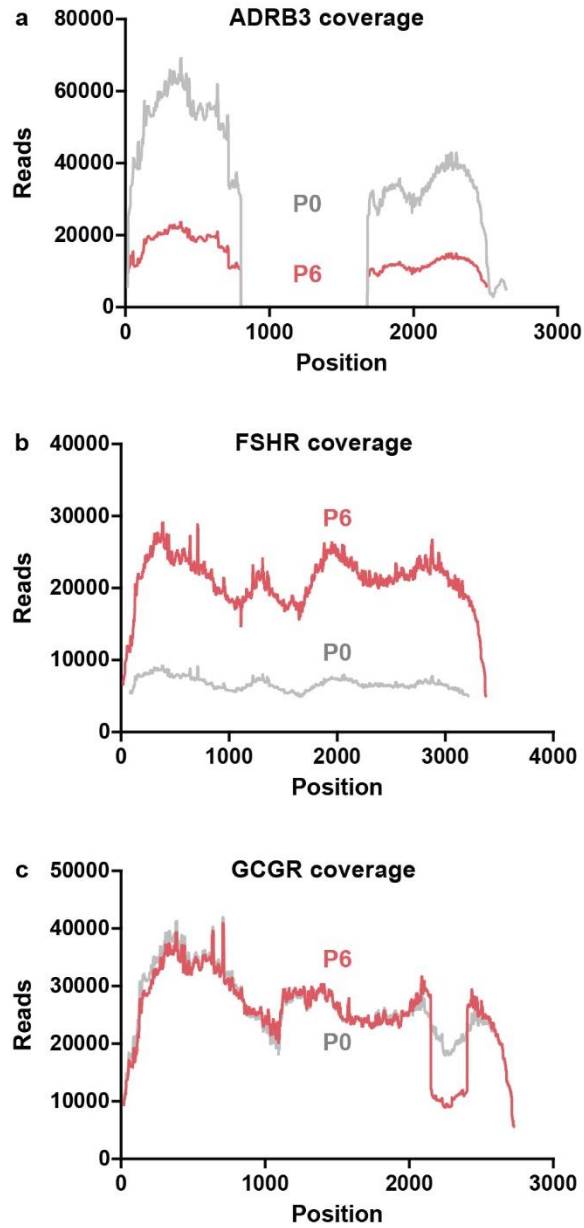
Figure 3.9 Aberrant coverage maps in next-generation sequencing of pre-selection (P0) and post-selection (P6) GPCR populations. Next-generation sequencing coverages were aberrant for three GPCRs: (a) ADRB3, (b) FSHR, (c) and GCGR.

In the vast majority of samples (except ADRB3-P6, AVPR2-P0, and FSHR-P0), we saw >85% of reads mapped to our amplified region of DNA. Even for the three samples with relatively low mapping rates, using conservative read cutoff of ~5,000 reads at each position includes all GPCR coding sequences. However, the coverage maps of three GPCRs stood out as requiring further investigation. The first coverage maps that stood out were those of ADRB3-P0 and ADRB3-P6, in which there is a large gap starting at the beginning of the coding region (**Figure 3.9a**). Upon further inspection, we observed that the gaps aligned perfectly with two KpnI cut sites, a restriction enzyme we used to clone all of the wild-type GPCRs into pENTR1A plasmids. KpnI restriction rendered the gene inoperable. This large deletion explained our failure to clone ADRB3 into bacterial plasmids (**Table 3.2**). We nonetheless continued to analyze the ADRB3 populations, as they provided an intriguing negative control (i.e., a virus that does not carry a functional gene of interest and therefore presumably is not subject to any meaningful selection pressure). The next coverage map of concern was of FSHR-P0, which had relatively low coverage compared to the rest of the samples (**Figure 3.9b**). Since the coverage across the gene was higher than 5,000 reads at every position along the gene of interest, we decided to continue with analysis of FSHR-P0, taking care to ensure that conclusions about this sample were sufficiently supported by the low coverage data. The final coverage map that stood out was GCGR-P6 (**Figure 3.9c**). We observed an abrupt drop in mapped reads near the 5′ end of the GCGR coding region, and then an abrupt increase after ~200 bp. We noted the irregularity, and proceeded with analysis.

In Chapter 2, we identified positive selected variants by looking for point mutations that were highly enriched in the population after selection. Here, in our analogous next-generation sequencing data, we first looked for highly enriched point mutations in the GPCR coding sequence (**Figure 3.10**). Two P6 samples showed mutations that were highly enriched: FSHR P6, which enriched E259STOP to 26% of the population and a synonymous R282R mutation to 24% of the population (**Figure 3.10a**); and GCGR, which enriched W68STOP to 51% of the population (**Figure 3.10b**). The enrichment of stop codons before the C-termini is surprising, since many GPCRs interact with G-proteins through their C-terminal tails. Perhaps surprisingly, the stop codon enriched in GCGR was not the same nonsense mutation observed through Sanger sequencing. The appearance of these stop codons reaffirmed our observation from Sanger sequencing that the virus seemed to be enriching for deactivating mutations.

Table 3.2 Sanger sequencing of P6 GPCRs. Insertions and deletions are marked with "+" and "–", respectively. We were unable to amplify ADRB3 due to a large internal deletion. An empty cell indicates a wild-type sequence.

Colony	AVPR2	CALCRb	CHRM3	FSHR
1	L94F, +T (287)	+A (35), C130C, A508T	+A (486)	L640I
2	+T (287), + C(347)	A422T		
3	P74L		D564D	
4	+T (166)	A375T		–A (73)
5	A72A			–G (63), –A (73), F911F, N118I
6			+AAA (486)	C25STOP
7	+T (287), + C(347)			
8				Q145R

Colony	GCGR	HTR6	MC1R	MC2R	MC3R
1			A149A		
2	–C (71), +T (368)				
3	C34S, Q131STOP, A159S				I40F
4	C34S, Q131STOP, A159S				
5			A64S		
6	+A (99), P444P	E427D			
7					
8	+A (72)	+T (283)		+T (156)	

Table 3.3 Comparing the sensitivity of next-generation sequencing for calling indels versus single base mismatches. The percentage of positions within the coding region of the MC2R P0 population that had indel or non-reference frequencies greater than various thresholds.

Threshold	Percentage of positions above threshold	
	Indels	**Non-reference**
> 0%	43.84%	100.00%
> 0.01%	16.88%	99.96%
> 0.05%	4.84%	93.42%
> 0.1%	0.04%	1.79%
> 0.5%	0.00%	0.54%
> 1%	0.00%	0.36%
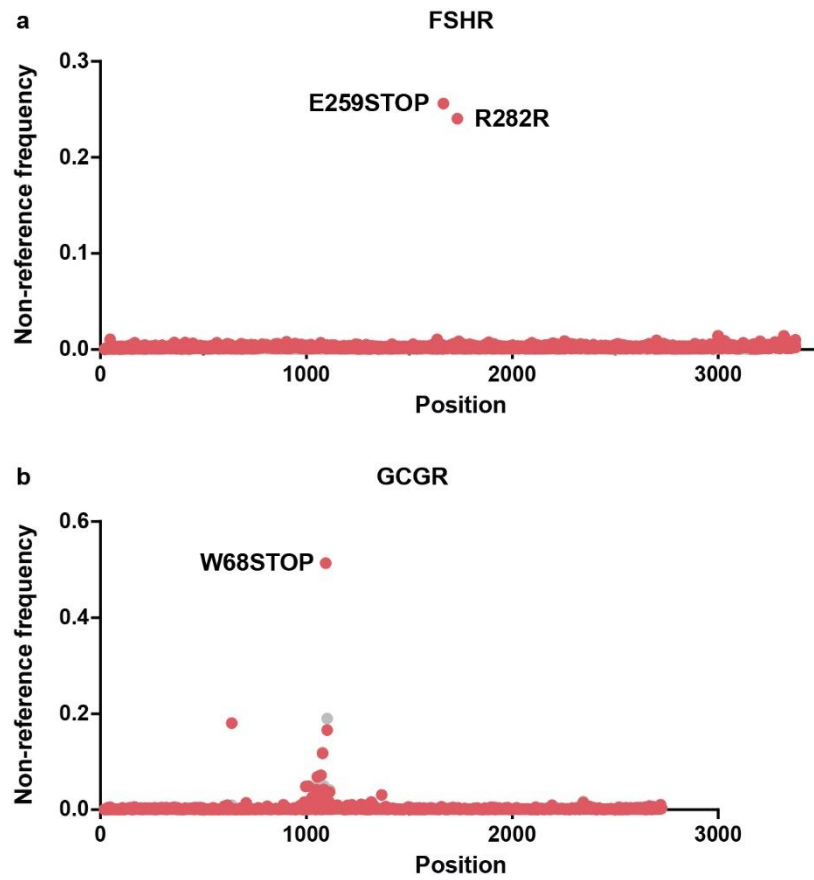
Figure 3.10 Enriched non-reference point mutations in post-selection GPCR populations. Two P6 samples showed highly enriched point mutations in P6 as compared to P0: (a) FSHR, and (b) GCGR.

In our preliminary Sanger sequencing data, we observed a high rate of deactivating insertions and deletions (indels). Following up on these observations, we looked for indels within our next-generation sequencing data. Due to the nature of Illumina's next-generation sequencing technology, we would anticipate a much lower error rate for indels than for non-reference base calls. We found a strikingly lower rate of indels are called than non-reference base calls, giving us confidence that we can use a lower cutoff gate for indel frequency than for non-reference frequency (**Table 3.3**).

As compared to P0 samples, all P6 samples showed considerable more highly enriched indels in the GPCR coding sequence. For each GPCR, we examined the top three indels in the GPCR coding region of the P6 population (**Table 3.4**). All samples show enriched indels, as high as 26.5% (CHRM3 +1A at nucleotide position 2346). As expected, all of the indels (except CHRM3 +3AAA at nucleotide position 2346) caused an extremely disruptive frameshift mutation. The indels observed through next-generation sequencing showed good agreement with the frameshifts observed through Sanger sequencing (**Table 3.2**), which encouragingly supports the use of next-generation and Sanger sequencing as complementary methods to identify enriched mutants in evolved populations. Interestingly, four of the enriched indels in **Table 3.4** were present in the P0 populations at >0.1% frequency. However, these indels were all massively enriched in the P6 population, a potential example of the founder effect observed in evolving populations.[48]

These tabulated indels show a few interesting patterns. First, all indels occur at stretches of at least three repeating bases. This finding comports with previous reports that DNA polymerases, including adenoviral polymerase, are more prone to slippage at repeating bases.[49-50] Second, A/T indels are more common in these data than G/C indels. Notably, due to the nature of adenoviral genome replication, we are unable to differentiate between slippage at pol-A or pol-T regions. Nonetheless, the prevalence of A/T indels agrees with previous observations of the mutagenesis profile of wild-type AdPol and could inform the mechanism of polymerase slippage.[50-51]

Table 3.4 Characterizing the most enriched indels for each post-selection GPCR population.
We characterized the top three most enriched indels for each P6 GPCR population. The type of indel is represented by "+" for an insertion and "–" for a deletion, followed by the number of bases inserted or deleted of the listed nucleotide. Also listed: the percentage of reads with the indel at P0 and at P6 (% P0, % P6); the immediate sequence context of the indel; and the resulting variant length of the coding sequence after the frameshift mutation (WT amino acid length provided as reference).

GPCR		Position	Type	% P0	% P6	Sequence context	Variant length	WT length
AVPR2	1	1748	+1T	1.20%	7.90%	CCG**TTTTTTTT**GGT	317	371
	2	1929	+1C	0.07%	4.00%	GGA**CCCCCCC**AAG	371	
	3	1384	+1T	0.00%	1.58%	GGC**TTTTT**CCC	197	
CALCRb	1	1078	+1A	0.01%	3.20%	ACG**AAAAAA**GAT	74	508
	2	906	+1A	0.00%	1.69%	GCG**AAAAAA**TTT	34	
	3	2144	+1T	0.01%	1.05%	GGG**TTTTTT**GTA	428	
CHRM3	1	2346	+1A	0.42%	26.54%	AAG**AAAAAAAA**GCA	547	590
	2	2346	+2A	0.00%	8.72%	AAG**AAAAAAAA**GCA	547	
	3	2346	+3A	0.00%	2.51%	AAG**AAAAAAAA**GCA	591	
FSHR	1	1107	-1A	4.52%	12.40%	TGG**AAAAAA**TCG	84	695
	2	1646	+1A	0.00%	3.60%	CTG**AAAAA**GCT	258	
	3	2797	+1T	0.04%	2.16%	AGA**TTTTTTT**ATC	647	
GCGR	1	1101	-1C	0.15%	5.45%	ATA**CCCCC**GCC	127	477
	2	1676	+1T	0.00%	2.09%	TCC**TTTTT**CTC	289	
	3	1844	+1T	0.00%	1.81%	AAC**TTTTT**CAT	335	
HTR6	1	1739	+1T	0.01%	0.94%	CCA**TTTTTT**GTG	380	440
	2	1075	+1T	0.00%	0.80%	TAA**TTTTTTT**CCT	105	
	3	933	+1G	0.00%	0.66%	CTT**GGGGGG**CTG	61	
MC1R	1	1326	+1T	0.01%	1.21%	GCA**TTTTTT**ATG	289	317
	2	1203	-1C	0.00%	0.34%	GGG**CCC**TTG	121	
	3	1422	+1T	0.00%	0.29%	CCC**TTTT**CAT	289	
MC2R	1	1070	+1T	0.00%	1.54%	TAC**TTTTTTT**ATT	68	297
	2	1357	+1T	0.00%	0.30%	GAC**TTTTT**GTA	213	
	3	1577	+1T	0.00%	0.28%	ATC**TTTT**GCT	267	
MC3R	1	961	+1T	0.01%	2.22%	CCC**TTTTTT**CTC	27	323
	2	1581	+1C	0.01%	1.64%	TCG**CCCCCC**AGC	235	
	3	1661	+1T	0.00%	0.36%	CCA**TTTTT**CCT	270	

Table 3.5 Quantifying the prevalence throughout the coding sequence of GPCRs pre- and post-selection. The number of positions within the coding sequence of GPCRs in pre-selection (P0) and post-selection (P6) populations with frequencies within the percentage ranges shown.

GPCR	≥ <	0% 0.01%	0.01% 0.1%	0.1% 1%	1% 10%	10% 100%
ADRB3	P0	1156	63	5	0	0
	P6	1137	80	6	1	0
AVPR2	P0	938	159	15	1	0
	P6	906	178	25	4	0
CALCRb	P0	1247	253	24	0	0
	P6	1202	288	29	5	0
CHRM3	P0	1491	259	19	1	0
	P6	1465	273	31	0	1
FSHR	P0	1638	414	33	0	0
	P6	1669	356	57	2	1
GCGR	P0	1162	252	17	0	0
	P6	1100	284	43	4	0
HTR6	P0	1115	180	25	0	0
	P6	1070	224	24	2	0
MC1R	P0	804	131	16	0	0
	P6	779	152	19	1	0
MC2R	P0	735	144	12	0	0
	P6	715	157	18	1	0
MC3R	P0	807	148	14	0	0
	P6	769	172	26	2	0

We were also interested in looking beyond the top three indels to quantify the prevalence of enriched indels throughout the entire GPCR coding sequence. In each GPCR coding sequence, we tabulated the number of positions with indel frequencies within different ranges: [0%, 0.01%), [0.01%, 0.1%), [0.1%, 1.0%), [1.0%, 10%), and [10%, 100%) (**Table 3.5**). These data show dramatic enrichment of indels in the P6 populations of all GPCRs. Every P6 population had more positions with indel frequencies greater than 0.1% than the respective P0 population. Furthermore, except for FSHR which suffered from low P0 coverage, every GPCR also had more positions with indel frequencies greater than 0.01% than the respective P0 population. Lastly, except for CHRM3 which suffered from the massive enrichment of indels at a single position, every GPCR also had more positions with indel frequencies greater than 1% than the respective P0 population. Correspondingly, except for FSHR, P0 populations had a greater number of positions with indels less than 0.01% than the respective P6 population. These data strongly suggest that a wide range of indels in the GPCR coding sequence were enriched in all 10 GPCR populations.

3.3.7 Quantifying selection pressure through viral replication assays. Although we initially observed that viral replication was favored when GPCRs were activated by an agonist, after observing the sequencing data we sought to revisit this observation more thoroughly. We previously relied on visual inspection when we observed an increase in viral replication during the course of an infection (**Figure 3.5d**). However, a more representative measurement of the extent of viral replication would be the actual measurement of viral titer, or the density of viral particles produced from an infection. For example, in the context of our GPCR selection circuit, we should observe an increase in titer upon GPCR activation.

To test our selection circuit, we infected Lenti-CRE-AdProt cells with a low MOI of CHRM3-encoding virus with or without 10 μM carbachol (**Figure 3.10a**). After 10 days of infection, we harvested the resulting viral population and measured the resulting viral titers using flow cytometry (**Figure 3.10b**). Surprisingly, CHRM3-encoding virus yields a 9.2-fold lower viral titer when treated with the CHRM3 agonist carbachol compared to no carbachol treatment. We ran the same experiment on producer cells (Chapter 2), which express AdProt constitutively without any cAMP response (**Figure 3.11a**). Perhaps even more surprisingly, CHRM3-encoding virus yields a 4.7-fold lower viral titer when treated with carbachol compared to no carbachol treatment (**Figure 3.11b**). These data caused us to look deeply at our system and ask: Is our directed evolution platform incompatible with a GPCR-based selection circuit?
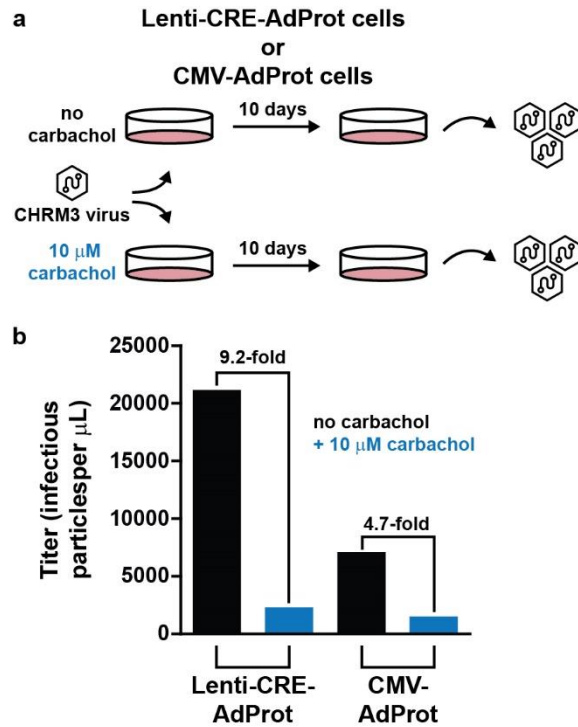
Figure 3.11 Quantifying selection pressure in lentivirally-transduced selection cells. (a) Lenti-CRE-AdProt cells and Lenti-CMV-AdProt cells (producer cells, Chapter 2) were infected with a low MOI of CHRM3-encoding virus with or without 10 μM carbachol. Viruses were harvested after 10 days of infection. (b) Resulting viral titers (in infectious particles per μL) were determined through flow cytometry.

3.4 Diagnosing the major issues of our GPCR evolution

We next sought to diagnose the major issues of the current selection circuit and re-engineer our platform toward the efficacious selection of active GPCRs. Three findings stood out to us regarding our selection circuit, suggesting a common link in the results of our directed evolution campaigns. First, we observed much higher titers after 10 days of viral replication in Lenti-CRE-AdProt cells as compared to producer cells (**Figure 3.11b**). Second, raw C_p values from qPCR of Lenti-CRE-AdProt cells and selector cells suggested that basal AdProt levels were higher in Lenti-CRE-AdProt cells (**Figure 3.5c, d**). Third, although we observed an increase in AdProt transcript levels comparable to those observed in the tTA-based evolution (**Figure 3.5c, d**), we reasoned that a ~4-fold increase in AdProt upon GPCR activation is relatively low. These findings suggest that there is likely considerable room for improvement within our selection circuit—namely, via reduction in basal AdProt levels and an increase in induced AdProt levels.

One of the most troubling findings from the previous work was that carbachol treatment decreased viral titers both in Lenti-CRE-AdProt cells and producer cells. One possibility these data suggest is that increases in intracellular cAMP levels negatively impact adenoviral growth. However, our own data do not support this hypothesis. The activity of the P6 population of the naturally constitutively active GPCR HTR6 was roughly equivalent to the activity of the P0 population, implying that cAMP induction by itself is not fundamentally detrimental to viral replication. We also did not observe the enrichment of any missense mutations, nonsense mutations, or insertions or deletions at populations higher than ~3% in the next-generation sequencing data of the P6 viral population of the HTR6 GPCRs. These data indicate that constitutively active GPCRs do not present such a significant burden to the replication of our adenoviral populations that the functional gene must be evolved to be nonfunctional, even though such selection pressure would have existed in our selection circuit. These findings suggest that the surprising results observed from the viral titering experiments are likely not generalizable to GPCR activity as a whole, and that constitutive GPCR activity is not necessarily an overwhelming burden to adenoviral replication. While these data remained puzzling, we reasoned that only way we would be able to fully diagnose this issue would be to first deconvolute problems potentially stemming from the cAMP signaling pathway from the issue of the elevated basal AdProt levels.

Therefore, we next set out to identify different methods of mammalian cell engineering to decrease basal AdProt levels and to increase the dynamic range of AdProt levels upon GPCR activation.

3.4.1 Engineering host cells with lower basal AdProt levels and higher dynamic ranges. We looked closely into our lentiviral transduction system to identify clues as to why basal

AdProt expression levels were elevated in Lenti-CRE-AdProt cells. First, the regulatory elements within the lenvitiral vector we used heavily promote the basal expression of any lentivirally-encoded transgene, regardless of whether the gene is preceded by a constitutive promoter. These regulatory elements include: the constitutively active PGK promoter immediately following AdProt;[52] the absence of any polyadenylation tail signal sequence following AdProt; and the woodchuck hepatitis virus posttranscriptional regulatory element following the puromycin resistance marker.[53] Second, it has been observed that lentiviruses preferentially integrate into actively transcribed regions in the host cell genome, which could lead to higher basal levels of AdProt expression.[54] Third, the long terminal repeats required for lentiviral integration themselves contain promoter and enhancer regions that can induce downstream transcription.[55-56] Taken together, these data suggest that our lentivirally transduced selection circuits should not necessarily be expected to provide low basal expression levels, and other cell engineering methods should be considered.

As opposed to Lenti-CRE-AdProt cells, which were engineered using lentivirus, Xfect-CRE-Luc cells were engineered using stable transfection. We rationalized that stably transfected Xfect-CRE-Luc cells would greatly reduce experimental variability, allowing us to more reliably compare results from different experiments. That we did not pursue lentiviral integration of this selection circuit was merely fortuitous—the presence of a hygromycin resistance marker in the original CRE-Luciferase plasmid simply eliminated any need to engineer any additional vector for our purposes.

We investigated genetic methods of decreasing AdProt expression. Stable transfection integrates at a low copy number[57] and does not show strong preference for actively transcribed regions of DNA,[54] suggesting that stable transfection of CRE-AdProt constructs may be a preferred approach to create selection circuits with low basal levels of activity and large dynamic ranges. Another common method to decrease basal protein levels is to use a peptide-based degradation tag, such as hPEST. However, since AdProt is a small, essential viral protein, modifying the AdProt cistron itself would likely harm viral replication. Instead, we searched for alternative regulatory elements that may cause a decrease in AdProt expression without modifying the polypeptide chain itself. A 2016 report by Geissler and co-workers identified a mammalian mRNA degradation sequence that is widespread in the mammalian genome.[58] This AT-rich sequence lowers mRNA transcript levels by promoting deadenylation through interactions with the CCR4-NOT complex. Geissler and co-workers showed that this mRNA degradation tag can be appended to mRNA transcripts in numerous genetic contexts to decrease the expression of exogenous reporter genes such as luciferase and GFP.
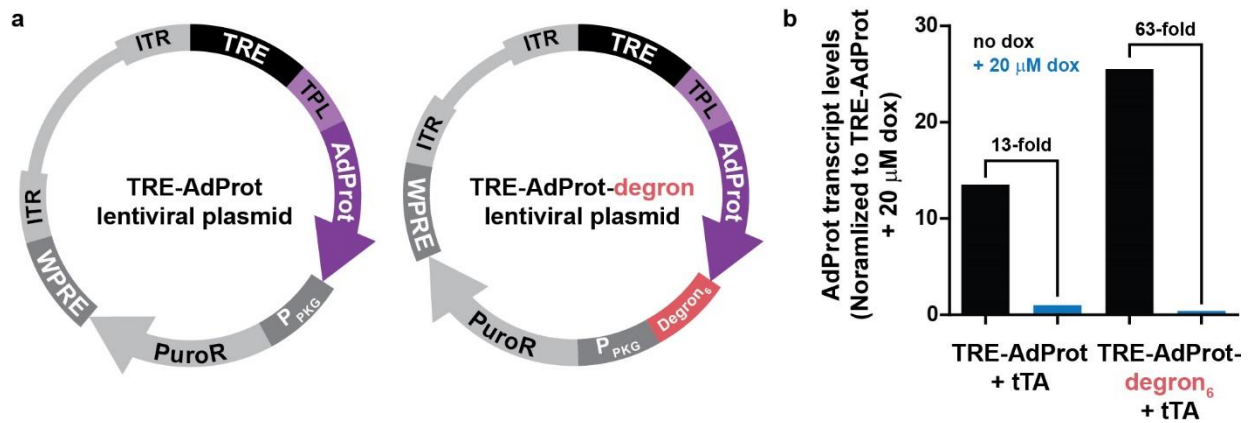
Figure 3.12 The effects of an mRNA degradation on basal AdProt expression levels. (a) Lentiviral plasmid maps with AdProt under control of the tet-responsive element (TRE, Chapter 2) with or without six copies of an mRNA degradation tag. (b) Cells were transfected with plasmids from (a) as well as the tet-transactivator, an engineered transcription factor that binds the TRE and is inhibited by doxycycline (dox). After transfection, AdProt levels were measured through qPCR.
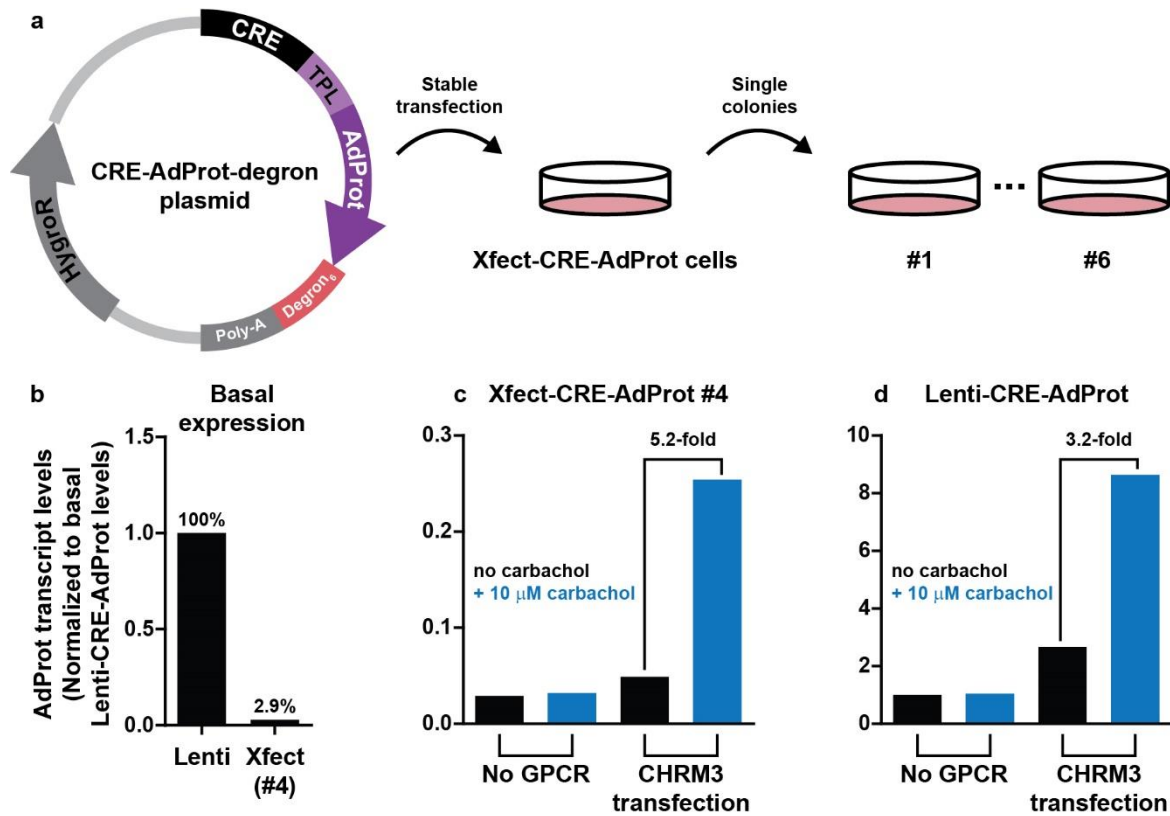
Figure 3.13 Generation and characterization of stably transfected CRE-AdProt cell lines. (a) A map of the plasmid used for creating Xfect-CRE-AdProt cells through stable transfection. A heterostable population was generated, from which homostable populations were identified through single colony picking. (b) Comparison of basal AdProt expression levels in Xfect-CRE-AdProt cells versus Lenti-CRE-AdProt cells, as determined by qPCR. (c–d) Comparison of dynamic range of AdProt levels upon carbachol-induced CHRM3 activation between (c) Xfect-CRE-AdProt cells and (c) Lenti-CRE-AdProt cells.

To initially test whether this mRNA degradation tag could be used to lower AdProt levels and increase AdProt dynamic range, we engineered a TRE-AdProt plasmid with or without six copies of the mRNA degradation tag (**Figure 3.12a**). Upon transfection with the construct and tTA, with or without 20 μM doxycycline treatment, we observed roughly 2-fold lower basal AdProt levels and roughly 2-fold higher induced AdProt levels (**Figure 3.12b**). The introduction of the degradation tag increased the dynamic range of AdProt in this assay from 13-fold to 63-fold.

Encouraged by these results, we sought to create new CRE-AdProt-degron cell lines through stable transfection. We modified the CRE-Luc plasmid to express AdProt preceded by the tripartite leader sequence upon cAMP induction, as well as introducing six copies of the mRNA degradation tag after the AdProt coding region but before the polyadenylation recognition sequence (Xfect-CRE-AdProt plasmid; **Figure 3.13a**). We transfected EP-Pol expressing HEK293A cells with the CRE-AdProt-degron plasmid and selected with hygromycin. After generating a heterostable population (Xfect-CRE-AdProt cells), we then generated stable homostable cell lines by picking single colonies. We compared the basal AdProt levels of best performing single colony (Xfect-CRE-AdProt-#4) and with pLVX-CRE-AdProt cells, and observed 34-fold lower AdProt levels in Xfect-CRE-AdProt-#4 cells (**Figure 3.13b**). When we transfected both cell populations with a CHRM3-encoding plasmid, we observed a slight improvement in AdProt level dynamic range with 10 μM carbachol treatment (**Figure 3.13c, d**). Overall, we find that Xfect-CRE-AdProt cells demonstrate dramatically lower basal AdProt levels and a modestly improved AdProt dynamic range upon GPCR activation.

3.4.2 Quantifying selection pressure through viral replication assays in stably transfected cells. We finally aimed to quantify selection pressure in Xfect-CRE-AdProt cells using a viral replication assay. In three single colonies (#3, #4, and #5), we infected cells with CHRM3-encoding adenovirus with or without 10 μM carbachol treatment (**Figure 3.14a**). After 10 days of infection, we quantified viral titers through flow cytometry (**Figure 3.14b**). We observed that single colonies #4 and #5 showed 2.1- and 2.5-fold increase in viral titer upon carbachol treatment. These demonstrate provide initial validation for our selection circuit, and provide a roadmap toward improving the CRE-based GPCR selection circuit in the future.
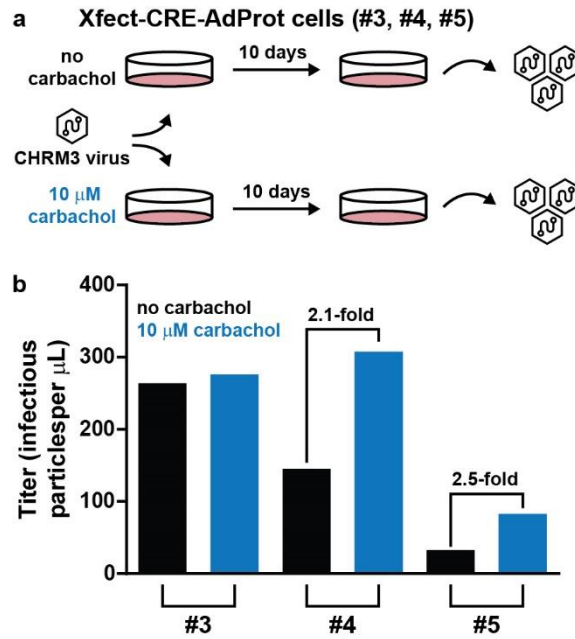
Figure 3.14 Quantifying selection pressure in Xfect-CRE-AdProt cells. (a) Selection scheme. Xfect-CRE-AdProt cells #3, #4, and #5 were infected with low MOI of CHRM3-encoding virus and treated with or without 10 μM carbachol. Viruses were harvested ten days post-infection and the resulting viral titers were measured through flow cytometry. (b) Flow cytometry-determined titers of resulting viral populations.

3.5 Discussion

During the pursuit of this work, another virus-based directed evolution platform in mammalian cells published the evolution of a constitutively active GPCR.[24] This platform, termed viral evolution of genetically actuating sequences (VEGAS), is described in detail in Chapter 1. Their approach was similar to our own, using the serum response element instead of the cAMP response element, and they evolved variants of the $G\alpha_q$-coupled GPCR MRGPRX2 to have higher constitutive activity in mammalian cells.

Our initial efforts to evolve GPCRs in Lenti-CRE-AdProt cells were marred by high basal levels of AdProt expression. Not only did we not observe increases of GPCR-based cAMP signaling activity after six rounds of evolution, we did observe the enrichment of nonsense mutations and frameshifts leading to early coding sequence termination. Learning from these results, we designed new plasmids designed for lower basal expression and created stronger selection circuits using stable transfection. We then confirmed that these selection circuits create a selection pressure that could, in theory, enrich for GPCRs with higher activity. By carefully examining our platform and diagnosing our unique modes of failure, we were able to go back to the drawing board and re-engineer cell lines that provided a selection pressure that would select for GPCR activity.

These results make clear three fundamental features of our platform. First, if basal AdProt levels are high enough to promote viral replication, then viruses will likely escape the designed selection pressure. In the case of many of our GPCRs, the viruses may even reject the transgene altogether, presumably because overexpression of an unnecessary and difficult-to-fold transgene causes a fitness defect on viral replication. Second, merely observing selection couples through luciferase assays, or even observing AdProt levels directly through qPCR, is not nearly sufficient evidence that a selection circuit is functional. The gold-standard, and perhaps the only standard, for determining the efficacy of a selection circuit is the measurement of viral titers after selection. This is perhaps the most important practical lesson learned. Third, and perhaps most speculatively, problems arising from difficult endogenous circuits or cellular engineering problems *can be solved*. Although the initial data we observed were discouraging, by carefully identifying the stumbling blocks and methodically re-engineering our system to work around them, we were able to eventually demonstrate a viral selection circuit based on GPCR activity.

Further engineering efforts would likely improve selection. These improvements include: employing improved CRE sequences;[59] enhancing endogenous cAMP signaling responses;[60] inhibiting or removing endogenous negative feedback loops inherent to cAMP signaling

pathways;[61] expression of effectors that modulate arms of the ER proteostasis network;[62] and overexpression of additional cAMP signaling effectors.[63]

These strategies and more may improve the dynamic range of the GPCR-based selection circuits, and ultimately allow us to evolve new functional GPCR variants. Never before have there been more tools to engineer mammalian cells—from precise genomic engineering to site-specific recombination and beyond—to find creative host cell engineering solutions that can be used to tackle difficult selection circuit problems.

3.6 Methods

3.6.1 Cloning methods. All PCR reactions for cloning cassettes were performed using Q5 High-Fidelity DNA Polymerase (New England BioLabs). Restriction cloning was performed using restriction enzymes and Quick Ligase from New England BioLabs. GPCR expression vectors and adenoviral vectors were engineered using Gateway cloning. All primers are listed in **Table 3.6**.

The CRE-Luciferase plasmid (pGL4.29[luc2P/CRE/Hygro] vector) used for generating Xfect-CRE-Luc cells was purchased from Promega.[37] Plasmids encoding all human GPCRs were obtained from the PRESTO-Tango plasmid kit from Addgene, deposited by Bryan Roth (Addgene kit # 1000000068).[64]

GPCR expression vectors and adenoviral vectors were created as follows: GPCR genes from the PRESTO-Tango plasmid kit were amplified using GPCRPT-F-KpnI and GPCRV2-R-NotI and cloned into pENTR1A (ThermoFisher). These vectors were then cloned into pcDNA-DEST40 expression vectors and AdEvolveDEST (Chapter 2) vectors through Gateway cloning (ThermoFisher).

The Lenti-CRE-AdProt lentiviral plasmid was created by modifying the Lenti-TRE-AdProt lentiviral plasmid through restriction cloning. The CRE from the CRE-Luciferase plasmid was amplified using primers CRE-FWD and CRE-REV and inserted into Lenti-TRE-AdProt through restriction cloning.

The Lenti-TRE-AdProt-degron lentiviral plasmid was created by modifying the Lenti-TRE-AdProt plasmid through restriction cloning. Primers mRNA-Degron-TOP and mRNA-Degron-BOTTOM were annealed to form a genetic region encoding six copies of an mRNA degron, which was inserted into Lenti-TRE-AdProt through restriction cloning.

Table 3.6 Table of primers.

Primer name	Sequence
GPCRPT-F-KpnI	ATATATGGTACCGCCACCATGAAGACGATCATCGCCC
GPCRV2-R-NotI	ATATATGCGGCCGCCTACGATGAAGTGTCCTTGGCCAGGG
CRE-FWD	AAAAAAAAATCGATCACCAGACAGTGACGTCAGCT
CRE-REV	AAAAAAAAAGGATCCCTCCCATTGACGTCAATGGGG
mRNA-Degron-TOP	CGCGTATTGAATAATTGAATAATTGAATAATTGAATAA TTGAATAATTGAATAATTGAATAATTGAATAGC
mRNA-Degron-BOTTOM	GGCCGCTATTCAATTATTCAATTATTCAATTATTCAA TTATTCAATTATTCAATTATTCAATTATTCAATA
AdProt-FseI-FWD	AAAAAAGGCCGGCCGGTAGAATTCACTCGAGACGCTTAGATCCAGA
AdProt-degron-REV	AAAAAACAAGCTTAGACACTAGAGGGTATATAATGGA AGCTCGACTTCCAGCTTGGCAATCCGGTACTGTTGG TAAAGCCACCACTCTCTTCCGCATCGCTGTC
AdProt-qPCR-FWD	GAAAAGTCCACCCAAAGCGT
AdProt-qPCR-REV	CTGTTGAGCATGGAGTTGGG
RPLP2-qPCR-FWD	CGTCGCCTCCTACCTGCT
RPLP2-qPCR-REV	CCATTCAGCTCACTGATAACCTT
AdV-GOI-FWD	CTACATAAGACCCCCACCTTATATATTCTTTCC
AdV-GOI-REV	AGCGGGAAAACTGAATAAGAGGAAGTGAAATC

The CRE-AdProt-degron plasmid used for generating Xfect-CRE-AdProt cells was created by modifying the CRE-Luciferase plasmid. A genetic region encoding the tripartite leader sequence, AdProt, the mRNA degrons, and a SV40 polyadenylation tail was amplified from the Lenti-TRE-AdPRot-degron lentiviral plasmid using primers AdProt-FseI-FWD and AdProt-degron-REV and cloned into the CRE-Luciferase plasmid using restriction cloning.

3.6.2 General cell culture. Cells were cultured at 37 °C and 5% CO_2(g). New cell lines were derived from a parent HEK293A cell line (Thermo Fisher) 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," "mutator" (Chapter 2),[38] Xfect-CRE-AdProt, and Xfect-CRE-Luc cell lines were cultured in 50 µg/mL hygromycin (Thermo Fisher) to stably maintain transgenes. The "selector," "phenotyping" (Chapter 2), and Lenti-CRE-AdProt cell lines were cultured in 1 µg/mL puromycin (Corning) for the same purpose.

3.6.3 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: 15 µg of RRE, 6 µg of REV, 3 µg of VSVG, and 15 µg of transfer vector using 60 µL of TransIT-Lenti transfection reagent (Mirus Bio). After 72 h, the medium was harvested and centrifuged for 5 min at 3200 × g to clear the cell debris. The supernatant was used to transduce HEK293A-derived cells supplemented with 4 µg/mL polybrene (Sigma-Aldrich). After 48 h, the medium was exchanged for DMEM containing appropriate antibiotics to select stable cell lines.

3.6.4 Generation of cell lines and single colonies by stable transfection. In a typical protocol, ~9 × 10^6 HEK293A cells (Thermo Fisher) were plated on a tissue culture-treated 10 cm plates. The next day, cells were transfected with 10 µg of the respective plasmid using 30 µL of TransIT-Lenti transfection reagent (Mirus Bio). After 72 h, the medium was exchanged for DMEM containing appropriate antibiotics to select stable cell lines. After ~two weeks, surviving cells were then plated at 1:10000 dilution onto 10 cm plates with appropriate selection media. After an additional two weeks, single colonies were physically detached from the plate and expanded in appropriate selection media.

3.6.5 Luciferase assay with Xfect-CRE-AdProt cells. In a typical protocol, ~5 × 10^5 Xfect-CRE-Luc cells are plated on tissue culture-treated 12-well plates. The next day, wells

were either: transfected with 1 μg of the respective plasmid using 3 μL of TransIT-Lenti transfection reagent (Mirus Bio); infected with GPCR-encoding adenovirus at an MOI ~ 1.0; or transfected with PCR-amplified regions of DNA encoding the CMV promoter, gene of interest, and polyadenylation tail. 48 h post-expression, cells were washed once with 1× PBS and then lysed with 500 μL luciferase assay buffer (0.1 M Tris-HCl, 0.04 M Tris-base, 75 mM NaCl, 3 mM MgCl$_2$, 0.25% Triton X-100, 5 mM DTT, 0.2 mM coenzyme A, 0.15 mM ATP, 1.4 mg/mL luciferin) for 10 minutes. Cell lysates were then transferred to a white-bottomed 96-well plate and luminescence is measured.

3.6.6 Quantitative PCR (qPCR). In a typical protocol, ~5 × 10^5 cells are plated on tissue culture-treated 12 well plates. The next day, wells were either: transfected with 1 μg of the respective plasmid using 3 μL of TransIT-Lenti transfection reagent (Mirus Bio); or infected with GPCR-encoding adenovirus at an MOI ~ 1.0. 48–72 h post-expression, 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 a high-capacity cDNA reverse transcription kit (Applied Biosystems). qPCR analysis for AdProt (primers AdProt-qPCR-FWD and AdProt-qPCR-REV) and the housekeeping gene RPLP2 (primers RPLP2-qPCR-FWD and RPLP2-qPCR-REV) was performed on a LightCycler 480 II (Roche). AdProt transcript levels were normalized to intrasample RPLP2 levels before comparing between samples.

3.6.7 Generation of adenovirus. Adenoviruses were produced by transfecting a PacI (New England BioLabs)-linearized vector into appropriate trans-complementing HEK293A cells ("producer" cells; Chapter 2). A 10 μg sample of PacI-linearized adenovirus vectors mixed with 30 μL of TransIT-Lenti transfection reagent (Mirus Bio) was added to a 10 cm plate of producer cells (~8 × 10^6 cells). The medium was intermittently replaced every 2–3 days until plaques were observed (typically ~3 weeks). Once plaques were detected, full 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 3200 × g for 15 min and the supernatant was stored at −80 °C. Viruses were then passaged on producer cells to produce a high-titer stock of virus which was used to initiate directed evolution experiments.

3.6.8 Directed evolution workflow. Ten 10 cm plates of Lenti-CRE-AdProt cells were each infected with ~250 μL of high-titer stock of one of ten GPCR-encoding ΔAdPolΔAdProt viruses. Once spreading infection was observed and the majority of cells were infected, 500 μL of viral media was passaged from a previous round of evolution to a new 10 cm plate of fresh

Lenti-CRE-AdProt cells. 2 mL of viral media was also saved from each round of evolution. This procedure was repeated for six rounds of evolution.

3.6.9 Next-generation sequencing of GPCR populations. Viral DNA populations from passage zero and passage six of each of the ten evolutions were harvested using a NucleoSpin Virus DNA isolation kit (Macherey-Nagel). A region of DNA encompassing the CMV promoter, GPCR coding sequence, and polyadenylation tail was amplified from 3 µL of harvested viral DNA and PCR-amplified with Q5 High-Fidelity DNA Polymerase (New England BioLabs) for 16 rounds of amplification using primers AdV-GOI-FWD and AdV-GOI-REV. The resulting PCR product was purified and prepared for Illumina sequencing via the NexTera XL DNA Library Prep protocol (Illumina). A 150 bp paired-end sequencing was performed on a MiSeq (Illumina). Sequencing reads were aligned to the amplicon sequence, which was derived from the GPCR adenovirus sequences using BWA and allele pileups were generated using samtools v1.5.[65] For all samples except a small number of positions in FSHR P0, each position within the GPCR gene had at least 5000-fold coverage.

3.6.10 Viral replication assay. For viral replication assays, ~5×10^5 cells were plated on tissue culture-treated 6 well plates. The next day, media was exchanged and 10 µM carbachol was added if appropriate, and then cells were infected with 20 µL of a high titer stock of CHRM3-encoding adenovirus. Pictures were taken every two days. After 10 days, viral media was harvested and subjected to three freeze-thaw cycles. The cell debris was removed by centrifugation at 3200 × g for 15 min and the supernatant was stored at −80 °C.

3.6.11 Viral titering through flow cytometry. Known volumes of viral supernatants were added to AdPol-expressing HEK293A cells. 72 h post-infection, the cells were washed once with medium, stained with 0.2 µg/mL DAPI (4′,6-diamidino-2-phenylindole, Thermo Fisher), 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 0.1–10% of the cells were infected.

3.7 References

1.   Edward Zhou, X.; Melcher, K.; Eric Xu, H., Structural biology of G protein-coupled receptor signaling complexes. *Protein Sci.* **2019,** *28* (3), 487-501.
2.   Weis, W. I.; Kobilka, B. K., The molecular basis of G protein–coupled receptor activation. *Annu. Rev. Biochem.* **2014,** *87* (1), 897-919.

3.   Hauser, A. S.; Attwood, M. M.; Rask-Andersen, M.; Schiöth, H. B.; Gloriam, D. E., Trends in GPCR drug discovery: new agents, targets and indications. *Nat. Rev. Drug Discov.* **2017,** *16* (12), 829-842.

4.   DeMaria, S.; Ngai, J., The cell biology of smell. *J. Cell Biol.* **2010,** *191* (3), 443-52.

5.   Chaudhari, N.; Roper, S. D., The cell biology of taste. *J. Cell Biol.* **2010,** *190* (3), 285-96.

6.   Sung, C. H.; Chuang, J. Z., The cell biology of vision. *J. Cell Biol.* **2010,** *190* (6), 953-63.

7.   Rosenbaum, D. M.; Rasmussen, S. G. F.; Kobilka, B. K., The structure and function of G-protein-coupled receptors. *Nature* **2009,** *459* (7245), 356-363.

8.   Hauser, A. S.; Chavali, S.; Masuho, I.; Jahn, L. J.; Martemyanov, K. A.; Gloriam, D. E.; Babu, M. M., Pharmacogenomics of GPCR drug targets. *Cell* **2018,** *172* (1), 41-54.e19.

9.   Smit, M. J.; Vischer, H. F.; Bakker, R. A.; Jongejan, A.; Timmerman, H.; Pardo, L.; Leurs, R., Pharmacogenomic and structural analysis of constitutive G protein-coupled receptor activity. *Annu. Rev. Pharmacol. Toxicol.* **2007,** *47*, 53-87.

10.   Stoy, H.; Gurevich, V. V., How genetic errors in GPCRs affect their function: possible therapeutic strategies. *Genes Dis.* **2015,** *2* (2), 108-132.

11.   García-Nafría, J.; Tate, C. G., Cryo-electron microscopy: moving beyond X-ray crystal structures for drug receptors and drug development. *Annu. Rev. Pharmacol. Toxicol.* **2020,** *60* (1), 51-71.

12.   Nichols, C. D.; Roth, B. L., Engineered G-protein coupled receptors are powerful tools to investigate biological processes and behaviors. *Front. Mol. Neurosci.* **2009,** *2*, 16-16.

13.   Conklin, B. R.; Hsiao, E. C.; Claeysen, S.; Dumuis, A.; Srinivasan, S.; Forsayeth, J. R.; Guettier, J.-M.; Chang, W. C.; Pei, Y.; McCarthy, K. D.; Nissenson, R. A.; Wess, J.; Bockaert, J.; Roth, B. L., Engineering GPCR signaling pathways with RASSLs. *Nat. Methods* **2008,** *5* (8), 673-678.

14.   Armbruster, B. N.; Li, X.; Pausch, M. H.; Herlitze, S.; Roth, B. L., 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.* **2007,** *104* (12), 5163-8.

15.   Roth, B. L., DREADDs for Neuroscientists. *Neuron* **2016,** *89* (4), 683-694.

16.   Dong, S.; Allen, J. A.; Farrell, M.; Roth, B. L., A chemical-genetic approach for precise spatio-temporal control of cellular signaling. *Mol. Biosyst.* **2010,** *6* (8), 1376-80.

17.   Aldrin-Kirk, P.; Bjorklund, T., Practical considerations for the use of DREADD and other chemogenetic receptors to regulate neuronal activity in the mammalian brain. *Methods Mol. Biol.* **2019,** *1937*, 59-87.

18.   Raper, J.; Morrison, R. D.; Daniels, J. S.; Howell, L.; Bachevalier, J.; Wichmann, T.; Galvan, A., Metabolism and distribution of clozapine-*N*-oxide: implications for nonhuman primate chemogenetics. *ACS Chem. Neurosci.* **2017,** *8* (7), 1570-1576.

19.   Manvich, D. F.; Webster, K. A.; Foster, S. L.; Farrell, M. S.; Ritchie, J. C.; Porter, J. H.; Weinshenker, D., The DREADD agonist clozapine-*N*-oxide (CNO) is reverse-metabolized to clozapine and produces clozapine-like interoceptive stimulus effects in rats and mice. *Sci. Rep.* **2018,** *8* (1), 3840.

20.   Baldessarini, R. J.; Centorrino, F.; Flood, J. G.; Volpicelli, S. A.; Huston-Lyons, D.; Cohen, B. M., Tissue concentrations of clozapine and its metabolites in the rat. *Neuropsychopharmacology* **1993,** *9* (2), 117-124.

21.   Gomez Juan, L.; Bonaventura, J.; Lesniak, W.; Mathews William, B.; Sysa-Shah, P.; Rodriguez Lionel, A.; Ellis Randall, J.; Richie Christopher, T.; Harvey Brandon, K.; Dannals Robert, F.; Pomper Martin, G.; Bonci, A.; Michaelides, M., Chemogenetics revealed: DREADD occupancy and activation via converted clozapine. *Science* **2017,** *357* (6350), 503-507.

22.   Vardy, E.; Robinson, J. E.; Li, C.; Olsen, Reid H. J.; DiBerto, Jeffrey F.; Giguere, Patrick M.; Sassano, Flori M.; Huang, X.-P.; Zhu, H.; Urban, Daniel J.; White, Kate L.; Rittiner, Joseph E.; Crowley, Nicole A.; Pleil, Kristen E.; Mazzone, Christopher M.; Mosier, Philip D.; Song, J.; Kash, Thomas L.; Malanga, C. J.; Krashes, Michael J.; Roth, Bryan L., A new

DREADD facilitates the multiplexed chemogenetic interrogation of behavior. *Neuron* **2015,** *86* (4), 936-946.

23.	Goutaudier, R.; Coizet, V.; Carcenac, C.; Carnicella, S., Compound 21, a two-edged sword with both DREADD-selective and off-target outcomes in rats. *PLoS One* **2020,** *15* (9), e0238156.

24.	English, J. G.; Olsen, R. H. J.; Lansu, K.; Patel, M.; White, K.; Cockrell, A. S.; Singh, D.; Strachan, R. T.; Wacker, D.; Roth, B. L., VEGAS as a Platform for Facile Directed Evolution in Mammalian Cells. *Cell* **2019,** *178* (3), 748-761.e17.

25.	Versele, M.; Lemaire, K.; Thevelein, J. M., Sex and sugar in yeast: two distinct GPCR systems. *EMBO Rep.* **2001,** *2* (7), 574-579.

26.	Lengger, B.; Jensen, M. K., Engineering G protein-coupled receptor signalling in yeast for biotechnological and medical purposes. *FEMS Yeast Res.* **2020,** *20* (1).

27.	Leurs, R.; Pena, M. S. R.; Bakker, R. A.; Alewijnse, A. E.; Timmerman, H., Constitutive activity of G protein coupled receptors and drug action. In *Pharmacochemistry Library*, Gulini, U.; Gianella, M.; Quaglia, W.; Marucci, G., Eds. Elsevier: 2000; Vol. 31, pp 327-331.

28.	Seifert, R.; Wenzel-Seifert, K., Constitutive activity of G-protein-coupled receptors: cause of disease and common property of wild-type receptors. *Naunyn Schmiedebergs Arch. Pharmacol.* **2002,** *366* (5), 381-416.

29.	Wootten, D.; Christopoulos, A.; Marti-Solano, M.; Babu, M. M.; Sexton, P. M., Mechanisms of signalling and biased agonism in G protein-coupled receptors. *Nat. Rev. Mol. Cell Biol.* **2018,** *19* (10), 638-653.

30.	Peterson, Y. K.; Luttrell, L. M., The diverse roles of arrestin scaffolds in G protein-coupled receptor signaling. *Pharmacol. Rev.* **2017,** *69* (3), 256-297.

31.	May, L. T.; Leach, K.; Sexton, P. M.; Christopoulos, A., Allosteric modulation of G protein–coupled receptors. *Annu. Rev. Pharmacol. Toxicol.* **2007,** *47* (1), 1-51.

32.	Sassone-Corsi, P., The cyclic AMP pathway. *Cold Spring Harb. Perspect. Biol.* **2012,** *4* (12), a011148.

33.	Pierce, K. L.; Premont, R. T.; Lefkowitz, R. J., Seven-transmembrane receptors. *Nat. Rev. Mol. Cell Biol.* **2002,** *3* (9), 639-650.

34.	White, A. D.; Jean-Alphonse, F. G.; Fang, F.; Peña, K. A.; Liu, S.; König, G. M.; Inoue, A.; Aslanoglou, D.; Gellman, S. H.; Kostenis, E.; Xiao, K.; Vilardaga, J. P., G(q/11)-dependent regulation of endosomal cAMP generation by parathyroid hormone class B GPCR. *Proc. Natl. Acad. Sci. U. S. A.* **2020,** *117* (13), 7455-7460.

35.	Mayr, B.; Montminy, M., Transcriptional regulation by the phosphorylation-dependent factor CREB. *Nat. Rev. Mol. Cell Biol.* **2001,** *2* (8), 599-609.

36.	Guo, H.; Cheng, Y.; Wang, C.; Wu, J.; Zou, Z.; Niu, B.; Yu, H.; Wang, H.; Xu, J., FFPM, a PDE4 inhibitor, reverses learning and memory deficits in APP/PS1 transgenic mice via cAMP/PKA/CREB signaling and anti-inflammatory effects. *Neuropharmacology* **2017,** *116*, 260-269.

37.	Cheng, Z.; Garvin, D.; Paguio, A.; Stecha, P.; Wood, K.; Fan, F., Luciferase reporter assay system for deciphering GPCR pathways. *Curr. Chem. Genomics* **2010,** *4*, 84-91.

38.	Berman, C. M.; Papa, L. J., 3rd; Hendel, S. J.; Moore, C. L.; Suen, P. H.; Weickhardt, A. F.; Doan, N. D.; Kumar, C. M.; Uil, T. G.; Butty, V. L.; Hoeben, R. C.; Shoulders, M. D., An adaptable platform for directed evolution in human cells. *J. Am. Chem. Soc.* **2018,** *140* (51), 18093-18103.

39.	Montminy, M. R.; Bilezikjian, L. M., Binding of a nuclear protein to the cyclic-AMP response element of the somatostatin gene. *Nature* **1987,** *328* (6126), 175-178.

40.	Montminy, M. R.; Sevarino, K. A.; Wagner, J. A.; Mandel, G.; Goodman, R. H., Identification of a cyclic-AMP-responsive element within the rat somatostatin gene. *Proc. Natl. Acad. Sci. U. S. A.* **1986,** *83* (18), 6682-6686.

41.   Rogers, S.; Wells, R.; Rechsteiner, M., Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* **1986,** *234* (4774), 364-368.

42.   Li, X.; Zhao, X.; Fang, Y.; Jiang, X.; Duong, T.; Fan, C.; Huang, C. C.; Kain, S. R., Generation of destabilized green fluorescent protein as a transcription reporter. *J. Biol. Chem.* **1998,** *273* (52), 34970-5.

43.   Zeng, F. Y.; Wess, J., Identification and molecular characterization of M3 muscarinic receptor dimers. *J. Biol Chem.* **1999,** *274* (27), 19487-97.

44.   Bhartiya, D.; Patel, H., An overview of FSH-FSHR biology and explaining the existing conundrums. *J. Ovarian Res.* **2021,** *14* (1), 144.

45.   Kristiansen, K., Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function. *Pharmacol. Ther.* **2004,** *103* (1), 21-80.

46.   Yang, D.; Zhou, Q.; Labroska, V.; Qin, S.; Darbalaei, S.; Wu, Y.; Yuliantie, E.; Xie, L.; Tao, H.; Cheng, J.; Liu, Q.; Zhao, S.; Shui, W.; Jiang, Y.; Wang, M.-W., G protein-coupled receptors: structure- and function-based drug discovery. *Signal Transduct. Target. Ther.* **2021,** *6* (1), 7.

47.   Banaszynski, L. A.; Chen, L.-C.; Maynard-Smith, L. A.; Ooi, A. G. L.; Wandless, T. J., A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* **2006,** *126* (5), 995-1004.

48.   Templeton, A. R., The theory of speciation via the founder principle. *Genetics* **1980,** *94* (4), 1011-1038.

49.   Viguera, E.; Canceill, D.; Ehrlich, S. D., Replication slippage involves DNA polymerase pausing and dissociation. *EMBO J.* **2001,** *20* (10), 2587-2595.

50.   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.* **2016,** *12* (11), e1006013.

51.   Sanjuán, R.; Nebot, M. R.; Chirico, N.; Mansky, L. M.; Belshaw, R., Viral mutation rates. *J. Virol.* **2010,** *84* (19), 9733-48.

52.   Norrman, K.; Fischer, Y.; Bonnamy, B.; Wolfhagen Sand, F.; Ravassard, P.; Semb, H., Quantitative comparison of constitutive promoters in human ES cells. *PloS One* **2010,** *5* (8), e12413-e12413.

53.   Donello, J. E.; Loeb, J. E.; Hope, T. J., Woodchuck hepatitis virus contains a tripartite posttranscriptional regulatory element. *J. Virol.* **1998,** *72* (6), 5085-5092.

54.   Yant, S. R.; Wu, X.; Huang, Y.; Garrison, B.; Burgess, S. M.; Kay, M. A., High-resolution genome-wide mapping of transposon integration in mammals. *Mol. Cell Biol.* **2005,** *25* (6), 2085-94.

55.   Berkhout, B.; Jeang, K. T., Functional roles for the TATA promoter and enhancers in basal and Tat-induced expression of the human immunodeficiency virus type 1 long terminal repeat. *J. Virol.* **1992,** *66* (1), 139-49.

56.   Shah, S.; Alexaki, A.; Pirrone, V.; Dahiya, S.; Nonnemacher, M. R.; Wigdahl, B., Functional properties of the HIV-1 long terminal repeat containing single-nucleotide polymorphisms in Sp site III and CCAAT/enhancer binding protein site I. *Virol. J.* **2014,** *11*, 92-92.

57.   Baron, U.; Bujard, H., Tet repressor-based system for regulated gene expression in eukaryotic cells: principles and advances. *Meth. Enzymol.* **2000,** *327*, 401-21.

58.   Geissler, R.; Simkin, A.; Floss, D.; Patel, R.; Fogarty, E. A.; Scheller, J.; Grimson, A., A widespread sequence-specific mRNA decay pathway mediated by hnRNPs A1 and A2/B1. *Genes Dev.* **2016,** *30* (9), 1070-1085.

59.   Davis, J. E.; Insigne, K. D.; Jones, E. M.; Hastings, Q. A.; Boldridge, W. C.; Kosuri, S., Dissection of c-AMP response element architecture by using genomic and episomal massively parallel reporter assays. *Cell Syst.* **2020,** *11* (1), 75-85.e7.

60.   Iourgenko, V.; Zhang, W.; Mickanin, C.; Daly, I.; Jiang, C.; Hexham, J. M.; Orth, A. P.; Miraglia, L.; Meltzer, J.; Garza, D.; Chirn, G. W.; McWhinnie, E.; Cohen, D.; Skelton, J.; Terry, R.; Yu, Y.; Bodian, D.; Buxton, F. P.; Zhu, J.; Song, C.; Labow, M. A., Identification of a family of cAMP response element-binding protein coactivators by genome-scale functional analysis in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **2003,** *100* (21), 12147-52.

61.   Molina, C. A.; Foulkes, N. S.; Lalli, E.; Sassone-Corsi, P., Inducibility and negative autoregulation of CREM: an alternative promoter directs the expression of ICER, an early response repressor. *Cell* **1993,** *75* (5), 875-86.

62.   Shoulders, M. D.; Ryno, L. M.; Genereux, J. C.; Moresco, J. J.; Tu, P. G.; Wu, C.; Yates, J. R., 3rd; Su, A. I.; Kelly, J. W.; Wiseman, R. L., Stress-independent activation of XBP1s and/or ATF6 reveals three functionally diverse ER proteostasis environments. *Cell Rep.* **2013,** *3* (4), 1279-92.

63.   Zhuang, H.; Matsunami, H., Synergism of accessory factors in functional expression of mammalian odorant receptors. *J. Biol. Chem.* **2007,** *282* (20), 15284-93.

64.   Kroeze, W. K.; Sassano, M. F.; Huang, X. P.; Lansu, K.; McCorvy, J. D.; Giguère, P. M.; Sciaky, N.; Roth, B. L., PRESTO-Tango as an open-source resource for interrogation of the druggable human GPCRome. *Nat. Struct. Mol. Biol.* **2015,** *22* (5), 362-9.

65.   Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R., The sequence alignment/map format and SAMtools. *Bioinformatics* **2009,** *25* (16), 2078-9.

Chapter 4: Directed evolution of CRISPR systems

4.1 Author contributions

Samuel J. Hendel, Amanuella A. Mengiste, and Matthew D. Shoulders conceived the project and designed the experiments. S.J.H. performed the cloning, virus and cell line engineering, characterization assays, directed evolution experiments. A.A.M. performed plaque assays, plasmid engineering, and characterization assays. Andrew Sabol performed the Xfect-CRE-AdProt cell line genomic integration assay. Amit Choudhary and Vedagopuram Sreekanth created the AcrIIA4-CSD constructs.

4.2 Overview of CRISPR systems

The development of RNA-guided endonucleases, such as CRISPR (clustered regularly interspaced short palindromic repeats) systems, has rapidly transformed molecular biology by enabling an immensely broad range of precisely targeted genetic perturbations.[1] CRISPR systems have been used to engineer genomes across all kingdoms of life, from single-celled organisms such as bacteria[2] and yeast,[3] to complex multicellular organisms such as plants[4-6] and mammals.[7] CRISPR-based gene editing has entered the clinic for the treatment of human disease, including sickle cell disease,[8] β-thalassemia,[8] Duchennes muscular dystrophy,[9] and Leber congenital amaurosis.[10] Furthermore, researchers can expand the capabilities of CRISPR gene editing systems by fusing different effector proteins with additional functionalities,[11] including destabilizing domains,[12] transcriptional activators,[13] epigenetic modifiers,[14] nucleotide deaminases, transposases, and prime editing systems.[15] Taken together, there are few fields in molecular biology that have not been impacted by the advent of CRISPR technologies.

Because CRISPR systems evolved as adaptive immune systems for bacteria[16] to recognize and fight off viral infections, CRISPR systems tend to work most efficiently in bacteria. While some CRISPR systems can function in mammalian cells, most notably *Streptococcus pyogenes* CRISPR associated protein 9 (Cas9, also referred to as SpCas9 or SpyCas9), most CRISPR systems found in nature do not function in mammalian cells.[17] The reasons are as diverse as the proteins themselves.[18] For example, from a biophysical perspective, the protein may get trapped in complex with undesirable partners, aggregate, fail to fold owing to lack of the correct cognate chaperones, or be directed to degradation. From a functional perspective, the protein may be post-translationally modified in a deleterious manner, may mislocalize, or may not interact properly with complex signaling pathways or structural features that differ between organisms.
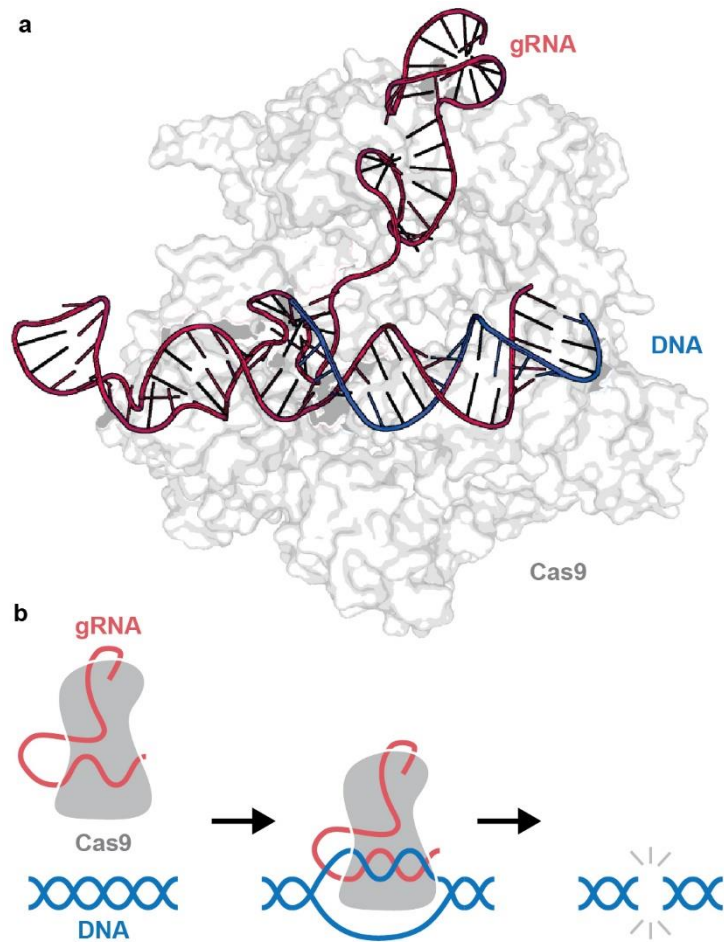
Figure 4.1 Structure and function of CRISPR-Cas9. (a) Crystal structure of Cas9 bound to a guide RNA (gRNA; red) and target DNA (blue). (b) CRISPR-Cas9 gene editing. First, Cas9 (grey) binds a gRNA to form a Cas9–gRNA complex. If this complex finds an appropriate DNA region complementary to the gRNA and with the correct PAM sequence, DNA cleavage occurs and a double strand break is formed.

As a result, researchers interested in using CRISPR in mammalian cells are limited to a small handful of CRISPR systems from which to choose. By far the most widely-used CRISPR system, and indeed the only one that has currently entered the clinic, remains Cas9 (structure[19] shown in **Figure 4.1a**; mechanism shown in **Figure 4.1b**[20]). However, Cas9 presents significant limitations for biomedical applications. The cistron encoding Cas9 is ~4.1 kb, while the current state-of-the-art gene delivery system—adeno-associated virus (AAV)—can package at most ~4.7 kb (**Figure 4.2**).[21] Combining SpCas9 and one or more sgRNAs, alongside regulatory sequences for proper gene expression, into a single AAV genome is challenging at best.[22] Moreover, the limited DNA packaging capacity of AAV precludes encoding fusions of SpCas9 with effector domains into a single vector, preventing critical therapeutic applications. Beyond AAV, other delivery systems that continue to be developed will likewise benefit from a smaller cargo. Furthermore, wild-type Cas9 also suffers from a fundamental limitation in which genome sites can be targeted.[2, 19-20] Canonically, CRISPR systems can only act at genetic sites that contain a Cas-specific protospacer adjacent motif (PAM). For Cas9, this means that only genetic sites that end in a GG dinucleotide can be targeted. While this is likely not a limitation for most gene editing applications, it may present a challenge for base editing approaches, which require the Cas9-deaminase fusion protein to be targeted to a small window of DNA sequences.

Thousands of Cas proteins have been discovered through genomic mining of bacterial genomes.[23] Of particular interest for biomedical applications are mini-CRISPRs, or CRISPR-Cas systems between 400–600 amino acids in length (**Figure 4.2**).[24-26] Despite efficacious activity *in vitro* and in bacteria, these mini-CRISPRs almost universally show poor activity in mammalian cells, erasing their therapeutic potential. Engineering new mini-CRISPRs with improved activity in mammalian cells would enable AAV delivery of CRISPR fusion proteins, thereby furnishing previously unachievable therapeutic modalities.
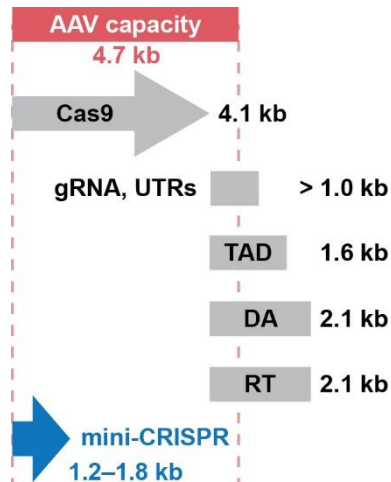
Figure 4.2 Challenges associated with delivering CRISPR systems using adeno-associated virus. Adeno-associated virus (AAV) has a packaging capacity of ~4.7 kb. It is challenging to package Cas9 with a gRNA and appropriate regulatory untranslated regions (UTRs) into a single AAV vector—and currently impossible to package Cas9-fusion proteins with valuable additional effector domains such as transcriptional activation domains (shown here, a highly active 1.6 kb transcriptional activation domain termed VPR), deaminase proteins (DA; shown here, a 2.1 kb adenosine deaminase gene), or reverse transcriptase (RT, used in prime editing). Mini-CRISPRs are Class II CRISPR systems 400-600 amino acids in length that could enable the packaging of CRISPR fusion proteins into a single AAV vector.
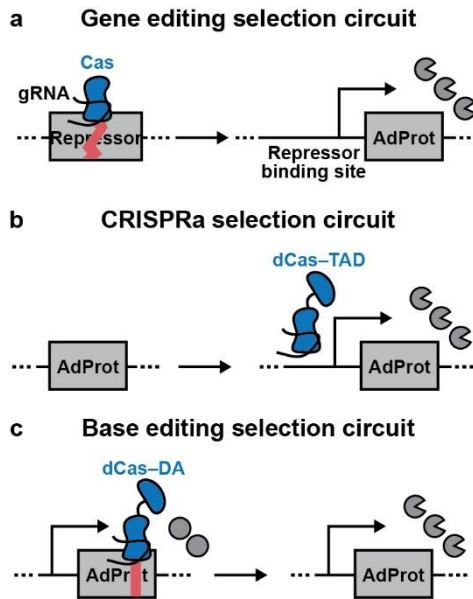
4.3 Directed evolution of CRISPR systems

Efforts to engineer new CRISPR activities through directed evolution have been pursued in bacteria. Continuous directed evolution methods in bacteria were used to create Cas9 proteins with expanded PAM sequences. The resulting expanded Cas9 variants, termed xCas9-3.6 and xCas9-3.7,[27] work well in prokaryotes but lose their expanded functionalities in mammalian cells.[28] As a result, efforts to engineer CRISPR systems usually rely on structural information, limiting researchers to a small number of structurally characterized CRISPR systems including the overwhelming majority of mini-CRISPRs.[19, 27, 29-30] Yet even when structural information is used, the resulting engineered proteins also tend to have decreased activity in mammalian cells, once again highlighting the necessity to perform biomolecule optimization in the intended environment of activity.

The setbacks and limitations of previous CRISPR engineering efforts to created engineered CRISPR systems that maintain high activity in mammalian cells present a compelling use-case for our directed evolution platform. Our system evolves proteins directly in the mammalian cellular context, ensuring that protein variants are highly active in the desired environment.[31] Our system has a large trans-gene packaging capacity, enabling us to evolve large genes such as Cas9 and even some Cas9 fusions. The flexibility of wide-ranging activity of CRISPR and the CRISPR toolbox writ large provide ample room to creatively engineer selection circuits that can be modified to evolve a variety of desired functionalities. Lastly, the parallelizable nature of our platform allows us to pursue many CRISPR systems and CRISPR activities simultaneously, including those without prior structural information.

4.4 Adapting our virus-based platform for the directed evolution of CRISPR systems

Due to the wide variety of activities CRISPR systems have been engineered to perform,[11] there are countless synthetic genetic circuits that could be used to evolve CRISPR systems.[32] Before adapting our virus-based platform toward the directed evolution of CRISPR systems, we first evaluated the benefits of three potential selection circuits (**Figure 4.3**).

a   **Gene editing selection circuit**

b   **CRISPRa selection circuit**

c   **Base editing selection circuit**

Figure 4.3 Adapting our platform toward the directed evolution of CRISPR systems. (a) A gene editing selection circuit. A stably integrated repressor gene prevents constitutive transcription of AdProt. If an active Cas protein can disrupt the repressor gene, transcriptional repression will be relieved and AdProt will be expressed. (b) A CRISPR activation (CRISPRa) selection circuit. AdProt is stably integrated into the host cell genome without being actively transcribed. Active CRISPRa systems—such as a dead Cas protein (dCas) fused to a transcriptional activation domain (TAD)—will induce transcription and ultimately expression of AdProt. (c) A base editing selection circuit. A stably integrated non-functional AdProt variant is constitutively expressed from the host cell genome. If an active base editor—such as a dCas protein fused to a deaminase domain—restores the functional AdProt gene.

First, we considered a gene editing-based selection circuit (**Figure 4.3a**). In one implementation of such a circuit, the host cell encodes a repressor that binds upstream of AdProt, preventing constitutive transcription. If the virally-encoded CRISPR system deactivates the repressor through gene editing, AdProt will be constitutively expressed and the virus will replicate. The main benefits for using gene editing with our platform are: (1) the Cas protein requires minimal engineering; (2) the circuit depends upon natural CRISPR activity; (3) the flexibility in repressor choice allows additional levers of control, including small-molecule control over basal expression levels.

Second, we considered a transcriptional activation-based selection circuit (**Figure 4.3b**). In one implementation of such a circuit, the viral genome encodes a CRISPR activation (CRISPRa)[13] construct and the host cell encodes AdProt without constitutive transcription. Viruses will replicate upon CRISPRa-based transcription of AdProt. The main benefits for using CRISPRa are: (1) it is compatible with all of our previously developed cell lines with minimal cell engineering required; (2) CRISPRa systems demonstrate tremendous dynamic range; (3) CRISPRa has previously been used to drive a selection circuit for virus-based directed evolution in bacteria.[27]

Third, we considered a base editing-based selection circuit (**Figure 4.3c**). In one implementation of such a circuit, the host cell encodes a constitutively express but catalytically dead (or truncated) AdProt gene, upon which the virally encoded base editor[15] will act and restore AdProt to functionality. The main benefits for base editing are: (1) theoretically, extremely low basal rates of replication; (2) selects for precise editing with high activity; (3) could be implemented to select for new base editing activities, such as generating transversions as opposed to transitions.

Ultimately, we decided to initially pursue a CRISPRa-based selection circuit, since such a circuit would be readily compatible with our previously engineered cell lines and could be easily adapted to mini-CRISPR systems. However, future implementations of CRISPR evolution should consider these and other selection circuits to enable the evolution of new activities or to improve upon our current circuits.

4.4.1 Validating CRISPRa in our engineered cell lines. Previous efforts to evolve CRISPR systems using continuous viral evolution employed CRISPR activation (CRISPRa) to create a functional selection circuit. In the development of xCas9-3.6 and xCas9-3.7,[27] Liu and co-workers fused a bacterial transcriptional activation domain (TAD) on a catalytically dead variant of Cas9 (dCas9), and made the propagation of M13 bacteriophage dependent upon

dCas9-TAD induced transcription of the essential M13 protein pIII. We rationalized that a similar concept could be adapted to create a selection circuit in mammalian cells.

Numerous implementations of CRISPRa have been previously developed, with engineering and activity trade-offs. The first reports of CRISPRa involved fusing dCas9 to VP64,[33] a common mammalian TAD consisting of four copies of the widely-used transcriptional activation domain of herpes simplex virus protein 16. Even though VP64 resulted in high activity for numerous other engineered transcription factors, dCas9-VP64 activated transcription relatively poorly, sparking a search for more active systems. Three systems, all with similar levels of activity,[34] eventually proved to work most efficiently in mammalian cells: SAM,[35] SunTag,[36] and VPR.[37] In choosing between these three systems, we ultimately decided upon VPR—a fusion of VP64, the TAD of the human transcription factor NF-κB termed p65, and the TAD of the Epstein-Barr virus R transactivator termed rTA—because of its modularity, its capacity to be immediately implemented into our system without significant host cell engineering, and the fact that it does not rely on structural information for proper design.

To validate that CRISPRa could be implemented in our cell lines, our first goal was to demonstrate that we could use dCas9-VPR to activate transcription in our HEK293A cell lines. This initial test required us to optimize gRNA cloning, dCas9-VPR and gRNA expression, and CRISPRa activity in general. We used previously verified gRNA sequences[34] to activate transcription of TTN and ASCL1, successfully optimizing our transfection protocol and observing high levels of transcriptional activation in HEK293A cells (**Figure 4.4**). Observing a >700-fold increase in ASCL1 levels was greatly encouraging, and indicated that we may be able to observe high levels of AdProt induction with CRISPRa.
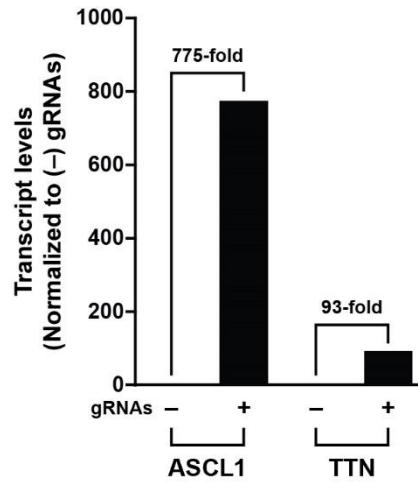
Figure 4.4 CRISPR-based transcriptional activation of endogenous genes. A CRISPRa system, dCas9-VPR, activated transcription of two endogenous human genes: Achaete-scute homolog 1 (ASCL1) and titin (TTN).

4.4.1 A suitable host cell for CRISPRa-based viral selection. As a result of generating various selection circuits in Chapters 2 and 3, we had previously developed a number of cell lines that we could potentially use as a host for a CRISPRa-based selection circuit.

In Chapter 3, we observed that Xfect-CRE-AdProt cells #3, #4, and #5 showed remarkably low levels of AdProt expression, which we realized was key to the successful demonstration of a selection circuit. Therefore, we rationalized that these cells could be ideal for a CRISPRa-based directed evolution circuit. However, because we generated these cells through random plasmid integration, we were unsure to what extent the plasmid incorporated into the genome. If the plasmid integrated incompletely or with substantial defects, we would have to accordingly modify our design of gRNAs that target AdProt. We harvested genomic DNA from these cellular populations and PCR amplified regions of DNA that encoded successively larger regions of the stably transfected CRE-AdProt plasmid (**Figure 4.5a**). We observed that all three cell lines had at least 1,000 bp of DNA upstream of AdProt integrated into cellular genome (**Figure 4.5b**), permitting us to proceed with designing gRNAs for all three cell lines.

We next tested whether we could activate AdProt from Xfect-CRE-AdProt #3, #4, and #5 cells (Chapter 3) using CRISPRa. We designed nine gRNAs to target dCas9-VPR to the 5′ UTR of AdProt between 80–250 bp before the transcription start site. We transfected a dCas9-VPR-encoding plasmid and all nine gRNAs, and measured AdProt levels through qPCR (**Figure 4.6**). While saw modest activation for cell lines #3 and #4, we saw a 97-fold increase in AdProt levels in cell line #5. From these three cells, we observed not only the lowest basal levels of AdProt expression in cell line #5, but also the highest levels of AdProt expression upon activation.
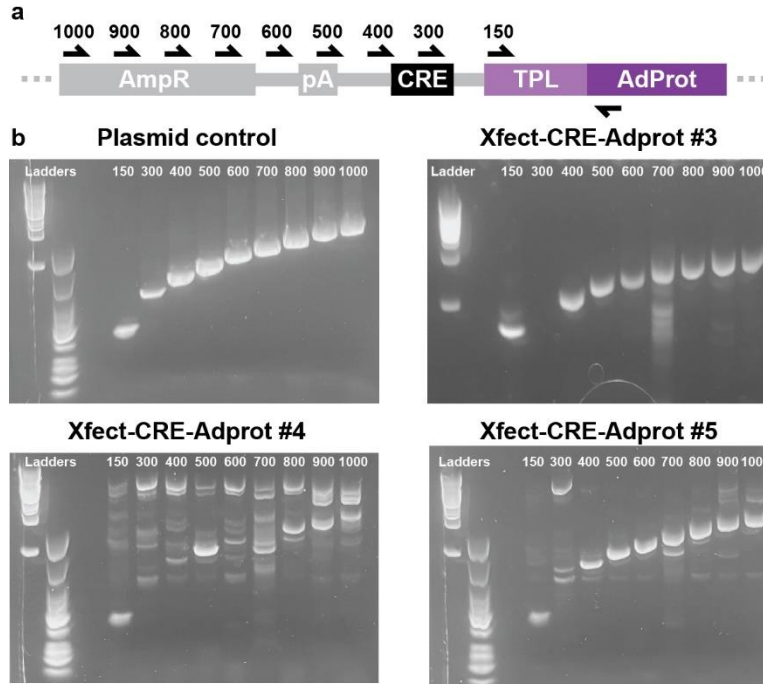
Figure 4.5 Characterizing the extent of genomic integration of a stably transfected plasmid in Xfect-CRE-AdProt cells. (a) A portion of the plasmid map for the genetic construct used to stably transfect Xfect-CRE-AdProt cells. DNA primers were spaced throughout the 5′ untranslated region and in the AdProt gene itself. PCR amplification was performed on prepared genomic DNA to determine the extent of genomic integration. (b) DNA agarose gels of the PCR amplification products for the plasmid by itself and three homostable populations of Xfect-CRE-AdProt cells (#3, #4, and #5). Lanes in each gel, from left to right: 1 kb ladder and 100 bp ladder (except #3), then PCR amplification products ranging from 150 bp to 1000 bp.
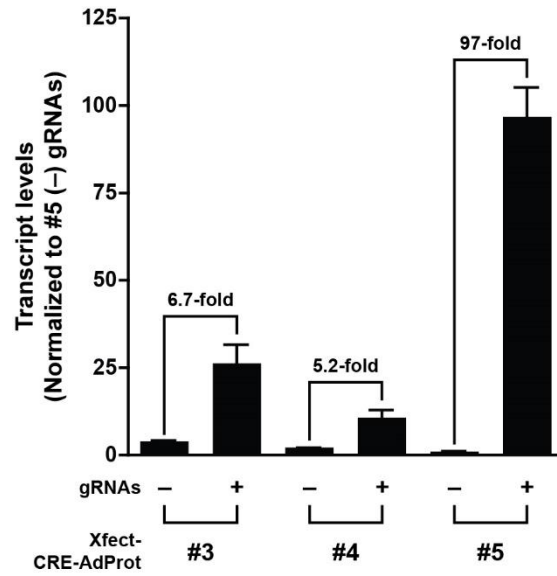
Figure 4.6 CRISPR-based transcriptional activation of the exogenous adenoviral protease.
A CRISPRa system, dCas9-VPR, activated transcription of exogenously encoded AdProt in three single colonies of Xfect-CRE-AdProt cells (#3, #4, and #5).

We next tested whether dCas9-VPR-encoding adenoviral replication on these cell lines was dependent upon gRNA expression. We infected cell lines #3, #4, and #5 with virus with or without gRNA expression, and monitored viral replication for 10 days before harvesting the virus (**Figure 4.7a**). We observed that the dCas9-VPR-encoding virus (which also encodes mCherry, a red fluorescent protein) could replicate robustly dependent on the presence of the nine gRNAs. We then quantified viral replication through plaque assay (**Figure 4.7b**), and observed that the virus was highly dependent on gRNA expression in all three cell lines. Cell line #5 stood out, with extremely low viral replication in the absence of gRNAs, as we were unable to observe any viral replication in our plaque assays. In an effort to quantify the low background rate of cell line #5, we used flow cytometry to quantify the concentration of infectious particles, and observed a 3,500-fold difference in viral titer between cells with and without gRNA expression (**Figure 4.7c**).
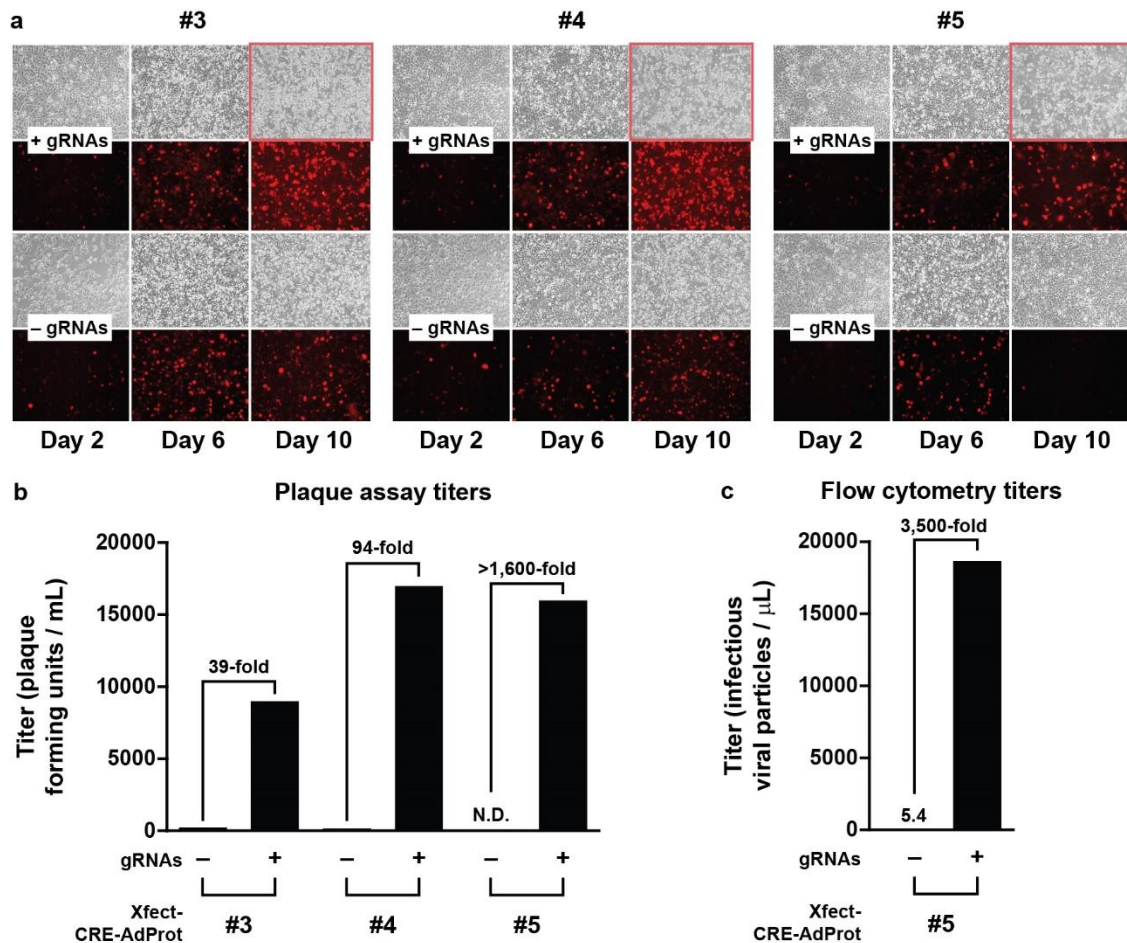
Figure 4.7 CRISPRa-dependent viral replication. (a) Xfect-CRE-AdProt cells were infected with dCas9-VPR- and mCherry-encoding adenovirus, either in the presence (top) or absence (bottom) of gRNAs targeted to AdProt. Images shown were taken on day 2, day 6, and day 10. The appearance of cell death in brightfield images (see red boxes) and the spreading of fluorescent over the course of the ten-day infection period indicates viral replication only in the presence of gRNAs. (b) After ten days of infection, viruses were harvested and the resulting viral titers were measured through a plaque assay. (c) For cell line #5, flow cytometry was used to determine viral titer, owing to the technique's higher sensitivity.

4.4.2 Implementation of two CRISPRa-based selection circuits. With these data in hand, we next sought to create evolutionary selection circuits that would select for valuable new activities of Cas9. Evolving Cas9 would be valuable not only as a proof-of-principle for our ultimate targets, such as mini-CRISPRs, but also because Cas9 remains the best-performing system in mammalian cells, and new engineering efforts of Cas9 for activity specifically in mammalian cells would be impactful in its own right. Toward this end, we created two selection circuits for evolving Cas9 toward new activities (**Figure 4.8**).

The first selection circuit is designed to evolve Cas9 variants with expanded PAM capacities (**Figure 4.8a**). Cas9 shows a strong preference for gRNAs that target sequences with an NGG PAM, and shows poor activity at non-NGG PAMs. Previous efforts to evolve Cas9 variants with expanded PAMs in bacteria have created variants that do not retain the expanded PAM preferences and have lower activity in mammalian cells.[28] Therefore, PAM expansion is an attractive target for directed evolution in mammalian cells, where such issues of translation are less likely to occur. We engineered 9–10 gRNAs to target AdProt for each the following non-NGG PAMs: NGA, NAG, NAA, and NGT. We then tested the ability of dCas9-VPR-encoding viruses to replicate on Xfect-CRE-AdProt cell lines #3, #4, and #5 in the presence of gRNAs targeting NGG or non-NGG PAMs, and measured the resulting viral titers through plaque assay (**Figure 4.9**). We observed robust replication for cells expressing gRNAs that target NGG PAMs and much poorer replication for all other PAM sequences. Most notably, cell line #5 showed the highest level of selection—both producing the highest titer with NGG gRNAs and the lowest titers with all other non-NGG PAM gRNAs. Encouragingly, we also observed that cell lines #3 and #4 showed more permissive replication at non-NGG PAMs, demonstrating that we can modulate selection pressure by choosing the appropriate cell line. These data clearly show that significant selection pressure exists for dCas9-VPR-encoding viruses to evolve activities expanded PAM.
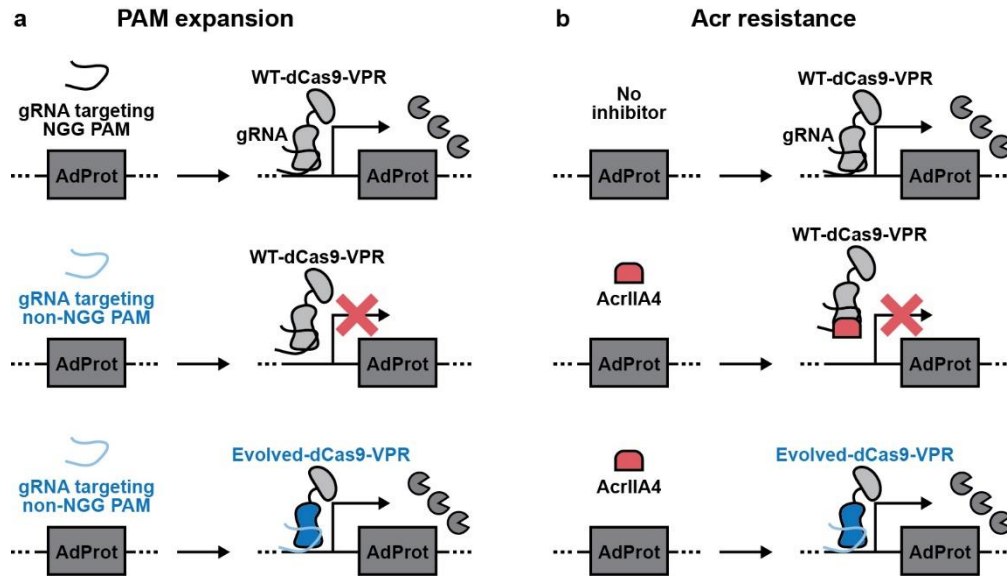
Figure 4.8 Selection circuits for the directed evolution of Cas9 in mammalian cells. (a) A selection circuit to expand the PAM sequences targetable by Cas9. Wild-type (WT) dCas9-VPR activates transcription of AdProt when complexed to gRNAs that target NGG PAMs. While WT-dCas9-VPR will not activate transcription at non-NGG PAMs, an evolved dCas9-VPR variant with expanded PAM activity would be able to activate transcription of AdPot, thereby enabling viral replication. (a) A selection circuit to create AcrIIA4-resistant Cas9 variants. While WT-dCas9-VPR is inhibited from activating transcription in the presence of the anti-CRISPR protein AcrIIA4, an evolved dCas9-VPR variant would be able to activate transcription of AdPot in the presence of AcrIIA4, thereby enabling viral replication.
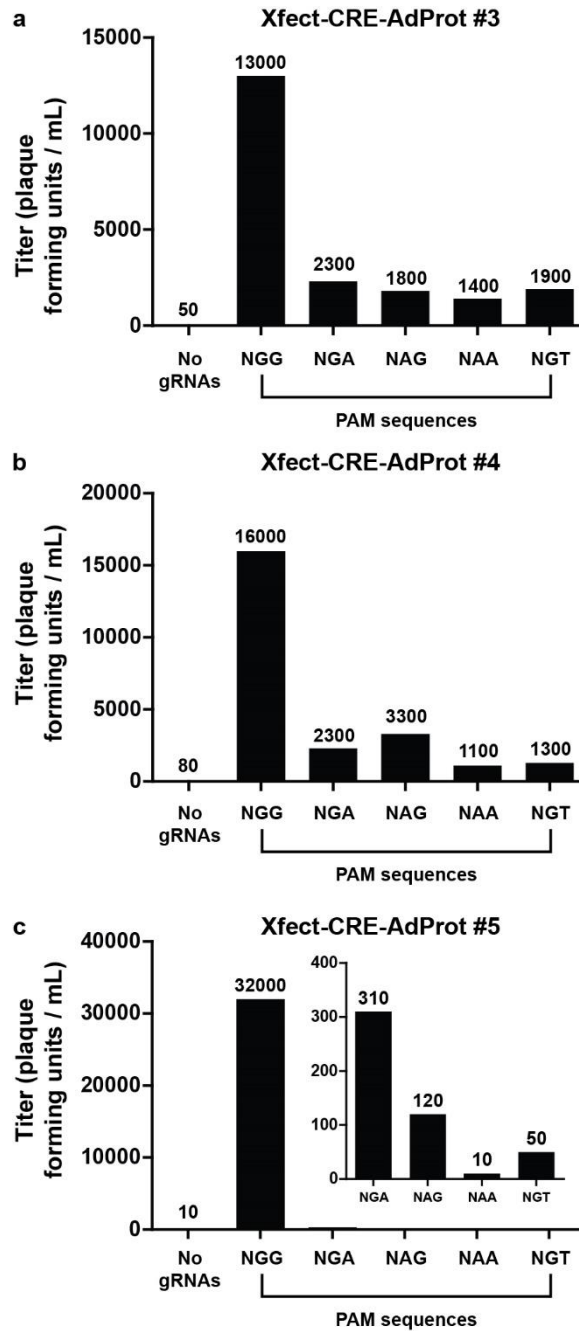
Figure 4.9 Characterizing selection pressure in a PAM expansion selection circuit. Xfect-CRE-AdProt cells (a) #3, (b) #4, and (c) #5 were infected with dCas9-VPR-encoding adenovirus in the presence of gRNAs targeting different PAM sequences. After ten days of infection, viruses were titered using a plaque assay.
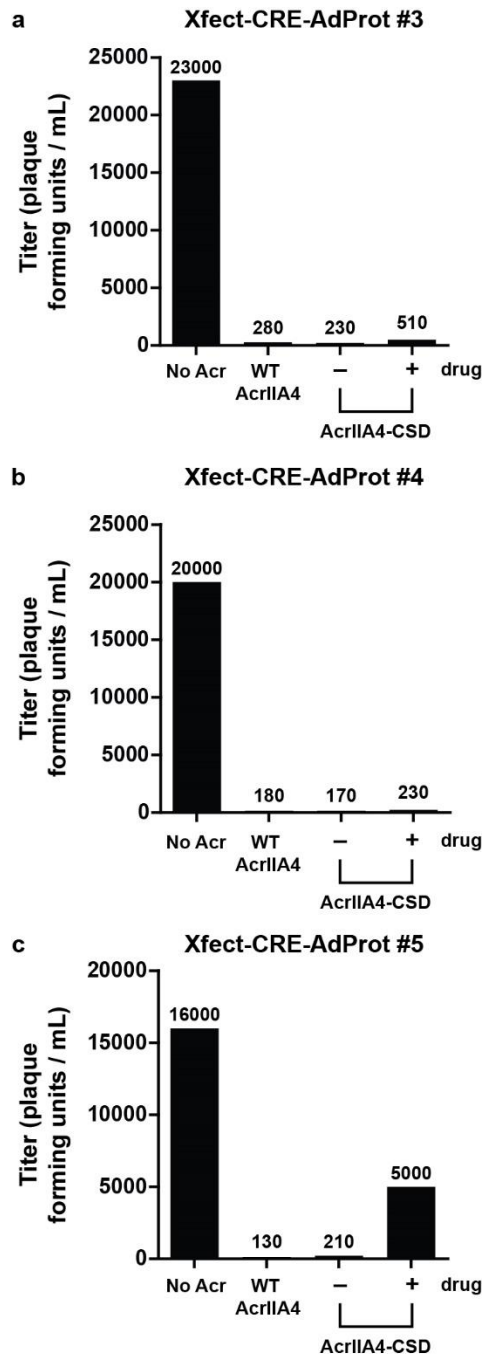
Figure 4.10 Characterizing selection pressure in an AcrIIA4 inhibition selection circuit. Xfect-CRE-AdProt cells (a) #3, (b) #4, and (c) #5 were infected with dCas9-VPR-encoding adenovirus in the presence of various AcrIIA4 constructs, including wild-type AcrIIA4 and AcrIIA4-CSD, a variant of AcrIIA4 that is rapidly degraded in the presence of pomalidomide (drug). After ten days of infection, viruses were tittered using a plaque assay.

The second selection circuit is designed to evolve Cas9 variants that are resistant to anti-CRISPR proteins (Acrs; **Figure 4.8b**). Acrs, proteins that inhibit the activity of CRISPR systems,[38] have emerged as an intriguing aspect of the evolutionary arms race between bacteriophage and the CRISPR-encoding bacteria they infect.[39] We were intrigued by the possibility of evolving Cas9 to resist AcrIIA4[40] (the most potent Acr against Cas9), both as a straightforward proof-of-principle evolution experiment and as a way to glean insight into the mechanism of AcrIIA4 inhibition. We tested the ability of dCas9-VPR-encoding viruses to replicate on Xfect-CRE-AdProt cell lines #3, #4, and #5 in the presence of AcrIIA4, and measured the resulting viral titers through plaque assay (**Figure 4.10**). We observed strong inhibition of viral replication in the presence of AcrIIA4. We also tested the ability of a variant of AcrIIA4 (AcrIIA4-CSD) to inhibit Cas9 activity. AcrIIA4-CSD is degraded in the presence of the drug pomalidomide in a dose-dependent manner, potentially providing small molecule control of selection pressure. While we observed that viruses replicated poorly in the presence of AcrIIA4-CSD, once again we demonstrated that we can modulate selection pressure, in this case through drug treatment.

Taken together, these data strongly confirm that we are able to adapt our platform toward the directed evolution of CRISPR systems.

4.5 Discussion

While CRISPR systems have enormous potential to impact human health, current systems require improvements—improvements that also critically depend on activity in mammalian cells. We have demonstrated the viability of a directed evolution circuit for evolving CRISPR activities directly in human cells with a continuous, virus-based method. We demonstrated two selection circuits to evolve valuable CRISPR functions, expanded PAM compatibility and Acr resistance.

The CRISPRa-based selection circuits we have developed dramatically outshine those developed in Chapters 2 and 3. While the successful tTA-based selection circuit showed ~10-fold change in viral replication, and the GPCR-based selection circuit showed ~2-fold change, the two CRISPRa-based selection circuits showed a remarkable >1,000-fold change in viral titers upon applying selection pressure. Even the increases in AdProt levels are striking—whereas tTA and GPCRs demonstrated at most ~10-fold increases in AdProt levels, CRISPRa demonstrated ~100-fold increase in AdProt levels. Unlike tTA and GPCRs, we were not required to make bespoke cell lines to demonstrate these selection circuits, rapidly accelerating our experimental timelines and allowing us to quickly characterize a broad range of extant cell lines.

Even with these promising developments, improvements to our CRISPR-based selection circuits are well within reach. Simply considering CRISPRa, improvements abound. The 775-fold

increase in ASCL1 levels (Figure 4.4) is nearly an order of magnitude higher than the 97-fold increase in AdProt levels in Xfect-CRE-AdProt cells, and increases of 10,000-fold have been previously observed with the dCas9-VPR system.[34, 37] New stably transfected cells could be created that allow for small molecule control over AdProt levels, allowing for the precise modulation of selection pressure over the course of an evolution. Lastly, simply picking new homostable single colonies from the transfected heterostable population may reveal cells with ever wider dynamic ranges. Beyond CRISPRa, additional selection circuits (as in **Figure 4.3** and beyond) could provide improved selection pressure and would allow for the evolution of different CRISPR activities.

Even without these improvements, the selection circuits position us well to perform many valuable directed evolution campaigns, from Acr resistance to expanded PAM sequences to the evolution of highly efficacious mini-CRISPRs. Most importantly, by performing these evolutions directly in mammalian cells, we ensure that the evolved proteins will function in the environment in which they are intended to function. Undoubtedly, the future of this project is bright.

4.5 Methods

4.5.1 Cloning methods. All PCR reactions for cloning cassettes were performed using Q5 High-Fidelity DNA Polymerase (New England BioLabs). The dCas9-VPR construct was obtained from Addgene, deposited by George Church (Addgene plasmid #63798).[37] The parent gRNA expression plasmid was obtained from Addgene, deposited by Charles Gersbach (Addgene plasmid #47108).[41] The AcrIIA4 plasmid was obtained from Amit Choudhary.[42] The AcrIIA4-CSD plasmid, with a C-terminal degradation tag termed superdegron,[43] was also obtained from Amit Choudhary. Primers are listed in **Table 4.1**.

Table 4.1 Table of primers.

Primer name	Sequence
dCas9-FRAG-FOR	TGGATCCGCCACCATGGACAAGAAGTACT
dCas9-FRAG-REV	GGGTCTAGATATTCAAAACAGAGATGTGTCGAAGATGG
pENT-VECT-FOR	CTCTGTTTTGAATATCTAGACCCAGCTTTCTTGTACAAAGT
pENT-VECT-REV	ATGGTGGCGGATCCAGTCGACTGAATTGG
TTN-qPCR-FWD	TGTTGCCACTGGTGCTAAAG
TTN-qPCR-REV	ACAGCAGTCTTCTCCGCTTC
ASCL1-qPCR-FWD	CGCGGCCAACAAGAAGATG
ASCL1-qPCR-REV	CGACGAGTAGGATGAGACCG
AdProt-qPCR-FWD	GAAAAGTCCACCCAAAGCGT
AdProt-qPCR-REV	CTGTTGAGCATGGAGTTGGG
RPLP2-qPCR-FWD	CGTCGCCTCCTACCTGCT
RPLP2-qPCR-REV	CCATTCAGCTCACTGATAACCTT

The dCas9-VPR construct was first cloned into pENTR1A using Gibson cloning with the HiFI Assembly Kit (New England BioLabs; primers dCas9-FRAG-FOR and dCas-FRAG-REV for dCas9-VPR amplification and pENT-VECT-FOR and pENTR-VECT-REV for pENTR1A vector amplification). Then, dCas9-VPR was cloned into DEST40 (ThermoFisher) and AdEvolveDEST (Chapter 2) using Gateway cloning (ThermoFisher).

4.5.2 General cell culture. Cells were cultured at 37 °C and 5% CO_2 (g). New cell lines were derived from a parent HEK293A cell line (Thermo Fisher) 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). The "producer," (Chapter 2), Xfect-CRE-AdProt, and Xfect-CRE-Luc (Chapter 3) cell lines were cultured in 50 μg/mL hygromycin (Thermo Fisher) to stably maintain transgenes.

4.5.3 Generation of cell lines. Producer cells were generated as described in Chapter 2.[31] Xfect-CRE-AdProt and Xfect-CRE-AdProt cells were generated as described in Chapter 3.

4.5.4 Quantitative PCR (qPCR). In a typical protocol, $\sim$5 × 10^5 cells are plated on tissue culture-treated 12-well plates. The next day, wells were transfected with 1 μg of total DNA using 3 μL of TransIT-Lenti transfection reagent (Mirus Bio). For experiments using dCas9 and gRNAs (including gRNAs targeting NGG and non-NGG PAMs), 0.5 μg of dCas9-VPR-DEST40 plasmid and 0.5 μg of a pooled panel of 9–10 gRNA expression plasmids were transfected. For experiments transfecting dCas9, gRNAs and AcrIIA4 constructs, 0.5 μg of dCas9-VPR-DEST40 plasmid, 0.3 μg of a pooled panel of 9–10 gRNA expression plasmids, and 0.2 μg of plasmids encoding AcrIIA4 constructs were transfected. 72 h post-expression, 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 a high-capacity cDNA reverse transcription kit (Applied Biosystems). qPCR analysis for TTN (primers TTN-qPCR-FWD and TTN-qPCR-REV), ASCL1 (primers ASCL1-qPCR-FWD and ASCL1-qPCR-REV), AdProt (primers AdProt-qPCR-FWD and AdProt-qPCR-REV), and the housekeeping gene RPLP2 (primers RPLP2-qPCR-FWD and RPLP2-qPCR-REV) was performed on a LightCycler 480 II (Roche). Transcript levels were normalized to intrasample RPLP2 levels before comparing between samples.

4.5.5 Generation of adenovirus. Adenoviruses were produced by transfecting a PacI (New England BioLabs)-linearized AdEvolveDEST vector into appropriate trans-complementing HEK293A cells ("producer" cells; Chapter 2). A 10 μg sample of PacI-linearized adenovirus vectors mixed with 30 μL of TransIT-Lenti transfection reagent (Mirus Bio) was added to a 10 cm plate of producer cells ($\sim$8 × 10^6 cells). The medium was intermittently replaced every 2–3 days until plaques were observed (typically $\sim$3 weeks). Once plaques were detected, full 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 3200 × g for 15 min and the supernatant was stored at −80 °C. Viruses were then passaged on producer cells to produce a high-titer stock of virus which was used to initiate directed evolution experiments.

4.5.6 Viral replication assay. For viral replication assays, ~5 × 10^5 cells were plated on tissue culture-treated 6 well plates. The next day, media was exchanged and 10 µM carbachol (Sigma Aldrich) was added if appropriate, and then cells were infected with 20 µL of a high titer stock of dCas9-VPR-encoding adenovirus. Pictures were taken every two days. After 10 days, viral media was harvested and subjected to three freeze-thaw cycles. The cell debris was removed by centrifugation at 3200 × g for 15 min and the supernatant was stored at −80 °C.

4.5.7 Viral titering through flow cytometry. For viral replication assays, ~5 × 10^5 cells were plated on tissue culture-treated 12-well plates. The next day, known volumes of viral supernatants were added to cells. 72 h post-infection, the cells were washed once with medium, stained with 0.2 µg/mL DAPI (4′,6-diamidino-2-phenylindole, Thermo Fisher), 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 0.1–10% of the cells were infected.

4.5.8 Viral titering through plaque assay. For viral replication assays, ~5 × 10^5 producer cells were plated on tissue culture-treated 6-well plates. The following day, 100 µL of viral supernatants of various dilutions (across six orders of magnitude from 10^0 to 10^{-6}) were added to cells. Six hours post-infection, media was taken off the wells and replaced with media with 0.4% agarose. Infected wells were monitored to observe the spreading viral plaques until they were large enough to be visible bye eye (roughly 3 weeks). Once plaques were observable, the wells were stained with MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium Bromide, 50 mg/10 mL in PBS; Research Products International) and developed for 2 h, after which macroscopically observable viral plaques were counted. To maximize the accuracy of the viral titer, data were only considered if they fell within the linear range, which typically encompassed dilutions where 5–50 viral plaques were counted in an individual well.

4.6 References

1.   Doudna Jennifer, A.; Charpentier, E., The new frontier of genome engineering with CRISPR-Cas9. *Science* **2014,** *346* (6213), 1258096.

2.   Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E., A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **2012,** *337* (6096), 816-21.

3.   Jacobs, J. Z.; Ciccaglione, K. M.; Tournier, V.; Zaratiegui, M., Implementation of the CRISPR-Cas9 system in fission yeast. *Nat. Commun.* **2014,** *5*, 5344.

4.   Li, J.-F.; Norville, J. E.; Aach, J.; McCormack, M.; Zhang, D.; Bush, J.; Church, G. M.; Sheen, J., Multiplex and homologous recombination–mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nat. Biotechnol.* **2013,** *31* (8), 688-691.

5.   Nekrasov, V.; Staskawicz, B.; Weigel, D.; Jones, J. D. G.; Kamoun, S., Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **2013,** *31* (8), 691-693.

6.   Shan, Q.; Wang, Y.; Li, J.; Zhang, Y.; Chen, K.; Liang, Z.; Zhang, K.; Liu, J.; Xi, J. J.; Qiu, J.-L.; Gao, C., Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* **2013,** *31* (8), 686-688.

7.   Cong, L.; Ran, F. A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P. D.; Wu, X.; Jiang, W.; Marraffini, L. A.; Zhang, F., Multiplex genome engineering using CRISPR/Cas systems. *Science* **2013,** *339* (6121), 819-23.

8.   Frangoul, H.; Altshuler, D.; Cappellini, M. D.; Chen, Y.-S.; Domm, J.; Eustace, B. K.; Foell, J.; de la Fuente, J.; Grupp, S.; Handgretinger, R.; Ho, T. W.; Kattamis, A.; Kernytsky, A.; Lekstrom-Himes, J.; Li, A. M.; Locatelli, F.; Mapara, M. Y.; de Montalembert, M.; Rondelli, D.; Sharma, A.; Sheth, S.; Soni, S.; Steinberg, M. H.; Wall, D.; Yen, A.; Corbacioglu, S., CRISPR-Cas9 gene editing for sickle cell disease and β-thalassemia. *N. Engl. J. Med.* **2020,** *384* (3), 252-260.

9.   Min, Y. L.; Bassel-Duby, R.; Olson, E. N., CRISPR correction of Duchenne muscular dystrophy. *Annu. Rev. Med.* **2019,** *70*, 239-255.

10.   Ledford, H., CRISPR treatment inserted directly into the body for first time. *Nature* **2020,** *579* (7798), 185.

11.   Adli, M., The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* **2018,** *9* (1), 1911.

12.   Moore, C. L.; Dewal, M. B.; Nekongo, E. E.; Santiago, S.; Lu, N. B.; Levine, S. S.; Shoulders, M. D., Transportable, chemical genetic methodology for the small molecule-mediated inhibition of heat shock factor 1. *ACS Chem. Biol.* **2016,** *11* (1), 200-210.

13.   Casas-Mollano, J. A.; Zinselmeier, M. H.; Erickson, S. E.; Smanski, M. J., CRISPR-Cas activators for engineering gene expression in higher eukaryotes. *CRISPR J.* **2020,** *3* (5), 350-364.

14.   Goell, J. H.; Hilton, I. B., CRISPR/Cas-based epigenome editing: advances, applications, and clinical utility. *Trends Biotechnol.* **2021,** *39* (7), 678-691.

15.   Anzalone, A. V.; Koblan, L. W.; Liu, D. R., Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **2020,** *38* (7), 824-844.

16.   Koonin, E. V.; Makarova, K. S., Origins and evolution of CRISPR-Cas systems. *Philos. Trans. R. Soc. B: Biol. Sci.* **2019,** *374* (1772), 20180087.

17.   Zetsche, B.; Gootenberg, J. S.; Abudayyeh, O. O.; Slaymaker, I. M.; Makarova, K. S.; Essletzbichler, P.; Volz, S. E.; Joung, J.; van der Oost, J.; Regev, A.; Koonin, E. V.; Zhang, F., Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **2015,** *163* (3), 759-71.

18.   Hendel, S. J.; Shoulders, M. D., Directed evolution in mammalian cells. *Nat. Methods* **2021,** *18* (4), 346-357.

19.   Nishimasu, H.; Ran, F. A.; Hsu, Patrick D.; Konermann, S.; Shehata, Soraya I.; Dohmae, N.; Ishitani, R.; Zhang, F.; Nureki, O., Crystal Structure of Cas9 in complex with guide RNA and target DNA. *Cell* **2014,** *156* (5), 935-949.

20.   Jiang, F.; Doudna, J. A., CRISPR–Cas9 structures and mechanisms. *Annu. Rev. Biophys.* **2017,** *46* (1), 505-529.

21.   Wang, D.; Tai, P. W. L.; Gao, G., Adeno-associated virus vector as a platform for gene therapy delivery. *Nat. Rev. Drug Discov.* **2019,** *18* (5), 358-378.

22.   Chew, W. L.; Tabebordbar, M.; Cheng, J. K. W.; Mali, P.; Wu, E. Y.; Ng, A. H. M.; Zhu, K.; Wagers, A. J.; Church, G. M., A multifunctional AAV–CRISPR–Cas9 and its host response. *Nat. Methods* **2016,** *13* (10), 868-874.

23.   Shmakov, S.; Smargon, A.; Scott, D.; Cox, D.; Pyzocha, N.; Yan, W.; Abudayyeh, O. O.; Gootenberg, J. S.; Makarova, K. S.; Wolf, Y. I.; Severinov, K.; Zhang, F.; Koonin, E. V., Diversity and evolution of class 2 CRISPR–Cas systems. *Nat. Rev. Microbiol.* **2017,** *15* (3), 169-182.

24.   Kim, D. Y.; Lee, J. M.; Moon, S. B.; Chin, H. J.; Park, S.; Lim, Y.; Kim, D.; Koo, T.; Ko, J.-H.; Kim, Y.-S., Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by adeno-associated virus. *Nat. Biotechnol.* **2022,** *40* (1), 94-102.

25.   Wu, Z.; Zhang, Y.; Yu, H.; Pan, D.; Wang, Y.; Wang, Y.; Li, F.; Liu, C.; Nan, H.; Chen, W.; Ji, Q., Programmed genome editing by a miniature CRISPR-Cas12f nuclease. *Nat. Chem. Biol.* **2021,** *17* (11), 1132-1138.

26.   Xu, X.; Chemparathy, A.; Zeng, L.; Kempton, H. R.; Shang, S.; Nakamura, M.; Qi, L. S., Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. *Mol. Cell* **2021,** *81* (20), 4333-4345.e4.

27.   Hu, J. H.; Miller, S. M.; Geurts, M. H.; Tang, W.; Chen, L.; Sun, N.; Zeina, C. M.; Gao, X.; Rees, H. A.; Lin, Z.; Liu, D. R., Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **2018,** *556* (7699), 57-63.

28.   Kim, H. K.; Lee, S.; Kim, Y.; Park, J.; Min, S.; Choi, J. W.; Huang, T. P.; Yoon, S.; Liu, D. R.; Kim, H. H., High-throughput analysis of the activities of xCas9, SpCas9-NG and SpCas9 at matched and mismatched target sequences in human cells. *Nat. Biomed. Eng.* **2020,** *4* (1), 111-124.

29.   Kleinstiver, B. P.; Prew, M. S.; Tsai, S. Q.; Topkar, V. V.; Nguyen, N. T.; Zheng, Z.; Gonzales, A. P.; Li, Z.; Peterson, R. T.; Yeh, J. R.; Aryee, M. J.; Joung, J. K., Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **2015,** *523* (7561), 481-5.

30.   Walton Russell, T.; Christie Kathleen, A.; Whittaker Madelynn, N.; Kleinstiver Benjamin, P., Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **2020,** *368* (6488), 290-296.

31.   Berman, C. M.; Papa, L. J., 3rd; Hendel, S. J.; Moore, C. L.; Suen, P. H.; Weickhardt, A. F.; Doan, N. D.; Kumar, C. M.; Uil, T. G.; Butty, V. L.; Hoeben, R. C.; Shoulders, M. D., An adaptable platform for directed evolution in human cells. *J. Am. Chem. Soc.* **2018,** *140* (51), 18093-18103.

32.   Santos-Moreno, J.; Schaerli, Y., CRISPR-based gene expression control for synthetic gene circuits. *Biochem. Soc. Trans.* **2020,** *48* (5), 1979-1993.

33.   Mali, P.; Aach, J.; Stranges, P. B.; Esvelt, K. M.; Moosburner, M.; Kosuri, S.; Yang, L.; Church, G. M., Cas9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **2013,** *31* (9), 833-8.

34.   Chavez, A.; Tuttle, M.; Pruitt, B. W.; Ewen-Campen, B.; Chari, R.; Ter-Ovanesyan, D.; Haque, S. J.; Cecchi, R. J.; Kowal, E. J. K.; Buchthal, J.; Housden, B. E.; Perrimon, N.; Collins, J. J.; Church, G., Comparison of Cas9 activators in multiple species. *Nat. Methods* **2016,** *13* (7), 563-567.

35.   Konermann, S.; Brigham, M. D.; Trevino, A. E.; Joung, J.; Abudayyeh, O. O.; Barcena, C.; Hsu, P. D.; Habib, N.; Gootenberg, J. S.; Nishimasu, H.; Nureki, O.; Zhang, F., Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **2015,** *517* (7536), 583-8.

36.   Tanenbaum, M. E.; Gilbert, L. A.; Qi, L. S.; Weissman, J. S.; Vale, R. D., A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **2014,** *159* (3), 635-646.

37.   Chavez, A.; Scheiman, J.; Vora, S.; Pruitt, B. W.; Tuttle, M.; E, P. R. I.; Lin, S.; Kiani, S.; Guzman, C. D.; Wiegand, D. J.; Ter-Ovanesyan, D.; Braff, J. L.; Davidsohn, N.; Housden, B. E.; Perrimon, N.; Weiss, R.; Aach, J.; Collins, J. J.; Church, G. M., Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* **2015,** *12* (4), 326-8.

38.   Davidson, A. R.; Lu, W.-T.; Stanley, S. Y.; Wang, J.; Mejdani, M.; Trost, C. N.; Hicks, B. T.; Lee, J.; Sontheimer, E. J., Anti-CRISPRs: protein inhibitors of CRISPR-Cas systems. *Annu. Rev. Biochem.* **2020,** *89* (1), 309-332.

39.   Rauch, B. J.; Silvis, M. R.; Hultquist, J. F.; Waters, C. S.; McGregor, M. J.; Krogan, N. J.; Bondy-Denomy, J., Inhibition of CRISPR-Cas9 with bacteriophage proteins. *Cell* **2017,** *168* (1-2), 150-158.e10.

40.   Dong, D.; Guo, M.; Wang, S.; Zhu, Y.; Wang, S.; Xiong, Z.; Yang, J.; Xu, Z.; Huang, Z., Structural basis of CRISPR-SpyCas9 inhibition by an anti-CRISPR protein. *Nature* **2017,** *546* (7658), 436-439.

41.   Perez-Pinera, P.; Kocak, D. D.; Vockley, C. M.; Adler, A. F.; Kabadi, A. M.; Polstein, L. R.; Thakore, P. I.; Glass, K. A.; Ousterout, D. G.; Leong, K. W.; Guilak, F.; Crawford, G. E.; Reddy, T. E.; Gersbach, C. A., RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods* **2013,** *10* (10), 973-6.

42.   Maji, B.; Gangopadhyay, S. A.; Lee, M.; Shi, M.; Wu, P.; Heler, R.; Mok, B.; Lim, D.; Siriwardena, S. U.; Paul, B.; Dančík, V.; Vetere, A.; Mesleh, M. F.; Marraffini, L. A.; Liu, D. R.; Clemons, P. A.; Wagner, B. K.; Choudhary, A., A high-throughput platform to identify small-molecule inhibitors of CRISPR-Cas9. *Cell* **2019,** *177* (4), 1067-1079.e19.

43.   Jan, M.; Scarfò, I.; Larson, R. C.; Walker, A.; Schmidts, A.; Guirguis, A. A.; Gasser, J. A.; Słabicki, M.; Bouffard, A. A.; Castano, A. P.; Kann, M. C.; Cabral, M. L.; Tepper, A.; Grinshpun, D. E.; Sperling, A. S.; Kyung, T.; Sievers, Q. L.; Birnbaum, M. E.; Maus, M. V.; Ebert, B. L., Reversible ON- and OFF-switch chimeric antigen receptors controlled by lenalidomide. *Sci. Transl. Med.* **2021,** *13* (575), eabb6295.

Chapter 5: A look at the future of directed evolution in mammalian cells

5.1 Overview of targets for mammalian cell-based directed evolution

5.1.1 Introduction. With the development of broadly useful and efficient platforms for mammalian cell-based directed evolution, new targets for directed evolution are now within reach. Below, we highlight examples of targets for which mammalian cell-based directed evolution should prove particularly impactful, categorized into broadly conserved cellular processes, metazoan-specific processes, or mammalian-specific processes (**Figure 5.1**). Within each of these categories, targets include biomolecules that natively perform relevant biological functions and non-native proteins that can be evolved to perturb those same native functions.

5.1.2 Considering targets for directed evolution in mammalian cells. Genes unique to mammals provide perhaps the most obvious examples of targets that would benefit from mammalian cell-based directed evolution. Even when a given conserved process is shared between kingdoms, it is often the case that directed evolution experiments to understand or perturb the mammalian version of that process still need to be performed in mammalian cells. As discussed in Chapter 1, protein variants evolved and optimized in lower organisms often fail to function properly once they are reintroduced into mammalian cells. Further, even for well-conserved biological processes, mammalian protein families often contain numerous paralogs that have highly differentiated functions. Likewise, there are many distinctive mammalian cell types with unique features that can greatly influence protein folding, modification, or activity.[1] Finally, even successful instances of directed evolution experiments involving conserved targets in lower organisms should not necessarily preclude similar experiments being performed in mammalian cells. On the contrary, they facilitate such experiments by providing a roadmap for researchers interested in evolving the same or improved protein activities in a vastly different environment.

5.2 Directed evolution targets conserved across all kingdoms

Although all cells share some common processes, the specific mechanisms by which these processes are carried out are typically not readily transferable between organisms. Some illustrative examples (**Figure 5.1a**) of general cellular processes that present compelling targets for mammalian cell-based directed evolution are discussed below.

a Cellular life			
Transcription	RNA Pol II	Multicomponent complex Numerous diverse mammalian isoforms Many components do not function in yeast
	Repressors/activators	Common tools are non-optimized, exogenous proteins Many mammalian activators/repressors are understudied
Translation	Aminoacyl tRNA synthetases	Require orthogonality to mammalian machinery Usually evolved in the relevant environment
	Ribosome	Mammalian ribosomes do not function in yeast Mammalian ribosomes are diverse and highly heterogeneous

b Multicellular life			
Communication	Signaling pathways	More responses to extracellular signals than single-celled organisms Requires numerous endogenous factors
	Neuronal processes	Tools to study and manipulate neurology are impactful but lacking Requires numerous endogenous factors
Development	Epigenetic modifiers	Current tools are unoptimized endogenous proteins
	Pioneer factors	Vital for differentiating cell lines in vitro Toolbox needs improved small-molecule control or enhanced activities

c Mammals			
Evolution	Historical evolutionary trajectories	Exploring available evolutionary paths Determining factors that impact evolution
Disease	Therapeutics	Biologics evolved specifically for function in mammalian cells

Figure 5.1. Selected examples of high-impact targets for mammalian cell-based directed evolution. Targets for mammalian cell-based directed evolution encompass targets across the phylogenetic tree of life, from (**a**) processes conserved in all kingdoms of cellular life to (**b**) processes specific to multicellular life to (**c**) processes specific to mammals.

5.2.1 Transcription and translation. Tools to regulate and control transcription and translation are crucially important for molecular biology and are active areas of research for the treatment of diseases such as cancer.[2-4] While RNA polymerase II is highly conserved from yeast to mammals,[5-6] many mammalian transcription factors, activators, and repressors do not function appropriately in yeast.[7] Moreover, mammalian cells have many more RNA polymerase II subunits than do yeast—subunits with important and diverse functions.[5-6] Studying these subunits, and the complex transcriptional machinery altogether, requires numerous known and unknown components already present in mammalian cells but absent in lower organisms.

Tools to control and enhance translation are also invaluable, both for fundamental and medical research. For example, genetic code expansion through unnatural amino acid incorporation has profoundly impacted our understanding of translation and spurred the growth of the genetically encoded biological toolbox.[8] Although each organism requires tools that are orthogonal to endogenous machinery, the current standard practice to create engineered amino acyl tRNA synthetases (aaRSs) for use in mammalian cells is to transport functional aaRSs from yeast and hope they retain orthogonality.[9] As a result, the toolkit for unnatural amino acid incorporation in mammalian cells is limited as compared to yeast or bacterial systems, although efforts of Chatterjee and co-workers highlight the potential of new strategies for generation of such tools.[10-11] Likewise, efforts to engineer the ribosome have seen exciting progress, but are almost exclusively focused on ribosomes from lower organisms like *Escherichia coli*.[12] Successfully and reliably evolving new mammalian cell-compatible tools for enhancing both transcription and translation will often require mammalian cell-based directed evolution.

5.2.2 Post-translational engagement. Almost as soon as translation initiates, nascent proteins begin to interact with many cellular components that aid in attaining proper structure and function. These include interactions with the proteostasis network, functional partners, and post-translational protein modification machineries. The differences in these systems between mammals and lower organisms can dramatically impact protein function.

Many high-impact targets for directed evolution require the mammalian proteostasis machinery to assist their folding and maturation.[13] This challenge is particularly acute for complex and large proteins, such as membrane proteins[14] or protein trafficking machineries. Beyond just the clients of mammalian proteostasis networks, the directed evolution of chaperones[15] or chaperone modulators in mammalian cells could lead to a much deeper understanding of the mammalian proteostasis network, as well as to strategies for enhanced production of well-folded mammalian proteins. Mammalian proteins also show a much larger diversity of post-translational modifications than yeast or bacteria, but the tools to genetically encode and control post-

translational modification in mammalian cells are primitive at best.[16] Although directed evolution schemes have been developed in simple systems to evolve kinases,[17] phosphatases,[18] lectins,[19] glycosyltransferases,[20] and proteases,[21] these targets are very rarely evolved in mammalian cells. Developing specialized tools to control and enhance mammalian-specific proteostasis machinery and post-translational modifications, particularly through mammalian cell-based directed evolution, would have enormous fundamental and biomedical research impacts.

5.3 Directed evolution targets from metazoan-specific processes

Metazoans have evolved many biological functions that do not exist in single-celled organisms. Pathways involved in these functions will provide fertile ground for many important mammalian cell-based directed evolution experiments, both for natively functioning proteins as well as for non-native proteins that perturb these processes (**Figure 5.1b**).

5.3.1 Differentiation and development. All metazoan organisms develop from a single cell, the descendants of which create a vast diversity of cell types. The underlying regulatory pathways that drive differentiation and development cannot be easily replicated in single-celled organisms such as yeast. Efforts to study or control differentiation and development, for example by employing epigenetic modifiers[22] and pioneer factors,[23] would benefit greatly from mammalian cell-based directed evolution.

Efforts to control epigenetic modification include orthogonal write-read epigenetic markers[24] and CRISPR-Cas9-based epigenome modifiers,[25] and the application of naturally occurring pioneer factors allows researchers to generate differentiated cell lines from pluripotent stem cells. Directed evolution of modified versions of these tools with improved small molecule control, more or less promiscuous activity, and expanded scope of activity would dramatically improve our ability to study and control cellular development and differentiation.

5.3.2 Communication. Unlike single-celled organisms, metazoans must implement higher-order communication to interpret environmental and endogenous cues. Engineering biomolecules in these pathways can create fundamental insights into the neurological pathways that drive behavior and may even allow precision engineering of organismal behavior. G-protein coupled receptors, involved in the sensing of light, smell, and aspects of taste,[26] are well-suited for mammalian cell-based directed evolution (Chapter 3).[27] The study of ion channels, involved in the sensing of heat, touch, and aspects of taste,[28] has benefited tremendously from engineering efforts but has yet to be significantly impacted by directed evolution. Conveniently, successfully engineering biomolecules in the complex senses will directly impact organism behavior, allowing researchers to easily employ these engineered biomolecules in model mammalian organisms.

Along similar lines, neuroscientists often rely on electrophysiology and imaging techniques to identify biochemical underpinnings of neurological activity. Engineered proteins that allow precise control through small molecules or light can help elucidate many neurological pathways. Indeed, evolved neurological receptors integral to neurological pathways have been important for advances in molecular neuroscience.[29] Further development of these proteins and others associated with neural function will undoubtedly provide even greater levels of control and expand the molecular neuroscientist's toolbox even further.

5.4 Directed evolution targets for mammalian-specific processes and needs

Some processes are important to study simply because we, as mammals, find them to be particularly important (**Figure 5.1c**).

5.4.1 Historical evolution. Directed evolution is an accelerated laboratory mimic of natural evolution. Thus, it can be used to study factors that impact natural evolutionary history. Protein fitness landscapes can be probed by directed evolution, yielding evolutionary insights and detailed information about protein function and stability.[30] Such studies are rarely performed in mammalian cells but, when they are, they often identify important cellular factors that may have impacted historical evolution.[31-32] Studying some of the fastest evolving genes in the human genome may help characterize the traits that evolved during the recent speciation of *Homo sapiens*. By evolving mammalian genes in the mammalian cell setting, but on a laboratory-timescale, we may be able to glimpse into our own history.

5.4.2 Treatments for human diseases. Directed evolution has already transformed the way we treat disease. One class of proteins, antibodies, demonstrates the tremendous therapeutic potential of evolving mammalian proteins.[33] The clever application of directed evolution of antibodies has also spawned the development of chimeric antigen receptors, where immune cell receptors can be engineered to target specific antigens.[34] But opportunities in this space go far beyond just antibodies. Many other mammalian proteins can have therapeutic impacts through other modalities. By using directed evolution, we may even be able to generate new proteins to solve problems for which natural evolution did not provide any answer. A compelling modern example is CRISPR enzymes that may be valuable for disease treatment (Chapter 4). Optimization of gene editing and other key functions directly in mammalian cells may prove critical for successful development of CRISPR-based therapeutics.

5.5 Concluding remarks

The development of biotechnologies purpose-built for application in mammalian cells have produced remarkable advances in mammalian cell-based directed evolution platforms. With numerous established techniques and a wide range of useful targets to evolve, there is ample opportunity for growth in this field. Impactful and potentially transformative targets abound in the mammalian genome and beyond, many of which have never been evolved or engineered before. Looking forward, improvements in library size and library complexity could be achieved by employing large synthetic DNA libraries or less biased random mutagenesis techniques. Furthermore, improved strategies for negative selection, more efficient methods for cellular screening, and strategies for the routine implementation of effective selection couples should expand the scope of activities that can be efficiently and reliably evolved in mammalian cells. I expect that the method described in this thesis, and its descendants, will serve to democratize mammalian cell-based directed evolution, such that researchers may soon routinely use virus-based directed evolution to augment their own specific fields of study without needing to invent or develop the techniques. If so, we will more rapidly enter a new age of mammalian biology—one not only of discovery, but of invention.

5.6 References

1. Ellgaard, L.; Molinari, M.; Helenius, A., Setting the standards: quality control in the secretory pathway. *Science* **1999,** *286* (5446), 1882-8.
2. Warren, L.; Lin, C., mRNA-based genetic reprogramming. *Mol. Ther.* **2019,** *27* (4), 729-734.
3. Trepotec, Z.; Lichtenegger, E.; Plank, C.; Aneja, M. K.; Rudolph, C., Delivery of mRNA therapeutics for the treatment of hepatic diseases. *Mol. Ther.* **2019,** *27* (4), 794-802.
4. Chen, A.; Koehler, A. N., Transcription factor inhibition: lessons learned and emerging targets. *Trends Mol. Med.* **2020,** *26* (5), 508-518.
5. Conaway, R. C.; Conaway, J. W., Origins and activity of the Mediator complex. *Semin. Cell Dev. Biol.* **2011,** *22* (7), 729-34.
6. Conaway, J. W.; Florens, L.; Sato, S.; Tomomori-Sato, C.; Parmely, T. J.; Yao, T.; Swanson, S. K.; Banks, C. A.; Washburn, M. P.; Conaway, R. C., The mammalian Mediator complex. *FEBS Lett.* **2005,** *579* (4), 904-8.
7. Kennedy, B. K., Mammalian transcription factors in yeast: strangers in a familiar land. *Nat. Rev. Mol. Cell Biol.* **2002,** *3* (1), 41-9.
8. Wang, L.; Xie, J.; Schultz, P. G., Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* **2006,** *35*, 225-49.
9. Chin, J. W., Expanding and reprogramming the genetic code of cells and animals. *Annu. Rev. Biochem.* **2014,** *83*, 379-408.
10. Italia, J. S.; Peeler, J. C.; Hillenbrand, C. M.; Latour, C.; Weerapana, E.; Chatterjee, A., Genetically encoded protein sulfation in mammalian cells. *Nat. Chem. Biol.* **2020,** *16* (4), 379-382.
11. Italia, J. S.; Zheng, Y.; Kelemen, R. E.; Erickson, S. B.; Addy, P. S.; Chatterjee, A., Expanding the genetic code of mammalian cells. *Biochem. Soc. Trans.* **2017,** *45* (2), 555-562.
12. Hammerling, M. J.; Fritz, B. R.; Yoesep, D. J.; Kim, D. S.; Carlson, E. D.; Jewett, M. C., *In vitro* ribosome synthesis and evolution through ribosome display. *Nat. Commun.* **2020,** *11* (1), 1108.
13. Sebastian, R. M.; Shoulders, M. D., Chemical biology framework to illuminate proteostasis. *Annu. Rev. Biochem.* **2020,** *89*, 529-555.
14. Marinko, J. T.; Huang, H.; Penn, W. D.; Capra, J. A.; Schlebach, J. P.; Sanders, C. R., Folding and misfolding of human membrane proteins in health and disease: from single molecules to cellular proteostasis. *Chem. Rev.* **2019,** *119* (9), 5537-5606.
15. Sachsenhauser, V.; Bardwell, J. C., Directed evolution to improve protein folding *in vivo*. *Curr. Opin. Struct. Biol.* **2018,** *48*, 117-123.
16. Chuh, K. N.; Batt, A. R.; Pratt, M. R., Chemical methods for encoding and decoding of posttranslational modifications. *Cell Chem. Biol.* **2016,** *23* (1), 86-107.
17. Christians, F. C.; Scapozza, L.; Crameri, A.; Folkers, G.; Stemmer, W. P., Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling. *Nat. Biotechnol.* **1999,** *17* (3), 259-64.
18. Xu, H.-f.; Zhang, X.-e.; Zhang, Z.-p.; Zhang, Y.-m.; Cass, A. E. G., Directed evolution of *E. coli* alkaline phosphatase towards higher catalytic activity. *Biocatal. Biotransfor.* **2009,** *21* (1), 41-47.
19. Hu, D.; Tateno, H.; Kuno, A.; Yabe, R.; Hirabayashi, J., Directed evolution of lectins with sugar-binding specificity for 6-sulfo-galactose. *J. Biol. Chem.* **2012,** *287* (24), 20313-20.
20. Yang, G.; Rich, J. R.; Gilbert, M.; Wakarchuk, W. W.; Feng, Y.; Withers, S. G., Fluorescence activated cell sorting as a general ultra-high-throughput screening method for directed evolution of glycosyltransferases. *J. Am. Chem. Soc.* **2010,** *132* (30), 10570-7.

21.	Gaudelli, N. M.; Komor, A. C.; Rees, H. A.; Packer, M. S.; Badran, A. H.; Bryson, D. I.; Liu, D. R., Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* **2017,** *551* (7681), 464-471.

22.	Atlasi, Y.; Stunnenberg, H. G., The interplay of epigenetic marks during stem cell differentiation and development. *Nat. Rev. Genet.* **2017,** *18* (11), 643-658.

23.	Iwafuchi-Doi, M.; Zaret, K. S., Pioneer transcription factors in cell reprogramming. *Genes Dev.* **2014,** *28* (24), 2679-92.

24.	Park, M.; Patel, N.; Keung, A. J.; Khalil, A. S., Engineering epigenetic regulation using synthetic read-write modules. *Cell* **2019,** *176* (1-2), 227-238.

25.	Pulecio, J.; Verma, N.; Mejia-Ramirez, E.; Huangfu, D.; Raya, A., CRISPR/Cas9-Based engineering of the epigenome. *Cell Stem Cell* **2017,** *21* (4), 431-447.

26.	Julius, D.; Nathans, J., Signaling by sensory receptors. *Cold Spring Harb. Perspect. Biol.* **2012,** *4* (1), a005991.

27.	English, J. G.; Olsen, R. H. J.; Lansu, K.; Patel, M.; White, K.; Cockrell, A. S.; Singh, D.; Strachan, R. T.; Wacker, D.; Roth, B. L., VEGAS as a platform for facile directed evolution in mammalian cells. *Cell* **2019,** *178* (3), 748-761.

28.	Subramanyam, P.; Colecraft, H. M., Ion channel engineering: perspectives and strategies. *J. Mol. Biol.* **2015,** *427* (1), 190-204.

29.	Piatkevich, K. D.; Jung, E. E.; Straub, C.; Linghu, C.; Park, D.; Suk, H. J.; Hochbaum, D. R.; Goodwin, D.; Pnevmatikakis, E.; Pak, N.; Kawashima, T.; Yang, C. T.; Rhoades, J. L.; Shemesh, O.; Asano, S.; Yoon, Y. G.; Freifeld, L.; Saulnier, J. L.; Riegler, C.; Engert, F.; Hughes, T.; Drobizhev, M.; Szabo, B.; Ahrens, M. B.; Flavell, S. W.; Sabatini, B. L.; Boyden, E. S., A robotic multidimensional directed evolution approach applied to fluorescent voltage reporters. *Nat. Chem. Biol.* **2018,** *14* (4), 352-360.

30.	Romero, P. A.; Arnold, F. H., Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **2009,** *10* (12), 866-76.

31.	Geller, R.; Pechmann, S.; Acevedo, A.; Andino, R.; Frydman, J., Hsp90 shapes protein and RNA evolution to balance trade-offs between protein stability and aggregation. *Nat. Commun.* **2018,** *9* (1), 1781.

32.	Phillips, A. M.; Gonzalez, L. O.; Nekongo, E. E.; Ponomarenko, A. I.; McHugh, S. M.; Butty, V. L.; Levine, S. S.; Lin, Y. S.; Mirny, L. A.; Shoulders, M. D., Host proteostasis modulates influenza evolution. *eLife* **2017,** *6*, e28652.

33.	Lu, R. M.; Hwang, Y. C.; Liu, I. J.; Lee, C. C.; Tsai, H. Z.; Li, H. J.; Wu, H. C., Development of therapeutic antibodies for the treatment of diseases. *J. Biomed. Sci.* **2020,** *27* (1), 1.

34.	Guedan, S.; Calderon, H.; Posey, A. D., Jr.; Maus, M. V., Engineering and design of chimeric antigen receptors. *Mol. Ther. Methods Clin. Dev.* **2019,** *12*, 145-156.