universität wien

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Higher-order functional genomics with Cas12a-based platforms

verfasst von | submitted by

lic. Leon Anthony Schmidt

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 830
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Molecular Microbiology, Microbial Ecology and Immunobiology
Betreut von | Supervisor:	Dipl.-Biochem. Dr. Marco Yannic Hein

Higher-order functional genomics with Cas12a-based platforms

Abstract

CRISPR-Cas proteins have become extremely efficient and versatile tools in functional genomics. Many advanced approaches require the targeting of multiple loci in parallel in the same cell. This poses challenges for lentivirus-based CRISPR-Cas9 screening implementations because lentiviral template switching produces chimeric proviruses, which limit the scalability of this method for higher-order multiplexing. A possible solution is to capitalize on the inherently multiplexed nature of endogenous CRISPR loci, where homologous regions are minimal in length.

For this purpose, we tested the utility of multiAsCas12a, a nickase version of a Class II, Type V CRISPR effector. multiAsCas12a is capable of processing its own array of crRNAs expressed from a single RNA Polymerase III promoter while containing mutations in key residues that turn the Cas protein into a DNase-slow version to minimize DNA damage. We found that while the platform needed minimal improvements, the cell line used in the experiment drastically influenced the efficacy of the perturbation at a given locus. Furthermore, we found that the levels of the Cas12 protein can be the limiting factor, especially when combined with an array of multiple crRNAs. This thesis highlights important caveats of the Cas12a gene perturbation platform which, if these issues are addressed, holds the potential to become a highly efficient and straightforward approach for multiplexed and combinatorial gene targeting.

Zusammenfassung

CRISPR-Cas-Proteine haben sich zu äußerst effizienten und vielseitigen Instrumenten der funktionellen Genomik entwickelt. Viele fortgeschrittene Ansätze erfordern die parallele Bearbeitung mehrerer Loci in derselben Zelle. Dies stellt eine Herausforderung für Lentivirus-basierte CRISPR-Cas9-Screening-Implementierungen dar, da das lentivirale Template-Switching chimäre Proviren erzeugt, die die Skalierbarkeit dieser Methode für Multiplexing einschränken. Eine mögliche Lösung besteht darin, die von Natur aus multiplexe Eigenschaft endogener CRISPR-Loci zu nutzen, bei denen homologe Regionen minimiert sind.

Zu diesem Zweck testeten wir multiAsCas12a, eine Nickase-Version eines CRISPR-Effektors der Klasse II, Typ V. multiAsCas12a ist in der Lage, seine eigene Anordnung von crRNAs zu verarbeiten, die von einem einzigen RNA-Polymerase-III-Promotor exprimiert werden, und enthält Mutationen in Schlüsselbereichen, die das Cas-Protein in eine DNase-langsame Version verwandeln, um DNA-Schäden zu minimieren. Wir fanden heraus, dass die Plattform zwar nur minimale Verbesserungen benötigte, die im Experiment verwendete Zelllinie jedoch die Wirksamkeit der Störung drastisch beeinflusste. Darüber hinaus stellten wir fest, dass die Menge des Cas12-Proteins die Effizienz einschränkt, insbesondere in Kombination mit einer Anordnung von mehreren crRNAs. Diese Arbeit hebt wichtige Nachteile der Cas12a-Plattform für Gen-Perturbation hervor, die, wenn diese Probleme behoben werden, das Potenzial hat, ein hocheffizienter und unkomplizierter Ansatz für multiplexes und kombinatorisches Gen-Targeting zu werden.

Acknowledgments

With this, I want to express my sincerest gratitude to the members of the Hein Lab for all their cumulative support throughout my stay. I want to thank Marco for his time and support, for his encouragement and guidance, and for fostering an amazing scientific spirit in the lab, Sebastian for being a patient and attentive supervisor as well as a remarkable scientist, Kavya for all the constructive discussions and Indian snacks, and Miriam for being a sensational gossip-partner and bench mate. I also want to thank Julius and Mahan for making life in the lab just a bit easier, and for sharing my passion for coffee. I want to thank Max, Kerstin, and Patrick for all the informative Journal Clubs and (para)philosophical lunch discussions.

Finally, I want to thank my treasured TF, my dearest family and friends for their endless support, and also myself, for never being *quite* satisfied. This work is dedicated to you.

Introduction

The battle between pathogens and their hosts is a formidable driver of evolution and a rich source of genes and proteins with the potential for biotechnological exploitation. Adaptive immunity, the ability to store a molecular memory of prior pathogen infection for the benefit of counteracting a future encounter, was long thought to be limited to higher organisms. Single-celled organisms were thought to be limited to more primitive, innate immune pathways. However, the discovery of clustered, regularly interspaced, short palindromic repeats (CRISPR) loci in many bacteria, opened up the field of bacterial adaptive immunity.

The CRISPR system was first extensively investigated in the 1990s, when researchers found a repetitive palindromic DNA sequence of 30 base pairs which alternated with a 36 base pair spacer region, while studying *Haloferax mediterranei*. Only years after, the spacer sequences found in the CRISPR loci have been linked to bacteriophage sequences, preluding to the idea of an acquired infection memory and thus a previously undescribed bacterial adaptive immune system [1–3]. Pivotal follow-up studies have described the roles of CRISPR-associated (*cas*) genes such as *cas1*, *cas2,* and *cas7*, and highlighted their importance for spacer acquisition and selection [2,4–6], whilst also identifying Cas9 as the system's primary nuclease, capable of cutting DNA in an RNA-dictated manner [2,4,5,7–9]. Further studies would elaborate on the specific location of CRISPR-induced DNA double-strand breaks (DSB), finding that individual CRISPR RNA (crRNA) sequences would get processed out of the spacer array and dictate the DSB nearby a protospacer-adjacent motif (PAM) [8–10].

CRISPR systems in biotechnology

Soon after, a study in which the authors managed to create a programmable CRISPR system *in vitro* and pioneered a fused version of the tracrRNA and crRNA, called single-guide RNA (sgRNA), paved the way for a new, more efficient way to perform targeted genome editing [11]. This finding would later be awarded with the 2020 Nobel Prize in Chemistry.

Ulterior optimization attempts managed to use CRISPR for precise genome editing in human and mouse cells, while also showing that the system has the potential to target multiple sites simultaneously if multiple crRNAs are provided in an array [12].

Researchers also became heavily invested in the discovery of alternative CRISPR enzymes, with pioneering research on Cpf1 (currently known as Cas12a), which similarly to Cas9 is an RNA-guided DNA endonuclease with a single protein effector, but lacks the need for a tracrRNA sequence as the enzyme can cleave and process

its own pre-crRNAs, [13] and C2c2 (currently known as Cas13a) which is an RNA-guided RNase [13,14]. A considerable advantage of these two CRISPR systems is their inherent ease of use in higher-order combinatorial targeting, which will be of relevance later in this dissertation.

Development of genetic screens

A genetic screen (or mutagenesis screen) refers to an experimental procedure that links genetic perturbations with phenotypic analyses. There are two main strategies for performing genetic screens: (i) forward genetics, which starts from a phenotype of interest and tries to elucidate the associated genetic perturbation, and (ii) reverse genetics, which starts from a known gene and aims to analyze the phenotype after its perturbation.

Historically, because of its rather simplistic setup, scientists first used forward genetic screens, in which a model system was challenged with a mutagen. If a phenotype of interest was found as a result of genetic mutations, the focus would start to shift to identifying the underlying causative gene. In this way, scientists were able to identify genes relevant to embryonic development by using pioneering forward genetics approaches in *Drosophila melanogaster* [15]. Forward genetics, as described by the panelists of the 2018 International Mammalian Genome Conference, is a re-evaluated application of classical Mendelian genetics [16].

Reverse genetics, being on the opposite side of the spectrum, was only possible once programmable genetic perturbations were achieved. By using these methods, geneticists were now able to directly link a specific perturbation of a candidate gene to a specific phenotype. As these programmable genetic mutation platforms were developed contemporaneously with The Human Genome Project, which resulted in the successful sequencing of the entire human genome, reverse genetics found its ideal framework, providing a powerful platform for precise gene manipulation and functional studies [17].

The CRISPR/Cas technology has found its ideal application in this context, all but replacing previous gold standard RNAi-based screening methods [18]; CRISPR/Cas allows for high-throughput studies with a straightforward sgRNA library design, can be upscaled to a wide repertoire of targets and organisms, is multiplexable, highly efficient and cost-effective, while constantly being improved in terms of PAM range, library design, and protein size. Remarkably, it only took approximately a year from the demonstration of human gene editing to the first genome-wide knockout screen based on this platform [19].

From genetics to functional genomics with CRISPR

Driven by the need for comprehensive and high-throughput analyses that go beyond studying individual molecules at a time, but instead study a biological system as a whole, the development of omics methods accelerated and rapidly became relevant in almost every area of biology. Today, omics approaches come in a wide array of flavors, each focusing on the complete set of its eponymous molecules: genomics, transcriptomics, proteomics, metabolomics, epigenomics, lipidomics, and glycomics, among many others.

A more recent overarching concept is functional genomics, which focuses on the dynamic expression of genes and their products in defined circumstances by taking into account the resulting phenotypes, all while keeping a sub-genomic or genome-wide approach. As the concept of functional genomics can be applied to virtually every level of the central dogma, I will specifically focus on transcriptomics moving forward, as it is the core context of my dissertation.

There currently are two main experimental models for CRISPR screening: pooled and arrayed [20]. Both require a cell line of interest which was engineered to express CRISPR-Cas proteins, and a gRNA library targeting genes of interest conveyed in a deliverable vehicle (most commonly using a lentivirus or piggybac [21] background for stable integration or plasmids for transient transfections) to create genetic perturbations in a Poisson-dependent distribution of cells. Depending on the context of the experiment, a challenge (such as drug treatment, pathogen infection, or metabolic challenge) can also be integrated.

A. The pooled format assumes Cas9-expressing cells to be seeded in bulk, in a single tissue culture dish. There, the total gRNA library is transduced (or transfected). After an optional challenge, the readout usually consists of gRNA enrichment or depletion, based on the selective advantage or disadvantage conveyed by the gRNA.
B. The arrayed format assumes Cas9-expressing cells to be seeded in separate wells of a multi-well dish (depending on the number of different gRNAs + conditions). Each well will receive a distinct gRNA. After the challenge, the readout will be more complex, as molecular phenotypes can be interpreted from distinct perturbations while assessing cell nonautonomous phenotypes.

As pooled screens tend to be scalable and more straightforward, while arrayed screens provide the most molecular insights, pooled screens are largely utilized for discovery, while the arrayed format is usually chosen for validation and small-scale further investigation [20]. A promising compromise, combining the scalability of pooled screens with the high-dimensional data typical for arrayed screens, is the use of single-cell CRISPR screens.

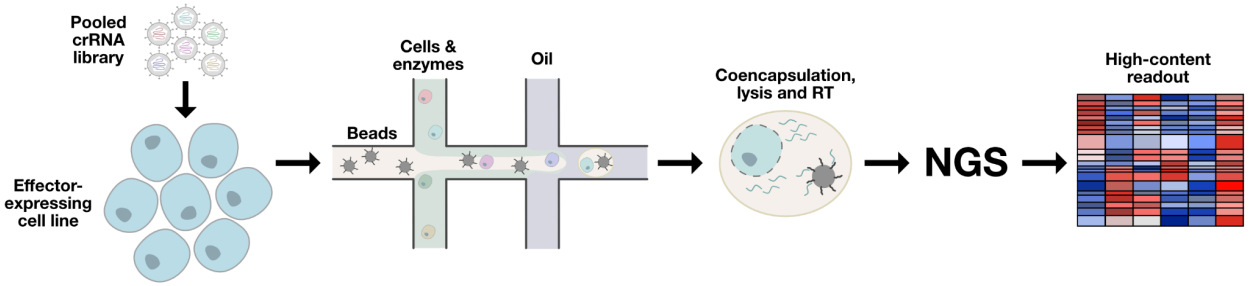


Figure 1. Schematized workflow overview for a pooled CRISPR screen with single-cell transcriptomic readout.

Single-cell CRISPR screening approaches were developed in parallel by multiple labs, with differences in the details of their technical implementation, and hence different acronyms (e.g. CRISP-seq [22], CROP-seq [23], or Perturb-seq [24],[25]. They all rely on the same common concept: combining CRISPR screens with transcriptome analysis at single-cell resolution (scRNA-seq), to determine gene regulation and cell states as a function of the genetic perturbation in each cell.

As shown in Fig. 1, a single-cell CRISPR screen usually employs a pooled format for cell seeding and gRNA delivery, while the readout consists of a total transcriptome analysis (alternatively surface protein abundance [26] or chromatin accessibility [27] for individual cells. To facilitate the physical separation of cells, isolated cells are co-encapsulated in nanodroplets with specialized beads. Depending on the experimental setup, the beads can present different oligonucleotides on their surfaces, such as ones with sequencing adaptors, barcodes, and unique molecular identifiers (UMIs), as well as a poly-T stretch, to which the polyA tails of the mRNA sequences present from each cell will bind to.

A key challenge in single-cell CRISPR screens is linking a cell's entire transcriptomic readout to its specific perturbation, i.e., identifying which gRNA it received. Here, the previously mentioned common single-cell CRISPR screening methods start to differ substantially from each other.

In the original Perturb-seq, the library vector is designed so that each vector also contains a specific guide barcode (GBC) which is placed in the polyadenylated transcript of a selection marker, thus allowing the GBCs to be captured by polyT primer-based reverse transcription [24,25].

The updated direct-capture Perturb-seq improves on this front by engineering a capture sequence into the stem-loop of the sgRNAs with a corresponding reverse transcription primer added to scRNA-seq workflow, enabling sgRNAs to be sequenced directly and not relying on GBCs as proxies [28].

CROP-seq handles the indexing of sgRNAs in a sample differently, by exploiting the variability in the long terminal repeat (LTR) sequences, which are vital for provirus formation. An RNA polymerase III promoter along with the sgRNA sequence is cloned into the 3′LTR. During lentiviral reverse transcription, the expression cassette will be duplicated into the 5′LTR from which the sgRNA will be transcribed. Additionally, the sgRNA will be recorded in RNA-seq because of an RNA polymerase II promoter positioned upstream of the 3′LTR [23]. Despite this design being suitable for CRISPR screens with one single perturbation per cell, multiplexing is severely limited due to the restricted space in the LTRs and the large size of the sgRNA expression cassette.

Complications of lentiviral recombination

Lentiviruses are a genus of retroviruses with a pseudodiploid RNA genome, harboring two viral genome copies per virion. In molecular biology, replication-incompetent HIV-based versions of the virus have been engineered [29] and have become the method of choice for gene transfer systems, by allowing stable integration of externally introduced DNA into the recipient cell's genome in a controlled fashion. Already when the existence of an RNA-dependent DNA Polymerase (i.e. Reverse Transcriptase) was first hypothesized, the considerable variances in lentiviral genomes suggested that the actively polymerizing reverse transcriptase regularly enters a metastable state from which it can either jump, creating mutations or frameshifts, or transfer to the second RNA strand, producing a chimeric provirus once integrated into the host genome [30]. The latter phenomenon is also termed lentiviral template switching, or lentiviral recombination. Template switching can occur between homologous sequences of only 5bp [31,32], however, the larger the homologous region, the likelier it is for template switching to occur, with regions of 1 kb inducing on average one template switching event [33]. An important consequence of this recombination process is that only one viral genome copy can ever be integrated into the host genome.

Current higher-order single-cell CRISPR screening approaches are severely limited by lentiviral template switching. A bottleneck of the original Perturb-seq lies in its indirect indexing of sgRNAs via GBC sequences, which are separated by a ~2.7kb homologous region [34]. The lentivirus library was produced in arrayed format for one of the original Perturb-seq papers to circumvent the issue of template switching [25], however, if it was produced in pooled format at least 50% of the proviruses would have contained sgRNA/barcode mismatches [33]. While direct-capture Perturb-seq removes the necessity of having a GBC by incorporating sgRNA capture sequences, its design is largely limited to dual sgRNA expression cassettes. This is because each sgRNA requires a preceding RNA Polymerase III promoter,

thus increasing homologous regions and recombination rates between the two sgRNAs, as well as reducing lentiviral titers due to larger construct size. The CROP-seq design faces similar challenges [23,34].

All lentivirus-based pooled CRISPR screens are affected by lentiviral recombination events, not only screens with a single cell readout [33,35]. Due to pooled packaging of a sgRNA library into lentiviral particles, co-packaging of different RNA templates, bearing different sgRNA and barcode combinations, is unavoidable.

Thus, an alternative platform that has the potential to reduce occurrences of lentiviral recombination whilst offering multiplexed gene targeting for higher-order combinatorial studies is warranted.

Cas12a as a means of multiplexed functional genomics

Cpf1, later known as Cas12a, was first described in 2015 by the lab of Feng Zhang [13]. They introduced a new RNA-guided nuclease, which similar to Cas9 functions as a single-component effector, but differs from it by not requiring a tracrRNA, classifying it as a Class II Type V CRISPR enzyme (Fig. 2). Two Cpf1 enzymes were identified as suitable for genome editing in human cells, derived from the bacteria *Acidaminococcus sp.* and *Lachnospiraceae bacterium*. The *Acidaminococcus* variant later became the primary form of Cpf1, also known as AsCas12a.

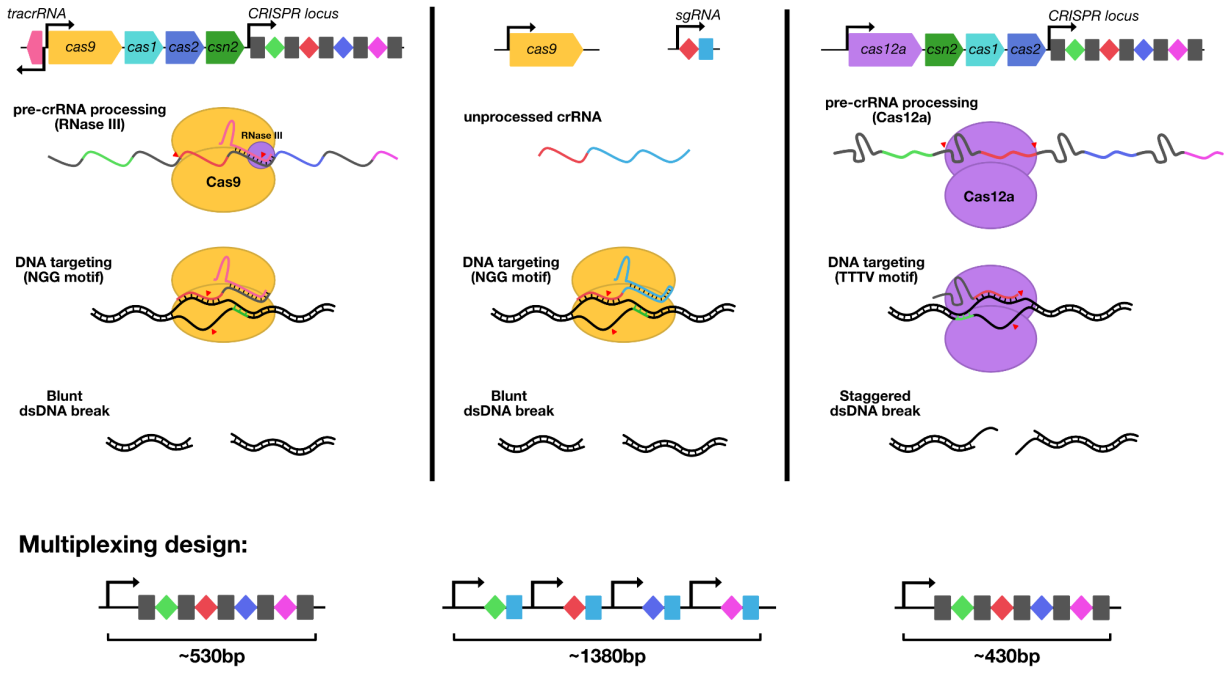


Figure 2. Schematic representation of expression and interference stages in type II-A and V-A CRISPR–Cas systems. Highlighted in the bottom part is a modeled design for targeting four different genes with each system. Adapted from Swarts & Jinek, 2018.

Intriguingly, Cas12a complements Cas9 by recognizing a T-rich PAM (5′-TTTV-3′), in contrast to Cas9's G-rich PAM (5′-NGG-3′) preference, and cleaving DNA in a staggered fashion rather than Cas9's blunt end cuts. Researchers have also highlighted an additional advantage of Cas12a: its ability to autonomously process its own crRNAs, via a distinct RNase-site in the WED-III subdomain [36], from a crRNA array without requiring external RNases like RNase III, which Cas9 relies on [13]. In a subsequent 2017 study, Cas12a's potential for multiplexed gene targeting was demonstrated, as it could independently process an array of crRNAs to target multiple genes simultaneously [37].

Significant advancements in Cas12a established it as a powerful functional genomics platform for combinatorial applications. In 2019, an optimized version of AsCas12a, engineered with multiple mutations to broaden the PAM targeting range (from 5′-TTTV-3′ to 5′-TTTN-3′) and increase efficiency, was introduced as "enhanced *Acidaminococcus* sp. Cas12a" or enAsCas12a [38]. Later, enAsCas12a was applied to genomics, when researchers developed optimizations to further improve its efficacy for combinatorial genetic screens, primarily by incorporating additional nuclear localization signals, defining crRNA design principles, and refining direct repeat (DR) sequences to resemble wild-type activity [36,39]. More recently, enAsCas12a's multiplexing capabilities have been used at a genome-wide scale, providing valuable insights into perturbation combinatorics, such as efficient paralog targeting and crRNA array reliability for up to seven targets, while also noting genotoxicity when multiple essential genes are targeted in tandem [40].

With growing concerns about DSB-induced genotoxicity and off-target effects in nuclease-based genome editing, there is increasing awareness that these methods may face serious limitations in high-precision applications, such as gene therapy or genomic screens [41–45]. These concerns would only intensify in the context of higher-order functional genomics [40]. This has driven the search for alternative techniques that can minimize DNA damage while maintaining multiplexability, allowing for more reliable editing in complex applications.

Multiplexing with Cas12a-based CRISPRi systems

CRISPR interference (CRISPRi) describes a non-catalytic, reversible, gene repression method that originally relied on a sgRNA-guided nuclease-dead Cas9 (dCas9) variant to interfere with RNA Polymerase-Promoter binding, either via proximity blocking or transcriptional repressor domains, such as the Krüppel associated box (KRAB) domain [46,47]. The major advantages of CRISPRi systems compared to their nuclease-active counterparts, are the reversibility of the genetic perturbation and lack of genotoxicity caused by DNA DSBs.

Relying on the concepts of CRISPRi, recent efforts aimed to create a highly efficient and reliable interference platform using Cas12a that could support high-throughput screening in a pooled format [31]. The authors began by evaluating the limitations of previously developed nuclease-dead Cas12a (dCas12a) variants [38], which were highly inefficient, probably due to reduced Cas12a-DNA complex stability. They then reported a more effective nickase variant, termed multiAsCas12a, which incorporates the R1226A mutation to enhance the stability of the ribonucleoprotein-DNA complex, without introducing indels. Their findings demonstrated that this new variant enabled efficient gene perturbations using arrays of up to six crRNAs in pooled screens. Additionally, multiAsCas12a demonstrated improved performance with lower crRNA multiplicity of infection (MOI) and more reliable target knockdown compared to denAsCas12a, two features that are essential for advancing combinatorial functional genomics, particularly in single-cell formats.
While these findings provide valuable insights into the potential of Cas12a CRISPRi platforms for pooled CRISPR screens, there is currently no proof of concept yet, supporting its suitability for single-cell CRISPR screening applications.

Aims of the dissertation

The goal of my Master's thesis is to contribute to the understanding and advancement of Cas12a systems in the scope of functional genomics. By introducing new optimization attempts for the multiAsCas12a system and testing previously undescribed positional effects of direct repeats within crRNA arrays, we aim to leverage and compatibilize the Cas12a system for higher-order combinatorial single-cell CRISPR screens.
For this, we first aimed to improve the efficiency of multiAsCas12a by swapping the original Krüppel-associated box (KRAB) domain from a Kox1 motif to a Zim3 motif. This decision was based on findings from an extensive comparison of KRAB domains optimized for dCas9-mediated CRISPRi, which identified Zim3 as providing the strongest on-target knockdown with minimal off-target effects [48].
Secondly, we investigated the impact of direct repeat (DR) positioning within a multiplexed crRNA array, a concept hypothesized but previously unexplored [31]. Recent advances in optimizing the Cas12a platform have highlighted the critical influence of crRNA position on targeting efficiency within arrays [31,37,40]. Still, clear insights into the positional effects of the DRs and the effect of neighboring crRNA sequences have yet to be provided.
Cas12a-based functional genomics offers a significant advantage for conducting highly multiplexed and combinatorial perturbations. By reducing component size and minimizing lentiviral recombination risks, Cas12a (including engineered variants

like multiAsCas12a) paves its way to become an optimal choice for complex applications, such as single-cell CRISPR screens. Its ability to lower the cost of single-cell CRISPR screens while boosting throughput further underscores its suitability for such experimental designs.

In this thesis, we contribute meaningful insights applicable across current Cas12a platforms. We tested various optimization strategies, identified rate-limiting steps, and evaluated how cell line selection impacts Cas12a-based perturbation experiments. In our opinion, Cas12a will be integral to future high-throughput screens that demand scalability, compact design, combinatorial flexibility, and precision.

Results

First, we tried to improve the efficiency of the original multiAsCas12a fusion, which consisted of 6xNLS-multiAsCas12a-XTEN80-Kox1 (Fig. 3). To implement the proposed optimizations, we designed two new plasmids, each replacing the C-terminal Kox1 KRAB domain with an N-terminal Zim3 domain. These designs differed in the placement of the 6xNLS sequence: one version (also referred to as “pLS12”) placed the NLS at the C-terminus, additionally serving as a linker, while the other (also referred to as “pLS13”) retained the NLS at the N-terminus before the Zim3 KRAB domain. Due to unresolved technical challenges, we were unable to establish cell lines using the second design, so all subsequent experiments are based on the first construct (Fig. 3).

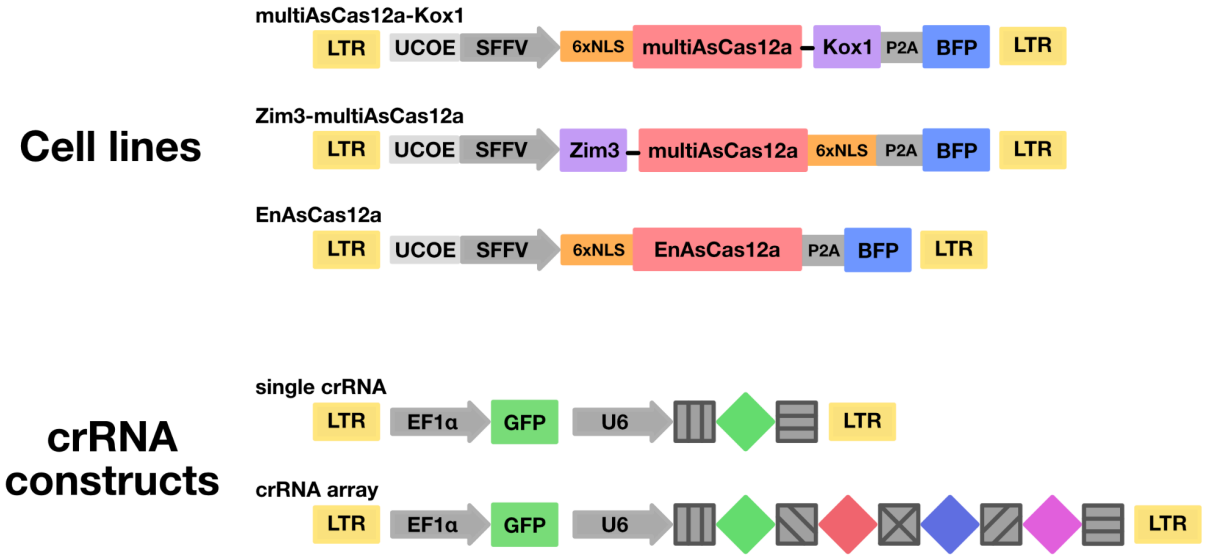


Figure 3. Schematic representation of the different Cas12a-engineered cell lines and crRNA constructs.

Secondly, we aimed to elucidate the potential positional effects of DR sequences within arrays. Furthermore, perturbation efficiency appears to depend not only on crRNA position but also on the sequence of the neighboring crRNAs [31], suggesting that additional, uncharacterized factors in Cas12a pre-crRNA processing may influence outcomes. To contribute to the understanding of these phenomena, we designed seven distinct 4-plex arrays, comprehensively swapping the order of crRNAs and DRs to assess the effects on targeting efficiency (Fig. 4).

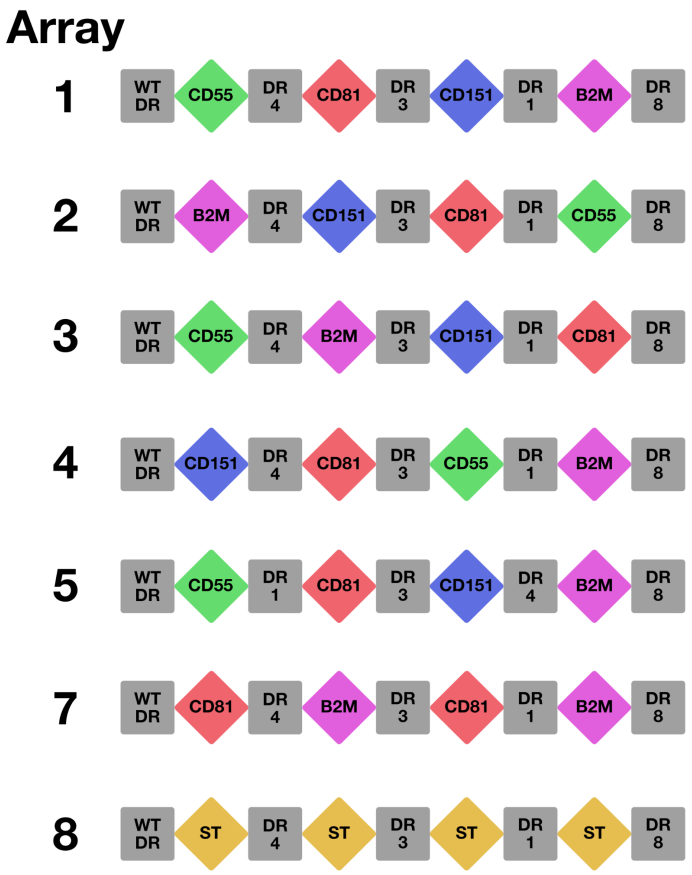


Figure 4. Design of the seven tested crRNA arrays, with emphasis on guide targets and DR order.

To compare Cas12a-based CRISPRi and CRISPRn systems more extensively, we also created a cell line expressing the nuclease-active enAsCas12a (Fig. 3). The effector-expressing plasmids were compatible with lentiviral delivery and designed to also include P2A-BFP connected to the Cas effector. We first transfected them into the HEK293T cell line (ATCC CRL-3216), to produce lentivirus. This lentivirus was then used to transduce A549 cells (ATCC CCL-185). We monitored the expression of Cas effectors by tracking BFP levels as a proxy. Additionally, crRNA lentiviruses were generated, designed to target cell surface markers CD55, CD81, CD151, and B2M, which included a separate expression cassette for GFP to serve as a selection marker and also as an indicator of crRNA expression. Effector-expressing cells were then transduced with the crRNA lentiviruses at a low MOI, resulting in around 20% GFP-positive cells. To validate the efficiency of the perturbations, we used APC-conjugated antibodies specific to the target epitopes and conducted flow cytometry analysis. A schematic of the workflow and data analysis is presented in Fig 5.

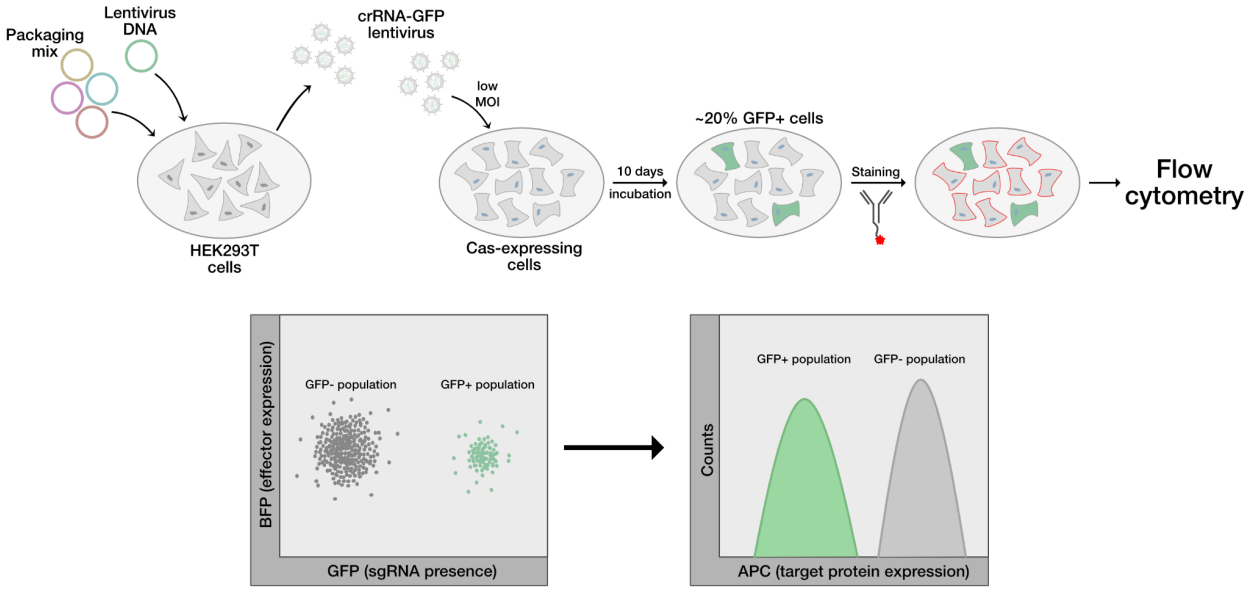


Figure 5. Schematic representation of the experimental workflow, from crRNA-containing lentivirus generation and transduction into effector-expressing cell lines to target-specific antibody staining and successive flow cytometry analysis principles.

Single crRNA system validation

To validate the functionality of the different systems, we first pursued quantifying the perturbation efficiency of the candidate Cas12a effectors using a single crRNA sequence for each of the target cell surface proteins (Fig. 6). Knockdown efficiencies can, in principle, be monitored by RT-qPCR, whole cell transcriptomics, or flow cytometry. We chose targets that localize to the plasma membrane and can be detected via fluorescently labeled antibodies, allowing us to monitor protein levels in a straightforward fashion and with single-cell resolution by flow cytometry.

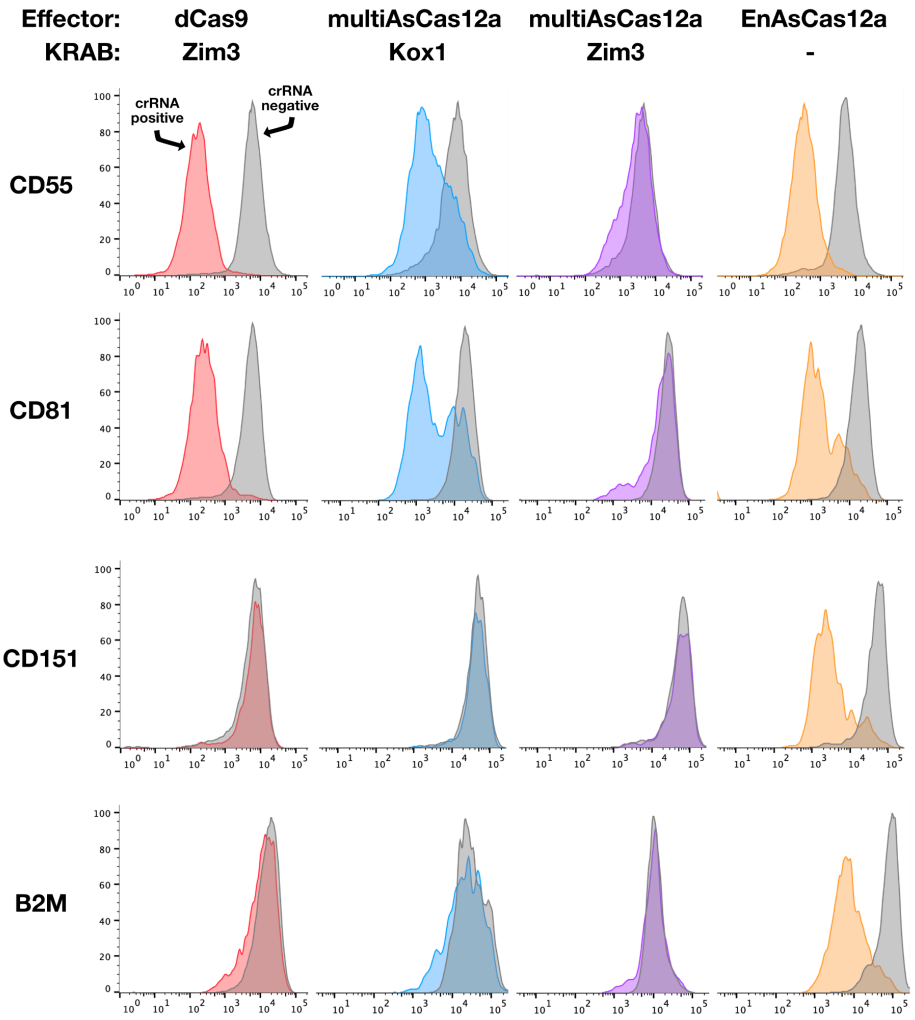


Figure 6. Comparative analysis of perturbation efficiencies via target staining. Colors indicate the crRNA transduced population in an effector-specific manner (Red: Zim3-dCas9, Blue: multiAsCas12a-Kox1, Purple: Zim3-multiAsCas12a, Orange: EnAsCas12a). Grey indicates the sample-specific untransduced population.

The histograms reveal that, when compared to the previously reported multiAsCas12a-Kox1 as a baseline for efficiency (2nd column, Fig. A), our Zim3-multiAsCas12a variant (3rd column, Fig. A) performed noticeably worse across all tested targets. Although we have no exact explanation as to why the Zim3 KRAB domain outperforms the Kox1 domain in the dCas9 system apparently completely inhibits the CRISPR interference activity of multiAsCas12a, we have gotten ulterior structural insights of the engineered fusion proteins using AlphaFold [49,50] and noticed a prominent change in the structure of the 6xNLS when comparing multiAsCas12a-Kox1 with Zim3-multiAsCas12a. Despite being highly debatable, a potential loss of function of the NLS might explain the severe drop in efficiency, considering that the protein effector has to be localized to the nucleus to

function properly [51]. Adding to this, it has been previously reported that the efficiency of Cas12a effectors is directly proportional to the number of NLS sequences fused to it [52].
Surprisingly, multiAsCas12a also showed limited ability to perturb CD151 and B2M expression, which contrasts with previously reported efficiencies in K562 cells [31]. To investigate this discrepancy, we also evaluated the perturbation efficiency of Zim3-dCas9, a system previously validated in K562 cells for the same targets [48]. Unexpectedly, Zim3-dCas9 displayed a similar efficiency pattern to multiAsCas12a-Kox1. This led us to hypothesize that the cell line may play a significant role, as both systems have previously shown strong knockdown capabilities in K562 cells at the same loci, using the same crRNAs.
To further explore this, we analyzed ATAC-seq data for both K562 and A549 cells to compare chromatin accessibility landscapes across the four targets (supplemental data) [33,53]. While we observed a similarly opened chromatin landscape for the CD55 promoter region in both cell lines, which is in accord with our results (Fig. 6), the CD81 promoter region seemed considerably less accessible in A549 than in K562 cells, which would suggest a lower perturbation efficiency using CRISPRi systems. As previously shown (Fig. 6), this didn't seem to be the case, as CD81 was one of the two efficient knockdown targets. On a similar note, ATAC-seq data for CD151 shows less chromatin accessibility in A549 and is consistent with our results, while B2M, which only shows a slightly reduced accessibility in the promoter region, is a highly inefficient perturbation target in our data (Fig. 6).
From this, we infer that no single factor can fully account for the variation in target perturbation efficiency of CRISPRi systems across different cell lines. Instead, it is likely a complex, multifaceted issue that remains largely unexplored.

To extract further insights from the histograms, we created a visualization scheme that allows us to precisely quantify perturbation efficiencies. This approach first measures the distance between the modes of the GFP+ and GFP- populations, reflecting the reduction in cell surface protein levels, plotted on the x-axis as perturbation efficiency. We then quantify the variance between the populations, highlighting the heterogeneity within the crRNA-transduced population relative to the control group, and plotting the result on the y-axis. This dual-metric approach allows us to evaluate the relationship between perturbation strength and population heterogeneity, making it easier to identify crRNAs that achieve both high knockdown efficiency and low variability and vice versa. Figure 7 illustrates how we extrapolate the perturbation efficiency and populational width ratio from the previously shown FACS data, while Figure 8 provides a detailed guide on interpreting different population distributions in the scatter plot.

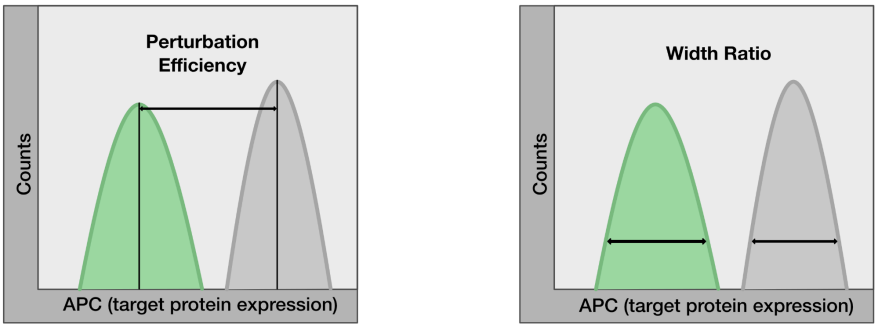


Figure 7. Schematized overview of the dual efficiency quantification system. “Perturbation Efficiency” quantifies the absolute distance between the modes of the crRNA+ and crRNA- populations (expressed in %)(left). “Width Ratio” creates a ratio between the interdecile ranges (IDR) of the crRNA+ and crRNA- populations (right).

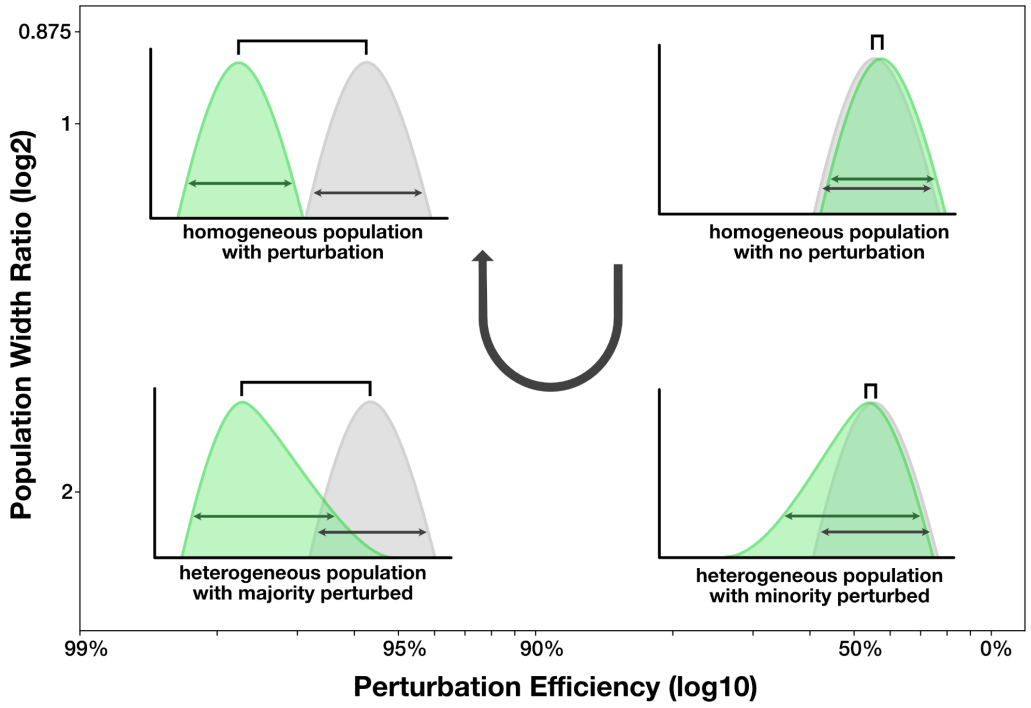


Figure 8. Interpretation principles for the two previously mentioned perturbation efficiency metrics in a scatterplot, with histograms showing where different populational distributions would place comparatively within the two-dimensional space.

By applying these metrics, we systematically quantified the perturbation efficiencies observed in the histograms, providing a clear and comparable view of each crRNA’s performance in the following scatter plot (Fig. 9).

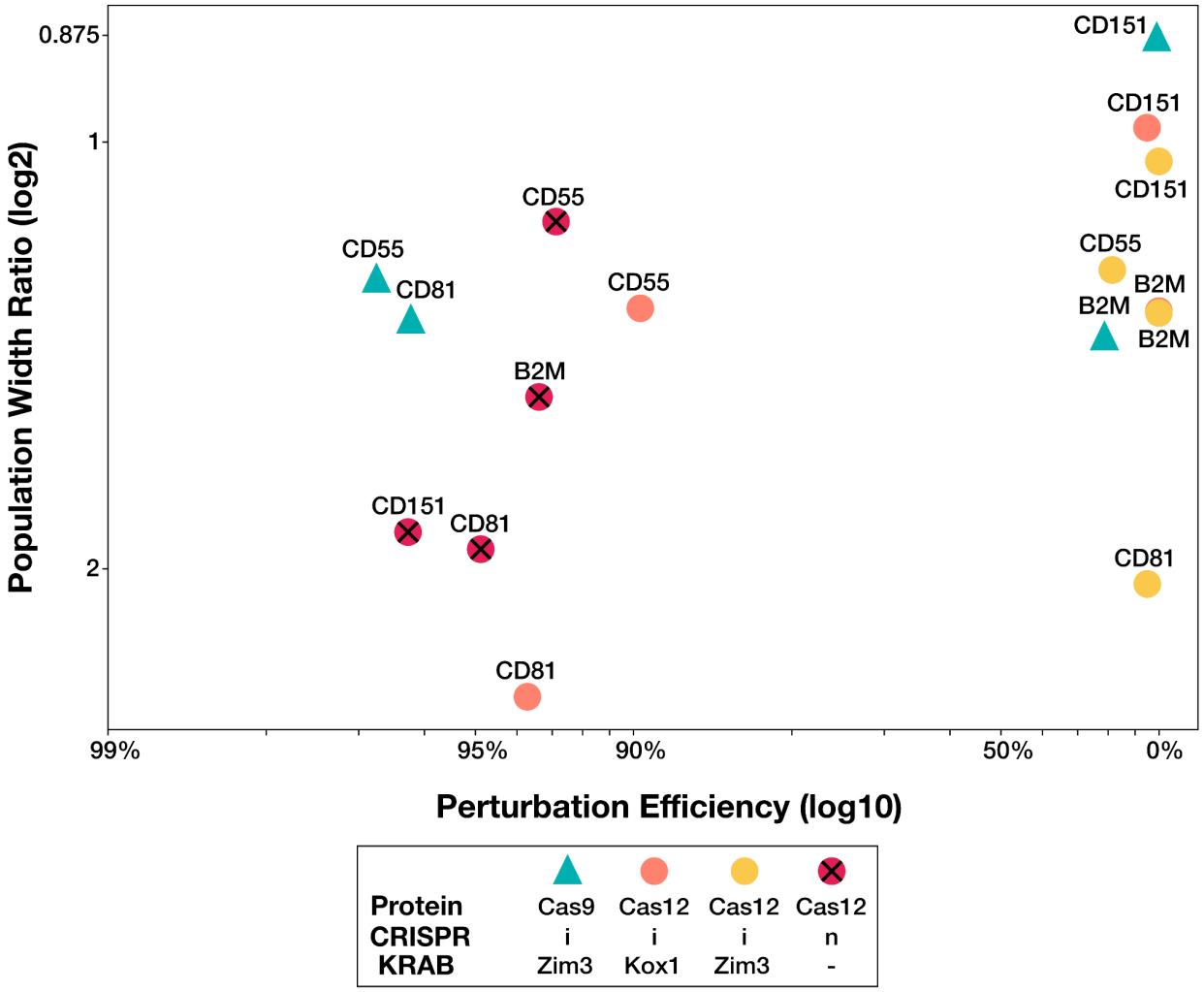


Figure 9. Scatter plot results for the single guide system validation experiment.

Fig. 9 clearly shows that the Zim3-multiAsCas12a version performs worse than any of the other systems tested here, with a maximal perturbation efficiency of ~20%. As previously mentioned, both the Zim3-dCas9 and multiAsCas12a-Kox1 showed inconsistent efficiencies, where both systems were only effective for the targets CD55 and CD81, where they achieved ~90% knockdown efficiencies. Comparing these two systems further, Zim3-dCas9 shows less populational variance than multiAsCas12a-Kox1, especially when targeting CD81, making it overall a slightly more desirable CRISPRi effector. As a result of this, we initiated the design for the following experimental arrays based on the nuclease active EnAsCas12a effector, as this was the only one showing robust results of >90% editing efficiencies across the board.

Multiplexing quantification via crRNA arrays

With the EnAsCas12a confirmed to reliably knockout all proposed cell surface targets when separately targeted, we proceeded to design crRNA 4-plex arrays, each containing a crRNA sequence for each target. As previous studies suggest that both the crRNA sequence position within the array and the position of the direct repeat (DR) sequences may influence perturbation efficiency, we developed a comprehensive set of seven distinct 4-plex arrays to explore these factors (Fig. 4). Array 1 represents a baseline arrangement, while Array 2 reverses the crRNA order. Arrays 3 and 4 test specific positional effects by swapping crRNAs between positions 2 and 4, and positions 1 and 3, respectively. Array 5 maintains the same crRNA sequence order as Array 1 but reverses the DR sequence arrangement to isolate potential impacts from DR positioning. Array 7 introduces a dual-target approach, assigning two crRNAs each to two cell surface markers, which lets us evaluate both the impact of doubling guides for one target and the change in available effector for each target. Array 8 serves as a safe-targeting (ST) control with crRNAs that target intergenic regions, providing both an off-target effect baseline and genotoxicity assessment. Perturbation efficiency across these configurations is illustrated in the accompanying histograms (Fig. 10).

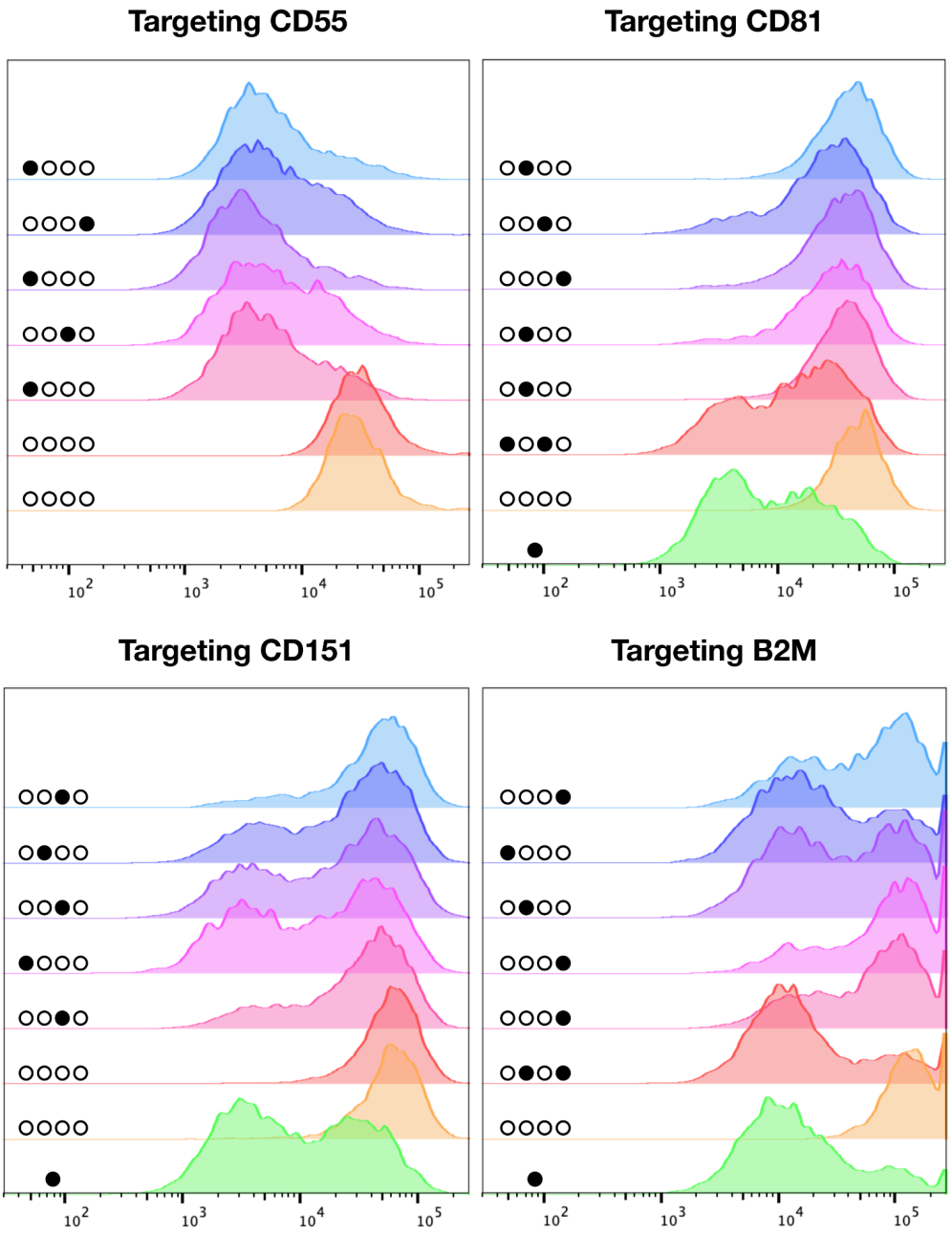


Figure 10. Comparative analysis of target perturbation efficiencies between the panel of designed crRNA arrays. Black dots represent the crRNA position for a specific target within the 4plex array.

When comparing Fig. 10 and Fig. 6, CD55 consistently appears as an easily perturbable target, regardless of crRNA positioning within the array.
In the cases of CD81 and CD151, however, a distinct decline in perturbation efficiency is observed in the array format, contrasting with their effective targeting seen in the single crRNA setup.

For CD81, a single crRNA in any array position proved ineffective; only the dual-crRNA configuration in array 7 produced a bimodal distribution, which still overlapped significantly with the control population and therefore was not considered efficient.

The targeting of CD151 revealed varied outcomes, suggesting a clear positional effect for the selected crRNA. Positions one and two in the 4-plex array facilitated more effective knockout. Interestingly, array 3, which places the CD151 crRNA in the third position, has shown a considerably higher efficiency compared to arrays 1 and 5, which also positioned it third. This pattern aligns with findings from prior studies [31] where the activity of crRNAs was influenced by the swapping of neighboring crRNAs in the array. A satisfactory mechanistic explanation for this is still missing.

A similar positional dependency was noted for B2M, with the fourth position consistently performing worse than positions one or two. Among the three arrays that place B2M in the final position (arrays 1, 4, and 5), array 1 displayed superior knockout efficacy. Interestingly, the only difference between arrays 1 and 5 lies in the position of the DRs, as the crRNA order is identical. Although the impact of DR positioning is subtle, this observation could represent the first clear indication of a DR-based positional effect within crRNA arrays. Additionally, the dual-crRNA design for B2M (guides in positions 2 and 4) achieved strong perturbation, surpassing even the single crRNA experiment in efficacy.

Following the single crRNA method, we also systematically quantified both perturbation efficiency and intrapopulation variance for each crRNA array configuration, visualized in a scatter plot. To keep the chart clear and focused, results are displayed sequentially, focusing on one target at a time (Fig. 11).

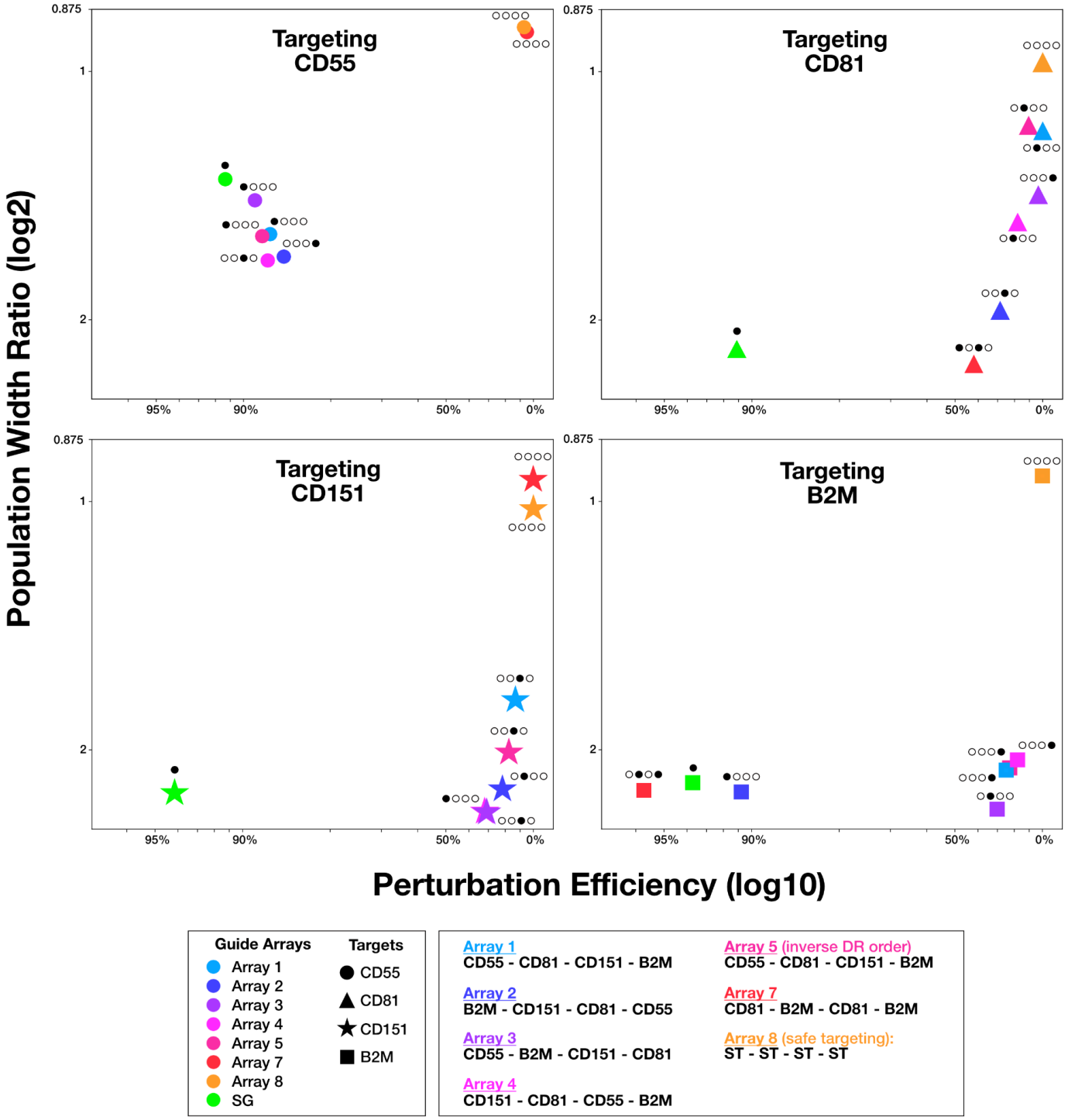


Figure 11. Comparative scatter plots quantifying the target-specific Perturbation Efficiency and Population Width Ratio across the tested array designs.

We can notice that CD55 is successfully targeted across nearly all positions within the 4-plex array, with perturbation efficiencies ranging from 86% to 89% for all array designs. Interestingly, the population width ratio increases as crRNAs move toward

later positions in the array, although overall population homogeneity remains largely consistent.

In contrast, CD81 is poorly targeted across all array formats, particularly when compared to its performance in the single crRNA setup. The dual-crRNA format (array 7) increases targeting efficiency from array 2's 28% to 42%, but this improvement comes at the expense of heightened population heterogeneity.

Targeting CD151 displays a similar trend, where only the single crRNA configuration consistently produces robust perturbation, while the arrays' perturbation efficiency range between 17% and 32%. However, this is accompanied by broad population distribution, leading to high heterogeneity even in the single crRNA format.

Targeting B2M, however, shows a distinct result: the dual-crRNA array outperforms the single-crRNA construct with a 96% perturbation efficiency, and only a slight increase in heterogeneity. Notably, a clear positional effect is observed, with crRNAs in the first position of a 4-plex array showing significantly better performance (91% perturbation efficiency) than those in other positions (ranging between 17% and 30%). However, in every array targeting B2M, the GFP+ cell distribution is nearly twice as wide as the GFP- population, indicating considerable heterogeneity.

Protein availability is limiting in both single guide and array formats

Intrigued by the difference in target perturbation efficiency and population variability between single crRNA setups and crRNA arrays, we aimed to identify potential causes. Hsiung et al. reported in their multiAsCas12a study that their system exhibited reduced sensitivity to low MOI crRNA-lentivirus transductions compared to prior Cas12a-based CRISPRi systems [31], making it attractive for single-cell CRISPR screens. To determine if similar behavior was present in our data, we examined GFP+ cells in various samples and found notable heterogeneity in perturbation efficiency linked to relative BFP expression. We observed a similar trend when analyzing cells by quantification of the indirect crRNA proxy, GFP expression. To further explore this, we divided the GFP+ guide-expressing cells into four subgroups: "BFP high," "BFP low," "GFP high," and "GFP low." This quadrant-based division of the GFP+ population is schematized below (Fig. 12).

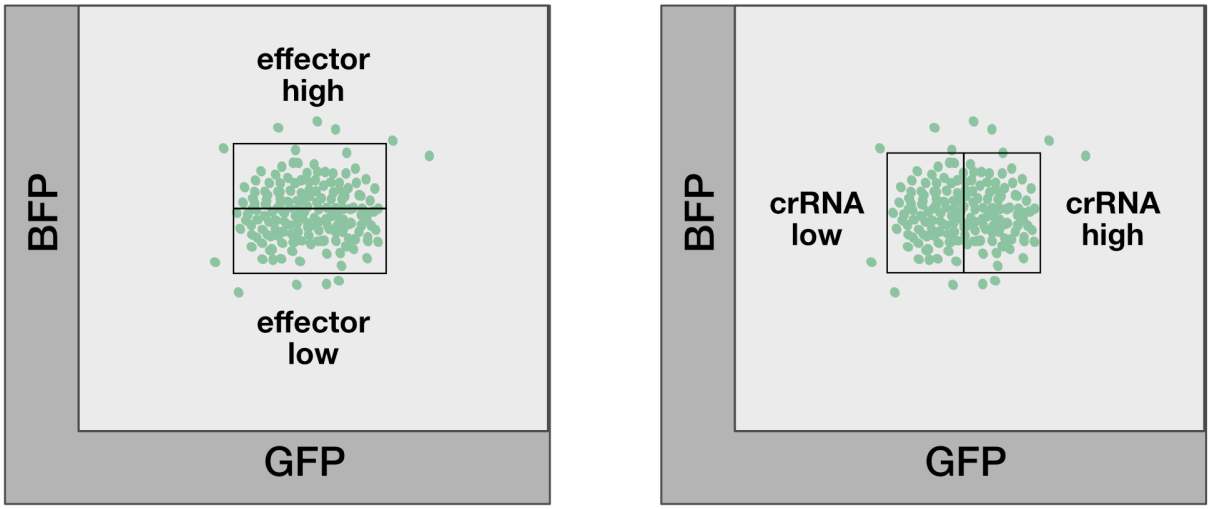


Figure 12. Schematization of the quadrant-based subpopulational analysis, with effector-based division (left) and crRNA-based division (right).

When we applied this analysis to each population within the array experiment that exhibited unexplained heterogeneity or a reduction in perturbation efficiency relative to the single-crRNA setup, we found clear evidence pointing to the limitations imposed by both crRNA and Cas-effector expression levels (Fig. 13).

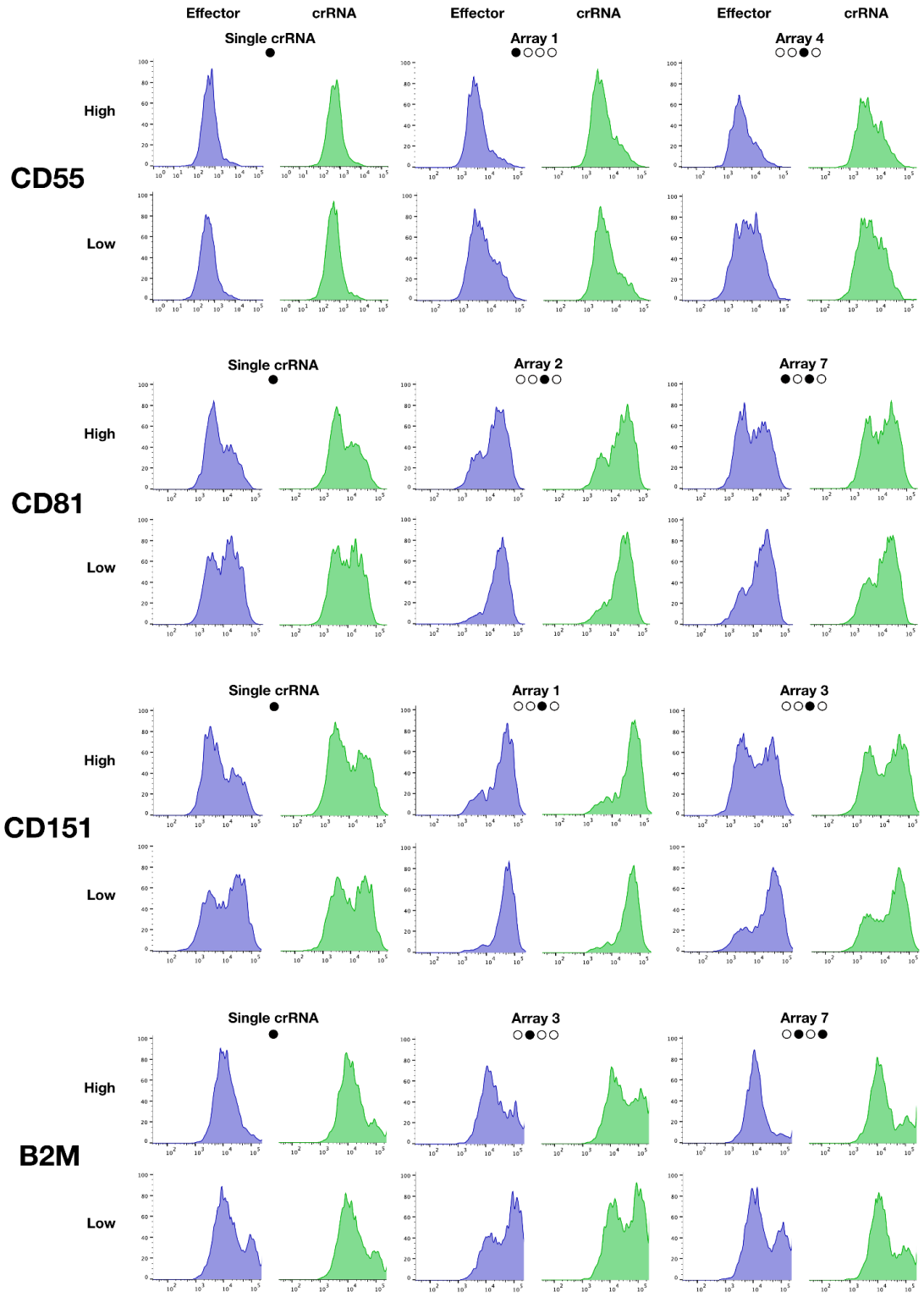


Figure 13. Comparative analysis of selected effector & crRNA high & low subpopulations.

Consistent with previous results where CD55 was a readily targetable gene, all the subpopulations of the crRNA-transduced population showed a high perturbation efficiency for CD55. However, this pattern shifted in array 4, which in terms of CD55

target knockout was the least efficient crRNA array. Here both GFP low and BFP low subpopulations seemed to form a multimodal distribution, whereas only the BFP high subpopulation maintained a clear unimodal histogram with strong perturbation efficiency. This observation suggests that, in this particular array configuration, both the CD55-targeting crRNA and overall Cas-effector expression appear as limiting factors, with Cas-effector levels being especially restrictive.

For CD81, quadrant subpopulation data for the single crRNA experiment, array 2 and array 7 are presented. In the single guide setup, each paired high and low fluorescent-protein (FP) subpopulation shows distinct distribution patterns, pointing to a limiting effect from both the Cas protein and crRNA levels. A similar trend appears in array 2, though with a key difference: both FP low subpopulations display an unimodal distribution overlapping with the crRNA-negative control, suggesting that target knockout was ineffective. In contrast, the FP high subpopulations exhibit a slight bimodal pattern, indicating partial success in perturbing CD81. In array 7, which uses a dual-crRNA strategy to target CD81, a subtle shift emerges. The BFP high subpopulation outperforms the GFP high subpopulation, with a more pronounced difference in distribution when comparing BFP high vs. low to GFP high vs. low. This suggests that for CD81 targeting, the dual-crRNA setup in array 7 shows increased dependency on Cas-effector levels rather than on crRNA abundance.

For CD151, the results of the single crRNA approach to those of array 3 are compared. In the single-guide configuration, the BFP high subpopulation exhibits a clear unimodal distribution, contrasting with the GFP high group, while the BFP low cells show reduced efficiency relative to the GFP low cells. This pattern implies limitations in both effector and crRNA levels, with the Cas-effector appearing more restrictive. In array 3, a noticeable decline in perturbation efficiency is observed across all subpopulations, even in the most effective BFP+ group, underscoring the impact of array design on target disruption efficiency.

The quadrant subpopulation analysis for B2M offered particularly insightful findings, leading us to evaluate the single-guide setup, the dual-guide array 7, and array 3. In the single-crRNA configuration, a marked difference emerged between the BFP high and BFP low subpopulations, while the GFP high and GFP low groups showed only subtle variations. This strongly suggests that, for effective B2M targeting, the availability of Cas-effector protein is the main limiting factor, even in a single-crRNA setting. A similar trend is seen in array 7, where B2M is targeted by two distinct crRNAs, underscoring the role of effector abundance in achieving efficient knockdown. In array 3, however, a shift in the pattern appears: the BFP low subpopulation shows significant limitations in knockdown efficiency, which is not as pronounced in the BFP high cells. When comparing GFP high to GFP low, a parallel trend emerges. Cross-referencing BFP high with GFP high and BFP low with GFP

low further reinforces the conclusion that high effector expression is a key, rate-limiting factor in achieving consistent and effective perturbation.

Taking these findings together, it seems plausible that with our experimental setup, the protein effector expression level doesn't reach the saturation level to efficiently knockout the candidate cell surface proteins, even with a single crRNA. Using a 4-plex crRNA configuration increases the likelihood of the Cas effector as the rate-limiting factor, highlighting the limited availability of Cas protein as the bottleneck. While this limitation does not seem to impact highly responsive targets or those with optimally designed crRNAs - such as CD55 in our experiments - three out of the four tested targets are affected by constraints in Cas-protein levels or a combination of protein and crRNA levels.

Discussion

In summary, we were able to recapitulate most of the previous findings which underscored Cas12a-based perturbation systems' promise in the context of functional genomics, while also providing new molecular insights relevant to potential system improvements. Our results indicate that the Zim3 KRAB domain, at least in the domain arrangement that is effective for dCas9, appears to completely inhibit CRISPRi activity in multiAsCas12a. However, the inability to perturb even the most accessible cell surface markers within our experimental setup suggests that factors beyond the Zim3 KRAB domain may contribute to this outcome. Additionally, we confirmed the previously suggested crRNA direct repeat (DR) arrangement bias, though the observed changes in perturbation efficiency were minimal [31]. Lastly, we report that both crRNA availability and, even more critically, effector protein levels are key limiting factors in Cas12a-mediated perturbations. While previous studies have highlighted inherent sgRNA potency, sgRNA and effector expression levels, and R-loop formation as rate-limiting steps in Cas9-mediated target perturbation [54,55], similar confounding factors - such as R-loop initiation and crRNA binding affinity - have been reported in Cas12a systems [56,57]. Our findings, however, emphasize a previously underexplored aspect of Cas12a functionality that, to our knowledge, has not been explicitly addressed in the literature and is crucial for complex applications, such as single-cell functional genomics.

Since the initial study of the Class II Type V Cas-like endonuclease "Cpf1" in 2015, significant advancements have been made in developing Cas12a-based perturbation systems. Nevertheless, several mechanisms critical to optimizing gene-editing efficiency with Cas12a remain unexplored. Early studies using crRNA arrays with Cas12a observed differences in target perturbation efficiencies, which they attributed to the relative position of each crRNA within an array, a phenomenon now known as the positional effect [31,37]. Subsequent research has consistently highlighted the existence of this effect, while newer studies are now proposing that DR sequences may introduce their own positional biases. Furthermore, currently unexplainable examples where neighboring guide sequences influence the efficacy of a crRNA have also been reported. We and others [31] think that such phenomena are closely linked to Cas12a's pre-crRNA array processing mechanism, which is still an active area of research [57].

There also are unanswered questions regarding the CRISPRi-based multiAsCas12a system. First, half of our chosen targets (CD55 and CD81) underperformed significantly compared to what was previously reported. Our subsequent tests with

dCas9 produced similar results, suggesting that the bottleneck may not be due to the effector itself but rather to factors specific to the cell line used. Cell line-specific factors that could impact CRISPRi perturbation efficiency include chromatin landscapes, epigenetic regulation, and target promoter activity. This emphasizes the need for thorough validation of crRNA activity and compatibility in a specific system and cell line, before attempting to answer a biological question.

Secondly, our attempt to fuse the Zim3 KRAB domain to the N-terminus of multiAsCas12a appeared ineffective. While the Zim3 sequence has demonstrated strong interference activity when fused to dCas9, this was not observed in our specific construct. When choosing the linker sequence and KRAB domain location for the Zim3-multiAsCas12a fusion protein, we were inspired by a previously validated effector, which placed the Zim3 domain at the N-terminus of dCas9 [48]. Future trials, such as positioning the NLS sequence upstream of the Zim3 domain or relocating the Zim3 KRAB domain to the C-terminus of multiAsCas12a, may yield improved results.

In both our CRISPRi and CRISPRn systems, the levels of Cas protein and crRNA appear to be limiting, significantly affecting perturbation efficiency. To circumvent this, several improvements could be beneficial. For effector protein delivery, we used a medium-high transduction MOI of around 2 (50-70% BFP+ cells). In contrast, previous studies have achieved higher Cas protein titers using a lentiviral MOI of 5 or high-titer transient protein expression via plasmid transfection. These approaches, as well as the use of a PiggyBac delivery system, could reduce protein-related bottlenecks by increasing effective Cas protein levels. Additional strategies include testing various high-expression promoter sequences, employing additional gene regulatory elements other than the UCOE to prevent transgene silencing, or knocking in the effector protein downstream of ubiquitously expressed endogenous genes.

Another approach to minimize protein bottlenecks is engineering the effector proteins at key residues to enhance their stability, both in unbound and crRNA-bound forms. Increased stability could, in turn, enhance the stability of expressed crRNAs, potentially reducing bottlenecks for both crRNA and Cas protein levels simultaneously. To determine whether gene silencing or protein stability poses a greater challenge, a comparison of transcript and protein abundances would be useful: unusually low transcript levels could indicate gene silencing, while abnormally low protein levels would point to protein stability issues.

As mentioned throughout this manuscript, the main goal of this dissertation was to explore the possible use case for the multiAsCas12a platform in single-cell CRISPR screen settings. In our view, two primary approaches exist for integrating multiplexed perturbations with single-cell transcriptome readout: direct-capture Perturb-seq, which identifies perturbations through specially designed capture

sequences, and the original CROP-seq, which identifies perturbations via RNA polymerase II transcripts.
Building on the example of CarPool-seq [58], which combines Cas13d with Perturb-seq, it seems feasible to design a barcoded gRNA (bcgRNA) for placement at the last position in a crRNA, paired with PCR and reverse transcription handles. This would allow the crRNA array to be identified via specific RT primers, and link them to individual cell transcriptomes. Though effective with Cas13d, this strategy has yet to be validated with Cas12a.
A CROP-seq approach may also be viable since Cas12a, unlike Cas9, doesn't require repetitive RNA polymerase III promoters. Its compact system could potentially accommodate a 2-plex crRNA array instead of a single gRNA with Cas9, or up to a 9-plex crRNA array instead of two sgRNAs with Cas9 [31]. However, higher-order combinatorics might be challenging to achieve, as a recent preprint [59] using Cas9 with the CROP-seq vector for multiplexed gene targeting reported a 10-fold decrease in effective lentiviral titer after adding approximately 100 bp to the LTR region compared to the original CROP-seq design.

We think combinatorial screening will revolutionize the future scientific landscape, as single genetic perturbations will not be sufficient to study most complex genetic systems. Cas12a's multiplexing capabilities could be especially valuable in contexts involving redundant pathways, combinatorial genetic interaction mapping, or studies of synthetic lethality. Additionally, the ability to use crRNA arrays also enables group testing, allowing multiple grouped targets to be screened in a single array to reduce costs. Previous studies have leveraged multiplexed screening approaches to identify gene interaction partners in myeloid leukemia differentiation [58], to perform higher-order combinatorial perturbations of cis-regulatory elements [31], to identify novel synthetic lethality in apoptotic genes [39] and leukemia epigenetic regulators [52].

Based on our findings and those of prior studies, we see Cas12a technology as a highly promising tool for multiplexed functional genomics. With further optimizations, it could become a powerful choice for perturbation-based CRISPR screens with single-cell transcriptomic readout.

Materials and methods

Plasmid design and construction

The original protein-expressing and crRNA-expressing vectors (pCH4/45 and pCH67, respectively) were a kind gift from CC Hsiung and Luke Gilbert. Cloning of protein-expressing vectors was performed via Gibson Assembly reactions using NEBuilder HiFi DNA Assembly Master Mix (NEB Cat# E2621). Cloning of crRNA-expressing vectors was performed via Golden Gate Assemblies. Final plasmid sequences were checked via Sanger sequencing and/or nanopore sequencing services provided by Microsynth and Plasmidsaurus, respectively. All oligonucleotides and gene fragments were ordered from Integrated DNA Technologies.

Protein construct expression vectors

The multiAsCas12a-expressing plasmid (UCOE-SFFV-6x(Myc)NLS-multiAsCas12a--XTEN80-KOX1-P2A-mTagBFP2) (from now on referred to as pCH45) was used without additional modifications. For the Gibson Assembly reactions, the pCH45 plasmid served as the PCR template for the modified versions pLS12 (UCOE-Zim3-multiAsCas12a--6x(Myc)NLS-P2A-mTagBFP2) and pLS13 (UCOE-6x(Myc)NLS-Zim3-multiAsCas12a--P2A-mTagBFP2). MluI and BamHI restriction enzymes were used to insert the newly generated constructs into the pCH45 backbone. The sequence for Zim3 was PCR amplified from pNM1128, a kind gift from the Jonathan Weissman lab.
The EnAsCas12a-expressing plasmid (UCOE-SFFV-6x(Myc)NLS-EnAsCas12a-P2A--mTagBFP2) (from now on referred to as pCH4) initially harbored the D908A mutation to inactivate the enzyme. To obtain the active nuclease version of EnAsCas12a, this mutation was reversed (A908D) via site-directed mutagenesis using the NEBuilder HiFi DNA Assembly Master Mix (NEB Cat# E2621), with the overlapping primers being designed to contain the A908D mutation.

crRNA expression vectors

All crRNA constructs were cloned into the human U6 promoter-driven expression vector pCH67 (EF1a-Flag-EGFP-P2A-Puro-WPRE-U6-WT_DR-cargo-DR_8). For individual single crRNA expression vectors, oligonucleotides encoding for the specific target gene and the backbone-specific overhangs were designed, annealed, and cloned into the vector via a one-pot Golden Gate Reaction using BsmBI/Esp3I restriction enzyme. For multiplex crRNA arrays, double-stranded DNA

gene blocks were ordered, which contained four different crRNA sequences, four different WT-like direct repeat sequences, BsmBI/Esp3I recognition sites, and PCR-handles. 7 different multiplex crRNA arrays were designed. Their description is as follows, in 5' to 3' direction:
Array 1: WT_DR, crRNA1, DR4, crRNA2, DR3, crRNA3, DR1, crRNA4, DR8
Array 2: WT_DR, crRNA4, DR4, crRNA3, DR3, crRNA2, DR1, crRNA1, DR8
Array 3: WT_DR, crRNA1, DR4, crRNA4, DR3, crRNA3, DR1, crRNA2, DR8
Array 4: WT_DR, crRNA3, DR4, crRNA2, DR3, crRNA1, DR1, crRNA4, DR8
Array 5: WT_DR, crRNA1, DR1, crRNA2, DR3, crRNA3, DR4, crRNA4, DR8
Array 6: WT_DR, crRNA2, DR4, crRNA4, DR3, crRNA2.2, DR1, crRNA4.2, DR8
Array 7: WT_DR, ST1, DR4, ST2, DR3, ST3, DR1, ST4, DR8
crRNA 1 is targeting CD55, crRNAs 2 & 2.2 are targeting CD81, crRNA 3 is targeting CD151, crRNAs 4 & 4.2 are targeting B2M. ST 1 & 2 & 3 & 4 are non-targeting crRNAs. The targets were chosen based on high-quality knock-down results for multiAsCas12a [47].
Sequences for crRNAs 4/4.1, ST 1, ST 2, ST 3, and ST 4 were taken from [38]. Sequences for crRNAs 1, 2/2.1 were taken from [49]. Sequences for crRNAs 2.2, 3, and 4.2 were the top results from the CRISPick algorithm [38].

crRNA sequences:
crRNA1: TCTTCTGTAACCTGGACGGCACT
crRNA2: AGTGCCTGCTGGGGACGGTAAGG
crRNA2.2: TATGACCAGGCCCTACAGCAGGC
crRNA3: GACGCTCCTTGAAGGTGGCGCAG
crRNA4: CCGATATTCCTCAGGTACTCCAA
crRNA4.2: CTGAATTGCTATGTGTCTGGGTT
ST1: GCCCGGACTAAATCCGACTGAGA
ST2: CCGACAAGAGACGGTCCTAATGT
ST3: GCTTACGATCGTTGGAGGGCAGC
ST4: GAACGCTATATACAGCGGTAACA

Direct repeat sequences:
WT_DR: TAATTTCTACTCTTGTAGAT
DR4: TAATTTCTACTATTGTAGAT
DR3: AAATTTCTACTCTAGTAGAT

DR1: TAATTTCTACTGTCGTAGAT
DR8: TAATTTCTCCTCTAGGAGAT

Molecular cloning

Depending on compatibility, if not otherwise specified, all molecular cloning steps were done via Gibson Assembly or Golden Gate Assembly, for the construction of protein-expressing plasmids or insertion of crRNA sequences into cargo plasmids, respectively.
For Gibson Assembly reactions, suitable sites for unique cutting enzymes were identified. An enzyme-digested version (MluI + BamHI; NEB cat. nr. R3198 and R3136, respectively) of pCH45 was chosen as the backbone for the Gibson Assembly of the modified variants pLS12 and pLS13. The sequences for multiAsCas12a and 6x(Myc)NLS were PCR-amplified from the original pCH45 plasmid. As already mentioned, the sequence for the Zim3 KRAB domain was PCR-amplified from pNM1127. Overlapping primers were designed automatically by SnapGene, with at least 20bp overlap. In line with standardized protocols, we aimed for a total fragment concentration of ~0.3pmol, each inserts being in 2-fold molar excess relative to the 50ng of backbone used. The reaction was incubated at 50°C for 1h. An aliquot of the reaction was transformed in Stellar chemically-competent *E. coli* (Takara Bio cat. nr. 636766).
To insert crRNA sequences into the pCH67 backbone, we resorted to Golden Gate assembly reactions. For this, a one-pot reaction model was chosen, where T4 DNA ligase (400U/mL; NEB cat. nr. M0202) was mixed with the type II restriction enzyme BsmBI_v2 (or its isoschizomer Esp3I), together with their respective buffers, additionally to ATP. To this, 50ng of backbone and the diluted crRNA-encoding double-stranded oligonucleotide were added. The oligonucleotides were ordered separately (coding sequence and its reverse complement) and were annealed in-house; the final reaction had a concentration of 0.1mM annealed oligonucleotide. Depending on the efficiency, thermocycler profiles were chosen between 12-60 cycles of 55°C (or 37°C with Esp3I) for 5 minutes followed by 16°C for 5 minutes and ended with a 45-minute incubation at 55°C (or 37°C). An aliquot of the reaction was transformed in Stellar chemically-competent *E. coli* (Takara Bio cat. nr. 636766).

Cell culture, plasmid transfection, lentiviral transduction

All cell lines were cultured at 37°C with 5% CO2 in cell culture incubators. HEK293T and A549 cells were cultured in media containing a mixture of DMEM, high glucose, pyruvate (Gibco cat. nr. 419660290; with 4.5g/L glucose, 580mg/mL glutamine, 110mg/mL sodium pyruvate added), supplemented with 10% FBS (Sigma cat. nr. F7524), and 1% 100X PSQ (10000 units/mL penicillin, 10000mg/mL streptomycin,

29.2mg/mL glutamine) (This mixture will from now on be referred to as cell culture media). Unless otherwise specified, cells were routinely passaged by incubation with 0.25% Trypsin-EDTA (Gibco cat. nr. 25200072) at 37°C with 5% CO2, for approximately 5 minutes, followed by quenching with cell culture media.
Lentiviral particles were produced by transfecting HEK293T which haven't exceeded 90% confluency for several passages. 24h before transfection, 9e5 cells are seeded into a well of a 6-well plate. On the transduction day, per well of a 6-well plate, 250mL Opti-MEM (Gibco cat. nr. 31985062) and 6mL TransIT-293 transfection reagent (Mirus cat. nr. MIR2700) are mixed and incubated for 15 minutes at room temperature. In the meanwhile, 1.5mg lentivirus DNA is combined with a mixture of packaging plasmids (100ng gag/pol, 100ng REV, 100ng TAT, 200ng VSVG). After incubation, the plasmids are mixed with the media-lipid emulsion and incubated for another 15 minutes at room temperature. The complexes are then added drop-wise to the wells and mixed by gentle back-and-forth movements of the plate. 6h post-transfection, the media is changed and replaced with warmed cell culture media. 24h and 48h after the media change, the lentivirus-rich supernatant was harvested, pooled, centrifuged and filtered using 0.45mm filters, aliquoted and snap-frozen in liquid nitrogen.
For stable protein-expressing cell line creation, A549 cells were transduced with 50% lentiviral supernatant in cell culture media supplemented with 5mg/mL Polybrene (Hexadimethrine bromide). To increase the transduction efficiency, a spinfection step was added (1h, 1000g, 37°C). Based on BFP(BV421) expression, which serves as a proxy for successful Cas12a-protein integration and expression, cells which were at least one log-fold brighter than WT A549 were sorted and reseeded. BFP expression was routinely measured for two weeks while passaging the cells, until the signal stayed within one tenth of a log. The delivery of guide lentiviruses to the stable protein-expressing cell line was done similarly, instead using 2-5% lentiviral supernatant and without a spinfection step.

Antibody staining and flow cytometry

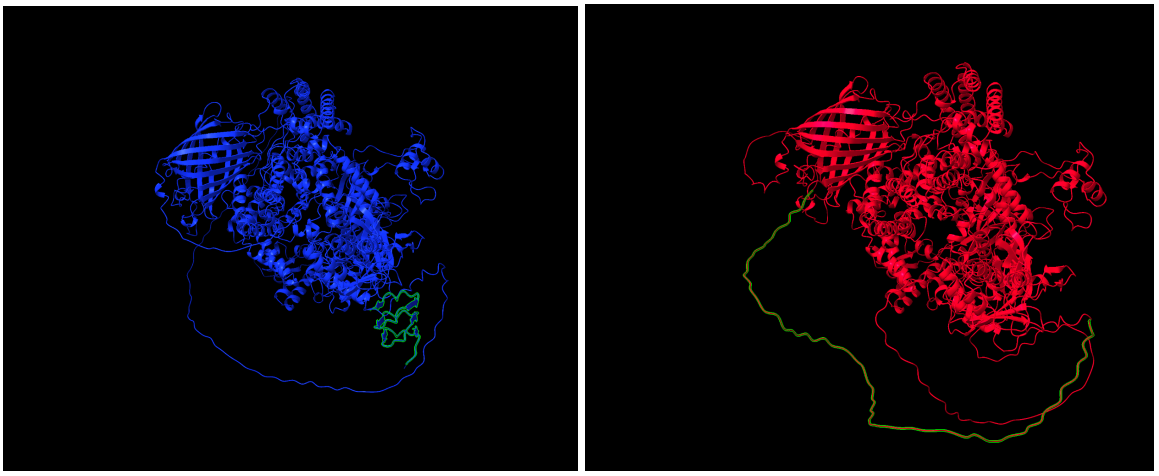
Cells were detached by incubating for 15 minutes at 37°C and 5% CO2, with Accutase (Invitrogen cat. nr. 00455556). After counting, 100.000 cells were seeded into a low-attachment 96-well plate (Corning Costar cat. nr. CLS3474) and washed with 150mL FACS buffer (PBS + 1% BSA) (Gibco cat. nr. 14190; Sigma-aldrich cat. nr. A2153), followed by centrifugation at 500g, 4°C, for 5 minutes and decantation. Cells were resuspended in 50mL antibody dilution (1:100 dilution in FACS buffer, according to the manufacturer's protocol; antibodies are target (CD55,CD81,CD151

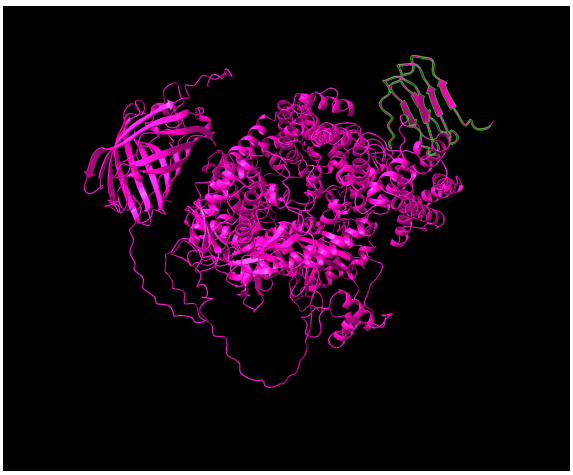
or B2M)-specific and fluorophore-conjugated) and incubated in dark for 30 minutes at 4°C and 600rpm. 150mL FACS buffer was added, followed by centrifugation and decantation. In preparation for the flow cytometry readout, the cells were again resuspended in 200mL FACS buffer.
Antibodies used: CD55-APC (Biolegend cat. nr. 311311), CD81-APC (Biolegend cat. nr. 349509), CD151-APC (Biolegend cat. nr. 350405), HLA-APC (Biolegend cat. nr. 311409).
For all flow cytometry experiments, the results were first gated for alive cells based on FSC-A:SSC-A, then for single cells based on SSC-A:SSC-H. Finally, the population that presented a clear shift in terms of FITC-A/EGFP-A signal was considered to represent cells where crRNA integration was successful. APC-A levels of GFP+ cells were compared to GFP- cells, where lower levels of APC-A served as a proxy for successful target perturbation. Flow cytometry was performed on the BD Fortessa and Biorad ZE5; cell sorting was performed on the BD Aria.

Data analysis

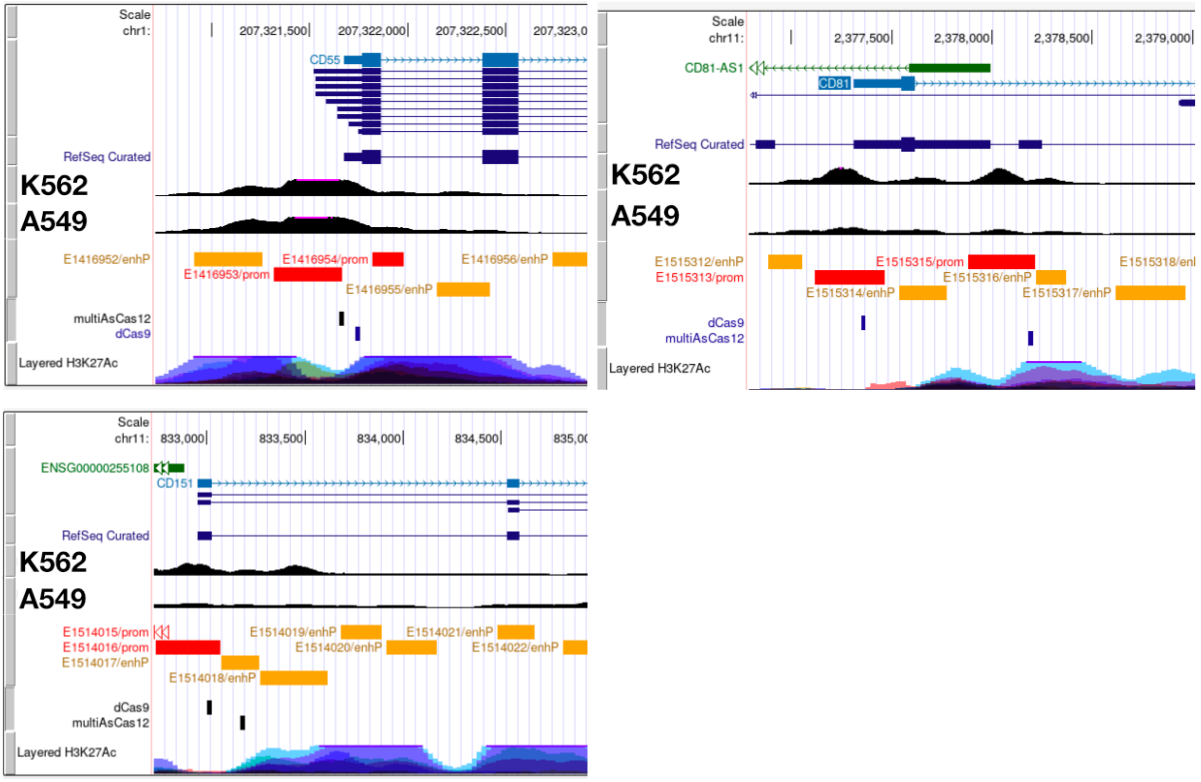
The majority of the initial flow cytometry data analysis was done using FlowJo™ v10.10 Software (BD Life Sciences). Perturbation Efficiency and Populational Width values and the respective scatter plots were generated using JupyterLab 4.0.11 (python 3.0). The figures were designed with Affinity Designer 2.

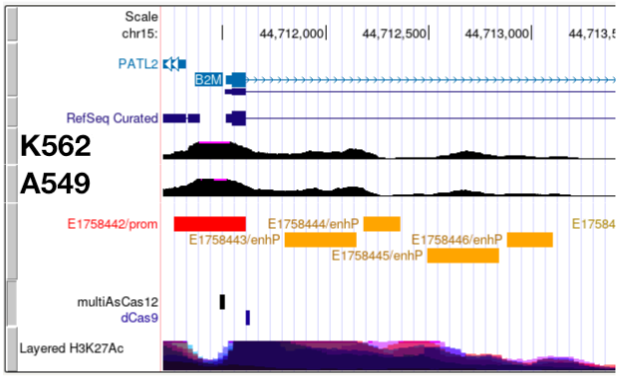
Supplementary data



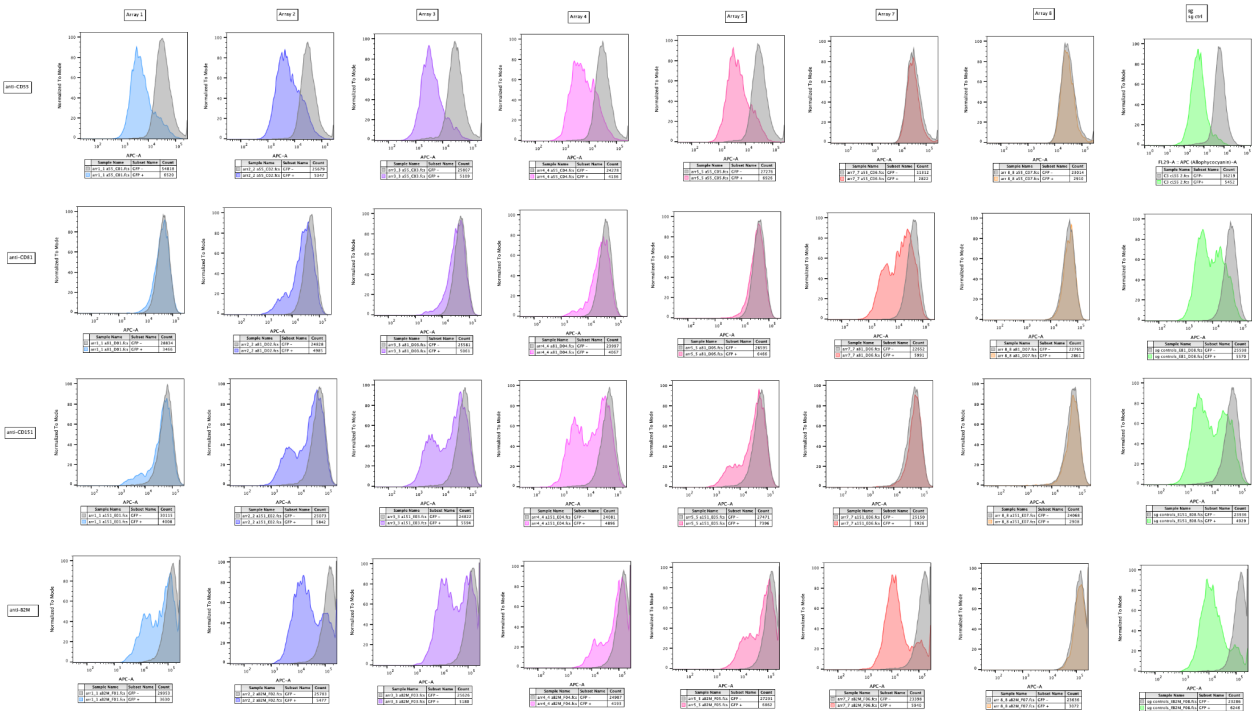


Supplementary Figure 1. Predicted protein structures of the original NLS-multiAsCas12a-Kox1 (Blue), Zim3-multiAsCas12a-NLS (Red) and NLS-Zim3-multiAsCas12a (Pink). Highlighted in green is the 6xNLS sequence, which for the Zim3-multiAsCas12a-NLS shows distinctive folding.





Supplementary Figure 2. ATAC-seq profiles for CD55 (top left), CD81 (top right), CD151 (bottom left), B2M (bottom right) in K562 and A549 cells. The crRNA-targeting sites in multiAsCas12a and Cas9 are also shown.



Supplementary Figure 3. Individual histograms with respective untransduced control populations. Columns indicate array designs (arrays 1,2,3,4,5,7,8 + single crRNA control), and rows indicate different targets (CD55, CD81, CD151, B2M).

References

1. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol. 2005;60: 174–182.

2. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology. 2005;151: 2551–2561.

3. Pourcel C, Salvignol G, Vergnaud G. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology. 2005;151: 653–663.

4. Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJH, Snijders APL, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science. 2008;321: 960–964.

5. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol Direct. 2006;1: 7.

6. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science. 2007;315: 1709–1712.

7. Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science. 2008;322: 1843–1845.

8. Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, et al. Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. J Bacteriol. 2008;190: 1390–1400.

9. Garneau JE, Dupuis M-È, Villion M, Romero DA, Barrangou R, Boyaval P, et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature. 2010;468: 67–71.

10. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature. 2011;471: 602–607.

11. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012;337: 816–821.

12. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013;339: 819–823.

13. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell. 2015;163: 759–771.

14. Abudayyeh OO, Gootenberg JS, Konermann S, Joung J, Slaymaker IM, Cox DBT, et al. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science. 2016;353: aaf5573.

15. Nüsslein-Volhard C, Wieschaus E. Mutations affecting segment number and polarity in Drosophila. Nature. 1980;287: 795–801.

16. Ali Khan A, Raess M, de Angelis MH. Moving forward with forward genetics: A summary of the INFRAFRONTIER Forward Genetics Panel Discussion. F1000Res. 2021;10: 456.

17. Hardy S, Legagneux V, Audic Y, Paillard L. Reverse genetics in eukaryotes. Biol Cell. 2010;102: 561–580.

18. Morgens DW, Deans RM, Li A, Bassik MC. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. Nat Biotechnol. 2016;34: 634–636.

19. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelson T, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science. 2014;343: 84–87.

20. Bock C, Datlinger P, Chardon F, Coelho MA, Dong MB, Lawson KA, et al. High-content CRISPR screening. Nat Rev Methods Primers. 2022;2. doi:10.1038/s43586-021-00093-4

21. Wu X, Scott DA, Kriz AJ, Chiu AC, Hsu PD, Dadon DB, et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nat Biotechnol. 2014;32: 670–676.

22. Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, David E, et al. Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. Cell. 2016;167: 1883–1896.e15.

23. Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, Klughammer J, et al. Pooled CRISPR screening with single-cell transcriptome readout. Nat Methods. 2017;14: 297–301.

24. Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, et al. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell. 2016;167: 1853–1866.e17.

25. Adamson B, Norman TM, Jost M, Cho MY, Nuñez JK, Chen Y, et al. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. Cell. 2016;167: 1867–1882.e21.

26. Mimitou EP, Cheng A, Montalbano A, Hao S, Stoeckius M, Legut M, et al. Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. Nat Methods. 2019;16: 409–412.

27. Pierce SE, Granja JM, Greenleaf WJ. High-throughput single-cell chromatin accessibility CRISPR screens enable unbiased identification of regulatory networks in cancer. Nat Commun. 2021;12: 2969.

28. Replogle JM, Norman TM, Xu A, Hussmann JA, Chen J, Cogan JZ, et al. Combinatorial

single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. Nat Biotechnol. 2020;38: 954–961.

29. Naldini L, Blömer U, Gallay P, Ory D, Mulligan R, Gage FH, et al. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science. 1996;272: 263–267.

30. Temin HM. Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. Proc Natl Acad Sci U S A. 1993;90: 6900–6903.

31. Hsiung CC-S, Wilson CM, Sambold NA, Dai R, Chen Q, Teyssier N, et al. Engineered CRISPR-Cas12a for higher-order combinatorial chromatin perturbations. Nat Biotechnol. 2024. doi:10.1038/s41587-024-02224-0

32. Pfeiffer JK, Telesnitsky A. Effects of limiting homology at the site of intermolecular recombinogenic template switching during Moloney murine leukemia virus replication. J Virol. 2001;75: 11263–11274.

33. Hill AJ, McFaline-Figueroa JL, Starita LM, Gasperini MJ, Matreyek KA, Packer J, et al. On the design of CRISPR-based single-cell molecular screens. Nat Methods. 2018;15: 271–274.

34. Adamson B, Norman TM, Jost M, Weissman JS. Approaches to maximize sgRNA-barcode coupling in Perturb-seq screens. bioRxiv; 2018. doi:10.1101/298349

35. Xie S, Cooley A, Armendarez D, Zhou P, Hon GC. Frequent sgRNA-barcode recombination in single-cell perturbation assays. bioRxiv; 2018. doi:10.1101/255638

36. Paul B, Montoya G. CRISPR-Cas12a: Functional overview and applications. Biomed J. 2020;43: 8–17.

37. Zetsche B, Heidenreich M, Mohanraju P, Fedorova I, Kneppers J, DeGennaro EM, et al. Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. Nat Biotechnol. 2017;35: 31–34.

38. Kleinstiver BP, Sousa AA, Walton RT, Tak YE, Hsu JY, Clement K, et al. Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. Nat Biotechnol. 2019;37: 276–282.

39. DeWeirdt PC, Sanson KR, Sangree AK, Hegde M, Hanna RE, Feeley MN, et al. Optimization of AsCas12a for combinatorial genetic screens in human cells. Nat Biotechnol. 2021;39: 94–104.

40. Esmaeili Anvar N, Lin C, Ma X, Wilson LL, Steger R, Sangree AK, et al. Efficient gene knockout and genetic interaction screening using the in4mer CRISPR/Cas12a multiplex knockout platform. Nat Commun. 2024;15: 3577.

41. Aguirre AJ, Meyers RM, Weir BA, Vazquez F, Zhang C-Z, Ben-David U, et al. Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. Cancer Discov. 2016;6: 914–929.

42. Kosicki M, Tomberg K, Bradley A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. Nat Biotechnol.

2018;36: 765–771.

43. Cullot G, Boutin J, Toutain J, Prat F, Pennamen P, Rooryck C, et al. CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. Nat Commun. 2019;10: 1136.

44. Leibowitz ML, Papathanasiou S, Doerfler PA, Blaine LJ, Sun L, Yao Y, et al. Chromothripsis as an on-target consequence of CRISPR-Cas9 genome editing. Nat Genet. 2021;53: 895–905.

45. Álvarez MM, Biayna J, Supek F. TP53-dependent toxicity of CRISPR/Cas9 cuts is differential across genomic loci and can confound genetic screening. Nat Commun. 2022;13: 4520.

46. Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat Protoc. 2013;8: 2180–2196.

47. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell. 2013;152: 1173–1183.

48. Replogle JM, Bonnar JL, Pogson AN, Liem CR, Maier NK, Ding Y, et al. Maximizing CRISPRi efficacy and accessibility with dual-sgRNA libraries and optimal effectors. Elife. 2022;11. doi:10.7554/eLife.81856

49. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596: 583–589.

50. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024;630: 493–500.

51. Conti E, Kuriyan J. Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin alpha. Structure. 2000;8: 329–338.

52. Gier RA, Budinich KA, Evitt NH, Cao Z, Freilich ES, Chen Q, et al. High-performance CRISPR-Cas12a genome editing for combinatorial genetic screening. Nat Commun. 2020;11: 3455.

53. Luo Y, Hitz BC, Gabdank I, Hilton JA, Kagda MS, Lam B, et al. New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. Nucleic Acids Res. 2020;48: D882–D889.

54. Yuen G, Khan FJ, Gao S, Stommel JM, Batchelor E, Wu X, et al. CRISPR/Cas9-mediated gene knockout is insensitive to target copy number but is dependent on guide RNA potency and Cas9/sgRNA threshold expression level. Nucleic Acids Res. 2017;45: 12039–12053.

55. Gong S, Yu HH, Johnson KA, Taylor DW. DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. Cell Rep. 2018;22: 359–371.

56. Strohkendl I, Saifuddin FA, Rybarski JR, Finkelstein IJ, Russell R. Kinetic Basis for DNA Target Specificity of CRISPR-Cas12a. Mol Cell. 2018;71: 816–824.e3.

57. Sinan S, Appleby NM, Chou C-W, Finkelstein IJ, Russell R. Kinetic dissection of pre-crRNA binding and processing by CRISPR-Cas12a. RNA. 2024;30: 1345–1355.

58. Wessels H-H, Méndez-Mancilla A, Hao Y, Papalexi E, Mauck WM 3rd, Lu L, et al. Efficient combinatorial targeting of RNA transcripts in single cells with Cas13 RNA Perturb-seq. Nat Methods. 2023;20: 86–94.

59. Walton RT, Qin Y, Blainey PC. CROPseq-multi: a versatile solution for multiplexed perturbation and decoding in pooled CRISPR screens. bioRxiv. 2024. doi:10.1101/2024.03.17.585235