Title

Improving the Safety and Efficacy of CRISPR-Cas Genome Editing

Author

Tsuchida, Connor

Publication Date

2023

Improving the Safety and Efficacy of CRISPR-Cas Genome Editing

By

Connor A. Tsuchida

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Joint Doctor of Philosophy

with the University of California, San Francisco

in

Bioengineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jennifer A. Doudna
Professor David V. Schaffer
Professor Alexander Marson
Professor Ross C. Wilson

Spring 2023

ABSTRACT

Improving the Safety and Efficacy of CRISPR-Cas Genome Editing

by

Connor A. Tsuchida

Joint Doctor of Philosophy in Bioengineering

with the University of California, San Francisco

University of California, Berkeley

Professor Jennifer A. Doudna, Chair

In just over a decade, clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR-associated (Cas) systems have advanced from first demonstration as a programmable RNA-guided DNA nuclease to use in human clinical trials to provide durable cures for previously incurable genetic diseases. CRISPR-Cas genome editing relies on a guide RNA (gRNA) to target a specific sequence in the genome prior to the generation of a double-strand DNA break (DSB). This mechanism can be exploited to disrupt, insert, or replace nucleotide sequences of interest, providing a powerful method for modulating the human genome for therapeutic benefit. Despite CRISPR-Cas genome editing holding tremendous promise, clinical efficacy and safety concerns over making permanent changes to the human genome remain. This work describes multiple efforts to advance CRISPR-Cas genome editors for improved translation as a human therapeutic, with a focus on increased safety and precision.

Two challenges for translating CRISPR-Cas systems for genome editing in humans are delivery and immunogenicity. While commonly used CRISPR effectors, such as Cas9, have shown great promise, their large size makes therapeutic delivery difficult. Moreover, pre-existing human adaptive immune responses to Cas9 may limit efficacy *in vivo*. To address these challenges, we engineered CasX, a novel and miniature CRISPR effector from non-human-associated microbes, which shows promise for improved delivery and reduced immunogenicity. To achieve this, we determined the cryo-electron microscopy (cryo-EM) structure of a CasX ortholog and used these structural insights to rationally engineer multiple CasX proteins and its gRNA for improved human cell genome editing activity.

In addition to safety concerns regarding CRISPR-Cas genome editing at unintended sites in the genome, another concern is the precision of genomic outcomes at the intended target site. Using two orthogonal assays, we identified that single-arm and entire-chromosome loss was a result of Cas9 DSBs in primary human T cells. Cas9-induced chromosome loss was specific to the targeted chromosome but was a generalizable phenomenon across different gRNAs targeted throughout the genome. T

cells with chromosome loss persisted for weeks during *ex vivo* cell culture but chromosome loss was surprisingly mitigated by simply changing the order of operations during the genome editing protocol. This unexpected finding may explain the lack of Cas9-induced chromosome loss we observed in T cells from patients in a first-in-human phase I clinical trial, and provides the first method for avoiding this potential genotoxicity.

Before ensuring that unintended genomic outcomes are avoided during genome editing, targeting of unintended cell types *in vivo* needs to be avoided. We developed virus-like particles (VLP) that package transient Cas9 ribonucleoproteins (RNP) with or without a transgene of interest. VLPs were capable of complex genome engineering, including both gene insertion and gene disruption, in primary immune cells to yield therapeutic chimeric antigen receptor (CAR) T cells. VLP-based manufacturing of CAR T cells can occur in a single step, greatly simplifying the standard manufacturing of these approved therapies. In addition, we showed that VLPs can be pseudotyped to enable cell-type specific genome editing of $CD4^+$ T cells within a mixed cell population.

Finally, during the COVID-19 pandemic our ongoing efforts to improve the safety and efficacy of CRISPR-Cas genome editing were paused. However, in response to a shortage of clinical testing in our community, we swiftly pivoted our scientific skills and expertise toward building a SARS-CoV-2 clinical testing facility. We established organizational, safety, scientific, regulatory, and clinical practices to ultimately test thousands of patent samples and provide clinical COVID-19 diagnoses.

ACKNOWLEDGEMENTS

To Jennifer Doudna, who has had an enormous impact on my graduate school experience. When I look back 30 years from now, I'll still be in awe that I got to pursue my PhD under your guidance. I still remember writing you an asking to rotate in your lab. I never expected a response but instead I got one less than an hour later. Thank you for taking a chance on a young bioengineering student, the amount I've grown as a scientist and as a person under your mentorship is more than I could have ever expected before I started my PhD. The lab environment you have cultivated is truly remarkable. Spending my formative scientific years in a lab that encourages freedom to think creatively and passionately, and with world-class scientists who are both critical and supportive, has been the opportunity of a lifetime. Thank you for always pushing me when I needed it, whether it was to think bigger, abandon ideas when they weren't panning out, or thoughtfully craft a compelling scientific story. Not many people are lucky enough to witness their mentor win the Nobel Prize – especially in the middle of a global pandemic. Watching you carry yourself throughout all of the media attention and hectic scheduling, while never losing an ounce of passion for the science and the scientists you train, will forever be the model I aspire to emulate.

To Kaihong Zhou, the unsung hero of the Doudna Lab. What would the lab be without your organization, tenacity, and spirit? Your determination to make sure the lab is equipped with anything we need to do our best science can never be appreciated as much as it deserves to be. Thank you for always looking after me these past six years and thank you for always organizing food in lab so we never go hungry.

To Jenny Hamilton, my JD Knockouts co-captain, fellow UW Husky (Go Dawgs!), fellow Washingtonian, and one of the first people I met when I joined the Doudna lab. Six years later I could not be more thankful for that opportune timing, you have been an incredible mentor and friend to a young and naïve PhD student who hardly knew what a lentivirus was before he joined. You have taught me so much in lab and your scientific rigor, creativity, and leadership have been a constant example for me as I have grown as a scientist.

To David Colognori, who a month after joining the lab told me "Connor, we're going to be best friends. I just know it." If there's two things I've learned about you it's that you know nearly everything there is to know about RNA, and that your persistence is unmatched. And to Joy Wang, my day one graduate school work wife, tennis partner, and pandemic 3 PM to 3 AM evening shift companion. I think I was more honored and nervous to give my bridesman speech at your wedding than any scientific conference talk I have every given. When I joined the Doudna lab I never expected to meet two of my best friends along the way. Through all the highs and lows, you two have always been my go-to support system, gossip buddies, and confidants – for that I am eternally grateful for our friendship. I'll always cherish our memories of playing board games (with David's "house rules") and traveling around Las Vegas and Bryce Canyon National Park, but I won't look back with too much longing because I know there's more adventures and memories in our futures.

To Abby Stahl, who has taught me everything from how to inject Cas9 into a mouse brain to how to properly use "yinz" in a sentence. I think I know more about Pittsburg than I do about most cities in America without ever having stepped foot in the city – I can't wait to visit one day. I have always been in awe of your tenacity in lab. Thank you for always being there for the happy celebrations and begrudged complaining sessions, for venturing into Slovenian cave systems, and for the therapeutic relief of petting Fitz and Shady. And to Enrique Lin Shiao, my "twin" and brother-from-another-mother. There's no one I'd rather get mistaken for looking similar to. I'll never forget the haircut you gave me in your backyard during those early days in the pandemic. For that and many other reasons, you have always been a massive support system for me – someone I know I can talk to for career and life advice. I'm supremely grateful to call you a friend. Along with Jenny, I'll forever remember us as the "Big 4" (voiced in my head with Fyodor's distinctive and excited cadence). How could I ever forget those early pandemic days setting up a SARS-CoV-2 clinical testing lab, having Zoom meetings seven days a week, validating assays in full PPE with Dirk Hockemeyer, ordering free coffee on "Joe's tab" at Yali's, and watching the Hamilton liquid handling robot accidentally pick up our 96-well plate full of clinical samples.

To Laura Hofman, my favorite Dutchie. They say graduate school is a marathon not a sprint; your "life is a gift" mentality and support truly helped conquer that last mile (1.61 km). Thank you for your steadfast support and positive energy. I'm so excited for you to begin graduate school, you'll do amazing things. If this acknowledgments section shows you anything, I hope you take away that graduate school is just as much about the people and relationships as it is about the science. Hopefully I can visit the Netherlands one day – I hear Uppsala is nice this time of the year and I can't wait to tell everyone "*Ik spreek een beetje Nederlands.*" And to Kevin Wasko, the person I feared most playing slap cup during the lab retreat. Thank you for all the lively discussions, whether they were about risky scientific ideas or just the latest standup comedian. Thank you both for putting up with my jaded senior grad student energy. Thanks for helping put snow chains on my car during a Tahoe "blizzard" and for the long philosophical hot tub conversations.

To the JD Knockouts MCB summer league softball team. Co-captaining our team to two back-to-back championship games and teaching so many friends which way to run around the bases has been the true highlight of my graduate school experience. To *Bayoncé*, the undisputed best lab bay in the Doudna Lab. Thank you Jenny Hamilton, Joy Wang, Audrey Lapinaite, Kasia Soczek, Riley McGarrigle, Cindy Sandoval Espinoza, and Chenglong Xia for always making coming to work a supportive and enjoyable time. And to the rest of the Doudna Lab, especially Josh Cofsky, Enbo Ma, Liz O'Brien, Gogo Liu, Hardiha Shivram, Tina Liu, Gavin Knott, Patrick Pausch, Brady Cress, Kai Chen, Matthew Kan, Carolyn Huang, Mohammad Saffari Doost, Amy Eggers, Marena Trindad, Keana Lucas, Araz Vartoumian, Johanna Staples-Ager and Jenny Ye, I am immensely grateful for the community of incredible scientific minds, thoughtful mentors, collaborative spirits, and empathetic humans. The biggest strength of the lab isn't the Hawaiian retreats, vendor sponsored pizzas arranged by Kaihong, or pre-poured LB plates, it's the amazing people that make up the group.

To Ross Wilson and Lorena de Oñate Monje, my second lab family and first home in the Innovative Genomics Institute. Ross, thank you for taking in a green rotation student, and for advocating for me every step of the way. Listening to your protein expression Lady Gaga karaoke cover will forever be a top 10 grad school memory for me. Lorena, thank you for being the *hermana mayor* I never imagined I'd ever have. We've been through so many professional and personal highs and lows together, and through it all your love and support have always been constant. A month into the pandemic when it felt like the world had ended, you were on my doorstep on my birthday with a Cheeseboard muffin and a candle.

To Professors Alex Marson, Dave Schaffer, Fyodor Urnov, James Nuñez, and Jimmie Ye. As collaborators, mentors, and colleagues I'm grateful for your academic rigor and for your continuous support of my career. I feel fortunate to have had as many academic examples to look up to as I have.

To Joao Ascensao, the best roommate I've ever had and a tremendous friend. I still think we could have skipped graduate school altogether, got casted on Survivor, and won the whole thing. I am so grateful for our reality TV binges during the pandemic; all the adventures hiking, skiing, tubing, and exploring Slovenian cities; and for our deep conversations on science and life. I truly can't imagine how I would've survived some parts of graduate school without you. And to Jamin Liu, my fellow BioE cohort mate and a friend who has always been there for me in the toughest times. Thanks for being my "San Francisco friend" that gives me a great excuse to visit the city and hang out with you. I'm glad I can always be your "friend with a car" who helps move your giant whale stuffed animal and cans of Spam® across the city. I know when I need someone to talk to and someone to tell me the tough truth, you're always there for me. Thank you for the hours of listening to me and consoling me over the phone and in person, I couldn't ask for a better friend. And also to Juan Hurtado, another BioE cohort mate and another incredible friend. I'm glad we could continuously run into each other on the 5th floor of IGIB, checking in on each other's sanity, complaining about the Sony Sorter, and organizing spontaneous lunch breaks. Our science talks always ignite my scientific creativity and I'm looking forward to more rigorous games of pickleball.

To Lift and Sprint and Grassroots Crossfit. Even though there were a couple times I thought I might pass out or vomit after an especially tough workout, the exercise and community was probably the single biggest blessing for my mental health while in grad school.

To my brothers, Kevin and Michael "Mikey" Tsuchida. Not many people can say they have not one but two lifelong best friends (even though Mikey was a little late to the party). No matter where we are in the world, I can always count on funny memes in my Instagram DMs and that you both are just a phone call away. Regardless of how much older we get or how much time has passed since we last talked on the phone or in person, things always default back to normal when we're together. The roasting, the deep conversations, the inside jokes. And to my parents, Pam and Craig Tsuchida. Thank you for raising me with such a deep respect for education, while never pushing me and always letting me discover my passions on my own. I've always known that you'll support me no matter what, that you're proud of me, and that you love me – and I'm incredibly grateful those have never been in question. The frequent care packages you mailed, whether for my birthday or Lunar New Year or Halloween, kept me going all these years. One of the silver linings of the pandemic was that we all learned how to Zoom. I loved being able to video call every few weeks and feel close to home even when I'm miles away. I'm so thankful for our family and I love you all deeply.

And finally, to everyone else I may have forgotten or didn't have the room to acknowledge here. Whether you played a small part or a big part in my graduate school journey, I’m thankful for every single thing that has happened along the way. I’ve learned so much and grown so much from every triumph and tribulation. Thank you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

CRISPR-Cas genome editing for the treatment of human genetic diseases

A portion of the work presented in this chapter is part of the following manuscript under preparation: Tsuchida C.A.*, Wasko K.*, Hamilton J.R.*, Doudna J.A. Targeted delivery strategies for *in vivo* CRISPR-Cas genome editing. (2023).

*These authors contributed equally.

1.1 Abstract

The programmability and efficacy of CRISPR-Cas systems for genome editing has revolutionized the field and provided enormous hope for providing genetic therapies for previously incurable diseases. While early clinical trials using CRISPR-Cas systems has shown both safety and efficacy, the current scope of diseases that can be feasibly treated is limited. Two major challenges impeding the expansion in the scope of diseases are the delivery as well as the precision and accuracy of CRISPR-Cas genome editors. Here, we discuss the current toolbox of CRISPR-Cas technologies, as well as the established and emerging human delivery methods, with a particular focus on cell-type specific targeting. Additionally, we review the precision and accuracy of CRISPR-Cas effectors at both intended and unintended regions of the genome.

1.2 Introduction

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins have evolved in microbes as a mechanism for adaptive immunity against foreign nucleic acids[1,2]. In nature, these systems contain a CRISPR array composed of repeated DNA sequences alternating with DNA spacer sequences. These spacer sequences are continuously acquired during the adaptation phase from invading foreign nucleic acids and are integrated into the CRISPR array[3]. During the RNA biogenesis phase, the CRISPR array is transcribed as a long pre-CRISPR RNA (crRNA) prior to cleavage into individual crRNAs, each containing a single spacer sequence specific to a foreign nucleic acid sequence. In Class 2 Type II CRISPR systems, a second trans-activating CRISPR RNA (tracrRNA) is separately transcribed and associates with the crRNA to form a duplexed crRNA:tracrRNA[4]. Finally, during the interference phase, one or more Cas effectors form a ribonucleoprotein (RNP) complex with its crRNA or crRNA:tracrRNA. This complex can recognize DNA or RNA sequences based on base pairing with the spacer region of its crRNA, before cleaving the nucleic acid substrate. Together, these three steps allow microbes with CRISPR-Cas systems the ability to recognize and destroy invading foreign nucleic acids, while also providing memory to prevent future infection.

While CRISPR-Cas systems naturally provide adaptive immunity, they have been co-opted for programmable genome editing. Synthetic guide RNAs (gRNA) resemble the natural crRNA or crRNA:tracrRNA but contain a user defined spacer. Complexing this gRNA with a Cas effector generates an RNP capable of recognizing and cleaving a target DNA or RNA. The ability to reprogram this gRNA has proven immensely powerful because it allows for user-defined targeting of the genome in numerous organisms. With CRISPR-Cas systems, the ability to delete, replace, or insert sequences of interest has become more tractable at nearly any site in the genome.

1.3 Therapeutic CRISPR-Cas genome editing

1.3.1 Divergent CRISPR-Cas systems for human genome and transcriptome editing

The continuous arms race between prokaryotes and viruses has provided evolutionary pressure to diversify CRISPR immune systems, many of which have been harnessed for genome or transcriptome editing[5–7]. CRISPR-Cas systems are assigned to two classes (Class 1 and Class 2) along with six types (Type I-VI), based on their unique composition of Cas proteins and nucleic acid target specificity during interference (Figure 1.1).

Class 1 systems rely on multiprotein interference complexes and are the most ubiquitously found CRISPR systems in prokaryotes, accounting for approximately 90% of all systems in bacteria and archaea[8,9]. Class 1 Type I systems are comprised of a crRNA-bound CRISPR-associated complex for antiviral defense (Cascade) which recruits the endonuclease Cas3 for target degradation. Cascade-Cas3 has been utilized for human cell genome editing, showing high processivity at the target site that results in large deletions ranging from hundreds to thousands of base pairs[10,11]. Another Class 1 system, Type III CRISPR effectors, also utilizes a multiprotein complex but targets RNA for destruction[12]. Type III CRISPR-Csm effectors have been efficiently employed in human cells to target both nuclear and cytoplasmic RNAs, offering a robust system for transcriptome modification[13]. While Class 1 systems are the most ubiquitously occurring CRISPR systems, their large size and complexity have limited their translation to therapeutic human genome or transcriptome engineering.

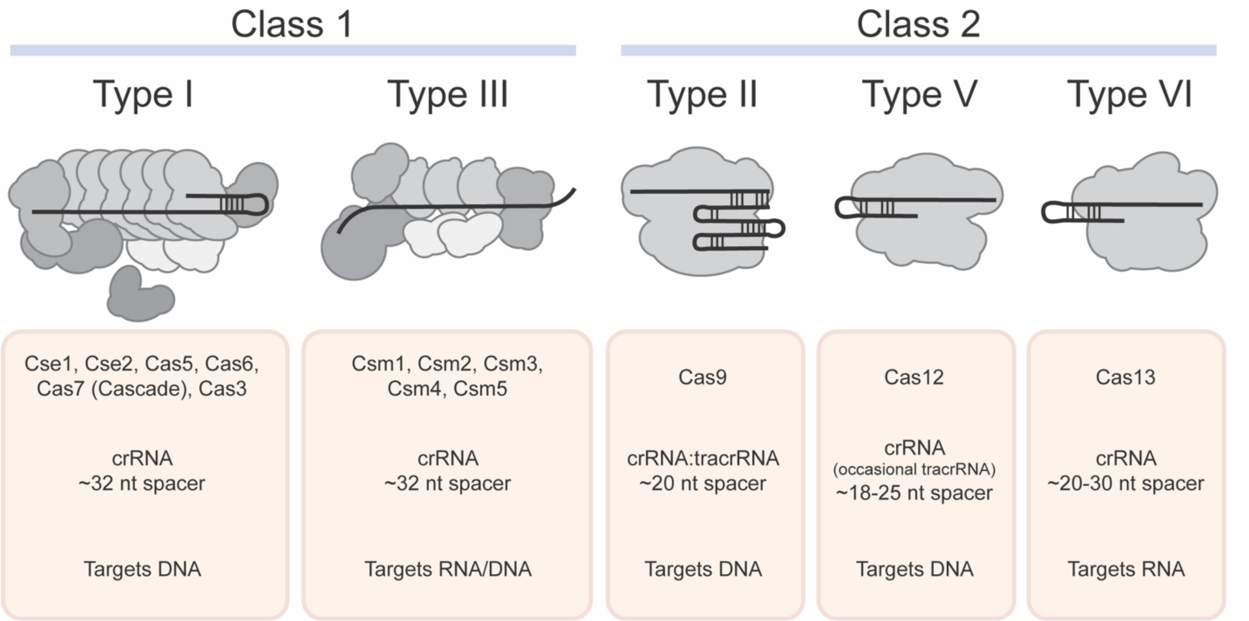


Figure 1.1: Diversity of CRISPR-Cas systems harnessed for human genome editing. Class 1 (left) and Class 2 (right) systems differ by their use of multiple or single protein interference complexes, respectively. Proteins in the complex, RNA requirement, and target specificity are displayed below each Type.

Class 2 systems evolved with a single protein for nucleic acid interference. Class 2 Type II systems rely on Cas9 complexed with a crRNA:tracrRNA to bind and cleave DNA. The first CRISPR system to be reprogrammed for genome editing was Cas9 from *S. pyogenes*, which continues to be the most popularly used CRISPR effector because of its well-studied mechanism and high efficacy[14–17]. More recently, Type V systems, which utilize Cas12, have begun to be discovered and used for human genome editing. More diverse than Type II systems, the wide variance in Type V systems has been exploited to discover protein variants with beneficial properties such as reduced size and trans-cleavage activity[18–23]. Finally, Type VI systems employ the single protein effector Cas13 for targeted degradation of RNA[24–26]. While RNA targeting with Type III and Type VI CRISPR systems[27] holds great promise for modulating the expression of genes in human disease without creating permanent alterations to the genome, this review will be solely focused on DNA targeting Type II and Type V systems.

1.3.2 DNA repair mechanisms after CRISPR-Cas cleavage

The Type II and Type V CRISPR effectors Cas9 and Cas12 utilize two or one catalytic domain to cleave the phosphodiester DNA backbone, respectively. Cleavage by Cas9 results in blunt DNA ends[28] while cleavage by Cas12 results in staggered DNA ends with 5' overhangs of ~5-9 nucleotides (nt) in length[18,20,29]. After generation of a DSB, genome editing relies on endogenous cellular DNA repair pathways to resolve the lesion[30,31]. Despite the observation that Cas9 remains bound to its target DNA for hours after cleavage *in vitro*, its resonance time on cleaved DNA is just seconds to minutes in cells, likely because of displacement by DNA repair proteins[32–34].

DNA DSB repair in human cells predominately occurs through end joining or templated repair. The two predominate end joining pathways are non-homologous end joining (NHEJ) and microhomology-mediated end joining (MMEJ)[35]. NHEJ is often considered the predominate outcome of DSB repair, as it involves rapid re-ligation of the DNA ends and is active in all phases of the cell cycle except mitosis (Figure 1.2A). DSBs generated by CRISPR effectors are usually repaired scarlessly by NHEJ, with one study estimating that NHEJ perfectly repaired DSBs 75% of the time[36], which is then re-targetable by the CRISPR effector. On occasion, NHEJ will result in small insertions or deletions (indels) which terminates this cycle of perfect repair and cleavage by changing the target sequence. These indels can be used to change the reading frame of a targeted gene, disrupting expression of the associated protein.

MMEJ involves short resection of the cleaved DNA ends, which exposes regions of homology used to repair the two ends (Figure 1.2B)[37]. End joining repair between regions of homology often results in larger deletions than the indels created NHEJ, and has been harnessed to collapse pathogenic microduplications in cells with Limb-Girdle Muscular Dystrophy 2G and Hermansky-Pudlak Syndrome Type 1[38].

While NHEJ and MMEJ are error-prone DNA repair pathways, templated repair provides a mechanism for the incorporation of specific sequences between the cleaved DNA fragments. Homologous recombination involves long range resection of the DNA ends, followed by templated repair using another DNA molecule with homologous sequences – often a sister chromatid. Homology-directed repair (HDR) exploits this mechanism by using an exogenous DNA template with homology to either end of the DSB to introduce a new sequence between the cleavage site (Figure 1.2C)[39]. HDR is considered to be less efficient than end joining pathways, takes longer, and is restricted to the S/G2/M phases of the cell cycle[30].

Importantly, DSB DNA repair pathways are in competition to resolve the DNA lesion. Several factors including cell cycle, presence or absence of a homologous sequence, and cell type can drastically bias the repair toward one pathway over another. While the DNA repair pathway chosen to resolve a single lesion is mutually exclusive, within a bulk population of cells with DSBs a variety of DNA repair pathways and resulting genetic outcomes can be present. With this knowledge, several studies have sought to artificially increase the frequency of HDR compared to NHEJ or MMEJ for accurate genome editing. Inhibition of DNA ligase IV, a key component of NHEJ, with Scr7 and small molecule-mediated cell cycle arrest in the HDR-favored S phase both improved rates of HDR[40–42].

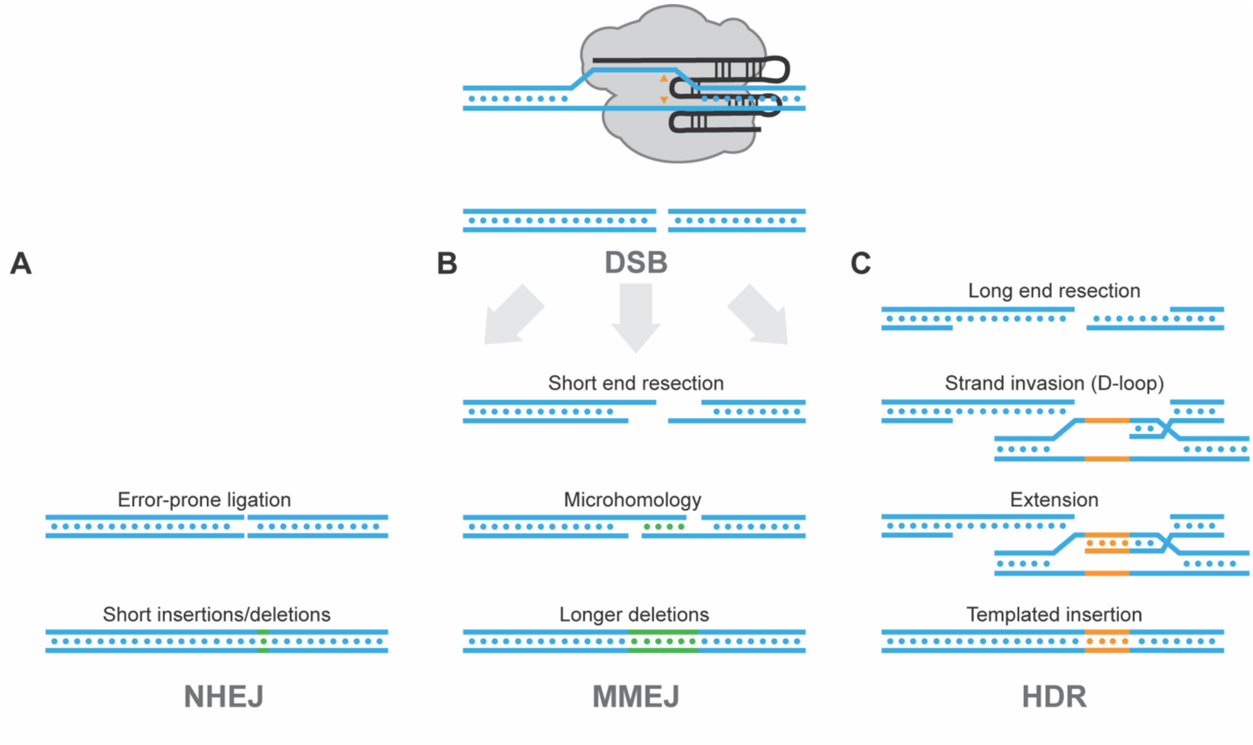


Figure 1.2: DNA repair outcomes after CRISPR-Cas cleavage. (A) NHEJ involves error-prone ligation that results in short insertions or deletions (indels). **(B)** MMEJ involves short end resection and results in longer deletions. **(C)** HDR involves long end resection, strand invasion with an exogenous template, and templated insertion.

1.3.3 CRISPR-Cas genetic engineering without DSBs

1.3.3.1 Base editing for single DNA base pair conversion

CRISPR-Cas genome editing relies on DSBs to generate indels capable of disrupting a gene of interest, or to insert a gene of interest. Disrupting a gene can be therapeutically beneficial in the case of a dominant negative genetic mutation or a pathogenic allele that is non-essential. Inserting a genetic sequence is preferred in the case of correcting a pathogenic mutation or imparting new genetic function. However, gene insertion via HDR is inefficient and DSB repair leads to a heterogeneity of different genetic outcomes. To improve both the efficacy and purity of genome editing outcomes, several new CRISPR-

Cas-based technologies have been developed that modify the human genome without DSBs including base editors, prime editors, epigenome editors, and CRISPR-associated transposons[43,44].

Base editors harness the programmable targeting of CRISPR-Cas9 combined with deaminase enzymes capable of deaminating one nucleotide into another (Figure 1.3A)[45,46]. The majority of known human genetic variants associated with disease are cause by a single nucleotide mutation, which inspired the need for a more efficient technology that could instill a single base change with high purity. The first base editor developed utilized a catalytically dead Cas9 (dCas9) that could still use a gRNA to bind a target sequence and generate an R-loop, but lacked the catalytic ability to cleave either DNA strand. The rat APOBEC1 cytidine deaminase accepts single-stranded DNA (ssDNA) substrates and deaminates the amine of cytosine (C) into uracil (U). Fusion of this cytidine deaminase to dCas9 resulted in C to U conversion, within a specific ~5 bp window generated by the dCas9 R-loop structure[47]. This cytosine base editor (CBE) worked efficiently *in vitro* but remained inefficient in human cells, which was hypothesized to be because of mammalian DNA repair mechanisms. Uracil DNA glycosylase can excise the uracil from a U:G mismatched DNA duplex, which would be counterproductive to the CBE. Addition of a uracil DNA glycosylase inhibitor (UGI) significantly improved CBE efficacy[47]. To further improve CBE efficacy, one catalytic domain on Cas9 was restored to create a nicking Cas9 (nCas9) capable of cleaving only the non-base edited strand of the DNA duplex (originally termed BE3 or CBE3). This nick biased DNA repair toward resolving the U:G mismatched DNA duplex into a U:A DNA duplex, which results in a T:A duplex after replication or repair, rather than restoring the C:G DNA duplex.

The creation of adenine base editors (ABE) capable of converting adenosine (A) to guanosine (G) required additional engineering, because no known adenine deaminases work on DNA substrates. Directed evolution was used on *E. coli* TadA, which naturally converts adenine to inosine (I) in the single-stranded anticodon loop of tRNAs, to accept ssDNA substrates[48]. Fusion of this engineered TadA to nCas9 resulted in targeted conversion of A to I, which is read as G within the cell. Continued evolution on the TadA deaminase has resulted in increasingly efficient A to G base editing activity, with the current ABE8 showing the highest rates to date[49,50]. Interestingly, further directed evolution of TadA fused to nCas9 has resulted in TadA-based CBEs capable of cytosine base editing, along with cytosine and adenine base editors (CABE) capable of performing both deamination reactions[51].

While the development of base editors that do not rely on DSBs limits some of the potential genotoxicities associated with nuclease-active CRISPR-Cas genome editing (discussed later in this review), extensive research has elucidated several off-target outcomes. Even with base editors, low levels of indels are observed[47,48]. In addition, base editors have shown both gRNA-dependent and gRNA-independent off-target DNA editing. gRNA-dependent editing occurs when the gRNA shares perfect or partial sequence complementarity with an unintended region of the genome; in this case, both ABEs and CBEs are capable of modifying unintended bases. Engineered high fidelity Cas9 variants have been utilized to minimize gRNA-dependent off-target editing[45,52], along with computational methods to determine the genome-wide specificity of individual gRNAs[53]. gRNA-independent off-target editing occurs at sites with no homology to the Cas9 gRNA, likely because of the overexpression of the deaminase. Whole genome sequencing discovered that CBEs generate infrequent gRNA-independent off-target DNA edits, though engineered CBEs have mitigated these events[54–56]. More recently, it was discovered that both ABEs and CBEs displayed off-target editing of RNA within the cell. Again engineering of both the APOBEC1 and TadA deaminases have yielded ABEs and CBEs that reduce this unwanted RNA editing activity[57–59].

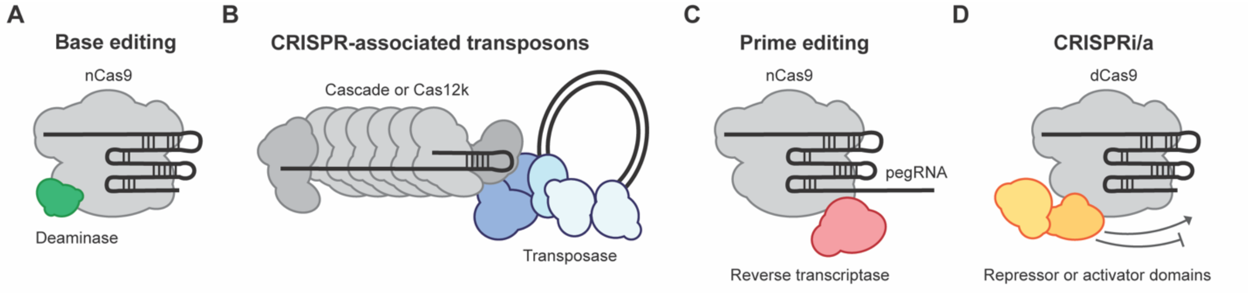


Figure 1.3: Novel CRISPR-Cas-based technologies for genetic engineering without DSBs. **(A)** Adenine (ABE) and cytosine (CBE) base editors utilize a deaminase domain fused to nicking Cas9. **(B)** CRISPR-associated transposons (CAST) utilize nuclease-deficient Type I or Type V Cas effectors that have evolved to associate with a targeting-deficient transposase. **(C)** Prime editing (PE) utilizes an extended gRNA encoding a template (pegRNA) and nicking Cas9 fused to a reverse transcriptase. **(D)** CRISPR interference and activation (CRISPRi/a) utilizes dead Cas9 fused to genetic repressor or activator domains.

1.3.3.2 Gene insertion with CRISPR-associated transposons and prime editing

The ability to introduce new genetic sequences within the human genome has been a long-standing goal in the fields of genome editing, cell engineering, and synthetic biology. Gene insertion promises the ability to replace entire genes containing multiple disease-causing mutations or to introduce an entirely new gene and cellular phenotype. The current clinical standard for gene insertion is modified retroviral and lentiviral vectors, which semi-randomly integrate transgenes as large as ~10 kb[60,61]. While this method has yielded several FDA-approved therapies, the semi-random integration poses a risk of insertional-oncogenesis[62,63]. Targeted integration can avoid the risk of insertional-oncogenesis and additionally can improve the expression and cellular phenotype of engineered cells. CRISPR-Cas9 genome editing via HDR has yielded gene insertion in numerous cell types[64,65], including T cells in clinical trials[66,67], yet efficacy, size constraints, and product purity remain a challenge. Novel CRISPR technologies, both discovered natural and engineered, have promised more efficient gene insertion without the generation of DSBs.

Naturally, CRISPR systems provide defense against invading nucleic acids such as transposons, which act as selfish genetic elements replicating themselves in the genome. While transposition is normally random, insertion into an essential gene is detrimental to both the host and transposon, as neither continues to replicate. This selective pressure could have driven the evolution of CRISPR-associated transposons (CAST) which mediate site-specific transposition. CASTs were discovered bioinformatically, through the identification of both Type I and Type V CRISPR systems within transposons (Figure 1.3B)[68,69]. Interestingly, many of these Tn7-like transposons lacked the domains necessary for DNA targeting and many of the CRISPR-Cas systems lacked the domains necessary for catalytic cleavage of the genome; this suggests a potential cooperation between these two systems. Since this discovery, both the Type I CAST from *V. cholerae* and the Type V CASTs from *S. hofmanni* and *A. cylindrica* have been demonstrated to achieve programmable gene insertion[70,71]. The Type I CAST utilizes Cascade and a 32 bp gRNA to direct the transposase complex to insert sequences with a left and right transposon sequence[70]. In *E. coli*, gene insertion was highly specific to 46-55 bp from the Cascade gRNA target site, and preferred insertions of approximately 800 bp in size, but could insert cargo as large as 10 kb[70]. The Type V CASTs utilize Cas12k as the DNA targeting mechanism, as opposed to the multi-subunit Cascade complex. Similarly to the Type I CAST, the Type V CASTs showed gene insertion in *E. coli* that was approximately 60-66 bp from the target site, however subsequent work revealed poor genome-wide specificity of Type V CASTs compared to the Type I CAST[72]. While these systems have already proven to be powerful methods for gene insertion into microbial genomes, efficient use in human cells has not been demonstrated. Recently, another Type I CAST from *Pseudoalteromonas* provided the first demonstration of gene integration into the human genome when supplemented with the protein dissociation factor ClpX, albeit at a maximum of ~1%[73].

Prime editing relies on an extended gRNA that serves as a template for targeted short sequence insertion or deletion (Figure 1.3C)[44]. The protein component of prime editors is a nCas9 fused to an engineered reverse transcriptase[74]. The extended gRNA, termed the pegRNA, encodes both the nCas9 target sequence, along with a primer binding site (PBS) and the desired edit on the 3' end. After nicking the target DNA, the 3' PBS of the pegRNA binds the liberated DNA strand and the edit is reverse transcribed from the pegRNA into the DNA. Supplying an additional gRNA to nick the non-edited strand (termed PE3) drives DNA repair and incorporation of the edit onto the second DNA strand. Prime editing is capable of creating all 12 point mutations, with templated insertions or deletions less than 50 bp in size; the use of two prime editors targeting opposite strands has been utilized for templated insertions or deletions up to 100-500 bp in size[75,76]. While these targeted insertions are smaller than seen with CAST systems, prime editing has proven efficacious in a variety of human cell types, including to treat phenylketonuria in the liver[77], sickle cell disease in hematopoietic stem cells[78], and Duchenne muscular dystrophy in cardiomyocytes[79].

1.3.3.3 Gene silence and activation with epigenome editing

Genome editing promises the potential for single dose, permanent cures for numerous genetic diseases. However, this permanence is also a point of caution for the field; along with permanently introducing the desired genomic edit, any potential unintended edits will also be permanent. Additionally, our understanding of gene networks is far from complete, meaning the editing of a target gene could have unknown epistatic effects on other genes within the cell. Epigenetics involves the expression of genes without changes to the underlying sequence. Epigenome editing, therefore, has become an increasingly popular method to alter the cellular phenotype because of its tunability, reversibility, and non-permanence[80,81].

CRISPR interference (CRISPRi) relies on modifying the epigenetic context of genes to silence gene expression (Figure 1.3D). dCas9 fused to the Krüppel-associated box

(KRAB) domain of Kox1 pairs the programmable targeting of dCas9 with an effector for epigenetic silencing[82]. Later studies showed that dCas9-KRAB induced H3K9 tri-methylation (H2K9me3) at targeted sites, generating heterochromatin and reducing gene accessibility and expression[83]. While CRISPRi is favorable because it does not make changes to the underlying genome, continued expression of the effector is necessary for durable gene repression. To overcome this challenge, several methods have been devised to instill lasting epigenetic silence. In both cases, addition of DNA methyltransferases (DNMT), which naturally silence endogenous retroviruses in the human genome, generated targeted gene silencing that was inherited for up to 450 cell divisions[84,85].

Multiple technologies have also been developed to epigenetically activate a gene of interest, termed CRISPR activation (CRISPRa) (Figure 1.3D). While the initial fusion of four copies of the VP16 transcriptional activator (VP64) or the p65 activation domain (p65AD) to dCas9 yielded modest transcriptional upregulation[82], further engineering was necessary to demonstrate robust CRISPRa. A tripartite fusion of VP64, p65, and Rta (VPR) significantly improved targeted gene activation with a single effector[86]. In addition, recruitment of high copy numbers of activation domains has further improved CRISPRa technology. The synergistic activation mediator (SAM) system utilizes dCas9-VP64 along with a gRNA with modified stem loops that include MS2 aptamer sequences[87]. These aptamer sequences can recruit multiple MS2-p65-heat shock factor 1 (HSF1) domains that are co-delivered, enhancing transcriptional activation. Similarly, dCas9 fused to a series of novel epitope tags, termed SunTag, can recruit multiple copies of the cognate antibody fused to VP64, which also showed improved transcriptional activation[88].

While epigenome editing offers enhanced safety by avoiding permeant changes to a patient's genome, it relies on modulating endogenous gene expression. Therefore, its therapeutic benefit is not well suited for diseases where the protein product of a gene needs to be fixed but still expressed, or for introducing new genetic sequences to enhance cellular function.

1.3.4 Early clinical trials with CRISPR-Cas genome editing

The first demonstration that CRISPR-Cas9 could be harnessed as a programmable nuclease for genome editing occurred in 2012[14], with application in human cells shortly thereafter[15–17]. In less than a decade, the repurposing of this prokaryotic immune system for human therapeutic genome editing has reached human clinical trials[89]. Thus far, Cas9 genome editing in humans has been largely limited to the simplest targets, where delivery to the cells of interest is feasible and gene knockout is therapeutically beneficial. Still, early observations have shown safety, efficacy, and therapeutic benefit, giving patients living with previously incurable genetic diseases a reason to be hopeful.

Sickle cell disease (SCD) arises from a single point mutation in the beta-globin gene *HBB*, which reduces the oxygen carrying capacity of red blood cells (RBC) and leads to stiff crescent-shaped RBCs that are prone to aggregate, causing vaso-occlusion, severe pain, and organ failure. Both the well-studied nature of this genetic disease, as well as the accessibility of the target cells made SCD one of the early targets for CRISPR-Cas9 genome editing. Hematopoietic stem and progenitor cells (HSPC) give rise to all blood cells in humans, including RBCs. HSPCs can be mobilized out of the bone marrow, collected, genome edited, and returned to the autologous patient[90]. In this way, HSPCs can be efficiently targeted with Cas9 *ex vivo*, which negates the concern over editing unintended cell types. Several methods for curing SCD with Cas9 have been demonstrated, including correcting the disease-causing mutation via HDR with a synthetic template[65] or adeno-associated virus (AAV) encoding an HDR template[91]. An alternative approach is inducing expression of fetal-globin, which is normally silenced in adults but can phenotypically replace beta-globin. BCL11A is repressor of fetal-globin expression[92]; Cas9 disruption of the *BCL11A* enhancer restores fetal-globin expression and ameliorates the sickled phenotype associated with SCD[93,94].

This latter strategy has been employed in multiple clinical trials for patients with SCD, along with the closely related disease beta thalassemia, including by CRISPR Therapeutics and Vertex Pharmaceuticals in the United States[95] and Bioray Laboratories and several academic institutions in China[96]. In both cases, Cas9 RNP was electroporated into isolated HSPCs and autologously re-infused into patients. Increased levels of fetal-globin were observed in patients, along with a concurrent reduction in vaso-occlusive episodes and transfusion independence for over a year. While the number of patients treated with these therapies is relatively small and clinical follow up needs to continue for additional years, the promising results suggests a durable cure for these genetic diseases. It is expected that the therapy *exa-cel* from CRISPR Therapeutics and Vertex Pharmaceuticals will be the first FDA-approved CRISPR-based therapy in late 2023 or 2024[97].

Significant clinical progress has also been achieved with CRISPR-Cas9 engineered T cells for immuno-oncology. T cells retargeted toward cancer-associated antigens, either through the introduction of an exogenous T cell receptor (TCR) or synthetic chimeric antigen receptor (CAR), can lead to durable remissions of blood cancers[98]. However, challenges remain including efficacy and persistence of the engineered cells, resulting in significant numbers of patients without therapeutic benefit[99,100]. Genome editing is being explored to further engineer these T cells with the goal of improving potency and preventing exhaustion.

CRISPR-Cas9 genome editing was first tested in a phase I clinical trial in T cells retargeted with a TCR against the NY-ESO-1 cancer antigen in three patients with advanced, refractory cancer[101]. Along with the lentivirally-integrated transgene, Cas9 RNPs were electroporated into autologous T cells to disrupt *TRAC, TRBC,* and *PDCD1*. Disruption of *TRAC* and *TRBC*, which encode parts of the TCR, prevents chimeric pairing of endogenous and exogenous TCR components, and disruption of *PDCD1* which encodes the immune checkpoint protein PD-1, improves T cell persistence. While single-cell RNA sequencing (scRNA-seq) revealed a mixed population of T cells with different genetic changes – just ~10% of cells had all three Cas9 edits – the engineered cells showed improved persistence and were detectable as far out as 9 months[101]. Translocations between the Cas9 cut sites were detectable but declined *in vivo* to the limit of detection by qPCR. In addition to no significant therapy-related toxicities, reduction in the tumor-associated antigens was also observed.

More recently, base editors have been applied to conduct multiplexed engineering in T cells. A major advantage of base editors for simultaneous editing of multiple loci is that the absence of DSBs lowers the incidence of translocations (discussed later in this review)[102,103]. A recent phase I clinical trial for the treatment of T cell leukemia and lymphoma is employing T cells engineered with an anti-CD7 CAR and edited with CBEs to disrupt *TRAC*, *CD7*, and *CD52*[104–106]. Since CD7 is expressed normally on T cells, manufacturing CAR T cells against CD7 induces fratricide, the destruction of neighboring engineered T cells. Disrupting *CD7* prevents "friendly fire" between CAR T cells while preserving the ability to target cancerous T cells. *CD52* was also disrupted so that lymphodepleting anti-CD52 antibodies could be used as a combination therapy.

Along with using Cas9 to disrupt genes within T cells, recent clinical progress has utilized HDR for non-viral integration of the retargeted TCR or CAR in a site-specific manner. Cas9-mediated HDR to integrate a CAR under the native *TRAC* promoter displays uniform expression, improved potency, and minimal tonic signaling associated with pre-mature T cell exhaustion[107]. Following up on this discovery, computationally derived neoantigen TCRs were integrated into the *TRAC* locus, simultaneously disrupting endogenous TCR expression and retargeting the T cells[66]. In sixteen patients, these therapies proved clinically safe, with some evidence of engineered T cell trafficking to the tumor site. Similarly, a CAR was integrated into the *PDCD1* gene using Cas9 HDR[67]. Non-viral integration and disruption of the immune checkpoint protein led to an increased number of memory T cells and an 87.5% remission rate in eight cancer patients.

HSPCs and T cells were good candidates for early clinical trials because editing these cell types can occur *ex vivo*, minimizing the challenge of delivery and concern over editing an unintended tissue. However, non-hematopoietic cell types are not as easily isolated or amenable to transplantation back into patients, necessitating *in vivo* delivery. The liver is an attractive first target for *in vivo* delivery because nanoparticles passively accumulate within the tissue[108].

Transthyretin amyloidosis occurs when misfolded transthyretin (TTR) proteins aggregate in the nervous and cardiovascular system, causing cardiomyopathy and heart failure. Since TTR is produced in the liver, a strategy was devised by Intellia Therapeutics and Regeneron Pharmaceuticals to systemically administer lipid nanoparticles packaging an mRNA encoding Cas9 and gRNA targeting the wildtype and mutant TTR gene[109]. TTR reduction of 52% was observed in patients receiving a low dose, while TTR reduction of 87% was observed in patients receiving a high dose, along with no adverse events[109].

Another clinical approach to liver genome editing is being led by Verve Therapeutics. Individuals with loss-of-function mutations in *PCSK9* have low levels of low-density lipoprotein (LDL), which is a major contributor to cardiovascular disease[110]. To replicate this genetic variation, lower the levels of LDL, and concurrently lower the risk of cardiovascular disease, ABEs encoded within an mRNA and a gRNA targeting *PCSK9* have been packaged within a lipid nanoparticle for systemic injection targeting the liver[111]. In non-human primates, base editing was highly efficacious and observed almost entirely in the liver. PCSK9 expression was almost entirely eliminated, which resulted in a ~50% reduction in LDL levels[111,112]. Based on these promising results in non-human primates, clinical trials have begun in patients with heterozygous familial hypercholesterolemia.

1.4 Delivery of CRISPR-Cas effectors in humans

Rapid progress characterizing, improving, and engineering Cas effectors for genome editing has resulted in robust tools for gene disruption, insertion, replacement, and silencing. While these tools work well in *in vitro* cell culture experiments, translation toward a therapeutic medicine for patients is a much larger challenge. The largest component of this challenge is specific delivery of genome editing effectors to desired cell types, tissues, or organs within a patient. For many genetic diseases, only a subset of cells or a specific organ show phenotypic signs of disease and thus would be the intended target for genome editing. Editing of unintended cells or organs would substantially increase the risk of inadvertent, deleterious consequences. The first clinical trials using CRISPR-Cas genome editing have focused on hematopoietic cell types, such as T cells and HSPCs, since they can be manipulated *ex vivo*, along with the liver, which passively accumulates injected particles. Thus, targeted delivery of genome editors remains an unaddressed necessity for the successful translation of genome editing toward a wider range of diseases.

Viral vectors, stripped of their native genome and ability to replicate, have become a popular delivery method for gene therapy and more recently CRISPR-Cas genome editing. Two of the most utilized viral vectors include lentivirus and AAV[113,114]. Lentiviral vectors express large transgenes (~10 kb) after genomic integration while AAV express a smaller transgene (~4.5 kb) from a long-lived episome; both viral vectors are capable of transducing dividing and non-dividing cells. Pseudotyping of lentiviral vectors, engineering of new AAV tropisms, and the use of tissue specific promoters has enabled cell-specific delivery with these technologies, yet viral delivery also introduces new safety concerns. Immunogenicity to the viral vector as well as to the overexpressed transgene product may counteract therapeutic efficacy and stimulate an immune response[115–117]. The long-term expression of CRISPR nucleases from lentiviral integration or AAV episomes also increases the risk of off-target genomic edits. Additionally, while predominantly non-integrating, AAV has been shown to integrate into sites of DSBs including preferentially at Cas9 target sites[118,119]. Viral integration, either by lentivirus at a high level or AAV at a low level, presents a serious concern over disrupting an essential gene. Previously, a lentiviral gene therapy resulted in four out of nine patients developing leukemia, which was attributed to insertional oncogenesis[63].

Transient delivery of CRISPR-Cas genome editors could eliminate the safety concerns over toxicity, off-target editing, and insertional oncogenesis. A growing number of researchers and clinicians have turned toward using RNPs, where the Cas protein is complexed with its gRNA, or encoding the Cas nuclease as an mRNA. Both methods are inherently transient in cells and have zero risk of insertional oncogenesis. However, unlike viruses with naturally evolved cell-specific tropisms, RNPs and mRNA have no intrinsic ability to target particular cell types or gain entry into the cell, necessitating significant engineering. Here, we review emerging strategies for targeted delivery of CRISPR-Cas genome editors, with a specific focus on delivering transient RNP or mRNA.

1.4.1 Targeted delivery of engineered Cas9 ribonucleoproteins

Over the past 40 years, protein therapeutics have exploded in use, with proteins such as human growth hormone for hormone deficiency and viral proteins as vaccines becoming standards of care[120]. Similarly, the use of CRISPR-Cas proteins as a therapeutic, has become increasingly popular. Unlike genetic encoding, Cas9 RNPs exist transiently within a cell, with degradation occurring after ~24 hours[121]. This transient nature minimizes off-target editing and immunogenicity while retaining high genome editing efficacy. In addition, the production of recombinant Cas9 is significantly cheaper than producing viral vectors encoding Cas9, which requires costly and laborious mammalian or insect cell culture[122]. While Cas9 RNPs offer numerous advantages for therapeutic genome editing, a major barrier to their successful translation is targeted delivery to cell types and organs of interest.

1.4.1.1 Physical isolation of cells for targeted delivery

One method for cell type specific delivery is to physically isolate the target cells for *ex vivo* – outside of the body – genome editing. This physical isolation guarantees the prevention of Cas9 delivery to unintended cell types, and also allows for monitoring and quality control of the genome edited cells while in *ex vivo* culture.

Hematopoietic cells are perhaps the easiest cell types in the human body to physically isolate; both lymphoid and myeloid cells can be isolated from peripheral blood through centrifugation or surface marker-based cell sorting. T cells have been widely engineered to reprogram their antigen specificities to combat cancer and autoimmune diseases. Since plasmid DNA was found to be inefficient and toxic when transfected into T cells, researchers have largely turned to Cas9 RNPs for genome editing[123,124]. Electroporation of Cas9 RNPs, where high voltage temporarily increases membrane permeability, results in highly efficient gene knockout as well as gene insertion, including in numerous clinical trials[64,66,67,101,125]. HSPCs are also a major target for genome editing in patients for a number of hematologic diseases. Again, Cas9 RNP electroporation into HSPCs has proven to be highly efficacious, though their isolation requires chemical mobilization from the bone marrow niche[126].

The advent of *in vitro* fertilization (IVF) in the late 20th century pioneered the isolation; cryopreservation; and handling of human gametes, zygotes, and embryos. While numerous advocates have voiced concerns over human germline editing, there remains interest in this research in order to better understand embryogenesis and fetal development, along with to treat genetic diseases prior to birth. Since both sperm and egg can be collected from patients and remain viable *ex vivo*, the majority of Cas9 genome editing in reproductive cells has been targeted via physical isolation. Isolated embryos have been microinjected with Cas9 RNP, which showed higher efficacy compared to mRNA and reduced mosaicism[127,128]. Since microinjection requires highly specialized equipment and personnel, methods for electroporation of Cas9 RNP into mouse or rat zygotes have also been developed[129–131].

The physical isolation of target cells, such as hematopoietic and reproductive cells, prior to Cas9 delivery is a powerful method for cell specific genome editing. Physical isolation completely avoids the risk of delivery to unintended cell types, and the transient lifetime of Cas9 RNP means the genome editing machinery is completely degraded prior to the cells being reintroduced into the patient. Additional advantages to *ex vivo* genome editing include being able to expand the cells as well as monitor and sequence the cells prior to reintroduction; this can increase the therapeutic dose and provide quality assurance. However, it is unlikely that this physical isolation method for cell specific targeting with Cas9 RNP delivery will expand in scope, since many cell types and tissues are unable to be cultured *ex vivo* or successfully implanted back into the patient.

1.4.1.2 Targeted genome editing *in vivo* via direct delivery

In vivo rather than *ex vivo* genome editing is desirable because it does not require additional equipment and labor to maintain the cell product in culture. Additionally, *in vivo* delivery can theoretically target any cell type, tissue, or organ within a living patient. However, herein also lies the crux of *in vivo* delivery; without physical isolation to target specific cell types, *in vivo* administration can indiscriminately target any cell type or organ. This level of cellular uncertainty is likely intolerable for genomic medicines, where a subset of cells or diseased tissue is the desired target. One method for targeted *in vivo* delivery is direct injection into the desired site. Since CRISPR-Cas effectors like Cas9 are relatively large (~10 nm across) their biodistribution is minimal and limited to immediately neighboring the injection site.

The brain is a major focus for therapeutic genome editing, especially for neurodegenerative diseases with limited current treatment options[132]. Targeting the brain is a unique challenge because of the blood-brain barrier (BBB) which is composed of tightly packed cells around the neurovasculature that highly regulate what can cross from the bloodstream into the brain. The BBB has historically been a challenge when brain targeting therapeutics are administered systemically; however, this cellular barrier also means that molecules in the brain are unlikely to spread to other tissues[133]. Direct *in vivo* injection into the hippocampus, striatum, and cortex has been achieved with Cas9 RNPs endowed with cell penetrating capabilities by fusing six SV40 nuclear localization sequences (NLS) to the protein, as well as by conjugating a neuron axonal import peptide to Cas12a[134,135]. Interestingly, both of these methods selectively targeted neurons over glial cells, but only within several hundred micrometers around the injection site. This biodistribution has been enhanced by conjugating polyethylene glycol (PEG) to Cas9 RNPs via a reduction cleavable linker[136]. PEG conjugation improved the area of genome editing by Cas9 RNP 3-4-fold after direct injection into the striatum. Another method for efficient brain editing is CRISPR-Gold. CRISPR-Gold is composed of gold nanoparticles covered in DNA oligonucleotides and Cas9 or Cas12a RNP, along with a layer of silica and a cationic endosomal disruptive polymer[137]. CRISPR-Gold was directly injected into the dente gyrus, hippocampus, or striatum which resulted in efficient genome editing of neurons, glia, and astrocytes and was observable 1-2 mm away from the site of direct injection[138].

The eye and skin are also easily accessible for direct injection *in vivo* of Cas9 RNP, presenting the ability to directly target these tissues. Subretinal injection has become a routine medical procedure in vitrectomies and is used to deliver Luxturna, the first FDA-approved *in vivo* gene therapy. Subretinal injection of Cas9 RNP nanocapsules, composed of anionic and cationic monomers along with PEG, facilitated genome editing in the retinal pigment epithelium (RPE) and was further enhanced when incorporated with all-*trans* retinoic acid (ATRA), which interacts with the interphotoreceptor retinoid-binding protein on RPE[139]. These Cas9 nanocapsules were also active after intramuscular injection into the tibialis anterior. CRISPR-Gold has also been used for *in vivo* genome editing via intramuscular injection[137]. Co-packaging Cas9 RNP and DNA HDR template, CRISPR-Gold was injected into the gastrocnemius and tibialis anterior muscles of *mdx* mice, which model Duchenne muscular dystrophy. This delivery strategy resulted in 2-6% HDR, restoration of dystrophin expression, improved muscle function, no detectable off-target editing, and no immunogenicity. Additionally, porous silica nanoparticles were shown to efficiently load and release Cas9 RNP[140]. Direct injection of these Cas9 nanoparticles into the tibialis anterior allowed for specific muscle targeting, where 20-30% editing efficacy of the myostatin gene, *MSTN,* was observed along with improvement in muscle function and mass.

1.4.1.3 Targeted genome editing *in vivo* via systemic delivery

While targeting specific tissues via direct mechanical injection has been utilized to introduce Cas9 RNPs, it also presents multiple challenges and limitations. Namely, this method often involves specialized equipment such as stereotaxic injectors, and the procedures are highly invasive. Systemic delivery, where therapeutics are administered via routine and minimally invasive injection into the circulatory system, avoids these limitations of direct delivery.

In order to achieve targeted genome editing after systemic delivery, researchers have relied on passive and active mechanisms. Passive targeting relies on the body’s natural tendency to accumulate injected molecules in tissues like the liver and spleen. For example, gold nanoparticles containing Cas9 RNPs injected into mice via the tail vein accumulated 10-100-fold greater in the liver and spleen compared to other organs such as the brain, lung, heart, intestine, and kidney[141]. Within the liver and spleen it was observed that macrophages were selectively targeted and edited compared to T or B cells.

Solid tumors have also been the target of genome editing using both passive and active methods of systemic delivery. The enhanced permeability and retention effect (EPR) has long been hypothesized to allow macromolecules to passively accumulate within the tumor microenvironment after systemic injection[142]. PEG-coated nanoparticles and gold nanorods containing Cas9 RNPs have exploited this phenomenon, showing preferential accumulation and retention in the tumor, even over the liver, via whole body *in vivo* imaging after 48 hours[143,144]. In either case, tumor targeting was further achieved by constructing the delivery vehicles with acid-degradable or hypoxia-responsive linkers. The acidic and hypoxic environment of the tumor allowed for selective release of Cas9 RNPs and *in vivo* genome editing. The overexpression of surface molecules on cancerous cells has also been utilized for active targeting of Cas9 RNPs. Nanoparticles encapsulating Cas9 RNPs were decorated with hyaluronic acid, which binds CD44 overexpressed in certain melanoma, colorectal, breast, and lung cancers[145]. These targeted Cas9 RNP nanoparticles lead to reduction in tumor volume in both xenograft and lung metastasis models after targeting *KRAS*. Active targeting with a molecular and physical basis have also been combined. NIR-responsive PEG nanoparticles were decorated with iRGD, a tumor homing peptide[146]. Inclusion of this tumor-targeting peptide increased *in vitro* Cas9 RNP genome editing 2-fold compared to non-targeted particles. After targeting the tumor via intravenous injection of iRGD nanoparticles, NIR was applied to the tumor site to induce release of Cas9 RNPs targeting *NRF2*, resulting in decreased tumor volume and prolonged survival. This multiplexed method for targeting, using molecular ligands and NIR-mediated release, displayed highly selective genome editing of the tumor over other organs such as the liver.

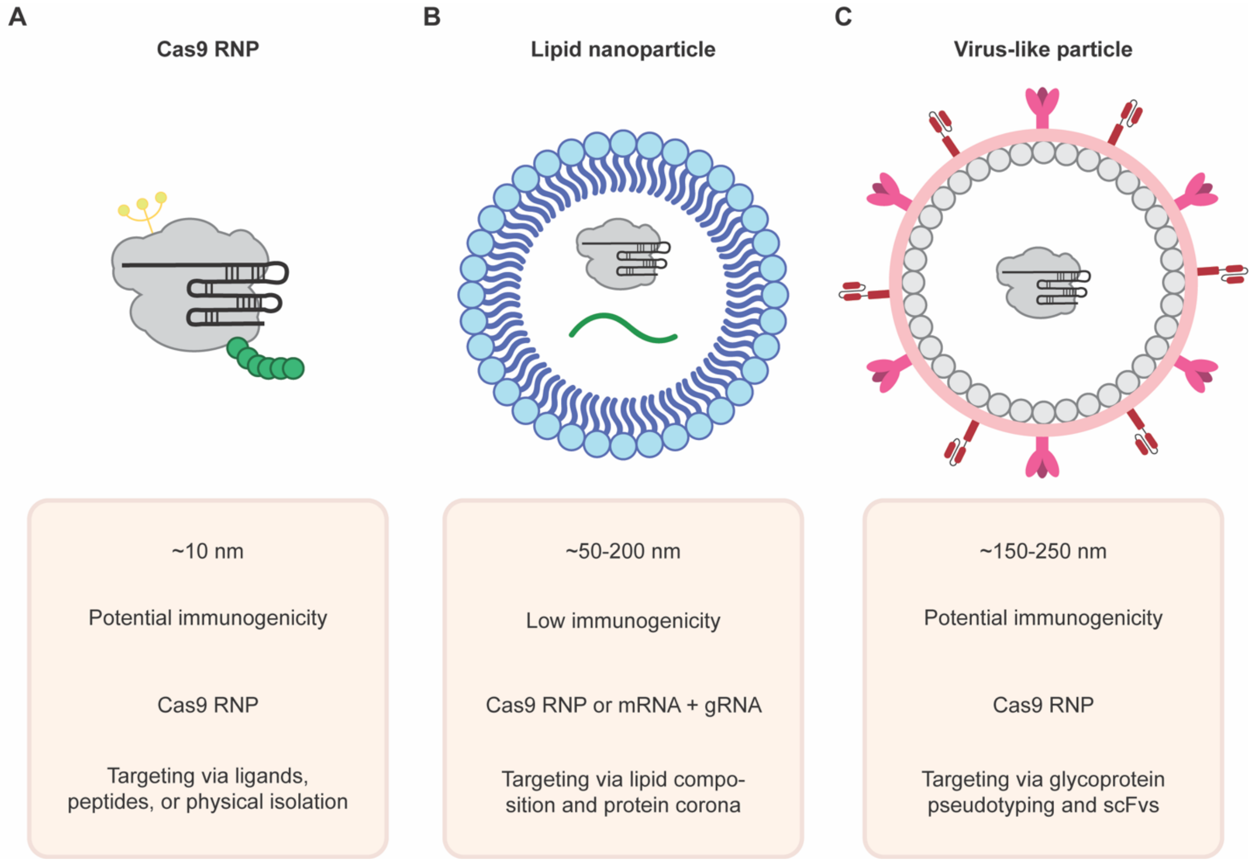


Figure 1.4: Novel delivery strategies for transient and targeted Cas9 genome editing. **(A)** Cas9 RNP can be directly fused to ligands and peptides to target specific cell types or cells can be physically isolated prior to delivery. **(B)** Lipid nanoparticles can package Cas9 RNP or mRNA encoding Cas9 along with its gRNA. Organ-specific targeting can be engineered by adapting the lipid composition to recruit different protein coronas. **(C)** Virus-like particles are membranous particles that can package Cas9 RNP. Cell-specific tropism can be programmed through glycoproteins and/or scFvs on the membrane surface.

1.4.2 Targeted delivery with synthetic lipid nanoparticles

Lipid nanoparticles (LNP) are composed of amphiphilic lipids formed into spherical micelles or liposomes. The major advantage to LNPs is their high degree of controllability; the lipid composition can be finely tuned to vary the property of the LNP or to package different cargo[147,148]. LNPs are also minimally toxic and able to be re-dosed if necessary. One practical advantage of LNPs is that their chemical synthesis translates well to larger-scale manufacturing. This is a major benefit over cell-derived vehicles such as exosomes and viruses, where the cost of cell culture media, sterility, and bio-reactors results in immensely expensive therapies[122]. The COVID-19 pandemic provided proof-of-principle for the rapid and affordable generation of safe and efficacious LNPs, as the mRNA vaccines against SARS-CoV-2 developed by Pfizer-BioNTech and Modern utilized LNPs[149–151]. Now one of the most widely administered therapeutics in the world, much of the concern over safety and efficacy of LNPs has been addressed. The bigger challenge for LNPs going forward will be cell-type or tissue specificity. LNPs are a blank canvas, without any inherent tropism, significant engineering is required to target specific cell types or organs.

1.4.2.1 Passive and active targeting of the liver

Among its many functions, the liver is involved in clearing both endogenous and exogenous particles from the bloodstream. Highly fenestrated endothelium (150-200 nm), the cell type comprising the inner layer of blood vessels, means that extravasation and accumulation of particles is high within the liver[108]. LNPs are no exception and have a bias toward passively accumulating in the liver without additional targeting mechanisms. Targeting of hepatocytes within the liver with LNPs is thought to occur from a protein corona mechanism[148]. Apolipoprotein E (APOE), which naturally transports cholesterol in the bloodstream, coats LNPs in the circulation before binding with the low-density lipoprotein receptor (LDL-R) expressed on hepatocytes. After receptor-mediated endocytosis, ionizable lipids in the LNP become cationic during the acidification, resulting in endosomal escape and cargo delivery into the cytoplasm[152].

CRISPR-Cas genome editing of the liver has routinely been some of the earliest demonstrations of *in vivo* genome editing, largely because this passive mechanism circumvents many challenges with targeted delivery. LNPs can be loaded with mRNA encoding Cas9 as well as a gRNA; this strategy has mediated efficient genome editing for hypercholesteremia[153] as well as hemophilia A and B in mouse models[154]. In both cases, several combinations of lipids comprising the LNP were screened, biodistribution after systemic delivery was highly localized to the liver, and *in vivo* editing rates were ~20-40%. ABEs encoded as mRNA have similarly been delivered via LNP into non-human primates to target *PCSK9* with the aim of lowering LDL[111,155]. Again, systemically administered LNPs accumulated in the liver to mediate robust base editing. In non-human primates, large doses and re-dosing of the LNP were well tolerated without adverse effect[155]. LNPs were tested in the first systemic, *in vivo* CRISPR-Cas genome editing human clinical trial when Intellia Therapeutics delivered mRNA encoding Cas9 to treat transthyretin amyloidosis, which has shown tremendous safety and efficacy thus far[109,156].

Along with the passive mechanism that leads to liver targeting with LNPs, additional methods for active targeting of the liver have further increased efficacy and specificity. Most prominently, N-acetylgalactosamine (GalNAc) has been functionalized onto the surface of LNPs to target the asialoglycoprotein receptor that is highly expressed and frequently recycled on the hepatocyte surface[148]. GalNAc targeting has been well studied[157], and a small interfering RNA (siRNA) conjugated to GalNAc was FDA approved in 2019 to treat acute hepatic porphyria[158]. This success has been translated to genome editing, where LNPs were functionalized with GalNAc to deliver ABE mRNA to the liver of non-human primates with homozygous familial hypercholesteremia (HoFH)[159]. HoFH is characterized be a loss of LDL-R, the main target of LNPs via the passive protein corona mechanism, resulting in unfunctionalized LNPs mediating minimal editing in the liver. However, active targeting of the liver with GalNAc functionaized LNPs resulted in 60% editing of the entire liver.

1.4.2.1 Organ-specific delivery beyond the liver

Naturally, most LNPs heavily target the liver with some targeting of the spleen. In order to achieve targeting of organs other than the liver, substantial engineering of the LNP needs to take place. Thus far, beyond the liver, *in vivo* targeting of the lungs and spleen have been demonstrated for CRISPR-Cas genome editing.

The most promising method for targeting extrahepatic organs involves altering the chemical composition of an LNP so that different proteins, other than the liver biased APOE, coat the particle surface and mediate uptake by other organs. This method, termed selective organ targeting (SORT), has been successfully applied to target Cas9 mRNA or RNP to the lungs or spleen[160–162]. For lung targeting, an addition quaternary ammonium headgroup-containing lipid was included in the LNP formulation. This headgroup recruits vitronectin to coat the LNP, which can interact with the αVβ3 integrin receptor expressed on the lungs[163]. Introducing the anionic phosphatidic acid into LNPs resulted in a protein corona enriched in β2-glycoprotein I and genome editing of T and B cells specifically in the spleen[163].

Robust LNPs specific to extrahepatic organs other than the lung and spleen have yet to be developed. However, the ease of manufacturing LNPs has enabled the high throughput generation of unique nanoparticles for *in vivo* biodistribution tests[164,165]. Generating libraries of LNPs with different chemical compositions may identify formulations with unique organ targeting preferences beyond the liver, lung, and spleen.

1.4.3 Targeted delivery with membranous delivery vehicles

Viral delivery vehicles, like lentivirus or AAV, or highly advantageous for the delivery of genome editors because years of evolution have endowed the ability to target specific cell types and deliver cargo into the intracellular cytoplasm. However, the long-lasting nature of genetically encoding nucleases like Cas9 creates a safety concern over off-target editing. On the other end of the spectrum, LNPs can package transient mRNA or RNP yet have no intrinsic ability to target specific cell types or escape the endosome into the cytoplasm. Recently, a new class of membranous delivery vehicles has sought to combine the advantages of both these methods. Both exosomes and virus-like particles (VLP) utilize the membranes of mammalian cells to package CRISPR-Cas effectors as mRNA or RNP and have begun to show promise for *in vivo* targeted genome editing.

1.4.3.1 Engineered virus-like particles for transient genome editing

Just as prokaryotic viruses have coevolved with microbes, starting the arms race that necessitated CRISPR systems for immunity, eukaryotic viruses have also coevolved with humans. Over these years, viruses have evolved complex and efficient mechanisms to target specific cell types, gain entry into the cell, and delivery their genome in order to replicate within the host. Enveloped viruses are a class of viruses that bud from the plasma membrane of the infected cell, packaging their genome along with structural and accessory proteins. On the outside of these membranous viruses are the tropism-specifying glycoproteins encoded within the viral genome and any other cell-encoded membrane proteins that were already present during budding. Harnessing the engineering nature has already done, enveloped viruses have been co-opted as effective delivery vehicles for genome editing. By replacing the viral genome with a transgene of interest, virus-like particles (VLP) can no longer replicate within cells but can still deliver nucleic acids or proteins of interest.

The first demonstration of Cas9 delivery via VLP came from the serendipitous finding that expressing Cas9 RNP within cells also producing lentivirus, a type of enveloped virus, resulted in the spontaneous encapsulation of the RNP within budding particles[166]. To promote the active loading of budding VLPs with genome editing molecules, several approaches have been explored. Aptamer binding proteins have been fused to the lentiviral Gag polyprotein, which includes structural proteins that associate with the inner plasma membrane before viral budding. By encoding the corresponding aptamer sequence in mRNA encoding Cas9 or within the gRNA of a Cas9 RNP, active loading of VLPs has been achieved[167,168]. Cas9 RNP or base editor RNP has also been directly fused to Gag, relying on natural proteolytic cleavage to be freed within the VLP[169,170]. This strategy has proven highly efficient for the delivery and genome editing of primary T cells to generate CAR T cells[171], along with the liver, brain, and eye[169,170].

One major delivery advantage for VLPs is that their cell-type specificity is wholly dictated by the glycoprotein displayed on the outside of the host cell membrane and eventually on the budded envelope of the VLP[171]. Enveloped viruses can be pseudotyped with exogenous glycoproteins with different cell specific tropisms. We showed that Cas9-containing VLPs could be pseudotyped for CD4 specific delivery by expressing the HIV-1 *Env* within the budding producer cells[171]. HIV-1 is specific for CD4+ T cells and macrophages; this same cell-type specificity was endowed in VLPs for the delivery of Cas9 RNP. This work has been further expanded for more modular cell-type specific programming. VLPs expressing a mutated VSV-G glycoprotein lose the ability to bind to their natural receptor, LDL-R, but retain the ability to release their cargo into cells once bound[172]. Co-expression of an antibody-derived single chain variable fragment (scFv), along with the mutated VSV-G glycoprotein, generated engineered VLPs specific for a user-defined antigen[173]. Using this method, VLPs could be modularly adapted to target cells expressing the immune cell markers CD19, CD20, CD4, or CD28. This approach holds tremendous potential toward being a facile approach to deliver transient Cas9 RNP with programmable single antigen specificity.

1.4.3.2 Exosomes mediate low toxicity targeting

Exosomes are membrane-bound extracellular vesicles that naturally bud from eukaryotic cells. While exosomes are normally responsible for extracellular trafficking and signaling, they are also released by cells in *ex vivo* cell culture, where they can be harvested in the cell culture medium supernatant. Advantages to exosomes as a delivery vehicle are their large packaging capacity, low immunogenicity, biocompatibility, and stability in the bloodstream[174]. Expression of Cas9 RNP within cell lines results in spontaneous encapsulation in budding exosomes[175], though active loading mechanisms can improve this packaging. Expression of an exosomal protein like CD63 fused to an epitope, along with Cas9 RNP fused to an epitope-binding domain resulted in increased packaging into budding exosomes, which mediated genome editing once these harvest exosomes were mixed with target cells[175,176].

Exosomes packaging CRISPR-Cas effectors have also been developed in a more synthetic fashion for cell- and organ-specific targeting. Purified exosomes derived from hepatic stellate cells have been electroporated to load Cas9 RNP, similar to how electroporating can transiently allow Cas9 RNP to enter the phospholipid bilayer of mammalian cells[177]. Tail vein injection in mice showed preferential accumulation in the liver, negligible toxicity, and genome editing of disease specific loci. Another synthetic exosome-like delivery strategy utilized Cas9 RNPs packaged within metal organic frameworks (MOF) – highly ordered and porous 3D structures[178]. Cell membranes were extracted via lysis, and Cas9-MOFs were coated in this cell-derived lipid membrane via vortexing and extrusion[179]. The resulting membrane encapsulated Cas9 RNP MOF showed biased uptake by target cells of the same origin as the cell lines used for membrane lysis. Additionally, Cas9-MOFs coated in membranes from MCF cancerous cells showed targeting to implanted MCF tumors in mice after intravenous injection, whereas Cas9-MOFs coated in non-cancerous membranes showed minimal *in vivo* tumor targeting.

1.5 Precision and accuracy of CRISPR-Cas genome editing

Traditional small molecule therapies such as antibiotics, biologics like antibodies or recombinant protein, and emerging nucleic acid therapies like siRNA or antisense oligonucleotides, provide treatment only for a finite period. CRISPR-Cas genome editing, on the other hand, can make permanent changes to a patient's genome that will remain for the lifetime of the cell, tissue, or organ. While this has the benefit of potentially being a single dose cure, it comes with the added risk of creating permanent genomic errors during the process.

1.5.1 Accuracy of CRISPR-Cas genome editing

The accuracy of Cas9 during genome editing refers to the ability to bind and cleave an intended target region of the genome. Cas9 relies on a 20 nt spacer sequence on the 5' end of its gRNA as the sole determinant of the RNP's target sequence[180]. While a perfect genomic DNA match to this spacer sequence in the gRNA is often considered a requirement for Cas9 cleavage, a certain degree of promiscuity exists. In fact, along with editing at the intended "on-target" genomic sequence, it is now well characterized that editing at unintended "off-target" genomic sequences can also occur[181].

Along with the protospacer target sequence in the genomic DNA, CRISPR-Cas systems also require the presence of a protospacer adjacent motif (PAM). This PAM naturally allows Cas effectors to distinguish self (the CRISPR array contains the 20 nt spacer target sequence but no PAM) from non-self (an invading foreign nucleic acid contains a 20 nt protospacer target sequence directly next to the PAM) and prevent autoimmunity[182]. For the widely used Cas9 from *S. pyogenes* this genomic motif is adjacent to the 3' end of the protospacer sequence and is 5'-NGG-3', where N is any nucleotide. While this PAM requirement is tightly regulated, several studies have found that Cas9 can tolerate 5'-NAG-3' to a limited extent[183]. In addition, the requirement for a perfect 20 bp match between the gRNA spacer and DNA target sequence is not always strictly required. After making systematic mismatches in the gRNA sequence relative to the DNA target sequence, Cas9-induced indels were still observed both *in vitro* and in human cell culture[183–185]. While wide variance is observed between different gRNAs and their corresponding target sequences, some patterns in the promiscuity of individual RNA:DNA mismatches have emerged. A seed region, approximately 5-10 nt upstream from the 3’ end of the spacer sequence (immediately next to the PAM) is less tolerant of mismatches with the target sequence[183,184]. Outside of this seed region, RNA:DNA mismatches may be tolerated, sometimes with cleavage efficiencies equivalent to perfectly matched sequences, and seems to correlate with distance from the PAM. That is, the closer the nucleotide is to the 5’ end of the spacer, the more amenable it is to mismatches.

The promiscuity of Cas9 to target addition sequences other than those with perfect complementarity to its gRNA raised concerns that cleavage may occur at unintended off-target sites in the genome which contain partial complementarity to the gRNA. Indeed, it has been demonstrated numerous times and is widely accepted that Cas9 can cleave at off-target sites in the genome, usually with high but imperfect complementarity (Figure 1.5A)[65,183]. The advent of more sensitive next generation sequencing methods has also improved the ability to detect rare but bonified off-target editing by Cas9.

1.5.1.1 Methods for off-target detection during CRISPR-Cas genome editing

Since it was first discovered that Cas9 could cleave at unintended regions in the genome, several methods for predicting these off-target sites have been developed. An initial algorithm to score on-target and off-target Cas9 editing efficacy was developed, though it was based on a limited set of experimental data and targets[183]. More recently, high-throughput CRISPR screens using thousands of gRNAs and targets along with machine learning have generative improved predictive models for both on-target and off-target Cas9 activity[186,187]. Now, multiple web-based servers exist that compile these algorithms to allow users to generate *in silico* predictions of on-target efficacy and off-target specificity[188]. Users can input long sequences of a target gene and generate rank ordered lists of potential gRNAs based on specificity, with potential genome-wide off-targets annotated[188].

Along with *in silico* predictions, experimental methods for identifying possible off-target sites in the genome have been developed both *in vitro* and in human cells. CIRCLE-seq uses purified, sheared, and circularized genomic DNA mixed with Cas9 *in vitro*[189]. Cleaved genomic DNA circles have sequencing adaptors ligated to the free ends, which enables sequencing and identification of all cleaved sequences. The use of cell free DNA and the ability to use a high concentration of Cas9 *in vitro*, minimizes the cost-associated with this off-target detection and allows for the detection of rare off-targets. However, other cellular factors influence Cas9’s ability to cleave the genome within a cell, including chromatin accessibility[190,191]. Therefore, *in vitro* CIRCLE-seq predictions are likely an overestimation of potential off-target sites. To address this, cell-based methods for off-target predictions have also been developed, including GUIDE-seq, which relies on the insertion of a short tag within a DSB site in the cell[192]. More recently, methods for identifying *in vivo* off-targets within animal models have been developed. DISCOVER-seq uses chromatin-immunoprecipitation sequencing (ChIP-seq) to identify sequences bound by MRE11, a DNA repair protein associated with Cas9-induced DSBs. This method is compatible with *in vivo* Cas9 delivery and can faithfully identify off-targets in cells and in animal models.

To date, several off-target prediction methods have been developed, both *in silico, in vitro,* in cells, and *in vivo*. Thus far, there is still no method that can predict Cas9 off-target sites with perfect specificity and sensitivity. Because of this, researchers often use multiple approaches, prioritizing the investigation of potential off-target sites that are identified through multiple means.

1.5.1.2 Engineered high-fidelity CRISPR-Cas systems

The natural promiscuity of Cas9 and its gRNA has spurred a large effort toward engineering enhanced specificity. Once high-throughput methods were developed to detect off-target edits, engineered Cas9 systems could be compared to the wildtype in terms of on-target efficacy and off-target specificity.

The Cas9 gRNA has been the focus of a limited number of studies hoping to mitigate off-target editing. Rather than a 20 nt spacer, shorter spacer sequences of ~17-18 nt reduced off-target editing at numerous target sites[193]. While shorter spacers may reduce off-targets by increasing the requirement for correct base pairing over a shorter sequence length, new off-target sites also arose because of the shorter stretch of complementarity. Alternatively, gRNA spacers have been extended in length on the 5' end to form short hairpin structures, which resulted in decreased off-target editing[194]. Using ChIP-seq with dCas9, it was found that hairpin-containing gRNAs bound DNA at a similar rate to normal gRNAs, meaning reduced cleavage was the main contributor to improved specificity. Another strategy to engineer the gRNA for increased specificity is to substitute RNA bases with DNA bases in a chimeric RNA-DNA guide. Replacing the first 10 nt on the 5' end of the gRNA spacer with DNA bases reduced off-target cleavage by Cas9 while preserving on-target cleavage[195,196].

A larger body of effort has been placed on protein engineering, both structure-guided and directed evolution, of Cas9 for increased specificity. Structures of *S. pyogenes* Cas9 have been exploited by rationally mutating nucleic acid interacting residues. Three separate efforts developing eSpCas9(1.1)[197], SpCas9-HF1[198], and HyphaCas9[199] were all shown to improve specificity through amino acid substitutions that raise the threshold for a conformational domain rearrangement necessary for catalytic activity, not by decreasing the overall binding activity as originally thought. In order to probe the full sequence space of Cas9 for mutations that may improve specificity, high throughput directed evolution in *E. coli* and yeast has been performed. *E. coli* were engineered to contain a toxic plasmid as well as potential off-target sequences within the genome. Cas9 editing of the toxic plasmid (positive selection) is required for cell survival, while cleavage of the off-target sequences in the genome will lead to cell death (negative selection). Screening a library of 10^7 Cas9 variants yielded Sniper-Cas9, which preserves on-target efficacy with improved off-target specificity[200]. A similar, but colorimetric strategy was employed in yeast. Engineered on-target and off-target sequences in the yeast genome yielded logic gating that meant no cutting of either sequence would be toxic for the cell, cutting of both sequences would yield white colonies, and cutting of the on-target but not the off-target sequence would yield red colonies; this method yielded evoCas9, which again displayed improved specificity[201]. Recently, a large effort to compare these high-fidelity Cas9 variants was conducted at thousands of target sites[202]. The overall activity ranked SpCas9 ≥ Sniper-Cas9 > eSpCas9(1.1) > SpCas9-HF1 > HypaCas9 >>

evoCas9, while the overall specificity ranked evoCas9 >> HypaCas9 ≥ SpCas9-HF1 ≈ eSpCas9(1.1) > Sniper-Cas9 > SpCas9[202]. These near perfectly opposite trends show that cleavage at on-target and off-target sites is inherently linked, and that improving or reducing one will likely improve or reduce the other.

1.5.2 Precision of CRISPR-Cas genome editing

The precision of Cas9 during genome editing refers to reproducibility of the genomic outcomes after cleavage at the intended on-target site. Cas9-induced DSBs, which generate a blunt breakage 3 bp upstream from the PAM, have been shown to result in predictable indels based on the target sequence. Empirical testing of thousands of gRNAs showed that small deletions and 1 bp insertions were the most common indels, and that the base 4 bp upstream from the PAM was most likely to be repeated as the 1 bp insertion[203,204]. This data was used to train a machine learning algorithm, called inDelphi, which is able to predict the indel spectrum based on a given Cas9 target sequence with high accuracy[203]. While the genomic outcomes involving indels are highly predictable after Cas9 cleavage, rare and unintended chromosomal abnormalities have recently been characterized at on-target sites. These potential genotoxicities are an area of concern for genome editing, as activating, inactivating, or forming a new gene product could affect cell phenotype or potentially be oncogenic.

1.5.2.1 Chromosomal rearrangement after Cas9-induced DSBs

DNA translocations occur when a cleaved portion of a chromosome is reattached to a different chromosome. While it was originally thought that homology between sites on different chromosomes may drive translocations, end-joining DNA repair mechanisms such as NHEJ and MMEJ were eventually found to be involved[205]. In a seminal study, it was shown that cleavage on two different chromosomes with an I-*Sce*I meganuclease induced significant rates of translocation between the two[206]. Predictably, using Cas9 and with multiple gRNAs targeted to different parts of the genome can result in precise translocations between the two cleavage sites (Figure 1.5B). This method was originally exploited to generate models of cancers that are driven by chromosomal translocations[207,208]. However, translocations have also been observed in a therapeutic context; genome editing with multiple gRNAs can drive translocation frequencies as high as ~1% in cells[209,210]. In fact, in a first-in-human phase I clinical trial involving three Cas9 gRNAs delivered simultaneously into T cells, several variations of translocations were observed in ~1-5% of cells, though this rate decreased over time *in vivo*, suggesting a possible fitness advantage within patients[101]. Several methods for mitigating chromosomal translocations have been devised, including using non-DSB-inducing base editors[103] and doing multiple Cas9 edits in series rather than in parallel[211].

Recently chromothripsis, an extensive rearrangement of sequence on a given chromosome, was identified as a consequence of Cas9 DSBs (Figurer 1.5C)[212]. Cas9 cleavage resulted in aberrant micronuclei and/or chromosome bridge formation, which can drive chromothripsis. Similar to translocations, Cas9 inducing this genotoxicity raises concern over changes to the cellular phenotype as well as cancerous transformation.

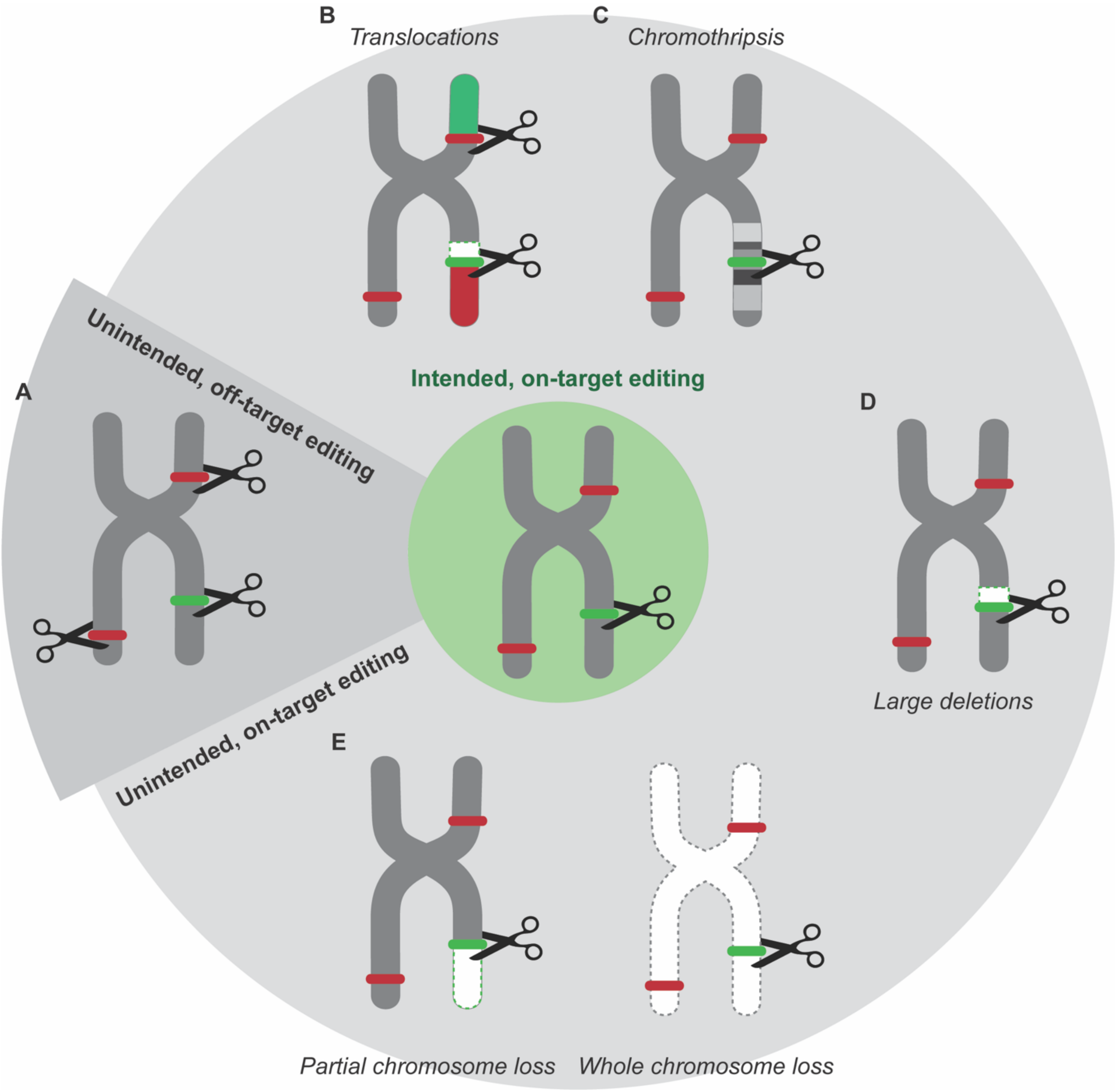


Figure 1.5: Precision and accuracy of CRISPR-Cas9 genome editing. Green lines represent intended, on-target sites; red lines represent unintended, off-target sites; and scissors represent Cas9 cleavage. **(A)** Cas9 can cleave unintended, off-target sites containing partial sequence complementarity with the gRNA spacer. **(B)** Translocations between two or more portions of the genome can occur during multiplexed Cas9 genome editing. **(C)** Cas9 cleavage can cause chromothripsis, or complex genomic rearrangement, at the intended target site. Cas9 cleavage can also result in **(D)** large deletions, as well as **(E)** partial or whole chromosome loss, initiated at the intended target site.

1.5.2.2 Large deletions and chromosome loss

Although small insertions and deletions (indels) are the most common consequence of Cas9 DSBs, numerous studies have now identified the presence of much longer deletions and in some cases the loss of entire chromosomes. Since next generation sequencing methods utilize short amplicons, large deletions traditionally went undetected. However alternative methods such as fluorescence *in situ* hybridization (FISH), droplet digital PCR (ddPCR), long read sequencing, and scRNA-seq have enabled a better quantification of these large deletions (Figure 1.5D) and chromosomal loss events (Figure 1.5E). Long read sequencing and PCR-free whole genome sequencing revealed kilobase-long deletions attributed to Cas9 cleavage[213,214]. Even larger, megabase scale deletions originating at the Cas9 target site and continuing past the telomere were also identified with FISH, ddPCR, and scRNA-seq[215,216]. Finally, we (discussed in Chapter 3) and others have found that rare but detectable loss of a whole chromosome can occur after cleavage by a single Cas9 effector[217–219]. To date, these potential genotoxicities have not caused a deleterious cell phenotype in research studies or clinical trials, however continued investigation is necessary given the permanence of genome editing.

CHAPTER 2

Structural-guided engineering of CRISPR-CasX for improved genome editing efficacy

A portion of the work presented in this chapter has been published previously as part of the following paper: Tsuchida C.A.*, Zhang S.*, Saffari Doost M.*, Zhao Y.*, Wang J., O'Brien E., Fang H., Li C-P., Li D., Hai Z-Y., Chuck J., Brötzmann J., Vartoumian A., Burstein D., Chen X-W., Nogales E., Doudna J.A., Liu J-J.G. Chimeric CRISPR-CasX enzymes and guide RNAs for improved genome editing activity. *Mol Cell* **82**, 1199–1209, (2022).

*These authors contributed equally.

2.1 Abstract

A compact protein with a size of <1,000 amino acids, the CRISPR-associated protein CasX is a fundamentally distinct RNA-guided nuclease compared to Cas9 and Cas12a. Although it can induce RNA-guided genome editing in mammalian cells, the activity of CasX is less robust than that of the widely used *S. pyogenes* Cas9. Here, we show that structural features of two CasX homologues and their guide RNAs affect the R-loop complex assembly and DNA cleavage activity. Cryo-EM-based structural engineering of either the CasX protein or the guide RNA produced two new CasX genome editors (DpbCasX-R3-v2 and PlmCasX-R1-v2) with significantly improved DNA manipulation efficacy. These results advance both the mechanistic understanding of CasX and its application as a genome editing tool.

2.2 Introduction

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins comprise adaptive immune systems used by prokaryotes and some giant phages to fight against invading nucleic acids[5,220]. The entire immune response is typically composed of three steps: integration of fragments from invading nucleic acids, synthesis of a ribonucleoprotein (RNP) interference complex, and nucleic acid interference[2,221]. During the last step of nucleic acid interference, a Cas protein is guided by its CRISPR RNA (crRNA), which is synthesized from the CRISPR array, to cleave a complementary DNA or RNA target. The programmability of CRISPR systems thus holds tremendous potential as transformative tools for genome editing[43,222,223]. After years of effort, only a few types of CRISPR-Cas nucleases have been widely used for efficient genome editing, such as Cas9 and Cas12a[6,18,180]. While efficient for genome editing, the large size of Cas9 and Cas12a (1,000–1,500 amino acids [aa]) precludes their ability to be delivered via an adeno-associated virus (AAV), which is useful for therapeutic delivery but has a limited transgene size of just 4.7 kilobase pairs (kbp).

A subtype of compact CRISPR nucleases, CasX (type V Cas12e, <1000 aa) has two homologous systems, CasX from *Deltaproteobacteria* (hereafter DpbCasX) and CasX from *Planctomycetes* (hereafter PlmCasX), that share 68.5% sequence similarity and expand the CRISPR-Cas genome editing family by offering a class of smaller, programmable nucleases as additional therapeutic options[21,224,225]. Compared to Cas9 or Cas12a, CasX is small enough to be delivered via a single AAV, with additional room for multiplexed single guide RNAs (sgRNA) or protein domain fusions[29,226]. Previous biochemical analysis showed that DpbCasX cleaves double-stranded DNA (dsDNA) with a protospacer adjacent motif (PAM) of 5’-TTCN[21,29]. Structural analysis further showed that DpbCasX cuts the non-target strand (NTS) DNA and target strand (TS) DNA sequentially, using a single nuclease active site with the help of a large sgRNA scaffold (hereafter sgRNAv1)[29]. Though DpbCasX is highly effective for bacterial interference, the genome editing activity in mammalian cells is modest relative to the widely used *S. pyogenes* Cas9. PlmCasX, although not well explored *in vitro* due to difficulty in protein expression and purification, showed equivalent or sometimes greater genome editing activity in mammalian cells compared to DpbCasX[29]. Therefore, we aimed to determine the biochemical and structural mechanism of DNA cleavage by PlmCasX and further improve the genome editing capacity of CasX nucleases by structure-based engineering.

In this study, we expressed and purified the PlmCasX protein with similar quality as DpbCasX via an improved workflow. While PlmCasX showed minimal dsDNA cleavage *in vitro*, consistent with our previous observation, PlmCasX efficiently disrupted GFP expression in a HEK293 fluorescent reporter cell assay at a similar or even higher rate compared to DpbCasX[29]. Cryo-EM studies of the dPlmCasX-sgRNAv1-dsDNA ternary complex identified three distinct conformational states, including one that displays high flexibility of the Helical-II domain. The existence of this dynamic state suggests that the Helical-II domain assists with assembly of the ternary (R-loop) complex and ensures effective dsDNA cleavage via direct interaction with the sgRNA scaffold stem. Structural comparison of DpbCasX and PlmCasX suggests that three nucleotide-binding loops within CasX may play beneficial roles for PAM-proximal region recognition, sgRNA interaction, and DNA substrate loading, which may contribute to the different biochemical and mammalian cell DNA cleavage efficacies between the two systems. Chimeric versions of CasX containing those beneficial loops showed improved DNA cleavage activity *in vitro*.

Further, by rational sgRNA design based on new structural information, we improved the genome editing activities of both DpbCasX and PlmCasX using a sgRNA we have termed sgRNAv2. With synergetic improvement to both the protein and sgRNA, the new CasX nucleases (DpbCasX-R3-v2 and PlmCasX-R1-v2) showed 10-fold and 20-fold improvement in biochemical dsDNA cleavage kinetics, and 53% and 78% median editing efficacy (2 to 3-fold improvement) for ten different GFP-targeting sgRNAs within human cells, respectively. In summary, these results yield fundamental knowledge and a practical improvement of CasX nucleases. Given the compact protein size of less than 1000 amino acids and the unique domain architecture relative to other Cas nucleases, CasX nucleases offer substantial advantages that expand the genome editing toolbox.

2.3 Results

2.3.1 PlmCasX shows minimal biochemical activity but functions robustly in mammalian cells

We used an improved protocol (see Method Details) to purify wild-type (wt) PlmCasX with similar purity and yield as wtDpbCasX (Figures 2.1A and 2.1B). PlmCasX eluted 0.3 mL earlier via size exclusion chromatography (Figure 2.1A), which suggests apo-PlmCasX (112.66 kDa) is less compact than apo-DpbCasX (112.93 kDa) and may lead to the increased difficulty observed during expression and purification. *In vitro*, PlmCasX cleaved just 10% of both the NTS and TS DNA (Figures 2.2A and 2.2B; Figures 2.1C and 2.1D) compared to DpbCasX with the previously reported sgRNA scaffold, sgRNAv1[29]. However, DpbCasX and PlmCasX showed similar linearization activity on pUC19 (Figure 2.1E), which may be due to the supercoiling-induced denaturation bubbles within plasmids[227].

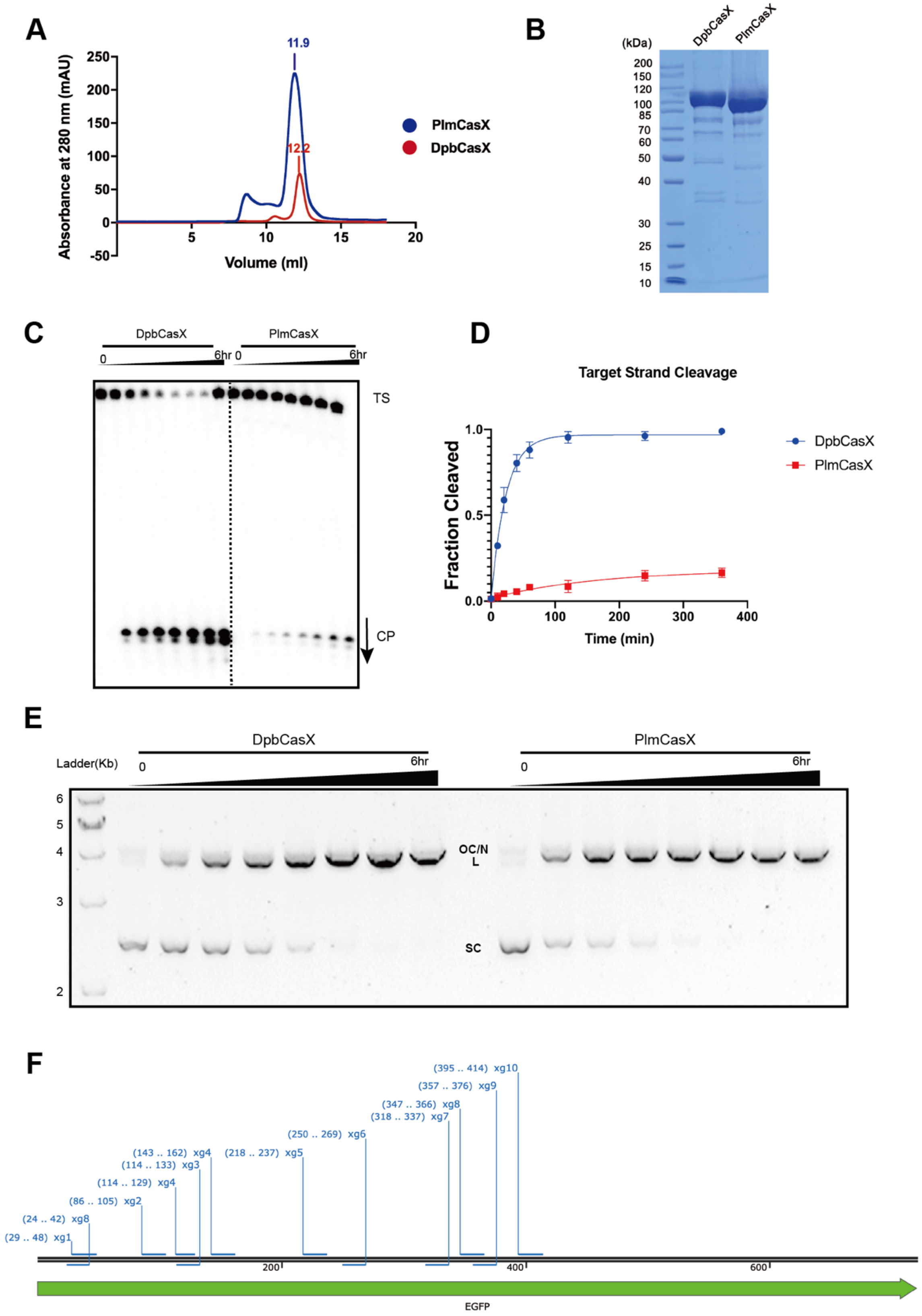

A
Absorbance at 280 nm (mAU)
250
200
150
100
50
0
-50
5
10
15
20
Volume (ml)
11.9
12.2
PlmCasX
DpbCasX
B
(kDa)
DpbCasX
PlmCasX
200
150
120
100
85
70
60
50
40
30
25
15
10
C
DpbCasX
PlmCasX
0
6hr
0
6hr
TS
CP
D
Target Strand Cleavage
Fraction Cleaved
1.0
0.5
0.0
0
100
200
300
400
Time (min)
DpbCasX
PlmCasX
E
Ladder(Kb)
DpbCasX
PlmCasX
0
6hr
0
6hr
6
5
4
3
2
OC/NL
SC
F
(395 .. 414) xg10
(357 .. 376) xg9
(347 .. 366) xg8
(318 .. 337) xg7
(250 .. 269) xg6
(143 .. 162) xg4
(218 .. 237) xg5
(114 .. 133) xg3
(114 .. 129) xg4
(86 .. 105) xg2
(24 .. 42) xg8
(29 .. 48) xg1
200
400
600
EGFP

Figure 2.1: Targeted DNA cleavage by DpbCasX and PlmCasX. **(A)** Purification of DpbCasX and PlmCasX by size exclusion chromatography. The representative S200 size exclusion traces by UV280 absorbance are aligned and shown. DpbCasX eluted at 12.2 mL and PlmCasX eluted at 11.9 mL. **(B)** SDS-PAGE for DpbCasX and PlmCasX samples taken from the elution peak after size exclusion chromatography. **(C)** *In vitro* dsDNA cleavage activity comparison between DpbCasX and PlmCasX revealed by denaturing PAGE. TS denotes the target strand DNA which was ^{32}P labeled on the 5' end. CP indicates the cleavage product. The fractions were collected at 0 min, 5 mins, 10 mins, 20 mins, 40 mins, 1 hr, 2 hrs, 4 hrs and 6 hrs, respectively. **(D)** Plot of DNA cleavage kinetics based on the TS band density from the cleavage fractions compared to the input TS band density at the reaction time of 0 min (n = 3, mean ± SD). The rate constant k values for DpbCasX and PlmCasX cleavage of the TS were 0.04351 and 0.004978 (fraction/minute), respectively. **(E)** *In vitro* plasmid cleavage activity comparison between DpbCasX and PlmCasX revealed by agarose gel. OC/N indicates the open-circle or nicked plasmid, L indicates the linearized plasmid and SC indicates the super-coiled plasmid (n = 3, representative gel shown). The fractions were collected at 0 min, 10 mins, 20 mins, 40 mins, 1 hr, 2 hrs, 4 hrs and 6 hrs, respectively. **(F)** The distribution of ten sgRNAv1 spacers across the genomically integrated GFP gene.

In HEK293 cells stably expressing GFP, plasmid transfection of PlmCasX showed adequate, and in some cases even higher, genome editing activity compared to DpbCasX using different GFP-targeting sgRNAv1s (Figures 2.2C and 2.2D; Figure 2.1F), which suggests PlmCasX is more proficient for genome editing by plasmid transfection. The vastly different *in vitro* and cell-based behavior further motivated us to understand the molecular difference between DpbCasX and PlmCasX. We therefore explored the structural details of PlmCasX and used this information to improve its biochemical and genome editing capacity through molecular engineering.

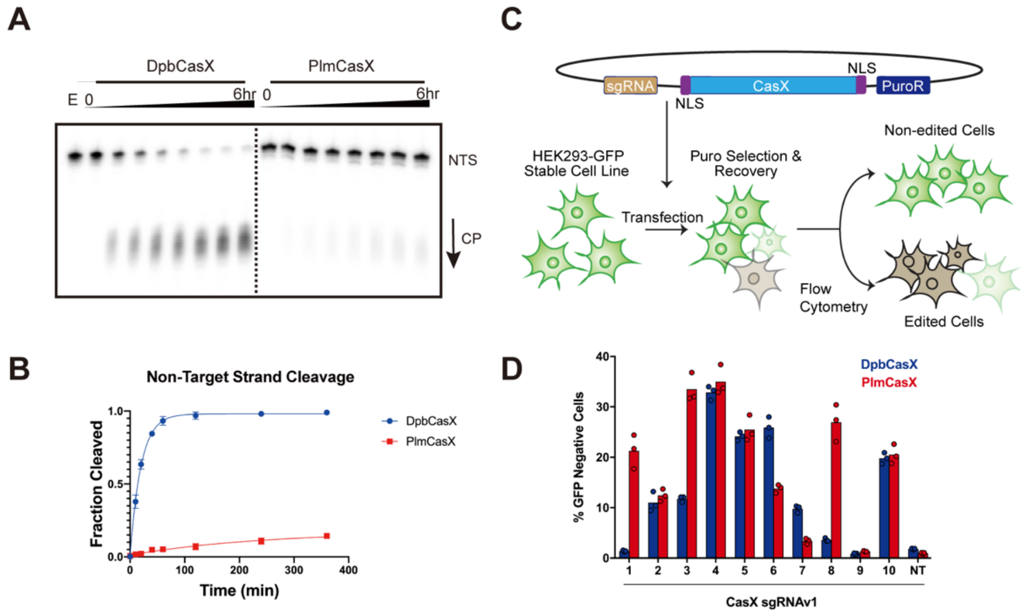


Figure 2.2: Comparison of DNA cleavage efficacy between DpbCasX and PlmCasX. (A) *In vitro* dsDNA cleavage activity comparison between DpbCasX and PlmCasX revealed by denaturing PAGE. NTS denotes the non-target strand which was ^{32}P labeled on the 5' end. CP indicates the cleavage product. The fractions were collected at 0 min, 10 mins, 20 mins, 40 mins, 1 hr, 2 hrs, 4 hrs and 6 hrs, respectively. E indicates an empty well with labeled DNA but no CasX enzyme. **(B)** The plot of DNA cleavage kinetics analyzed based on the NTS band density from fractions compared to the input NTS band density at the

reaction time of 0 min (n = 6, mean ± SD). One-phase association in Prism7 was used to model the kinetics here and in following experiments. The single turnover rate constant k values (fraction cleaved per minute) for DpbCasX and PlmCasX were 0.05031 and 0.004137 (fraction/minute), respectively. **(C)** The workflow for human cell genome editing experiments, which were based on the disruption of constitutive GFP expression in HEK293 cells. **(D)** Human cell genome editing by DpbCasX and PlmCasX with sgRNAv1, measured 10 days after plasmid transfection. The GFP disruption efficacies for 10 GFP-targeting guides both for DpbCasX and PlmCasX are shown (n = 3, the mean of three technical replicates is shown). NT indicates the non-targeting sgRNAv1.

2.3.2 The mobility of the Helical-II domain impairs DNA cutting by PlmCasX

We reconstituted a ternary complex containing deactivated PlmCasX (D659A, E756A, D922A; dPlmCasX), sgRNAv1 (122 nucleotides [nt]) and a complementary DNA substrate (40 base pairs [bp]) but found that the majority of ternary complex disassembled during cryo-EM grid preparation (Figure 2.3A). Crosslinking the complex using BS3 significantly improved the holo-complex stability for single particle cryo-EM analysis (Figure 2.3B). 3D classification and refinement identified three conformational populations of the cross-linked complex that were resolved at resolutions of 2.9 Å, 3.4 Å and, 3.2 Å (State I, State II, and State III, respectively) (Figure 2.4A; Figures 2.3B–2.3F; Table 2.1). The cryo-EM density maps for States I and II both accounted for the entire complex, with all seven CasX protein domains (Figure 2.4B). They correspond to a NTS DNA cleavage state and a TS DNA cleavage state, respectively (Figure 2.4A; Figure 2.5A). Comparison of these two conformations revealed a large structural rearrangement of the Helical-II (H2) domain, which may help to bend the sgRNA-DNA duplex and push the TS DNA into the RuvC catalytic domain. This structural rearrangement and stepwise DNA-loading mechanism is highly similar to the mechanism we previously described for the DpbCasX[29].

In State III of the dPlmCasX ternary complex, the NTS DNA appears loaded into the RuvC domain as in State I, but the density for the H2 domain is missing, most likely due to high flexibility (Figure 2.4A; Figures 2.3E and 2.5A. By losing the interaction with the H2 domain, the sgRNA scaffold stem in State III is fully exposed and bent about 20 and 23 downward relative to States I and II, respectively (Figure 2.4B). Notably, State III accounted for 41% of the entire population of dPlmCasX ternary complexes (Figure 2.4A; Figure 2.3B). For many type V CRISPR nucleases, a stable H2 domain (also termed the REC2 domain) in the ternary complex is structurally important to maintain the active DNA R-loop conformation and assist with DNA cleavage[29,228,229]. Thus, we hypothesized that the presence of State III, with its highly mobile H2 domain, could explain the reduced DNA editing capability of PlmCasX in vitro (Figures 2.2A and 2.2B). To test this hypothesis, we truncated the H2 domain in DpbCasX (DpbCasX ΔH2), which resulted in decreased DNA cleavage activity down to a level similar to wtPlmCasX (Figures 2.5B–2.5E). On the other hand, truncation of the H2 domain in PlmCasX (PlmCasX ΔH2) had little to no effect on DNA cleavage as compared to wtPlmCasX (Figures 2.5B–2.5E). These results suggest that the high mobility of the H2 domain in wtPlmCasX largely decreases its *in vitro* cleavage capability to a minimal level similar to H2 truncation constructs. We then tested whether PlmCasX ΔH2 and DpbCasX ΔH2 were still capable of genome editing in human cells. Truncation of the H2 domain in both CasX enzymes led to insignificant GFP disruption in HEK293 cells, which demonstrated the necessity of the H2 domain for effective genome editing in cells (Figures 2.5F and 2.5G).

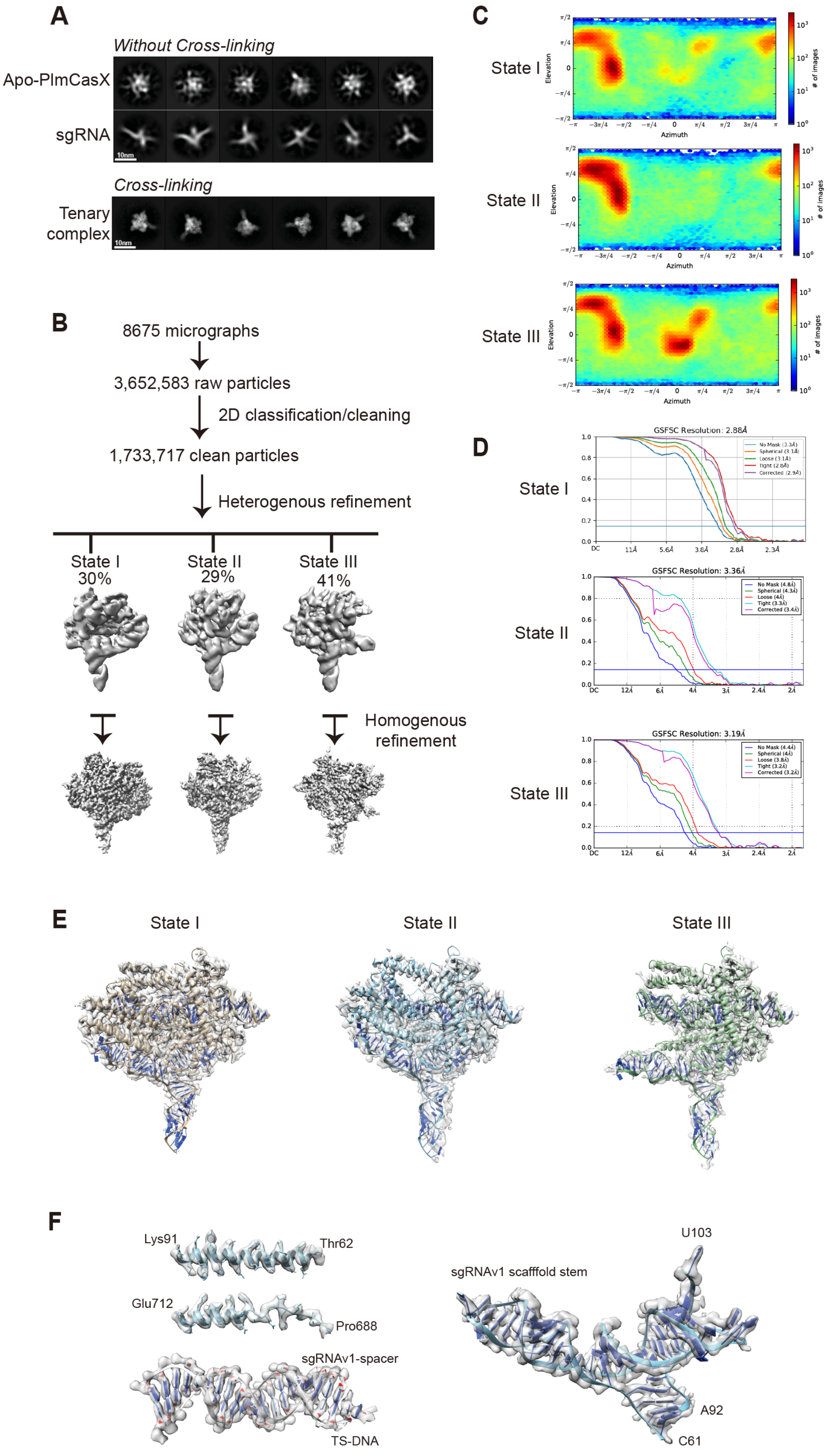

A
Without Cross-linking
Apo-PlmCasX
sgRNA
10nm
Cross-linking
Tenary complex
10nm
B
8675 micrographs
3,652,583 raw particles
2D classification/cleaning
1,733,717 clean particles
Heterogenous refinement
State I 30%
State II 29%
State III 41%
Homogenous refinement
C
State I
State II
State III
Elevation
Azimuth
# of images
D
State I
GSFSC Resolution: 2.88Å
No Mask (3.3Å)
Spherical (3.1Å)
Loose (3.1Å)
Tight (2.8Å)
Corrected (2.9Å)
State II
GSFSC Resolution: 3.36Å
No Mask (4.8Å)
Spherical (4.3Å)
Loose (4Å)
Tight (3.3Å)
Corrected (3.4Å)
State III
GSFSC Resolution: 3.19Å
No Mask (4.4Å)
Spherical (4Å)
Loose (3.8Å)
Tight (3.2Å)
Corrected (3.2Å)
E
State I
State II
State III
F
Lys91
Thr62
Glu712
Pro688
sgRNAv1-spacer
TS-DNA
sgRNAv1 scafffold stem
U103
A92
C61

Figure 2.3: Single particle cryo-EM analysis of the dPlmCasX-sgRNAv1-dsDNA complex. (A) Representative 2D class averages of native and cross-linked (BS3 crosslinking) complexes. **(B)** The workflow for single particle cryo-EM analysis in CryoSparc. The particles in State **I**, State **II** and State **III** account for 30%, 29% and 41%, respectively, of all the particles used for 3D refinement. **(C)** Euler angle distribution of the refined particles belonging to the three states. **(D)** Fourier shell correlation (FSC) curves calculated using two independent half maps. The final resolutions for B-factor corrected maps were 2.9 Å (State **I**), 3.4 Å (State **II**) and 3.2 Å (State **III**). Panels **C** and **D** were directly taken from the standard output of CryoSparc. **(E)** Overall fitting between the atomic models and the EM maps from the three states of the dPlmCasX-sgRNAv1-dsDNA complex. The atomic models are shown by ribbon cartoon and EM maps are shown by transparent surface. **(F)** Details of the fitting between the atomic model and the EM map of State **I**. The amino acid sidechains are shown.

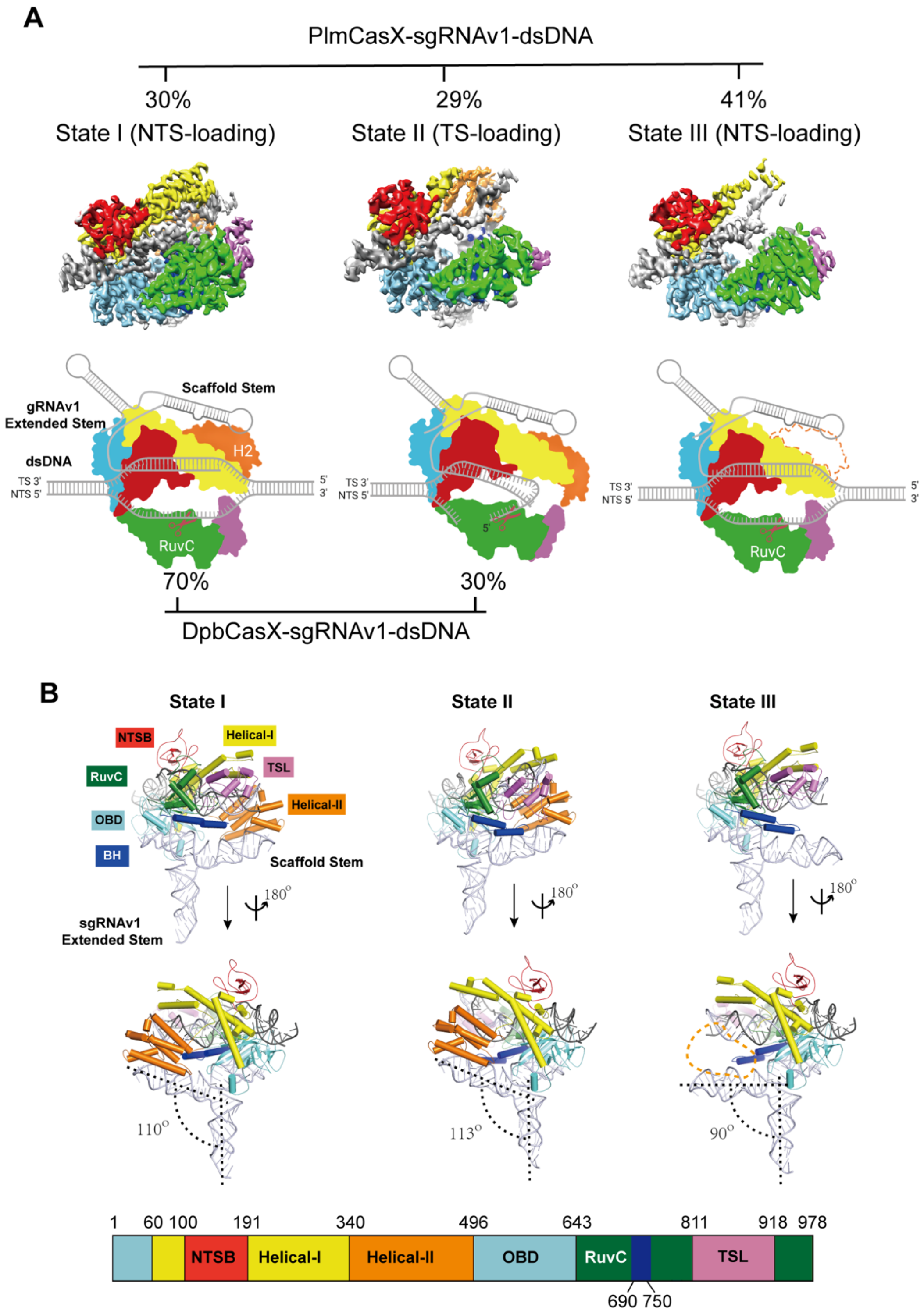


Figure 2.4: Overall structures of the dPlmCasX-sgRNAv1-dsDNA complex. (A) The different structural states of the dPlmCasX ternary complex with the sgRNAv1 scaffold revealed by single particle cryo-EM. The top views of refined EM maps for States I, II and III are shown in the top panel. The three maps are shown at contour thresholds of 6 to 9 times sigma. The cartoon model for each map is presented in the

bottom panel for better elucidation of substrate DNA loading and cleavage. Referring to the published DpbCasX maps (Liu et al., 2019), the NTSB domain is colored in red, Helical-I in yellow, Helical-II in orange, OBD in aquamarine, RuvC in green, TSL in pink and the bridge helix (BH) in blue. The sgRNAv1 is in light gray and the dsDNA is in dark gray. The invisible Helical-II (H2) domain in State III is represented with a dashed line. The particle proportions for all functional states within the PlmCasX complex (determined in this study) and DpbCasX complex (Liu et al., 2019) are presented with percentages. **(B)** The atomic models of the dPlmCasX-sgRNAv1-dsDNA complex in three states shown in a front and back view. The domain architecture of the PlmCasX amino acid sequence is shown in the bottom panel. The protein domains in the atomic models share the same color codes as in **A**. The angle between the sgRNAv1 scaffold stem and extended stem (defined by RNA helix rotation axis, black dashed line) was calculated in PyMol. The Helical-II domain region is outlined with an orange dashed line in State III.

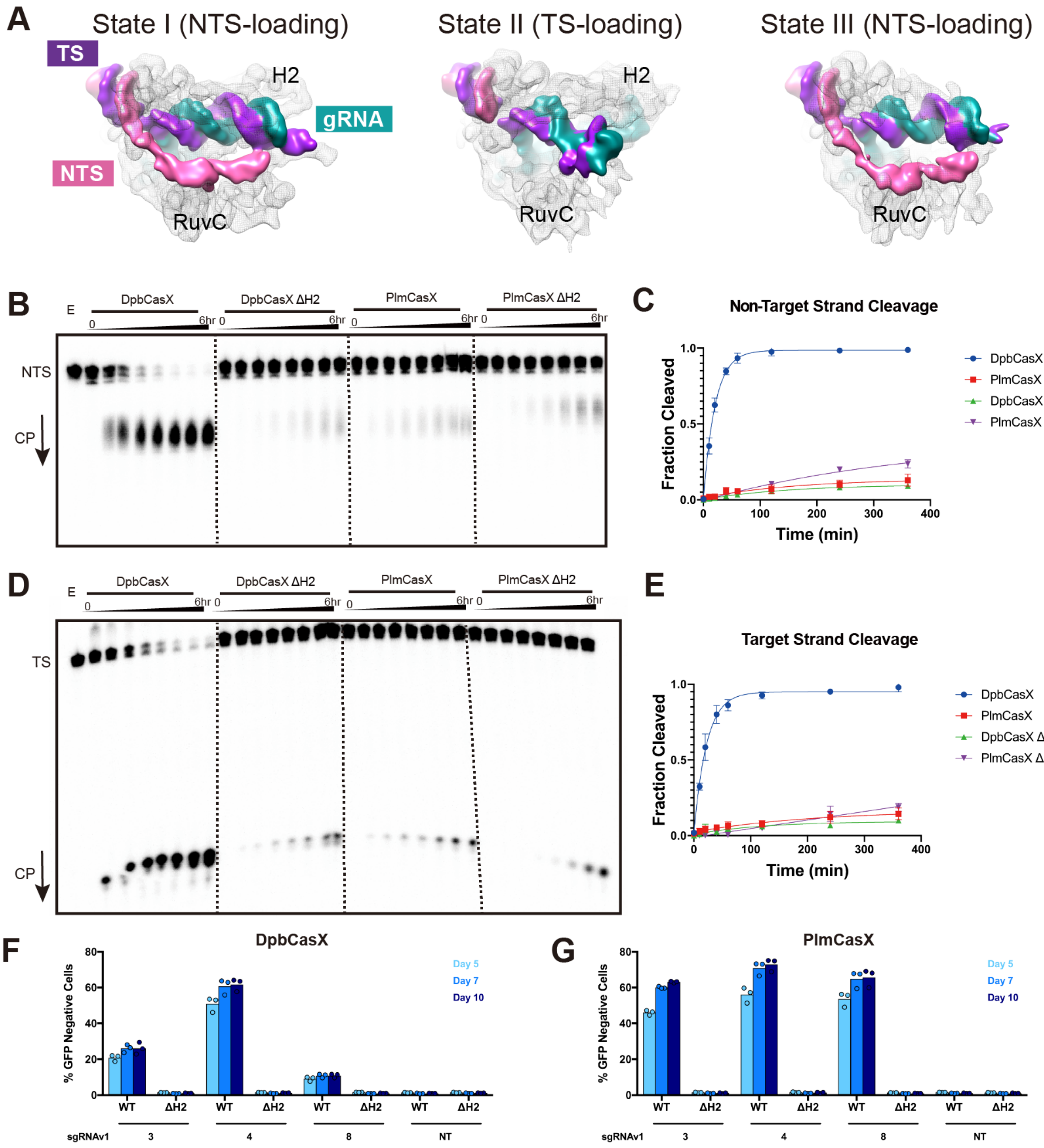


Figure 2.5: Effect of the Helical-II domain in DpbCasX and PlmCasX on dsDNA cleavage. (A) Conformational states within the dPlmCasX-sgRNAv1-dsDNA complex revealed by cryo-EM. The protein density is shown as a transparent mesh, the TS density is colored purple, the NTS is colored pink and the sgRNA density is colored teal. The RuvC and H2 domains are indicated in each map. All maps are low-pass filtered to 4.5 Å. **(B)** *In vitro* biochemical dsDNA cleavage activity comparison between DpbCasX, DpbCasX with a Helical-II truncation (DpbCasX ΔH2), PlmCasX and PlmCasX with a Helical-II truncation (PlmCasX ΔH2). NTS denotes the non-target strand DNA which was ^{32}P labeled on the 5' end. CP indicates the cleavage product. The fractions were collected at 0 min, 10 mins, 20 mins, 40 mins, 1 hr, 2 hrs, 4 hrs and 6 hrs, respectively. E indicates an empty well with labeled DNA but no CasX enzyme. **(C)** Plot of DNA cleavage kinetics analyzed based on the cleaved NTS band density compared to the input NTS band density at the reaction time of 0 min (n = 3, mean ± SD). **(D)** *In vitro* biochemical dsDNA cleavage activity comparison among DpbCasX, DpbCasX ΔH2, PlmCasX and PlmCasX ΔH2. TS denotes the target strand DNA which was ^{32}P labeled on the 5' end. CP indicates the cleavage product. The fractions were collected at 0 min, 10 mins, 20 mins, 40 mins, 1 hr, 2 hrs, 4 hrs and 6 hrs, respectively. E indicates an empty well with labeled DNA but no CasX enzyme. **(E)** The cleavage fraction analysis based on the cleaved TS band density compared to the input TS band density at the reaction time of 0 min (n = 3, mean ± SD). **(F)** Genome editing efficacy comparison between wild type DpbCasX (WT) and DpbCasX ΔH2 (ΔH2) constructs with sgRNAv1. **(G)** Genome editing efficacy comparison between wild type PlmCasX (WT) and PlmCasX ΔH2 (ΔH2) constructs with sgRNAv1. sgRNAv1 with spacers 3, 4 and 8 targeting the GFP gene were tested. NT denotes the non-targeting sgRNAv1 control (n = 3, mean). Cells were collected and analyzed by flow cytometry at 5, 7 and 10 days after plasmid transfection.

2.3.3 Nucleotide-binding loops in CasX contribute to R-loop assembly and DNA cutting

To further understand the structural details that led to unstable assembly and a mobile H2 domain within the PlmCasX ternary complex, we conducted a comprehensive analysis of the sequence and structural differences between PlmCasX and DpbCasX in State I. PlmCasX and DpbCasX share 68.5% sequence identity overall, with a structural similarity Z score of 33.8 as calculated by the Dali Server[230]. We identified the protein domains (NTSB, OBD, Helical-I, Helical-II, RuvC, TSL, and BH) of PlmCasX that correspond to those in the DpbCasX structure and redefined the protein sequence corresponding to the BH domain based on the better resolved structural details in PlmCasX (Figure 2.4B). Within the context of the same protein architecture, we found three nucleotide-binding loops that exist exclusively in either PlmCasX or DpbCasX and could have relevance to R-loop complex assembly and DNA cleavage (Figures 2.6A and 2.6B).

We found that the region 1 loop (R1, K390~L396) in the DpbCasX H2 domain, which, together with the H2 domain helices, forms a deep pocket for tight binding of the sgRNA scaffold stem, likely contributes to the stable assembly of the R-loop complex (Figure 2.6A; Figures 2.7A and 2.7B). R1 is shortened in wtPlmCasX, giving rise to a shallower binding pocket that likely leads to weaker H2 domain-sgRNA binding and eventually the assembly of a less stable R-loop complex (Figure 2.6A; Figures 2.7A and 2.7B). A chimeric PlmCasX with the DpbCasX R1 loop (PlmCasX-R1) showed about 3-fold higher DNA cleavage kinetics *in vitro* (Figure 2.6C; Figures 2.7C; Figures 2.8A-C), but similar DNA-editing activity in HEK293 cells compared to that of wtPlmCasX (Figure 2.8D).

The region 2 loop (R2, G520I526) is only present in the DpbCasX OBD domain and structurally interacts with the PAM proximal region (Figure 2.6A; Figures 2.7A and 2.7B), which may be important for initial steps of dsDNA substrate loading. However, adding R2 to PlmCasX-R1 (PlmCasX-R1-R2) completely disrupted DNA cleavage *in vitro* and editing in mammalian cells (Figure 2.6C; Figure 2.7C; Figure 2.8A-2.8D). This result suggests that for R2, interactions with both DNA and the surrounding protein elements are likely important for proper ternary complex assembly (Figure 2.6A).

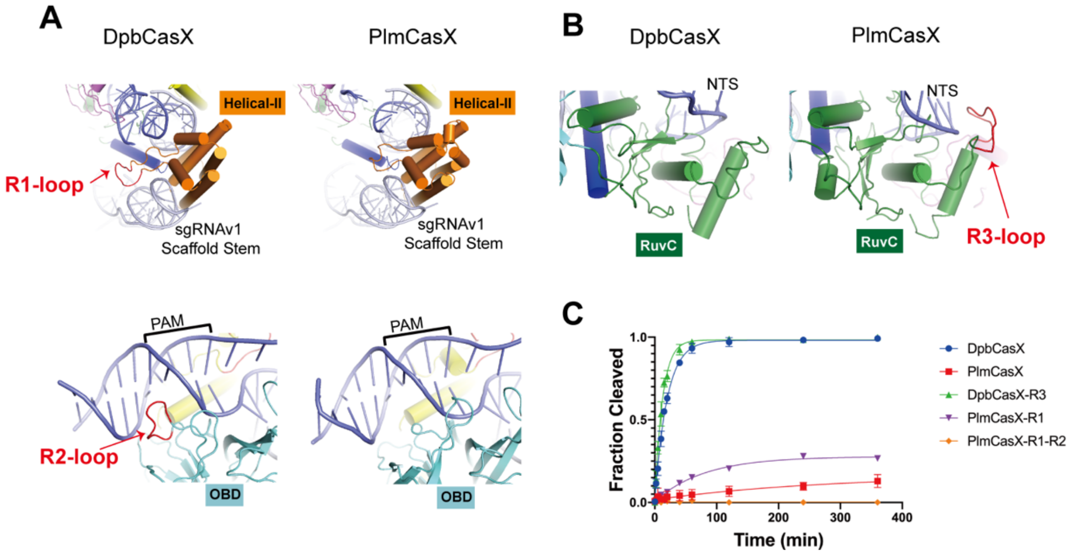


Figure 2.6: Structural comparison between DpbCasX and PlmCasX. (A) Region 1 (R1) and region 2 (R2) loops are located within DpbCasX but are absent from PlmCasX. The protein domains are colored as seen in Figure 2; the sgRNAv1 is colored in light gray and the dsDNA (with PAM region labeled) in blue. **(B)** The region 3 (R3) loop is located within PlmCasX but is absent from DpbCasX. **(C)** Biochemical dsDNA cleavage activity comparison between CasX chimeras with sgRNAv1 (n = 3, mean ± SD), based on cleavage of the NTS DNA. The rate constant k values for DpbCasX, PlmCasX, DpbCasX-R3 and PlmCasX-R1 were 0.05042, 0.003569, 0.07993 and 0.012503 (fraction/minute), respectively.

The region 3 loop (R3, Q945~G951) is exclusively present in PlmCasX, and similar to R1, forms a deep active pocket together with the remaining part of the RuvC domain that likely helps to faithfully accommodate and degrade ssDNA substrates (Figure 2.6B; Figures 2.7A and 2.7B). In contrast, the DpbCasX RuvC lacks R3 and contains a shallow active pocket that may have a lower affinity interaction with a ssDNA substrate (Figure 2.6B; Figure 2.7B). A chimeric DpbCasX with the PlmCasX R3 (DpbCasXR3) had about 1.6-fold higher DNA cleavage kinetics in our biochemical cleavage assays (Figure 2.6C; Figure 2.7C; Figures 2.8A-2.8C), and a 1.6-fold increase in median genome editing efficacy of HEK293 cells across three sgRNAv1s, compared to wtDpbCasX (Figure 2.8E).

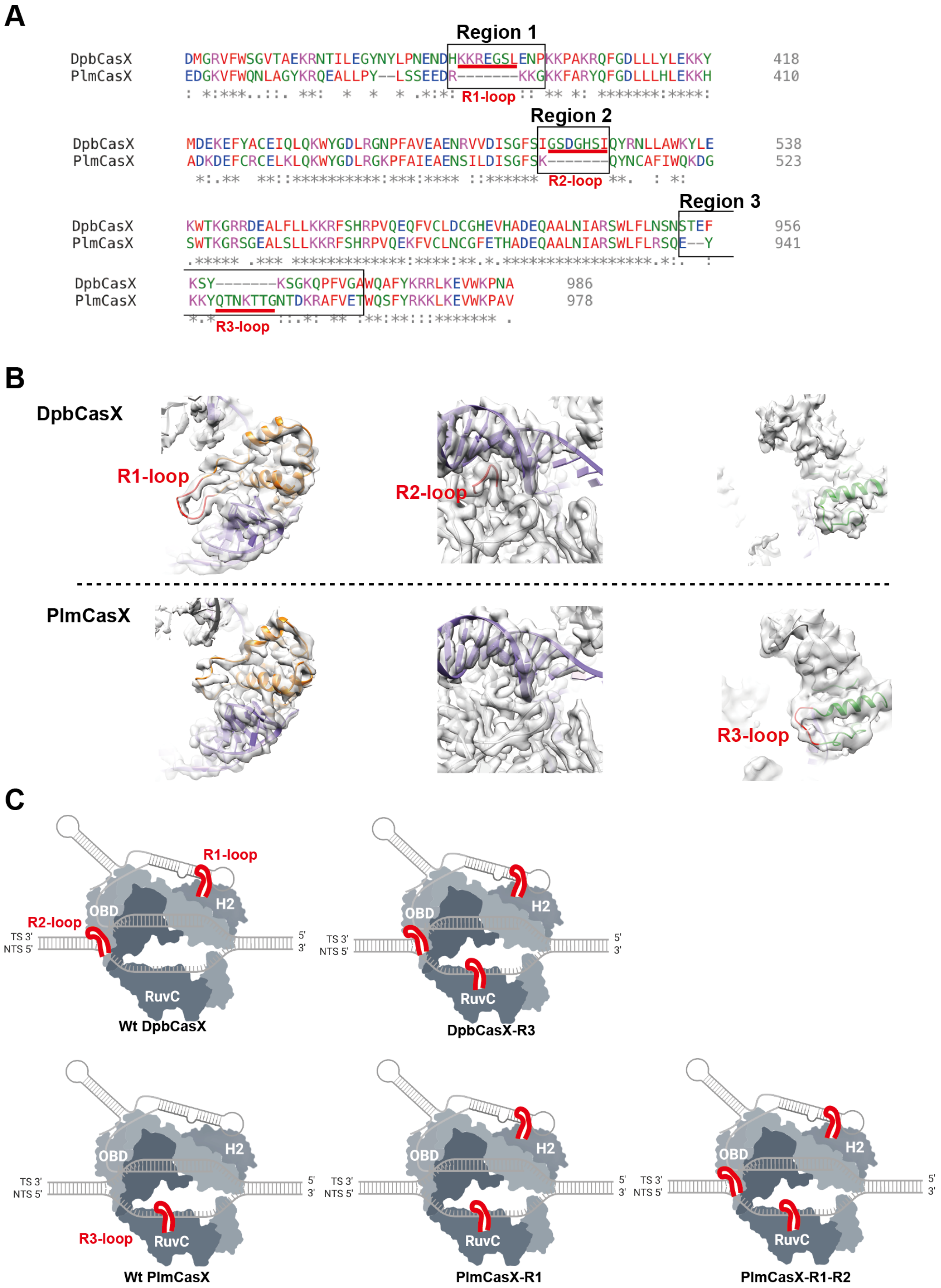

A
Region 1
DpbCasX DMGRVFWSGVTAEKRNTILEGYNYLPNENDHKKREGSLENPKKPAKRQFGDLLLYLEKKY 418
PlmCasX EDGKVFWQNLAGYKRQEALLPY--LSSEEDR-------KKGKKFARYQFGDLLLHLEKKH 410
R1-loop
Region 2
DpbCasX MDEKEFYACEIQLQKWYGDLRGNPFAVEAENRVVDISGFSIGSDGHSIQYRNLLAWKYLE 538
PlmCasX ADKDEFCRCELKLQKWYGDLRGKPFAIEAENSILDISGFSK-------QYNCAFIWQKDG 523
R2-loop
Region 3
DpbCasX KWTKGRRDEALFLLKKRFSHRPVQEQFVCLDCGHEVHADEQAALNIARSWLFLNSNSTEF 956
PlmCasX SWTKGRSGEALSLLKKRFSHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQE--Y 941
DpbCasX KSY-------KSGKQPFVGAWQAFYKRRLKEVWKPNA 986
PlmCasX KKYQTNKTTGNTDKRAFVETWQSFYRKKLKEVWKPAV 978
R3-loop
B
DpbCasX
R1-loop
R2-loop
PlmCasX
R3-loop
C
R1-loop
R2-loop
OBD
H2
RuvC
TS 3'
NTS 5'
5'
3'
Wt DpbCasX
DpbCasX-R3
R3-loop
Wt PlmCasX
PlmCasX-R1
PlmCasX-R1-R2

Figure 2.7: Design of CasX protein chimeras. (A) Amino acid sequence alignment between DpbCasX and PlmCasX using Clustal. Only regions in proximity to loops R1, R2 and R3 are shown for clarity. **(B)** Details of the fitting between the atomic model and EM map at region 1, 2 and 3. The R1, R2 and R3-loops are colored in red and labeled in each model. For DpbCasX, the published model (PDB code 6NY2) with re-built R1-loop and EM map (EMDB code EMD-8996) was used. In the bottom right panel, the EM map is shown at a low contour threshold of 3.68 times sigma due to the weak density of the R3-loop in the PlmCasX reconstruction. The EM maps in other panels are shown at a contour threshold of 6 to 9 times sigma. **(C)** Cartoon models for wild type CasX proteins and chimeric designs. The R1, R2 and R3 regions are presented as red loops on the 3D structure.

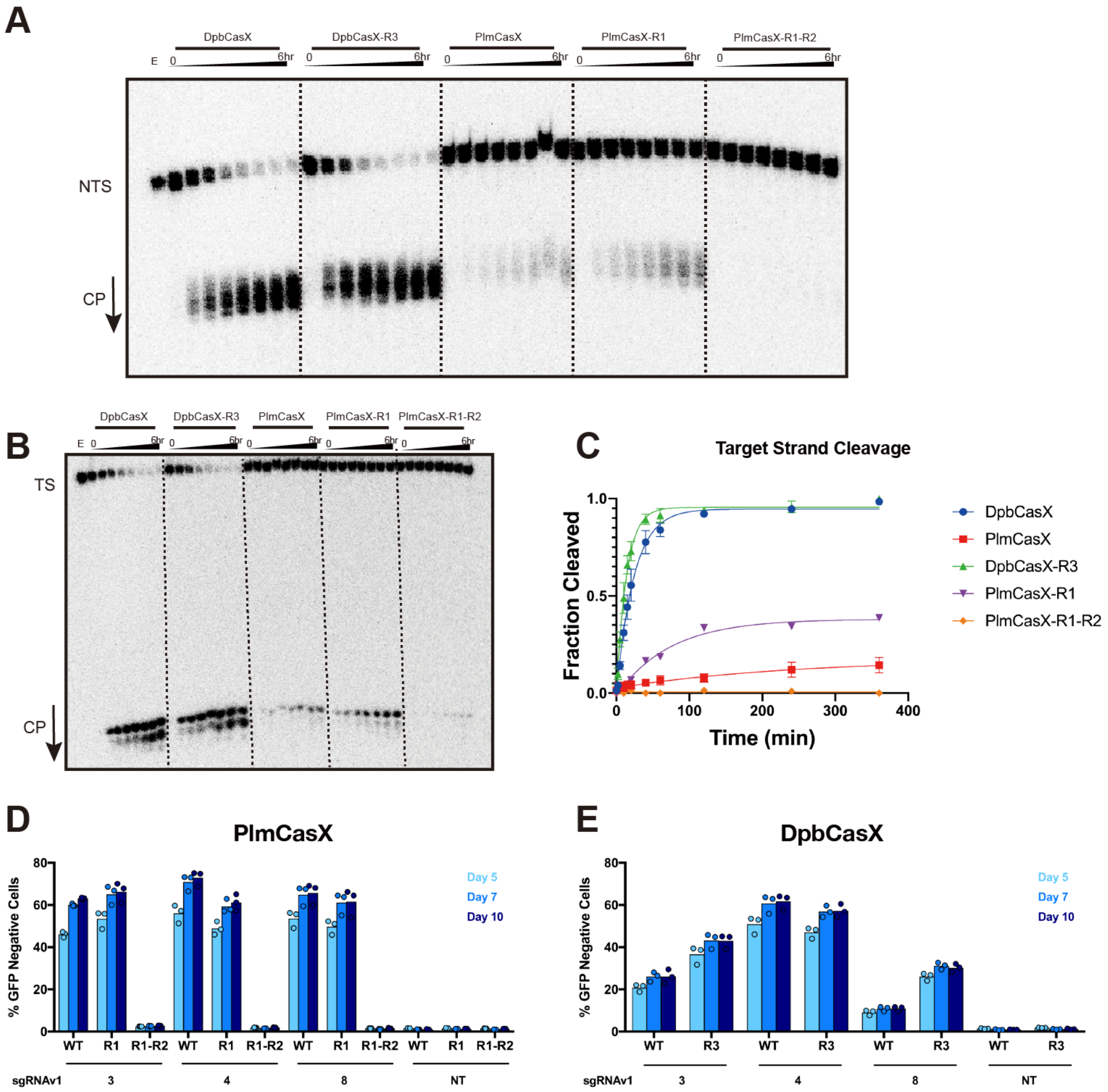


Figure 2.8: DNA cleavage by CasX protein chimeras. (A) *In vitro* dsDNA cleavage activity comparison among wild type and CasX chimeras using the sgRNAv1 scaffold revealed by denaturing PAGE. NTS denotes the non-target strand DNA which was ^{32}P labeled on the 5' end. CP indicates the cleavage product. The fractions were collected at 0 min, 10 mins, 20 mins, 40 mins, 1hr, 2hrs, 4hrs and 6hrs, respectively. E indicates an empty well with labeled DNA but no CasX enzyme. **(B)** *In vitro* dsDNA cleavage activity comparison among wild type and CasX chimeras using the sgRNAv1 scaffold revealed by denaturing PAGE. TS denotes the target strand DNA which was ^{32}P labeled on the 5' end. The fractions were collected

at 0 min, 10 mins, 20 mins, 40 mins, 1hr, 2hrs, 4hrs and 6hrs, respectively. E indicates an empty well with labeled DNA but no CasX enzyme. **(C)** Plot of DNA cleavage kinetics analyzed based on the TS band density from the cleaved fractions compared to the input TS band density at the reaction time of 0 min (n = 3, mean ± SD). The rate constant k values for DpbCasX, PlmCasX, DpbCasX-R3 and PlmCasX-R1 were 0.04189, 0.004144, 0.07236 and 0.01364 (fraction/minute), respectively. **(D)** Genome editing efficacy comparison between wild type PlmCasX (WT), PlmCasX-R1 (R1), and PlmCasX-R1-R2 (R1-R2) chimeras with sgRNAv1. **(E)** Genome editing efficacy comparison between wild type DpbCasX (WT) and DpbCasX-R3 chimera (R3) with sgRNAv1. sgRNAv1 with spacers 3, 4 and 8 targeting the GFP gene were tested. NT denotes the non-targeting sgRNAv1 control (n = 3, mean). Cells were collected and analyzed by flow cytometry at 5, 7 and 10 days after plasmid transfection.

2.3.4 A new sgRNA scaffold promotes CasX R-loop assembly and DNA cleavage

Our cryo-EM structures indicate that the weak interaction between the H2 domain and sgRNA scaffold stem likely interferes with R-loop complex assembly and thus decreases the DNA cleavage activity of PlmCasX. In addition to engineering the CasX protein, we were curious as to whether we could redesign the sgRNA sequence to stabilize the scaffold stem for better interaction with the H2 domain and further improve DNA cleavage activity. Based on secondary structure prediction and available atomic structures, adding an additional U at the 5' end of sgRNAv1 could form a new base pairing interaction with A29 and thus limit the mobility of the scaffold stem without changing the structure (hereafter sgRNAv1-2) (Figures 2.9A-2.9C). However, DpbCasX showed lower DNA cleavage activity with sgRNAv1-2 (Figures 2.9D and 2.9E). Instead, CasX may require a certain level of flexibility within the sgRNA to adopt the necessary conformational changes during the multi-step assembly of the ternary complex[29]. By structural inspection, disruption of the G30-C54 base pairing and adding nucleotides after G23 to increase the single stranded linker may increase the flexibility of the sgRNA scaffold stem while preserving its predicted secondary structure (Figures 2.9B and 2.9F). RNA profiling showed that the native PlmCasX tracrRNA sequence also contains additional nucleotides compared to sgRNAv1, which was designed based on the native DpbCasX tracrRNA sequence (Figures 2.9A). Referring to this structural interpretation and the PlmCasX tracrRNA sequence, we revised the sgRNA design by adding an additional nucleotide A after G23 and swapping the G30-C54 pair to U31-U55 (Figures 2.9C and 2.9F). The new sgRNA (hereafter sgRNAv2) enhanced both DpbCasX and PlmCasX dsDNA cleavage kinetics by 5.6 and 11-fold, respectively (Figures 2.10A and 2.10B). Again, adding a U or more nucleotides to the 5' end of sgRNAv2 decreased the dsDNA cleavage activity of PlmCasX (sgRNAv2-2 and sgRNAv2-3) (Figures 2.9D-2.9F). Both DpbCasX and PlmCasX also showed increased plasmid linearization activity using sgRNAv2 compared to sgRNAv1 (Figure 2.1E; Figure 2.9G).

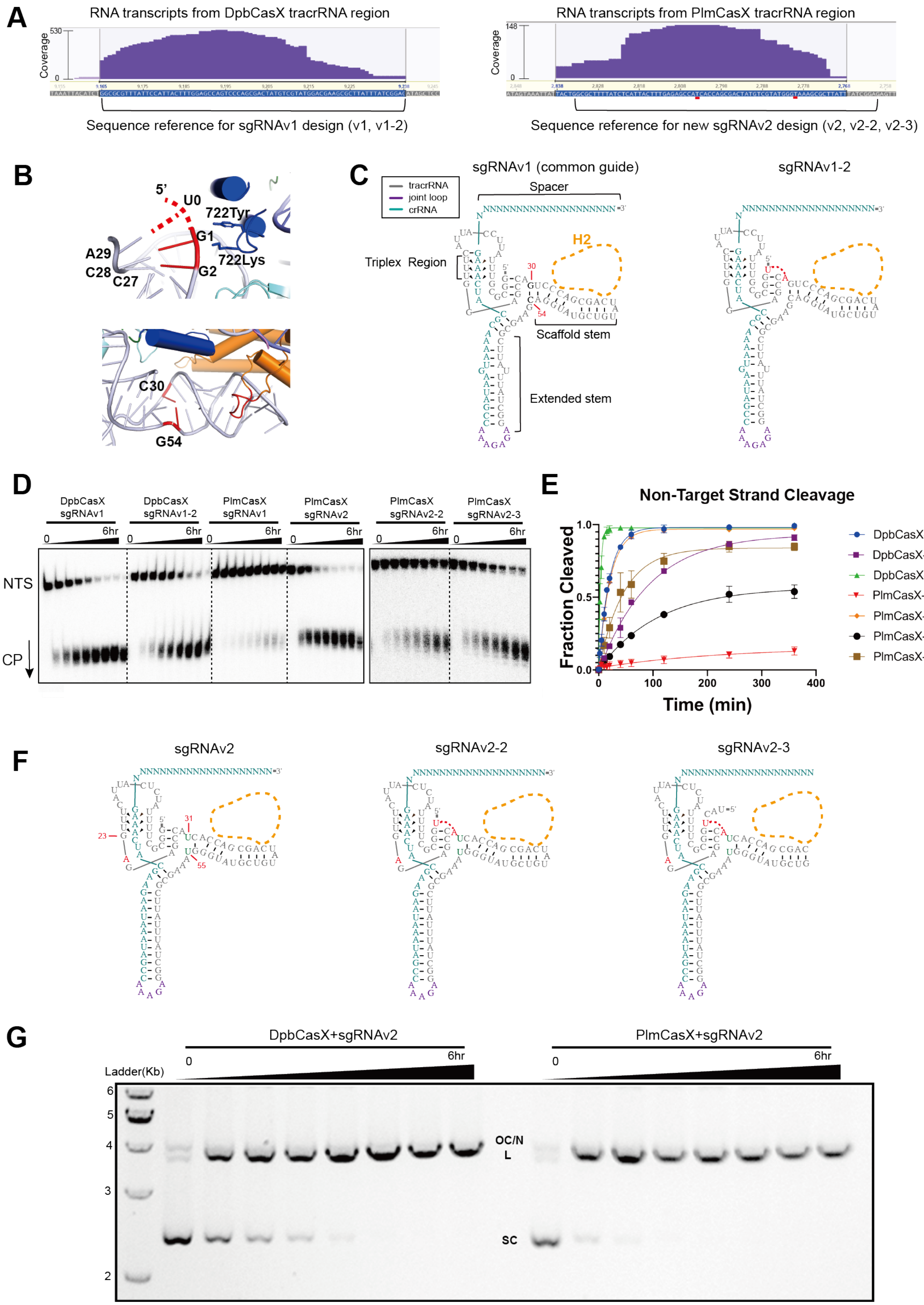

A
RNA transcripts from DpbCasX tracrRNA region
Coverage
530
0
Sequence reference for sgRNAv1 design (v1, v1-2)
RNA transcripts from PlmCasX tracrRNA region
Coverage
148
0
Sequence reference for new sgRNAv2 design (v2, v2-2, v2-3)
B
5'
U0
722Tyr
G1
722Lys
A29
C28
C27
G2
C30
G54
C
tracrRNA
joint loop
crRNA
sgRNAv1 (common guide)
Spacer
Triplex Region
H2
30
54
Scaffold stem
Extended stem
sgRNAv1-2
D
DpbCasX sgRNAv1
DpbCasX sgRNAv1-2
PlmCasX sgRNAv1
PlmCasX sgRNAv2
PlmCasX sgRNAv2-2
PlmCasX sgRNAv2-3
0
6hr
NTS
CP
E
Non-Target Strand Cleavage
Fraction Cleaved
1.0
0.5
0.0
0
100
200
300
400
Time (min)
DpbCasX-v1
DpbCasX-v1-2
DpbCasX-v2
PlmCasX-v1
PlmCasX-v2
PlmCasX-v2-2
PlmCasX-v2-3
F
sgRNAv2
sgRNAv2-2
sgRNAv2-3
23
31
55
G
DpbCasX+sgRNAv2
PlmCasX+sgRNAv2
0
6hr
Ladder(Kb)
6
5
4
3
2
OC/NL
SC

Figure 2.9: Updated design of the CasX sgRNA. (A) Transcripts from the hypothetical CasX tracrRNA loci revealed by meta-transcriptome sequencing. The coverage axis denotes the number of sequencing reads. The abscissa denotes the genome sequence of the hypothetical tracrRNA region. **(B)** The structural details for sgRNAv1. The additional U at the 5' end was hypothetically modeled in the top panel. The structural details of the G30-C54 base-paring region in sgRNAv1 are also shown in the bottom panel. **(C)** The secondary architecture of sgRNAv1 revealed by cryo-EM. The sgRNAv1-2 architecture is modeled based on sgRNAv1. The potential base pairing between the additional 5' U and A29 is labeled with a dashed line. The Helical-II domain was modeled to indicate the interaction interface with the sgRNA scaffold stem. **(D)** *In vitro* dsDNA cleavage activity comparison among different sgRNAs by DpbCasX and PlmCasX. NTS denotes the non-target strand DNA which was ^{32}P labeled on the 5' end. The fractions were collected at 0 min, 10 mins, 20 mins, 40 mins, 1 hr, 2 hrs, 4 hrs and 6 hrs, respectively. **(E)** Plot of DNA cleavage kinetics analyzed based on the NTS band density from the cleaved fractions compared to the input NTS band density at the reaction time of 0 min (n = 3, mean ± SD) (DpbCasX-v2 from Figure 4A). The rate constant k values for DpbCasX-v1, DpbCasX-v1-2, DpbCasX-v2, PlmCasX-v1, PlmCasX-v2, PlmCasX-v2-2 and PlmCasX-v2-3 were 0.05065, 0.01087, 0.2817 and 0.004433, 0.04858, 0.009191 and 0.02169 (fraction/minute), respectively. **(F)** The hypothetical secondary architectures of sgRNAv2 designs. **(G)** *In vitro* plasmid cleavage activity comparison between DpbCasX and PlmCasX with sgRNAv2 revealed by agarose gel. The fractions were collected at 0 min, 10 mins, 20 mins, 40 mins, 1 hr, 2 hrs, 4 hrs and 6 hrs, respectively. OC/N indicates the open-circle or nicked plasmids, L indicates the linearized plasmids and SC indicates the super-coiled plasmids (n = 3, representative gel shown).

To further investigate whether and how sgRNAv2 helped with the overall stability of the R-loop complex (Figure 2.10C), we performed single particle cryo-EM analysis on the dPlmCasXsgRNAv2-dsDNA (40 bp) complex. Indeed, the new complex appeared more stable without the need for crosslinking during cryo-EM sample preparation (Figures 2.11A). 3D classification showed that only 14% of the dPlmCasX-sgRNAv2-dsDNA complexes were present in State III, a sharp decrease from 41% for the dPlmCasX-sgRNAv1-dsDNA complexes, presumably due to the higher affinity interaction of the H2 domain with sgRNAv2 (Figure 2.10D; Figures 2.11B). Further 3D variability analysis for particles from State I of dPlmCasX-sgRNAv2-dsDNA indicated that the extended stem of sgRNAv2 adopts a continuum of states, that may contribute to the limited resolution of the EM map (Figures 2.11C and 2.11D; Table 2.1). Structural comparison of dPlmCasX-sgRNAv1-dsDNA and dPlmCasX-sgRNAv2- dsDNA showed that the addition of an A after G23 increased the curvature in the single stranded RNA linker, and swapping the G30-C54 pair to U31-U55 generated a minor distortion at the end of the sgRNAv2 scaffold stem (Figures 2.10E and 2.10F; Figures 2.11E and 2.11F). Meanwhile, the angle between the sgRNA extended stem and scaffold stem decreased from 110 to 90 (Figure 2.4B; Figure 2.10E). Notably, the structures of the PlmCasX proteins appear indistinguishable between the two complexes (Figure 2.10F). Overall, sgRNAv2 increases the stability of the R-loop complex, which could explain the observed increase in DNA cleavage activity when complexed with both DpbCasX and PlmCasX.

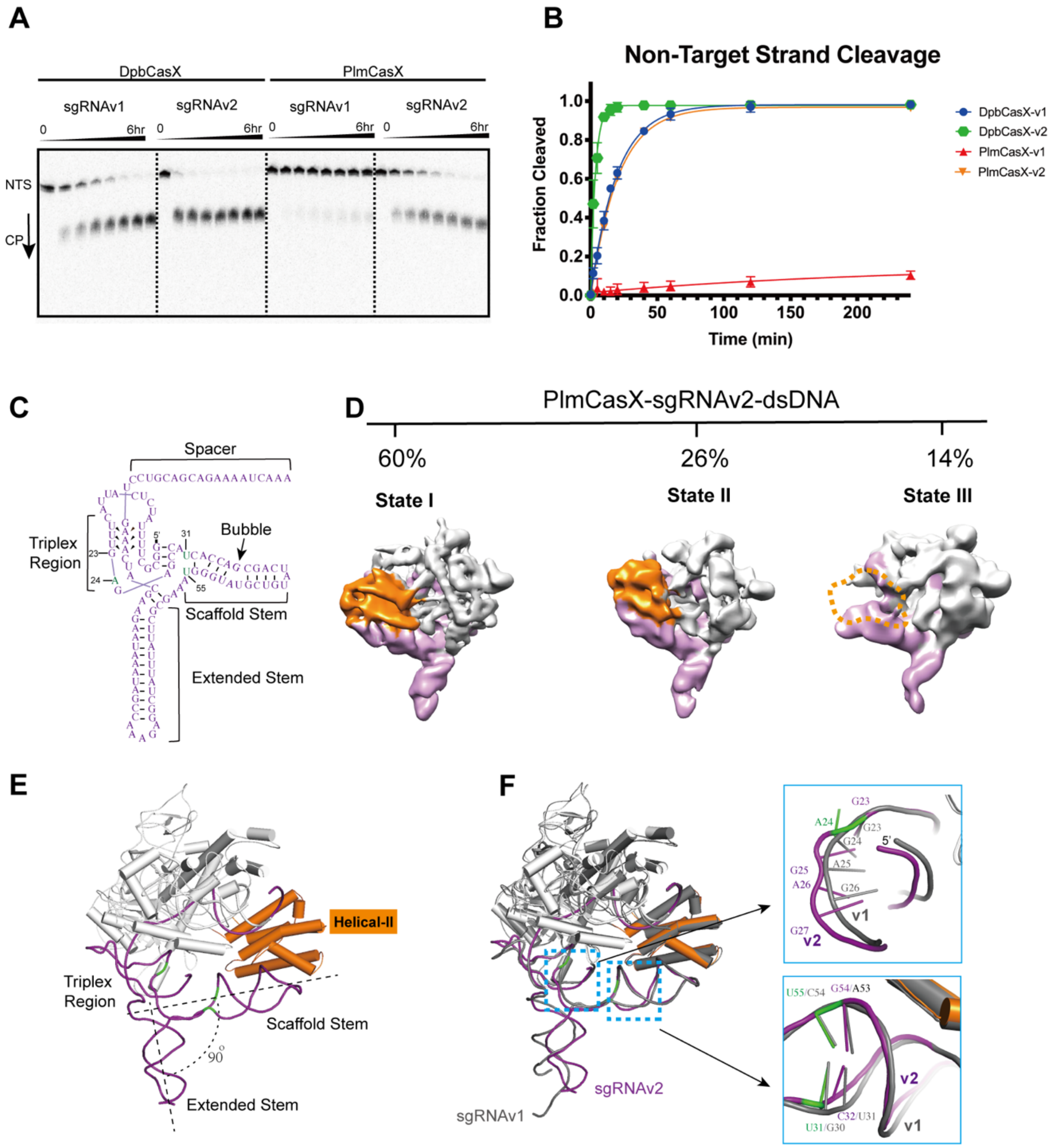


Figure 2.10: In vitro biochemical cleavage behavior of CasX using sgRNAv2. (A) *In vitro* dsDNA cleavage activity comparison between DpbCasX and PlmCasX using sgRNAv1 and sgRNAv2 revealed by denaturing PAGE. The fractions were collected at 0 min, 10 mins, 20 mins, 40 mins, 1 hr, 2 hrs, 4 hrs and 6 hrs, respectively. **(B)** Cleavage fraction analysis based on the NTS band density compared to the input NTS band density at the reaction time of 0 min (n = 5, mean ± SD). CasX-v1 denotes the CasX complex using sgRNAv1, while CasX-v2 denotes the CasX complex using sgRNAv2. The rate constant k values for DpbCasX-v1, PlmCasX-v1, DpbCasX-v2 and PlmCasX-v2 were 0.05065, 0.004433, 0.2817 and 0.04858 (fraction/minute), respectively. **(C)** The secondary architecture of sgRNAv2 revealed by cryo-EM. The key nucleic acid variants in sgRNAv2 compared to sgRNAv1 are marked in green. The nucleotide numbers for G23, A24, U31 and U55 are labeled. **(D)** The different structural states of the dPlmCasX ternary complex with the sgRNAv2 scaffold revealed by single particle cryo-EM. The back views of refined EM maps for State I, State II and State III are shown in the top panel. The three maps were low-pass filtered at 6 Å and

shown at contour thresholds of 6 to 9 times sigma for clear presentation and comparison. The Helical-II domain is colored in orange and the sgRNAv2 in purple. Other parts of the complex are colored in light gray. The invisible Helical-II domain in State III is represented with a dashed outline. The particle proportions for all functional states within the dPlmCasX-sgRNAv2-dsDNA complex are presented with percentages. **(E)** Atomic model of dPlmCasX-sgRNAv2-dsDNA in State I. The CasX protein is colored in light gray and the sgRNAv2 is shown in purple. The Helical-II domain is emphasized by highlighting in orange. **(F)** Structural comparison between dPlmCasX-sgRNAv1-dsDNA (all in gray) and dPlmCasX-sgRNAv2-dsDNA (CasX in light gray and sgRNAv2 in purple) complexes in State I. The two structures were aligned in PyMol referring to the PlmCasX protein and dsDNA. The dsDNA models are hidden for better presentation. The zoomed in features for the sgRNA triplex region (top) and scaffold stem (bottom) are shown in the right panels, with the number of key nucleotides labeled.

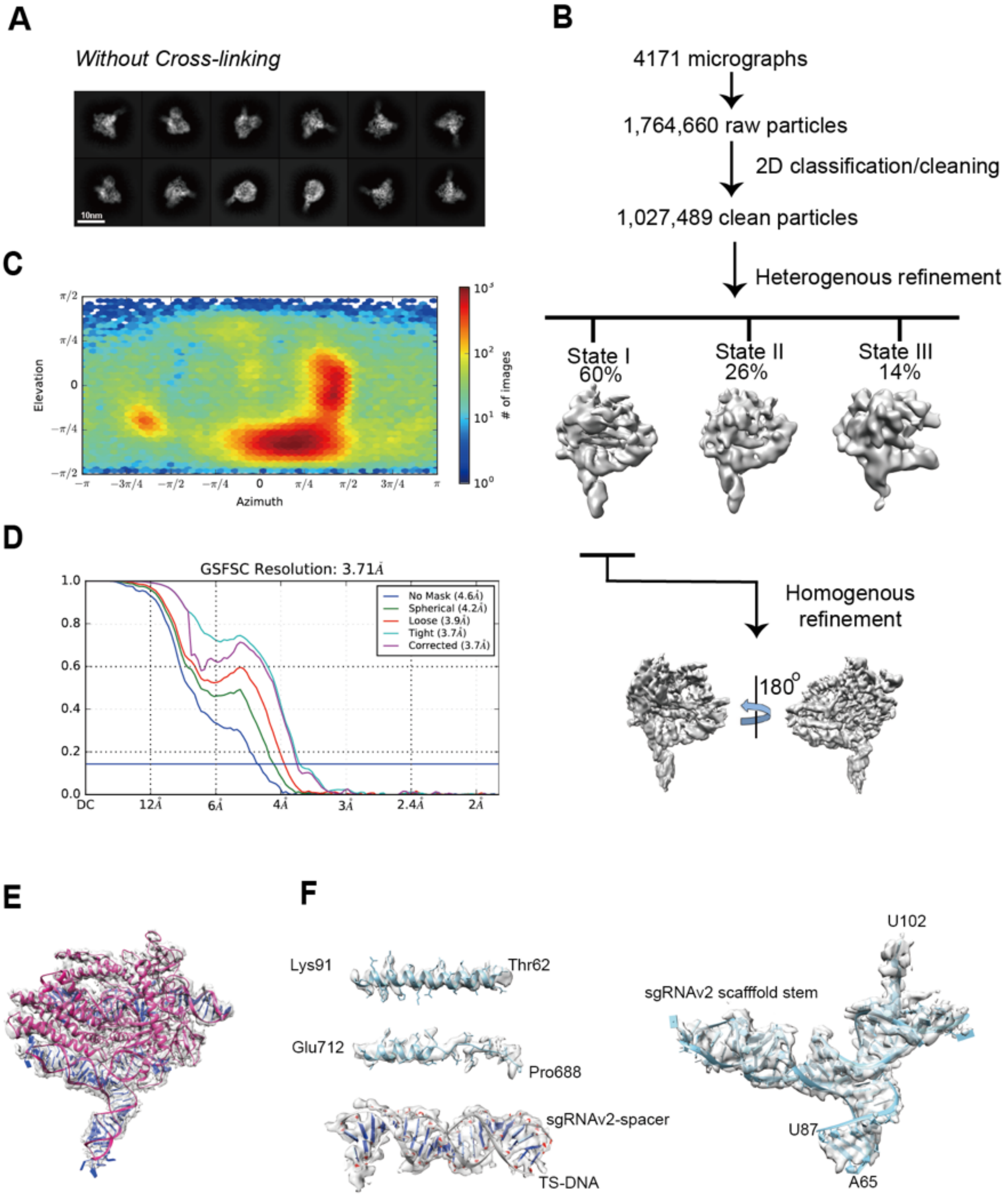


Figure 2.11: Single particle cryo-EM analysis of the dPlmCasX-sgRNAv2-dsDNA complex. (A) Representative 2D class averages of the native complex (without BS3 crosslinking); the scale bar is 10 nm. **(B)** The workflow for single particle cryo-EM analysis in CryoSparc. The particles in State **I**, State **II** and State **III** account for 60%, 26% and 14%, respectively, of all the particles used for 3D refinement. **(C)** Euler angle distribution of the refined particles belonging to State **I**. **(D)** Fourier shell correlation (FSC) curves calculated using two independent half maps. The final resolutions for the B-factor corrected State **I** map was 3.7 Å. Panels **C** and **D** are directly taken from the standard outputs of CryoSparc. **(E)** Overall fitting between the atomic model and the EM map of the dPlmCasX-sgRNAv2-dsDNA complex in State **I**. The

atomic model is shown by ribbon cartoon and EM map is shown by transparent surface. **(F)** Details of the fitting between the atomic model and the EM map of State **I**. The amino acid sidechains are shown.

2.3.5 Improved versions of CasX for mammalian genome editing

Using structure-based engineering of both the CasX protein and sgRNA, we were able to improve DNA cleavage by CasX *in vitro* (Figure 2.8C; Figure 2.10B). We further tested the newly designed sgRNA for mammalian cell genome editing and observed a considerable improvement in DNA editing efficacy for both DpbCasX and PlmCasX using ten different sgRNAs targeting HEK293 cells stably expressing GFP (Figure 2.12A). The median editing efficacy for DpbCasX and PlmCasX with sgRNAv2 (DpbCasX-v2 and PlmCasX-v2) was 43.50% and 77.25%, respectively, a significant improvement from 31.45% for DpbCasX and 32.95% for PlmCasX when using sgRNAv1 (DpbCasX-v1 and PlmCasX-v1) (Figure 2.12B).

Next, we were curious as to whether combining both protein chimeras and the new sgRNAv2 could make a yet more effective CasX genome editing tool. Indeed, a combination of DpbCasX-R3 and sgRNAv2 (DpbCasX-R3-v2) outperformed all other combinations of CasX and sgRNA constructs in *in vitro* dsDNA cleavage activity (Figures 2.13A and 2.13B) and works robustly for genome editing (median editing efficacy of 56.60%) (Figure 2.12B). A combination of PlmCasX-R1 and sgRNAv2 (PlmCasX-R1-v2) showed improved dsDNA cleavage kinetics *in vitro* (20-fold increase compared to PlmCasX with sgRNAv1 [PlmCasX-v1]) (Figures 2.13A and 2.13B) and showed the highest median editing efficacy (78.20%) and smallest interquartile range (18.33%) across multiple spacers compared to all other combinations of CasX and sgRNAs in HEK293 cells (Figure 2.12B). Unlike type II CRISPR nucleases like Cas9, sequence specific cis-cleavage by type V Cas12 nucleases activates non-specific ssDNA trans-cleavage[19,20,231]. Our previous data indicated that DpbCasX with sgRNAv1 (DpbCasX-v1) shows minimal trans activity compared to LbCas12a[29]. Cleavage assays investigating indiscriminate ssDNA trans-cleavage revealed that the trans-activities of the new CasX enzymes and sgRNAs remain minimal, similar to the original DpbCasX-v1 (Figures 2.13C).

We further explored the capacity of PlmCasX-R1-v2, which showed the highest editing efficacy in our fluorescent reporter assay, for endogenous genome editing by targeting the *EMX1* gene and clinically relevant *B2M* and *TTR* genes via plasmid transfection. Next generation sequencing revealed that PlmCasX-R1-v2 generated insertions and deletions (indels) at the targeted gene, and notably, showed as high as 10-fold higher activity than PlmCasX-v1 (Figure 2.12C). Interestingly, at the two endogenous targets with the highest levels of indels, PlmCasXR1-v2 generated larger indels than seen with other class II CRISPR nucleases, such as Cas9, Cas12a, or Cas12f, with the most prevalent indel being a 15 or 19 bp deletion (Figures 2.13D and 2.13E)[232,233].

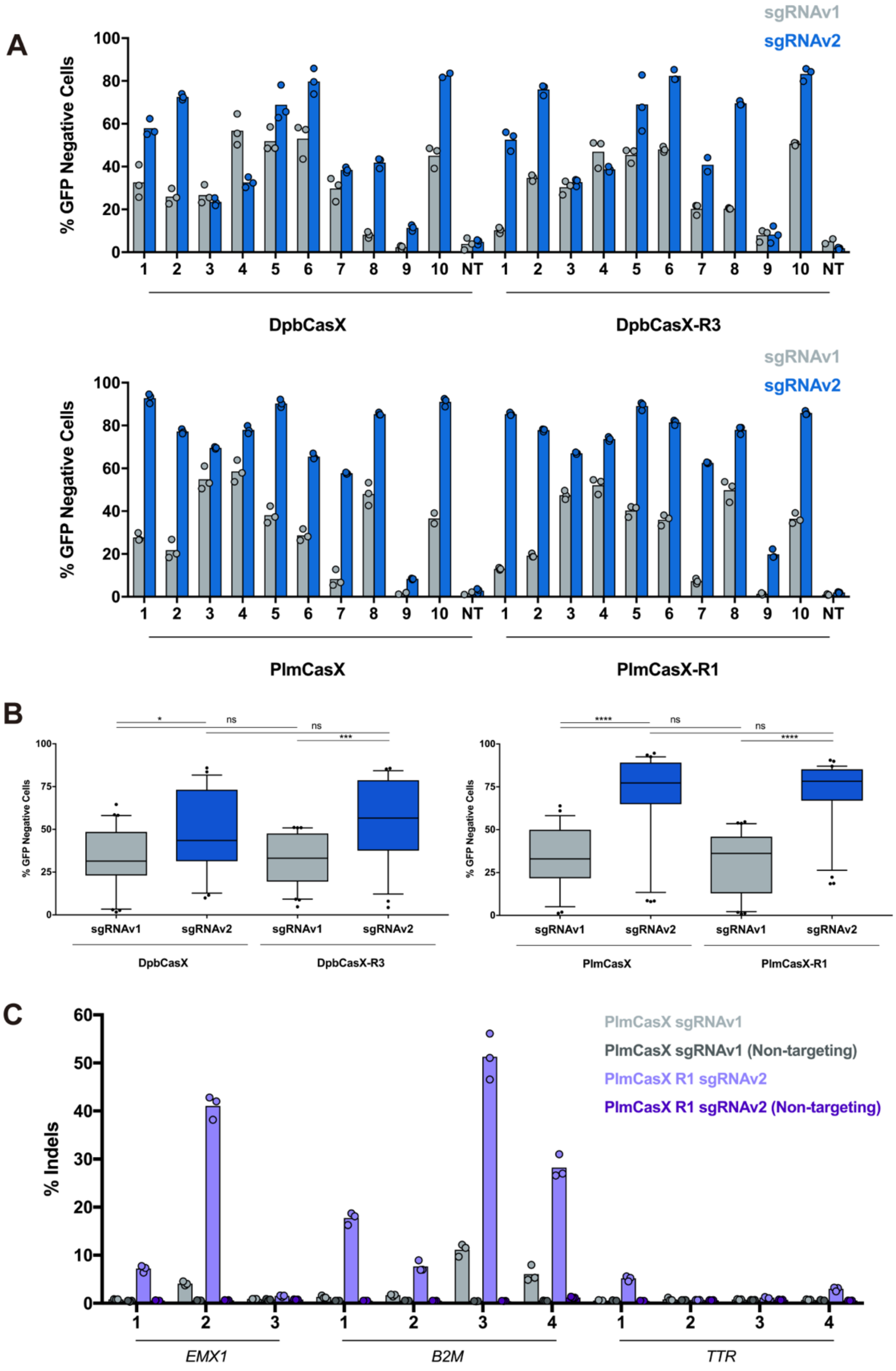


Figure 2.12: Improved genome editing by engineered DpbCasX and PlmCasX. (A) Human cell genome editing by DpbCasX, DpbCasX-R3, PlmCasX and PlmCasX-R1 using sgRNAv1 or sgRNAv2 revealed by disruption of genetically encoded GFP. The GFP disruption efficacies for all ten GFP guides are shown (n = 3 (except PlmCasX sgRNAv1 spacers 9, 10, NT; DpbCasX-R3 sgRNAv1 NT; DpbCasX sgRNAv2 spacer 10 and DpbCasX-R3 sgRNAv2 spacer 7; n = 2), mean). NT indicates the non-targeting sgRNA. (**B)** Genome editing efficacies for all ten GFP-targeting sgRNAs as a box and whisker plot: the box represents the 25th,

50th, and 75th percentile, the whiskers represent the 10th and 90th percentile, and outliers are plotted individually. Significances were determined via one-way ANOVA followed by Tukey's multiple comparisons test. ns = not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$. **(C)** Editing of the human genes *EMX1, B2M,* and *TTR* by PlmCasX-v1 or PlmCasX-R1-v2 with multiple spacer sequences in HEK293T cells. The sequences of all spacers are listed in Table 2.2.

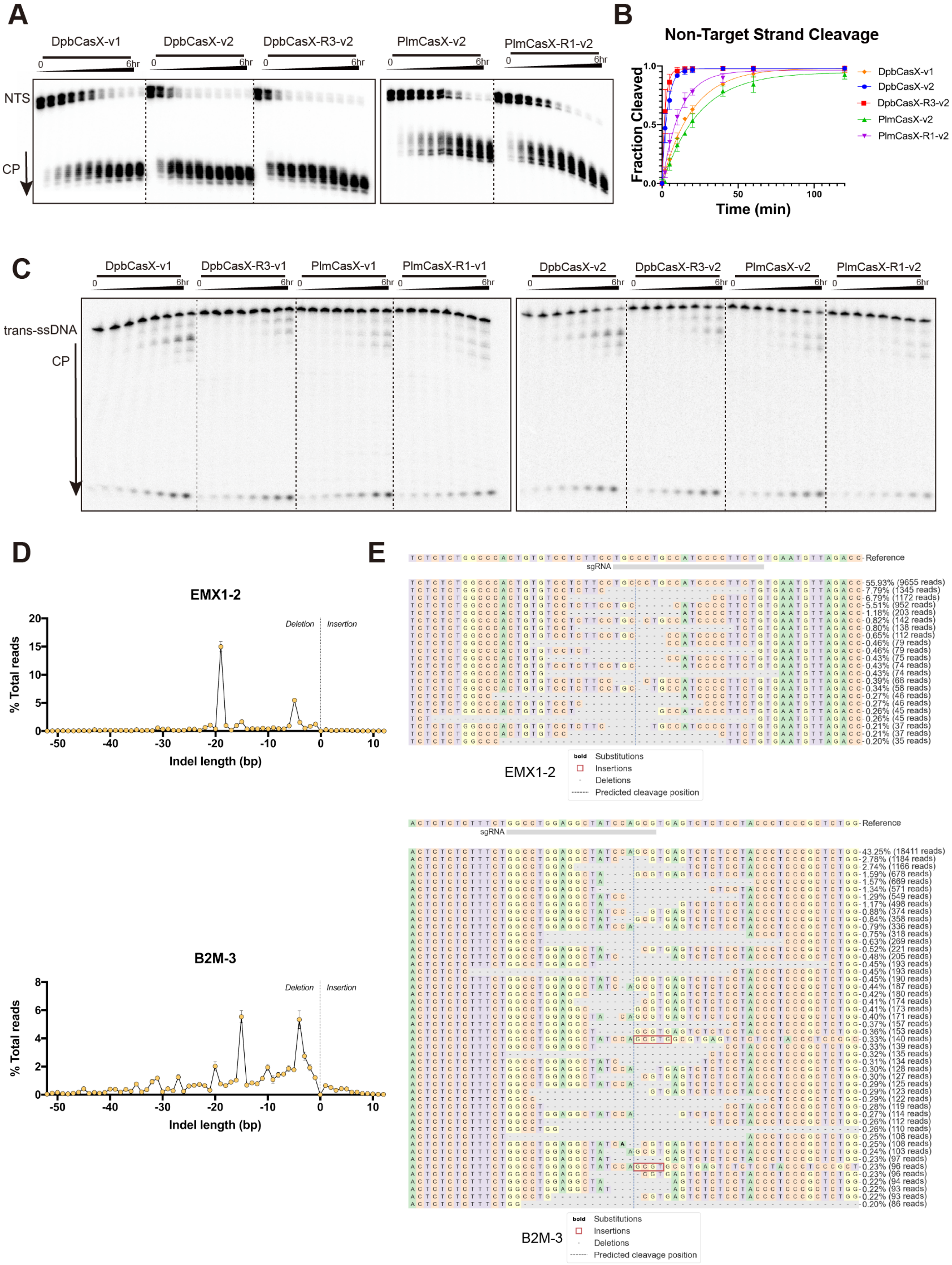

A
DpbCasX-v1
DpbCasX-v2
DpbCasX-R3-v2
PlmCasX-v2
PlmCasX-R1-v2
0
6hr
NTS
CP
B
Non-Target Strand Cleavage
Fraction Cleaved
Time (min)
0.0
0.5
1.0
0
50
100
DpbCasX-v1
DpbCasX-v2
DpbCasX-R3-v2
PlmCasX-v2
PlmCasX-R1-v2
C
DpbCasX-v1
DpbCasX-R3-v1
PlmCasX-v1
PlmCasX-R1-v1
DpbCasX-v2
DpbCasX-R3-v2
PlmCasX-v2
PlmCasX-R1-v2
trans-ssDNA
CP
D
EMX1-2
% Total reads
Deletion
Insertion
Indel length (bp)
B2M-3
E
Reference
sgRNA
bold Substitutions
Insertions
Deletions
Predicted cleavage position
EMX1-2
B2M-3

Figure 2.13: CasX nucleases with largely enhanced genome editing efficacy. (A) *In vitro* dsDNA cleavage activity comparison among different sgRNAs by wild type CasX and chimeras. NTS denotes the non-target strand DNA which was ^{32}P labeled on the 5' end. The fractions were collected at 0 min, 2 mins, 5 mins, 10 mins, 15 mins, 20 mins, 40 mins, 1 hr, 2 hrs, 4 hrs and 6 hrs, respectively. **(B)** Plot of DNA cleavage kinetics analyzed based on the NTS band density from the cleaved fractions compared to the input NTS band density at the reaction time of 0 min (n = 3, mean ± SD). The rate constant k values for DpbCasX-v1, DpbCasX-v2, DpbCasX-R3-v2, PlmCasX-v2 and PlmCasX-R1-v2 are 0.05065, 0.2817, 0.4730, 0.04182 and 0.08680 (fraction/minute), respectively. **(C)** *In vitro* trans-ssDNA cleavage activity comparison among different constructs. The trans-ssDNA was ^{32}P labeled on the 5' end. The fractions were collected at 0 min, 10 mins, 20 mins, 40 mins, 1 hr, 2 hrs, 4 hrs and 6 hrs, respectively (n = 3, representative gel shown). **(D)** Indel size distribution by PlmCasX-R1-v2 and an *EMX1*-targeting sgRNAv2 spacer 2 (top) or *B2M*-targeting sgRNAv2 spacer 3 (bottom) (n = 3, mean ± SD). Deletions are plotted as negative in length and insertions are plotted as positive in length. **(E)** Representative summary of indels generated by PlmCasX-R1-v2 and an *EMX1*-targeting sgRNAv2 spacer 2 (top) or *B2*M-targeting sgRNAv2 spacer 3 (bottom). Only indels with frequencies ≥ 0.2% of the total sequencing reads are shown for brevity.

2.4 Discussion

In this study, we explored the biochemical and structural mechanism of DNA cleavage by PlmCasX and revealed the structural differences between DpbCasX and PlmCasX that correlate with their genome editing behaviors. By designing chimeric versions of CasX and new sgRNAs, we created two significantly improved versions of CasX as a DNA editing tool (DpbCasXR3-v2 and PlmCasX-R1-v2) that offer small, yet efficient, RNA-guided nucleases. PlmCasX-R1-v2 worked robustly in human cells, showing up to 90.5% editing in our fluorescent reporter assay and 56.1% editing at an endogenous human gene. In addition, CasX may offer substantial advantages compared to other CRISPR nucleases. First, the compact size of CasX would allow for delivery via a single AAV. The safety, efficacy, and cell-specific tropism of AAVs have made them the leader for *in vivo* gene delivery, culminating in around 150 clinical trials and two FDA-approved therapies within the United States alone[234–236]. However, a major limitation for AAV delivery is the minimal DNA packaging size, which prevents the ability of encoding *S. pyogenes* Cas9 within a single vector, let alone Cas9 fused to other functional domains. The compact size and structural flexibility of CasX could also be beneficial for functional domain insertions, creating tools such as epigenetic editors and base editors that still fit within this packaging capacity[224,237,238]. Previous studies have additionally shown the presence of pre-existing humoral and cellular immunity against the commonly used Cas9 nucleases from *S. pyogenes* and *S. aureus* in patients, presumably because these enzymes originate from common human commensal or pathogenic bacteria[116,239]. While the extent to which this pre-existing immunity may be a challenge for *in vivo* genome editing has yet to be fully elucidated, CRISPR nucleases from non-human-associated sources such as CasX from *Deltaproteobacteria* or *Planctomycetes* could circumvent this potential issue. Moreover, though the off-target specificity has yet to be validated using these new CasX genome editing tools, recently published work has shown that DpbCasX has a lower mismatch tolerance compared to Cas9 and Cas12a, suggesting that DpbCasX, and likely also PlmCasX, has high fidelity and low off-target editing, an important property within the burgeoning clinical genome editing field[240].

Recently, our group and others have described additional hypercompact CRISPR-Cas nucleases, including CasΦ-2 (Cas12j; 757 aa), AsCas12f1 (422 aa), and Cas14a1 (Un1Cas12f1; 537 aa). While CasΦ-2 is smaller than PlmCasX (984 aa), the CasΦ-2 nuclease showed only up to 33% editing of GFP in our fluorescent reporter assay compared to PlmCasX-R1-v2, which reached as high as 90.5% editing[20]. Extensive engineering of both the protein and sgRNA has dramatically increased the editing seen by Cas12f nucleases, which represent some of the smallest CRISPR effectors to date[233,241,242]. However, the editing efficacy at endogenous genes by Cas12f in human cells varied greatly, ranging from a mean of 5% to 26%, which is comparable to the mean of 15% seen with PlmCasX-R1-v2. Regarding delivery of the protein and sgRNA as an RNP, recent work has demonstrated that the Cas12f nuclease functions as an asymmetric homodimer to cleave dsDNA, which makes the effective RNP complex similar in size to CasX (Cas12f dimer: 800–1000 aa; CasX monomer: 984 aa)[243,244]. Similar sequence-wide, high-throughput screening approaches to further engineer the CasX protein and sgRNA have yet to be explored. Based on the success of this strategy with other CRISPR nucleases, we anticipate this could be a promising approach to further minimize the size and improve the editing efficacy of CasX.

2.4.1 Limitations of the study

In this study, structural analysis revealed three nucleotide binding loops that contribute to DNA cleavage by CasX. While chimeric designs with the R1 or R3 loop insertion increased the DNA cleavage activity of CasX, the R2 loop insertion eliminated activity. Our current design of the R2 insertion is therefore non-optimal, and most likely disrupted the interaction between PlmCasX and the PAM DNA instead of stabilizing the interaction. In future work, the R2 region could be a potential spot for further improvement of PlmCasX by rational design or directed evolution screening. Additionally, our data suggests sgRNAv2 stabilizes the R-loop (ternary) complex and increases DNA cleavage activity by CasX; however, the mechanism by which sgRNAv2 affects the RNP (binary) complex assembly and thereby DNA unwinding and loading is unknown. We hypothesized the increase in flexibility of sgRNAv2 compared to sgRNAv1 was responsible for the significant improvement in activity, though more detailed studies are required to explore the structural states of the sgRNAs alone. Finally, while PlmCasX-R1-v2 proved to be a significantly improved genome editor at a fluorescent reporter gene and at endogenous genes, all experiments performed in this study were done in transformed cell lines. Future work is needed to test these improved versions of CasX within more difficult environments such as primary cells or animal models, along with delivery by methods such as AAV.

2.5 Materials and methods

2.5.1 Culture of human cell lines

GFP HEK293 and HEK293T cells (UC Berkeley Cell Culture Facility) were cultured in Dulbecco's Modified Eagle Medium (DMEM) (GIBCO) supplemented with 10% fetal bovine serum (FBS) (VWR) and 1% penicillin-streptomycin (GIBCO). Cells were maintained at 37 °C and 5% CO2 at sub-confluent conditions. The MycoAlert Mycoplasma Detection Kit (Lonza) was used to routinely test cells for mycoplasma.

2.5.2 CasX protein expression and purification

The wildtype and engineered CasX proteins were expressed using Rosetta E.coli cells. The E. Coli cells were transformed by mixing competent cells (100 mL) with CasX encoding plasmids (100 ng) and incubating for 30 min on ice. The tube containing the plasmid and cells was incubated at 42 °C for 35 s to induce the transformation. After 5 min of resting on ice, Luria broth (LB) (950 mL) was added to the solution and incubated at 37 °C for 1 h to recover. The cells were transferred to a flask containing LB and 50 mg/mL ampicillin (1:1000) and incubated at 37 °C overnight. 2.7% of the grown culture was added to the main culture containing Terrific broth and 50 mg/mL ampicillin (1:1000). The main culture was incubated at 37 °C until it reached an OD of 0.5-0.6. The culture was cooled on ice and protein expression was induced by addition of IPTG to a final concentration of 1 mM and incubated at 16 °C overnight. Cells were harvested by centrifugation (4000 rpm, 4C) and resuspended in lysis buffer (600 mM sodium chloride, 20 mM HEPES, pH 7.5, 10% glycerol, 50 mM imidazole, 1 mM TCEP). PMSF (0.5 mM) and 4 tablets of protease inhibitor cocktail (Roche) were added per 100 mL of mixture. The cell suspension was lysed by sonication and pelleted by ultra-centrifugation at 35,000 g for 30 min. The soluble lysate was mixed with equilibrated Ni-NTA agarose beads at 4 °C for 2 h. Using a gravity-flow column, the Ni-NTA agarose beads were washed using lysis buffer. To elute the construct, the Ni-NTA beads were incubated overnight at 4 °C in lysis buffer and TEV protease (final concentration of 1 mg protease/ 20 mg purified protein). Using the gravity-flow column, the protein of interest was eluted using lysis buffer with 300 mM imidazole. The flow-through was collected and concentrated using a 30 kDa MWCO concentrator (Amicon Ultra, Merck). The solution containing the protein was mixed with lower salt buffer (200 mM sodium chloride, 20 mM HEPES, pH 7.5, 10% glycerol, 1 mM TCEP) and applied to a heparin column on an Akta FPLC (GE). The protein was eluted using a potassium chloride gradient up to 1 M. The combined fractions were concentrated using a 30 kDa MWCO concentrator (Amicon Ultra, Merck) and applied to a Superdex 200 10/300 column (GE healthcare/Cytiva) using SEC buffer (400 mM potassium chloride, 20 mM HEPES, pH 7.5, 10% glycerol, 1 mM TCEP). The protein was concentrated, flash frozen in liquid nitrogen and stored at -80C to use in assays. All the engineered and wildtype CasX proteins were expressed and purified using the same method. Compared to the original purification protocol for DpbCasX, we increased the sodium chloride concentration from 500 mM to 600 mM, and added 50 mM imidazole in the lysis buffer, which helped to decrease the non-specific protein and nucleic acid contamination. Since apo PlmCasX is less stable than DpbCasX, it should be kept in buffer with ≥400 mM sodium chloride or potassium chloride during the entire purification process, and the purification should be ideally finished within 24 h.

2.5.3 sgRNA preparation

All the sgRNAs were produced using in vitro transcription. First, to make the DNA template, primers were ordered from Integrated DNA Technologies (IDT) and PCR amplified using Q5 polymerase (New England Biolabs). The DNA template (50 mg) along with 10x IVT buffer (300 mM Tris-HCl (pH 8.1), 250 mM MgCl2, 0.1% Triton X-100, 20 mM Spermidine, add 100 mM DTT before use), 5x NTPs (25 mM NTP mixture, pH 7.5), T7 polymerase, RNase inhibitor (Promega) and DEPC-treated water were incubated on a 37 °C heat block for 3-4 h. The solutions were then treated with RNase-Free DNase I (Promega) by addition of 10x Reaction Buffer (400 mM Tris-HCl, pH 8.0 at 25C, 100 mM MgSO4, 10 mM CaCl2) and RNase-free DNase I and incubated on a 37 °C heat block for 30 min. The sample was spun down at 4 °C and the soluble fraction was moved to a new tube. After adding 2x formamide (95% formamide, 0.02% SDS, 0.02% bromophenol blue, 0.01% xylene cyanol FF, 1 mM EDTA), samples were gel purified using a 15% urea-PAGE gel. The band containing the sgRNA was cut out and incubated in water and 1/30 NaOAc at 4 °C overnight. Samples were then filtered using a 0.22 mm Corning filter into 50 mL tubes. sgRNA samples were concentrated using a 3 kDa MWCO concentrator (Amicon Ultra, Merck). 100% ethanol was added to sgRNA samples to precipitate the sgRNA. Precipitated sgRNA was pelleted via centrifugation and washed using 70% ethanol. sgRNA samples were resuspended in DEPC-treated water and stored at -80C to be used for cleavage assays.

2.5.4 *In vitro* cleavage assays

For dsDNA cleavage assays, DNA substrates were 50 labeled using T4 PNK (NEB) by adding ɣ-32P-ATP. CasX proteins were diluted to 2 mM using 1x reaction buffer (400 mM potassium chloride, 5% glycerol, 20 mM Tris-HCl, pH 7.8, 5 mM magnesium chloride, 1 mM DTT). sgRNAs were diluted to 3 mM with 1x reaction buffer. The sgRNA and protein samples were then mixed and incubated at room temperature for 1 h to reconstitute the RNP complex. The final concentration of the CasX-sgRNA was 300 nM and the concentration of radiolabeled probe was 2 nM. Reactions were initiated by mixing CasX-sgRNA and radiolabeled DNA on a 37 °C heat block. Sample aliquots were taken at the following time points: 0, 2, 5, 10, 15, 20, 30, 60, 120, 240, and 360 min. The aliquots were mixed with 2x formamide loading buffer (95% formamide, 10 mM EDTA, 0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene cyanol FF) and quencher (50 mg/mL heparin, 25 mM EDTA) and were incubated in 95 °C heat blocks for 5 min to stop the cleavage reaction. Samples were run on 12% urea-PAGE gels before being dried and visualized using a phosphoimager (Amersham Typhoon, GE Healthcare).

For plasmid cleavage assays, the target DNA sequence was cloned into the pUC19 plasmid. For each 100 mL cleavage reaction, 400 nM CasX-sgRNA RNP and 20 nM pUC19 plasmid DNA were incubated in 1x reaction buffer (500 mM sodium chloride, 5% glycerol, 20 mM Tris-HCl, pH 7.8, 10 mM magnesium chloride, 1 mM DTT) at 37 °C. Sample aliquots were taken at the following time points: 0, 10, 20, 30, 60, 120, 240, and 360 min. The aliquots were mixed with 6x DNA loading buffer (30 mM EDTA, 36% (v/v) glycerol, 0.05% (w/v) bromophenol blue, 0.05% (w/v) xylene Cyanol FF) and then digest with 100 mg/mL Proteinase K (from *Tritirachium album*, Sangon Biotech) for 1 h at 37 °C to quench the reaction. A 1% agarose gel was used to analyze cleavage products.

For the trans-cleavage activity assay, a random 50 nucleotide oligonucleotide substrate was labeled using T4 PNK (NEB) by adding ɣ-32P-ATP. Each reaction included 300 nM CasX protein, 360 nM sgRNA, 450 nM activator, and 2 nM substrate. The trans-cleavage assay was performed and analyzed similarly to the dsDNA cleavage assay, described above.

2.5.5 Plasmid construction

For human genome editing experiments, DpbCasX plasmid pBLO62.4 (Addgene plasmid #123123) and PlmCasX plasmid pBLO62.5 (Addgene plasmid #123124) were utilized or modified, which were codon-optimized for expression in human cells and contain a SV40 nuclear localization sequence on both termini[29]. Short oligonucleotides (IDT) containing the sgRNA spacer sequence were annealed and phosphorylated prior to Golden Gate assembly (BbsI restriction sites) for insertion just downstream of the CasX guide RNA scaffold within the plasmids. CasX protein mutants were constructed by PCR amplification of the CasX sequence in two pieces, with primers containing the deletion or insertion sequences. pBLO62.4 and pBLO62.5 were digested with AgeI and BamHI (NEB) and gel electrophoresis was utilized to separate the digested components. The plasmid backbone was excised from the gel and purified with the QIAquick PCR Purification Kit (QIAGEN) or the NucleoSpin Gel and PCR Cleanup Kit (Takara) according to the manufacturer's protocol. In-Fusion cloning (Takara) with the Cloning Enhancer was used to insert the PCR amplified mutant CasX sequences within the digested backbone according to the manufacturer's protocol. Plasmids encoding mutant CasX sgRNA scaffolds were constructed similarly to CasX mutant protein plasmids. Plasmids encoding engineered or wildtype CasX proteins were digested using KpnI and PciI (NEB). Gel electrophoresis was used to isolate the digested plasmid backbone. Digested backbone was excised from the gel and purified with the PCR QIAquick PCR Purification Kit (QIAGEN) or the NucleoSpin Gel and PCR Cleanup Kit (Takara) according to the manufacturer's protocol. Mutant sgRNA scaffolds were ordered as gBlocks from IDT and cloned into the digested backbone using In-Fusion cloning (Takara). Cloned plasmids were sequence verified by capillary Sanger sequencing (UC Berkeley DNA Sequencing Facility). For the endogenous genome editing experiments, an mNeonGreen fluorescent protein was genetically encoded between the CasX gene and puromycin resistance gene, each separated by self-cleaving 2A peptide sequences. Plasmids were cleaved with BamHI and In-Fusion cloning was utilized as described above to insert a gBlock (IDT) encoding mNeonGreen. Plasmids were propagated in Mach1 T1 competent cells (Thermo Fisher) and purified using a QIAprep Spin Miniprep kit (QIAGEN) according to the manufacturer's protocol.

2.5.6 Genome editing in fluorescent reporter human cells

GFP HEK293 reporter cells were seeded into 96-well plates and transfected 12-18 h later at 60%–70% confluency according to the manufacturer's protocol with lipofectamine 3000 (Life Technologies) and 200 ng of plasmid DNA encoding the wildtype or engineered CasX plasmids. 24 h post-transfection, GFP HEK293 reporter cells that were successfully transfected were selected for by adding 1.5 mg/mL puromycin to the cell culture media for 48 h. Cell culture media was replaced with media containing fresh 1.5 mg/mL puromycin for an additional 24 h before replacing with cell culture media without puromycin. Cells were passaged regularly to maintain sub-confluent conditions and then analyzed in 96-well round bottom plates on an Attune NxT Flow Cytometer with an autosampler. Cells were analyzed on the flow cytometer after 5, 7, and 10 days to track the disruption of the GFP gene in cells. The sequences of all spacers used in this study are listed in Table 2.2.

2.5.7 Endogenous genome editing

HEK293T cells (UC Berkeley Cell Culture Facility) were cultured in DMEM (GIBCO) supplemented with 10% FBS (VWR) and 1% penicillin-streptomycin (GIBCO). The MycoAlert Mycoplasma Detection Kit (Lonza) was used to routinely test cells for mycoplasma. HEK293T cells were plated in 96-well plates and allowed to grow overnight to 60%–70% confluency before transfecting with 200 ng of plasmid and lipofectamine 3000 according to the manufacturer's protocol. 24 h post-transfection, HEK293T cells that were successfully transfected were selected for by adding 1.5 mg/mL puromycin to the cell culture media for 48 h. Cell culture media was replaced with media containing fresh 1.5 mg/mL puromycin for an additional 24 h before replacing with cell culture media without puromycin. Media was removed from the cells and 50 mL of QuickExtract (Lucigen) was added to each well and incubated at room temperature for 10-15 min. Cell extracts were then thermocycled at 65 °C for 20 min followed by 95 °C for 20 min. Amplicons containing the targeted site were amplified via PCR with Q5 polymerase (NEB) and primers containing Illumina adaptor sequences. Amplicons were cleaned with magnetic solid phase reversible immobilization (SPRI) beads (UC Berkeley Sequencing Core) and were further library prepped and loaded onto an Illumina MiSeq by the Center for Translational Genomics (Innovative Genomics Institute, UC Berkeley). Over 20,000 reads per sample were routinely achieved. 300 bp paired-end reads were analyzed using CRISPResso2 (https://crispresso.pinellolab.partners.org/submission), using a quantification window centered at -3 bp, a quantification window size of 8 bp (to account for the large, staggered cleavage pattern of CasX), and a plot window size of 30 bp (to visualize large indels)[53]. Cells treated with PlmCasX-v1 or PlmCasX-R1-v2 with a non-targeting sgRNA were evaluated at every spacer sequence within every amplicon as a control. Percentage of indels plotted was based on the percentage of modified reads from the CRISPResso2 output. For the indel size distribution plots, sequencing reads of a particular deletion length (regardless of insertions or substitutions) were grouped and plotted. The remaining reads were grouped and plotted based on insertion length (regardless of substitutions). For clarity, unmodified reads (indel length of 0 bp) were plotted as 0% of the total reads. The sequences of all spacers used in this study are listed in Table 2.2.

2.5.8 Cryo-EM sample preparation and data collection

The PlmCasX-sgRNA complex was assembled by incubating protein with a 1.25-fold excess of sgRNA for 30 min at room temperature. The ternary complexes were assembled by incubating dPlmCasX-sgRNA with a 1.5-fold excess of annealed dsDNA target for 30 min at room temperature. After the complexes were assembled, they were purified by size-exclusion chromatography using a Superdex 200 10/300 column. PlmCasX complexes at 10 mM concentration in a buffer containing 20 mM HEPES, pH 7.5, 300 mM potassium chloride, 1 mM DTT, and 0.25% glycerol were aliquoted and stored in liquid nitrogen for further usage.

For EM sample preparation of dPlmCasX-sgRNAv1-dsDNA, the complex (final concentration 1 mM) was mixed with BS3 crosslinker (final concentration 1 mM) and incubated on ice for 1 h. 3.7 µL droplets of the sample were placed onto Quantifoil grids (1.2/1.3 µm) with freshly coated graphene-oxide film[245]. After a 1 min incubation, the grids were blotted for 3 s with a blot force of 4 and immediately plunged into liquid ethane using a FEI Vitrobot MarkIV maintained at 8 °C and 100% humidity. Data was acquired using a Thermo Fisher Titan Krios transmission electron microscope operated at 300 keV with an energy filter (GIF quantum 1967), and images were taken at a nominal magnification of x135,000 (0.9 Å pixel size) with defocus ranging from -0.7 to -2.1 µm. Micrographs were recorded using SerialEM on a Gatan K3 Summit direct electron detector operated in super-resolution mode[246]. We collected a 5 s exposure fractionated into 50, 100 ms frames with a dose of 10 $e\text{-}Å^{-2}s^{-1}$. In total, 8,675 movies were collected for this sample.

For EM sample preparation of dPlmCasX-sgRNAv2-dsDNA, complex (non-crosslinked) at a concentration of 5 mM was used. Immediately after glow-discharging the grid for 14 s using a Solaris plasma cleaner, 3.6 µL droplets of the sample were placed onto C-flat grids (2/2 mm). The grids were blotted for 4 s with a blot force of 8 and rapidly plunged into liquid ethane using a FEI Vitrobot MarkIV maintained at 8 °C and 100% humidity. Data was acquired by following the same protocol as described above but using 3 exposures per hole. In total, 4,171 movies were collected for this sample.

2.5.9 Single particle cryo-EM analysis

46 frames (the first 2 and last 2 frames were skipped) of each image stack in super-resolution mode were aligned, decimated, summed and dose-weighted using Motioncor2[247]. They were then imported into cryoSparc[248] for patched CTF estimation and particle picking using 2D class-averages of DpbCasX from our previous study[29] as templates. 3,652,583 raw particles were picked from the dPlmCasX-sgRNAv1-dsDNA dataset, and 1,764,600 particles were picked from the dPlmCasX-sgRNAv2-dsDNA dataset. Particle extraction, ab-initio reconstruction, and 3D classification were performed without 2D classification. Good models from 3D classification were further refined using homogeneous refinement. In cases when the post-processing in cryoSparc over-sharpened the map, half-maps generated by cryoSparc were imported into Relion[249] for post-processing. The workflows and more details are summarized in Figures 2.3 and 2.5.

2.5.10 Atomic model building and refinement

For dPlmCasX-sgRNAv1-dsDNA, an initial model of PlmCasX was first constructed using homology modeling in the Swiss-model server with the DpbCasX structure (PDB:6NY2) as reference. The sgRNAv1-DNA part was adopted from the DpbCasX structure (PDB:6NY2) with manual revision in Coot. The two parts were fitted into the density map of State I of dPlmCasX-sgRNAv1-dsDNA (2.9 Å resolution) and then manually modified in Coot to better fit the density. The entire model was subjected to PHENIX real space refinement (global minimization and ADP refinement) with secondary structure, Ramachandran, rotamer, and nucleic-acid restraints[250]. The final model was validated using Molprobity[251]. The atomic model of dPlmCasXsgRNAv1-dsDNA State II was obtained by running flexible fitting of the State I atomic model against the State II cryo-EM map (3.4 Å resolution) with secondary structure restraints using MDFF[252]. The output model was manually rebuilt in Coot[253] and PHENIX real space refinement was used to improve backbone geometry. The State III atomic model was directly adopted from State I by deleting the Helical-II domain, followed by PHENIX real space refinement against the State III cryo-EM map (3.2 Å resolution).

For model building of dPlmCasX-sgRNAv2-dsDNA in State I, the dPlmCasX-sgRNAv1-dsDNA model in State I was used as the starting model. Then, the sgRNA sequence was modified and the structures were manually rebuilt in Coot. PHENIX real space refinements against dPlmCasX-sgRNAv2-dsDNA EM maps were used to improve the models. The final model was validated using MolProbity.

2.5.11 Quantification and statistical analysis

All statistical analysis was performed using GraphPad Prism 7. The number of independent technical replicates (n) for each experiment is listed in the respective figure legends. For cleavage kinetics plots, error bars represent the standard deviation between replicates and the data were fitted using one-phase association to yield the single turnover rate constant k values (fraction cleaved per minute). For cellular editing bar plots, individual technical replicates (n) were plotted with the bar representing the mean. For box and whisker plots the box represents the 25th, 50th, and 75th percentile, the whiskers represent the 10th and 90th percentile, and outliers are plotted individually. Significances were determined via one-way ANOVA followed by Tukey’s multiple comparisons test. ns = not significant, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and ****$p < 0.0001$.

2.6 Accession codes

The electron density maps have been deposited to the Electron Microscopy Data Bank (EMDB) under the accession numbers EMDB: EMD-32389, EMD-32390, EMD-32391, and EMD-32392 and are publicly available as of the date of publication. The atomic coordinates and structural data have been deposited to the Protein Data Bank (PDB) under the accession numbers PDB: 7WAY, 7WAZ, 7WB0, and 7WB1 and are publicly available as of the date of publication. All the accession numbers are also listed in Table 2.1. The raw cryo-EM micrographs and movies used in this study are available from the lead contact upon request.

2.7 Acknowledgments

EM data were collected at the Tsinghua Cryo-EM facility and the Cal-Cryo facility at UC Berkeley. The data were analyzed using the Bio-Computation platform at the Tsinghua University branch of the Chinese National Center for Protein Sciences (Beijing). We thank D.B. Toso, J.L. Lei, and X.M. Li for expert electron microscopy assistance. We thank T. Yang, Y.K. Wang, A. Chintangal, and P. Tobias for computational support. We thank J. Hurtado for providing the GFP HEK293 cell line. We thank N. Krishnappa and the Center for Translational Genomics (Innovative Genomics Institute, UC Berkeley) for assistance with Illumina sequencing. We thank Y. Xue and X.Y. Fang for help analyzing the sgRNA structure. This project was supported by the National Natural Science

Foundation of China (NSFC) (grant no. 32150018), Chunfeng Fund (project no. 2021Z99CFY020) and start-up funds from Tsinghua University, Beijing (J.J.G.L.); National Science Foundation (United States) grant no. 1244557 (J.A.D.); National Institutes of Health (NIH) (United States) grant no. P01GM051487 (J.A.D. and E.N.); NSFC grant no. 91957119, 91954001, and 32125021 (X.W.C.). C.A.T. is supported by Campus Executive Grants 2101705 and 1655264 through Sandia National Laboratories and an NIH Ruth L. Kirschstein National Research Service Award F31 Pre-doctoral Fellowship (United States) (National Heart, Lung, and Blood Institute, 1F31HL156468-01). J.A.D. and E.N. are Howard Hughes Medical Institute Investigators.

2.8 Author contributions

J.J.G.L., J.A.D., and E.N. supervised the project. J.J.G.L., C.A.T., and X.W.C. designed the experiments. Y.Z., M.S.D., J.C., and J.B. prepared the CasX and RNP complexes. Y.Z., M.S.D., and D.L. carried out the biochemical assays. C.A.T., M.S.D., Y.Z., E.O., Z.Y.H., H.F., and A.V. cloned CasX plasmids. C.A.T. performed the mammalian cell experiments and prepared samples for next-generation sequencing. J.J.G.L., S.Z., and J.W. carried out the cryo-EM analysis. S.Z. and C.P.L. built the atomic structures. D.B. and S.Z. did the analysis of the tracrRNA sequence. C.A.T., S.Z., M.S.D., Y.Z., E.N., J.A.D., and J.J.G.L. wrote the manuscript with revision from all authors.

2.9 Supplementary information

Complex and State	dPlmCasX-sgRNAv1-dsDNA State I	dPlmCasX-sgRNAv1-dsDNA State II	dPlmCasX-sgRNAv1-dsDNA State III	dPlmCasX-sgRNAv2-dsDNA State I	dPlmCasX-sgRNAv2-dsDNA State II	dPlmCasX-sgRNAv2-dsDNA State III
EMDB code	EMD-32389	EMD-32390	EMD-32391	EMD-32392	N/A	N/A
PDB code	7WAY	7WAZ	7WB0	7WB1	N/A	N/A
Data collection and processing						
Magnification	135,000	135,000	135,000	135,000	135,000	135,000
Voltage (kV)	300	300	300	300	300	300
Electron exposure (e-/Å^2)	~50	~50	~50	~50	~50	~50
Defocus range (μm)	0.5~2.0	0.5~2.0	0.5~2.0	0.5~2.0	0.5~2.0	0.5~2.0
Pixel size (Å)	0.94	0.94	0.94	0.94	0.94	0.94
Symmetry imposed	C1	C1	C1	C1	C1	C1
Final particle images (no.)	520,115	502,778	710,824	616,493	267,147	143,849
Map resolution (Å) **FSC threshold**	2.9 at 0.143	3.4 at 0.143	3.2 at 0.143	3.7 at 0.143	N/A	N/A
Map resolution range (Å)	2.5~6	3~7	3~7	3~7	N/A	N/A
Refinement						

Model resolution (Å)	2.9	3.4	3.2	3.7		
FSC threshold	0.143	0.143	0.143	0.143		
Model resolution range (Å)	2.5~6	3~7	3~7	3~7		
Map sharpening *B* factor (Å^2)	-70	-120	-100	-137		
Model composition						
Non-hydrogen atoms	11453	11034	10096	11524		
Protein residues	960	952	797	960		
Nucleotides	175	157	175	178		
***B* factors-Mean (Å^2)**						
Protein	65.72	116.41	83.97	123.99		
Nucleotide	106.17	161.06	157.90	157.80		
R.m.s. deviations						
Bond lengths (Å)	0.004	0.002	0.003	0.003		
Bond angles (°)	0.542	0.589	0.579	0.590		
Validation						
MolProbity score	1.45	2.43	1.65	1.90		
Clashscore	5.49	11.37	7	11.07		
Poor rotamers (%)	0.00	5.46	0	0		
Ramachandran plot						
Favored (%)	97.17	95.69	96.57	95.18		
Allowed (%)	2.83	4.10	3.3	4.72		
Disallowed (%)	0.00	0.21	0.1	0.1		

Table 2.1: Cryo-EM data collection, refinement, and validation statistics

Experiment	**Name**	**Sequence**
Preparation of the ternary complex for cryoEM	TS DNA	ATCGTTATACTTTGATTTTCTGCTGCAGGATGAAATCCCG
Preparation of ternary complex for cryoEM	NTS DNA	CGGGATTTCATCCTGCAGCATCCCCGACCCGTATAACGAT
sgRNAv1 used for EM complex preparations, and cleavage assays	sgRNAv1	ggCGCGUUUAUUCCAUUACUUUGGAGCCAGUCCCAGCGACUAUGUCGUAUGGACGAAGCGCUUAUUUAUCGGAGAGAAACCGAUAAGUAAAACGCAUCAAAGUCCUGCAGCAGAAAAUCAAA
In vitro DNA cleavage	NTS DNA	GCCCGCGGGATTTCATCCTGCAGCAGAAAATCAAAGACAATGAATATTTCGGCGC
In vitro DNA cleavage	TS DNA	GCGCCGAAATATTCATTGTCTTTGATTTTCTGCTGCAGGATGAAATCCCGCGGGC
The secondary architecture of sgRNAv1	sgRNAv1	ggCGCGUUUAUUCCAUUACUUUGGAGCCAGUCCCAGCGACUAUGUCGUAUGGACGAAGCGCUUAUUUAUCGGAGAGAAACCGAUAAGUAAAACGCAUCAAAG
The secondary architecture of sgRNAv1	sgRNAv1-2	UggCGCGUUUAUUCCAUUACUUUGGAGCCAGUCCCAGCGACUAUGUCGUAUGGACGAAGCGCUUAUUUAUCGGAGAGAAACCGAUAAGUAAAACGCAUCAAAG
The secondary architecture sgRNAv2	sgRNAv2	ggCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAAACCGAUAAAUAAGAAGCAUCAAAG
The secondary architecture sgRNAv2	sgRNAv2-2	UggCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAAACCGAUAAAUAAGAAGCAUCAAAG
The secondary architecture sgRNAv2	sgRNAv2-3	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAAACCGAUAAAUAAGAAGCAUCAAAG
sgRNAv2 used for cleavage assays	sgRNAv2	ggCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAAACCGAUAAAUAAGAAGCAUCAAAGUCCUGCAGCAGAAAAUCAAA
trans-cleavage assay	Random ssDNA	GTTTATTTACTTTAGTCACTCCAGGATTCCAATAGATATTTACTTTGAAG
CasX human GFP targeting	sgRNA spacer 1	CCGGGGTGGTGCCCATCCTG
CasX human GFP targeting	sgRNA spacer 2	GCGTGTCCGGCGAGGGCGAG
CasX human GFP targeting	sgRNA spacer 3	GGGTCAGCTTGCCGTAGGTG
CasX human GFP targeting	sgRNA spacer 4	TCTGCACCACCGGCAAGCTG
CasX human GFP targeting	sgRNA spacer 5	GCCGCTACCCCGACCACATG
CasX human GFP targeting	sgRNA spacer 6	GGCATGGCGGACTTGAAGAA
CasX human GFP targeting	sgRNA spacer 7	CCTCGGCGCGGGTCTTGTAG
CasX human GFP targeting	sgRNA spacer 8	AGGGCGACACCCTGGTGAAC
CasX human GFP targeting	sgRNA spacer 9	GCTCGATGCGGTTCACCAGG
CasX human GFP targeting	sgRNA spacer 10	AGGAGGACGGCAACATCCTG
CasX human GFP targeting	Non-targeting	CGTGATGGTCTCGATTGAGT
CasX human *EMX1* targeting	sgRNA spacer 1	CTCTGGCCCACTGTGTCCTC
CasX human *EMX1* targeting	sgRNA spacer 2	CAGAAGGGGATGGCAGGGCA

CasX human *EMX1* targeting	sgRNA spacer 3	CCTGGGCCAGGGAGGGAGGG
CasX human *B2M* targeting	sgRNA spacer 1	GGAATGCCCGCCAGCGCGAC
CasX human *B2M* targeting	sgRNA spacer 2	TGAAGCTGACAGCATTCGGG
CasX human *B2M* targeting	sgRNA spacer 3	GGCCTGGAGGCTATCCAGCG
CasX human *B2M* targeting	sgRNA spacer 4	TCTCCCGCTCTGCACCCTCT
CasX human *TTR* targeting	sgRNA spacer 1	CCGGTGCCCTGGGTGTAGAG
CasX human *TTR* targeting	sgRNA spacer 2	AGATGCTGTCCGAGGCAGTC
CasX human *TTR* targeting	sgRNA spacer 3	GAACACATGCACGGCCACAT
CasX human *TTR* targeting	sgRNA spacer 4	GAAAGGCTGCTGATGACACC
CasX human *EMX1*, *B2M*, and *TTR* targeting	Non-targeting	AAGTAAAACCTCTACAAATG
CasX human *EMX1*, *B2M*, and *TTR* targeting	*EMX1* Illumina library prep fwd primer (with **adaptor sequence**)	**GCTCTTCCGATCT**CAGCTCTGTGACCCTTTGTTTG
CasX human *EMX1*, *B2M*, and *TTR* targeting	*EMX1* Illumina library prep rev primer (with **adaptor sequence**)	**GCTCTTCCGATCT**CTGCCGTTTGTACTTTGTCCTC
CasX human *EMX1*, *B2M*, and *TTR* targeting	*B2M* Illumina library prep fwd primer (with **adaptor sequence**)	**GCTCTTCCGATCT**CTTGGAGACAGGTGACGGTC
CasX human *EMX1*, *B2M*, and *TTR* targeting	*B2M* Illumina library prep rev primer (with **adaptor sequence**)	**GCTCTTCCGATCT**GGGCCACCAAGGAGAACTTG
CasX human *EMX1*, *B2M*, and *TTR* targeting	*TTR* Illumina library prep fwd primer (with **adaptor sequence**)	**GCTCTTCCGATCT**AGTGTGTAATTCTTGTTTCGCTCCA
CasX human *EMX1*, *B2M*, and *TTR* targeting	*TTR* Illumina library prep rev primer (with **adaptor sequence**)	**GCTCTTCCGATCT**CAAGTGAGGGGCAAACGG

Table 2.2: DNA/RNA sequences used in this study

CHAPTER 3

Systematic analysis and mitigation of Cas9-induced chromosome loss in clinical T cells

A portion of the work presented in this chapter has been submitted for publication as part of the following paper: Tsuchida C.A.*, Brandes N.*, Bueno R.*, Trinidad M., Mazumder T., Yu B., Hwang B., Chang C., Liu J., Sun Y., Hopkins C.R., Parker K.R., Qi Y., Satpathy A.T., Stadtmauer E.A., Cate J.H.D., Eyquem J., Fraietta J.A., June C.H., Chang H.Y., Ye C.J., Doudna J.A. Mitigation of chromosome loss in clinical CRISPR-Cas9-engineered T cells. *BioRxiv* (2023).

*These authors contributed equally.

3.1 Abstract

CRISPR-Cas9 genome editing has enabled advanced T cell therapies, but occasional loss of the targeted chromosome remains a safety concern. To investigate whether Cas9-induced chromosome loss is a universal phenomenon and evaluate its clinical significance, we conducted a systematic analysis in primary human T cells. Arrayed and pooled CRISPR screens revealed that chromosome loss was generalizable across the genome and resulted in partial and entire loss of the chromosome, including in pre-clinical chimeric antigen receptor T cells. T cells with chromosome loss persisted for weeks in culture, implying the potential to interfere with clinical use. A modified cell manufacturing process, employed in our first-in-human clinical trial of Cas9-engineered T cells[101], dramatically reduced chromosome loss while largely preserving genome editing efficacy. Expression of p53 correlated with protection from chromosome loss observed in this protocol, suggesting both a mechanism and strategy for T cell engineering that mitigates this genotoxicity in the clinic.

3.2 Introduction

The precision of CRISPR-Cas9 genome editing is paramount to ensure clinical safety and avoid unintended and permanent genotoxicities. Promiscuous off-target genome editing at unintended sites has been extensively studied[28,189,254] and mitigated[198,199] *in vitro* and *in vivo*. However, unintended chromosomal abnormalities following on-target genome editing, such as chromosome loss, have not been systematically investigated or prevented. Thus, these potential concerns for genome editing continue to persist, including in the clinic, at an unknown frequency.

T cells have been extensively engineered using Cas9 to develop potent immunotherapies for cancer[66,67,101,255] and autoimmune diseases[256,257]. In a previous study, low-level chromosome 14 aneuploidy was detected in primary human T cells after Cas9-mediated genome editing of the *T cell receptor alpha constant* (*TRAC*) gene using one clinically relevant guide RNA (gRNA)[216]. However, the extent to which chromosome loss occurs at other target sites, the determinants of this phenomenon, the behavior of T cells with Cas9-induced chromosome loss, and the clinical relevance of these findings remain unknown. Along with understanding this phenomenon, strategies to reduce or eliminate chromosome loss as an outcome of genome editing would improve the safety of engineered T cell therapies for patients.

Here, we analyzed chromosome loss following Cas9-induced genome editing at hundreds of sites across every somatic chromosome in the human genome to determine the frequency, determinants, and consequences of this phenomenon. We found Cas9-induced chromosome loss was a generalizable phenomenon, although it was specific to the chromosome targeted by Cas9, and was prevalent at sites across the genome. T cells with Cas9-induced chromosome loss had a fitness and proliferative disadvantage, yet could persist over multiple weeks of *ex vivo* culture. Surprisingly, chromosome loss also occurred during pre-clinical production of chimeric antigen receptor (CAR) T cells but was minimal or undetectable in Cas9-edited patient T cells from a first-in-human phase 1 clinical trial. Further experimentation demonstrated that a modified T cell editing protocol employed in our clinical trial increased levels of the DNA damage response protein p53 while decreasing chromosome loss, suggesting a possible mechanism for Cas9-induced chromosome loss and an unexpected strategy to avoid this unintended outcome in patients.

3.3 Results

3.3.1 Single-cell RNA sequencing reveals chromosome loss in Cas9-edited T cells

The *TRAC* locus, which encodes the T cell receptor (TCR) responsible for CD4 and CD8 T cell reactivity to peptide-MHC complexes, is of immense interest for genome editing applications in human T cells. For engineered adoptive T cell therapies, where T cells are armed with a transgenic receptor for targeted immunotherapy, disrupting endogenous TCR expression limits graft-versus-host toxicity associated with mispairing of transgenic and endogenous TCRs[258]. Additionally, abrogating the TCR is an important step towards developing allogeneic “off-the-shelf” T cell therapies that could reduce cell manufacturing costs and expand patient accessibility[259].

To quantify chromosome loss after genome editing of *TRAC*, primary human T cells were electroporated with *S. pyogenes* Cas9 ribonucleoprotein (RNP) including one of 11 different gRNAs tiled across the first exon of *TRAC* or a non-targeting gRNA (Figure 3.1A). Reduced TCR expression occurred in 60-99% of all cells as measured by flow cytometry (Figure 3.2A) and editing at the *TRAC* locus was observed in 62-97% of all genomic DNA sequencing reads (Figure 3.2B), depending on the gRNA. Four days after Cas9 RNP introduction, T cells were subjected to multiplexed single-cell RNA sequencing (scRNA-seq) to detect reduced transcript levels resulting from chromosome loss caused by Cas9-mediated double-strand DNA breaks (DSBs) (Figure 3.2C)[260]. Transcriptome-wide analysis using existing computational methods revealed a reduction in gene dosage, specifically on chromosome 14, in cells with a *TRAC*-targeting gRNA compared to cells with a non-targeting gRNA (Figure 3.1B)[261]. We further estimated the distribution of breakpoint loci across chromosome 14, finding the highest frequency to be near the known genomic location of our *TRAC*-targeting gRNAs (Figure 3.1C). We observed cells with lower gene dosage originating at the Cas9 target site (partial chromosome loss) as well as cells with lower gene dosage across the entire chromosome (whole chromosome loss) (Figure 3.1D). Overall, ~5-20% of T cells exhibited partial or whole loss of chromosome 14 depending on the *TRAC*-targeting gRNA (Figure 3.1E; Figures 3.2D-F).

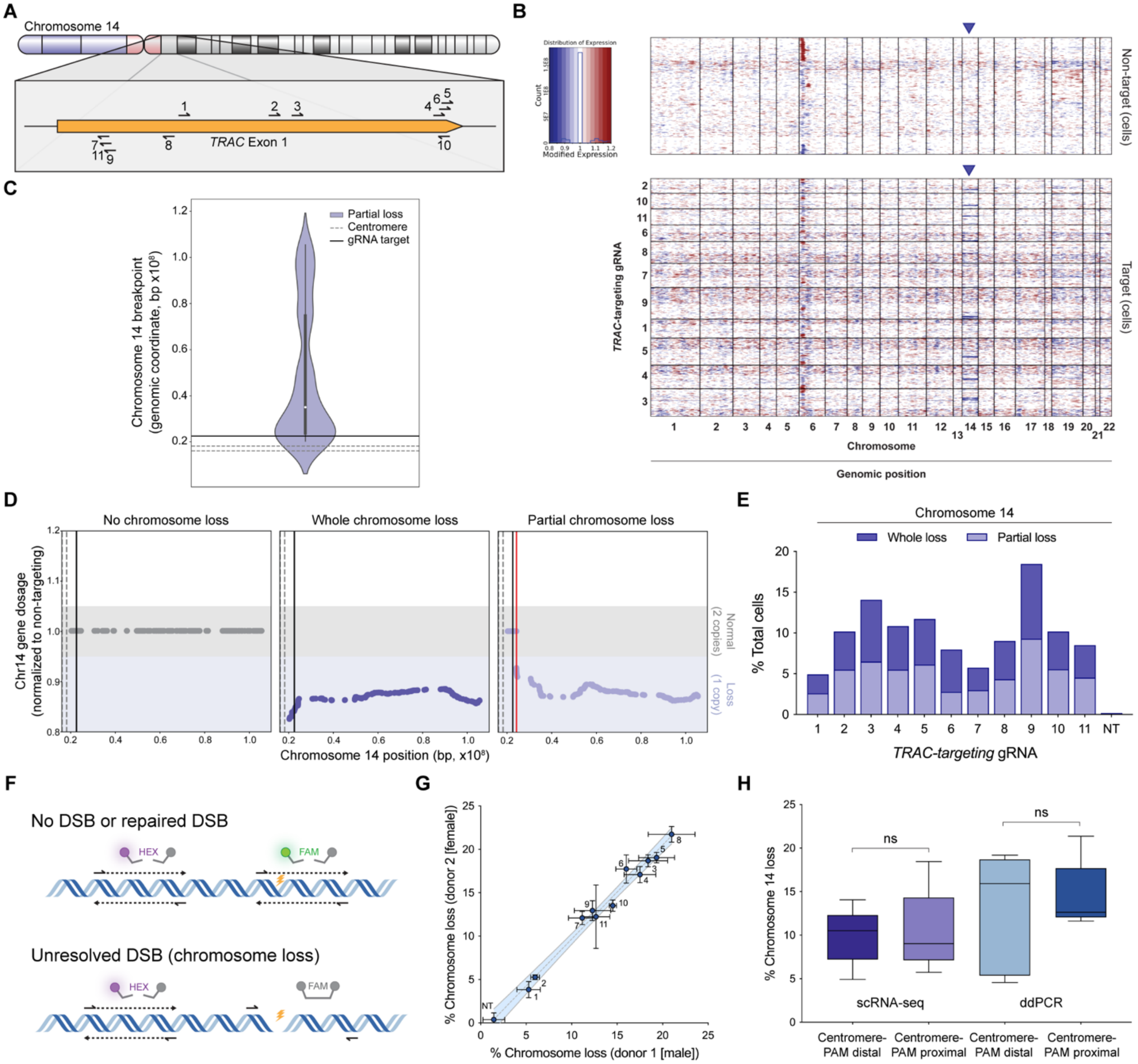


Figure 3.1: CRISPR-Cas9 genome editing of TRAC results in whole and partial chromosome loss. **(A)** Cas9 gRNA target sites tiled across the first exon of *TRAC* on chromosome 14. **(B)** Gene dosage from transcriptome-wide scRNA-seq of T cells treated with Cas9 and a non-targeting gRNA (top heatmap) or *TRAC*-targeting gRNA (bottom heatmap). Each individual row corresponds to a single cell and each column corresponds to a specific gene and its genomic position, grouped by chromosome (outlined in black). Red represents increase in gene dosage while blue represents decrease in gene dosage. Rows outlined in black represent cells treated with different *TRAC*-targeting gRNAs. Blue arrows highlight chromosome 14, where *TRAC* is located. **(C)** Distribution of computationally predicted chromosome 14 breakpoints in cells predicted to have a chromosomal loss event. The distribution is an aggregate of 11 different *TRAC*-targeting gRNAs (all within ~300 bp) in cells with partial chromosome loss. **(D)** Representative single cell chromosome 14 gene dosage plots illustrating a cell with no chromosome loss (left), whole chromosome loss (middle), or partial chromosome loss (right). Gene dosage was normalized to non-targeting samples. Gray shaded area (gene dosage of 0.95-1.05) represents normal gene dosage (2 copies). Blue shaded area (gene dosage of <0.95) represents reduction in gene dosage (1 copy). Dotted lines represent the centromere, black lines represent the Cas9 target site, and the red line represents the computationally predicted breakpoint. **(E)** Quantification of whole and partial chromosome 14 loss across all gRNAs from

scRNA-seq. NT indicates non-targeting gRNA. **(F)** Schematic of ddPCR assay to measure chromosome loss. The yellow lightning bolt represents the Cas9 target site. The detection of both HEX and FAM probes indicates no DSB or repaired DSB (top illustration). The detection of the HEX probe but not the FAM probe indicates an unresolved DSB that represents chromosome loss (bottom illustration). **(G)** Quantification of chromosome loss at the Cas9 target site across all gRNAs from ddPCR (n = 3, n = 2 biological donors). Numbers next to each point represent the *TRAC*-targeting gRNA. NT indicates non-targeting gRNA and represents samples from four different ddPCR amplicons. Error bars represent the standard deviation from the mean. Dashed line represents linear regression line of best fit and shaded region represent 95% confidence intervals (Slope = 1.082, $R^2 = 0.9853$). **(H)** Comparison of chromosome 14 loss between *TRAC*-targeting conditions where the PAM is distal (Centromere-PAM distal) or proximal (Centromere-PAM proximal) to the centromere relative to the gRNA spacer sequence. Chromosome 14 loss was measured by scRNA-seq (n = 1 biological donor) or ddPCR (n = 3, n = 2 biological donors). *P*-values are from Welch's unpaired t-tests and are from left to right 0.8689 and 0.7338. ns = not significant.

3.3.2 DNA-based droplet digital PCR is an orthogonal method to detect chromosome loss

As an orthogonal approach to detect chromosome loss, we used droplet digital PCR (ddPCR) to directly quantify genomic DNA copy number, eliminating potential interference from transcriptional or epigenetic factors that may affect the scRNA-seq results. Two primer/probe sets were designed to amplify nearby regions of the target gene, one as a control (HEX) and the other spanning the Cas9 gRNA target site (FAM) so that a DSB and chromosome loss would inhibit amplicon generation (Figure 3.1F). Three days after Cas9 electroporation, genomic DNA harvested from *TRAC*-targeted T cells resulted on average in ~4-22% chromosome loss; these losses were highly reproducible across biological T cell donors (Figure 3.1G). Importantly, primers and probes in the amplicon spanning the intended Cas9 target site were positioned to avoid binding site disruption by small insertions and deletions (indels), the most common outcome after Cas9 genome editing.

Based on the observation that Cas9 preferentially remains bound to the non-protospacer adjacent motif (PAM) side of the target DNA after cleavage[33,262], we wondered whether orientation of the PAM relative to the centromere influenced chromosome loss (Figure 3.2G). However, we found no significant difference between targets where the PAM was distal or proximal to the centromere, relative to the gRNA spacer sequence (Figure 3.1H). Furthermore, the rates of chromosome 14 loss measured by scRNA-seq or ddPCR did not correlate with the efficacy of TCR disruption or genomic position targeted by different gRNAs (Figures 3.2H and 3.2I).

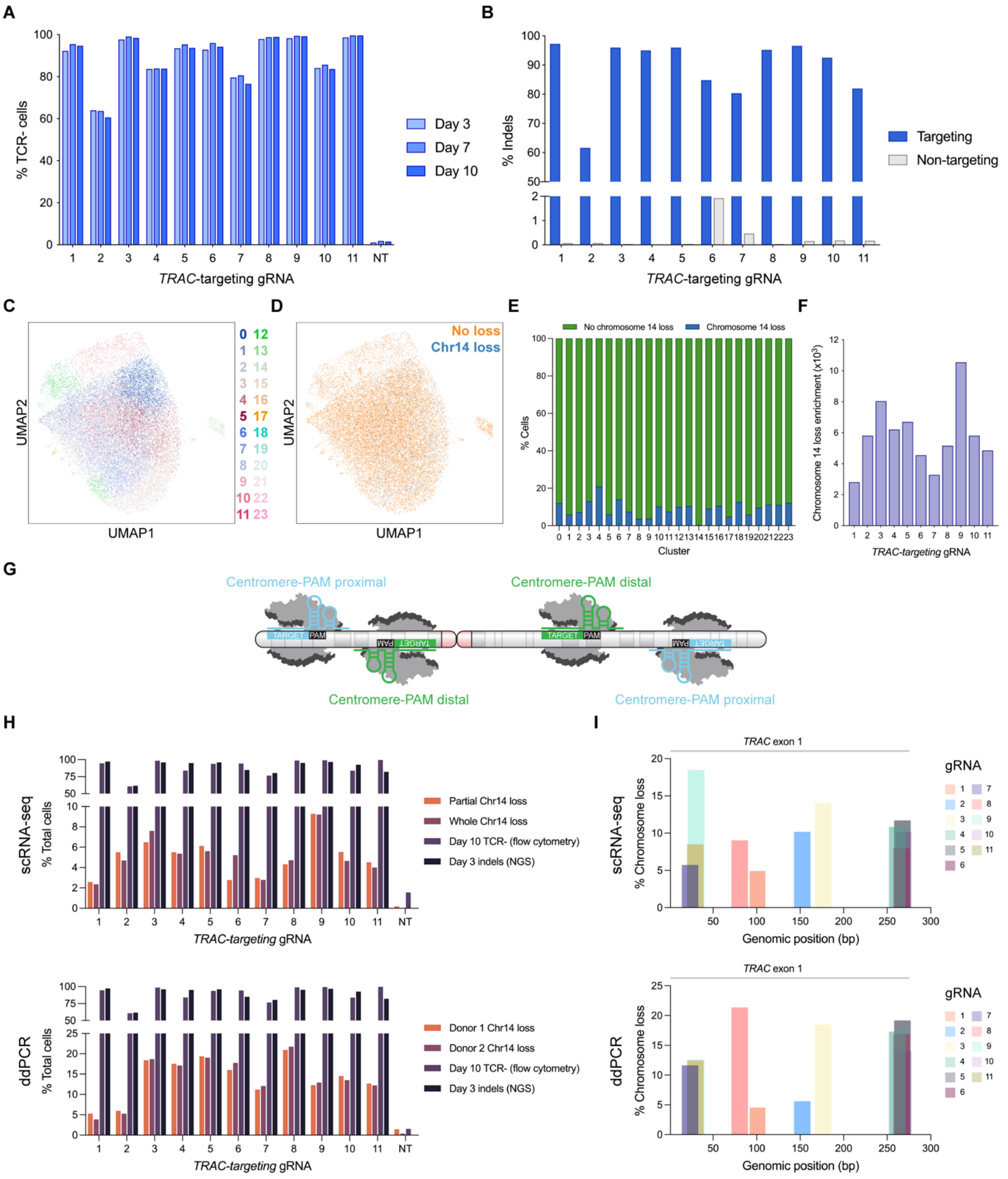


Figure 3.2: CRISPR-Cas9 genome editing results in chromosome loss regardless of guide RNA orientation or genomic position. (A) TCR disruption from Cas9 genome editing of *TRAC*. TCR expression was measured via flow cytometry at 3, 7, and 10 days post-electroporation. NT indicates non-targeting gRNA. **(B)** Indels at the *TRAC* locus (targeting) from Cas9 genome editing as measured by next-generation sequencing. Cells treated with a non-targeting gRNA were evaluated for indels at each of the *TRAC* target sequences (non-targeting). **(C)** UMAP projection of T cells edited with a *TRAC*-targeting or non-targeting gRNA. **(D)** The UMAP projection of T cells within the *TRAC* editing experiment was overlayed with

estimations of whether the cell had chromosome 14 loss or not (whole or partial chromosome loss). **(E)** Percentage of cells with chromosome 14 loss per cluster (see Fig. S1D). (**F)** Quantification of chromosome 14 loss enrichment across 11 different *TRAC*-targeting gRNAs. Chromosome 14 loss enrichment was calculated relative to T cells treated with Cas9 and a non-targeting gRNA.**(G)** Schematic of gRNA orientation relative to the centromere. Cas9 targets where the PAM was proximal to the centromere (red) relative to the target DNA sequence were considered centromere-PAM proximal (blue), while Cas9 targets where the PAM was distal to the centromere relative to the target DNA sequence were considered centromere-PAM distal (green). **(H)** Comparison of TCR disruption (by flow cytometry or next-generation sequencing) and chromosome 14 loss as measured by scRNA-seq (top) or ddPCR (bottom). **(I)** Chromosome 14 loss by scRNA-seq (combination of whole and partial chromosome 14 loss, top) or ddPCR (mean of n = 3 replicates, n = 2 biological donors, bottom) for each gRNA based on target position within the *TRAC* gene. Gray line indicates the first exon of *TRAC* (274 bp).

3.3.3 DNA-based droplet digital PCR is an orthogonal method to detect chromosome loss

To determine the generality of chromosome loss after genome editing, we conducted a comprehensive CRISPR screen using a library of 384 unique gRNAs targeting 3-7 genes on every somatic chromosome (92 genes in total) with four unique gRNA sequences targeting each gene (Figure 3.3A; Figures 3.4A, and 3.4B). CROP-seq was used to track gRNAs delivered to individual cells in our experiment because it avoids the template switching observed with other methods[263,264] and because it has been successfully deployed in primary human T cells[265]. Targets relevant to T cell genome engineering were prioritized, such as *TRAC*[64,101,107], *TRBC*[101], *PDCD1*[101]*, B2M*[171], *IL2RA*[64], *CXCR4*[266], and *CIITA*[267], as well as other targets of interest for clinical genome editing such as *BCL11A*[93,95] and *HBB*[65,91] for the treatment of sickle cell disease, *TTR* to treat transthyretin amyloidosis[109], *HTT* to treat Huntington's disease[268], and *SERPINA1* to treat alpha-1-antitrypsin deficiency[269].

Using our previously established computational pipeline on the CROP-seq dataset, we determined the breakpoints and gene dosage for 92 different genes targeted by Cas9 (Figures 3.4C-F). For numerous genes targeted in our screen, we observed significant enrichment for loss of the targeted chromosome in cells with a corresponding gRNA compared to cells with a gRNA targeting a different chromosome (Figure 3.3B). Chromosome loss above background levels occurred with 55% (201/364) of all gRNAs, for 89% (82/92) of all genes targeted, and in 100% (22/22) of all chromosomes targeted. Across all gRNAs, 3.25% of targeted cells had detectable whole or partial chromosome loss (Figure 3.5A). For cells with a non-targeting gRNA, no detectable chromosome loss was identified on any somatic chromosome. Enrichment for chromosome loss was much higher in chromosomes targeted by Cas9 compared to non-targeted chromosomes (Figure 3.3C), suggesting that this phenomenon is an outcome of target-specific cleavage during Cas9-mediated genome editing.

We validated this genome-scale chromosome loss by selecting 15 gRNAs from the library and individually electroporating them as Cas9 RNPs into T cells. ddPCR was used to measure chromosome loss at various sites in the genome and showed greater levels of chromosome loss at the targeted site compared to non-targeted sites for nearly all gRNAs (Figure 3.3D). Additionally, rates of chromosome loss were highly correlated (Spearman's correlation = 0.59) with those estimated by scRNA-seq (Figure 3.3E).

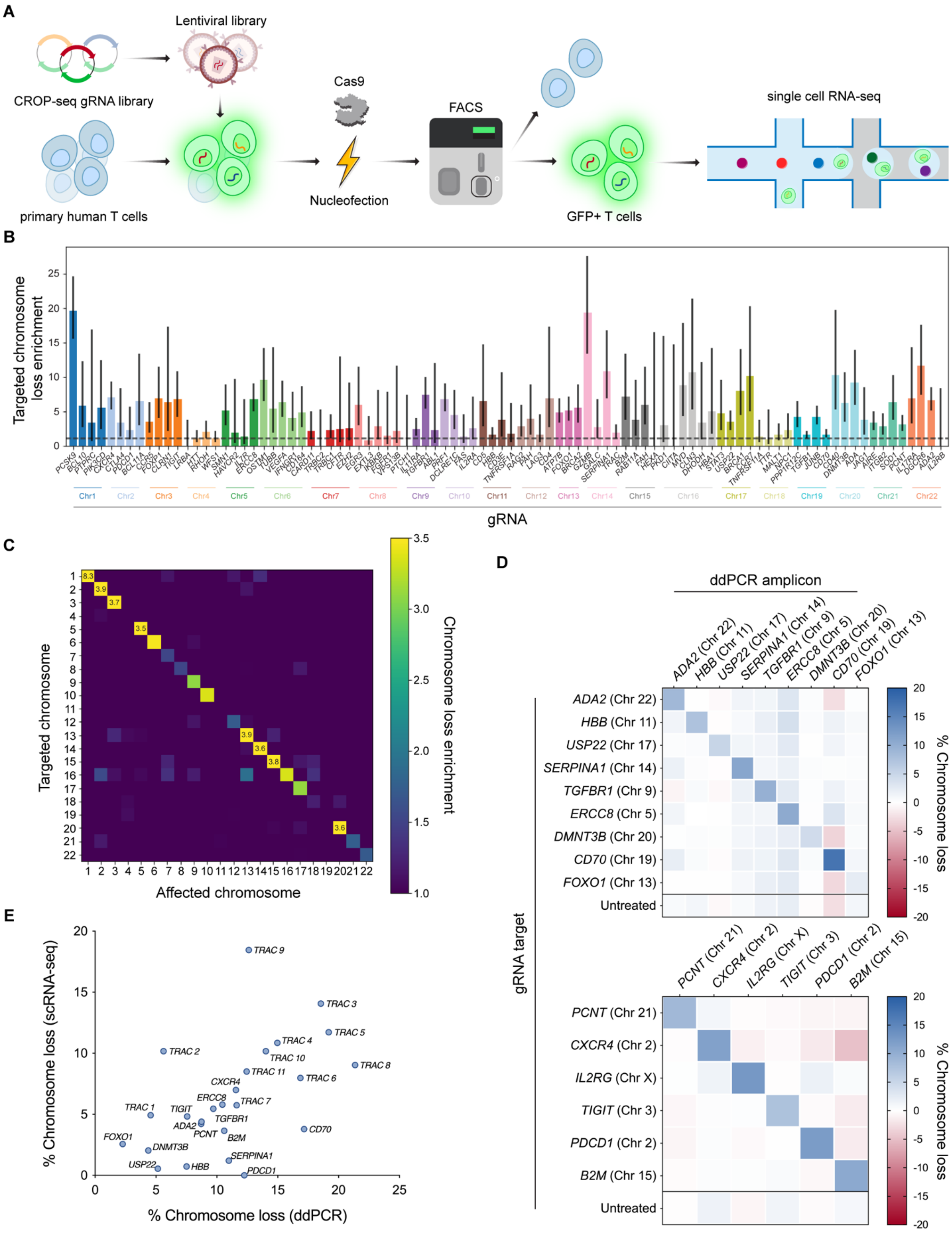


Figure 3.3: Genome-scale CRISPR-Cas9 screen reveals target-specific chromosome loss. (A) Workflow for a CRISPR-Cas9 screen to estimate chromosome loss in T cells. Primary human T cells were

transduced with a CROP-seq lentiviral library expressing one of 384 gRNAs. Cells were then electroporated with Cas9 protein, before GFP+ cells (co-expressed on the CROP-seq gRNA vector) were enriched via fluorescence-activated cell sorting. Enriched cells were subject to scRNA-seq and downstream analysis. **(B)** Quantification of targeted chromosome loss enrichment for each target gene. Each of the 92 bars represents the combination of four unique gRNAs targeting the same gene. Chromosome loss enrichment was calculated relative to the baseline loss per chromosome in cells containing a gRNA targeting a different chromosome. Error bars represent 95% confidence intervals. **(C)** Chromosomal loss enrichment at each somatic chromosome across all gRNAs. Rows represent the chromosome targeted by the Cas9 gRNA. Columns represent the chromosome analyzed for chromosome loss. **(D)** Chromosome loss measured by ddPCR at 15 different Cas9 target sites across the genome. Rows titles indicate the identity of the gRNA used. Column titles indicate the site in the genome that was analyzed via ddPCR. Heatmap values represent the mean of replicates (n = 3, except n = 2 for *B2M* target column). **(E)** Correlation between chromosome loss from 25 gRNAs as measured by scRNA-seq and ddPCR. Spearman's correlation = 0.59, ***P* = 0.0017 (two-tailed).

Further analyses of the CROP-seq screen to investigate the contribution of the Cas9 gRNA sequence revealed no influence by gRNA binding orientation (Figure 3.5B), nucleotide sequence (Figures 3.5C and 3.5D), or GC content on chromosome loss (Figure 3.5E). The computationally predicted dominant end-joining repair pathway, namely non-homologous end joining (NHEJ) or microhomology-mediated end joining (MMEJ), for each gRNA target also did not show a strong correlation with chromosome loss (Figures 3.5F and 3.5G). However, we did observe a moderate correlation between the distance of each targeted gene from the centromere and the rate of chromosome loss induced (Figure 3.5H), with gRNAs targeting closer to the centromere showing higher levels of chromosome loss.

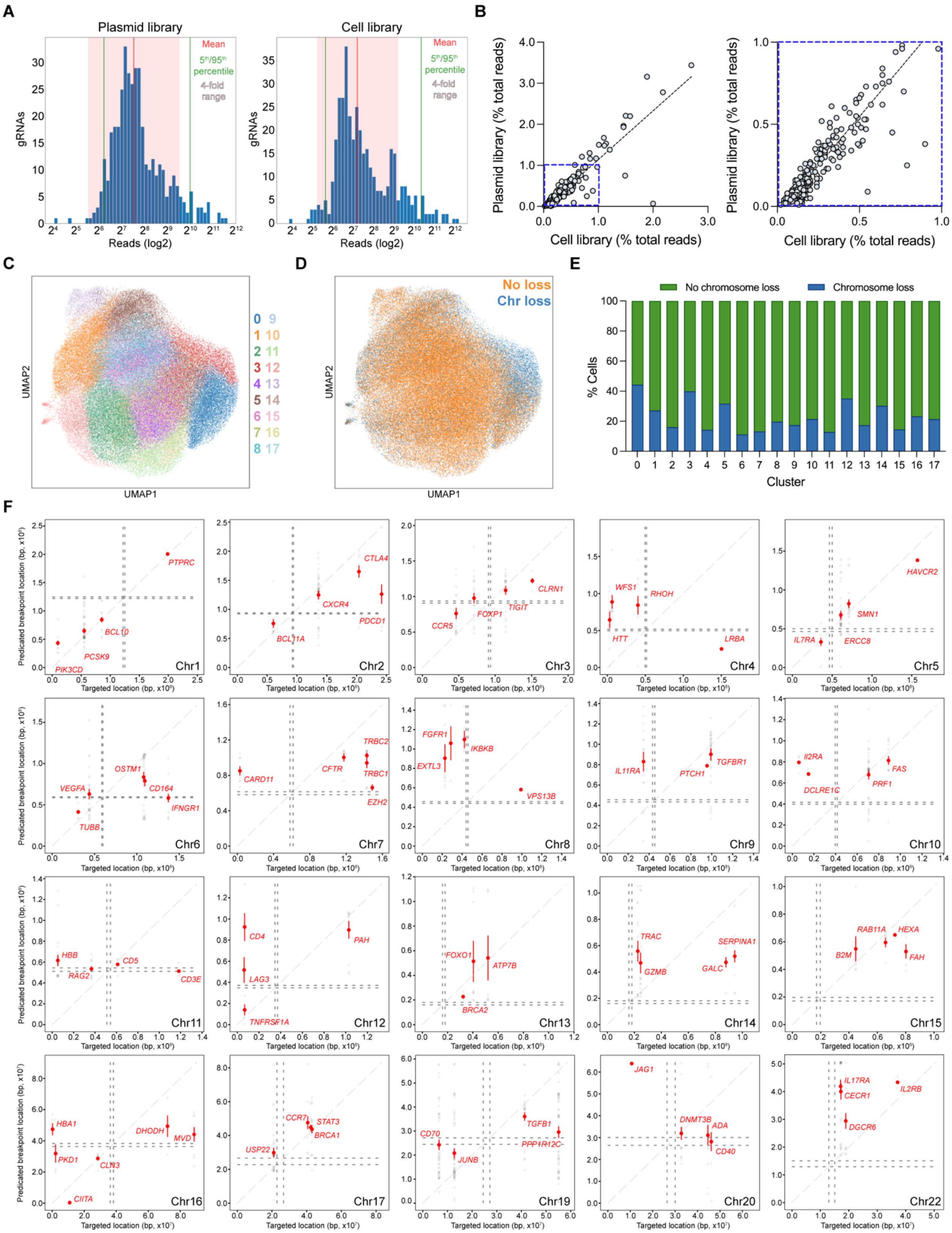


Figure 3.4: CROP-seq reveals genome-scale breakpoints and chromosome loss. (A) Distribution of next-generation sequencing reads for each gRNA in the CROP-seq gRNA library as a plasmid (left) or once

integrated into T cells via lentivirus (right). The mean is shown as a red line, 5th and 95th quartiles are shown as green lines, and a 4-fold range to either side of the mean is shown as a pink shaded area. **(B)** Correlation between gRNA reads in the plasmid library and cell library. A zoomed in perspective (blue dashed line, left) is shown in a separate panel (blue dashed line, right). Dashed gray line represents the linear regression line of best fit (Slope = 1.189, $R^2 = 0.8348$). **(C)** UMAP projection of T cells from the CROP-seq screen. **(D)** The UMAP projection of T cells within the CROP-seq screen was overlayed with estimations of whether the cell had a targeted chromosome loss or not (whole or partial chromosome loss). **(E)** Percentage of cells with targeted chromosome loss per cluster (see Fig. S2C). **(F)** Predicted breakpoint location versus intended gRNA target location from the CROP-seq screen. gRNAs are grouped by targeted chromosome. Chromosomes 18 and 21 are omitted because no partial chromosome loss was detected. Red data points represent the mean and 95% confidence interval.

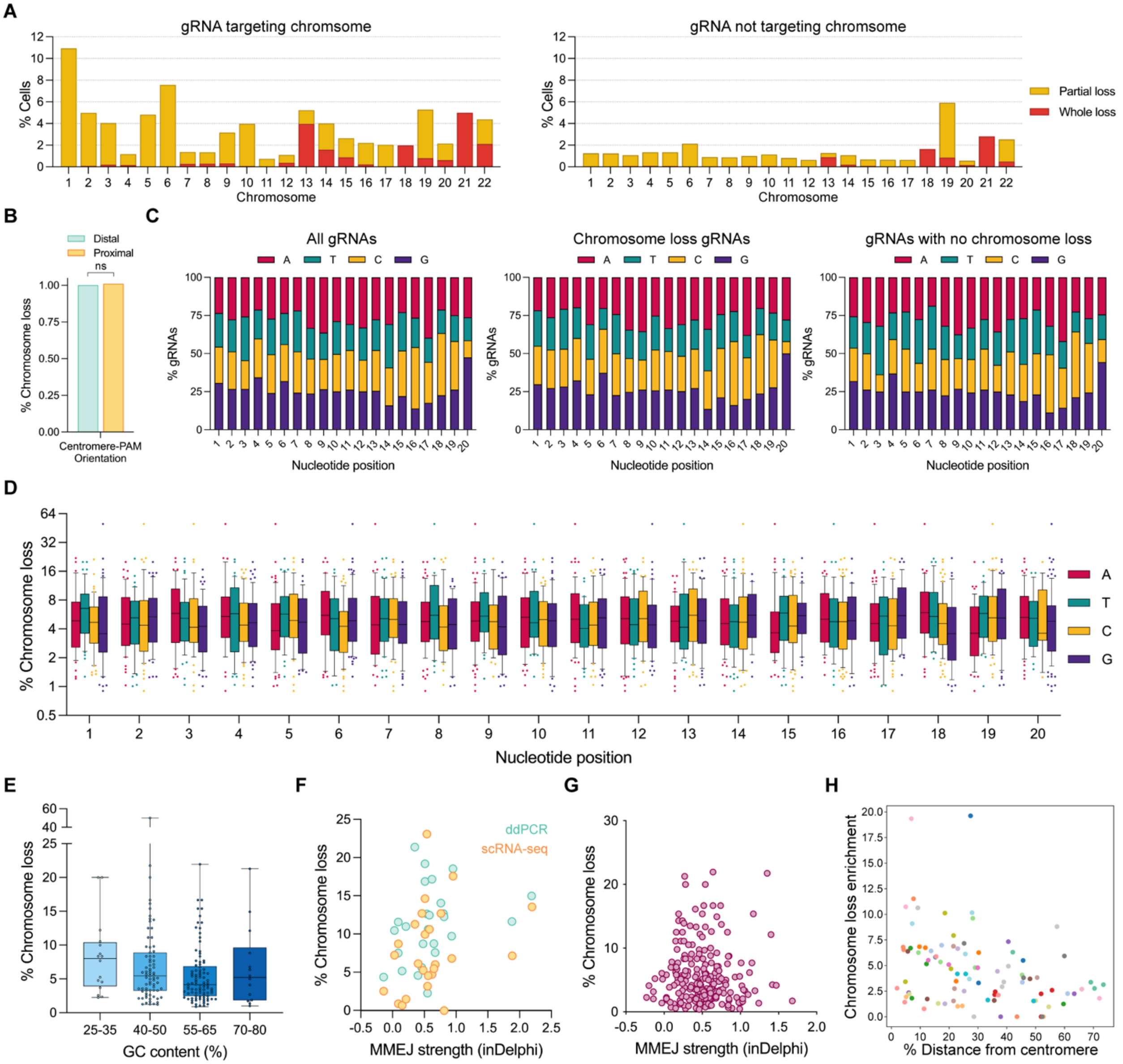


Figure 3.5: Influence of genetic context and Cas9 gRNA sequence on chromosome loss. (A) Partial and whole chromosome loss from the CROP-seq screen. Chromosome loss was quantified at chromosomes where the gRNA was targeting that specific chromosome (left) or at chromosomes not targeted by the gRNA (right). Chromosome loss at chromosomes not targeted by the gRNA was used to

calculate baseline noise. **(B)** Influence of gRNA orientation on chromosome loss. gRNAs where the PAM is distal to the centromere relative to the gRNA target sequence were compared against gRNAs where the PAM is proximal to the centromere relative to the gRNA target sequence. *P*-value was calculated using a two-sided Fisher's Exact Test and was 0.592413. ns = not significant. **(C)** Distribution of nucleotides across each position of the gRNA spacer within the CROP-seq library. Distribution for all gRNAs (left), gRNAs that resulted in chromosome loss (middle), and gRNAs that did not result in chromosome loss (right). **(D)** Chromosome loss by nucleotide identity across each position of the gRNA spacer within the library. **(E)** Chromosome loss by gRNA spacer sequence GC content. gRNAs were arbitrarily binned by varying levels of GC content. **(F)** Chromosome loss versus computationally predicted MMEJ influence for Cas9 RNP electroporation experiments (teal = ddPCR measurements, yellow = scRNA-seq measurements. Chromosome loss rates are identical to Fig. 2e). ddPCR Spearman's correlation = 0.40, **P* = 0.04 (two-tailed); scRNA-seq Spearman's correlation = 0.27, *P* = 0.19 (two-tailed). **(G)** Chromosome loss versus computationally predicted MMEJ influence for the CROP-seq screen experiment. gRNAs with non-zero chromosome loss were plotted. Spearman correlation = -0.08, *P* = 0.25 (two-tailed). **(H)** Chromosome loss by position along the target chromosome. Distance from the centromere was normalized to the length of the target chromosome. Spearman's correlation = -0.34, ****P* = 0.0009 (two-tailed).

3.3.4 Chromosome loss accompanies transcriptional signatures of DNA damage response, apoptosis, and quiescence

To assess the functional effects of Cas9-induced chromosome loss, we performed differential gene expression analysis between Cas9-edited T cells with or without chromosome loss. We identified genes that were differentially expressed across groups of cells with different chromosomes lost (Figure 3.6A; Figure 3.7A). *CD70*, for example, was significantly upregulated in every group of cells with chromosome loss regardless of which chromosome was lost, and *MDM2* was significantly upregulated in every group of cells with chromosome loss except those that lost chromosomes 13 or 18. Meanwhile, *PHGDH* was downregulated in every group of cells with chromosome loss except those that lost chromosomes 6, 10, 18, or 21. The numerous genes that were differentially expressed across cells with various chromosomes lost suggests that these findings are not a result of expression changes from lowered dosage of the target gene, but a general influence of Cas9-induced chromosome loss.

Previous studies have demonstrated that a single Cas9-induced DSB can lead to p53 upregulation[270]. Consistent with this finding, gene ontology analysis revealed the p53 DNA damage response and general apoptosis pathways were the most significantly overexpressed gene modules in cells with Cas9-induced chromosome loss (Figure 3.6B). We also observed an increase in cell cycle markers associated with the G0 phase and a decrease in those associated with the S phase for T cells with chromosome loss compared to those without (Figure 3.6C; Figures 3.7B and 3.7C), suggesting p53-induced cell cycle arrest. The results of both the differential gene expression analysis and cell cycle analysis indicate reduced fitness in T cells with Cas9-induced chromosome loss.

We further investigated the relationship between epigenetic markers and chromatin accessibility with Cas9-induced chromosome loss in primary human T cells; however, no significant correlation was found between these epigenetic factors and chromosome loss (Figure 3.6D).

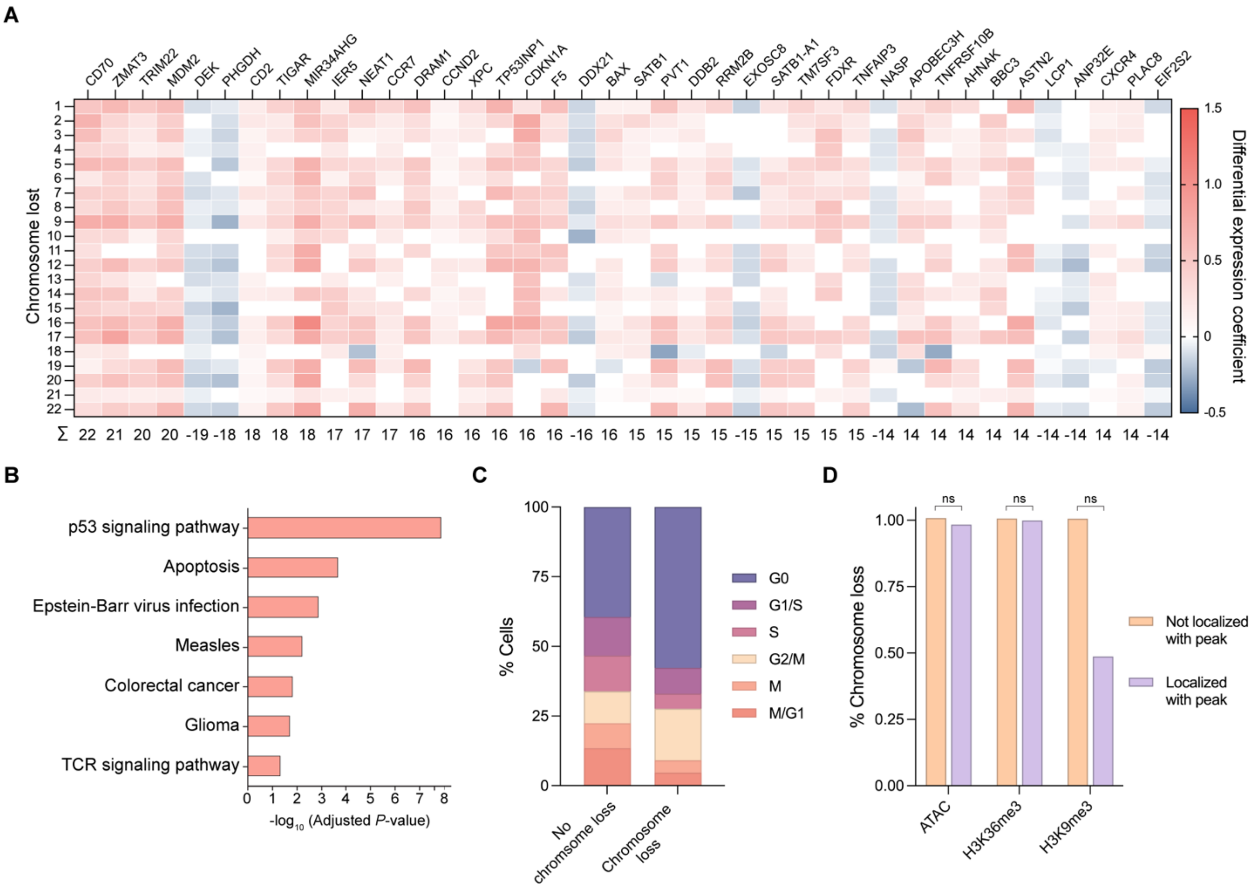


Figure 3.6: Genetic and epigenetic factors influence Cas9-induced chromosome loss. **(A)** Heatmap of differentially expressed genes in cells with chromosome loss compared to cells without chromosome loss. Cells with chromosome loss were divided into 22 groups depending on which somatic chromosome was lost (rows), and differentially expressed genes were individually investigated (columns). Upregulated genes are shown in red while downregulated genes are shown in blue. Genes were given a score of 1 (upregulated), -1 (downregulated), or 0 (no difference) for each chromosome loss group. Summed gene scores across all chromosome loss groups is shown below; genes with a score >|13| are displayed. **(B)** Gene ontology analysis based on differential gene expression. The most significantly upregulated modules are displayed. **(C)** Cell cycle analysis based on expression profiles. The percentage of cells in each cell cycle phase were quantified for cells with no chromosome loss or cells with chromosome loss. **(D)** Influence of epigenetic marks on chromosome loss. The gRNA sequence for cells with or without chromosome loss was analyzed for localization within ± 75 bp of an epigenetic marker peak. *P*-values were calculated using a two-sided Fisher's Exact Test and are from left to right (ATAC) 0.365496, (H3K36me3) 0.789824, and (H3K9me3) 0.305706. ns = not significant.

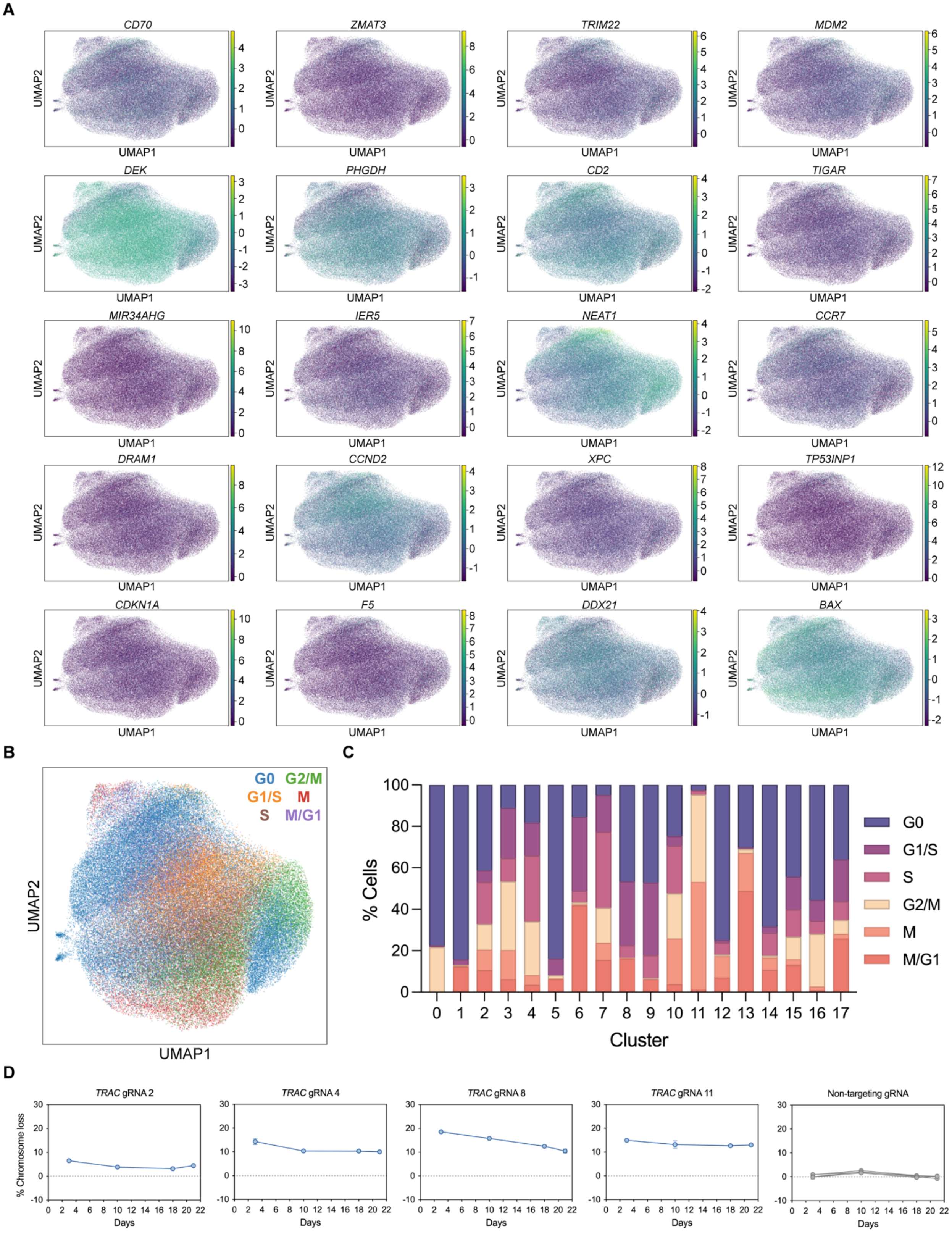


Figure 3.7: Cas9-induced chromosome loss is associated with differential gene expression and a fitness disadvantage. (A) UMAP projections of T cells within the CROP-seq screen. Gene expression was

overlayed onto the projections for the top 20 genes that were most differentially expressed across cells that had chromosome loss. **(B)** UMAP projection of T cells within the CROP-seq screen overlayed with cell cycle markers (see Figure 3.4C). **(C)** Quantification of cell cycle states across the different clusters (see Figure 3.4C). **(D)** ddPCR measurements of chromosome loss at the Cas9 *TRAC* target site over 21 days. Error bars represent the standard deviation from the mean (n = 3).

3.3.5 T cells with chromosome loss persist *ex vivo* but with reduced fitness and proliferation

To determine whether T cells with Cas9-induced chromosome loss persist over time, we used ddPCR to measure the extent of chromosome loss at various timepoints during *ex vivo* culture. We chose timepoints over 2-3 weeks, which is a similar length of *ex vivo* culture compared to the manufacturing protocols of clinical trials with Cas9-edited T cells[101,255]. As expected, T cells treated with Cas9 RNP targeting *TRAC* showed the highest levels of DSBs one day after electroporation (Figure 3.8A), when Cas9 RNP is still present within cells and DNA repair is ongoing[121,271]. By day three post-treatment, DSBs plateaued until day 14 with most conditions showing a slight downward trend (Figure 3.8A). Since Cas9 RNP-mediated cleavage and DNA repair go to completion within 24 hours[272], we posited that DSBs measured at day three and beyond are from unrepaired DSBs which we considered chromosome loss.

We evaluated the temporal dynamics of chromosome loss over a longer period by repeating the experiment over the course of 21 days using four of the 11 *TRAC*-targeting gRNAs. Again, levels of chromosome loss showed a slight reduction over the three weeks, with detectable chromosome loss at the last timepoint remaining above that of non-targeting controls (Figure 3.7D).

To test the possibility that targeting a specific gene or chromosome may affect chromosome loss or T cell viability, we repeated the Cas9 electroporation and genomic DNA ddPCR with 15 gRNAs targeting other genes throughout the genome. Culturing the genome edited T cells for 14 days and measuring chromosome loss at various timepoints throughout, we once again observed a gradual reduction in chromosome loss over time (Figure 3.8B). These findings show that chromosome loss in T cells is observable as far out as 2-3 weeks, across multiple targeted genes and chromosomes. However, the gradual decrease in chromosome loss over time suggests a fitness disadvantage for T cells with this genomic aberration.

We additionally tested whether chromosome loss was associated with a proliferative disadvantage in T cells. *TRAC*-edited T cells were stained with a cell proliferation dye and cultured for five days. Cells that underwent the highest and lowest amounts of proliferation, based on dye intensity, were sorted and chromosome loss was measured between the two groups. Chromosome loss in the highest proliferating quartile was identifiable but statistically lower than chromosome loss in the lowest proliferating quartile (Figure 3.8C), which suggests that Cas9-induced chromosome loss confers a proliferative disadvantage.

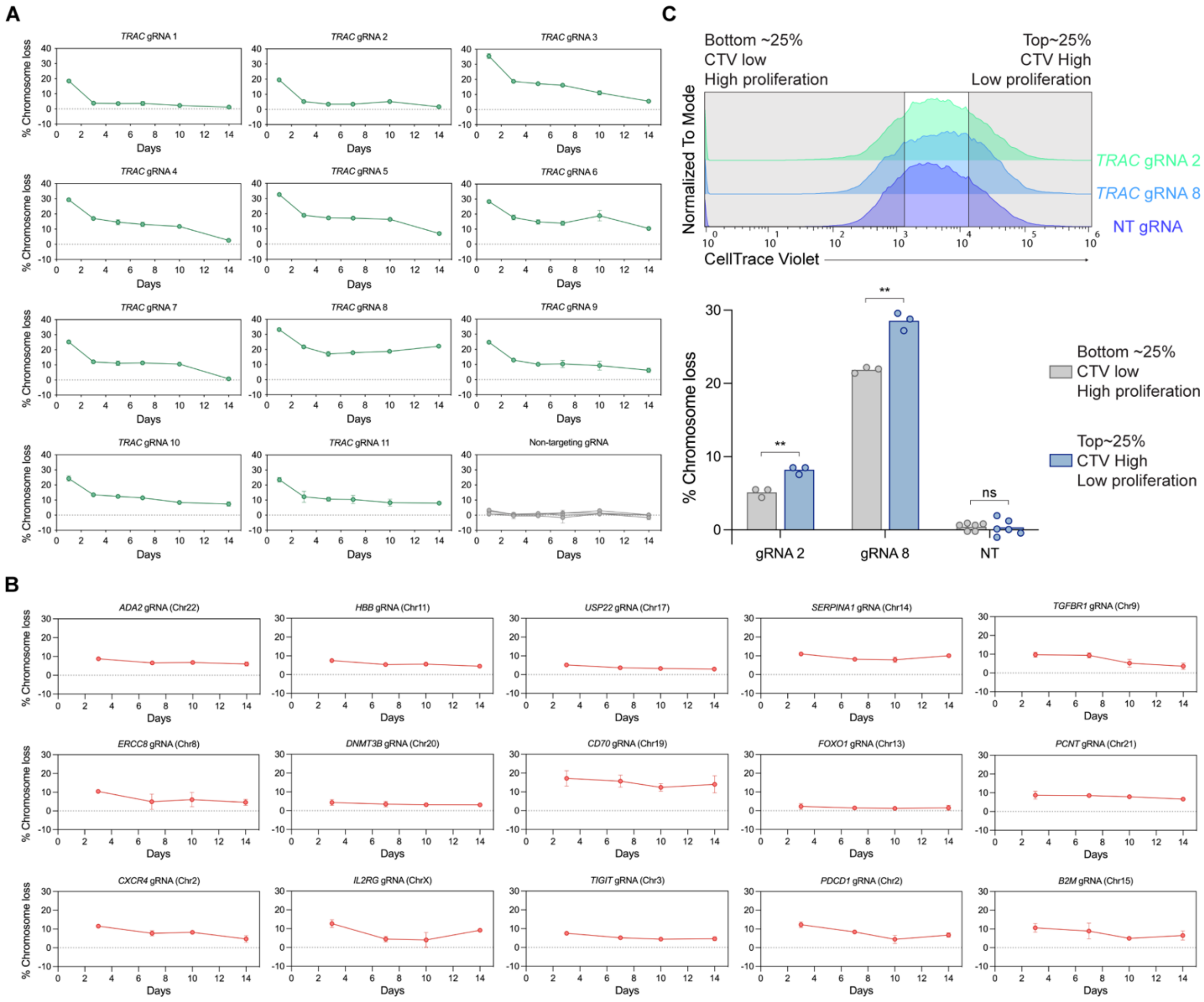


Figure 3.8: Cas9-induced chromosome loss persists for weeks but results in reduced fitness and proliferation. (A) ddPCR measurements of chromosome loss at the Cas9 *TRAC* target site over 14 days. Error bars represent the standard deviation from the mean (n = 3). Day 3 results were additional used as the donor 2 (female) results shown in Fig. 1g. **(B)** ddPCR measurements of chromosome loss for 15 different gRNAs targeted to sites across the genome over 14 days. Error bars represent the standard deviation from the mean (n = 3, except n = 2 for *B2M*). Day 3 results were additionally used for the diagonal values in the heatmaps of Fig. 2d. **(C)** Measurement of chromosome loss across T cells of varying proliferative capacity. T cells were stained with CellTrace Violet (CTV) and cultured for five days before sorting the top and bottom quartile (top panel). ddPCR was used to measure chromosome loss in lowly proliferative (CTV high) and highly proliferative (CTV low) populations (bottom panel). NT = non-targeting gRNA. Non-targeted samples evaluated for chromosome loss at the gRNA 2 or gRNA 8 amplicon were combined into a single column (n = 3 for each of the two different ddPCR amplicons). *P*-values are from Welch's unpaired t-tests and from left to right are 0.002970, 0.002970, and 0.275572.

3.3.6 Gene insertion via homology-directed repair results in chromosome loss

Thus far, we have shown that targeted chromosome loss can occur when using Cas9 to disrupt a desired gene, which predominantly occurs through end-joining DNA repair pathways. Tremendous effort has also been invested toward using Cas9 genome editing to correct a mutation or insert a gene by homology-directed repair (HDR). Since end-joining repair and HDR are divergent DNA repair pathways that involve different proteins, undergo different amounts of resection of the DSB ends, and occur in different stages of the cell cycle[30], we wanted to determine whether chromosome loss after Cas9-mediated genome editing also occurs during HDR. To explore this, we used Cas9 RNPs with a gRNA targeting *CD5* and various oligonucleotide HDR templates to integrate a short hemagglutinin (HA) tag in-frame with CD5[273]. Successful generation of CD5+/HA+ cells via HDR peaked as high as ~40% three or five after electroporation (Figure 3.9A). We performed ddPCR to quantify chromosome loss rates at both timepoints and observed similar levels of chromosome loss in *CD5*-targeted cells with an HDR template compared to *CD5*-targeted cells without an HDR template (Figure 3.9B). Additionally, using CD5 and other T cell surface markers, we attempted to use fluorescence-activated cell sorting to enrich for cells without chromosome loss; however, we observed no reduction in chromosome loss (Supplementary Note 3.1).

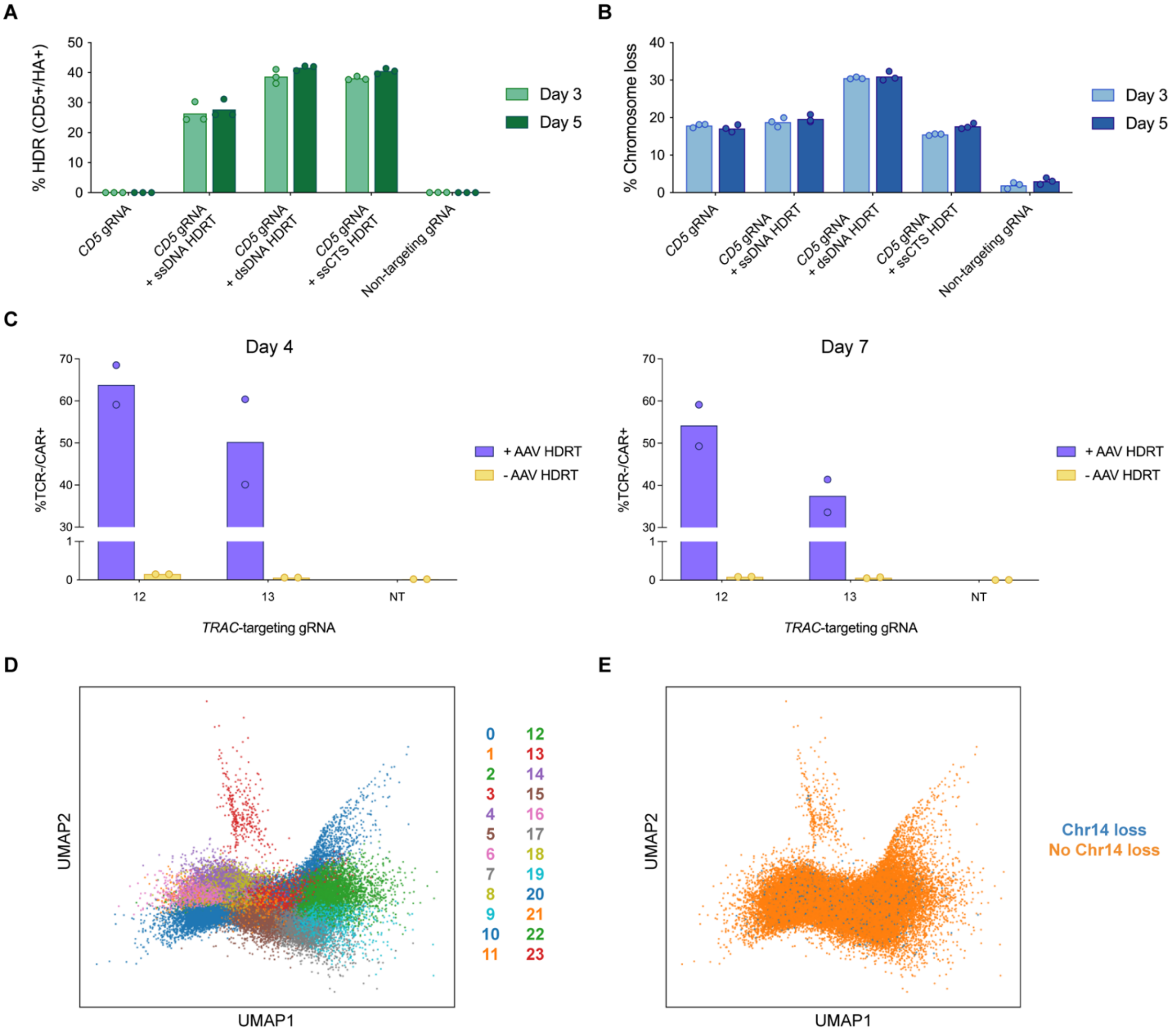


Figure 3.9: CRISPR-Cas9 homology-directed repair results in chromosome loss. (A) HDR efficacy 3 or 5 days post-electroporation, determined by measuring CD5+/HA+ T cells via flow cytometry. **(B)** Chromosome loss at the target *CD5* locus via ddPCR, 3 or 5 days post-electroporation. **(C)** HDR efficacy determined by measuring TCR-/CAR+ T cells via flow cytometry, four or seven days post-electroporation. Two separate electroporations/transductions were conducted for the different time points (n = 2 biological donors). **(D)** UMAP projection of CAR T cells generated via Cas9 HDR. Projection is an aggregate of two biological donors and multiple time points. **(E)** Distribution of CAR T cells with chromosome 14 loss across the UMAP projection (whole or partial chromosome loss).

3.3.7 Pre-clinical CAR T cell generation results in chromosome loss

While CAR T cells and transgenic TCR T cells are currently manufactured using a retrovirus or lentivirus to semi-randomly integrate the retargeting transgene, researchers have also used Cas9 and HDR to precisely insert the transgene within the *TRAC* locus[107,125,273]. This approach, which utilizes the native *TRAC* promoter to control expression of the CAR or retargeted TCR, offers advantages including uniform expression, minimal tonic signaling, and simultaneous disruption of the endogenous TCR. To explore whether chromosome loss occurs when generating pre-clinical CAR T cells via HDR, we electroporated primary human T cells with Cas9 complexed with one of two *TRAC*-targeting gRNAs or a non-targeting gRNA. Just after electroporation, recombinant adeno-associated virus 6 (AAV6) encoding a 1928ζ CAR as an HDR template was added to the T cells (Figures 5A and 5B)[274,275]. Both the reduction of TCR expression and the gain of CAR expression were observed in two independent electroporations; cells from one electroporation were subjected to scRNA-seq at day four post-manufacturing, while cells from the other were subjected to scRNA-seq at day seven post-manufacturing. Overall rates of CAR integration via HDR were ~34-69% (Figure S5C). In all conditions that received a *TRAC*-targeting gRNA, regardless of day or whether an HDR template was present, we observed an enrichment in chromosome 14 loss compared to conditions with a non-targeting gRNA (Figures 5C, S5D, and S5E). When chromosome 14 loss enrichment was normalized to editing efficacy, since day four and day seven CAR T cells were engineered independently, we observed a slight decrease in chromosome 14 loss enrichment over time (Figure 5D). Together, these data suggest that chromosome loss is a general phenomenon that occurs in Cas9-edited T cells, regardless of the DNA repair pathway involved. Our findings also indicate that cells with chromosome loss are present among pre-clinical, Cas9-edited CAR T cells, highlighting the importance of understanding and mitigating this genotoxicity in the context of engineered T cell therapies.

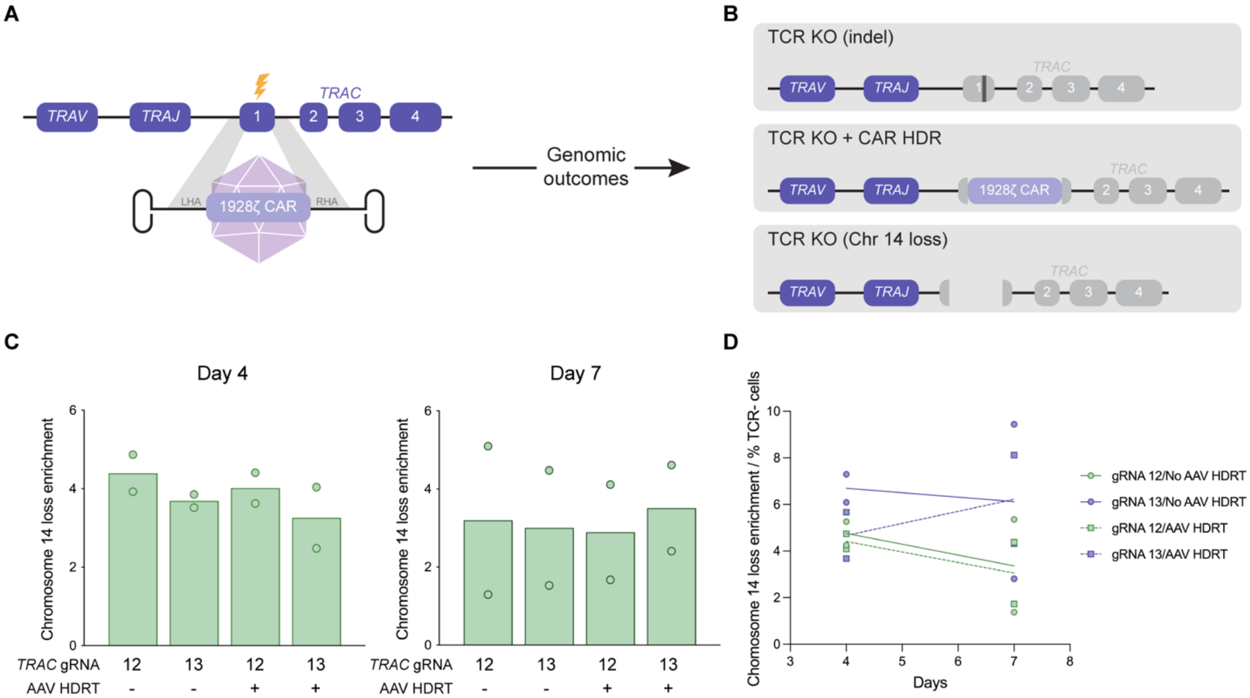


Figure 3.10: Pre-clinical CAR T cell production via homology-directed repair results in chromosome loss. (A) Strategy to generate CAR T cells via HDR with Cas9. AAV6 encoding a 1928ζ CAR between left and right homology arms (LHA and RHA, respectively) serves as a template for HDR after Cas9 cleavage (yellow lightning bolt) of *TRAC*. **(B)** Three potential genomic outcomes after Cas9 HDR: indels that disrupt TCR expression (top), insertion of the CAR transgene that simultaneously disrupts TCR expression (middle), and chromosome loss that disrupts TCR expression (bottom). **(C)** Quantification of chromosome 14 loss enrichment across two *TRAC*-targeting gRNAs with or without an AAV HDR template from scRNA-seq (n = 2 biological donors). Two separate batches of CAR T cells were manufactured, before being subjected to scRNA-seq four or seven days after generation. Chromosome 14 loss enrichment was calculated relative to T cells treated with Cas9 and a non-targeting gRNA. **(D)** Chromosome 14 loss enrichment over time, normalized to Cas9 editing efficacy (n = 2 biological donors). Editing efficacy was determined by the percentage of TCR negative cells as measured via flow cytometry (see Figure 3.12C).

3.3.8 Investigation of chromosome loss in Cas9-edited T cells from clinical trial patients

Our studies thus far have focused on *ex vivo* culturing of T cells; it is not yet known how these findings translate *in vivo*. We conducted a first-in-human phase 1 clinical trial (clinicaltrials.gov, trial NCT03399448) where Cas9 genome edited T cells were administered to patients with advanced, refractory cancer[101]. Autologous T cells from three cancer patients were collected and electroporated with three different Cas9 RNPs, simultaneously targeting *TRAC*, *TRBC*, and *PDCD1*. These edited T cells were then transduced with a lentivirus encoding an HLA-A2*0201-restricted TCR specific to a peptide from the NY-ESO1 and LAGE-1 cancer antigens, resulting in engineered T cells (NYCE). NYCE cells were infused back into the patients and were found to be well-tolerated.

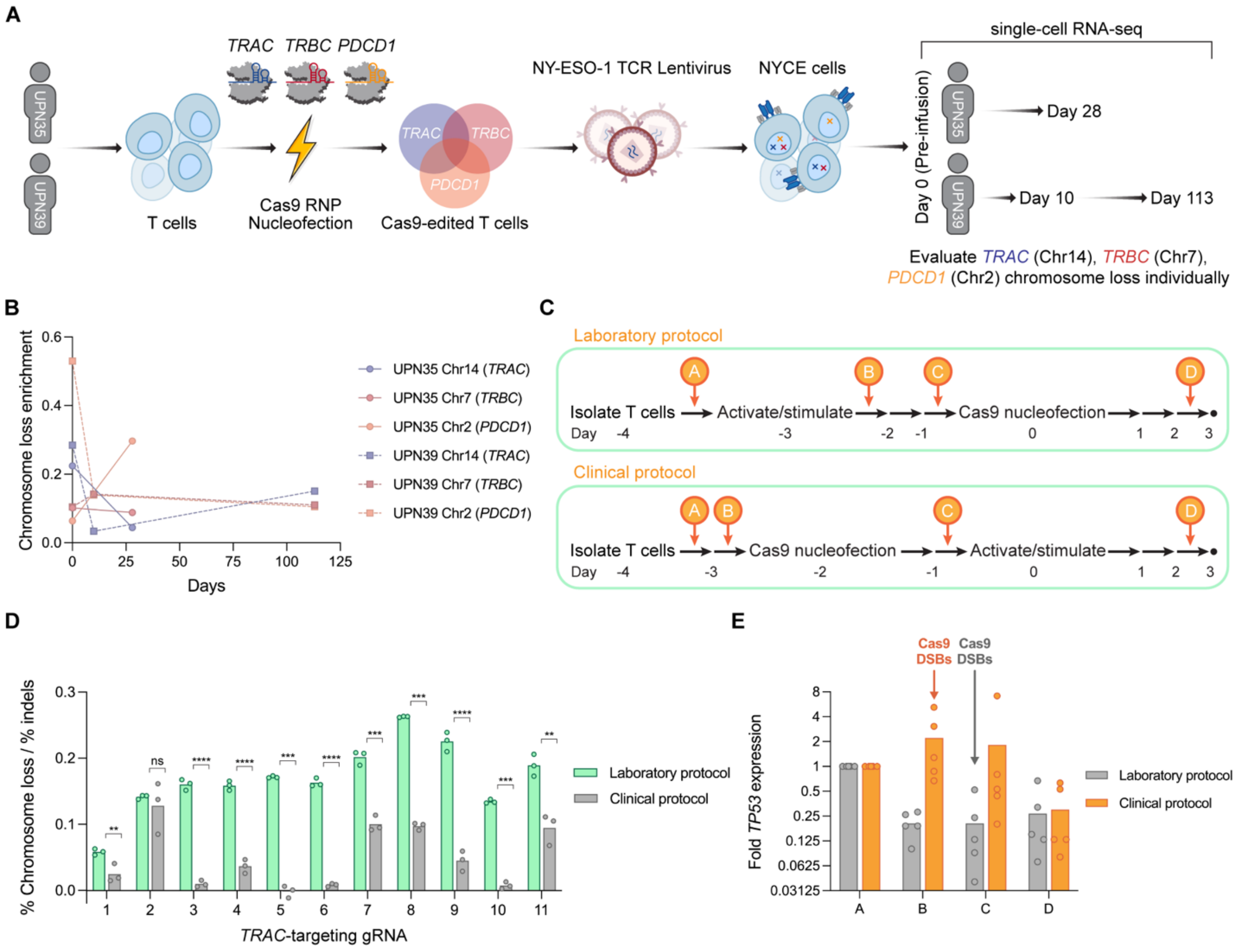


Figure 3.11: Clinical CRISPR-Cas9 genome editing protocol in patient T cells mitigates chromosome loss. **(A)** Strategy to investigate chromosome loss in two clinical trial patients with CRISPR-edited T cells. Two patients with refractory cancer had T cells isolated, electroporated with *TRAC, TRBC,* and *PDCD1*-targeting Cas9 RNPs, and transduced with a lentivirus encoding an NY-ESO-1 TCR (NYCE cells). Cells were subjected to scRNA-seq prior to infusion (Day 0) and as well as at different time points post-infusion (Days 10, 28, and/or 113). **(B)** Chromosome loss enrichment on chromosome 14 (*TRAC*), chromosome 7 (*TRBC*), or chromosome 2 (*PDCD1*) at different timepoints for both patients. Enrichment was calculated relative to non-targeted chromosomes (all chromosomes but 2, 7, and 14). Day 0 represents NYCE cells prior to infusion, while other later timepoints represent NYCE cells that were collected after circulation *in vivo*. **(C)** Diagram of two different protocols for Cas9 genome editing of primary human T cells. The laboratory protocol (top) consisted of activating/stimulating cells prior to Cas9 electroporation, and was used throughout this study. The clinical protocol (bottom) consisted of electroporating cells with Cas9 prior to activating/stimulating and is representative of our clinical trial. **(D)** Relative chromosome loss with 11 different *TRAC*-targeting gRNAs using the laboratory or clinical protocol in primary human T cells. Chromosome loss was normalized to the indel efficacy (see Extended Data Fig. 14b). *P*-values are from Welch's unpaired t-tests and from left to right are 0.008320, 0.111695, 0.000052, 0.000076, 0.000159, 0.000073, 0.000125, 0.000087, 0.000073, 0.000050, and 0.001416. **(E)** Fold *TP53* mRNA expression during the laboratory or clinical protocols for Cas9 genome editing of primary human T cells (n = 5 biological donors). Data points are the mean of two technical replicates. X-axis letters correspond to timepoints in Fig. 6c. "Cas9 DSBs" represents the timepoints in the laboratory or clinical protocols where Cas9 was electroporated into T cells to generate DSBs.

To investigate whether clinical manufacturing of a Cas9-edited adoptive T cell therapy results in levels of chromosome loss similar to those observed in our laboratory studies, we analyzed scRNA-seq data from NYCE cells of two patients at various timepoints throughout the clinical trial. Cells were collected from patient UPN35 prior to infusion and at day 28 post-infusion, while cells from patient UPN39 were collected prior to infusion as well as at days 10 and 113 post-infusion (Figure 3.11A). Similar to our laboratory experiments, we inferred gene dosage at each of the target chromosomes (*TRAC*, Chr14; *TRBC*, Chr7; *PDCD1*, Chr2) and looked for partial and whole chromosome loss. Surprisingly, we observed extremely low levels of chromosome loss at the targeted chromosomes, which were not enriched compared to background levels at non-targeted chromosomes (Figure 3.11B).

3.3.9 Order of operations during Cas9 genome editing impacts chromosome loss

We wondered whether the discrepancy between the high rates of chromosome loss in our laboratory studies versus the low rates in our clinical trial could be attributed to the engineered T cell manufacturing protocol. In our laboratory studies, we activated and stimulated T cells prior to introducing Cas9 RNP and generating DSBs, while in our clinical trial we introduced Cas9 RNP and generated DSBs prior to activating and stimulating the T cells (Figure 3.11C). We therefore performed Cas9 genome editing of *TRAC* in primary human T cells using our laboratory protocol (activation/stimulation followed by DSBs) and simulating our clinical protocol (DSBs followed by activation/stimulation) in parallel. Across all 11 *TRAC*-targeting gRNAs, we observed markedly lower levels of chromosome loss using our clinical protocol compared to our laboratory protocol (Figure 3.12A). However, genome editing with the clinical protocol on average resulted in ~14% lower indels compared to the laboratory protocol (Figure 3.12B). To control for this difference, we normalized the rate of chromosome loss to the rate of indels generated per gRNA and still observed a statistically significant reduction in chromosome loss with our clinical protocol as compared to our laboratory protocol with 10 out of 11 gRNAs (Figure 3.11D).

Previous studies have shown that p53, a key protein in cell cycle regulation and apoptosis, also regulates T cell activation. Downregulation of p53 is critical for murine T cell proliferation[276]. Additionally, p53-mediated apoptosis has been shown as a mechanism for selecting against aneuploid cells[277]. Therefore, we tested whether differences in manufacturing protocol influenced p53 levels, and how that related to the chromosome loss we observed. Expression of *TP53*, which encodes p53, was measured via RT-qPCR at different timepoints throughout both T cell genome editing protocols. Similar to what was observed in murine T cells, expression of p53 was lowered in both protocols after human T cell activation/stimulation (Figure 3.11E). The mean *TP53* expression across five biological donors was >10 times higher immediately prior to Cas9-induced DSBs in our clinical protocol compared to our laboratory protocol (Figure 3.11E). Thus, *TP53* expression during Cas9-induced DSBs inversely correlated with rates of chromosome loss in T cells between our two protocols. For our clinical protocol, the higher expression of this key DNA damage response factor during Cas9-induced DSBs could select against cells with chromosome loss and explain the dramatically lower rates we observed. Implementation of this modified protocol for Cas9 genome editing in T cells represents a simple adjustment that could substantially mitigate chromosome loss in future research and clinical studies.

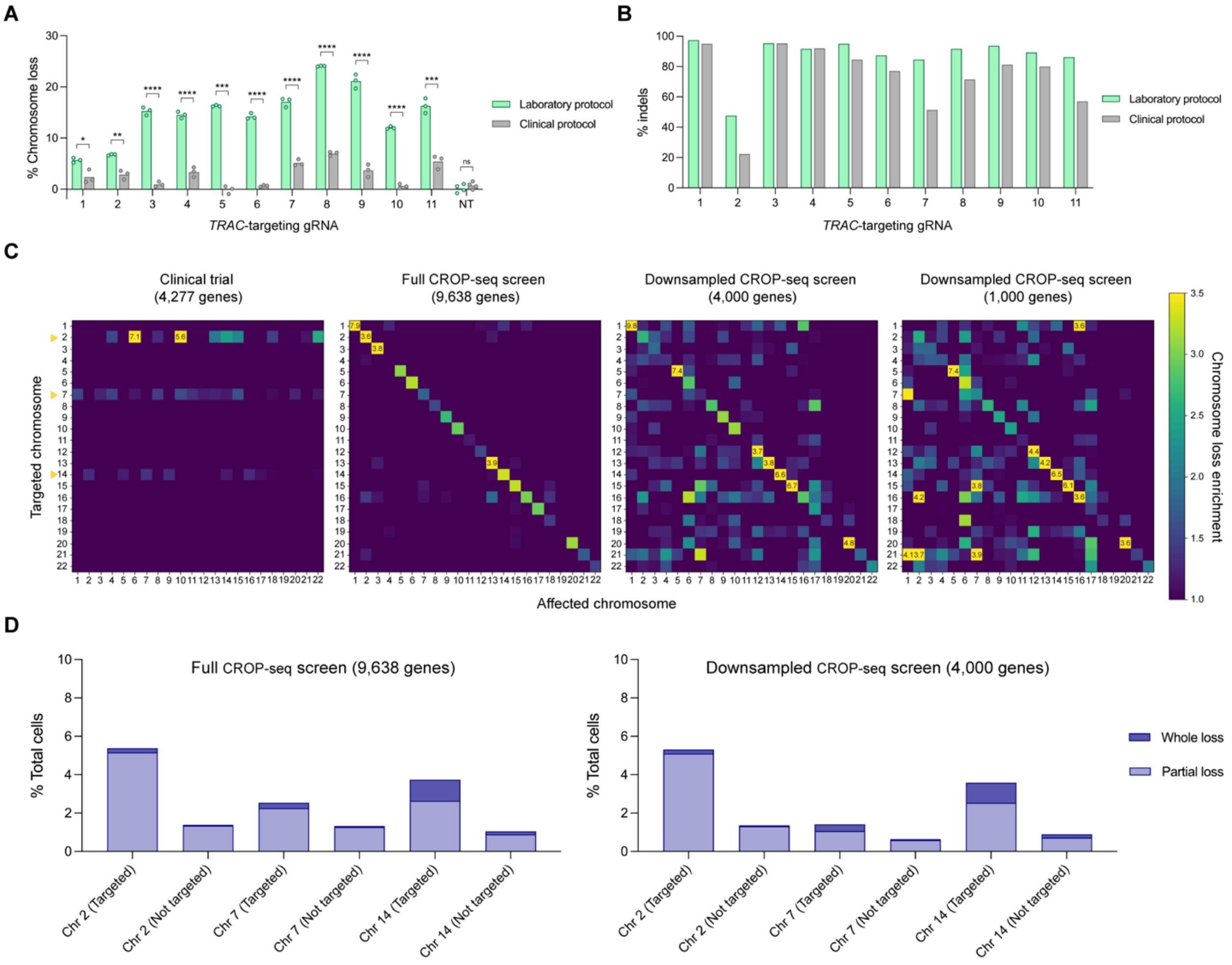


Figure 3.12: Clinical CRISPR-Cas9 genome editing protocol reduces chromosome loss in T cells. (A) Chromosome loss with 11 different *TRAC*-targeting gRNAs or a non-targeting gRNA (NT) using the laboratory or clinical protocol in T cells (n = 3). *P*-values are from Welch's unpaired t-tests and from left to right are 0.010220, 0.004303, 0.000063, 0.000083, 0.000170, 0.000083, 0.000063, 0.000063, 0.000063, 0.000031, 0.000224, and 0.079286. **(B)** Indels measured by next-generation sequencing at the *TRAC* locus by Cas9 genome editing with the laboratory or clinical protocol. **(C)** Downsampling analysis to investigate the influence of total genes on chromosome loss enrichment. Rows represent the chromosome targeted by Cas9 and its gRNA. Columns represent the chromosome analyzed for chromosome loss. Chromosomal loss enrichment for all clinical trial patients and timepoints was evaluated only when targeting chromosomes 2 (*PDCD1*), 7 (*TRBC*), and 14 (*TRAC*) (yellow arrows, left heatmap). 4,277 total genes were detected in the clinical trial dataset. The full CROP-seq screen dataset (9,638 genes, second from left) was downsampled to 4,000 genes (second from right) or 1,000 genes (right) to evaluate the influence of total genes on chromosomal loss enrichment. **(D)** Quantification of whole and partial chromosome loss at chromosomes 2, 7, and 14 from the CROP-seq screen. Chromosome loss was measured at the specific chromosome targeted by the Cas9 gRNA (targeted) or at chromosomes not targeted by the Cas9 gRNA (not targeted), from the full CROP-seq screen dataset (left) or the CROP-seq screen dataset downsampled to mimic the clinical trial dataset (right).

3.4 Discussion

In this study, we comprehensively investigated the frequency and consequences of Cas9-induced chromosome loss in primary human T cells, taking a genome-scale approach to understand what influences this phenomenon and investigating both pre-clinical and clinical T cell therapies. Targeting Cas9 across the *TRAC* locus, we estimated chromosome loss in ~5-20% of cells depending on the gRNA. We discovered that Cas9-induced chromosome loss was a generalizable phenomenon; chromosome loss was observed across the genome in an average of 3.25% of T cells that were targeted by Cas9. These T cells showed detectable levels of chromosome loss over 2-3 weeks of *ex vivo* culture, though they displayed a fitness and proliferative disadvantage. These disadvantages could cause cells without chromosome loss to outgrow those with chromosome loss, explaining the gradual reduction in this chromosomal aberration measured over the multi-week culture. Importantly, we still detected chromosome loss in nearly all conditions at our final timepoint, and this 2-3 week timeframe is similar to current clinical adoptive T cell therapy manufacturing protocols[101,255]. This suggests that T cells with Cas9-induced chromosome loss could persist throughout *ex vivo* manufacturing and end up in the final product administered to patients. In addition, continued efforts aim to further shorten the engineered T cell manufacturing process, which has been shown to improve T cell activity and persistence but could result in higher levels of chromosome loss[278].

To date, no study has investigated Cas9-induced chromosome loss in a clinical setting. In order to determine clinical significance, we generated CAR T cells using Cas9-mediated HDR, an approach being used in a growing number of clinical trials[66,67], and found a significant enrichment in chromosome loss compared to non-targeted cells. We also investigated Cas9-edited T cells of two patients enrolled in a first-in-human phase 1 clinical trial. We previously reported detectable levels of Cas9-induced translocations, another unintended genomic aberration, in these patient T cells, though levels reduced to the limit-of-detection after *in vivo* circulation[101]. Surprisingly, when we investigated patients' T cells for chromosome loss, we saw no enrichment above background levels, marking two of the few cases where we did not find Cas9-induced chromosome loss.

Comparing the results from our laboratory experiments (where substantial chromosome loss was detected) and the clinical trial (where we did not observe chromosome loss above background levels), there were multiple technical differences in the parameters used for chromosome loss estimation (Supplementary Note 3.2). We tried to account for these differences by downsampling the CROP-seq screen dataset so that its parameters were similar to the clinical trial dataset, which was sparser in data (Figure 3.12C). Even upon downsampling, our estimations of chromosome loss in the CROP-seq screen were comparable to the original complete dataset (Figure 3.12D). This supports the conclusion that biological rather than technical reasons explain the dramatic difference in chromosome loss estimation.

We considered and eliminated multiple factors that might correlate with or potentially explain Cas9-induced chromosome loss including Cas9 binding orientation, gRNA sequence, chromatin accessibility, and targeted gene or chromosome. Instead, we found that introducing Cas9-mediated DSBs prior to T cell activation/stimulation, a protocol used in our clinical trial but not in our laboratory experiments, influenced this phenomenon by significantly diminishing chromosome loss. It is possible that high levels of transcription in activated T cells during our laboratory protocol may predispose cells to chromosome loss due to genome instability caused by active transcription[279]. This effect could also be explained by levels of the DNA damage response protein p53 at the time of DSB generation, since we found *TP53* expression and chromosome loss were inversely correlated. Consistent with this hypothesis, a report in immortalized fibroblasts showed knockout of p53 increased Cas9-induced chromosomal truncations[215]. For engineering T cells, using the manufacturing protocol in which cells are activated after delivery of Cas9 could become standard practice to minimize chromosome loss in the manufactured product. This protocol adjustment does not require novel equipment, modification of Cas9 or its gRNA, or additional cost, meaning it can be easily and immediately integrated into clinical practice.

Recently, several other Cas9-mediated chromosomal abnormalities such as translocations[101], large deletions[213,214], loss of heterozygosity[217], and chromothripsis[212] have been reported. Of these, only methods and technologies for limiting translocations have been demonstrated, including serial rather than simultaneous multiplexed genome editing[211], use of Cas12 nucleases[280], fusion of Cas9 to an exonuclease to limit repeat cleavage[281], or utilizing base editors that do not generate DSBs[103]. Along with our modified clinical protocol, additional technologies could be developed to similarly mitigate chromosome loss.

CRISPR-based technologies that do not generate a DSB, such as base editors or epigenome editors, would likely avoid high levels of chromosome loss[45,81]. However, base editing can only modify one or a few nucleotides and epigenetic editing lacks permanence; neither of these technologies are ideal for permanent gene disruption or gene insertion. The use of CRISPR-Cas9 genome editing that creates DSBs is still highly advantageous and will continue to expand in clinical use. Therefore, mitigating genomic aberrations from DSBs, such as chromosome loss, will have substantial value to avoid potential genotoxicity in patients. Our comprehensive study suggests that although chromosome loss is a universal consequence of site-specific Cas9 genome editing, protocol adjustments and further exploration of underlying mechanisms can minimize its occurrence and impact.

3.4.1 Limitations of the study

To determine the generalizability of Cas9-induced chromosome loss, we performed a CROP-seq CRISPR screen targeting several genes on each somatic chromosome. The emergence of higher throughput scRNA-seq may allow this study to be expanded to a genome-wide screen in the future. Additionally, where possible we selected highly active and specific gRNAs from previously reported studies to include in our CROP-seq library. However, since we cannot reliably measure genome editing efficacy in our pooled format, it is possible that gRNAs with low chromosome loss detected simply had low cleavage activity. Finally, sequencing quality and computational gene calling varied between experiments (Supplementary Note 3.2). Since this could influence our chromosome loss measurements, we primarily displayed relative chromosome loss enrichment, which was normalized to a non-targeting gRNA or untreated sample, rather than absolute chromosome loss.

3.5 Materials and methods

3.5.1 Cell culture

Primary adult peripheral blood mononuclear cells (PBMCs) were obtained as cryopreserved vials from Allcells Inc. CD3+ T cells were isolated from PBMCs using EasySep Human T Cell Isolation Kits (StemCell Technologies) according to the manufacturer's instructions. Isolated CD3+ T cells were cultured in X-Vivo 15 medium (Lonza) with 5% fetal bovine serum (FBS) (VWR), 50 μM 2-mercaptoethanol, and 10 mM N-acetyl L-cysteine (Sigma-Aldrich). One day post-isolation, CD3+ T cells were activated and stimulated with a 1:1 ratio of anti-human CD3/CD28 magnetic Dynabeads (Thermo Fisher) to cells, as well as 5 ng/mL IL-7 (PeproTech), 5 ng/mL IL-15 (R&D Systems), and 300 U/mL IL-2 (PeproTech) for three days. After the initial activation and stimulation, magnetic beads were removed and T cells were cultured in medium with 300 U/mL IL-2. Medium was replaced every other day and T cells were maintained at a density of ~0.5-1x10^6 cells/mL.

For CAR T cell experiments, primary adult PBMCs were obtained as Leukopaks (StemCell Technologies) from deidentified healthy donors and cryopreserved in RPMI medium supplemented with 20% human serum and 10% DMSO. T cells were isolated as described before and cultured in X-Vivo 15 medium supplemented with 5% human serum, 5 ng/mL IL-7 (Miltenyi Biotec), and 5 ng/mL IL-15 (Miltenyi Biotec). Immediately after isolation, T cells were stimulated for two days with anti-human CD3/CD28 magnetic Dynabeads (Thermo Fisher) using a 1:1 bead-to-cell ratio.

3.5.2 Cas9 ribonucleoprotein electroporation

100 pmol Alt-R crRNA and 100 pmol Alt-R tracrRNA (IDT) were diluted in IDT Duplex Buffer, incubated at 90 °C for 5 min, and then slow cooled to room temperature (Supplementary Tables 1, 7-8). 50 pmol *S. pyogenes* Cas9 V3 (IDT) was diluted in RNP buffer (20 mM HEPES, 150 mM NaCl, 10% Glycerol, 1 mM $MgCl_2$, pH 7.5). Cas9 and duplexed gRNA (1:2 molar ratio) were incubated at 37° C for 15 min. Primary human T cells were washed once with PBS (-/-) before 250,000 cells were resuspended in P3 Buffer (Lonza). 50 pmol Cas9 RNP was added to the cells before electroporation in a Lonza 4D-Nucleofector with pulse code EH-115. X-Vivo 15 medium with 300 U/mL IL-2 was added to the electroporated cells before a 30 min recovery at 37 °C. Electroporated T cells were plated at a density of ~0.5-1x10^6 cells/mL in 96-well U-bottom plates.

3.5.3 TCR flow cytometry

T cells were resuspended in Cell Staining Buffer (BioLegend) with Ghost Dye Red 780 (1:1,000, TonboBio) and anti-human TCR α/β Brilliant Violet 421 (1:100, BioLegend). Cells were stained for 30 min at 4 °C in the dark. After staining, cells were washed with Cell Staining Buffer and analyzed on a Thermo Fisher Attune NXT Flow Cytometer with an autosampler. Over 20,000-100,000 cells were routinely collected and analyzed with FlowJo.

3.5.4 Next-generation sequencing of *TRAC* genome editing

Genomic DNA from T cells was extracted with QuickExtract DNA Extraction Solution (Lucigen) by incubating resuspended cells for 10 minutes at room temperature before heating lysates at 65 °C for 20 minutes then 95 °C for 20 min. The region of *TRAC* containing the Cas9 target site was amplified from genomic DNA with Q5 High-Fidelity DNA polymerase (NEB) to add universal adaptors (Supplementary Table 2). Amplicons were cleaned with SPRIselect beads (Beckman Coulter) before a second round of PCR was performed to add unique i5 and i7 Illumina indices to each sample. Subsequent amplicons were cleaned again, and libraries were sequenced on an Illumina iSeq 100 (2x150 bp). FASTQ files were trimmed, merged, and analyzed for indels with CRISPResso2 (crispresso2.pinellolab.org)[53]. For non-targeting conditions, sequencing from cells receiving Cas9 and a non-targeting gRNA was used as input to check for indels at a given region.

3.5.5 Single-cell RNA sequencing with MULTI-seq barcoding

Four days post-electroporation, T cells were labeled via MULTI-seq as previously described[260]. Briefly, a lipid-modified oligonucleotide (LMO) was combined with a unique oligonucleotide barcode (Supplementary Tables 3 and 6) at a 1:1 molar ratio in PBS (-/-). 500,000 cells or fewer were washed twice with PBS (-/-) and then resuspended in PBS (-/-). The LMO/barcode solution was mixed with each cell suspension and incubated on ice for 5 minutes before addition of a co-anchor LMO and incubation on ice for an additional 5 minutes. Cold 1% BSA in PBS (-/-) was added to sequester free LMOs before washing cells twice with cold 1% BSA in PBS (-/-). Uniquely labeled cells were pooled in equal numbers in 1% BSA in PBS (-/-) to a final concentration of ~1,600 cells/mL. 10x Genomics Chromium Next GEM Single Cell 3' Gene Expression Kits (v3.1) were utilized according to the manufacturer's instructions, with the following modifications. Lanes of a standard Chromium chip were "super loaded" with ~50,000 cells to yield a target cell recovery of ~25,000 cells. During the cDNA amplification, 1 μL of a 2.5 μM MULTI-seq primer (see McGinnis et al.) was added. Supernatant from the cDNA bead cleanup was saved because it contained the MULTI-seq barcode amplicon. Supernatants were further cleaned by addition of SPRIselect beads and isopropanol with a conventional magnetic bead cleanup protocol. 3.5 ng of each cleaned amplicon was used in a PCR reaction to add sequencing indices; the reactions included KAPA HiFi HotStart ReadyMix (Roche), a unique i5 primer, and a unique RPI i7 primer (Supplementary Table 3). The PCR reactions were cleaned with SPRIselect beads before final library QC. Gene expression and MULTI-seq barcode libraries were pooled 6:1 (molar ratio) and sequenced on an Illumina NovaSeq 600 S1 Flow Cell.

3.5.6 Single-cell RNA sequencing analysis

Cell Ranger (v7.0) was used to process Chromium single cell data. cellranger count was performed with the parameters --r1-length=28 and --r2-length=90. For the first *TRAC*-targeting experiment the --force-cells parameter was set to 15,000 cells. To demultiplex the different pools of cells using the MULTI-seq barcode, cellranger multi was performed. The results from the different pools were aggregated using cellranger aggr. The results from cellranger were parsed with scanpy and converted to a h5ad file format. The demultiplexed results were added as metadata to the h5ad file.

For the CROP-seq screen, we counted the number of reads within each cell aligning to each of the gRNAs used in the screen. To determine which cells were targeted by a single, unique gRNA, we tested whether the gRNA with the highest number of reads had significantly more reads than the second highest. Specifically, let c_1 and c_2 be the number of reads for the first and second most common gRNAs in a cell, respectively. To test whether c_1 is significantly greater than c_2, we calculated a *P*-value based on a binomial distribution with parameters $n = c_1 + c_2$ and $p = 0.5$ (i.e. $x \sim B(c_1 + c_2, 0.5)$). If the probability of $x \geq c_1$ was smaller than 0.05, the cell was determined to be transduced by a single gRNA.

3.5.7 Quantification of chromosome loss from scRNA-seq

To assess the dosage of each gene in each cell, inferCNV of the Trinity CTAT project (https://github.com/broadinstitute/infercnv) was executed in R (version 4.1) with default parameters over the h5ad dataset created by cellranger (see previous section) for every scRNA-seq dataset[261]. Each cell with a gRNA was labeled as a “treatment” and each cell with a non-targeted gRNA was labeled as a “control.” To successfully run inferCNV for the CROP-seq screen, inferCNV was performed in multiple batches of 30,000 cells.

The output of inferCNV was the estimated dosage for each gene; according to the software’s specifications, values below 0.95 were considered loss of at least one copy of the gene. inferCNV values for each gene were binarized as <0.95 or ≥ 0.95. Each chromosome in each cell was then searched for the interval between two genes that maximizes the difference between the average binarized inferCNV values on either side of the interval. This interval was the candidate breakage point for a particular chromosome in a cell.

We used the inferCNV values for all genes on a given chromosome within each cell (with respect to each of the 22 somatic chromosomes) to estimate the loss status of that chromosome. Specifically, we estimated whether there was 1) no chromosome loss, 2) whole chromosome loss, or 3) partial chromosome loss. If at least 70% of a minimum 150 genes to either the left or right of the candidate breakage point were below the 0.95 threshold, but less than 70% of the genes on the other side were below the threshold, the cell was labeled as partial chromosome loss for that chromosome. Otherwise, if at least 70% of all the genes throughout the entire chromosome were below the threshold, the cell was labeled as having whole chromosome loss for that chromosome. If neither were true, the cell was labeled as having no chromosome loss for that chromosome.

Downsampling of the CROP-seq screen dataset was performed to assess the dependency of chromosome loss enrichment on the number of genes in the inferCNV output. To do this, 4,000 or 1,000 genes contained within the 9,639 gene output of the full

CROP-seq inferCNV output were randomly sampled. Our chromosome loss calling pipeline, as described above, was then performed on these downsampled datasets.

3.5.8 Droplet digital quantitative PCR

Genomic DNA was collected from T cells at different times post-electroporation with QuickExtract DNA Extraction Solution, identical to as described earlier. The ddPCR setup was similar to what has been previously described[272]. For multiplexed ddPCR, two ~200 bp amplicons for each target gene were designed (Supplementary Tables 4, 7, and 8). Amplicon 1 was located proximal to the centromere and utilized a hexachlorofluorescein-labeled (HEX) oligonucleotide probe (PrimeTime qPCR probes, Zen double quencher, IDT). Amplicon 2 was located ~100-200 bp away from amplicon 1, was distal relative to the centromere, and utilized a 6-fluorescein-labeled (FAM) oligonucleotide probe (PrimeTime qPCR probes, Zen double quencher, IDT). Amplicon 1 served as a control, which should be unaffected by Cas9 genome editing or chromosome loss and would signal whether the gene of interest was in a given droplet. Amplicon 2 spanned the Cas9 target site, with the probe located ~30-60 bp away from the cleavage site. If the target site was not successfully repaired after Cas9 cleavage, amplicon 2 would not be able to be amplified and the FAM probe would be unable to dissociate from its quencher. ddPCR reactions were assembled with ddPCR Supermix for Probes (No dUTP, Bio-Rad), 900 nM of each primer, 250 nM of each probe, and 10-30 ng of genomic DNA. Droplets were formed using a Bio-Rad QX200 Droplet Generator following the manufacturer's instructions before thermal cycling. The following day, ddPCR droplets were analyzed on a Bio-Rad QX200 Droplet Reader. Data were analyzed with the QX Manager Software (Bio-Rad), and thresholds were set manually based on wells with untreated samples. The percentage of alleles with chromosome loss was calculated based on droplets that had the target amplicon 1 (HEX+) but were unable to produce the neighboring amplicon 2 (FAM-). The equation utilized is as follows: $\%\, Chromosome\ loss = 100\ \times \left(1 - \frac{[FAM]}{[HEX]}\right)$.

3.5.9 Genome-scale CROP-seq CRISPR screen design

The CROP-seq library was designed to contain multiple gRNAs that target multiple genes on every chromosome. When possible, validated gRNA sequences from previous publications were utilized (Supplementary Table 5). The gRNA library was ordered as an oPool oligo pool (IDT) and Golden Gate cloned into a custom CROP-seq vector that co-expressed GFP. To analyze the library, primers were used to amplify the gRNA spacer from either the plasmid library or genomic DNA library before sequencing on an Illumina iSeq. MAGeCK (https://sourceforge.net/p/mageck/wiki/Home/) was used to quantify the representation of each gRNA in the library[282].

3.5.10 CROP-seq CRISPR screen lentiviral production

For lentivirus production, Human Embryonic Kidney 293T (HEK293T) cells were cultured in Dulbecco's Modified Eagle Medium (Gibco) with 10% FBS and 1% penicillin/streptomycin (Gibco). HEK293Ts were transfected at 70-90% confluency with

10 μg CROP-seq gRNA plasmid, 10 μg Gag-pol expression plasmid (psPax2, gift from Didier Trono, Addgene plasmid #12260), and 1 μg pCMV-VSV-G plasmid (gift from Bob Weinberg, Addgene plasmid #121669) using polyethylenimine (PEI, Polysciences Inc.) at a 3:1 PEI:plasmid ratio. Approximately 6-8 hours after transfection, the medium was aspirated from cells and replaced with Opti-Mem (Gibco). Supernatant containing lentivirus was collected 48 hours after transfection, the medium was replaced, and medium was collected once more after an additional 48 hours. Viral supernatants were filtered through a 0.45 μm PES membrane bottle top filter (Thermo Fisher) and then concentrated with Lenti-X Concentrator (Takara) according to the manufacturer's instructions. Purified and concentrated lentivirus was used immediately or stored at -80° C. Lentivirus was titered by counting the number of initially transduced cells, adding serial dilutions of lentivirus to primary human T cells, and measuring the percentage of GFP+ cells after three days (only in conditions with <30% GFP+ cells to ensure a majority were single transduction events).

3.5.11 CROP-seq CRISPR screen

For the CROP-seq screen, primary human T cells were isolated and stimulated as stated previously. 24 hours after stimulation, lentivirus was added to the cells at a multiplicity of infection (MOI) of ~0.3. MOI was confirmed via flow cytometry two days later, at day three post-stimulation. Dynabeads were then removed from T cells and Cas9 was electroporated as stated previously. For the full CROP-seq library experiment, four days post-electroporation, T cells were subject to fluorescence-activated cell sorting (FACS) on a Sony SH800S cell sorter to enrich for GFP+ cells. Genomic DNA was harvested from a small number of cells, as previously described, to assess the library representation. The rest of the live/GFP+ cells were arbitrarily divided into six pools and subject to MULTI-seq barcoding and 10x Genomics scRNA-seq, as previously described. The CROP-seq gRNA was enriched from the resulting cDNA similar to what has been previously described (Supplementary Table 6)[264]. Briefly, 25 ng of cDNA was added to eight separate KAPA HiFi HotStart ReadyMix PCR reactions and amplified for 10 cycles with an annealing temperature of 65 °C to enrich for the gRNA. Individual PCR reactions were pooled together and cleaned with SPRIselect beads. 8 μL of cleaned PCR1 product was added to a second KAPA HiFi HotStart ReadyMix PCR reaction and amplified for 10 cycles with an annealing temperature of 65 °C to add Illumina sequencing adaptors. Gene expression, MULTI-seq barcode, and CROP-seq enrichment libraries were sequenced on an Illumina NovaSeq 600. Multiple iterations of library sequencing were concatenated to achieve the desired sequencing depth.

3.5.12 Strand and MMEJ analyses

Each gRNA was mapped using the GRCh38 genome assembly. A two-sided Fisher's Exact Test was performed to determine whether gRNAs binding distal or proximal to the centromere, relative to the gRNA spacer sequence, affected chromosome loss. MMEJ analyses were performed using inDelphi (indelphi.giffordlab.mit.edu)[203]. The cell type was set to K562s and the MMEJ strength was measured for each unique gRNA sequence.

3.5.13 Differential gene expression analysis

To identify genes differentially expressed between cells with and without chromosome loss, we used the memento algorithm with default parameters (capture rate = 0.07)[283]. We tested each gene for differential expression with respect to each of the 22 somatic chromosomes separately, and only reported genes that were consistently over- or underexpressed across most chromosomes. We further ensured that the differences we observed were specific to the tested gene and were not the result of overall lower or greater levels of gene expression in cells with chromosome loss. To do this we accounted for the total expression of transcripts sharing the same chromosome as the tested gene by including the overall count of all transcripts on that chromosome as a covariate. This covariate was defined with respect to the chromosome containing the gene tested for differential expression and not with respect for the chromosome determining the two compared groups (cells with or without loss of that chromosome). Since memento supports only discrete covariates, we discretized the total transcript count into 10 decile bins.

We corrected the results of memento for multiple testing using FDR over the combined set of all tested genes over all tested chromosomes. We considered a gene to be statistically significant with respect to a chromosome (i.e. the gene to be over- or underexpressed in cells losing that chromosome) if its corrected *P*-value was below 0.05. Accordingly, we assigned the significance status of that gene-chromosome combination to be 1, 0, or -1 if it was significantly overexpressed, not significant, or under expressed, respectively. We then assigned each gene a total score between -22 and 22 by summing the significance status of that gene with respect to each of the 22 somatic chromosomes. 613 genes obtained a total score ≥ 5 and were considered the top overexpressed genes, while 590 obtained a total score ≤ -5 and were considered the top underexpressed genes.

We identified pathways enriched among the 613 top overexpressed genes by searching through the pathway terms defined in the KEGG database using the GSEApy Python package[284,285].

3.5.14 Cell cycle analysis

Cell cycle states were defined using data and methods as previously described[286].

3.5.15 Epigenetic analyses

Datasets (ENCFF233TCT, ENCFF055FYI, and ENCFF129GAM) corresponding to activated T cells from a male donor (43 years old) were selected from the ENCODE Portal (www.enccodeproject.org)[287,288]. The presence of open chromatin from ATAC-seq data and the location of epigenetic marks (H3K9me3 and H3K36me3) from ChIP-seq data were determined within a 75 bp window around the GRCh38 coordinates of each gRNA and a *P*-value $<10^{-5}$, according to best practices[289].

3.5.16 T cell proliferation tracking

After isolation and stimulation, primary human T cells were electroporated with Cas9 RNPs identical to what was described earlier. Immediately after electroporation recovery, cells were pelleted and resuspended in 5 μM CellTrace Violet (Invitrogen). Cells were incubated in CellTrace Violet for 20 min at 37 °C, prior to diluting in 4x volume of complete X-Vivo 15 medium to absorb unbound dye and incubating again for 5 min at 37 °C. Cells were pelleted and resuspended in complete X-Vivo 15 with 300 U/mL IL-2. T cells were passaged every other day to refresh medium and maintain a density of ~0.5-1x10^6 cells/mL. Cells were sorted on a BD FACSAria II to obtain the approximate bottom and top quartile of cells according to CellTrace Violet signal.

3.5.17 Cas9-mediated CD5 homology-directed repair

HA tag insertion was achieved with either a single-stranded DNA HDR template (ssDNA HDRT), a double-stranded DNA HDR template (dsDNA HDRT), or a single-stranded DNA HDR template with Cas9 target sequences (ssCTS HDRT) (Supplementary Note 3)[273]. Equimolar HDRT oligonucleotides were diluted in IDT Duplex Buffer, heated to 95° C for 5 minutes, then allowed to slow cool to room temperature. 100 pmol of HDRT was added to Cas9 RNP electroporations of primary human T cells, identical to as described above. Cells were analyzed on a Thermo Fisher Attune NXT Flow Cytometer with an autosampler, identical to as described above, except with the antibodies anti-human CD5 (UCHT2)-PE (1:100, Invitrogen) and anti-HA (6E2)-AF647 (1:100, Cell Signaling Technology). Over 20,000-100,000 cells were routinely collected and analyzed with FlowJo.

3.5.18 Fluorescence-activated cell sorting of CD5, CD81, and CD3E

Primary human T cells were electroporated with Cas9 RNPs targeting *CD5* identical to what was described before. Seven days post-electroporation, cells were stained with anti-human TCR α/β Brilliant Violet 421 (BioLegend), anti-human CD5 (UCHT2)-PE (Invitrogen), anti-human CD81 (5A6)-FITC (BioLegend), and anti-human CD3E (SK7)-APC (Invitrogen), all at a 1:100 dilution. Cells were sorted on a BD FACSAria II to isolate different populations (Supplementary Note 1).

3.5.19 CAR adeno-associated virus production

An AAV transgene plasmid encoding the inverted terminal repeats, a 1928ζ CAR, a truncated human EGFR (EGFRt) tag, and *TRAC* homology arms for HDR was used as previously described (Supplementary Note 4)[274]. The AAV plasmid was packaged into AAV6 by transfection of HEK293T cells together with pHelper and pAAV Rep-Cap plasmids using PEI. The AAVs were purified using iodixanol gradient ultracentrifugation. The titration of the AAV was determined by quantitative PCR on DNaseI (NEB) treated and proteinase K (Qiagen) digested AAV samples, using primers against the left homology arm. The quantitative PCR was performed with SsoFast EvaGreen Supermix (Bio-Rad) on a StepOnePlus Real-Time PCR System (Applied Biosystems).

3.5.20 CAR T cell production

gRNAs targeting exon 1 of the *TRAC* locus (*TRAC* gRNA 12), the intron preceding the *TRAC* locus (*TRAC* gRNA 13), or a non-targeting control gRNA were purchased from Synthego and resuspended in TE buffer (Supplementary Table 9). Cas9 RNP was generated by incubating 60 pmol of Cas9 protein with 120 pmol sgRNA. T cells were counted, resuspended in P3 buffer at 2x10^6 per 20 µL, mixed with 3 µL of RNPs and added to a 96-well nucleofection plate. Cells were electroporated using a Lonza 4D-Nucleofector 96-well unit with the EH-115 protocol and immediately recovered by adding pre-warmed X-Vivo 15 medium without human serum. Recombinant AAV6 encoding the HDR template was added to the culture 30 to 60 min after electroporation at an MOI of 10^5, and incubated with the cells overnight. The day after the electroporation and transduction, edited cells were resuspended in medium and expanded using standard culture conditions, keeping a density of 10^6 cells/mL. TCR disruption and CAR HDR efficiency was evaluated by flow cytometry by staining the TCR with anti-TCRα/β (BW242/412)-PE (1:50, Miltenyi) and the CAR with goat anti-mouse IgG (H+L) AlexaFluor 647 Fab (1:100, Jackson ImmunoResearch).

3.5.21 CAR T cell scRNA-seq

CAR T cells were harvested at two time points after independent electroporations (day four and day seven post-electroporation). TotalSeq-A0251-1 anti-human Hashtag reagents (BioLegend) were used to label different cell conditions. For the experiment, 500,000 cells from each condition were labeled with the hash antibodies in Cell Staining Buffer at 4 °C for 30 min. After labeling, cells were washed three times with Cell Staining Buffer at 4 °C and then resuspended in PBS (-/-) containing 0.04% BSA. Labeled cells were pooled and 50,000 cells were "super loaded" into four lanes (two lanes for day four samples and the other two lanes for day seven samples) of a 10X Chromium Single-Cell G Chip. A 10x Genomics Chromium Next GEM Single Cell 3' Gene Expression Kit (v3.1) was utilized according to the manufacturer's instructions, and the subsequent library was sequenced on an Illumina NovaSeq 600 S4 Flow Cell.

3.5.21 CAR T cell scRNA-seq

CAR T cells were harvested at two time points after independent electroporations (day four and day seven post-electroporation). TotalSeq-A0251-1 anti-human Hashtag reagents (BioLegend) were used to label different cell conditions. For the experiment, 500,000 cells from each condition were labeled with the hash antibodies in Cell Staining Buffer at 4 °C for 30 min. After labeling, cells were washed three times with Cell Staining Buffer at 4 °C and then resuspended in PBS (-/-) containing 0.04% BSA. Labeled cells were pooled and 50,000 cells were "super loaded" into four lanes (two lanes for day four samples and the other two lanes for day seven samples) of a 10X Chromium Single-Cell G Chip. A 10x Genomics Chromium Next GEM Single Cell 3' Gene Expression Kit (v3.1) was utilized according to the manufacturer's instructions, and the subsequent library was sequenced on an Illumina NovaSeq 600 S4 Flow Cell.

3.5.22 Laboratory versus clinical T cell manufacturing

For the laboratory protocol, T cells were activated and stimulated identical to what was described earlier. After electrporation of Cas9 RNP, T cells were cultured in X-Vivo 15 medium with 5 ng/mL IL-7, 5 ng/mL IL-15, and 300 U/mL IL-2.

For the clinical protocol, we followed a protocol similar to what was used in our phase 1 clinical trial[101]. After T cell isolation, cells were cultured in X-Vivo 15 medium with 5 ng/mL IL-7 and 5 ng/mL IL-15 for two days. Non-activated T cells were electroporated with 50 pmol Cas9 RNP using a Lonza 4D-Nucleofector with pulse code EH-115. After electroporation, cells were incubated in X-Vivo 15 medium with 5 ng/mL IL-7 and 5 ng/mL IL-15 for a 30 min recovery at 37 °C. Electroporated T cells were plated at a density of ~0.5-1x10^6 cells/mL in 96-well U-bottom plates. Two days after electroporation, cells were counted and activated/stimulated with a 1:1 ratio of anti-human CD3/CD28 magnetic Dynabeads to cells, as well as 5 ng/mL IL-7, 5 ng/mL IL-15, and 300 U/mL IL-2 for an additional three days.

3.5.23 T cell RT-qPCR

During the laboratory or clinical T cell manufacturing protocols, 500,000 cells were periodically pelleted and resuspended in TRIzol (Invitrogen). RNA was isolated via phenol-chloroform extraction, precipitated by addition of isopropanol, washed with 75% ethanol, and resuspended in nuclease-free water. Isolated RNA was treated with TURBO DNase (Invitrogen) and SUPERase-In RNase inhibitor (Thermo Fisher) for 30 min at 37 °C before addition of DNase Inactivation Reagent according to the manufacturer’s instructions. DNA-free RNA underwent cDNA synthesis using SuperScript III Reverse Transcriptase (Invitrogen) and Random Primers (Promega) according to the manufacturer’s instructions. qPCRs were performed with the resulting cDNA using iTaq Universal SYBR Green Supermix (Bio-Rad) on a Bio-Rad CFX96 Real-Time PCR Detection System (Supplementary Table 10). *TP53* expression levels were normalized to the expression levels of the housekeeping gene *GAPDH,* and to timepoint A (where the laboratory and clinical protocols start identically) using the $\Delta\Delta$Ct method.

3.6 Accession codes

This work did not generate any data with accession codes.

3.7 Acknowledgments

The authors would like to thank Jennifer Hamilton, Matthew Kan, Elizabeth Abby Stahl, and David Colognori for thoughtful discussion throughout this project. We thank Christopher McGinnis and Zev Gartner for providing the MULTI-seq reagents; Yoon Gi Justin Choi (QB3 Functional Genomics Laboratory) for assistance with the scRNA-seq; Scarleth Chalen Ulloa for assistance with the FACSAria; Netra Krishnappa (IGI Center for Translational Genomics), Carrianne Miller (QB3 Genomics Sequencing Laboratory), and Eric Chow (UCSF Center for Advanced Technology) for assistance with NGS; MinCheol Kim for assistance with the differential gene expression analysis; Francois

Aguet for assistance with the cell cycle analysis; and Brian Shy for assistance with the CD5 HDR assay. We thank the patients, their families, and the clinical study team at the Abramson Cancer Center Clinical Research Unit for their involvement in the clinical trial. We thank all members of the Satpathy, Cate, Eyquem, Fraietta, June, Chang, Ye, and Doudna laboratories for their input and advice. C.A.T. is supported by a National Institutes of Health (NIH) Ruth L. Kirschstein National Research Service Award F31 Pre-Doctoral Fellowship (National Heart, Lung, and Blood Institute, F31HL156468-01) and the Siebel Scholarship (Siebel Foundation). B.Y. is supported by the Parker Institute for Cancer Immunotherapy Parker Bridge Fellowship and the V Foundation. This project was supported by the Parker Institute for Cancer Immunotherapy (A.T.S., J.E., C.H.J., H.Y.C., and C.J.Y.), the NIH National Institute of General Medical Sciences (R01-GM065050, J.H.D.C.), the NIH National Cancer Institute (R35CA209919, H.Y.C), the NIH Centers for Excellence in Genomic Science (RM1HG007735, H.Y.C; RM1HG009490, J.A.D.), the Howard Hughes Medical Institute (H.Y.C. and J.A.D.), the NIH National Institute of Arthritis and Musculoskeletal and Skin Diseases (R01AR071522, C.J.Y.), the NIH National Institute of Allergy and Infectious Diseases (R01AI136972, C.J.Y.), the Chan Zuckerberg Initiative and Chan Zuckerberg Biohub (C.J.Y.), and the NIH Somatic Cell Genome Editing Program of the Common Fund (U01AI142817-02, J.A.D.).

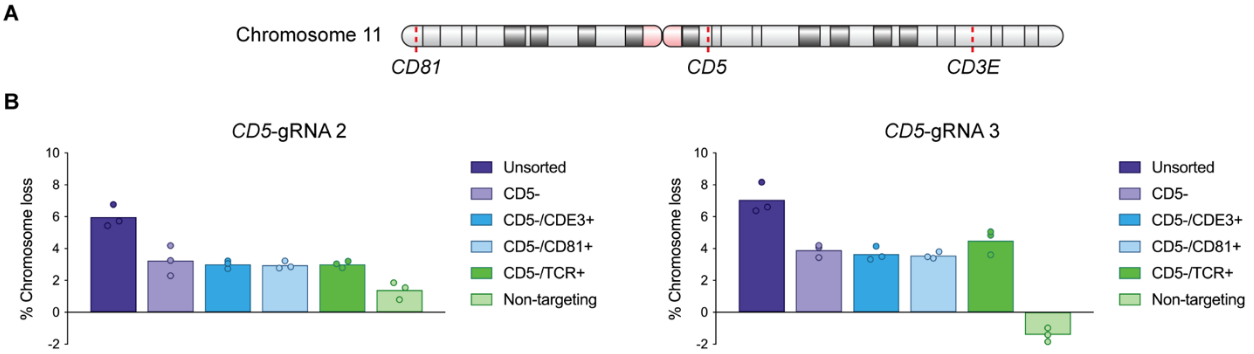
3.8 Author contributions

C.A.T., J.H.D.C., and J.A.D. conceived the study with subsequent input from N.B., R.B., and C.J.Y. C.A.T., B.H., C.C., J.L., and Y.S. performed the T cell experiments and generated the laboratory scRNA-seq libraries. C.R.H. provided technical support on the clinical protocol. N.B., R.B., M.T., and T.M. analyzed the laboratory scRNA-seq data. K.R.P. and Y.Q. generated the clinical trial scRNA-seq libraries. M.T., T.M., and B.Y analyzed the clinical trial scRNA-seq data. A.T.S., E.A.S., J.H.D.C., J.E., J.A.F., C.H.J., H.Y.C., C.J.Y., and J.A.D. supervised the project. C.A.T., N.B., C.J.Y., and J.A.D. wrote the manuscript with input from all other authors.

3.9 Supplementary information

Supplementary Note 3.1: T cell surface marker enrichment does not select for cells without chromosome loss

While targeting *CD5* on chromosome 11, we realized that the T cell surface proteins CD3E and CD81 were also encoded on chromosome 11. *CD5* is located near the centromere on the q arm, *CD3E* is located further down the q arm, and *CD81* is located on the p arm (Supplementary Note 1A). Therefore, we hypothesized that T cell surface markers could be used to identify cells with Cas9-induced chromosome loss. Targeting *CD5* with Cas9 should result in disruption of CD5 expression. Additionally, if partial chromosome loss occurs, *CD3E* should unintentionally be lost and CD3E should be disrupted. Finally, if whole chromosome loss occurred, *CD81* on the opposite arm should also be unintentionally lost and CD81 should be disrupted. After T cells were nucleofected with Cas9 and one of two different *CD5*-targeting gRNAs, we used FACS to sort different populations of cells based on CD5, CD3E, and/or CD81 expression. While we expected sorting for CD5-/CD3E+ or CD5-/CD81+ cells may enrich for *CD5*-targeted cells without chromosome 11 loss, we did not observe any significant difference in these populations (Supplementary Note 3.1B). Haplosufficiency at any of these loci could explain why protein expression does not correlate with chromosome loss.



Supplementary note 3.1: (A) Schematic of chromosome 11. *CD5* is located near the centromere on the q arm, *CD3E* is located further down the q arm, and *CD81* is located on the p arm. **(B)** Chromosome loss via ddPCR in T cells treated with one of two *CD5*-targeting gRNAs (*CD5*-gRNA 2, *CD5*-gRNA 3) (n = 3). Unsorted = bulk cells edited at *CD5* by Cas9; non-targeting = cells treated with Cas9 complexed with a non-targeting gRNA. For all other conditions, cells edited at *CD5* by Cas9 were subject to FACS to analyze the presence of various surface markers. The TCR (*TRAC*, chromosome 14) served as a non-chromosome 11 control.

Supplementary Note 3.2: scRNA-seq and gene dosage analyses to estimate chromosome loss

In total, we analyzed four scRNA-seq datasets of T cells edited with Cas9. We attempted to perform uniform analyses across all these experiments; however, there were several differences in the parameters of the scRNA-seq experiments. Our three laboratory experiments utilized 10x Genomics Chromium Next GEM Single Cell 3’ Kits while the clinical trial utilized 10x Genomics Chromium Next GEM Single Cell 5’ Kits. In addition to the two clinical trial patients reported in this study, we also analyzed scRNA-seq data from a third clinical trial patient, UPN07, but chose to omit them because extremely low sequence saturation and cell number made the results unreliable.

Our estimates of chromosome loss were heavily dependent on the computational definitions we applied. To estimate partial or whole chromosome loss, we required 70% of genes, with a minimum of 150 genes, to either side of the breakpoint to have a gene dosage <0.95. It should be noted that changing the required percentage or minimum number of genes would influence estimations of chromosome loss. Based on these definitions, it is no surprise that our ability to estimate chromosome loss was dependent on the number of genes reported by inferCNV (at least 150 genes per chromosome were required to call chromosome loss) (see tables below). Analyzing several parameters that could influence the number of genes reported by inferCNV, we found that the median number of genes per cell obtained from scRNA-seq was the best indicator (see tables and figure below).

One scenario with a limited number of genes was shorter chromosomes (mostly chromosomes 18-22). Accordingly, we observed more background noise in these chromosomes (see Figure 3.3C; Figure 3.5A). Other scenarios with a limited number of genes were the CAR T cell and clinical trial scRNA-seq datasets. In general, absolute quantification of chromosome loss is more sensitive to the computational definitions (mentioned earlier) and number of genes, compared to chromosome loss enrichment which only considers the ratio of chromosome loss between targeted and non-targeted cells. Therefore, we only estimated absolute percentages of chromosome loss in experiments with a large number of genes reported by inferCNV (*TRAC*-targeting experiment, Figure 3.1E; and CROP-seq screen, Figure 3.5A). For chromosome loss enrichment, we reported estimations for all datasets (*TRAC*-targeting experiment, Figure 3.2F; CROP-seq screen, Figure 3.3B; CAR T cell experiment, Figure 3.10C; clinical trial, Figure 3.11B).

TRAC-editing experiment (related to Figure 3.1)

Sample	Number of cells	Mean reads per cell	Median genes per cell	Saturation %	inferCNV genes detected
Pool 1	19,536	22,563	1,897	26.2	7,172
Pool 2	22,010	23,010	1,390	29.7	7,172
Pool 3	30,523	23,483	1,302	31.9	7,172

CROP-seq screen experiment (related to Figure 3.3 and Figure 3.6)

Sample	Number of cells	Mean reads per cell	Median genes per cell	Saturation %	inferCNV genes detected
	178,222	45,202	4,194	40.8	9,638

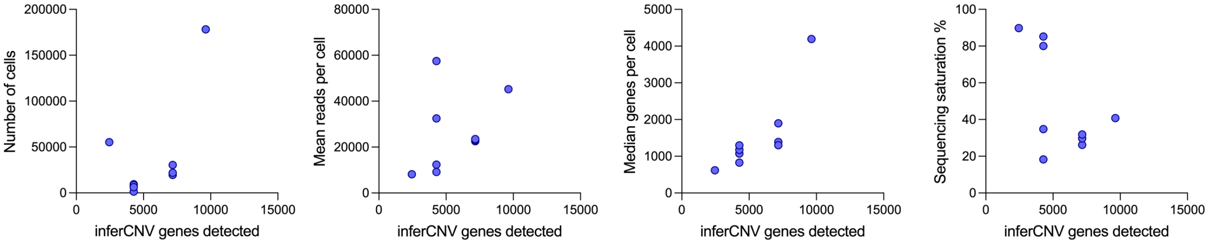
CAR T cell experiment (related to Figure 3.10)

Sample	Number of cells	Mean reads per cell	Median genes per cell	Saturation %	inferCNV genes detected
	55,213	8,146	617	89.8	2,456

Clinical trial experiment (related to Figure 3.11)

Sample	Number of cells	Mean reads per cell	Median genes per cell	Saturation %	inferCNV genes detected
UPN35 Day0	1,475	9,168	829	18.3	4,277
UPN35 Day28	9,429	32,510	1,073	80.0	4,277
UPN39 Day0	8,583	12,359	1,178	34.8	4,277
UPN39 Day10	6,200	57,502	1,297	85.2	4,277

UPN39 Day118	7,646	43,708	1,113	83.5	4,277



Spearman r = 0.4433
P = 0.2349

Spearman r = 0.4699
P = 0.2040

Spearman r = 0.9399
****P* = 0.0008

Spearman r = -0.4522
P = 0.2238

Supplementary Note 3.3: Cas9 HDR templates to insert an HA tag in-frame with CD5.

Left *CD5* homology arm
HA tag
Right *CD5* homology arm
Partial Cas9 target sequence

CD5-HA single-strand HDR template (ssHDRT)

ACCCTCCTCTCTTCTTTCTGCAGTCGCTTCCTGCCTCGGATACCCATACGATGTGCCTGATTACGCAGGATCACGGCTCAGCTGGTATGACCCAGGTAAGGAAGAGCCACATG

CD5-HA double-strand HDR template, top (dsHDRT)

ACCCTCCTCTCTTCTTTCTGCAGTCGCTTCCTGCCTCGGATACCCATACGATGTGCCTGATTACGCAGGATCACGGCTCAGCTGGTATGACCCAGGTAAGGAAGAGCCACATG

CD5-HA double-strand HDR template, bottom (dsHDRT)

CATGTGGCTCTTCCTTACCTGGGTCATACCAGCTGAGCCGTGATCCTGCGTAATCAGGCACATCGTATGGGTATCCGAGGCAGGAAGCGACTGCAGAAAGAAGAGAGGAGGGT

CD5-HA single-strand Cas9 target sequence HDR template (ssCTS)

AACAAGCAGCGCTTCCTGCCTCGGACGGACCCTCCTCTCTTCTTTCTGCAGTCGCTTCCTGCCTCGGATACCCATACGATGTGCCTGATTACGCAGGATCACGGCTCAGCTGGTATGACCCAGGTAAGGAAGAGCCACATGCCGTCCGAGGCAGGAAGCGCTGCTTGTT

CD5-HA single-strand Cas9 target sequence HDR, 5’ primer (ssCTS)

AGAGGAGGGTCCGTCCGAGGCAGGAAGCGCTGCTTGTT

CD5-HA single-strand Cas9 target sequence HDR, 3’ primer (ssCTS)

AACAAGCAGCGCTTCCTGCCTCGGACGGCATGTGGCTC

Supplementary Note 3.4: AAV transgene used as a Cas9 homology-directed repair template. The transgene encodes a 1928ζ chimeric antigen receptor and truncated EGFRt tag separated by 2A sequences and flanked by homology sequences to TRAC.

AAV ITR
TRAC homology arms
qPCR primers
P2A
EGFRt tag
1928z CAR

TTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTG
CCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGA
TAATCCTGATATGAATAAATTGCAGTTTCACTTGATGCTCGATGAGTTTTTCTAATGA
GGACCTAAATGTAATCACCTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTGCTGG
CGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGATGCTCAAGT
CAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAA
GCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGC
CTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCA
GTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCA
GCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGA
CACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGT
ATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAG
AAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTCGGAAAAAGAGT
TGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTT
GCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATTTTC
TACCGAAGAAAGGCCCACCCGTGTAAAACGACGGCCAGTTTATCCTGCAGGCAGC
TGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCG
ACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCC
AACTCCATCACTAGGGGTTCCTGCGGCCTCTAGCTCGAGTCATTTCTTTCTCAGAA
GAGCCTGGCTAGGAAGGTGGATGAGGCACCATATTCATTTTGCAGGTGAAATTCCT
GAGATGTAAGGAGCTGCTGTGACTTGCTCAAGGCCTTATATCGAGTAAACGGTAGT
GCTGGGGCTTAGACGCAGGTGTTCTGATTTATAGTTCAAAACCTCTATCAATGAGA
GAGCAATCTCCTGGTAATGTGATAGATTTCCCAACTTAATGCCAACATACCATAAAC
CTCCCATTCTGCTAATGCCCAGCCTAAGTTGGGGtGACCACTCCAGATTCCAAGAT
GTACAGTTTGCTTTGCTGGGCCTTTTTCCCATGCCTGCCTTTACTCTGCCAGAGTTA
TATTGCTGGGGTTTTGAAGAAGATCCTATTAAATAAAAGAATAAGCAGTATTATTAA
GTAGCCCTGCATTTCAGGTTTCCTTGAGTGGCAGGCCAGGCCTGGCCGTGAACGT
TCACTGAAATCATGGCCTCTTGGCCAAGATTGATAGCTTGTGCCTGTCCCTGAGTC

CCAGTCCATCACGAGCAGCTGGTTTCTAAGATGCTATTTCCCGTATAAAGCATGtGA
CCGTGACTTGCCAGCCCCACAGAGCCCCGCCCTTGTCCATCACTGGCATCTGGAC
TCCAGCCTGGGTTGGGGCAAAGAGGGAAATGAGATCATGTCCTAACCCTGgaattgg
ATCCTCTTGTCttACAGATGGATCTGGAGCAACAAACTTCTCACTACTCAAACAAGCA
GGTGACGTGGAGGAGAATCCCGGCCCcatggctctcccagtgactgccctactgcttcccctagcgcttc
tcctgcatgcagaggtgaagctgcagcagtctggggctgagctggtgaggcctgggtcctcagtgaagatttcctgcaag
gcttctggctatgcattcagtagctactggatgaactgggtgaagcagaggcctggacagggtcttgagtggattggacag
atttatcctggagatggtgatactaactacaatggaaagttcaagggtcaagccacactgactgcagacaaatcctccagc
acagcctacatgcagctcagcggcctaacatctgaggactctgcggtctatttctgtgcaagaaagaccattagttcggtag
tagatttctactttgactactggggccaagggaccacggtcaccgtctcctcaggtggaggtggatcaggtggaggtggat
ctggtggaggtggatctgacattgagctcacccagtctccaaaattcatgtccacatcagtaggagacagggtcagcgtca
cctgcaaggccagtcagaatgtgggtactaatgtagcctggtatcaacagaaaccaggacaatctcctaaaccactgattt
actcggcaacctaccggaacagtggagtccctgatcgcttcacaggcagtggatctgggacagatttcactctcaccatca
ctaacgtgcagtctaaagacttggcagactatttctgtcaacaatataacaggtatccgtacacgtccggaggggggacca
agctggagatcaaacgggcggccgcaattgaagttatgtatcctccttacctagacaatgagaagagcaatggaacc
attatccatgtgaaagggaaacacctttgtccaagtcccctatttcccggaccttctaagcccttttgggtgctggtggtggttg
gtggagtcctggcttgctatagcttgctagtaacagtggcctttattattttctgggtgaggagtaagaggagcaggctcctgc
acagtgactacatgaacatgactccccgccgccccgggcccacccgcaagcattaccagccctatgccccaccacgcg
acttcgcagcctatcgctccagagtgaagttcagcaggagcgcagacgcccccgcgtaccagcagggccagaaccag
ctctataacgagctcaatctaggacgaagagaggagtacgatgttttggacaagagacgtggccgggaccctgagatgg
ggggaaagccgagaaggaagaaccctcaggaaggcctgtacaatgaactgcagaaagataagatggcggaggcct
acagtgagattgggatgaaaggcgagcgccggaggggcaaggggcacgatggcctttaccagggtctAagtacagcc
accaaggacacctacgacgcccttcacatgcaggccctgccccctcgcGGAAGCGGAGCTACTAACTTC
AGCCTGCTGAAGCAGGCTGGAGACGTGGAGGAGAACCCTGGACCCatgcttctcctggtg
acaagccttctgctctgtgagttaccacacccagcattcctcctgatcccacgcaaagtgtgtaacggaataggtattggtg
aatttaaagactcactctccataaatgctacgaatattaaacacttcaaaaactgcacctccatcagtggcgatctccacatc
ctgccggtggcatttaggggtgactccttcacacatactcctcctctggacccacaggaactggatattctgaaaaccgtaa
aggaaatcacagggtttttgctgattcaggcttggcctgaaaacaggacggacctccatgcctttgagaacctagaaatca
tacgcggcaggaccaagcaacatggtcagttttctcttgcagtcgtcagcctgaacataacatccttgggattacgctccctc
aaggagataagtgatggagatgtgataatttcaggaaacaaaaatttgtgctatgcaaatacaataaactggaaaaaact
gtttgggacctccggtcagaaaaccaaaattataagcaacagaggtgaaaacagctgcaaggccacaggccaggtct
gccatgccttgtgctcccccgagggctgctggggcccggagcccagggactgcgtctcttgccggaatgtcagccgagg
cagggaatgcgtggacaagtgcaaccttctggagggtgagccaagggagtttgtggagaactctgagtgcatacagtgc
cacccagagtgcctgcctcaggccatgaacatcacctgcacaggacggggaccagacaactgtatccagtgtgcccac
tacattgacggcccccactgcgtcaagacctgcccggcaggagtcatgggagaaaacaacaccctggtctggaagtac
gcagacgccggccatgtgtgccacctgtgccatccaaactgcacctacggatgcactgggccaggtcttgaaggctgtcc
cacgaatgggcctaagatcccgtccatcgccactgggatggtgggggccctcctcttgctgctggtggtggccctggggat
cggcctcttcatgGGCAGCGGCGCGACCAACTTTAGCCTGCTGAAACAGGCGGGCGATG
TTGAAGAAAACCCGGGCCCGaaTATCCAGAACCCTGACCCTGCCGTGTACCAGCT
GAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCA
AACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCT
AGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAA
TCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGACACCTT
CTTCCCCAGCCCAGGTAAGGGCAGCTTTGGTGCCTTCGCAGGCTGTTTCCTTGCT
TCAGGAATGGCCAGGTTCTGCCCAGagctCTGGTCAATGATGTCTAAAACTCCTCTG
ATTGGTGGaCTCGGCCTTATCCATTGCCACCAAAACCCTCTTTTTACTAAGAAACAG

TGAGCCTTGTTCTGGCAGTCCAGAGAATGACACGGAAAAAAGCAGATGAAGAGA
AGGTGGCAGGAGAGGGCACGTGGCCCAGCCTCAGTCTCTCCAACTGAGTTCCTG
CCTGCCTGCCTTTGCTCAGACTGTTTGCCCCTTACTGCTCTTCTAGGCCTCATTCTA
AGCCCCTTCTCCAAGTTGCCTCTCCTTATTTCTCCCTGTCTGCCAAAAAATCTTTCC
CAGCTCACTAAGTCAGTCTCACGCAGTCACTCATTAACCCACCAATCACTGATTGT
GCCGGCACATGAATGCACCAGGTGTTGAAGTGGAGGAATTGCGGCCGCAGGAAC
CCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGC
CGGGCGACCAAAGGTCGCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGA
GCGAGCGAGCGCGCAGCTGCCtgcaggATAGCAGAAAGTCAAAAGCCTCCGCGTAA
GAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGT
TATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGAGCCATATTCAACGGGAAACG
TCTTGCTTGAAGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAA
TGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGA
AGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGA
TGTTACAGATGAGATGGTCAGGCTAAACTGGCTGACGGAATTTATGCCTCTTCCGA
CCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATC
CCAGGGAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATAT
TGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATT
GTCCTTTTAACGGCGATCGCGTATTTCGaCTCGCTCAGGCGCAATCACGAATGAAT
AACGGTTTGGTTGGTGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTG
AACAAGTCTGGAAAGAAATGCATAAACTCTTGCCATTCTCACCGGATTCAGTCGTC
ACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGT
TGTATTGATG

gRNA	Target	Spacer
1	*TRAC*	AACAAATGTGTCACAAAGTA
2	*TRAC*	TGTGCTAGACATGAGGTCTA
3	*TRAC*	CTTCAAGAGCAACAGTGCTG
4	*TRAC*	GACACCTTCTTCCCCAGCCC
5	*TRAC*	TTCTTCCCCAGCCCAGGTAA
6	*TRAC*	CTTCTTCCCCAGCCCAGGTA
7	*TRAC*	GCTGGTACACGGCAGGGTCA
8	*TRAC*	GAGAATCAAAATCGGTGAAT
9	*TRAC*	TCTCTCAGCTGGTACACGGC
10	*TRAC*	CTTACCTGGGCTGGGGAAGA
11	*TRAC*	CTCTCAGCTGGTACACGGCA
NT	N/A	ACGGAGGCTAAGCGTCGCAA

Table 3.1: *TRAC*-targeting gRNAs

gRNA	Target	PCR amplicon	Fwd primer	Rev primer
1	*TRAC*	TGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCAT	GCTCTTCCGATCTTGCCTATTCACCGATTTT	GCTCTTCCGATCTATGCAAAGTCAGATTTGTTG

2	*TRAC*	TGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCAT	GCTCTTCCGATCTTGCCTATTCACCGATTTT	GCTCTTCCGATCTATGCAAAGTCAGATTTGTTG
3	*TRAC*	TGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCAT	GCTCTTCCGATCTTGCCTATTCACCGATTTT	GCTCTTCCGATCTATGCAAAGTCAGATTTGTTG
4	*TRAC*	TCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGGTAAGGGCAGCTTTGGTGCCTTCGCAGGCTGTTTCC	GCTCTTCCGATCTCTATGGACTTCAAGAGCAAC	GCTCTTCCGATCTGGAAACAGCCTGCGAA
5	*TRAC*	TCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGGTAAGGGCAGCTTTGGTGCCTTCGCAGGCTGTTTCC	GCTCTTCCGATCTCTATGGACTTCAAGAGCAAC	GCTCTTCCGATCTGGAAACAGCCTGCGAA
6	*TRAC*	TCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGGTAAGGGCAGCTTTGGTGCCTTCGCAGGCTGTTTCC	GCTCTTCCGATCTCTATGGACTTCAAGAGCAAC	GCTCTTCCGATCTGGAAACAGCCTGCGAA
7	*TRAC*	ATGTCCTAACCCTGATCCTCTTGTCCCACAGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGA	GCTCTTCCGATCTATGTCCTAACCCTGATCCT	GCTCTTCCGATCTCAGAATCCTTACTTTGTGACAC
8	*TRAC*	ATGTCCTAACCCTGATCCTCTTGTCCCACAGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGA	GCTCTTCCGATCTATGTCCTAACCCTGATCCT	GCTCTTCCGATCTCAGAATCCTTACTTTGTGACAC
9	*TRAC*	ATGTCCTAACCCTGATCCTCTTGTCCCACAGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGA	GCTCTTCCGATCTATGTCCTAACCCTGATCCT	GCTCTTCCGATCTCAGAATCCTTACTTTGTGACAC
10	*TRAC*	TCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGGTAAGGGCAGCTTTGGTGCCTTCGCAGGCTGTTTCC	GCTCTTCCGATCTCTATGGACTTCAAGAGCAAC	GCTCTTCCGATCTGGAAACAGCCTGCGAA
11	*TRAC*	ATGTCCTAACCCTGATCCTCTTGTCCCACAGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGA	GCTCTTCCGATCTATGTCCTAACCCTGATCCT	GCTCTTCCGATCTCAGAATCCTTACTTTGTGACAC
NT	N/A	TGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCAT	GCTCTTCCGATCTTGCCTATTCACCGATTTT	GCTCTTCCGATCTATGCAAAGTCAGATTTGTTG
NT	N/A	TCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGGTAAGGGCAGCTTTGGTGCCTTCGCAGGCTGTTTCC	GCTCTTCCGATCTCTATGGACTTCAAGAGCAAC	GCTCTTCCGATCTGGAAACAGCCTGCGAA
NT	N/A	ATGTCCTAACCCTGATCCTCTTGTCCCACAGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGA	GCTCTTCCGATCTATGTCCTAACCCTGATCCT	GCTCTTCCGATCTCAGAATCCTTACTTTGTGACAC

Table 3.2: *TRAC* NGS primers

gRNA	Target	Anchor (5'-LMO)	Barcode	Co-anchor (3'LMO)	MULTI-seq i5 primer	MULTI-seq i7 primer
1	*TRAC*	TGGAATTCTCGGGTGCCAAGGgtaacgatccagctgtcact	CCTTGGCACCCGAGAATTCCAGGAGAAGAAAAAAAAA	AGTGACAGCTGGATCGTTAC	AATGATACGGCGACCACCGAGATCTACACtctttccCTACACGACGCTCTTCCGATCT	CAAGCAGAAGACGGCATACGAGATcgtgatGTGACTGGAGTTCCTTG

			AAAAAAAAAAAAAAAAAAAAAA			GCACCCGAGAATTCCA
2	*TRAC*	TGGAATTCTCGGGTGCCAAGGgtaacgatccagctgtcact	CCTTGGCACCCGAGAATTCCATCACAGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	AGTGACAGCTGGATCGTTAC	AATGATACGGCGACCACCGAGATCTACACtcgatcgCTACACGACGCTCTTCCGATCT	CAAGCAGAAGACGGCATACGAGATacatcgGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
3	*TRAC*	TGGAATTCTCGGGTGCCAAGGgtaacgatccagctgtcact	CCTTGGCACCCGAGAATTCCACCACAATGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	AGTGACAGCTGGATCGTTAC	AATGATACGGCGACCACCGAGATCTACACtctttccCTACACGACGCTCTTCCGATCT	CAAGCAGAAGACGGCATACGAGATcgtgatGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
4	*TRAC*	TGGAATTCTCGGGTGCCAAGGgtaacgatccagctgtcact	CCTTGGCACCCGAGAATTCCATGAGACCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	AGTGACAGCTGGATCGTTAC	AATGATACGGCGACCACCGAGATCTACACtctttccCTACACGACGCTCTTCCGATCT	CAAGCAGAAGACGGCATACGAGATcgtgatGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
5	*TRAC*	TGGAATTCTCGGGTGCCAAGGgtaacgatccagctgtcact	CCTTGGCACCCGAGAATTCCAGCACACGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	AGTGACAGCTGGATCGTTAC	AATGATACGGCGACCACCGAGATCTACACtctttccCTACACGACGCTCTTCCGATCT	CAAGCAGAAGACGGCATACGAGATcgtgatGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
6	*TRAC*	TGGAATTCTCGGGTGCCAAGGgtaacgatccagctgtcact	CCTTGGCACCCGAGAATTCCAAGAGAGAGAAAAAAAA	AGTGACAGCTGGATCGTTAC	AATGATACGGCGACCACCGAGATCTACACtcgatcgCTACACGACGCTCTTCCGATCT	CAAGCAGAAGACGGCATACGAGATacatcgGTGACTGGAGTTCCTTG

			AAAAAAAAAAAAAAAAAAAAAA			GCACCCGAGAATTCCA
7	*TRAC*	TGGAATTCTCGGGTGCCAAGGgtaacgatccagctgtcact	CCTTGGCACCCGAGAATTCCAGAAAAGGGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	AGTGACAGCTGGATCGTTAC	AATGATACGGCGACCACCGAGATCTACACtcgatcgCTACACGACGCTCTTCCGATCT	CAAGCAGAAGACGGCATACGAGATacatcgGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
8	*TRAC*	TGGAATTCTCGGGTGCCAAGGgtaacgatccagctgtcact	CCTTGGCACCCGAGAATTCCACGAGATTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	AGTGACAGCTGGATCGTTAC	AATGATACGGCGACCACCGAGATCTACACtcgatcgCTACACGACGCTCTTCCGATCT	CAAGCAGAAGACGGCATACGAGATacatcgGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
9	*TRAC*	TGGAATTCTCGGGTGCCAAGGgtaacgatccagctgtcact	CCTTGGCACCCGAGAATTCCAGTAGCACTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	AGTGACAGCTGGATCGTTAC	AATGATACGGCGACCACCGAGATCTACACagattgcCTACACGACGCTCTTCCGATCT	CAAGCAGAAGACGGCATACGAGATgcctaaGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
10	*TRAC*	TGGAATTCTCGGGTGCCAAGGgtaacgatccagctgtcact	CCTTGGCACCCGAGAATTCCACGACCAGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	AGTGACAGCTGGATCGTTAC	AATGATACGGCGACCACCGAGATCTACACagattgcCTACACGACGCTCTTCCGATCT	CAAGCAGAAGACGGCATACGAGATgcctaaGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
11	*TRAC*	TGGAATTCTCGGGTGCCAAGGgtaacgatccagctgtcact	CCTTGGCACCCGAGAATTCCATTAGCCAGAAAAAAAA	AGTGACAGCTGGATCGTTAC	AATGATACGGCGACCACCGAGATCTACACagattgcCTACACGACGCTCTTCCGATCT	CAAGCAGAAGACGGCATACGAGATgcctaaGTGACTGGAGTTCCTTG

			AAAAAAAAAAAAAAAAAAAAAA			GCACCCGAGAATTCCA
NT	N/A	TGGAATTCTCGGGTGCCAAGGgtaacgatccagctgtcact	CCTTGGCACCCGAGAATTCCAGGACCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	AGTGACAGCTGGATCGTTAC	AATGATACGGCGACCACCGAGATCTACACagattgcCTACACGACGCTCTTCCGATCT	CAAGCAGAAGACGGCATACGAGATgcctaaGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA

Table 3.3: scRNA-seq MULTI-seq

gRNA	Target	ddPCR amplicon 1 (control) fwd primer	ddPCR amplicon 1 (control) rev primer	ddPCR amplicon 1 (control) probe (5'-HEX)
1	*TRAC*	CGTGTACCAGCTGAGAG	TGCAAAGTCAGATTTGTTGC	TCCATAGACCTCATGTCTAGCACAG
2	*TRAC*	CGTGTACCAGCTGAGAG	TGCAAAGTCAGATTTGTTGC	TCCATAGACCTCATGTCTAGCACAG
3	*TRAC*	CGTGTACCAGCTGAGAG	TGCAAAGTCAGATTTGTTGC	TCCATAGACCTCATGTCTAGCACAG
4	*TRAC*	CGTGTACCAGCTGAGAG	TGCAAAGTCAGATTTGTTGC	TCCATAGACCTCATGTCTAGCACAG
5	*TRAC*	CGTGTACCAGCTGAGAG	TGCAAAGTCAGATTTGTTGC	TCCATAGACCTCATGTCTAGCACAG
6	*TRAC*	CGTGTACCAGCTGAGAG	TGCAAAGTCAGATTTGTTGC	TCCATAGACCTCATGTCTAGCACAG
7	*TRAC*	CGTGTACCAGCTGAGAG	TGCAAAGTCAGATTTGTTGC	TCCATAGACCTCATGTCTAGCACAG
8	*TRAC*	CGTGTACCAGCTGAGAG	TGCAAAGTCAGATTTGTTGC	TCCATAGACCTCATGTCTAGCACAG
9	*TRAC*	CGTGTACCAGCTGAGAG	TGCAAAGTCAGATTTGTTGC	TCCATAGACCTCATGTCTAGCACAG
10	*TRAC*	CGTGTACCAGCTGAGAG	TGCAAAGTCAGATTTGTTGC	TCCATAGACCTCATGTCTAGCACAG

11	*TRAC*	CGTGTACCAGCTGAGAG	TGCAAAGTCAGATTTGTTGC	TCCATAGACCTCATGTCTAGCACAG
NT	N/A	CGTGTACCAGCTGAGAG	TGCAAAGTCAGATTTGTTGC	TCCATAGACCTCATGTCTAGCACAG
NT	N/A	CGTGTACCAGCTGAGAG	TGCAAAGTCAGATTTGTTGC	TCCATAGACCTCATGTCTAGCACAG
NT	N/A	CGTGTACCAGCTGAGAG	TGCAAAGTCAGATTTGTTGC	TCCATAGACCTCATGTCTAGCACAG
NT	N/A	CGTGTACCAGCTGAGAG	TGCAAAGTCAGATTTGTTGC	TCCATAGACCTCATGTCTAGCACAG

gRNA	**Target**	**ddPCR amplicon 2 (target) fwd primer**	**ddPCR amplicon 2 (target) rev primer**	**ddPCR amplicon 2 (target) probe (5'FAM)**
1	*TRAC*	ACCTCCCATTCTGCTAATG	CTCAAGGAAACCTGAAATGC	CCAGCAATATAACTCTGGCAGAGTAA
2	*TRAC*	TGATTCTCAAACAAATGTGTCAC	AAAGCTGCCCTTACCTG	CGCCTTCAACAACAGCATTATTCCA
3	*TRAC*	TGATTCTCAAACAAATGTGTCAC	AAAGCTGCCCTTACCTG	CGCCTTCAACAACAGCATTATTCCA
4	*TRAC*	ACTTTGCATGTGCAAACG	TTGGTGGCAATGGATAAGG	TCTGCCCAGAGCTCTGGTCAAT
5	*TRAC*	ACTTTGCATGTGCAAACG	TTGGTGGCAATGGATAAGG	TCTGCCCAGAGCTCTGGTCAAT
6	*TRAC*	ACTTTGCATGTGCAAACG	TTGGTGGCAATGGATAAGG	TCTGCCCAGAGCTCTGGTCAAT
7	*TRAC*	GCAAAGAGGGAAATGAGATC	TTGTCTGTGATATACACATCAG	CTGCCTATTCACCGATTTTGATTCTCA
8	*TRAC*	CGTGTACCAGCTGAGAG	TGCAAAGTCAGATTTGTTGC	TCCATAGACCTCATGTCTAGCACAG
9	*TRAC*	GCAAAGAGGGAAATGAGATC	TTGTCTGTGATATACACATCAG	CTGCCTATTCACCGATTTTGATTCTCA
10	*TRAC*	ACTTTGCATGTGCAAACG	TTGGTGGCAATGGATAAGG	TCTGCCCAGAGCTCTGGTCAAT
11	*TRAC*	GCAAAGAGGGAAATGAGATC	TTGTCTGTGATATACACATCAG	CTGCCTATTCACCGATTTTGATTCTCA
NT	N/A	ACCTCCCATTCTGCTAATG	CTCAAGGAAACCTGAAATGC	CCAGCAATATAACTCTGGCAGAGTAA

NT	N/A	TGATTCTCAAACAAATGTGTCAC	AAAGCTGCCCTTACCTG	CGCCTTCAACAACAGCATTATTCCA
NT	N/A	ACTTTGCATGTGCAAACG	TTGGTGGCAATGGATAAGG	TCTGCCCAGAGCTCTGGTCAAT
NT	N/A	GCAAAGAGGGAAATGAGATC	TTGTCTGTGATATACACATCAG	CTGCCTATTCACCGATTTTGATTCTCA

Table 3.4: *TRAC* ddPCR

Pool	Anchor (5'-LMO)	Barcode	Co-anchor (3'LMO)
1	TGGAATTCTCGGGTGCCAAGGgtaacgatccagctgtcact	CCTTGGCACCCGAGAATTCCAGGAGAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	AGTGACAGCTGGATCGTTAC
2	TGGAATTCTCGGGTGCCAAGGgtaacgatccagctgtcact	CCTTGGCACCCGAGAATTCCACCACAATGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	AGTGACAGCTGGATCGTTAC
3	TGGAATTCTCGGGTGCCAAGGgtaacgatccagctgtcact	CCTTGGCACCCGAGAATTCCATGAGACCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	AGTGACAGCTGGATCGTTAC
4	TGGAATTCTCGGGTGCCAAGGgtaacgatccagctgtcact	CCTTGGCACCCGAGAATTCCAGCACACGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	AGTGACAGCTGGATCGTTAC
5	TGGAATTCTCGGGTGCCAAGGgtaacgatccagctgtcact	CCTTGGCACCCGAGAATTCCAAGAGAGAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	AGTGACAGCTGGATCGTTAC
6	TGGAATTCTCGGGTGCCAAGGgtaacgatccagctgtcact	CCTTGGCACCCGAGAATTCCATCACAGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	AGTGACAGCTGGATCGTTAC
Pool	**MULTI-seq i5 primer**	**MULTI-seq i7 primer**	**CROP-seq PCR1 Fwd**
1	AATGATACGGCGACCACCGAGATCTACACAGTGTTACCTACACTCTTTCCCTACACGACGCTCTTCCGATCT	CAAGCAGAAGACGGCATACGAGATcgtgatGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGcTTGTGGAAAGGACGAAACAC
2	AATGATACGGCGACCACCGAGATCTACACGCCAACCCTGACACTCTTTCCCTACACGACGCTCTTCCGATCT	CAAGCAGAAGACGGCATACGAGATacatcgGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGcTTGTGGAAAGGACGAAACAC
3	AATGATACGGCGACCACCGAGATCTACACTTAGACTGATACACTCTTTCCCTACACGACGCTCTTCCGATCT	CAAGCAGAAGACGGCATACGAGATgcctaaGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGcTTGTGGAAAGGACGAAACAC
4	AATGATACGGCGACCACCGAGATCTACACTATCTTCATCACACTCTTTCCCTACACGACGCTCTTCCGATCT	CAAGCAGAAGACGGCATACGAGATtggtcaGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGcTTGTGGAAAGGACGAAACAC
5	AATGATACGGCGACCACCGAGATCTACACGAGCATCTATACACTCTTTCCCTACACGACGCTCTTCCGATCT	CAAGCAGAAGACGGCATACGAGATcactgtGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGcTTGTGGAAAGGACGAAACAC
6	AATGATACGGCGACCACCGAGATCTACACCCCTAACTTCACACTCTTTCCCTACACGACGCTCTTCCGATCT	CAAGCAGAAGACGGCATACGAGATattggcGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGcTTGTGGAAAGGACGAAACAC
Pool	**CROP-seq PCR1 Rev**	**CROP-seq PCR 2 i5 Primer**	**CROP-seq PCR2 i7 Primer**
1	ACACTCTTTCCCTACACGACG	AATGATACGGCGACCACCGAGATCTACACagcgctagACACTCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATtcctctacGTCTCGTGGGCTCGG
2	ACACTCTTTCCCTACACGACG	AATGATACGGCGACCACCGAGATCTACACgatatcgaACACTCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATtcatgagcGTCTCGTGGGCTCGG
3	ACACTCTTTCCCTACACGACG	AATGATACGGCGACCACCGAGATCTACACcgcagacgACACTCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATcctgagatGTCTCGTGGGCTCGG
4	ACACTCTTTCCCTACACGACG	AATGATACGGCGACCACCGAGATCTACACtatgagtaACACTCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATtagcgagtGTCTCGTGGGCTCGG
5	ACACTCTTTCCCTACACGACG	AATGATACGGCGACCACCGAGATCTACACaggtgcgtACACTCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATgtagctccGTCTCGTGGGCTCGG

6	ACACTCTTTCCCTACACGACG	AATGATACGGCGACCACCGAGATCTACACgaacatacACACTCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATtactacgcGTCTCGTGGGCTCGG

Table 3.5: CROP-seq library prep

Target	Spacer	ddPCR amplicon 1 (control) fwd primer	ddPCR amplicon 1 (control) rev primer	ddPCR amplicon 1 (control) probe (5'-HEX)
CECR1	GCTCCGAATCAAGTTCCCCA	acccacacctaggtgagc	gcagacagccagttctaac	ctgtggtccaaactccctgaatgcc
HBB	TCTGCCGTTACTGCCCTGTG	TACCATCAGTACAAATTGCTAC	GCAGGAAGAGATCCATCTAC	CTTCAATATGCTTACCAAGCTGTGATTCC
USP22	GCCATTGATCTGATGTACGG	atgcgtggtatttaaaactgg	ccgatgagaaactattttccc	ccacctgacacggagtgactatgc
SERPINA1	ACTCACGATGAAATCCTGGA	gaaaactgaagaatccacgc	tttggtcaaggagcttgac	ccagcctgagctgttcccatag
TGFBR1	ATGGGCAAGACCGCTCGCCG	tccaggaacagccactg	tgtacacatcctttagttcctc	agataattggtatcagagctggcactc
ERCC8	GATGTTGAAAGAATCCACGG	cactttttatatttattttaccattagac	ccaaatgttgtcaaataaacatttc	agcacctaaaagccctctgcatc
DNMT3B	CCCAACAACACGCAACCAGG	aggagatgagatagtgctg	gtatgtgacagaggaaggag	ctagggaatgtcagtggtcacctccc
CD70	GAGCTGCAGCTGAATCACAC	ctctctacgctgcaaag	taggcacacatggaaatgg	ccatcccgtcctaggaggccc
FOXO1	GGTGGCGCAAACGAGTAGCA	aggcacttgtacaggtgtc	agccaggcatctcataac	ctggacatgctcagcagacatct
PCNT	CAGACTTTGAGGAACAACTG	aaatcccgaaacgatgacc	tgcaggaaggcatcttg	agttaaacatgcatcaggaagcaatgacc
CXCR4	GCTTCTACCCCAATGACTTG	ctggagtgaaaacttgaagac	gaggccctagctttcttc	ctggaccctctgctcacagagg
IL2RG	GGGCAGCTGCAGGAATAAGA	ggcagagtggaaacactg	tcctggcctctagtgatc	caccgtgcttggcctcctcc
TIGIT	CAGGCACAATAGAAACAACG	caatacaggcagtgaatgtg	tctccacacatctctactcc	tgccctatgcttccagagctga
PDCD1	GGCGCCCTGGCCAGTCGTCT	agtgaggaccaaggatgc	tcctgggttcctctctg	cctgcctgcccaggagcaa

B2M	CGCGAGCACAGCTAAGGCCA	CCAAGCTGTAGTTATAAACAG	TTCTAGGACTTCAGGCTG	CACCATTAGCAAGTCACTTAGCATCTC

Target	**Spacer**	**ddPCR amplicon 2 (target) fwd primer**	**ddPCR amplicon 2 (target) rev primer**	**ddPCR amplicon 2 (target) probe (5'FAM)**
CECR1	GCTCCGAATCAAGTTCCCCA	cttaacaggcagcccttc	ttgtgctctgtcggtgc	tggcttctctgtctgtctcacaga
HBB	TCTGCCGTTACTGCCCTGTG	attggtctccttaaacctgtc	CAGAGCCATCTATTGCTTac	ctgtgttcactagcaacctcaaacag
USP22	GCCATTGATCTGATGTACGG	aaaccttgcattttccaagc	acctcactgtgtttaactctg	cctccttggcgattatttccatgtct
SERPINA1	ACTCACGATGAAATCCTGGA	tgtctggctggttgagg	agtccaacagcaccaatatc	agccttcatggatctgagcctc
TGFBR1	ATGGGCAAGACCGCTCGCCG	ggtgtcagattatcatgagc	ccttctattttcatagacattattc	agaactgcttatagaattaccttgggtacc
ERCC8	GATGTTGAAAGAATCCACGG	gtataataaacttactatctcccttc	gatattttaaacatttatacaagatgtc	tgtctaaatcttctaccacacatcctaacg
DNMT3B	CCCAACAACACGCAACCAGG	gccctggagactcattg	cagctgctagaaagctg	ctgactctcccaagaagtggtccc
CD70	GAGCTGCAGCTGAATCACAC	ctctccccacttgtctttc	ggcacaggggacacatag	ttctctctctgtgcctcttcttctc
FOXO1	GGTGGCGCAAACGAGTAGCA	gtgtttgtataggcatctgg	cttttggataatctcaaccttc	ccaacatcattaactgtttcgacccagtc
PCNT	CAGACTTTGAGGAACAACTG	ccttgattctttggaatcctg	ccgcacagactgtagc	ccgcttcagtttctccaactcttga
CXCR4	GCTTCTACCCCAATGACTTG	gagatgataatgcaatagcag	tgttggcgtctggatcc	tggccttatcctgcctggtattgtc
IL2RG	GGGCAGCTGCAGGAATAAGA	ccacagccacccttctc	acagacagactacacccag	cctccagtcccagatttcccacca
TIGIT	CAGGCACAATAGAAACAACG	gcctcaaaggcccttag	ccagttgacctgggtcac	aggccagctgctgacccag
PDCD1	GGCGCCCTGGCCAGTCGTCT	gactgagggtggaaggtc	cctgagcagtggagaag	cctggctctgggacacctgacc
B2M	CGCGAGCACAGCTAAGGCCA	CCTTGTCCTGATTGGCTG	ggagaggaaggaccagag	CAGCGCGACGCCTCCACT

Table 3.6: Extended target ddPCR

Target	Spacer	Fwd primer	Rev primer
CECR1	GCTCCGAATCAAGTTCCCCA	GCTCTTCCGATCTcagcccttctgttcacag	GCTCTTCCGATCTacagatccaaagatgtggc
HBB	TCTGCCGTTACTGCCCTGTG	GCTCTTCCGATCTttgataccaacctgcccag	GCTCTTCCGATCTgttcactagcaacctcaaac
USP22	GCCATTGATCTGATGTACGG	GCTCTTCCGATCTatttccatgtctttgtcatagatg	GCTCTTCCGATCTcttccattgtggtttttaattcaac
SERPINA1	ACTCACGATGAAATCCTGGA	GCTCTTCCGATCTaagccttcatggatctgag	GCTCTTCCGATCTagtgagcatcgctacagc
TGFBR1	ATGGGCAAGACCGCTCGCCG	GCTCTTCCGATCTaacagatacacagttactgtg	GCTCTTCCGATCTaagcttaaataatagaactgcttatag
ERCC8	GATGTTGAAAGAATCCACGG	GCTCTTCCGATCTacttactatctcccttcaacag	GCTCTTCCGATCTttctccccttttatttttaagag
DNMT3B	CCCAACAACACGCAACCAGG	GCTCTTCCGATtgttggagtgggcccac	GCTCTTCCGATCTaagtggtcccaccagca
CD70	GAGCTGCAGCTGAATCACAC	GCTCTTCCGATCTctctctgtgcctcttcttc	GCTCTTCCGATCTtctgcttacgtccgtgc
FOXO1	GGTGGCGCAAACGAGTAGCA	GCTCTTCCGATCTtgctggattggccatatg	GCTCTTCCGATCTttaactgtttcgacccagtc
PCNT	CAGACTTTGAGGAACAACTG	GCTCTTCCGATCTtgaatttcagaccatccgtg	GCTCTTCCGATCTgtttctccaactcttgagtc

CXCR4	GCTTCTACCCCAATGACTTG	GCTCTTCCGATCTaggataaggccaaccatg	GCTCTTCCGATCTcgacttcatctttgccaac
IL2RG	GGGCAGCTGCAGGAATAAGA	GCTCTTCCGATCTgtgtcttcattcccattgg	GCTCTTCCGATCTgaatgaagagcaagcgc
TIGIT	CAGGCACAATAGAAACAACG	GCTCTTCCGATCTctgctgacccaggactc	GCTCTTCCGATCTaggagaggtgacattgtaag
PDCD1	GGCGCCCTGGCCAGTCGTCT	GCTCTTCCGATCTcccacctacctaagaaccatc	GCTCTTCCGATCTgagaaggcggcactctg
B2M	CGCGAGCACAGCTAAGGCCA	GCTCTTCCGATCTgggcattcctgaagctg	GCTCTTCCGATCTgggtaggagagactcacg

Table 3.7: Extended target NGS

Target	**Spacer**	**ddPCR amplicon 1 (control) fwd primer**	**ddPCR amplicon 1 (control) rev primer**	**ddPCR amplicon 1 (control) probe (5'-HEX)**
CD5-1	CAGTCGCTTCCTGCCTCGGA	GTAATCCTAGCTACTCAAGAG	TCACAATCAAGGTATACACCAG	CCACTGCACTCCAGCCTGGGT
CD5-2	CAGCATCTGTGAAGGCACCG	N/A	N/A	N/A
CD5-3	TTTCCTGAAGCAATGCTCCA	N/A	N/A	N/A
Target	**Spacer**	**ddPCR amplicon 2 (target) fwd primer**	**ddPCR amplicon 2 (target) rev primer**	**ddPCR amplicon 2 (target) probe (5'FAM)**
CD5-1	CAGTCGCTTCCTGCCTCGGA	CCCTAAAACCCATGATAGTGG	CAGGCCTTTCTCCATGTG	CACCCACAGCCCGAGCAATGG
CD5-2	CAGCATCTGTGAAGGCACCG	N/A	N/A	N/A
CD5-3	TTTCCTGAAGCAATGCTCCA	N/A	N/A	N/A

Table 3.8: CD5 HDR ddPCR

gRNA	Target	Spacer
12 (exon)	TRAC	CAGGGTTCTGGATATCTGT
13 (intron)	TRAC	CTGGATATCTGTGGGACAAG
NT	N/A	GCACUACCAGAGCUAACUCA

Table 3.9 CAR T HDR gRNAs

CHAPTER 4

Development of virus-like particles for targeted delivery of Cas9 ribonucleoproteins

A portion of the work presented in this chapter has been published previously as part of the following paper: Hamilton J.R.*, Tsuchida C.A.*, Nguyen D.N., Shy B.R., McGarrigle E.R., Sandoval Espinoza C.R., Carr D., Blaeschke F., Marson A., Doudna J.A. Targeted delivery of CRISPR-Cas9 and transgenes enable complex immune cell engineering. *Cell Reports* **35**, 109207, (2021).

*These authors contributed equally.

4.1 Abstract

As genome engineering advances cell-based therapies, a versatile approach to introducing both CRISPR-Cas9 ribonucleoproteins (RNPs) and therapeutic transgenes into specific cells would be transformative. Autologous T cells expressing a chimeric antigen receptor (CAR) manufactured by viral transduction are approved to treat multiple blood cancers, but additional genetic modifications to alter cell programs will likely be required to treat solid tumors and for allogeneic cellular therapies. We have developed a one-step strategy using engineered lentiviral particles to introduce Cas9 RNPs and a CAR transgene into primary human T cells without electroporation. Furthermore, programming particle tropism allows us to target a specific cell type within a mixed cell population. As a proof-of-concept, we show that HIV-1 envelope targeted particles to edit CD4$^+$ cells while sparing co-cultured CD8$^+$ cells. This adaptable approach to immune cell engineering *ex vivo* provides a strategy applicable to the genetic modification of targeted somatic cells *in vivo*.

4.2 Introduction

Engineering target specificity into immune cells enables the antigen-specific elimination of cells expressing cancer-associated epitopes[99]. Currently approved cell therapies require the isolation of patient T cells, the viral introduction of a chimeric antigen receptor (CAR) to redirect cytotoxic activity toward target cells, and subsequent reintroduction into the body. These autologous immune cell therapies can effectively combat blood cancers but remain largely ineffective for the treatment of solid tumors[290,291]. However, genome editing may have the potential to unleash engineered cell activity against solid tumors: knockout of transforming growth factor-β (TGF-β) receptor or programmed cell death protein-1 (PD-1) have both shown some promise for enhancing the activity of cell therapies[292–295]. In addition, genome editing may increase the accessibility of engineered cell therapies by enabling the production of “off-the-shelf” allogeneic cells with disrupted major histocompatibility complex class I (MHC class I) and endogenous T cell receptor expression, thereby minimizing the risks of rejection and graft-versus-host disease (GVHD)[296].Future immune cell therapies must therefore incorporate complex modifications of both targeted genetic disruption and stable gene addition. Current genetic engineering approaches, either to prevent premature exhaustion or enable allogeneic adoptive cell transfer, generally require viral transduction to program antigen specificity combined with the electroporation of nucleases to produce targeted genetic disruptions. Streamlining the generation of engineered immune cells with enhanced cytotoxic activity, resistance to cellular exhaustion, and minimized risk of rejection or GVHD would facilitate the next generation of universally accessible cellular therapies against solid tumor malignancies.

CRISPR-Cas9 genome editing enables the disruption of targeted genes but requires the effective delivery of genome editing tools into target cells[14,43,89,297,298]. The modification of Cas9 ribonucleoprotein (RNP) complexes with cell-penetrating and endosomolytic peptides has improved direct cellular uptake[134,299–301]; however, electroporation remains the predominant strategy for delivering Cas9 RNPs into the intracellular environment. Retroviral virus-like particles (VLPs) packaging Cas9 protein have been produced by fusing Cas9 directly to the group-specific antigen (Gag) structural protein[166,169]. This has proved an efficient strategy to promote Cas9 particle encapsidation and to couple the cell-entry mechanisms of an enveloped virus to the transient genome editing activity of Cas9 RNPs. However, VLPs are an unexplored strategy for linking the delivery of pre-formed Cas9 RNPs with a clinically relevant transgene and leveraging viral glycoprotein pseudotyping to direct genome editing to specific cell types.

Here, we optimize and demonstrate that engineered lentiviral particles can deliver Cas9 RNP complexes for genome editing, either tracelessly or while simultaneously integrating a lentiviral-encoded transgene (Cas9-VLPs) in immortalized cell lines and primary human T cells. Treatment of primary human T cells with Cas9-VLPs packaging a lentiviral-encoded CAR resulted in the simultaneous knockout of genetic targets relevant to allogeneic T cell production while effectively mediating CAR expression, an approach that was amenable to multiplexing. In addition, the treatment of T cells with broadly transducing Cas9-VLPs resulted in targeted genetic knockout in $CD4^+$ and $CD8^+$ T cells, while treatment with Cas9-VLPs pseudotyped with the CD4-tropic HIV-1 envelope glycoprotein drove the exclusive transduction and genome editing of $CD4^+$ T cells within a mixed cell population. These data establish Cas9-VLPs as an effective approach for mediating cell-type-specific genome editing using Cas9 RNPs. As Cas9-VLPs circumvent the requirement for *ex vivo* Cas9 RNP delivery via electroporation, this strategy suggests a path forward for leveraging the tropism of viral glycoproteins in targeting specific cell types for genome engineering *in vivo*.

4.3 Results

4.3.1 Engineering lentivirus-based VLPs for the controlled delivery of Cas9 RNP complexes

Lentiviral production involves the co-transfection of producer cells with plasmids encoding the viral structural components, viral glycoprotein, and lentiviral transfer plasmid with a transgenic sequence flanked by long terminal repeat (LTR) sequences. To promote the packaging of Cas9 protein in HIV-1 VLPs (Cas9-VLPs), we constructed a plasmid to express *Streptococcus pyogenes* Cas9 fused to the C terminus of the Gag polypeptide and included this during lentiviral production. A lentiviral transfer plasmid encoding the expression cassettes for both an mNeonGreen fluorescent reporter and a single guide RNA (sgRNA) was included (Figure 4.1A). To promote the separation of Cas9 from Gag during proteolytic virion maturation, we inserted an HIV-1 protease-cleavable linker between Gag and Cas9 (Figure 4.1B). We produced Cas9-VLPs pseudotyped with the broadly transducing vesicular stomatitis virus glycoprotein (VSV-G) and varied the ratio of Gag-pol and Gag-Cas9 plasmids to optimize Cas9 incorporation in budded particles. Bands corresponding to the expected size of Cas9 fused to Gag (55 kDa + 160 kDa = 215 kDa) and proteolytically released Cas9 (160 kDa) were detectable by western blot in all of the Cas9-VLP formulations tested (Figure 4.1C). A component of the Gag polypeptide, capsid (CA), was used for quantifying Cas9-VLP production by ELISA. CA-containing particles were detected for all of the formulations except for Cas9-VLP formulation F (Figure 4.1D). Formulation F is composed entirely of Gag-Cas9 polypeptides, which may interfere with the successful budding of Cas9-VLPs from transfected cells.

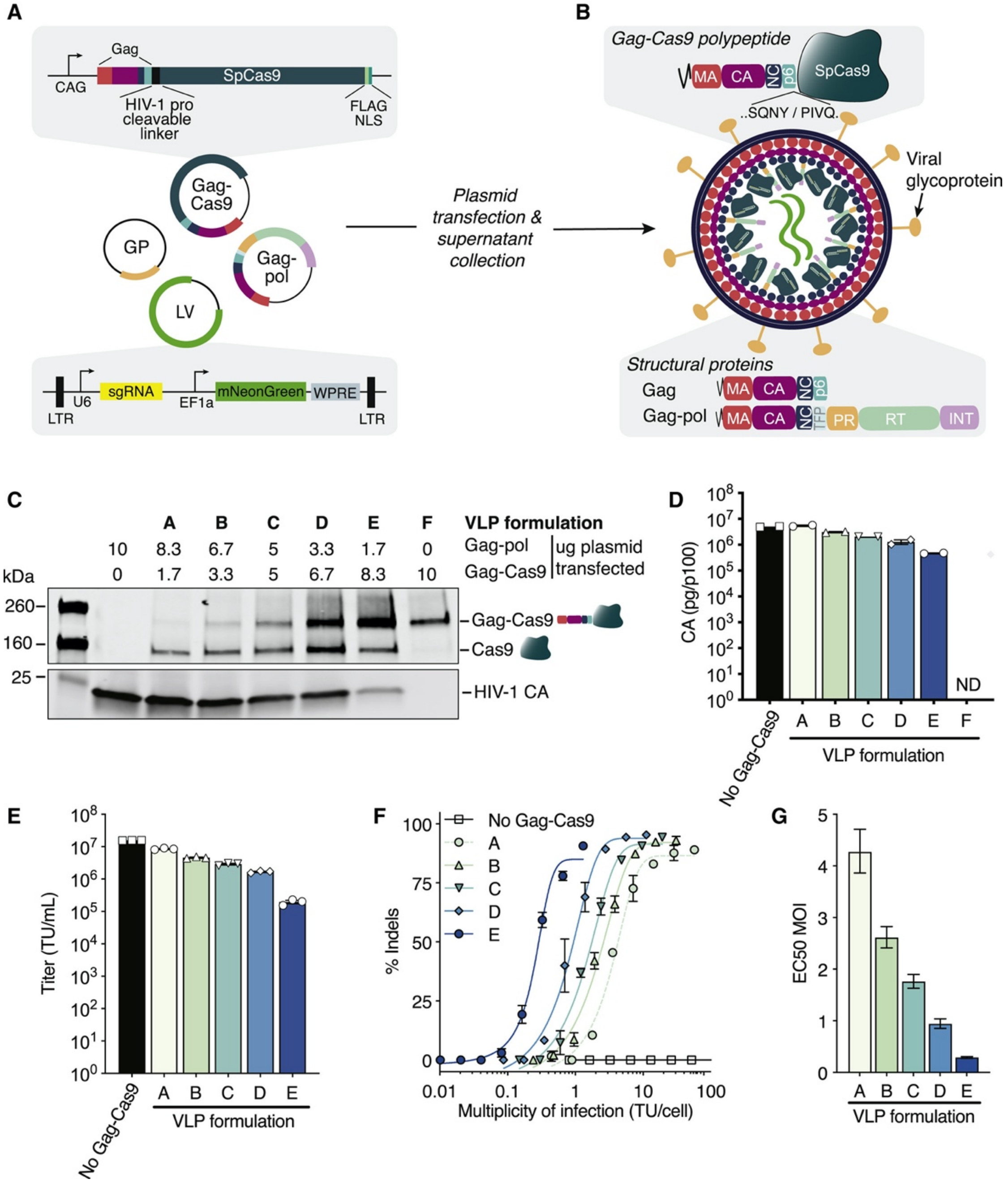


Figure 4.1: Production of Cas9-VLPs. (A) Schematic of plasmids used for the production of Cas9-VLPs. GP = glycoprotein. LV = lentiviral transfer plasmid. LTR = long terminal repeat. **(B)** Schematic of an immature, pre-proteolytically processed Cas9-VLP produced through transient transfection. An HIV-1 protease cleavable linker containing the amino acid sequence SQNY/PIVQ was inserted between Gag and Cas9 to promote separation during proteolytic virion maturation. **(C)** Western blot of Cas9-VLP content when various ratios of Gag-pol to Gag-Cas9 plasmids are used for production. An anti-Flag antibody was used for Cas9 detection and an anti-HIV-1 p24 capsid (CA) antibody was used to detect Cas9-VLP production. **(D)** Quantification of Cas9-VLP production by CA ELISA. Amount of CA produced per

transfected p100 dish plotted, n = 2 technical replicates. Error bars indicate standard error of the mean. ND = not detected.

We hypothesized that Cas9-VLPs packaging relatively high Gag-Cas9 polypeptide content would require fewer individual Cas9-VLPs to deliver a sufficient quantity of Cas9 RNPs for successful genome editing. To assess genome editing activity, Cas9-VLPs were produced with a lentiviral transfer plasmid expressing a sgRNA targeting the β-2 microglobulin (*B2M*) gene. The transduction-competent Cas9-VLP titer (transducing units TU/mL) was quantified for each Cas9-VLP preparation (Figure 4.1E) and used to calculate the multiplicity of infection (MOI, TU/cell) required to achieve 50% editing (effective concentration 50 [EC_{50}] MOI) in the Jurkat cells (Figure 4.1F). We confirmed that as increasing amounts of Gag-Cas9 are packaged per particle, a lower EC_{50} MOI is needed to achieve genome editing (Figure 4.1G), with an approximate MOI of 2.6 required to achieve 50% indels using Cas9-VLP formulation B and an MOI of 0.9 using Cas9-VLP formulation D.

4.3.2 Characterization of Cas9-VLPs for genome editing

We next assessed the kinetics of genome editing following Cas9-VLP treatment. Jurkat and A549 cells were treated with formulation D B2M-targeting Cas9-VLPs, and cell-surface-expressed B2M protein was assessed by flow cytometry at 3, 6, and 8 days post-treatment. We observed dose-dependent knockout of B2M protein by day 3 (Figure 4.2A), with a maximum loss of protein expression achieved by day 6. We further confirmed genetic knockout by next-generation sequencing and observed B2M-guide-specific indels at the *B2M* locus (Figure 4.2B). Similar to what has been observed for Cas9-packaging MLV VLPs[169], mixing Cas9-VLPs with a DNA template was sufficient to mediate homology-directed repair (HDR) in a fluorescence reporter assay[32]. We found that Cas9-VLP-directed HDR activity could be further enhanced by electroporating Cas9-VLPs with the DNA template before the treatment of target cells, which may promote the complexing of Cas9-VLPs with the HDR template (Figure 4.3).

Current RNP-based genome editing approaches have not permitted the quantification of cells edited as a function of the number of cells receiving RNPs. We reasoned that Cas9-VLPs co-delivering Cas9 RNPs and a lentiviral genome may enable tracking cells that receive Cas9 RNPs. To assess whether transduction is a marker of RNP-edited cells, we treated A549s and Jurkats with serial dilutions of B2M-Cas9-VLPs delivering the mNeonGreen transgene and quantified B2M expression at day 6 post-treatment (Figure 4.2C; Figure 4.4). For Jurkats, successfully edited cells largely correlated with the transduction marker mNeonGreen; however, we did observe a population of B2M-knockout cells that did not express mNeonGreen. We hypothesized that this discordance could be explained by a proportion of Cas9-VLPs not co-packaging both the lentiviral genome and Cas9 RNPs. However, in A549 cells treated with the same Cas9-VLP preparation, cells lacking B2M overwhelmingly expressed the transduction marker. This suggests that in the A549 cell line, transduction is a reliable marker for identifying the cell population edited by Cas9 RNPs and that Jurkats may use a cell-intrinsic mechanism restricting reverse transcription of the lentiviral genome, nuclear import, or integration independent of Cas9-mediated genome editing.

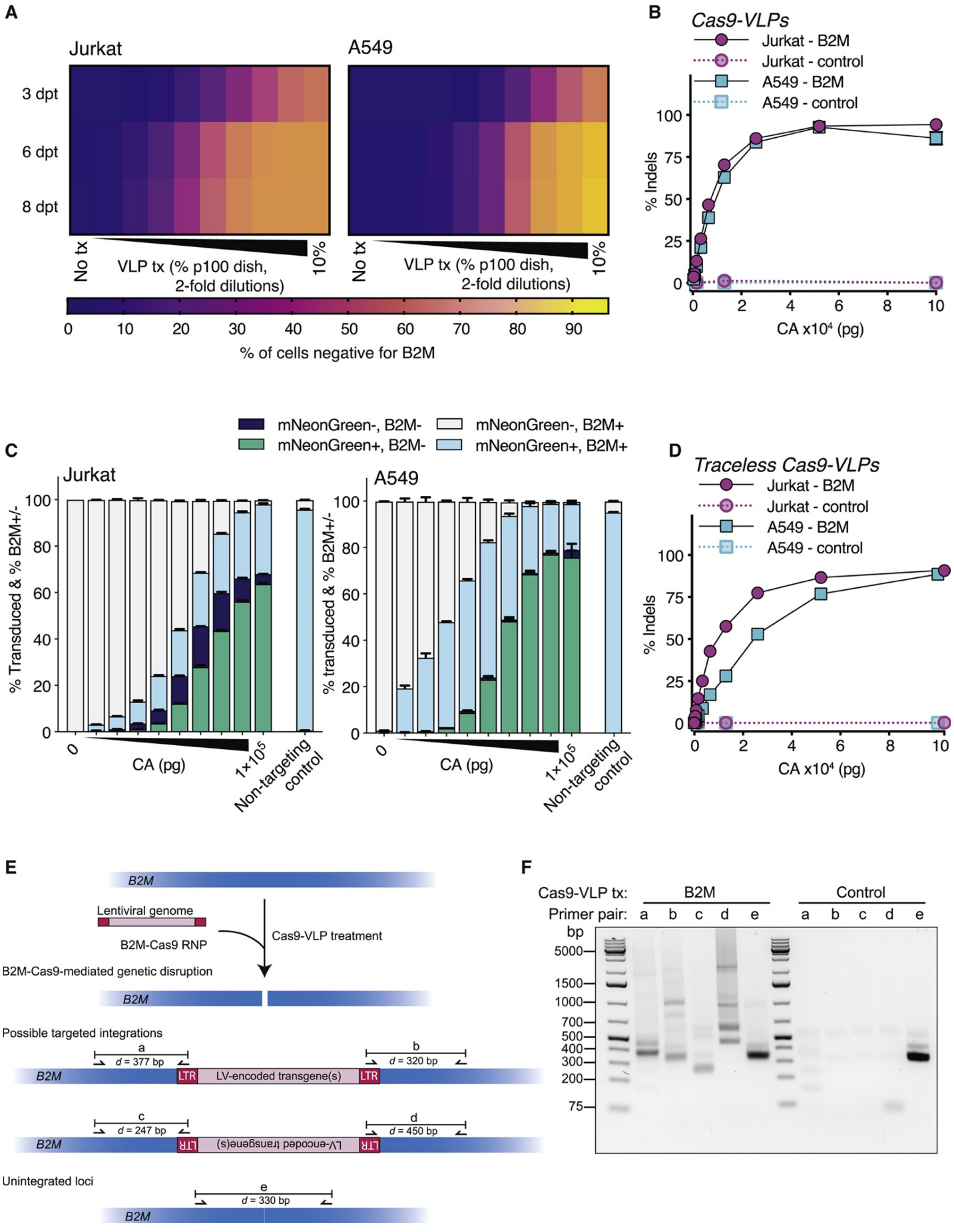


Figure 4.2: Increasing Gag-Cas9 content reduces the multiplicity of infection needed for Cas9-VLP-mediated genome editing. (A) Cas9-VLP formulations A-E and VLPs lacking Gag-Cas9 were produced

and used to treat Jurkat cells. Transducing units (TU) per mL were calculated by quantifying the number of mNeonGreen+ cells treated with a 1:256 Cas9-VLP dilution 6 days post treatment. **(B)** The percentage of indels plotted against the multiplicity of infection (MOI) for each Cas9-VLP formulation and a sigmoidal four parameter logistic fit was applied. Indels were quantified using Synthego's ICE analysis tool. **(C)** The predicted MOI for each Cas9-VLP formulation to achieve 50% indels, (interpolated from **B**). EC50 = effective concentration at which a drug gives a half-maximal response. n = 3 technical replicates were run (**A**, **B**), except for formulation A which was run as n = 2 (**B**). Error bars indicate standard error of the mean (**A**, **B**) and 95% confidence interval (**C**).

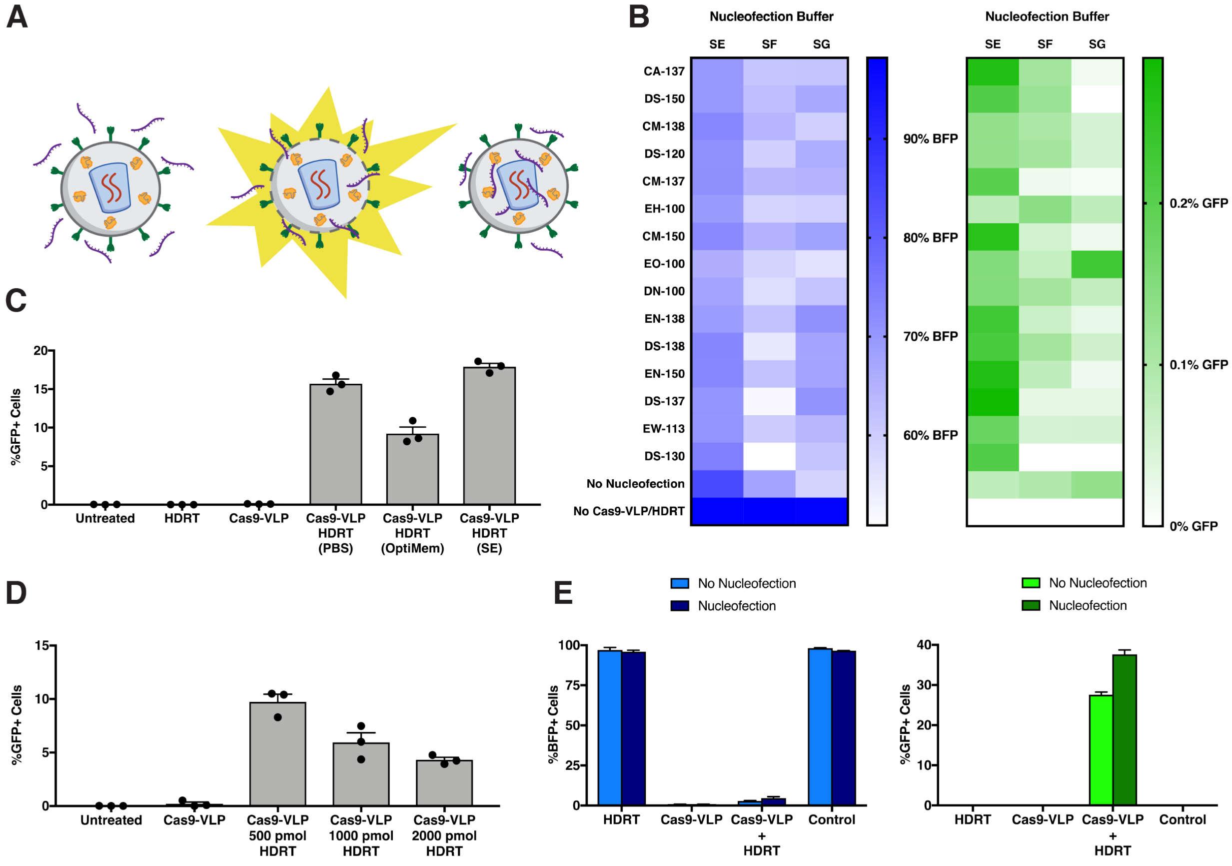


Figure 4.3: Cas9-VLPs mediate homology-directed repair (HDR). (A) Schematic of nucleofection of Cas9-VLPs and single-stranded DNA homology-directed repair templates (HDRT, purple). **(B)** Assessment of different Lonza nucleofection buffers and pulse codes, 5 days post treatment. Cas9-VLPs packaging BFP-targeting RNPs were mixed with 80 pmol HDRT and nucleofected using the indicated nucleofection buffers and pulse codes. Nucleofected HDRT/Cas9-VLPs were subsequently used to treat a BFP-to-GFP HDR reporter HEK293 cell line (Richardson et al., 2016) where BFP knockout is indicative of non-homologous end joining and GFP expression is representative of HDR. **(C)** HDR-mediated GFP expression induced treatment with Cas9-VLPs nucleofected (Lonza, CM-150) with 500 pmol HDRT in different buffers, 7 days post treatment. **(D)** HDR-mediated GFP expression with varying concentrations of HDRT nucleofected (Lonza, CM-150) with Cas9-VLPs in SE buffer (Lonza), 7 days post treatment. **(E)** Pre-nucleofection of Cas9-VLPs and HDRT enhances HDR activity. Cas9-VLPs (2.59x10^6 pg CA) and 500pmol HDRT were mixed in SE buffer and either directly added to BFP-to-GFP reporter cells or subjected to nucleofection (Lonza, CM-150) prior to cell treatment. BFP-positive and GFP-positive cells were quantitated by flow cytometry at 7 days post treatment. All error bars represent standard error of the mean.

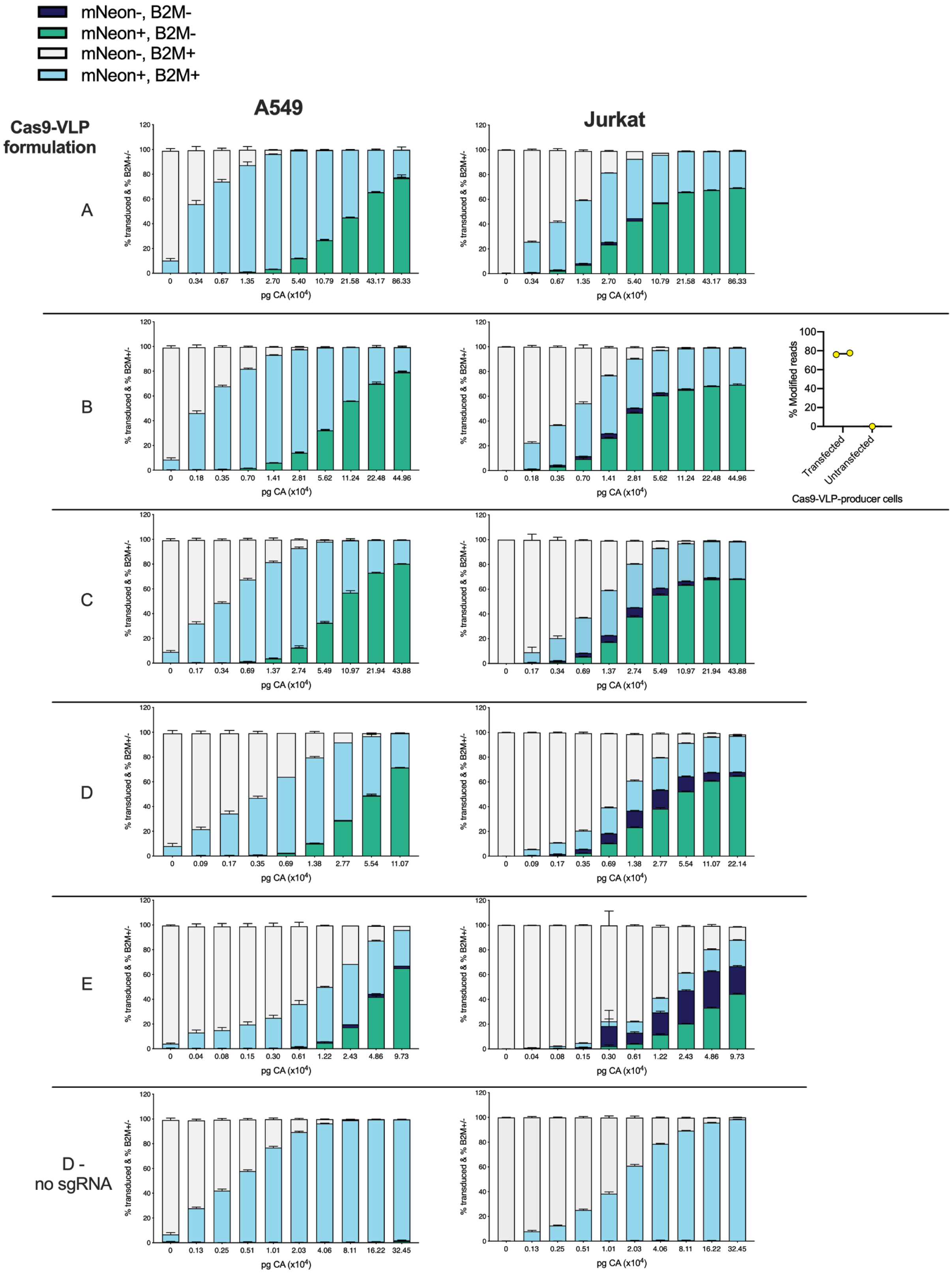


Figure 4.4: All Cas9-VLP formulations mediate genome editing. Jurkat or A549 cells were treated with B2M-Cas9-VLP formulations A-E and transduction (mNeonGreen+) and B2M-PE expression was assessed

by flow cytometry 6 days post treatment. Of note, cells transfected to produce B2M-Cas9-VLPs themselves undergo genome editing (DNA isolated 3 days post transfection). n = 3 technical replicates were performed at each Cas9-VLP treatment dose and error bars represent standard error of the mean.

As the sgRNA expression cassette is embedded within the lentiviral genome, sgRNA transcription could occur both in the packaging cell line during Cas9-VLP production and in transduced cells. To assess whether Cas9 RNP formation occurs predominantly in the packaging cells or in the treated cells, we produced Cas9-VLPs lacking a lentiviral genome and instead expressed the B2M sgRNA from an orthogonal expression plasmid. We found that treatment with “traceless” Cas9-VLPs mediated high levels of editing (Figure 4.2D; Figures 4.5A-D), suggesting that the majority of Cas9 RNPs are formed within the packaging cell line. We noted a slight increase in editing efficiency when a lentiviral genome including the sgRNA expression cassette was co-packaged within the Cas9-VLP (compare Figure 4.2B versus 4.2D), suggesting, at this concentration, that a fraction of VLP-packaged Cas9 may remain in the guideless apo-Cas9 state until sgRNA transcription occurs in treated cells. It was also possible to generate hybrid Cas9-VLPs that co-package a lentiviral genome but do not require a lentiviral-encoded guide RNA expression cassette (Figures 4.5E and 4.5F). The ability of Cas9-VLPs to deliver Cas9 RNPs without co-packaging a lentiviral genome enables Cas9-VLPs to mediate genome editing in the absence of transgene integration, which may be advantageous for clinical applications.

As Cas9-VLPs deliver the reverse-transcribed viral genome concomitantly with double-stranded DNA (dsDNA) break-inducing Cas9 RNPs, we reasoned that targeted lentiviral insertion may occur at the genomic site targeted for genome editing. To investigate this possibility, we isolated DNA from cells treated with either *B2M*-targeting or non-targeting Cas9-VLPs co-packaging a lentiviral genome. Amplification of cellular genomic DNA with primers specific to the *B2M* locus and the lentiviral LTR resulted in detectable viral insertion at the Cas9-targeted region (Figures 4.2E and 4.2F). This was further validated using primers specific to the *B2M* locus and the lentiviral Psi sequence, and next-generation sequencing confirmed targeted lentiviral integration (Figure 4.6).

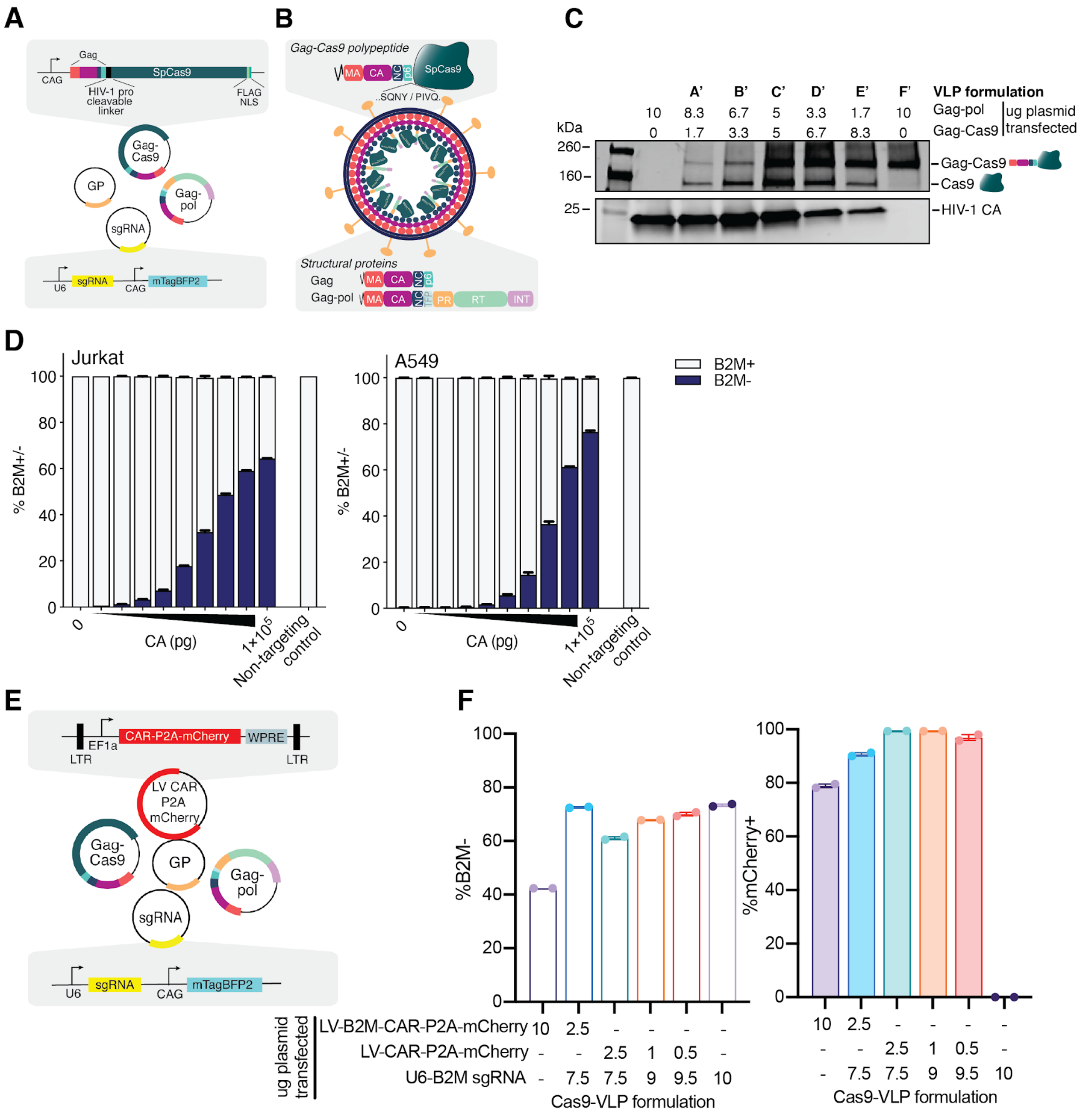


Figure 4.5: Traceless Cas9-VLPs mediate genome editing without viral transgene insertion and hybrid Cas9-VLPs do not require a lentiviral-encoded guide RNA expression cassette. (A) Schematic of plasmids used for the production of traceless Cas9-VLPs. GP = glycoprotein. **(B)** Schematic of an immature, pre-proteolytically processed Cas9-VLP, produced through transient transfection and lacking a lentiviral genome. An HIV-1 protease cleavable linker containing SQNY/PIVQ was inserted between the c-termini of Gag and the n-termini of Cas9 to promote the separation during proteolytic virion maturation. **(C)** Western blot of Cas9-VLP content when various ratios of Gag-pol to Gag-Cas9 plasmids are used for production. An anti-Flag antibody was used for Cas9 detection and an anti-HIV-1 capsid (CA) antibody was used to detect Cas9-VLP production. A' is used to indicate VLP formulation "A" lacking a packaged lentiviral genome. **(D)** Flow cytometry quantification of B2M expression in A549 and Jurkats 6 days post treatment with traceless Cas9-VLPs. Non-targeting control = Cas9-VLPs packaging the tdTom298 sgRNA. n = 3 technical replicates were performed at each Cas9-VLP treatment dose and error bars indicate standard error of the mean. **(E)** Schematic of plasmids used for the production of Cas9-VLPs that co-package Cas9 RNPs and a lentiviral genome that lacks a guide RNA expression cassette ("hybrid Cas9-VLPs"). **(F)**

Optimization of hybrid Cas9-VLPs. Cas9-VLPs were produced as indicated and used to treat Jurkat cells. Targeted protein disruption (% of cells negative for B2M expression) and transduction (% of cells mCherry positive) were quantified at day 7. LV-B2M-CAR-P2A-mCherry = lentiviral transfer plasmid that encodes the U6-promoter driven expression of a *B2M*-targeting guide RNA and the EF1a-promoter driven expression of a CAR-P2A-mCherry transgene. LV-CAR-P2A-mCherry = lentiviral transfer plasmid that encodes the CAR-P2A-mCherry expression cassette alone. U6-B2M = a transient guide RNA expression plasmid.

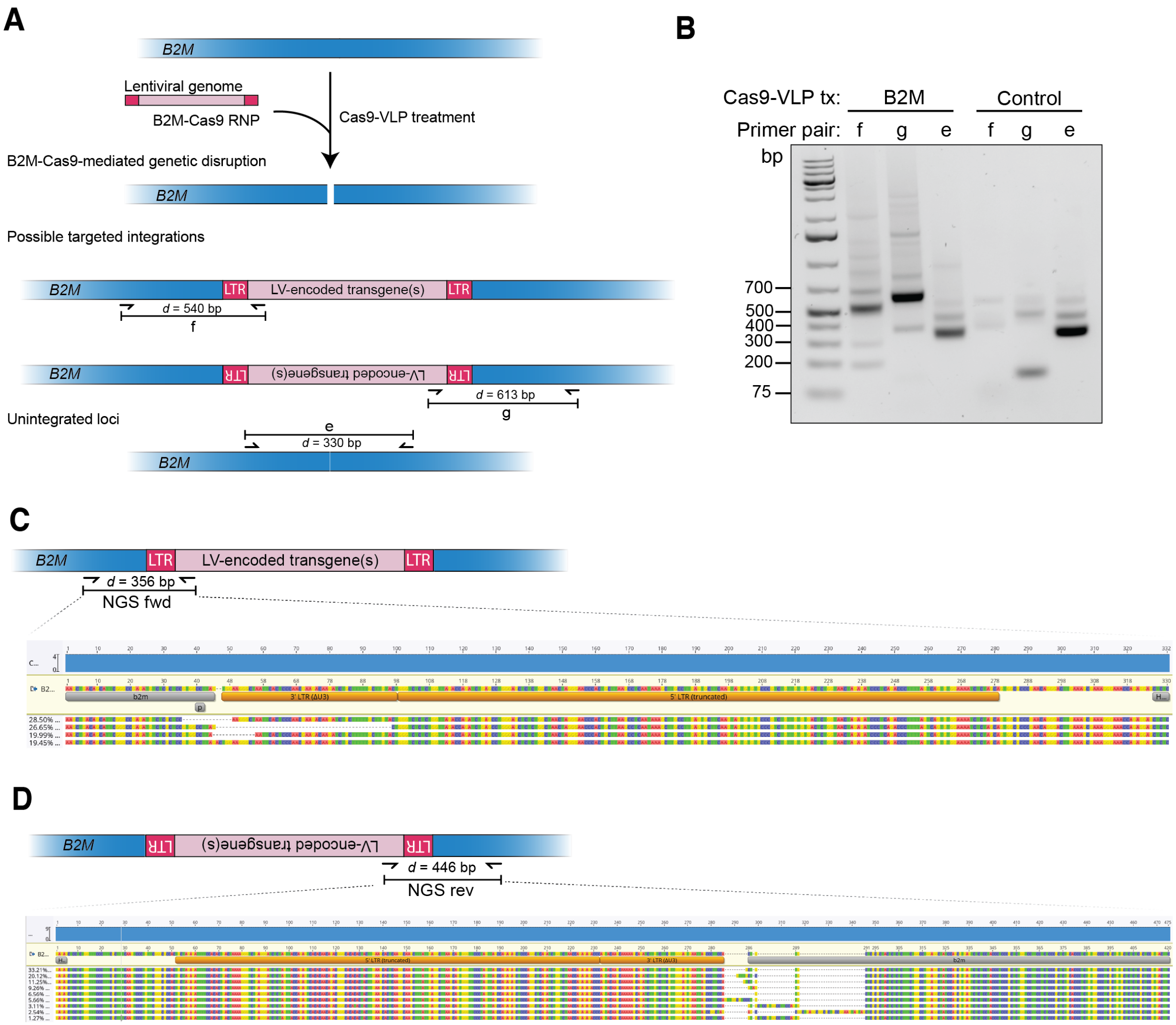


Figure 4.6: Targeted integration of the lentiviral genome into the Cas9 RNP target site. (A) Schematic of hypothetical lentiviral insertion at the Cas9 RNP-induced double-stranded DNA break. **(B)** PCR to assess targeted lentiviral integration. DNA was isolated from 293T cells 3 days post treatment with B2M-targeting or non-targeting Cas9-VLPs and the indicated primer pairs were used for analysis. **(C)** MiSeq analysis of the targeted “forward” lentiviral integration. Reads mapped to the hypothetical B2M-lentiviral junction are shown. **(D)** MiSeq analysis of the targeted “reverse” lentiviral integration. Reads mapped to the hypothetical B2M-lentiviral junction are shown.

4.3.3 Cas9-VLPs efficiently edit primary human T cells

Engineered T cell therapies are transforming the treatment of certain cancers by retargeting T cell activity through the introduction of antigen-specific receptors such as CARs[302,303]. We next tested whether Cas9-VLPs could mediate genome editing in primary human T cells. Transducing bulk $CD4^+$ and $CD8^+$ primary human T cells with Cas9-VLPs resulted in B2M knockout levels comparable to Cas9 RNP electroporation, the current clinical standard (Figures 4.7A and 4.7B; Figure 4.8). Cas9-VLP-mediated transduction and B2M knockout was dose dependent and cellular viability (as measured by relative cell count) was improved compared to previous reports using Cas9 RNP nucleofection (Figure 4.7B)[64].

Recently, transgenic T cell receptor (TCR) T cells modified by CRISPR-Cas9 were tested in the first phase I clinical trial[101]. The engineered T cell product was produced by the electroporation of Cas9 RNPs to first disrupt expression of PD-1 and the endogenous TCR (by targeting *PDCD1* and *TRAC*, respectively), followed by subsequent lentiviral transduction to integrate an exogenous TCR for retargeting antigen specificity. We hypothesized that Cas9-VLPs could simplify the production of multiply edited engineered T cells by simultaneously delivering Cas9 RNPs and a lentiviral genome encoding a transgenic TCR or CAR (Figure 4.7D). To test this, we assessed whether it was possible to multiplex genetic knockout by treating primary human T cells with Cas9-VLPs targeting two genetic loci for disruption. Treatment of primary human T cells with separate Cas9-VLPs targeting *B2M* or *TRAC* resulted in 23.9% $CD4^+$ and 9.55% $CD8^+$ double-knockout cells by 13 days post-treatment (Figure 4.7C; Figure 4.9A). We next optimized the production of Cas9-VLPs to maximize the simultaneous integration of a lentiviral-encoded CAR and knockout of B2M expression in Jurkats (“CAR-Cas9-VLPs”) (Figure 4.9B). To determine how capsid quantity correlates to MOI in primary T cells, we next assessed genome editing levels generated using both mNeonGreen and CAR Cas9-VLPs. An approximate MOI of 20 for mNeonGreen Cas9-VLPs resulted in ~7% of cells lacking B2M protein while the equivalent MOI for CAR-Cas9-VLPs resulted in ~28% $B2M^-$ cells (Figure 4.10). The enhanced editing efficiency of CAR-Cas9-VLPs may be explained by a higher proportion of VLP-packaged Cas9 being associated with guide RNA, as the optimized CAR-Cas9-VLP production involves the overexpression of guide RNA in VLP producer cells (Figure 4.9B). Finally, we generated CAR-Cas9-VLPs packaging Cas9 RNPs targeting either *B2M* or *TRAC* for disruption; the treatment of primary T cells exhibited dose-dependent CAR-P2A-mCherry expression and reduction in surface-expressed B2M or TCR (Figure 4.7E; Figures 4.9C and 4.9D). In addition, Cas9-VLP-engineered CAR T cells were functionalized to kill $CD19^+$ Nalm-6 target cells (Figure 4.7F) and stimulation resulted in effector profiles for cytokine production and activation marker expression (Figure 4.11). Cas9-VLPs provide a simplified workflow for manufacturing complex CRISPR-modified CAR T cells in a single step, which compares favorably to current clinical manufacturing methods.

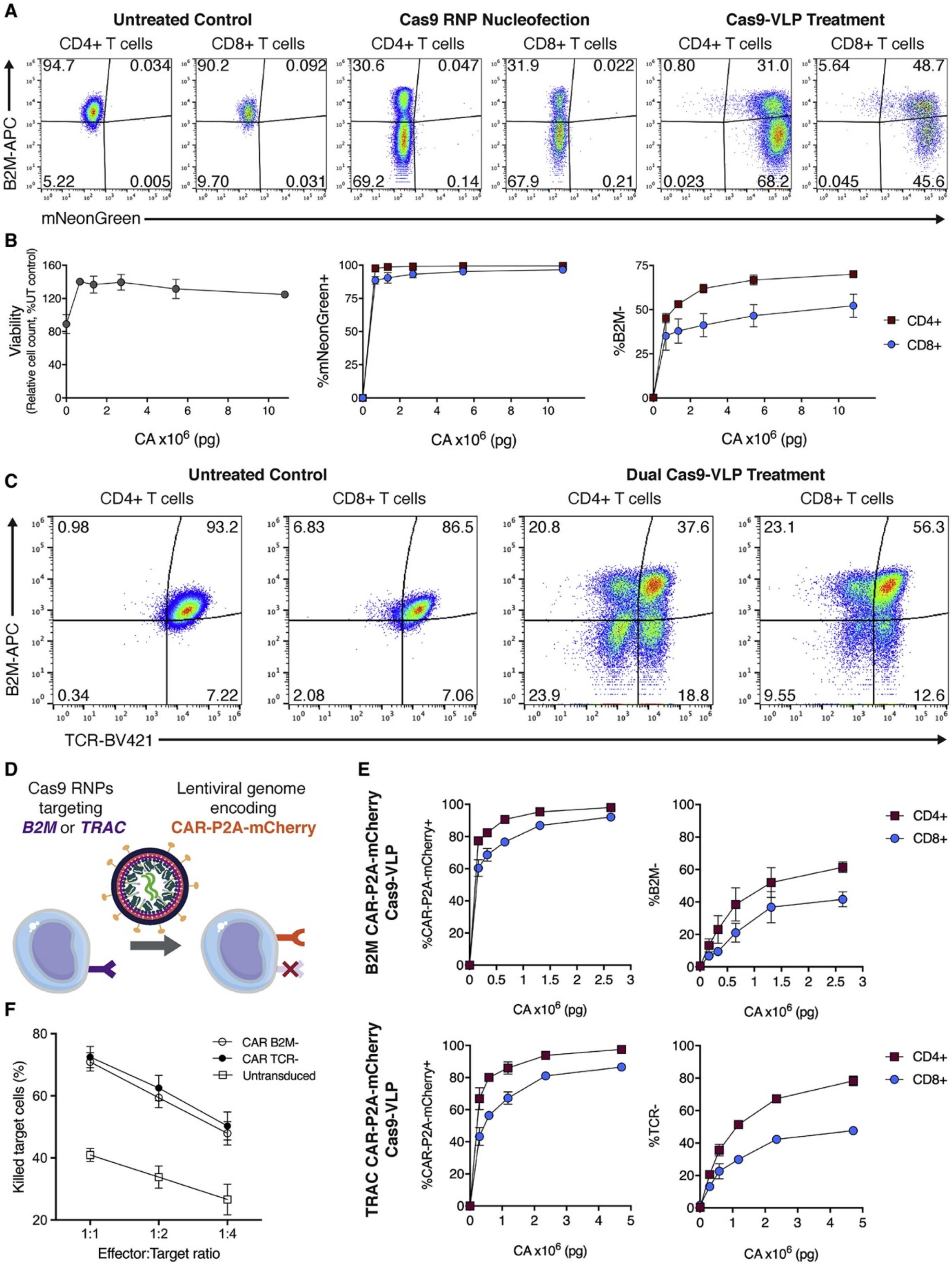


Figure 4.7: Genome editing and generation of highly engineered CAR-expressing primary human T cells using Cas9-VLPs. (A) *B2M* genome editing mediated by Cas9 RNP nucleofection and Cas9-VLP treatment in primary human T cells. Cas9-VLP transduction (mNeonGreen) and surface expression of B2M

protein were quantified by flow cytometry 7 days post treatment. **(B)** Viability, transduction, and B2M expression were measured in primary human T cells treated with Cas9-VLPs. B2M expression is plotted for both CD4+ (red squares) and CD8+ (blue circles) T cell subpopulations. **(C)** Simultaneous treatment with two Cas9-VLPs targeting distinct genomic targets results in multiplexed genome editing. Cas9-VLPs targeting *B2M* and Cas9-VLPs targeting *TRAC* were mixed and used to treat primary human T cells. Surface expression of B2M and TCR was assessed by flow cytometry at 13 days post treatment. n = 2 biological replicates from independent donors were used (**A**, **B**, **C**) and representative flow cytometry plots are shown for one donor (**A**, **C**). **(D)** Schematic of a single-step method to generate highly engineered CAR-T cells. Cas9-VLPs targeting *B2M* or *TRAC* for gene editing were packaged with a lentiviral genome encoding CAR-P2A-mCherry. **(E)** CAR-encoding Cas9-VLPs targeting *B2M* (top two panels) or *TRAC* (bottom two panels) were used to engineer bulk primary human T cells and analyzed 12 days post treatment. Knockout was assessed for both CD4+ (red squares) and CD8+ (blue circles) T cell subpopulations. **(F)** CAR-T cells generated by Cas9-VLP treatment, or untreated primary human T cells, were co-cultured in CD19+ Nalm-6 cells and cytotoxic killing activity was measured at 24 h. Error bars indicate standard error of the mean.

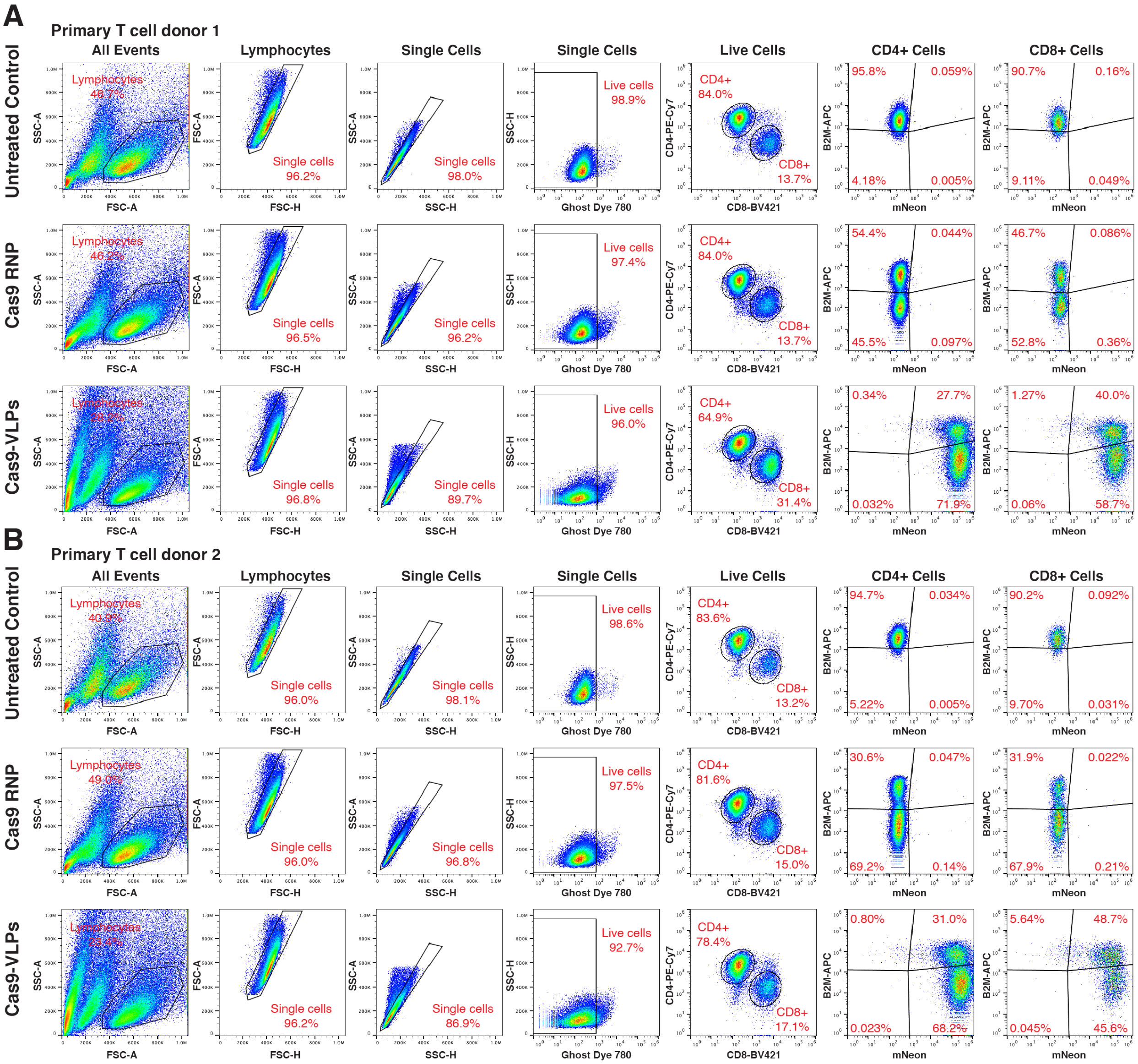


Figure 4.8: Representative flow cytometry gating strategy for quantifying genome editing in primary human T cells. **(A)** Flow cytometry gating strategy to assess surface-expressed B2M in primary human T cells after no treatment, nucleofection of Cas9 RNPs, and treatment with Cas9-VLPs from donor 1. **(B)** Flow cytometry gating strategy to assess surface-expressed B2M in primary human T cells after no treatment, nucleofection of Cas9 RNPs, and treatment with Cas9-VLPs from donor 2.

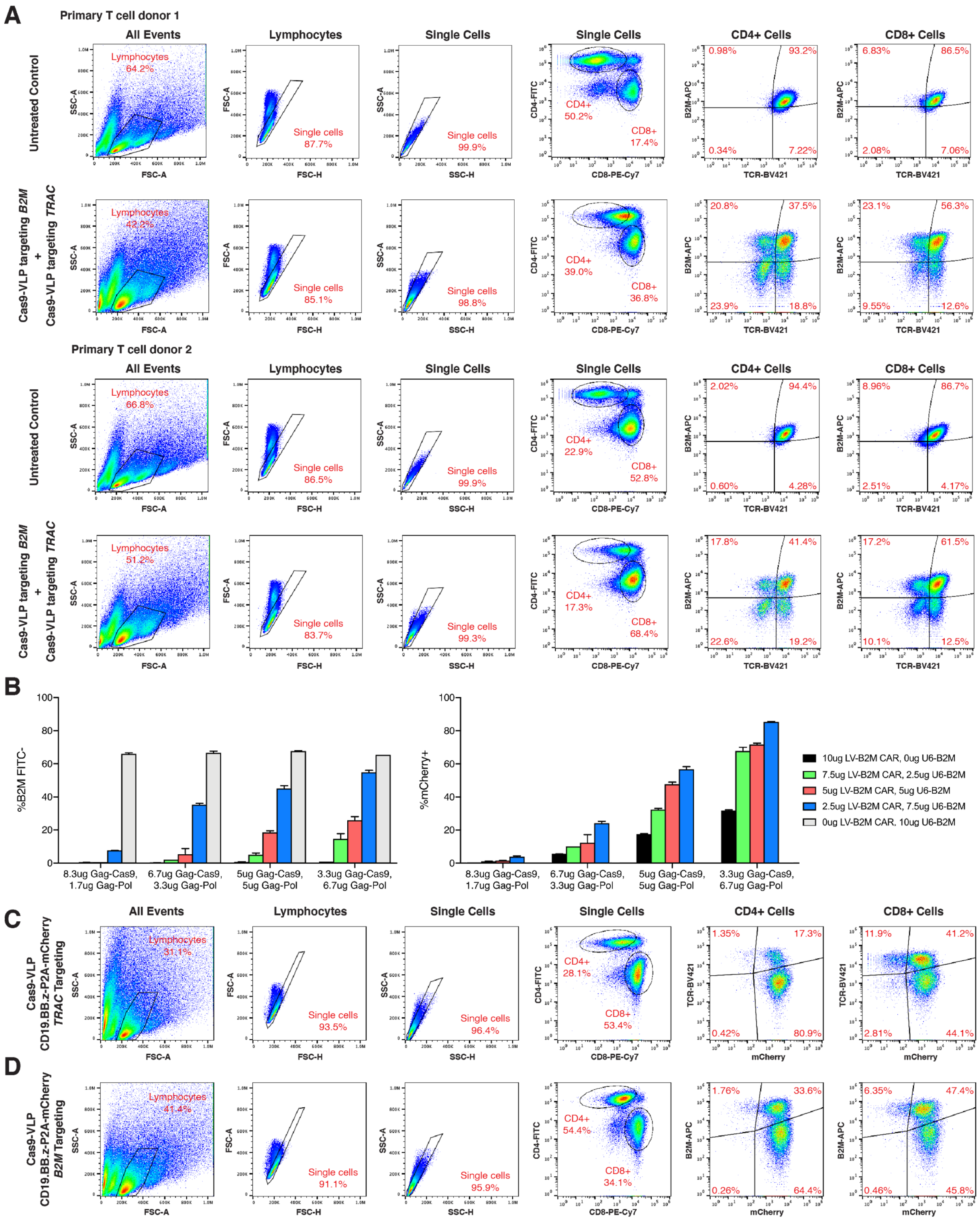


Figure 4.9: Optimization of CAR-Cas9-VLP production & representative flow cytometry gating strategy for Cas9-VLP-mediated multiplexed genome engineering of primary human CAR-T cells. (A) Flow cytometry gating strategy to assess the dual knockout of surface-expressed TCR and B2M by simultaneous treatment with Cas9-VLPs targeting *TRAC* and Cas9-VLPs targeting *B2M* in two independent

T cell donors. Cas9-VLPs optimized for simultaneous CAR transgene insertion and B2M knockout were used (**Supp. 5B**). **(B)** Optimization of Cas9-VLP production to maximize simultaneous CAR transgene integration and genome editing. Cas9-VLPs were produced with various ratios of plasmids encoding the Gag-Cas9 and Gag-pol structural proteins, and with various ratios of plasmids encoding a lentiviral transfer plasmid (encoding expression cassettes for U6-B2M CAR-P2A-mCherry) and a U6-B2M guide RNA expression plasmid. Jurkats were treated, passed at day 4 post treatment to maintain subconfluent culture conditions and flow cytometry was performed at 6 days post treatment to quantify B2M expression (**B,** left) and CAR-P2A-mCherry expression (**B,** right). Cas9-VLPs produced through transient transfection with the following plasmids were most efficient at mediating simultaneous knockout of B2M and CAR-P2A-mCherry transgene expression: 1µg VSV-G, 3.3µg Gag-Cas9, 6.7µg Gag-pol plasmid, 2.5µg LV-B2M, and 7.5µg U6-B2M. n = 2 replicates per treatment, error bars represent standard error of the mean. **(C)** Flow cytometry gating strategy to assess the knockout of surface-expressed TCR and expression of CAR-P2A-mCherry in primary human T cells by treatment with Cas9-VLPs. **(D)** Flow cytometry gating strategy to assess the knockout of surface-expressed B2M and expression of CAR-P2A-mCherry in primary human T cells by treatment with Cas9-VLPs.

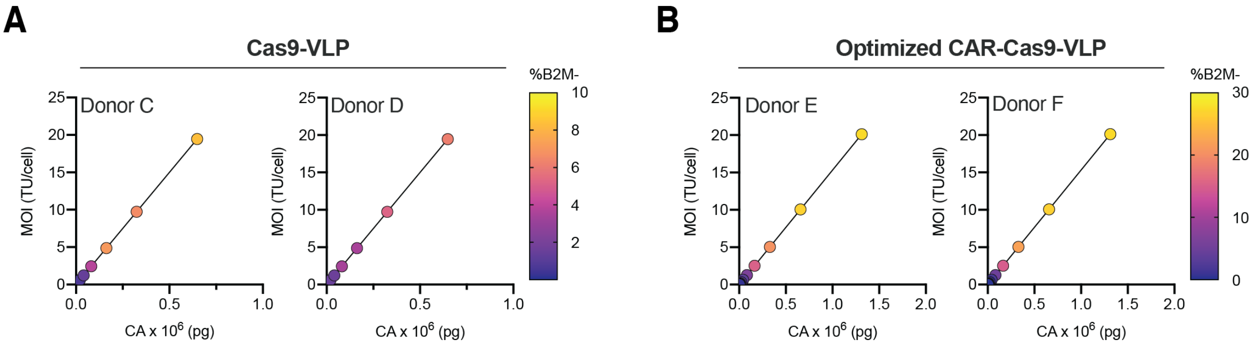


Figure 4.10: Cas9-VLP genome editing as a function of MOI and quantity of CA. (A) Cas9-VLPs co-packaging B2M-targeting Cas9 RNPs and a lentiviral genome encoding mNeonGreen were generated (as used in Figure 4) and **(B)** Cas9-VLPs optimized to co-package B2M-targeting Cas9 RNPs and a lentiviral genome encoding CAR-P2A-mCherry were produced. The transducing units/mL (TU/mL) titer and capsid (CA) content were quantified for each Cas9-VLP preparation. Primary T cells from two human donors were treated with indicated multiplicity of infection (MOI) and picogram (pg) CA and cells negative for B2M protein were quantified by flow cytometry at day 7.

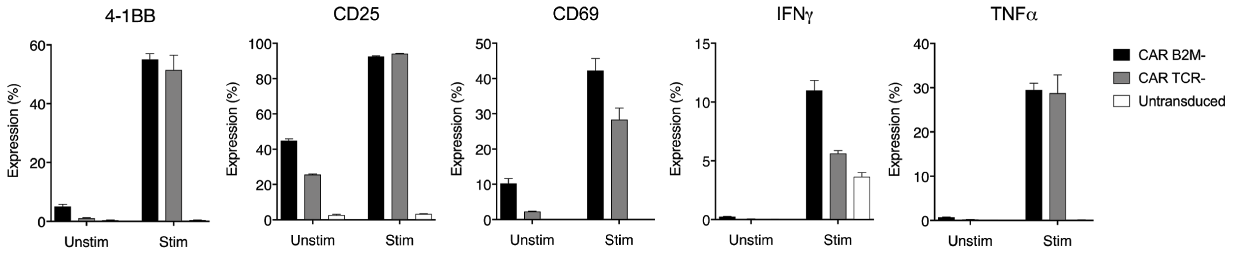


Figure 4.11: Functional cytokine production and surface receptor expression in Cas9-VLP generated CAR-T cells. Cytokine and surface receptor expression were quantified in stimulated and unstimulated CAR-T cells generated from Cas9-VLPs at 24 h. For all, n = 2 biological replicates from independent donors were used and error bars indicate standard error of the mean.

4.3.4 Cell-type-specific editing via pseudotyping of Cas9-VLPs

Virus and VLP cell-type specificity may be altered through pseudotyping with varied surface glycoproteins[304]. To test whether the Cas9-VLP glycoproteins were essential for the genome editing of mammalian cells, we produced Cas9-VLPs lacking viral glycoproteins ("bald" Cas9-VLPs) and assessed their ability to mediate genome editing. Bald Cas9-VLPs were effectively produced (Figures 4.12A-D), but cellular treatment resulted in <0.1% of reads containing indels by deep sequencing, a 3-log reduction in genome editing compared to treatment with VSV-G pseudotyped Cas9-VLPs (Figure 4.13A). Efficient delivery of VLP-packaged Cas9 RNPs is therefore dependent upon the expression of viral glycoproteins. To test whether we could engineer Cas9-VLPs to target a specific cell type for genome editing, we produced Cas9-VLPs pseudotyped with the HIV-1 envelope glycoprotein (Env), the viral determinant for the CD4$^+$ T cell tropism of HIV-1 (Figures 4.12E and 4.12F)[305]. Env-Cas9-VLPs were produced packaging Cas9 RNPs targeting the human *B2M* locus and an mNeonGreen-expressing lentiviral genome. A mixture of CD4$^+$ and CD8$^+$ T cells were treated with Env-Cas9-VLPs, and transduction and B2M protein expression were assessed. At the highest treatment dose, Env-Cas9-VLPs preferentially transduced CD4$^+$ cells over CD8$^+$ cells (53.20% versus 2.51%, respectively) (Figures 4.13B and 4.13C; Figure 4.12G). Concomitantly, Env-Cas9-VLP treatment resulted in the knockout of B2M in CD4$^+$ T cells while co-cultured CD8$^+$ T cells remained unmodified. This establishes Cas9-VLP pseudotyping as a promising approach to specifically retarget Cas9 RNP-mediated genome editing to predetermined cell types within a mixed cell population without the unintended modification of bystander cells.

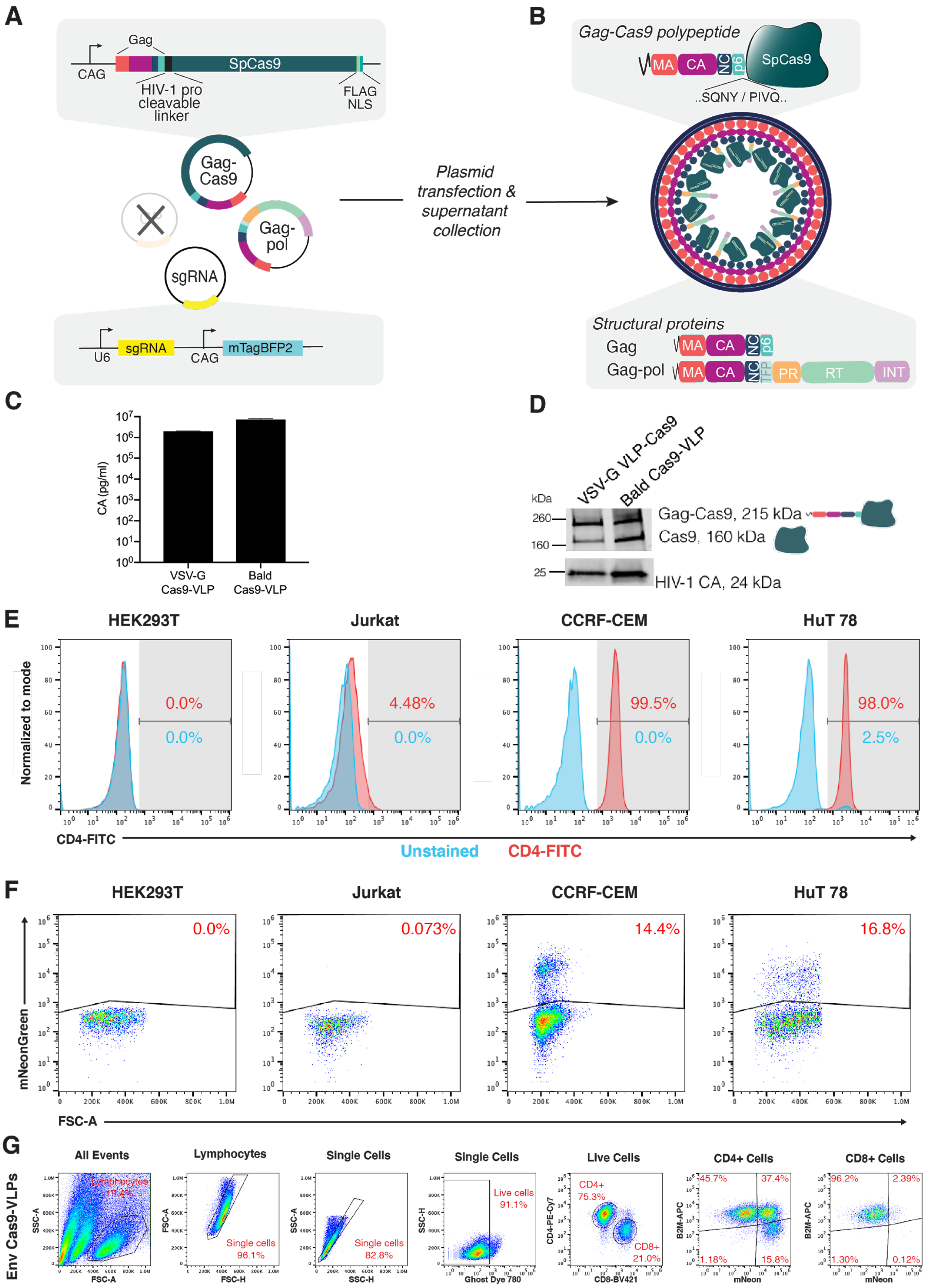

A
Gag
CAG
SpCas9
HIV-1 pro cleavable linker
FLAG NLS
Gag-Cas9
Gag-pol
sgRNA
U6
sgRNA
CAG
mTagBFP2
Plasmid transfection & supernatant collection
B
Gag-Cas9 polypeptide
MA
CA
NC
p6
SpCas9
..SQNY / PIVQ..
Structural proteins
Gag
MA CA NC p6
Gag-pol
MA CA NC TFP PR RT INT
C
CA (pg/ml)
10^0
10^1
10^2
10^3
10^4
10^5
10^6
10^7
VSV-G Cas9-VLP
Bald Cas9-VLP
D
VSV-G VLP-Cas9
Bald Cas9-VLP
kDa
260
160
25
Gag-Cas9, 215 kDa
Cas9, 160 kDa
HIV-1 CA, 24 kDa
E
HEK293T
Jurkat
CCRF-CEM
HuT 78
Normalized to mode
0.0%
0.0%
4.48%
0.0%
99.5%
0.0%
98.0%
2.5%
CD4-FITC
Unstained
CD4-FITC
F
HEK293T
Jurkat
CCRF-CEM
HuT 78
mNeonGreen
0.0%
0.073%
14.4%
16.8%
FSC-A
G
Env Cas9-VLPs
All Events
Lymphocytes 19.4%
SSC-A
FSC-A
Lymphocytes
Single cells 96.1%
FSC-A
FSC-H
Single Cells
Single cells 82.8%
SSC-A
SSC-H
Single Cells
Live cells 91.1%
SSC-H
Ghost Dye 780
Live Cells
CD4+ 75.3%
CD8+ 21.0%
CD4-PE-Cy7
CD8-BV421
CD4+ Cells
45.7%
37.4%
1.18%
15.8%
B2M-APC
mNeon
CD8+ Cells
96.2%
2.39%
1.30%
0.12%
B2M-APC
mNeon

Figure 4.12: Characterization of bald and HIV-1 Env pseudotyped Cas9-VLPs. (A) Production of “bald” Cas-VLPs. Schematic of plasmids used for the production of bald Cas9-VLPs that lack a glycoprotein. **(B)** Schematic of an immature, pre-proteolytically processed Cas9-VLP produced through transient transfection. **(C)** Quantification of Cas9-VLP production by CA ELISA. Amount of CA produced per transfected p100 dish is shown. **(D)** Western blot of Cas9-VLP content. An anti-Flag antibody was used for Cas9 detection and an anti-HIV-1 capsid (CA) antibody was used to detect Cas9-VLP production. **(E)** HIV-1 Env Cas9-VLPs are specific for CD4+ cells. Cell surface expression of CD4 in HEK293T, Jurkat, CCRF-CEM, and HuT 78 cell lines. **(F)** Transduction of Cas9-VLPs pseudotyped with the HIV-1 Envelope correlates with cellular CD4 expression. **(G)** Representative flow cytometry gating strategy to assess the cell-type specificity of B2M knockout by Env Cas9-VLPs within a mixed population of primary human T cells.

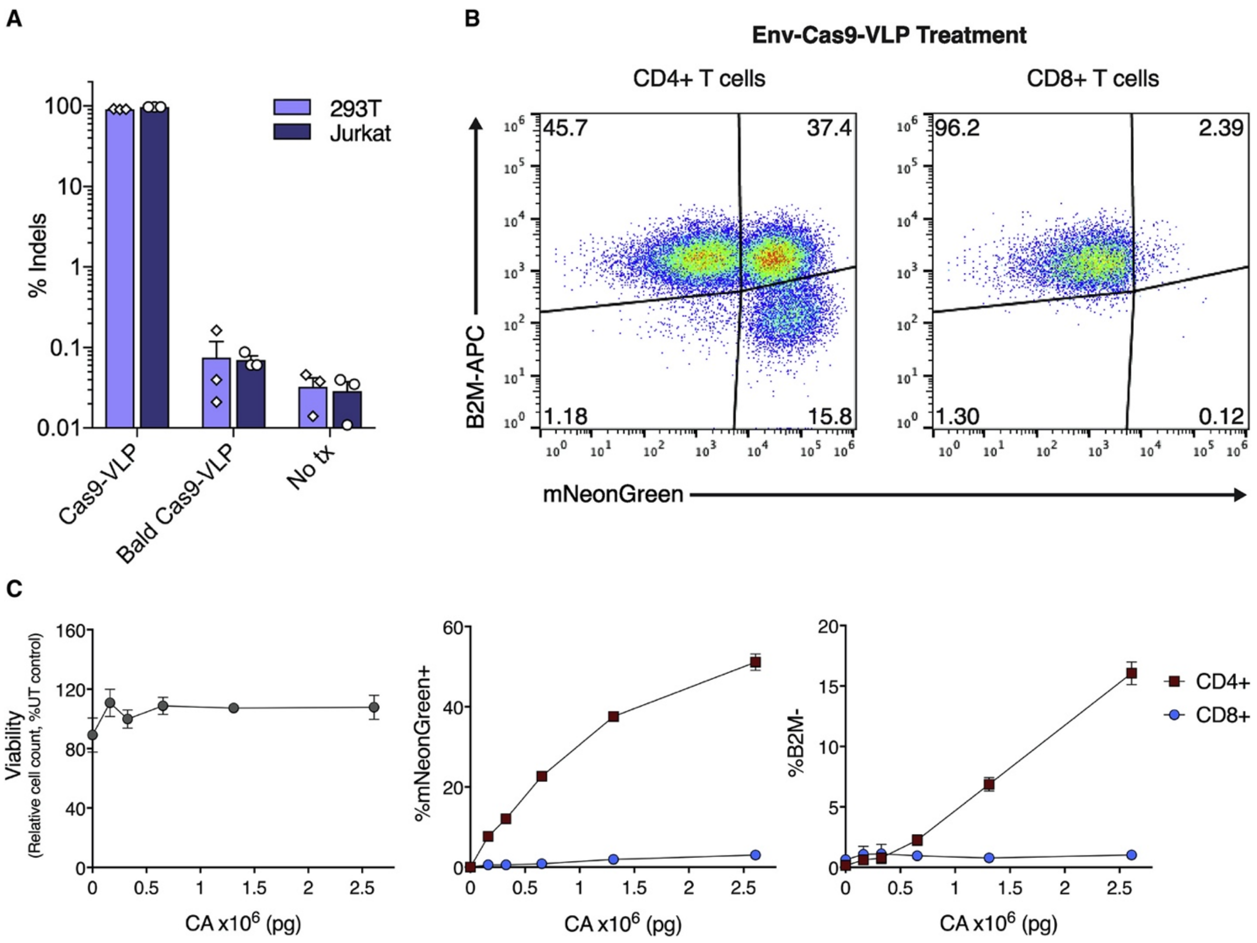


Figure 4.13: Cas9-VLP-mediated genome editing can be specifically targeted to primary human CD4+ T cells via HIV-1 Env glycoprotein pseudotyping. (A) A displayed viral glycoprotein is essential for Cas9-VLPs to mediate robust genome editing. 293T and Jurkat cells were treated with Cas9-VLPs pseudotyped with the VSV-G glycoprotein (Cas9-VLP), without the VSV-G glycoprotein (Bald Cas9-VLP), or left untreated (No tx). Indels at the *B2M* locus were quantified 3 days post treatment by next-generation amplicon sequencing, n = 3. **(B)** Cas9-VLPs pseudotyped with the HIV-1 envelope glycoprotein (HIV-1 Env Cas9-VLPs), and targeting *B2M*, were used to treat primary human T cells (a mixture of CD4+ and CD8+ cells). **(C)** Viability, transduction (mNeonGreen) and B2M knockout was assessed for both CD4+ (red squares) and CD8+ (blue circles) T cell subpopulations 7 days post treatment with HIV-1 Env-Cas9-VLPs. n = 2 biological replicates from independent donors were used (**B**, **C**) and representative flow cytometry plots are shown for one donor (**B**). Error bars indicate standard error of the mean.

4.4 Discussion

The therapeutic translation of genome editing requires the safe and effective delivery of CRISPR-Cas9 genome editing tools to therapeutically relevant cell types, either *ex vivo* or *in vivo*. Established viral delivery strategies generally result in the expression of genome editing tools for the lifetime of the cell, thereby increasing the risk of off-target genomic damage, malignant transformation, and potentially invoking adaptive immune responses against the edited cells *in vivo*. Most strategies for the non-viral delivery of genome editing tools rely upon electroporation, thus limiting therapeutic applications of genome editing to cells that can be manipulated outside the body. We sought to couple the delivery efficiency and cell type specificity of a virus with the transient genome editing activity of preassembled Cas9 RNPs. Such a delivery tool would enable laboratories to perform Cas9 RNP-mediated genome editing, targetable to any cell type susceptible to glycoprotein-mediated viral transduction, without the need for an electroporator, and amenable to both *ex vivo* and *in vivo* applications.

In this study, we engineered Cas9-VLPs as a delivery vehicle for Cas9 RNP complexes, with or without the co-delivery of a lentiviral genome for permanent transgene expression in treated cells. We demonstrate that Cas9-VLPs are a modular delivery system in which the genome editing efficiency, ability to codeliver a lentiviral-encoded transgene, and cellular tropism are programmable elements. Retroviral gag, integrase, Vpx, and Vpr have previously been engineered to direct particle packaging and delivery of enzymatic and reporter proteins, as well as I-SceI, zinc finger, and TAL effector nucleases[306–313]. Here, we directly fused Cas9 protein to the Gag structural protein to promote Cas9 incorporation during VLP assembly, an approach that has been successful at promoting the packaging of Cas9 protein in other engineered retroviruses. Specifically, lentiviral VLPs packaging a Cas9-N-terminal Gag fusion induced modest genome editing (14%–28%) but required guide RNA expression in target cells[166]. In contrast, murine leukemia virus-like particles (MLV VLPs) have also been engineered by fusing Cas9 to the C terminus of MLV Gag[169]. This protein fusion orientation allowed guide RNA to associate with fused Cas9 within the MLV VLPs, thereby allowing for the delivery of pre-formed Cas9 RNP complexes capable of mediating robust levels of genome editing in target cells. Recently it has been demonstrated that fusing Cas9 to the lentiviral accessory protein Vpr also promotes Cas9 packaging in budding particles[314]. Vpr-Cas9-containing lentivirus was highly effective at mediating genome editing in immortalized cell lines but only modestly effective (2.7%-15% indels) at mediating genome editing in primary human $CD4^+$ T cells. The reduced efficiency of genome editing in T cells may be due to Cas9-Vpr being packaged within the lentiviral core. While this intravirion localization may promote the delivery of Cas9 RNPs directly to the nuclear pore complex, antiviral restriction factors expressed by T cells may inhibit viral uncoating and, concomitantly, Cas9 RNP access to the nucleus. It has also been shown that Cas9 RNPs can be packaged into lentivirus by encoding RNA aptamers in the sgRNA tetraloop and directly fusing cognate aptamer binding proteins to the C terminus of Gag[167]. By expressing Cas9 during VLP production, aptamer-sgRNA-Cas9 complexes were incorporated, a strategy that ensures that all particle-packaged Cas9 protein is bound by sgRNA. Future studies will be needed to directly compare the effectiveness of viral engineering strategies, as

VLPs are an emerging strategy for coupling the cell-targeting and cell-fusion capabilities of enveloped viruses to the transient delivery of CRISPR-Cas9 tools.

By producing Cas9-VLPs simultaneously packaging Cas9 RNPs and a lentiviral-encoded CAR, we demonstrate a streamlined strategy for mediating gene knockout (of either the therapeutically relevant *B2M* or *TRAC* genes) and simultaneous CAR transgene integration for the production of genetically modified CAR T cells. Recently, lentiviral transduction combined with multiplexed CRISPR-Cas9 genome editing was used for the first time to treat three patients with refractory cancer in a phase I study[101]. Production of the infused T cell product required Cas9 RNP electroporation to mediate genetic knockouts, followed by cellular expansion and subsequent lentiviral transduction to introduce the NY-ESO-1 T cell receptor transgene. This multistep process significantly increases the complexity of clinical-grade manufacturing and the mixture of cellular products obtained highlights the challenge of generating a consistent outcome when combining multiple independent genome modifications. In contrast, by combining the genome editing and transduction capabilities into a single particle, Cas9-VLPs can couple lentiviral genome integration and Cas9-mediated knockout into one high-efficiency genome editing step allowing for coordinated transgene addition and endogenous gene knockout. This streamlined approach simplifies manufacturing, reduces the requirements for good manufacturing practice reagents and equipment, and may improve the consistency of the final therapeutic product.

Lastly, we leverage viral pseudotyping to target Cas9 RNP-mediated genome editing activity to a specific cell type within a mixed-cell population. By pseudotyping Cas9-VLPs with the HIV-1 viral glycoprotein Env, it was possible to exclusively direct Cas9 RNP genome editing to CD4$^+$ T cells, while leaving bystander CD8$^+$ T cells unmodified. Strategies for mediating cell-type-specific genome editing with Cas9 RNPs remain limited and Cas9-VLPs offer an approach for linking a payload of pre-formed Cas9 RNP complexes to the tropism of a viral glycoprotein. Our group has previously demonstrated that directly modifying the Cas9 RNP with asialoglycoprotein receptor ligands promotes Cas9 RNP uptake into hepatic cell types *in vitro*[315]. In addition, work has also been done to direct Cas9 RNP-loaded nanoparticles to specific cell types and tissues via targeting moieties. Decorating nanoparticles with the targeting ligand all-trans retinoic acid promotes genome editing in the retinal pigment epithelium of the eye[139], and directly altering the nanoparticle lipid composition could target genome editing activity to either the liver or lung *in vivo*[162]. Pseudotyping with viral glycoproteins is a strategy that has been successful for re-targeting retroviral/lentiviral transgene delivery to predetermined cell types; Cas9-VLPs can leverage this well-established field to achieve cell-type-specific delivery of Cas9 RNPs. Future Cas9-VLP studies will investigate additional viral glycoproteins and other targeting strategies for mediating specific genome editing of additional cell types, introduction of transgenes by homology-directed repair, and whether the targeting of CD4$^+$ T cells by Env-pseudotyped Cas9-VLPs will promote the generation of CAR T cells *in vivo*.

4.5 Materials and methods

4.5.1 Culture of human cell lines

Lenti-X, 293T, A549, CCRF-CEM, HuT 78 and Jurkat cell lines were obtained from the UC Berkeley Cell Culture Facility. All cells were cultured with 10% fetal bovine sera (VWR) and 100 U/mL penicillin-streptomycin (GIBCO). Lenti-X, 293T, and A549 cells were cultured in DMEM (Corning), Jurkat and CCRF-CEM cells were cultured in RPMI 1640 (Thermo Fisher) and 1 mM sodium pyruvate, while HuT 78 cells were cultured in IMEM (Thermo Fisher). Cell lines were routinely checked for mycoplasma using the MycoAlert mycoplasma detection kit (Lonza) according to the manufacturer's instructions.

4.5.2 Isolation and culture of human primary T cells

Primary adult blood cells were obtained from anonymous healthy human donors as a leukoreduction pack purchased from StemCell Technologies, Inc. or Allcells Inc, or as a Trima residual from Vitalant, under a protocol approved by the University of California, San Francisco Institutional Review Board (IRB). If needed, peripheral blood mononuclear cells were isolated by Ficoll-Paque (GE Healthcare) centrifugation. Bulk CD3+ T lymphocytes were then further isolated by magnetic negative selection using an EasySep magnetic Cell Isolation kit (STEMCELL, as per the manufacturer's instructions). 96-well flat bottom plates were primed for stimulation by incubating with anti-human CD3 (10 µg/mL) and anti-human CD28 (5 µg/mL) antibodies in PBS for 1 hour at 37°C prior to washing. Primary T cells were activated by plating at 250,000 cells/mL and culturing for one day in XVivo15 medium (Lonza) containing fetal bovine serum (5%), 2-mercaptoethanol (50 µM), N-acetyl L-cysteine (10 mM), IL-2 (300 U/mL), IL-7 (5 ng/mL), and IL-15 (5 ng/mL). Cas9-VLPs in RPMI 1640 were added to primary human T cells 24 hours later along with IL-2 (500 U/mL) and protamine sulfate (4 µg/mL). Media and growth factors were replaced as needed, approximately every 5–6 days. The number of unique primary human T cell donors used for each experiment is listed in Table 4.4.

4.5.3 Plasmid construction

The Gag-pol expression plasmid psPax2 was a gift from Didier Trono (Addgene plasmid #12260). pCMV-VSV-G was a gift from Bob Weinberg (Addgene plasmid #8454). Gag-Cas9 was constructed by amplifying Gag from psPax2 and Cas9 from pMJ920 (Addgene plasmid #42234). HIV-1 Env amino acid sequence was obtained from UniProt (P04578), human codon optimized (IDT), and ordered as a gBlock (IDT). In-Fusion (Takara Bio) cloning was used to clone Gag-Cas9 and Env into the pCAGGS expression vector. pCF221 (Addgene plasmid #121669) was modified to express mNeonGreen (Allele Biotechnology) or the α-CD19-4-1BBζ-P2A-mCherry (CAR-P2A-mCherry) construct[316,317] in place of mCherry and was used as the sgRNA-expressing lentiviral transfer plasmid. For generation of hybrid Cas9-VLPs, the guide RNA expression cassette was removed from the CAR-P2A-mCherry lentiviral plasmid via digestion with EcoRI and KpnI (NEB). The following primers (IDT) were phosphorylated, annealed, and ligated into the digested vector: 5′- cATCGATCTTAAGTCGCGACTCGAg and 5′ - aattcTCGAGTCGCGACTTAAGATCGATggtac. The U6-sgRNA CAG-mTagBFP2 expression plasmid used for traceless Cas9-VLP and CAR-Cas9 VLP production was a gift from Benjamin Oakes. Oligos encoding guide RNA spacers were ordered from IDT, phosphorylated, annealed and ligated into digested sgRNA expression vectors.

4.5.4 Cas9-VLP production

Cas9-VLPs were produced in mammalian cell culture by transient transfection of Lenti-X cells (Takara Bio). 3.5–4 million cells were seeded into 10 cm tissue culture dishes (Corning). The following day, cells were transfected with psPax2, Gag-Cas9, 1 µg pCMV-VSV-G or 0.2 µg HIV Env glycoprotein, and 10 µg of plasmid encoding the sgRNA-expression cassette (either transiently or in the context of a lentiviral transfer plasmid). Plasmids were diluted in Opti-MEM (GIBCO) and mixed with polyethylenimine (PEI, Polysciences Inc.) at a 3:1 PEI:plasmid ratio. Quantities of transfected Gag-Cas9 and psPax2 plasmid are listed in Figure 4.1C for VLP formulations A-F. A549 and Jurkat experiments used Cas9-VLP formulation D, unless indicated otherwise, and supernatant was harvested at 48 hours post transfection. Cas9-VLP experiments with primary human T cells used Cas9-VLP formulation B, where the Lenti-X media was replaced with Opti-MEM 6–18 hours post transfection. Cas9-VLP-containing Opti-MEM was collected at 48 and 96 hours post media change, with fresh Opti-MEM being added to the cells after 48 hours. Harvested supernatants were centrifuged at 1,500 rpm for 10 minutes and filtered through a 0.45 µM PES membrane bottle top filter (Thermo Fisher) or syringe filter (VWR). Cas9-VLPs were concentrated via ultracentrifugation by floating Cas9-VLP-containing supernatant on top of a cushioning buffer of 30% (w/v) sucrose in 100 mM NaCl, 10 mM Tris-HCl pH 7.5, 1 mM EDTA pH 8.0, at 25,000 rpm with a SW28 or SW41 Ti rotor (Beckman Coulter) for 2 hours at 4 °C in polypropylene tubes (Beckman Coulter). After ultracentrifugation, the Cas9-VLP pellet was resuspended in RPMI 1640 (GIBCO) or XVivo15 (Lonza) for treatment of primary T cells or Opti-MEM. Cas9-VLPs were either stored at 4 °C or frozen at -80 °C within an isopropanol-filled freezing container until use.

4.5.5 Cas9-VLP quantification

Western blots were performed to assess protein components of Cas9-VLPs. Cas9-VLPs were denatured by mixing with Laemmli buffer with 10% 2-mercaptoethanol and heating at 90 °C for 5 minutes. Samples were run on 4%-20% SDS-PAGE gels (Bio-Rad) prior to transfer onto a methanol soaked polyvinylidene difluoride (PVDF, Bio-Rad) membrane. PVDF membranes were blocked with 5% non-fat milk (Apex) in PBS (GIBCO) with 0.1% Tween (Sigma) (PBS-T) for one hour at room temperature (~22-25 °C). The solution was replaced with 1% non-fat milk in PBS-T and a 1:5000 primary antibody dilution containing anti-FLAG (Sigma) or a 1:2000 dilution of anti-p24 (Abcam) antibodies prior to shaking at 4 °C overnight. The following day, the solution was replaced with 1% non-fat milk in PBS-T and a 1:5000 secondary antibody dilution containing IR680 or IR800 conjugated antibodies (LI-COR) and shaken for 1 hour. Western blot membranes were washed with PBS-T three times prior to imaging on a LI-COR OdysseyCLx.

Lenti-X p24 rapid titer kits (Takara Bio) were used to quantify the titer of Cas9-VLPs after concentration. Cas9-VLPs were diluted 1:1,000-100,000 and the ELISA was performed according to the manufacturer's directions. Absorbance was measured at 450 nm on a BioTek plate reader. Cas9-VLP p24 content was calculated by comparison to serial dilution of a p24 standard (Takara Bio). To calculate transducing units per mL (TU/mL), Cas9-VLP preps were serially diluted and used to treat 15k Jurkats or 25k primary T cells in 96-well u-bottom plates. The percent of cells transduced (mNeonGreen+) was quantified at 6–7 days post treatment using an Attune NxT flow cytometer with a 96-well autosampler (Thermo Fisher Scientific) and titer was quantified as TU/mL = (number of cells transduced × percent mNeonGreen+) / (virus treatment volume). Wells where Cas9-VLP transduction was <25% were used for titer calculation. MOI was plotted against indels and a sigmoidal four parameter logistic fit was applied to each dataset to interpolate the MOI at which 50% indels would be expected, using a 95% confidence interval.

4.5.6 Cas9-VLP homology-directed repair

Cas9-VLPs targeting BFP were produced as previously described (see Table 4.1 for guide sequence). Cas9-VLPs were mixed with a single-stranded DNA template (IDT, Table 4.3) in either DPBS (Thermo Fisher Scientific), Opti-MEM (Thermo Fisher Scientific), or SE/SF/SG buffer (Lonza). Unless otherwise noted, SE buffer (Lonza) with pulse code CM-150 was utilized. The mixture was electroporated using a 4D-nucleofector (Lonza) before immediately adding to 293T cells stably expressing a BFP-to-GFP reporter (Addgene plasmid #71825). A three nucleotide conversion within the BFP gene results in GFP expression. Cells were analyzed for loss of BFP (non-homologous end joining) and gain of GFP (homology-directed repair) expression after 5–7 days on a Attune NxT flow cytometer with a 96-well autosampler (Thermo Fisher Scientific).

4.5.7 Targeted integration analysis

15k 293T cells treated with B2M-targeting or non-targeting Cas9-VLPs and DNA was isolated 3 days post treatment by resuspending in Quick Extract (Lucigen) and heating at 65 °C for 20 minutes followed by 95 °C for 20 minutes before storing at −20 °C. A nested PCR approach using PrimeStar GXL DNA polymerase (Takara Bio) was used to detect integration of the lentiviral genome into the *B2M* genomic site targeted by Cas9. For PCR analysis of lentiviral integration, the *B2M* targeted region was first amplified using nested primer set #1 and cleaned up (NucleoSpin Gel and PCR Clean-Up kit, Takara Bio) followed by amplification with primer sets a-g (Table 4.2). For MiSeq next generation sequencing analysis of targeted integration, the *B2M* targeted region was first amplified with nested primer set #2 and cleaned up (SPRI beads, UC Berkeley Sequencing Core) followed by amplification with primer sets to detect both integration orientations (primer pairs NGS Fwd and NGS Rev, Table 4.2). Pair-end reads were merged, trimmed, and aligned to the expected sequence of lentiviral insertion into the expected Cas9 target site in the *B2M* gene (Geneious).

4.5.8 RNP nucleofection

Cas9 RNPs were formed as previously described[125] at a 1:2 molar ratio between Cas9-NLS (UC Berkeley QB3 MacroLab) and annealed crRNA and tracrRNA (Horizon Discovery) in IDT duplex buffer with a polyglutamic acid electroporation enhancer, aliquoted, and stored frozen at -80 °C until use. Cas9 RNPs (50 pmol) were electroporated into primary human T cells using a 96-well format 4D-nucleofector (Lonza) with the P3 buffer and the EH-115 pulse code. Immediately after electroporation cells were rescued by adding growth media and incubating for 20 minutes prior to diluting to 0.5 to 1e6 cells/mL for culturing.

4.5.9 Flow cytometry

All flow cytometry was performed on an Attune NxT flow cytometer with a 96-well autosampler (Thermo Fisher Scientific). Cells were resuspended in FACS buffer (1%–2% BSA in PBS) and stained with the surface marker-targeting antibodies: B2M-FITC (Biolegend), B2M-PE (Biolegend), B2M-APC (Biolegend), CD4-FITC (Biolegend), CD8-PeCy7 (BD Biosciences), and TCRa/b-BV421 (Biolegend) and live/dead stains Ghost Dye red 780 (Tonbo) or Ghost Dye violet 450 (Tonbo), prior to analyzing. All analysis was done using the FlowJo v10 software. The gating strategy for flow cytometry can be seen in Figure 4.8; Figure 4.9; Figure 4.12.

4.5.10 Cytotoxicity assay

Nalm-6 target cells were labeled using CellTrace Violet Cell Proliferation Kit (Thermo Fisher Scientific) according to the supplier's information. T cells were co-cultured with labeled target cells at various Effector:Target ratios for 16–24 hours. The percent of transduced cells were normalized by adding untransduced T cells. Absolute count of remaining living target cells was analyzed and percent killing was calculated by comparing to control wells (target cells only). Measurement was performed on an Attune NxT Flow Cytometer (Thermo Fisher Scientific).

4.5.11 Intracellular cytokine and activation assay

Cells were stimulated with Nalm-6 target cells at an E:T ratio of 1:1. Transduction rates were normalized by adding untransduced T cells. 24 hours later, eBioscience Brefeldin A Solution (1000X) was added and incubated for 4 hours at 37 °C. Cells were stained with extracellular antibodies eBioscience Fixable Viability Dye eFluor 780 (Thermo Fisher Scientific), CD25 PE-Cy7 (BD), CD69 PerCP (BioLegend), 4-1BB BV711 (BioLegend) and intracellular antibodies TNF-a Pacific Blue (BioLegend), IL-2 APC (BD) and IFN-g FITC (BioLegend) using the FIX & PERM Cell Fixation & Cell Permeabilization Kit (Thermo Fisher Scientific). CAR samples were gated on mCherry+ cells. Measurement was performed on an Attune NxT Flow Cytometer (Thermo Fisher Scientific).

4.5.12 Amplicon sequencing

Genome editing was determined either by Sanger sequencing or next-generation sequencing; in both cases, the presence of insertions or deletions around the Cas9-targeted sequence was used to determine genome editing efficiency. Cells were pelleted and resuspended in QuickExtract (Lucigen) and heated at 65 °C for 20 minutes followed by 95 °C for 20 minutes before storing at -80 °C. An amplicon containing the target sequence was amplified via PCR with Q5 polymerase (NEB) or PrimeStar GXL DNA polymerase (Takara Bio) and the resulting sample was cleaned with magnetic SPRI beads (UC Berkeley Sequencing Core). PCR amplicons were analyzed via Sanger sequencing (UC Berkeley Sequencing Core) and the resulting traces were deconvolved with Synthego's Inference of CRISPR Edits (ICE) program (https://ice.synthego.com). NGS sequencing was prepared similarly, but with PCR primers containing Illumina adaptor sequences. PCR amplicons were analyzed on an Illumina MiSeq by QB3 Genomics at UC Berkeley. Paired-end NGS reads were analyzed for indels with CRISPResso2 (http://crispresso.pinellolab.partners.org/login).

4.5.13 Statistical analysis

Statistical analysis was performed in Prism v7, v8, and v9. Statistical details for all experiments, including value and definition of n, error bars, and significance thresholds can be found in the Figure Legends.

4.6 Accession codes

This work did not generate any data with accession codes.

4.7 Acknowledgments

We thank Yannick D. Muller for providing the α-CD19-4-1BBζ-P2A-mCherry plasmid, as well as Stacia Wyman for bioinformatic analysis of NGS amplicons. We thank all of the members of the Doudna and Marson laboratories for their thoughtful input and technical assistance. This research was supported by the Centers for Excellence in Genomic Science of the NIH (award no. RM1HG009490) and the Somatic Cell Genome Editing Program of the Common Fund of the NIH (award no. U01AI142817-02). J.R.H. is a Fellow of The Jane Coffin Childs Memorial Fund for Medical Research. C.A.T. is supported by Campus Executive Grants 2101705 and 1655264 through Sandia National Laboratories. D.N.N. is supported by NIH grants L40AI140341 and K08AI153767 and the CIRM Alpha Stem Cell Clinic Fellowship. B.R.S. is supported by the UCSF Herbert Perkins Cellular Therapy and Transfusion Medicine Fellowship. F.B. was supported by the Care-for-Rare Foundation and the German Research Foundation (DFG). A.M. holds a Career Award for Medical Scientists from the Burroughs Wellcome Fund, is an investigator at the Chan Zuckerberg Biohub, and is a recipient of The Cancer Research Institute (CRI) Lloyd J. Old STAR grant. The Marson lab has received funds from the Innovative Genomics Institute (IGI), the Simons Foundation, and the Parker Institute for Cancer Immunotherapy (PICI). J.A.D. is an investigator of the Howard Hughes Medical Institute (HHMI).

4.8 Author contributions

J.R.H. conceived the idea of engineering Cas9-VLPs. J.R.H., C.A.T., E.R.M., and C.R.S.E. designed and performed experiments optimizing Cas9-VLPs, with input from J.A.D. J.R.H. and C.A.T. conceived the ideas of pseudotyped Cas9-VLPs and Cas9-VLPs to generate CAR T cells, with input from D.N.N., B.R.S., A.M., and J.A.D. D.N.N., B.R.S., and D.C. designed and performed the experiments in primary human T cells, and F.B. designed and performed the cytotoxicity assays, all with input from J.R.H., C.A.T., A.M., and J.A.D. J.R.H., C.A.T., and J.A.D. wrote the manuscript, with input from all of the authors.

4.9 Supplementary information

Target	Protospacer sequence	PAM
B2M	GAGTAGCGCGAGCACAGCTA	AGG
TRAC	AGAGTCTCTCAGCTGGTACA	CGG
BFP	GCTGAAGCACTGCACGCCAT	GGG
Control (tdTom298)	AAGTAAAACCTCTACAAATG	TGG
Control (non-targeting guide used for integration site analysis)	GTATTACTGATATTGGTGGG	

Table 4.1: Protospacer sequences for mammalian genome editing

Target	Sequence
B2M_Sanger_F	TCACCCAGTCTAGTGCATGC
B2M_Sanger_R	GACGCTTATCGACGCCCTAA
TRAC_Sanger_F	CATCACTGGCATCTGGACTCCA
TRAC_Sanger_R	TGCTCTTGAAGTCCATAGACCTCA
B2M_NGS1_F	**GCTCTTCCGATCT**TGCGGGCCTTGTCCTGATTG
B2M_NGS1_R	**GCTCTTCCGATCT**AGATCCAGCCCTGGACTAGC
B2M_NGS2_F	**GCTCTTCCGATCT**AAGCTGACAGCATTCGGGC
B2M_NGS2_R	**GCTCTTCCGATCT**GAAGTCACGGAGCGAGAGAG
Integration_a_F	**GCTCTTCCGATCT**TGCGGGCCTTGTCCTGATTG
Integration_a_R	GTTCGGGCGCCACTGCTAGA
Integration_b_F	TTAAGCCTCAATAAAGCTTGCC
Integration_b_R	**GCTCTTCCGATCT**AGATCCAGCCCTGGACTAGC
Integration_c_F	**GCTCTTCCGATCT**TGCGGGCCTTGTCCTGATTG
Integration_c_R	TTAAGCCTCAATAAAGCTTGCC
Integration_d_F	GTTCGGGCGCCACTGCTAGA
Integration_d_R	**GCTCTTCCGATCT**AGATCCAGCCCTGGACTAGC
Integration_e_F	**GCTCTTCCGATCT**TGCGGGCCTTGTCCTGATTG
Integration_e_R	**GCTCTTCCGATCT**AGATCCAGCCCTGGACTAGC
Integration_f_F	**GCTCTTCCGATCT**TGCGGGCCTTGTCCTGATTG

Integration_f_R	TACTGACGCTCTCGCACCCAT
Integration_g_F	TACTGACGCTCTCGCACCCAT
Integration_g_R	**GCTCTTCCGATCT**AGATCCAGCCCTGGACTAGC
Integration_NGS fwd_F	**GCTCTTCCGATCT**AAGCTGACAGCATTCGGGC
Integration_NGS fwd_R	**GCTCTTCCGATCT**GAGAGCTCCTCTGGTTTCCC
Integration_NGS rev_F	**GCTCTTCCGATCT**GAGAGCTCCTCTGGTTTCCC
Integration_NGS rev_R	**GCTCTTCCGATCT**GAAGTCACGGAGCGAGAGAG
Integration_Nested_1_F	TCACCCAGTCTAGTGCATGC
Integration_Nested_1_R	GACGCTTATCGACGCCCTAA
Integration_Nested_2_F	**GCTCTTCCGATCT**AGGTCCGAGCAGTTAACTGG
Integration_Nested_2_R	**GCTCTTCCGATCT**ACTTAGCGGGCGCCTAGA

Table 4.2: Genomic amplification and sequencing primers. Illumina adapter sequences used for library prep are in **bold**.

Target	Sequence
BFP_GFP_HDRT	GCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACGTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGA

Table 4.3: HDR template

Treatment	Figures	Number of Donors
Nucleofection	Fig. 4.13, Fig. 4.8	2 (Donors A-B)
VSV-G Cas9-VLP	Fig. 4.13, Fig. 4.8, Fig. 4.10	4 (Donors A-D)
Env Cas9-VLP	Fig. 4.13, Fig. 4.12	2 (Donors A-B)
B2M Cas9-VLP + TRAC Cas9-VLP	Fig. 4.13, Fig. 4.9	2 (Donors E-F)
B2M CAR-P2A-mCherry Cas9-VLP	Fig. 4.13, Fig. 4.9, Fig. 4.10	4 (Donors C-D, E-F)
TRAC CAR-P2A-mCherry Cas9-VLP	Fig. 4.13, Fig. 4.9	2 (Donors E-F)

Table 4.4: Primary human T cell donors

CHAPTER 5

Establishing a SARS-CoV-2 testing facility

A portion of the work presented in this chapter has been published previously as part of the following papers: IGI Testing Consortium. Blueprint for a pop-up SARS-CoV-2 testing lab. *Nature Biotechnology* **38**, 791-797, (2020).

5.1 Abstract

When the World Health Organization declared the 2019 coronavirus disease (COVD-19) a global pandemic in March of 2020, countries, states, and communities struggled to implement sufficient clinical diagnostic testing. In response to this dire need, we established a brand new, certified clinical testing lab to provide patient diagnoses for infection by SARS-CoV-2, the disease causing COVID-19, at the University of California, Berkeley. Here, we describe the organizational, safety, scientific, regulatory, and clinical challenges and solutions we encountered in the three weeks it took to build this clinical testing lab and the time following providing diagnoses to thousands of patients.

5.2 Introduction

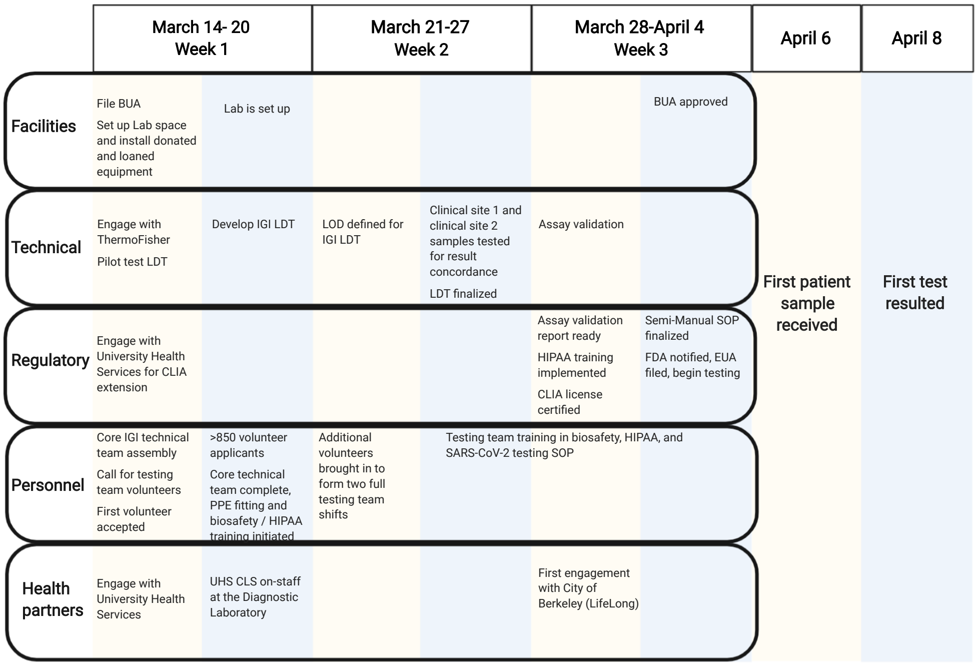
On 11 March 2020, the World Health Organization declared the 2019 coronavirus disease (COVID-19) a global pandemic[318]. As of 29 May, the virus that causes the disease, SARS-CoV-2, has infected over 5,813,000 people and killed more than 360,000 worldwide (https://coronavirus.jhu.edu/map.html). The virus continues to spread around the world, and at the time of writing there are no clinically validated medical interventions to prevent or cure COVID-19 (there are now several FDA-approved vaccines[149] for prevention as well as small molecule cures[319]). Public health measures in the United States and elsewhere focus on mitigating spread through diagnostic testing, self-isolation and shelter-in-place orders[320].

The presence of presymptomatic and mildly symptomatic individuals in the general population is a major driver in the accelerated and widespread outbreaks that have overwhelmed healthcare infrastructures worldwide, causing more deaths[320–323]. Extensive testing in countries such as Iceland, New Zealand, Germany and South Korea, among others, has proven an effective tool in controlling the spread of the disease[321–325].

At the start of our effort, on 14 March 2020, the turnaround time for testing for University of California (UC) Berkeley students through commercial labs exceeded seven days (UC Berkeley Tang Center, personal communication), and no rapid or surveillance testing was available to City of Berkeley first responders (City of Berkeley Fire Department Chief David Brannigan, personal communication) or to vulnerable populations in Berkeley, including those living in congregated settings and the unsheltered.

5.3 Results

To address the need for expanded testing capacity, the Innovative Genomics Institute (IGI) at UC Berkeley established a clinical testing laboratory for SARS-CoV-2 in three weeks (Figure 5.1). Timely setup presented formidable challenges, including navigating state and federal regulations, supply-chain and logistic obstacles, and challenges related to serving populations beyond UC Berkeley (Table 5.1). To tackle these hurdles, we partnered with UC Berkeley's University Health Services (UHS) and created specialized teams to execute the technical, operations, regulatory, human resources, data management, physician interface, sample collection and sample reporting processes for the IGI laboratory (Figure 5.2).



	March 14- 20 Week 1		March 21-27 Week 2		March 28-April 4 Week 3		April 6	April 8
Facilities	File BUA Set up Lab space and install donated and loaned equipment	Lab is set up				BUA approved	First patient sample received	First test resulted
Technical	Engage with ThermoFisher Pilot test LDT	Develop IGI LDT	LOD defined for IGI LDT	Clinical site 1 and clinical site 2 samples tested for result concordance LDT finalized	Assay validation			
Regulatory	Engage with University Health Services for CLIA extension				Assay validation report ready HIPAA training implemented CLIA license certified	Semi-Manual SOP finalized FDA notified, EUA filed, begin testing		
Personnel	Core IGI technical team assembly Call for testing team volunteers First volunteer accepted	>850 volunteer applicants Core technical team complete, PPE fitting and biosafety / HIPAA training initiated	Additional volunteers brought in to form two full testing team shifts	Testing team training in biosafety, HIPAA, and SARS-CoV-2 testing SOP				
Health partners	Engage with University Health Services	UHS CLS on-staff at the Diagnostic Laboratory			First engagement with City of Berkeley (LifeLong)			

Figure 5.1: Timeline of the IGI SARS-CoV-2 Laboratory Establishment. Establishing the IGI SARS-CoV-2 testing laboratory took approximately three weeks, from concept to receiving the first patient sample. The timeline for establishing the legislative, technical and operational requisites for the laboratory, from March 13 to April 4, are described in this figure. The IGI SARS-CoV-2 testing laboratory opened its doors on April 6th, and resulted its first patient sample on April 8th.

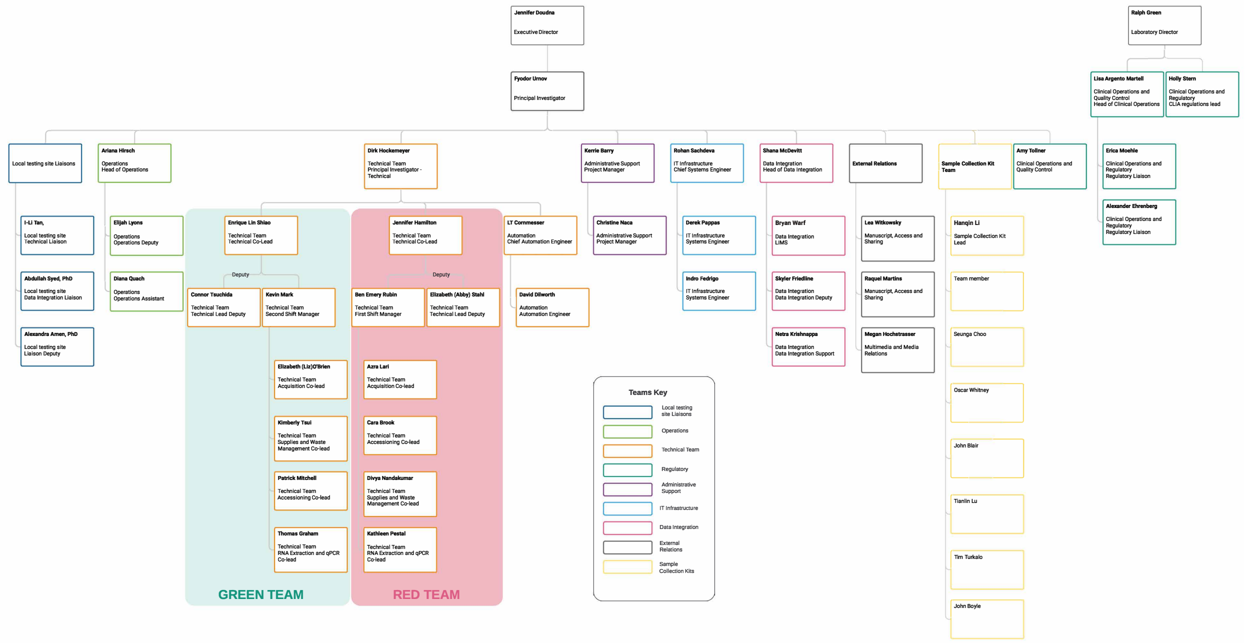


Figure 5.2: IGI testing facility team organizational chart. The IGI team divided into several specialized, lean, subteams to cover the different areas needed to establish our testing facility. The teams specialized

in operations, regulatory, technical, sample collection kits, IT infrastructure, data integration, administrative support, external relations, and liaising collaborative relationships with local testing facilities. The testing team worked in two shifts, noted in the figure as "Green Team" and "Red Team".

When we began, our campus did not have a clinical testing facility that would allow our testing lab volunteers to work at the level of biosafety required by our campus for SARS-CoV-2 diagnostics, and without a medical school with an affiliated medical center, our campus had no mechanism to provide medical services to patients from off campus. To serve populations beyond the campus, we established partnerships with community health centers and implemented an electronic portal compliant with Clinical Laboratory Improvement Amendments (CLIA) and Health Insurance Portability and Accountability Act (HIPAA) for requisitioning and providing results of tests. The portal integrates with our laboratory information management system (LIMS) (Figure 5.3).

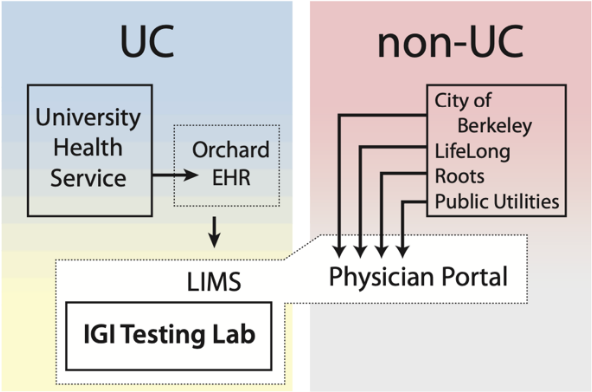


Figure 5.3: Interface between IGI testing facility LIMS system and UC and non- UC health partners. For communication between IGI LIMS and UC Berkeley UHS, patient identifying information and associated sample barcodes are exchanged in a fully integrated way between LIMS and the UHS Orchard electronic health record. For interactions with non-UC health partners the IGI built an interface physician portal application on our LIMS system. This portal receives test requisitions from non-UC physicians, and the test results from the LIMS, associated with the respective patient identifying information which are then communicated to the physician.

Three regulatory developments enabled our technical work. The first, California Governor Gavin Newsom’s 4 March Executive Order N-25-20, modified the requirements for clinical laboratory personnel running diagnostic tests for SARS-CoV-2 in a certified laboratory, allowing trained volunteer scientists to staff the operation. The second, the US Food and Drug Administration (FDA) 16 March “Policy for Diagnostic Tests for Coronavirus Disease-2019 during the Public Health Emergency”, simplified the authorization process for a SARS-CoV-2 test, enabling quick adoption of an existing authorized test kit. The third increased the speed and flexibility of state and federal licensing procedures for clinical laboratory facilities under the CLIA program.

The Centers for Medicare & Medicaid Services’ CLIA program regulates all US clinical laboratory testing, ensuring the accuracy and reliability of patient test results. Under the state’s temporarily relaxed regulatory requirements, we obtained CLIA certification for the IGI testing lab by extension of the existing CLIA license at UC Berkeley’s student health center. By partnering with the campus clinic, we were able to combine their medical and clinical laboratory expertise with the IGI's access to biosafety level (BSL)-2+ laboratory space, equipment, and technical expertise of IGI and UC Berkeley scientists. This combined expertise was critical for rapid SARS-CoV-2 test implementation with full regulatory oversight.

In addition to a facility license, other steps were required for regulatory compliance. First, the lab obtained a Biological Use Authorization from UC Berkeley's institutional biosafety committee (the Committee for Laboratory and Environmental Biosafety), which determined the level of biosafety and nature of personal protection equipment (PPE) necessary to receive patient samples and process them on site. Second, while Governor Newsom's executive order allowed people other than clinical laboratory staff (CLS) to develop and staff a SARS-CoV-2 testing facility, it preserved the essential requirement of oversight by licensed CLS and documentation of proper personnel training. Partnering with UHS also allowed us to work under the guidance of their clinical laboratory director and licensed CLS, and together we developed a rigorous training program and proficiency assessment for each member of the testing team that included biosafety and assay workflow training. Testing team leads (Figure 5.2) are trained in the entire test workflow, while other team members are trained and tested specifically on the task they perform. Documentation of this training, along with proof of education in a relevant field, was sent to the California Department of Public Health to satisfy personnel requirements under the executive order. Third, in keeping with CLIA requirements for continuous proficiency assessment, our technical leads were tested in a competency assessment by processing blinded samples provided by the American Proficiency Institute. Finally, given the existence of protected health information in the testing laboratory, HIPAA compliance was observed by establishing mandatory HIPAA training for all testing personnel. In total, the complete training, testing and PPE process takes an average of ten working days.

Additional precautionary measures were implemented to ensure the safety of our volunteer staff. Standard sample collection kits as recommended by the US Centers for Disease Control use COPAN tubes with universal transport medium (UTM), a buffer that stabilizes the virus before analysis. We developed and validated our own patient specimen collection kit, which uses a chaotropic agent (DNA/RNA Shield, Zymo) in place of UTM (Figure 5.4). This substitution not only preserves the integrity of the sample nucleic acid during transport, but also inactivates pathogens at the time of specimen collection. In this way, we minimize the potential for live virus to enter our facility. Additionally, as required by our Biological Use Authorization, our laboratory operates under BSL-2+, a higher standard of safety than the BSL-2 conditions typically employed with inactivated SARS-CoV-2. Stringent PPE requirements for all testing lab personnel include a disposable outer layer and N95 masks professionally fitted for each staff member. An added daily self-assessment ensures that symptomatic personnel are detected as early as possible to prevent transmission within the team.

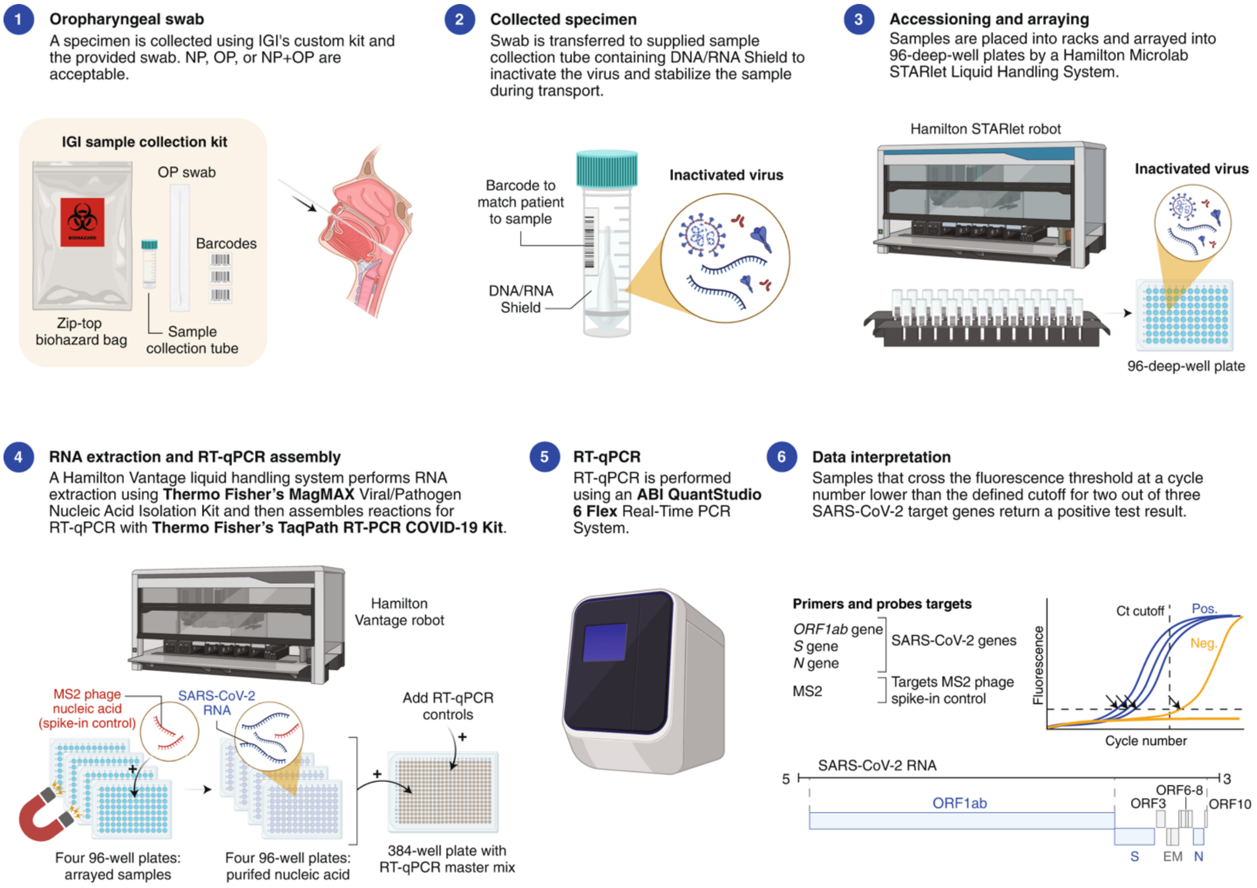
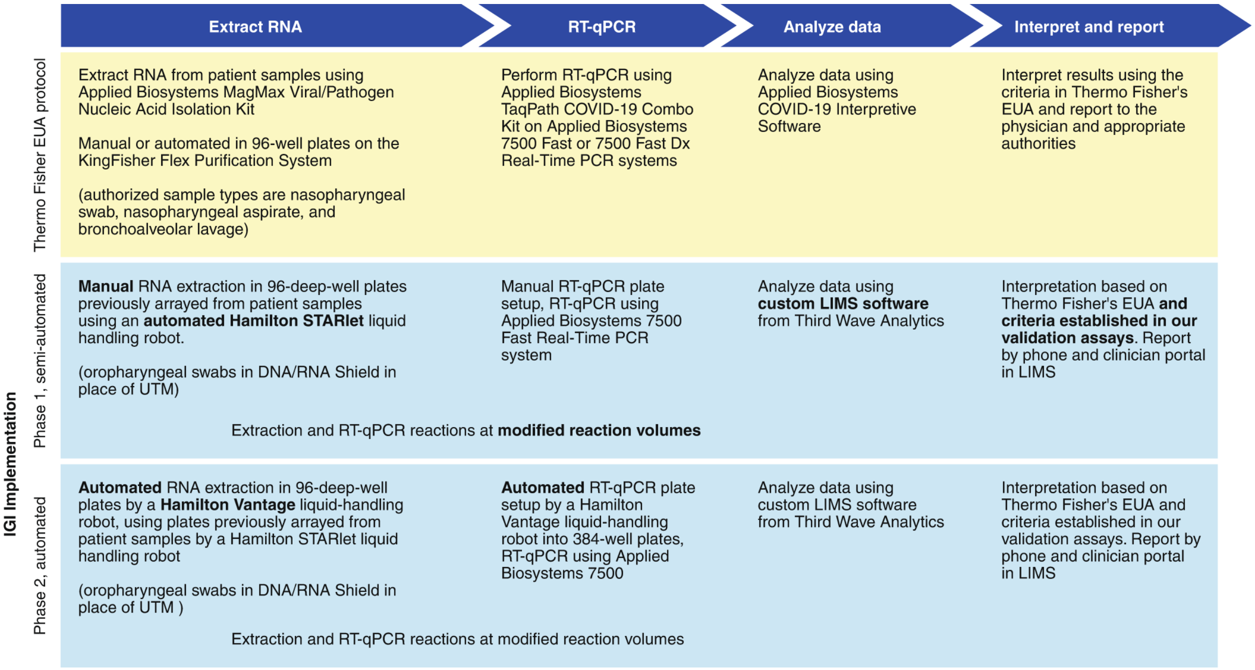


Figure 5.4: Overview of the IGI SARS-CoV-2 Testing Consortium assay. Diagram of the IGI's testing workflow and assay for SARS-CoV-2 using the automated method. Our test begins with a custom specimen collection kit for nasopharyngeal (NP) or oropharyngeal (OP) swabs (step 1). This kit uses alternative collection tubes and a chaotropic agent, DNA/RNA Shield, to inactivate the virus for transport to our facility (step 2). Barcodes are supplied to connect patient intake forms with specimens. Samples are transported to the IGI by courier where they are visually inspected for rejection criteria before being decontaminated in a biosafety cabinet and placed into trays for arraying. In step 3, a Hamilton STARlet liquid handler scans the sample barcodes, accessioning them into our LIMS, and arrays them into 96-well deep-well plates, recording the corresponding positions in the LIMS. Step 4 is conducted by a Hamilton Vantage liquid handler in our automated method and manually in our semi-automated method. The MS2 spike-in control is added and RNA extraction is performed using Thermo Fisher's MagMAX Viral/Pathogen Nucleic Acid Isolation Kit. Following extraction, the Hamilton Vantage performs RT-qPCR reaction setup using Thermo Fisher's TaqPath RT-PCR COVID-19 Kit by consolidating four 96-well plates into one 384-well plate preloaded with RT-qPCR master mix (this step performed manually in our semi-automated approach and using 96-well plates only). After the Hamilton Vantage adds positive and negative controls, the plate is moved to an ABI QuantStudio 6 Flex Real-Time PCR System for RNA detection. In our semi-automated approach, RT-qPCR is performed on an Applied Biosystems 7500 Fast Real-Time PCR system. Step 6 involves interpretation of the RT-qPCR data. Thermo Fisher's TaqPath RT-PCR COVID-19 Kit targets three SARS-CoV-2 genes shown in blue on the diagram of the SARS-CoV-2 genome (NCBI NC_045512.2). The open reading frame targeted by Thermo Fisher's kit, *ORF1ab*, encodes non-structural proteins for replication, whereas the spike (*S*) and nucleocapsid (*N*) genes encode two structural proteins. A diagram of qPCR amplification curves demonstrates the criteria by which sample data are interpreted; Ct, cycle threshold.

In response to the pandemic, the FDA enacted its authority to issue Emergency Use Authorizations (EUAs), reducing the time required to devise and implement a new diagnostic test. This shift enabled companies to develop new PCR-based diagnostics for SARS-CoV-2. Rather than starting from scratch, the IGI chose to adapt the test marketed by Thermo Fisher Scientific because of its reagent availability, equipment compatibility and robust test performance in our hands. However, to increase test throughput, reduce costs, improve staff safety and make use of existing equipment, we modified Thermo Fisher's EUA workflow. Since an EUA is granted to a specific protocol implemented on defined equipment, our modifications (reduced reagent use, different sample collection kit) made our implementation a laboratory developed test (LDT) requiring a new EUA (Figure 5.5). Nonetheless, our choice to adapt a test with an existing EUA allowed us to perform bridging studies, accelerating the path to our own EUA, further aided by technical support from Thermo Fisher during our effort.



	Extract RNA	RT-qPCR	Analyze data	Interpret and report
Thermo Fisher EUA protocol	Extract RNA from patient samples using Applied Biosystems MagMax Viral/Pathogen Nucleic Acid Isolation Kit Manual or automated in 96-well plates on the KingFisher Flex Purification System (authorized sample types are nasopharyngeal swab, nasopharyngeal aspirate, and bronchoalveolar lavage)	Perform RT-qPCR using Applied Biosystems TaqPath COVID-19 Combo Kit on Applied Biosystems 7500 Fast or 7500 Fast Dx Real-Time PCR systems	Analyze data using Applied Biosystems COVID-19 Interpretive Software	Interpret results using the criteria in Thermo Fisher's EUA and report to the physician and appropriate authorities
IGI Implementation Phase 1, semi-automated	**Manual** RNA extraction in 96-deep-well plates previously arrayed from patient samples using an **automated Hamilton STARlet** liquid handling robot. (oropharyngeal swabs in DNA/RNA Shield in place of UTM)	Manual RT-qPCR plate setup, RT-qPCR using Applied Biosystems 7500 Fast Real-Time PCR system	Analyze data using **custom LIMS software** from Third Wave Analytics	Interpretation based on Thermo Fisher's EUA **and criteria established in our validation assays**. Report by phone and clinician portal in LIMS
	Extraction and RT-qPCR reactions at **modified reaction volumes**			
IGI Implementation Phase 2, automated	**Automated** RNA extraction in 96-deep-well plates by a **Hamilton Vantage** liquid-handling robot, using plates previously arrayed from patient samples by a Hamilton STARlet liquid handling robot (oropharyngeal swabs in DNA/RNA Shield in place of UTM)	**Automated** RT-qPCR plate setup by a Hamilton Vantage liquid-handling robot into 384-well plates, RT-qPCR using Applied Biosystems 7500	Analyze data using custom LIMS software from Third Wave Analytics	Interpretation based on Thermo Fisher's EUA and criteria established in our validation assays. Report by phone and clinician portal in LIMS
	Extraction and RT-qPCR reactions at modified reaction volumes			

Figure 5.5: Workflow. Overview of Thermo Fisher's TaqPath COVID-19 Multiplex Diagnostic Solution[326,327] showing the steps as authorized in their EUA (yellow background) and IGI's implementation of this workflow (blue background). The IGI's SARS-CoV-2 testing lab was established in two phases. Phase 1 started with automated arraying of patient samples, followed by manual implementation of the Thermo Fisher kit. Phase 2 is a fully automated workflow. Bolded words indicate elements changed from the implementation above.

To lower the cost of our test and limit reagent use in the face of supply-chain shortages, we changed the quantity of reagents used for RNA extraction and reverse transcription and quantitative PCR (RT-qPCR) after verifying that the test maintained robust performance at half the reaction volume and with targeted changes in some reagent volumes. Additional modifications to Thermo Fisher's test included the development of a new sample collection kit with the goals of ensuring safety of our volunteer staff, as discussed above, and circumventing supply chain shortages and its deployment on liquid handlers for increased testing throughput (see below).

The Thermo Fisher kit uses primer–probe pairs targeting three SARS-CoV-2 genes: *ORF1ab*, the spike protein gene (*S*) and the nucleocapsid protein gene (*N*) (Figure 5.4). To return a positive result, the test must detect two of the three genes in a patient sample. Thermo Fisher's kit controls for RNA extraction and amplification by including an MS2 bacteriophage spike-in control together with the corresponding primer–probe pairs, which are added before RNA extraction and RT-qPCR, respectively.

With the goal of initiating diagnostic testing at the earliest possible time, we designed our research and development workstream to begin with a semi-automated approach with a capacity of 180 tests per day that could be implemented rapidly while we developed and validated a fully-automated assay to increase testing throughput to over 1,000 tests per day. Operating initially at reduced capacity also allowed the timely identification and resolution of process inefficiencies and technical issues, as well as the establishment of working relationships with non-university healthcare providers to meet the needs of a larger population outside our campus.

Although no specialized equipment is required to perform an RT-qPCR assay manually, minimizing the potential for human error is essential. To this effect, we deployed a liquid handler (Hamilton STARlet) to perform patient specimen consolidation into 96-well deep-well plates (semi-automated workflow) and a second liquid handler (Hamilton Vantage) to perform RNA extraction and generate plates ready for RT-qPCR in the fully automated workflow (Figures 5.4 and 5.5). Custom automation code for the Hamilton STARlet and Hamilton Vantage in the IGI setup are available on request (Figure 5.6). Early in our process we benefited from loaned laboratory equipment (qPCR machines, liquid handlers, biological safety cabinets, and extra cold storage to enable backup in case of equipment failure, as well as our expanding testing capacity) and purchased our own as the testing continued.

With the modifications implemented to the Thermo Fisher protocol, our LDT was validated within the CLIA framework, as described below, and our EUA was submitted to the FDA as a bridging study to the original EUA awarded to Thermo Fisher. Validating an LDT requires measuring specific Centers for Medicare & Medicaid Services and FDA metrics for analytical and clinical validity and meeting or exceeding benchmarks. For our semi-automated assay, these were (i) measuring the assay limit of detection (LOD); (ii) assessing clinical and analytical validity by running mock positive and negative samples at known concentrations, and (iii) performing our LDT on samples previously identified as positive and negative for SARS-CoV-2 provided by two local clinical diagnostic testing facilities.

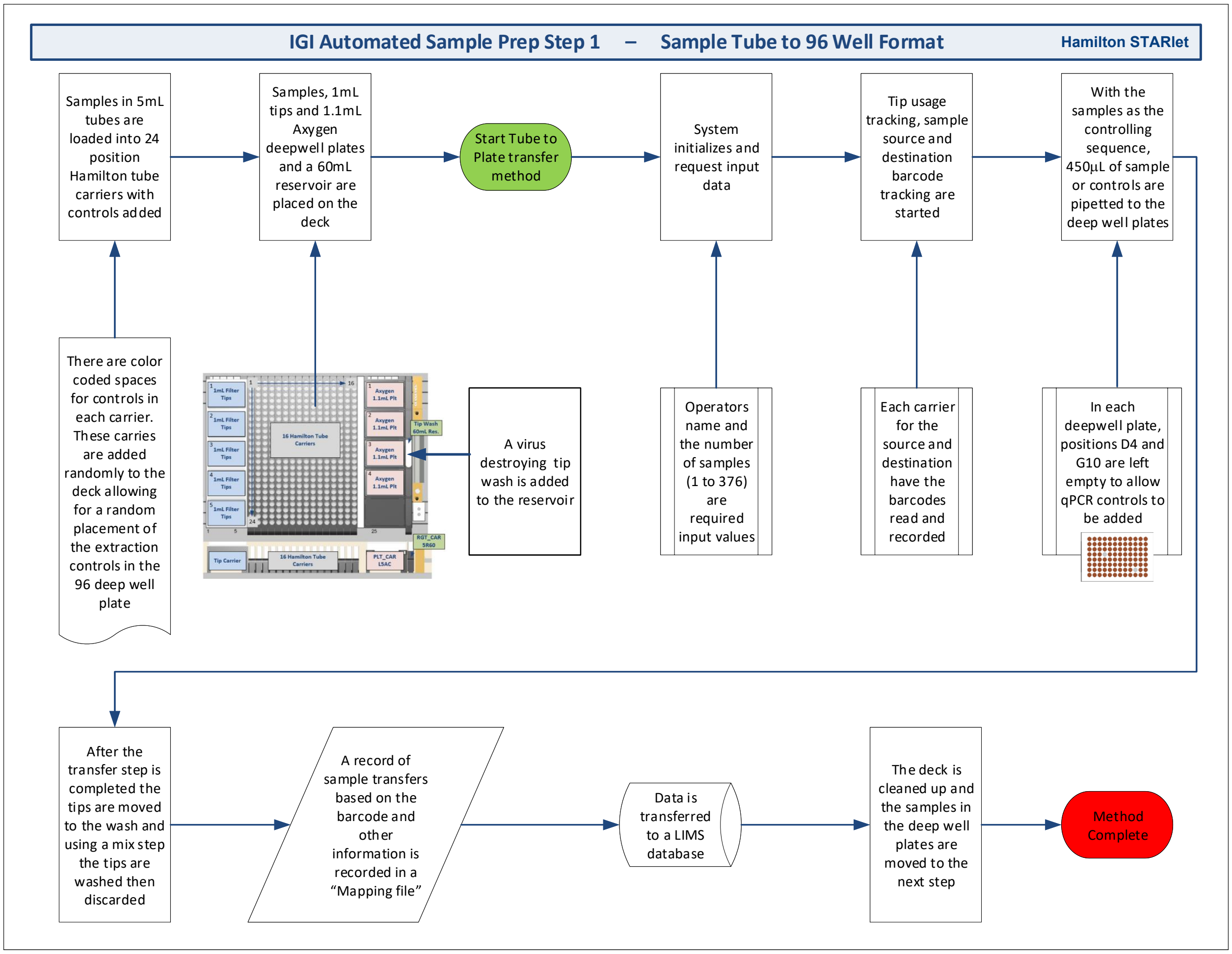
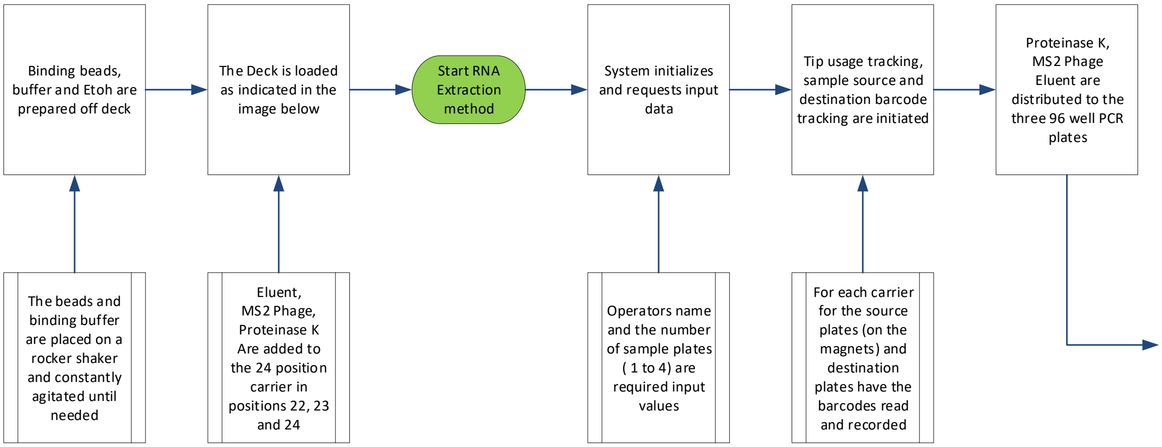


Figure 5.6: Hamilton Microlab STARlet and Hamilton Vantage automation process workflow. The following pages describe the automation process in the Hamilton STARlet and Hamilton Vantage in the IGI testing laboratory setup, from patient sample accessioning (STARlet), through RNA extraction and to sample arraying in RT-qPCR plate (Vantage).

IGI Automated Sample Prep Step 2 (pg. 1) – RNA Extraction

Hamilton Vantage



Binding beads, buffer and Etoh are prepared off deck

The Deck is loaded as indicated in the image below

Start RNA Extraction method

System initializes and requests input data

Tip usage tracking, sample source and destination barcode tracking are initiated

Proteinase K, MS2 Phage Eluent are distributed to the three 96 well PCR plates

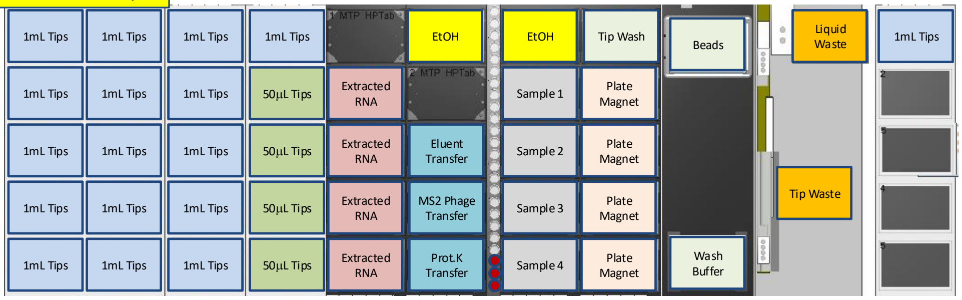
The beads and binding buffer are placed on a rocker shaker and constantly agitated until needed

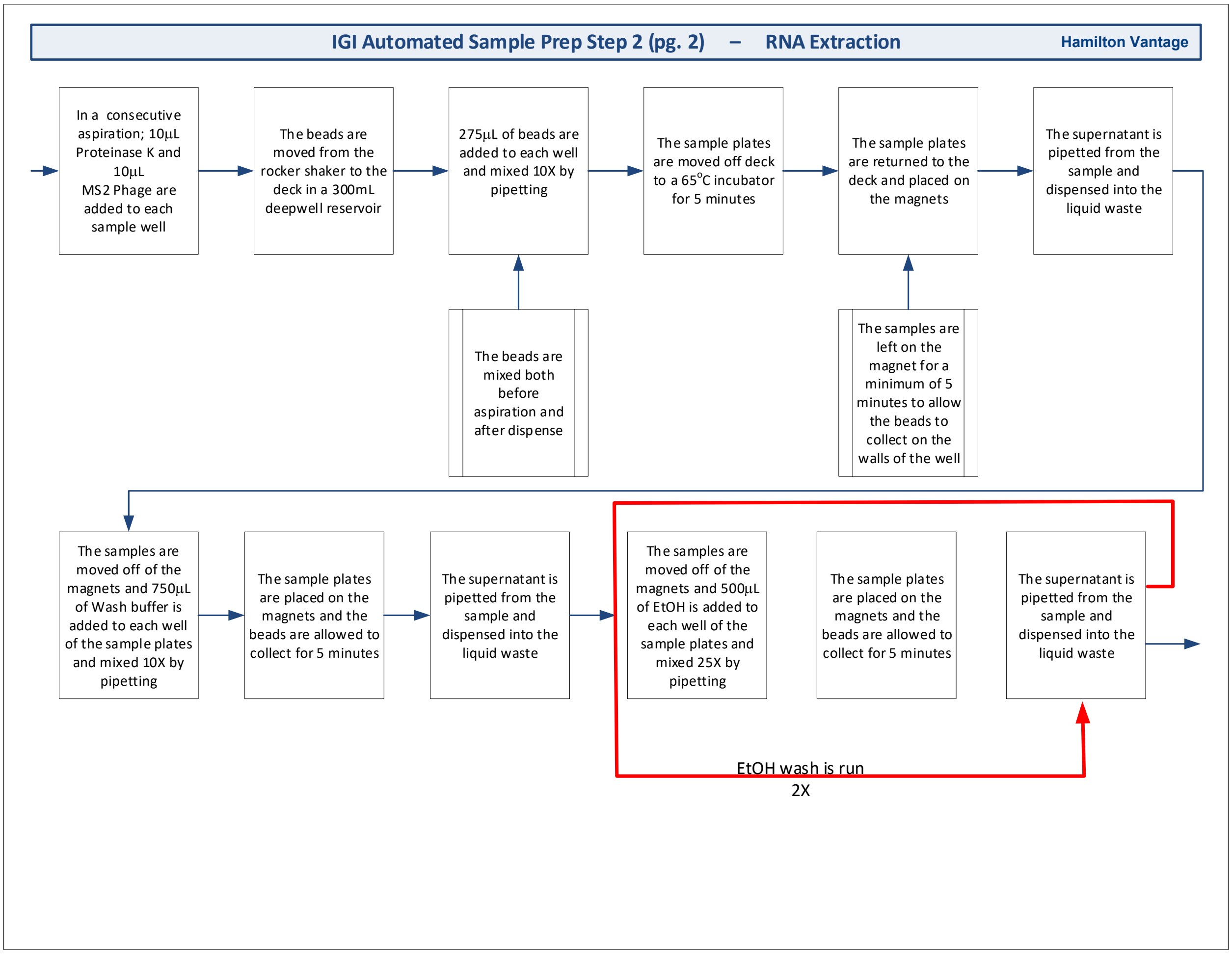
Eluent, MS2 Phage, Proteinase K Are added to the 24 position carrier in positions 22, 23 and 24

Operators name and the number of sample plates (1 to 4) are required input values

For each carrier for the source plates (on the magnets) and destination plates have the barcodes read and recorded

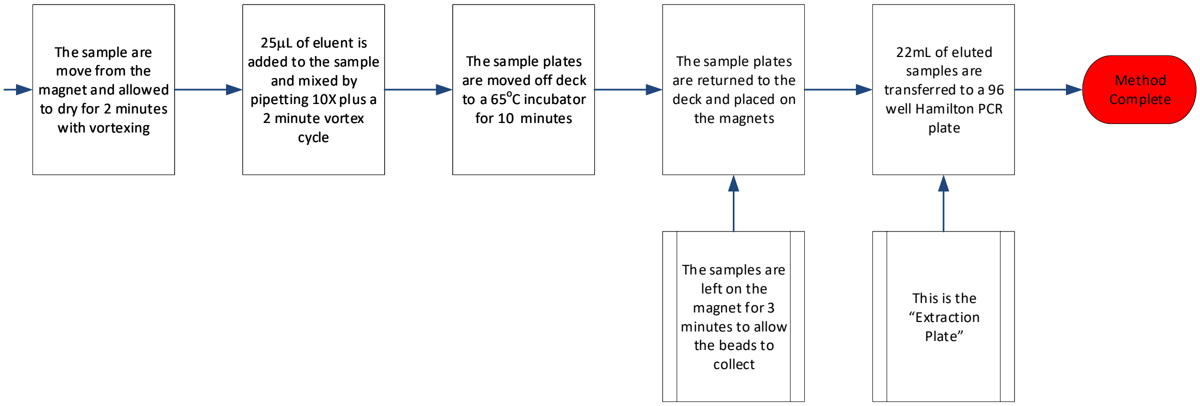
Twice the user will be prompted to reload the tips

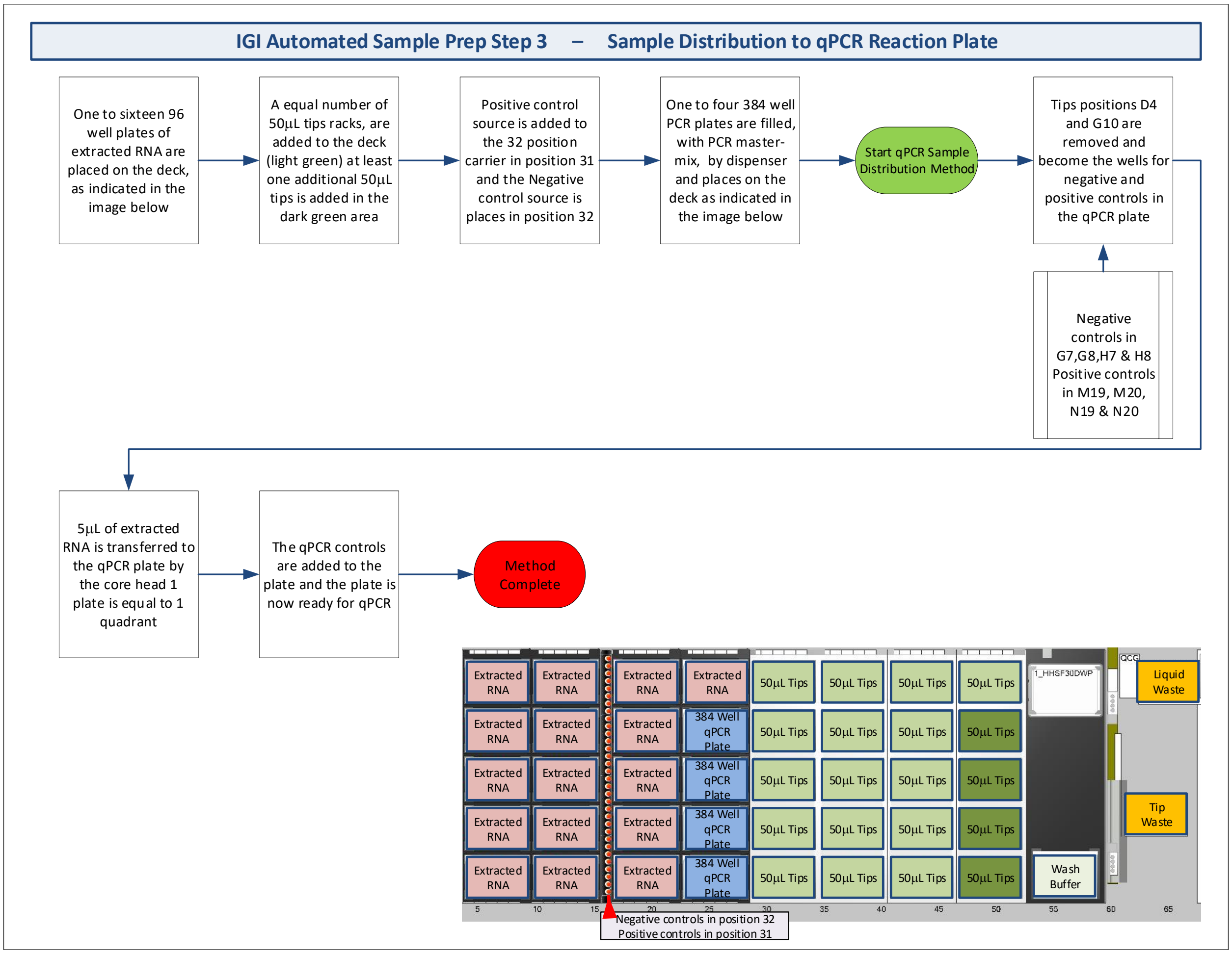



IGI Automated Sample Prep Step 2 (pg. 2) – RNA Extraction
Hamilton Vantage
In a consecutive aspiration; 10µL Proteinase K and 10µL MS2 Phage are added to each sample well
The beads are moved from the rocker shaker to the deck in a 300mL deepwell reservoir
275µL of beads are added to each well and mixed 10X by pipetting
The beads are mixed both before aspiration and after dispense
The sample plates are moved off deck to a 65°C incubator for 5 minutes
The sample plates are returned to the deck and placed on the magnets
The samples are left on the magnet for a minimum of 5 minutes to allow the beads to collect on the walls of the well
The supernatant is pipetted from the sample and dispensed into the liquid waste
The samples are moved off of the magnets and 750µL of Wash buffer is added to each well of the sample plates and mixed 10X by pipetting
The sample plates are placed on the magnets and the beads are allowed to collect for 5 minutes
The supernatant is pipetted from the sample and dispensed into the liquid waste
The samples are moved off of the magnets and 500µL of EtOH is added to each well of the sample plates and mixed 25X by pipetting
The sample plates are placed on the magnets and the beads are allowed to collect for 5 minutes
The supernatant is pipetted from the sample and dispensed into the liquid waste
EtOH wash is run 2X

IGI Automated Sample Prep Step 2 (pg. 3) – RNA Extraction

Hamilton Vantage




IGI Automated Sample Prep Step 3 – Sample Distribution to qPCR Reaction Plate

One to sixteen 96 well plates of extracted RNA are placed on the deck, as indicated in the image below

A equal number of 50µL tips racks, are added to the deck (light green) at least one additional 50µL tips is added in the dark green area

Positive control source is added to the 32 position carrier in position 31 and the Negative control source is places in position 32

One to four 384 well PCR plates are filled, with PCR master-mix, by dispenser and places on the deck as indicated in the image below

Start qPCR Sample Distribution Method

Tips positions D4 and G10 are removed and become the wells for negative and positive controls in the qPCR plate

Negative controls in G7,G8,H7 & H8 Positive controls in M19, M20, N19 & N20

5µL of extracted RNA is transferred to the qPCR plate by the core head 1 plate is equal to 1 quadrant

The qPCR controls are added to the plate and the plate is now ready for qPCR

Method Complete

Negative controls in position 32
Positive controls in position 31

To measure the LOD, the FDA recommends that “laboratories test a dilution series of three replicates per concentration, and then confirm the final concentration with 20 replicates.”[328] For the purposes of an EUA, the agency defines the LOD as “the lowest concentration at which 19/20 replicates are positive.” In accordance with this recommendation, we determined the LOD of our assay to be 1 genomic copy per microliter (Figure 5.7).

To ensure that diagnostic tests are clinically valid, the FDA recommends “that laboratories confirm performance of their assay with a series of contrived clinical specimens by testing a minimum of 30 contrived reactive specimens and 30 non-reactive specimens.” We assessed the clinical validity of our LDT by creating a range of different types of contrived SARS-CoV-2 RNA-positive and negative samples using SARS-CoV-2 positive control RNA from the TaqPath COVID-19 Control Kit, patient samples positive for SARS-CoV-2 from two local testing facilities, and human negative control RNA. Results showed 100% concordance with expected positive and negative samples (Figures 5.7 and 5.8).

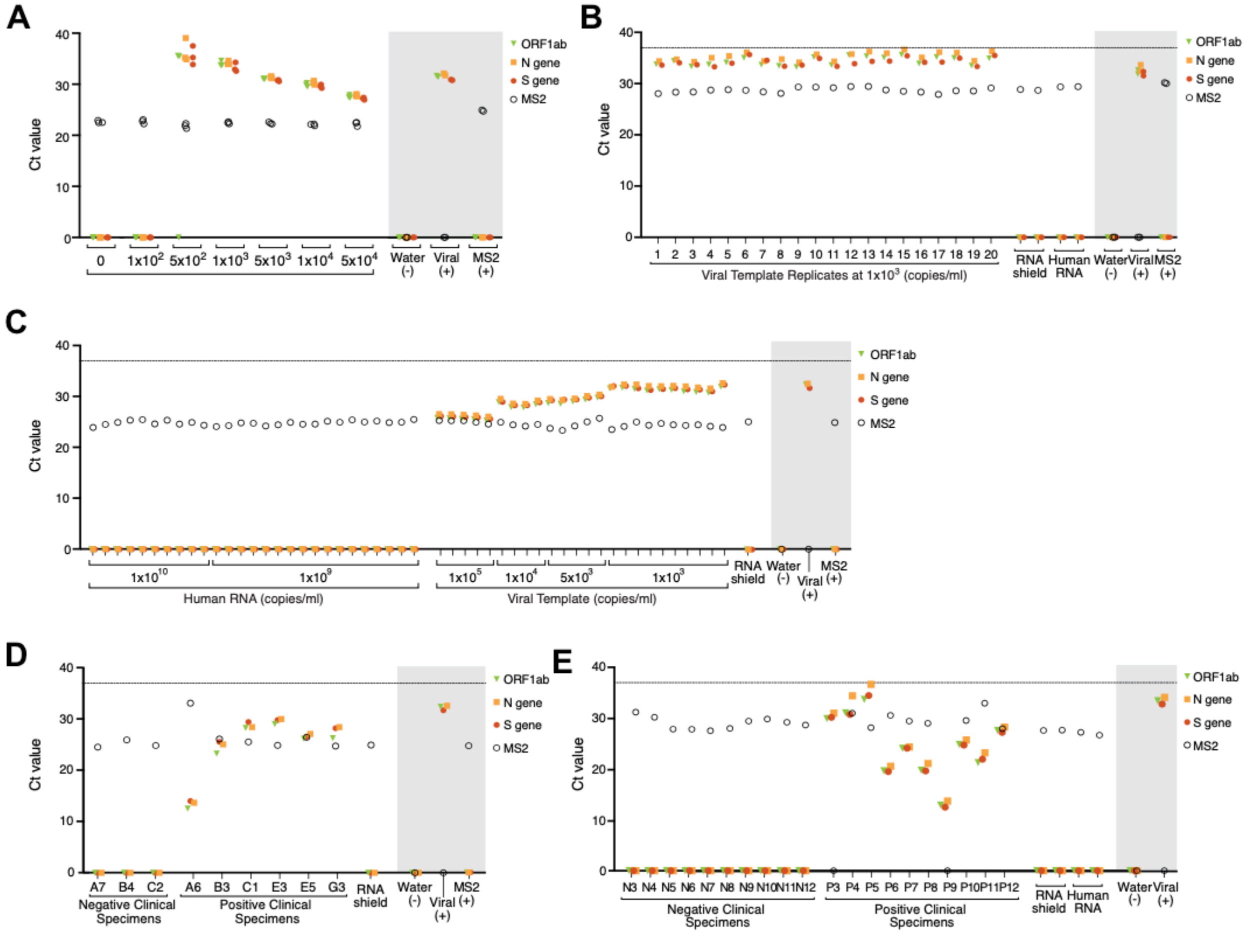


Figure 5.7: Validation of SARS-CoV-2 IGI LDT in Semi-Automated Method. (A) Limit of detection assay. Serial dilutions of SARS-CoV-2 positive control RNA with spiked-in MS2 internal control underwent RNA extraction and subsequent RT-qPCR amplification in triplicate using qPCR primers and probes targeting ORF1ab, N gene, S gene, and MS2 nucleic acid. RT-qPCR controls were performed in duplicate and are plotted in the shaded region of the graph: "Viral (+)" control contains only Thermo Fisher's SARS-CoV-2 positive control RNA (at 50 genomic copies per reaction) in RT-qPCR master mix, while "MS2 (+)" control contains only the MS2 nucleic acid in RT-qPCR master mix. All replicates are plotted as individual points, and an undetermined Ct value is plotted as Ct = 0. (**B)** Reproducibility at the limit of detection. Twenty replicate samples containing SARS-CoV-2 positive control RNA at 1 x 10^3 copies/ml underwent RNA extraction and RT-qPCR amplification as in **(A).** "RNA shield" and "Human RNA" represent full workflow controls and were processed from RNA extraction through RT-qPCR. The "RNA shield" control contained M2S internal control in sample collection buffer (DNA/RNA Shield in PBS), while the "Human RNA" control contained RNA extracted from HEK293T cells in sample collection buffer containing MS2 internal control. RT-qPCR controls (shaded region) were performed in duplicate and plotted as in **(A)**. An undetermined Ct value is plotted as Ct = 0. A Ct of 37 is indicated with a dotted line and represents the maximum Ct value that will return a valid result as defined by the uppermost Ct values in the LOD sample replicates. (**C**) Clinical Sample Evaluation Assay using Contrived Samples. Contrived positive "Viral Template" (SARS-CoV-2 positive control RNA at the indicated concentrations) and contrived negative "Human RNA" (HEK293T cell extracted RNA at the indicated concentrations) samples underwent RNA extraction and RT-qPCR amplification as in **(A)** in duplicate thermal cyclers. "RNA shield" control was performed as in **(B)**, and was composed of MS2 internal control in sample collection buffer. Single RT-qPCR controls (shaded region) were performed and plotted as in **a**. Results for "RNA shield" and RT-qPCR controls are duplicated in **(D)**, as experiments were run in parallel. Results from the duplicate RT-qPCR are shown in Fig. 5.8A. (**D**), Clinical Sample Evaluation for Samples from Clinical Testing Site 1. Clinical samples previously

determined to be positive or negative for SARS-CoV-2 underwent RNA extraction and RT-qPCR amplification as in **(A)** in duplicate thermal cyclers. “RNA shield” control was performed and plotted as in **(B)**. RT-qPCR controls (shaded region) were performed and plotted as in **(A)**. Results for “RNA shield” and RT-qPCR controls are duplicated in **(C)**, as experiments were run in parallel. The results from the second duplicate RT-qPCR are shown in Fig. 5.8B. **(E)** Clinical Evaluation of Samples from Kaiser Permanente. Samples in universal transport medium, previously determined to be positive or negative for SARS-CoV-2, underwent inactivation with DNA/RNA Shield, RNA extraction, and RT-qPCR amplification as in **a** in duplicate thermal cyclers. “RNA shield” control was performed and plotted as in **(B)**. RT-qPCR controls (shaded region) were performed and plotted as in **(A)**. The results from the second duplicate RT-qPCR are shown in Fig. 5.8C. For all panels, a dashed line marks Ct = 37, and undetermined Ct values are plotted as Ct = 0.

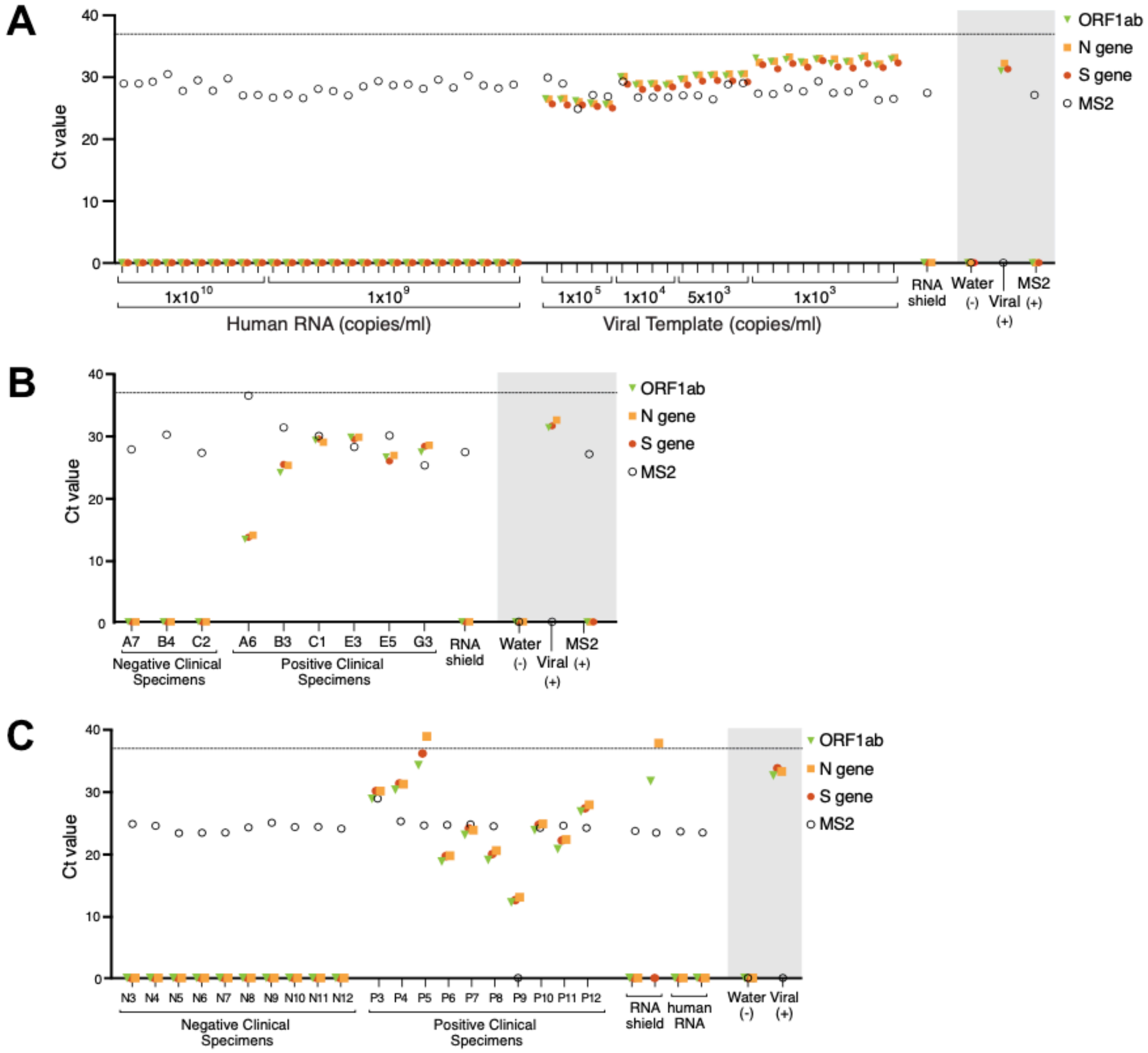


Figure 5.8: Clinical Validation, Contrived and Clinical Samples in Semi-Automated method - duplicate RT-qPCR plates. (A) Duplicate results corresponding to Fig 5.7C. Experimental layout as in corresponding figure legend. (**B)** Duplicate results corresponding to Supplementary Fig. 5.7D. Experimental layout as in corresponding figure legend. (**C)** Duplicate results corresponding to Fig 5.7E. Experimental layout as in corresponding figure legend. Amplification of ORF1ab was detected at a Ct <37 in one RNA

shield extraction control replicate on this plate. This likely was a technical error due to cross contamination between wells, as this was not detected in the duplicate plate (Fig. 5.7E).

Specificity of the primer–probe pairs and the potential for cross-reactivity with other common pathogens were previously assessed by Thermo Fisher in their EUA. To further confirm specificity in our pipeline, we showed that the primer–probe sets do not cross-react with human RNA from a virus-negative cell line and that they return negative results for patient samples from alternative testing facilities that were previously identified as SARS-CoV-2 negative using orthogonal primer–probe pairs (Figures 5.7 and 5.8).

Upon developing our automated workflow, we performed experiments to assess the LOD and clinical and analytical sensitivity (Figure 5.9). Our automated method showed comparable analytical and clinical sensitivity, with an LOD at or better than that of our semi-automated method. At the time of this Correspondence, we are preparing for EUA submission to begin the switch to an automated testing platform.

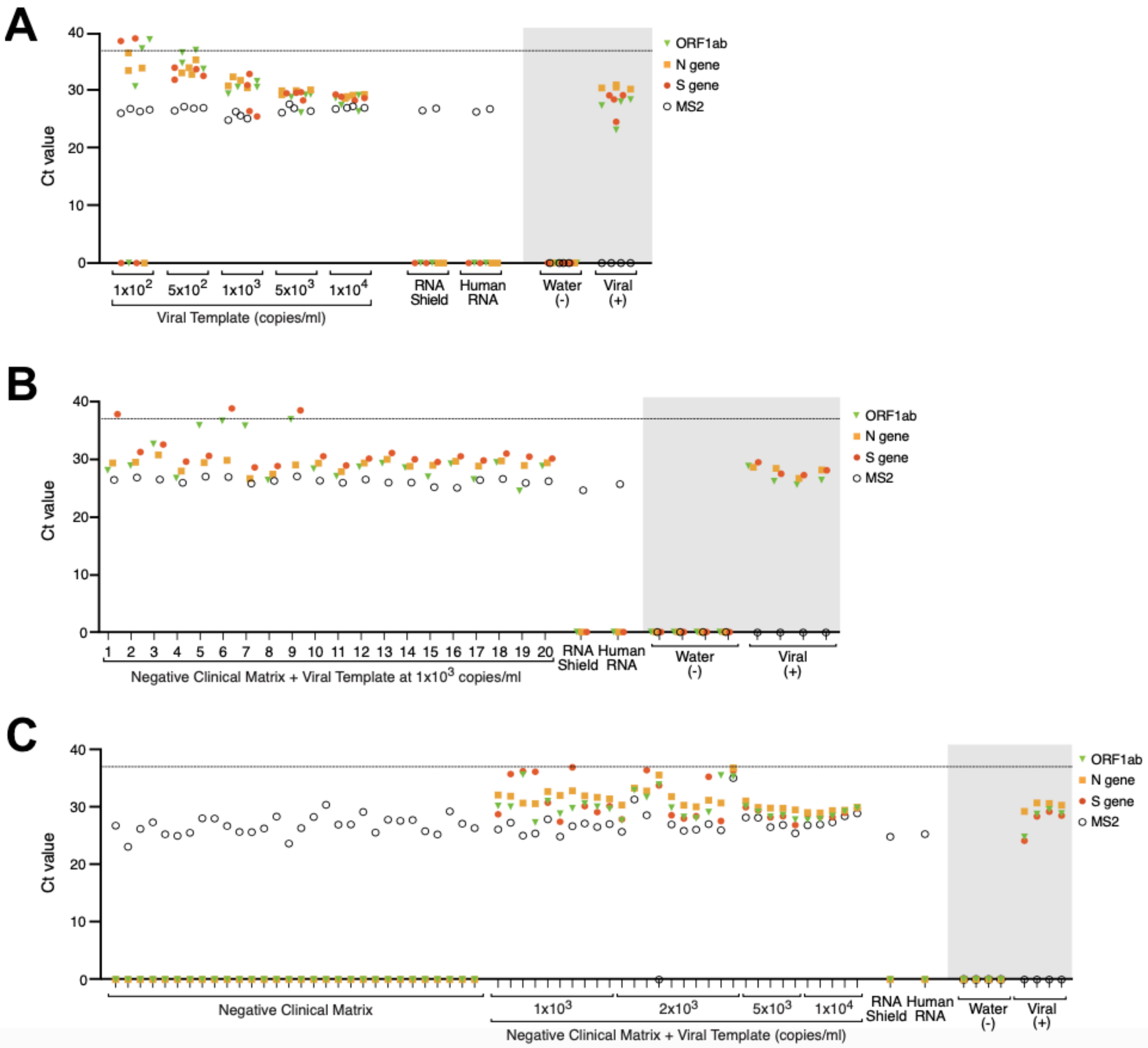
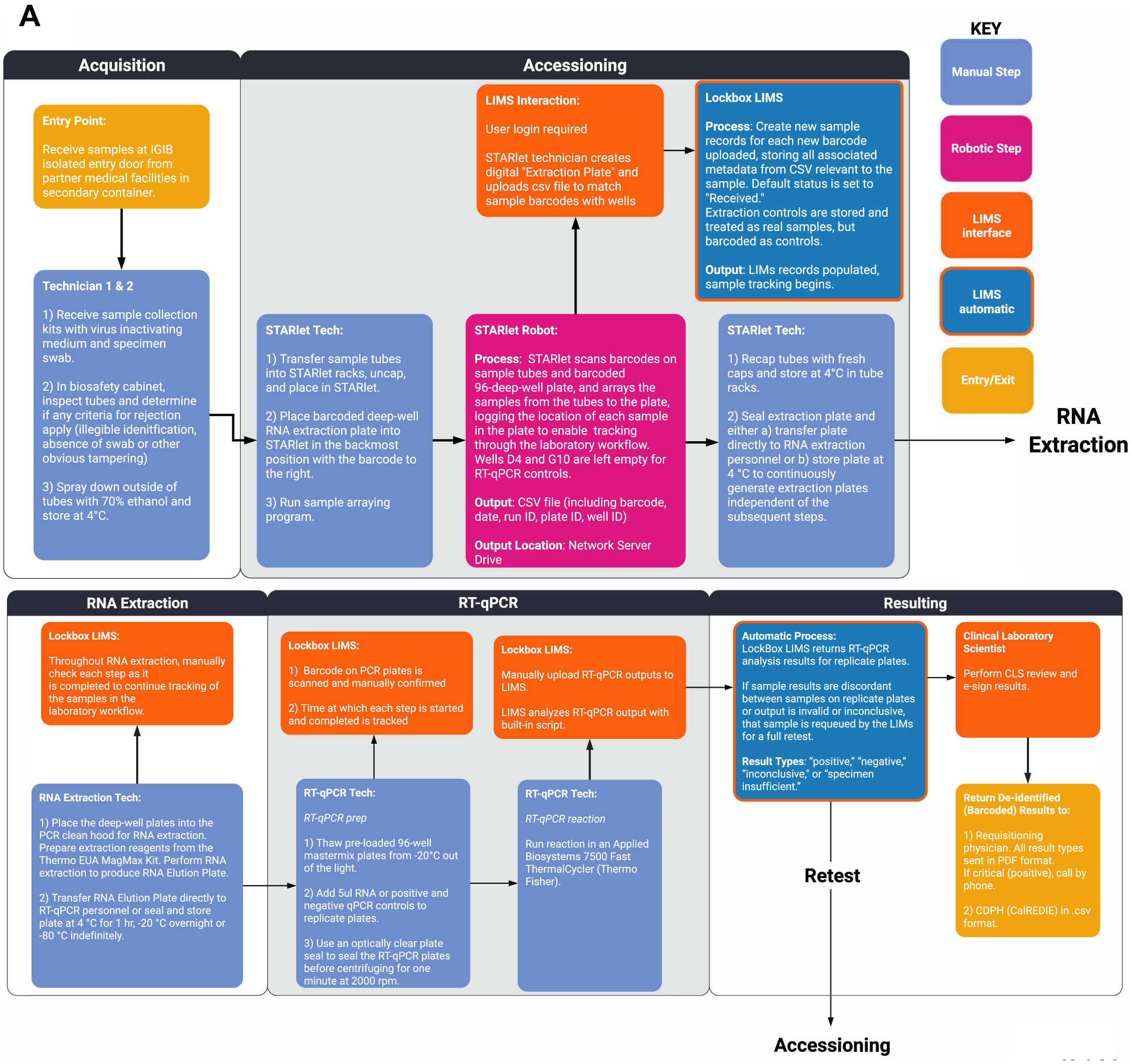


Figure 5.9: Establishing an automated SARS-CoV-2 IGI LDT - Limit of Detection and Clinical Sample Validation in Automated Method. (A) Limit of detection assay. Serial dilutions of SARS-CoV-2 positive control RNA with spiked-in MS2 internal control underwent RNA extraction and subsequent RT-qPCR amplification in quadruplicate using qPCR primers and probes targeting ORF1ab, N gene, S gene, and MS2 nucleic acid. RNA shield alone or Human RNA were included as above in duplicate. The "RNA shield" control contained M2S internal control in sample collection buffer (DNA/RNA Shield in PBS), while the "Human RNA" control contained RNA extracted from HEK293T cells in sample collection buffer containing MS2 internal control. RT-qPCR controls were performed in quadruplicate and are plotted in the shaded region of the graph: "Viral (+)" control contains only SARS-CoV-2 positive control RNA (at 50 copies per reaction) in RT-qPCR master mix. All replicates are plotted as individual points. An undetermined Ct value is plotted as Ct = 0. A Ct of 37 is indicated with a dotted line and represents the maximum Ct value that will return a valid result as defined by the uppermost Ct values in the LOD sample replicates. **(B)** Reproducibility at the limit of detection. Twenty replicate samples containing SARS-CoV-2 positive control RNA at $1\text{x}10^3$ copies/mL were diluted into leftover clinical specimens that were previously resulted as "Negative" using the IGI SARS-CoV-2 Manual LDT. These underwent RNA extractions and RT-qPCR amplification as in **(A)** alongside full workflow controls, "RNA shield" and "Human RNA". An undetermined Ct value is plotted as Ct = 0. A Ct of 37 is indicated with a dotted line and represents the maximum Ct value that will return a valid result. **(C)** Clinical Evaluation. Sixty leftover clinical specimens that were previously resulted as "Negative" using the IGI SARS-CoV-2 Manual LDT were selected; SARS-CoV-2 positive control RNA at the indicated concentrations were diluted into thirty of these negative clinical specimens. These underwent RNA extractions and RT-qPCR amplification as in **(A)** alongside full workflow controls, "RNA shield" and "Human RNA." An undetermined Ct value is plotted as Ct = 0. A Ct of 37 is indicated with a dotted line and represents the maximum Ct value that will return a valid result.

To support the technical workflow of the IGI testing lab and to track patient samples through our facility, we developed a custom laboratory information management software (LIMS) system with Third Wave Analytics. Our LIMS was designed in two phases to accommodate both our semi-automated and automated approaches in a cost-effective and HIPAA-compliant manner (Figure 5.10). In phase 1 of the LIMS build, de-identified patient barcodes are used for sample tracking. The LIMS returns de-identified barcoded results, which are sent to the UHS for integration into their electronic health system. There, the results are connected to a specific patient's file via standard unique identifier barcode matching. A template of our LIMS architecture, as customized for our semi-automated setup, is available for the cost of licensing alone, as described on the IGI website (https://innovativegenomics.org/sars-cov-2-testing-guide/).

In phase 2 of the LIMS development, corresponding to our automated approach, we built a HIPAA-compliant clinician access portal through which tests are requisitioned and reported, supporting non-UC Berkeley patient samples (Figure 5.3). To further enable non-UC Berkeley partnerships, we incorporated a payment interface within the portal. Partnering with non-UC Berkeley organizations enabled us to bringing testing to underserved, high-risk and other priority populations, such as the unsheltered, first responders and essential infrastructure workers. To meet these needs in our community, our partners include LifeLong Medical, the City of Berkeley, Roots Community Clinic and the State of California publicly owned utilities.

All testing results are reported to the requisitioning physician and the California Reportable Disease Information Exchange (CalREDIE) of the State of California Department of Public Health in .csv format. Data are transferred using UC Berkeley's Google e-mail service plus Virtru encryption.



Entry Point:

Receive samples at IGIB isolated entry door from partner medical facilities in secondary container.

Technician 1 & 2

1) Receive sample collection kits with virus inactivating medium and specimen swab.

2) In biosafety cabinet, inspect tubes and determine if any criteria for rejection apply (illegible ideinitfication, absence of swab or other obvious tampering)

3) Spray down outside of tubes with 70% ethanol and store at 4°C.

STARlet Tech:

1) Transfer sample tubes into STARlet racks, uncap, and place in STARlet.

2) Place barcoded deep-well RNA extraction plate into STARlet in the backmost position with the barcode to the right.

3) Run sample arraying program.

STARlet Robot:

Process: STARlet scans barcodes on sample tubes and barcoded 96-deep-well plate, and arrays the samples from the tubes to the plate, logging the location of each sample in the plate to enable tracking through the laboratory workflow. Wells D4 and G10 are left empty for RT-qPCR controls.

Output: CSV file (including barcode, date, run ID, plate ID, well ID)

Output Location: Network Server Drive

LIMS Interaction:

User login required

STARlet technician creates digital "Extraction Plate" and uploads csv file to match sample barcodes with wells

Lockbox LIMS

Process: Create new sample records for each new barcode uploaded, storing all associated metadata from CSV relevant to the sample. Default status is set to "Received."
Extraction controls are stored and treated as real samples, but barcoded as controls.

Output: LIMs records populated, sample tracking begins.

STARlet Tech:

1) Recap tubes with fresh caps and store at 4°C in tube racks.

2) Seal extraction plate and either a) transfer plate directly to RNA extraction personnel or b) store plate at 4 °C to continuously generate extraction plates independent of the subsequent steps.

Lockbox LIMS:

Throughout RNA extraction, manually check each step as it is completed to continue tracking of the samples in the laboratory workflow.

RNA Extraction Tech:

1) Place the deep-well plates into the PCR clean hood for RNA extraction. Prepare extraction reagents from the Thermo EUA MagMax Kit. Perform RNA extraction to produce RNA Elution Plate.

2) Transfer RNA Elution Plate directly to RT-qPCR personnel or seal and store plate at 4 °C for 1 hr, -20 °C overnight or -80 °C indefinitely.

Lockbox LIMS:

1) Barcode on PCR plates is scanned and manually confirmed

2) Time at which each step is started and completed is tracked

RT-qPCR Tech:

RT-qPCR prep

1) Thaw pre-loaded 96-well mastermix plates from -20°C out of the light.

2) Add 5ul RNA or positive and negative qPCR controls to replicate plates.

3) Use an optically clear plate seal to seal the RT-qPCR plates before centrifuging for one minute at 2000 rpm.

Lockbox LIMS:

Manually upload RT-qPCR outputs to LIMS.

LIMS analyzes RT-qPCR output with built-in script.

RT-qPCR Tech:

RT-qPCR reaction

Run reaction in an Applied Biosystems 7500 Fast ThermalCycler (Thermo Fisher).

Automatic Process:
LockBox LIMS returns RT-qPCR analysis results for replicate plates.

If sample results are discordant between samples on replicate plates or output is invalid or inconclusive, that sample is requeued by the LIMs for a full retest.

Result Types: “positive,” “negative,” “inconclusive,” or “specimen insufficient.”

Clinical Laboratory Scientist

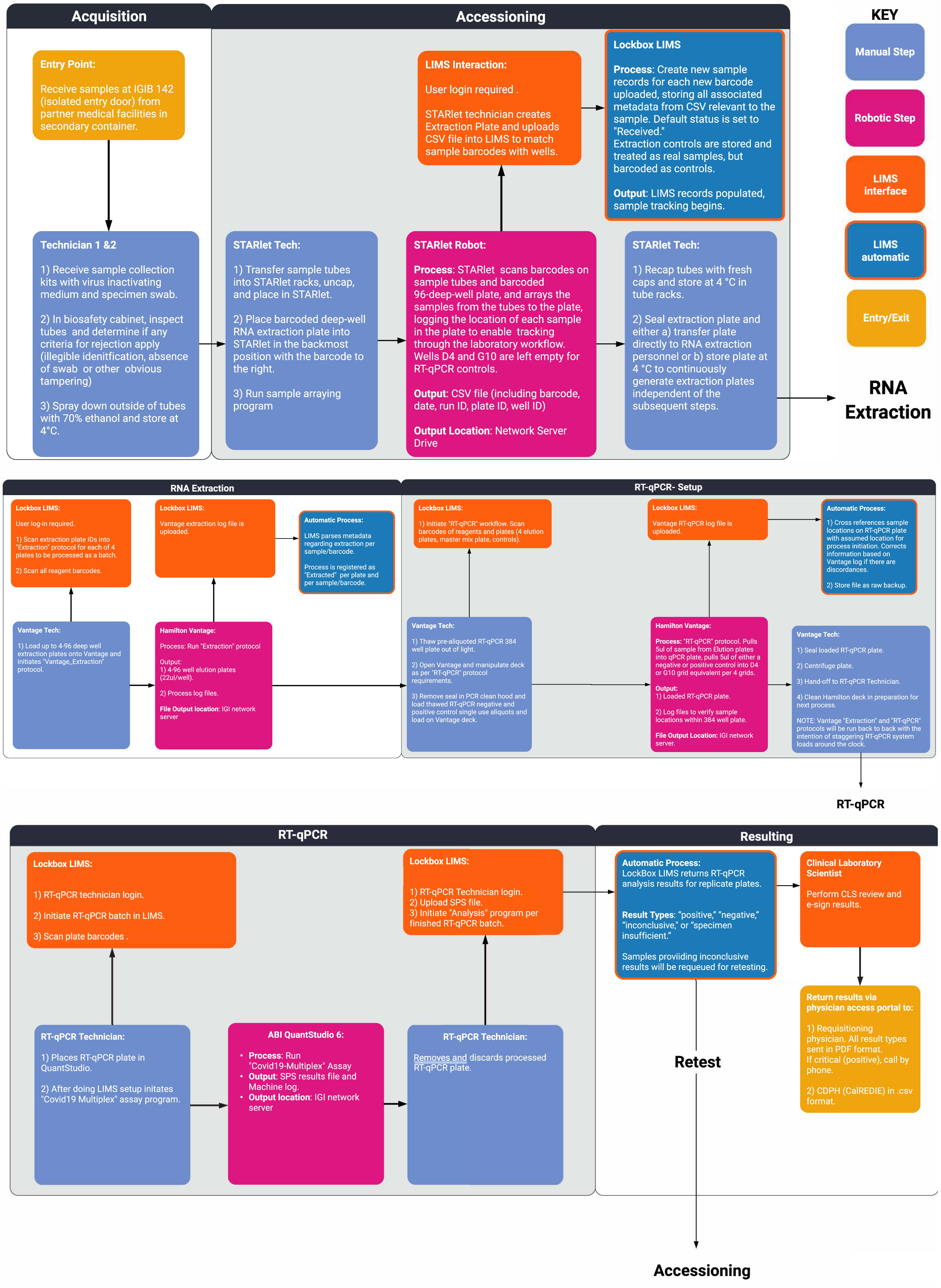
Perform CLS review and e-sign results.

Return De-identified (Barcoded) Results to:

1) Requisitioning physician. All result types sent in PDF format.
If critical (positive), call by phone.

2) CDPH (CalREDIE) in .csv format.

B



Entry Point:

Receive samples at IGIB 142 (isolated entry door) from partner medical facilities in secondary container.

Technician 1 &2

1) Receive sample collection kits with virus inactivating medium and specimen swab.

2) In biosafety cabinet, inspect tubes and determine if any criteria for rejection apply (illegible ideintification, absence of swab or other obvious tampering)

3) Spray down outside of tubes with 70% ethanol and store at 4°C.

STARlet Tech:

1) Transfer sample tubes into STARlet racks, uncap, and place in STARlet.

2) Place barcoded deep-well RNA extraction plate into STARlet in the backmost position with the barcode to the right.

3) Run sample arraying program

STARlet Robot:

Process: STARlet scans barcodes on sample tubes and barcoded 96-deep-well plate, and arrays the samples from the tubes to the plate, logging the location of each sample in the plate to enable tracking through the laboratory workflow. Wells D4 and G10 are left empty for RT-qPCR controls.

Output: CSV file (including barcode, date, run ID, plate ID, well ID)

Output Location: Network Server Drive

LIMS Interaction:

User login required .

STARlet technician creates Extraction Plate and uploads CSV file into LIMS to match sample barcodes with wells.

Lockbox LIMS

Process: Create new sample records for each new barcode uploaded, storing all associated metadata from CSV relevant to the sample. Default status is set to "Received."
Extraction controls are stored and treated as real samples, but barcoded as controls.

Output: LIMS records populated, sample tracking begins.

STARlet Tech:

1) Recap tubes with fresh caps and store at 4 °C in tube racks.

2) Seal extraction plate and either a) transfer plate directly to RNA extraction personnel or b) store plate at 4 °C to continuously generate extraction plates independent of the subsequent steps.

Lockbox LIMS:

User log-in required.

1) Scan extraction plate IDs into "Extraction" protocol for each of 4 plates to be processed as a batch.

2) Scan all reagent barcodes.

Vantage Tech:

1) Load up to 4-96 deep well extraction plates onto Vantage and initiates "Vantage_Extraction" protocol.

Hamilton Vantage:

Process: Run "Extraction" protocol

Output:
1) 4-96 well elution plates (22ul/well).

2) Process log files.

File Output location: IGI network server

Lockbox LIMS:

Vantage extraction log file is uploaded.

Automatic Process:

LIMS parses metadata regarding extraction per sample/barcode.

Process is registered as "Extracted" per plate and per sample/barcode.

Lockbox LIMS:

1) Initiate "RT-qPCR" workflow. Scan barcodes of reagents and plates (4 elution plates, master mix plate, controls).

Vantage Tech:

1) Thaw pre-aliquoted RT-qPCR 384 well plate out of light.

2) Open Vantage and manipulate deck as per "RT-qPCR" protocol requirements.

3) Remove seal in PCR clean hood and load thawed RT-qPCR negative and positive control single use aliquots and load on Vantage deck.

Hamilton Vantage:

Process: "RT-qPCR" protocol. Pulls 5ul of sample from Elution plates into qPCR plate, pulls 5ul of either a negative or positive control into D4 or G10 grid equivalent per 4 grids.

Output:
1) Loaded RT-qPCR plate.

2) Log files to verify sample locations within 384 well plate.

File Output Location: IGI network server.

Lockbox LIMS:

Vantage RT-qPCR log file is uploaded.

Automatic Process:

1) Cross references sample locations on RT-qPCR plate with assumed location for process initiation. Corrects information based on Vantage log if there are discordances.

2) Store file as raw backup.

Vantage Tech:

1) Seal loaded RT-qPCR plate.

2) Centrifuge plate.

3) Hand-off to RT-qPCR Technician.

4) Clean Hamilton deck in preparation for next process.

NOTE: Vantage "Extraction" and "RT-qPCR" protocols will be run back to back with the intention of staggering RT-qPCR system loads around the clock.

Lockbox LIMS:

1) RT-qPCR technician login.

2) Initiate RT-qPCR batch in LIMS.

3) Scan plate barcodes .

RT-qPCR Technician:

1) Places RT-qPCR plate in QuantStudio.

2) After doing LIMS setup initates "Covid19 Multiplex" assay program.

ABI QuantStudio 6:

- **Process**: Run "Covid19-Multiplex" Assay
- **Output**: SPS results file and Machine log.
- **Output location**: IGI network server

RT-qPCR Technician:

Removes and discards processed RT-qPCR plate.

Lockbox LIMS:

1) RT-qPCR Technician login.
2) Upload SPS file.
3) Initiate "Analysis" program per finished RT-qPCR batch.

Automatic Process:
LockBox LIMS returns RT-qPCR analysis results for replicate plates.

Result Types: "positive," "negative," "inconclusive," or "specimen insufficient."

Samples proviiding inconclusive results will be requeued for retesting.

Clinical Laboratory Scientist

Perform CLS review and e-sign results.

Return results via physician access portal to:

1) Requisitioning physician. All result types sent in PDF format.
If critical (positive), call by phone.

2) CDPH (CalREDIE) in .csv format.

Figure 5.10: Semi-automated and automated sample workflow. (A) Semi-automated sample workflow. **(B)** Automated sample workflow.

5.4 Discussion

In summary, we describe here a process for creating a CLIA-certified clinical testing laboratory at a non-medical institution. Extending the license from an existing CLIA-certified facility, using online HIPAA training and adapting a commercially available FDA-authorized test saved substantial time and resources. Supply-chain bottlenecks were managed by sourcing alternative collection tubes and swabs from a provider with adequate stocks, using donated equipment, validating half-volume reactions in our assay, adopting in-house sample barcoding, and adapting materials (for example, sampling tubes) to work with available equipment. Finally, PPE including masks, gloves, and gowns was obtained by donation to our facility.

The IGI testing laboratory is currently supported primarily by philanthropy, which has enabled our rapid deployment and access to populations that would not otherwise be able to obtain testing. While the CARES Act mandates insurance coverage for SARS-CoV-2 testing, including providing coverage for the uninsured, universities without an affiliated hospital lack the insurance claims department necessary to access this financial support. We encourage government entities to consider grants to universities with testing facilities to enable these critical services to continue uninhibited and to enable expansion of the range of institutions capable of responding to this crisis.

Some observations based on initial test results are pertinent here. First, the range of viral titer in patient specimens can vary by six orders of magnitude, and we consistently detect positive specimens with a viral load that approaches our LOD. This range will affect the test sensitivity for pooled samples, raising concerns about the utility of such surveillance testing if widely implemented. Second, we observe positive specimens obtained by oral, nasal and mixed oral–nasal methods, but our data do not address the comparative sensitivity of these methods. Finally, we have detected a ~3.5% positivity rate for our total population tested, which is enriched for symptomatic or potentially exposed UC Berkeley-affiliated patients and low socioeconomic status or vulnerable community members (Figure 5.11). Although in aggregate this positive test result percentage agrees with that of other testing facilities in the San Francisco Bay Area, these positive samples are not evenly distributed among the populations we are serving. This observation, while not part of a controlled study, nonetheless underscores findings elsewhere that this disease is disproportionately affecting communities experiencing existing health disparities[329–332]. Our observations emphasize the need for expanding access to testing and follow-up care for these communities.

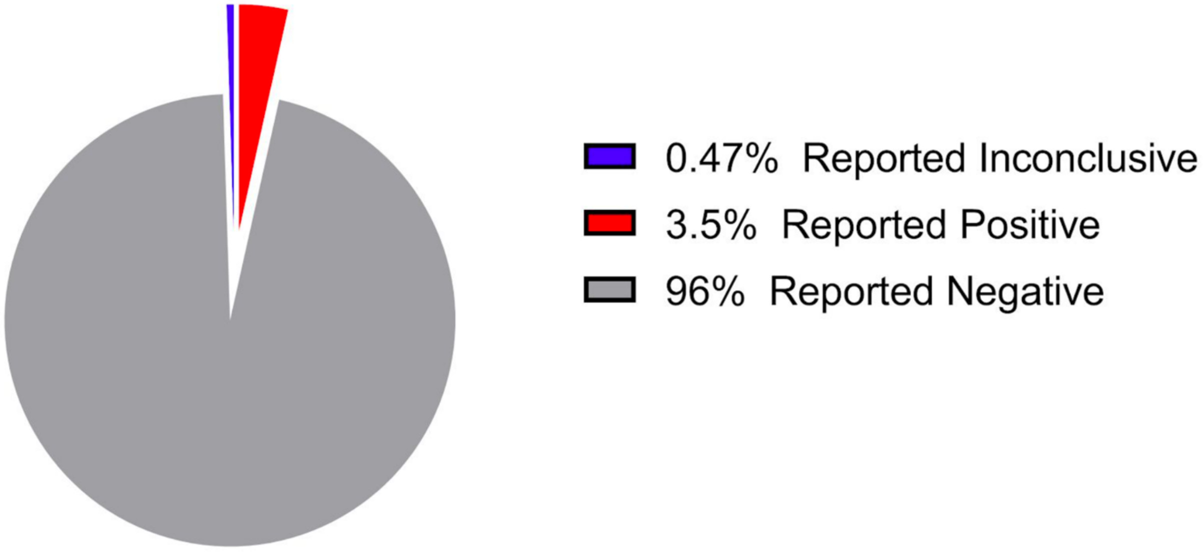


Figure 5.11: Resulting outputs for SARS-CoV-2 testing in the IGI clinical testing facility, by date of publication. At time of writing, the IGI clinical testing facility has noted a positivity rate of ~3.5% for our total population tested, which is enriched for symptomatic or potentially exposed UC Berkeley affiliates and low socioeconomic status or vulnerable community members.

To address continued large-scale surveillance needs, the IGI facility is developing saliva-based testing and eventually aims to implement serologic testing to enable better monitoring of population transmission and seroconversion rates. In keeping with our mission as a research institute, our facility also enables research on asymptomatic transmission and analysis of virus sequence evolution and provides benchmarking for new diagnostic technologies.

Although the challenges we faced were formidable, our experience and that of others demonstrates that they can be overcome. We encourage other institutions with a molecular biology department and health clinic with CLS staff to replicate or further amplify our approach and together create an invaluable resource for controlling this pandemic.

5.5 Materials and methods

5.5.1 Laboratory developed test validation

FDA-approved diagnostic testing for COVID-19 infection is primarily based on quantitative real-time polymerase chain reactions (RT-qPCR) to measure SARS-CoV-2 viral RNA present within a patient specimen. To do this, an intact specimen is collected by nasopharyngeal or oropharyngeal swab at a health care facility and is submitted to the diagnostic testing facility. RNA is extracted from the sample then RT-qPCR is used to detect viral S-gene, N-gene, and ORF1ab, along with an MS2 Phage control. Tests are scored as positive or negative for viral infection based on our validated limit of detection

Ct threshold (see below under *Limit of Detection (Semi-Automated Method)*), and using EUA approved criteria identified in Thermo Fisher's instructions (see table below).

ORF1ab	N gene	S gene	MS2	Status	Result	Action
NEG	NEG	NEG	NEG	Invalid	-	Repeat test. If the repeat result remains invalid, consider collecting a new specimen.
NEG	NEG	NEG	POS	Valid	SARS-CoV-2 Not Detected	Report result
POS for **one** SARS-CoV-2 target			POS or NEG	Valid	SARS-CoV-2 Inconclusive	Repeat test. If the repeat result remains inconclusive, additional confirmation testing should be conducted if clinically indicated.
POS for **two** or more SARS-CoV-2 targets			POS or NEG	Valid	Positive for SARS-CoV-2	Report result
Discordant results between PCR plate replicates, in terms of viral genes detected (used during IGI's manual protocol)				Invalid	-	Repeat test

5.5.2 Limit of detection (semi-automated method)

A serial dilution of the SARS-CoV-2 positive control RNA in the ThermoFisher TaqPath COVID-19 Control Kit was prepared in our sample collection medium (2X DNA/RNA Shield diluted to 1X in phosphate-buffered saline (PBS)), spanning 1×10^2 to 5×10^4 genomic copies of SARS-CoV-2 RNA per mL. 450 µl of each concentration was manually transferred to deep-well 96-well plates in triplicate for RNA extraction. Extraction followed by RT-qPCR were performed as in the semi-automated SOP. The RT-qPCR fluorescence threshold values were set manually based on background fluorescence of the reaction mixture. Supplied MS2 (Thermo Fisher) was used as an internal control for RNA extraction, reverse transcription, and PCR amplification. We then performed 20 replicates at our target LoD to confirm the limit of detection. Defining the LOD also provides a critical parameter for interpreting sample results. At the concentration of our limit of detection, our highest Ct value was 36.67 (Supplementary Fig. 3b, sample 15, N gene). This defines an upper Ct boundary for reliably identifying positive samples. From this, we set our reporting criteria such that a Ct above 37 will return a negative result as defined in our analysis software for that gene.

5.5.3 Clinical sample evaluation assay (semi-automated method)

For all samples, RNA extraction and RT-qPCR were performed as detailed in the semi-automated SOP and outlined in this document under “Patient Sample Testing”.

Contrived samples

24 mock positive samples were created by diluting (in a range near our LoD) Thermo Fisher’s SARS-CoV-2 positive control RNA from the TaqPath COVID-19 Control Kit into our sample collection medium (1X DNA/RNA Shield in PBS). 27 Mock negative samples were created by diluting RNA purified from cultured human 293T cells into IGI’s sample collection medium (1X DNA/RNA Shield in PBS). Human RNA concentration was calculated using UV absorbance, and the concentration was converted to an approximate copy number for comparison with positive SARS-CoV-2 RNA controls using the NEBioCalculator for ssRNA with an estimated average RNA fragment length of 100nt.

Samples from Clinical Testing Site 1

Samples from a local clinical testing facility, Clinical Testing Site 1, arrived at the IGI in a frozen state, as 200µl of leftover clinical samples already inactivated in DNA/RNA Shield. Since the local testing facility receives samples in UTM, they added equal volumes of 2X DNA/RNA Shield to inactivate their samples prior to entering their testing pipeline. We received 200 µl of these leftover inactivated samples. After thawing at IGI, sample collection medium (1X DNA/RNA Shield in PBS) was added to the 200 µl-samples to bring the final volume to 450 µl. These were then manually pipetted into 96-deep-well plates by transferring the full 450 µl, and RNA extraction and PCR reaction preparation were executed following the instructions in the semi-automated SOP.

Kaiser Permanente Clinical Specimen Preparation

Ten clinical samples known to be positive for SARS-CoV-2 and ten clinical samples known to be negative for SARS-CoV-2 were obtained from Kaiser Permanente. These samples arrived in COPAN clinical specimen tubes, frozen in universal transport medium. These samples were received into the UC Berkeley BSL-3 facility by Dr. Sarah Stanley’s laboratory. In the BSL-3, samples were thawed and 450 µl of sample was mixed with 450 µl 2X DNA/RNA Shield to reach a final concentration of 1X DNA/RNA Shield. This sample (900 µl) was then transferred from the primary tube into a new sample tube in a BSL-2 biosafety cabinet within the BSL-3, and the primary tube was discarded into BSL-3 waste. The outside of the new sample tube was decontaminated and moved out of the BSL-3 for analysis at the IGI. Once at the IGI, the s emi-automated SOP was followed to array the full contents (450 µl) of the clinical samples into 96-deep-well plates using the Hamilton STARlet, and procedures for manual pipetting of RNA extraction and RT-qPCR were followed exactly.

5.5.4 Limit of detection (automated method)

A serial dilution of the SARS-CoV-2 positive control RNA in the ThermoFisher TaqPath COVID-19 Control Kit was prepared in our sample collection medium (2X DNA/RNA Shield diluted to 1X in phosphate-buffered saline (PBS)), spanning 1x10^2 to 1x10^4 genomic copies of SARS-CoV-2 RNA per mL. 450 µl of each concentration was accessioned on the STARlet in quadruplicate for concentration. Extraction followed by RT-qPCR were performed.The RT-qPCR fluorescence threshold values were set manually based on background fluorescence of the reaction mixture. MS2 (Thermo Fisher) was used as an internal control for RNA extraction, reverse transcription, and PCR amplification. We then performed 20 replicates of SARS-CoV-2 positive control RNA diluted to our LoD in leftover clinical matrix. Twenty specimens were chosen from a pool of specimens that previously passed through the IGI SARS-CoV-2 manual LDT and were not resulted as “Positive”. These could theoretically have been “Negative” (no viral gene amplification) or “New Specimen Requested” (specimen insufficient or inconclusive), but random selection yielded 20 “Negative” specimens. These specimens were generated using our current kits preloaded with 2.5ml DNA/RNA Shield in PBS, but the Automated LDT will use kits preloaded with 4.0ml of the DNA/RNA Shield in PBS solution, so the leftover specimens were diluted to simulate the more dilute clinical marix the automated assay will encounter. Each specimen had 1 x 10^3 copies/ml SARS-CoV-2 RNA spiked in and was processed from accessioning through RT-qPCR, as above. In the automated method validation experiments, we sought to bridge to our semi-automated assay by directly comparing the LOD and thus adopted our Ct threshold to be 37 in order to allow a direct comparison to our validated semi-automated method.

5.5.5 Clinical sample evaluation assay (automated method)

Contrived samples

Thirty mock positive and 30 mock negative samples were generated using leftover clinical specimens as in the automated method LoD experiment above and, for the positive samples, SARS-CoV-2 viral RNA. The dilutions tested included 20 mock positives at a concentration of 1-2 times the LoD and the remaining at 5x10^3 and 1x10^4 copies/ml. Extraction followed by RT-qPCR were performed.

5.5.6 LIMS verification

The previously tested and verified Lockbox LIMS software v1.48 was used as the template for the LIMS. Significant customizations were implemented to the v1.48 LIMS package, per the laboratory user requirements of the SARS-CoV-2 detection workflow.

Prior to clinical sample testing, the customized Lockbox LIMS was verified by Third Wave Analytics personnel to ensure that all customizations function appropriately per the unique user requirements for the SARS-CoV-2 detection workflow. To that end, all requirements were documented in Third Wave Analytics’ internal software verification system, first detailing all user requirements for how the system should function for the end laboratory user. Secondly, each user requirement then had a list of all the functional requirements for how the LIMS was specifically customized to meet the user requirements. Test cases were then created and run for each specific functional requirement, to document that each specific functional requirement was met. Upon the verification of the customized Lockbox LIMS software by Third Wave Analytics, the end users in the laboratory performed additional user acceptance testing to verify that the software performed as anticipated.

5.5.7 Patient sample testing (semi-automated method)

While establishing a fully automated method, we have started testing patient samples with a semi-automated method at a smaller scale. Notably, steps of laboratory testing (*e.g.,* step start and completion time) are directly tracked by laboratory technicians using tablet computers that directly report to the LIMS.

5.5.8 Acquisition

Briefly, during sample acquisition, a kit is received with a patient sample tube and biohazard bag containing matching barcodes and patient identifying information. The kits are then examined to determine acceptance or rejection for testing. Rejection criteria include: fully or partially uncapped sample tube, absence of swab, compromised sample identification (illegible or tube identification not matching the biohazard bag), and visible liquid in the kit's biohazard bag. For rejected sample kits, the disposal will be made in biohazard waste and recollection from the respective patients under investigation will be pursued by our health partners when possible.

An accepted patient sample tube is sprayed with 70% molecular biology-grade ethanol, wiped down with a Kimwipe in the biosafety cabinet and stored at 4°C or immediately accessioned (below). The testing workflow consists of four main steps: Accessioning, RNA Extraction, RT-qPCR and Resulting.

5.5.9 Accessioning

The clean barcoded patient sample tube is placed in a Hamilton Microlab STARlet rack to be scanned and arrayed into a barcoded 96-deep-well plate, which will then either be stored at 4°C or transferred directly to an extraction personnel for the RNA Extraction phase. For details on Hamilton Microlab STARlet Automation process see Supplementary Materials. The STARlet output .csv file is uploaded into the LIMS to register the now de-identified samples into the database as received and pending extraction.

5.5.10 RNA extraction

Extraction personnel place the centrifuged and sealed deep-well plates into the PCR clean hood for RNA extraction. In the hood, the reagents are prepared for the Thermo Fisher EUA MagMAX Viral/Pathogen Nucleic Acid Isolation Kit using our reduced reaction volume approach (see *Validation Methods* section), and samples are processed by digestion with proteinase K and addition of nucleic acid-binding magnetic beads. Pipette tips are switched between samples throughout the process to avoid cross-contamination. MS2 internal control (from Thermo Fisher’s TaqPath RT-PCR COVID-19 Kit) is spiked into every well as an internal control for RNA extraction and RT-PCR. Negative controls are a 1:1 mix of 2X DNA/RNA Shield in phosphate-buffered saline (PBS) and human cell line RNA. Using the LIMS interface, extraction personnel manually indicate completion of each step in the RNA extraction protocol. The RNA extraction plate is either transferred directly to RT-qPCR personnel or stored at 4°C for 1h, -20°C overnight or -80°C indefinitely.

5.5.11 RT-qPCR

For RT-qPCR, the TaqPath RT-PCR COVID-19 Kit (Thermo Fisher) COVID-19 positive control RNA at 1 x 10^4 genomic copies/mL is used diluted at a volume of 50 genomic copies/reaction. Negative control is molecular biology grade water (Thermo Fisher or equivalent). RT-qPCR personnel thaw pre-loaded master mix plates from -20°C, out of the light, for 10 min. Replicate RT-qPCR plates are prepared with matching patient sample positions. Identification of RT-qPCR plates used are manually entered in the LIMS interface and RT-qPCR steps are manually checked as they are completed. The RT-qPCR is held in an Applied Biosystems 7500 Fast ThermalCycler (Thermo Fisher). Reaction outputs for both plates are manually uploaded to the LIMS system for analysis.

5.5.12 Resulting

The uploaded outputs are analyzed and synthesized in the LIMS. Negative and positive results are evaluated by a CLS in the LIMS. A new specimen will be requested for any samples that do not result in a negative or positive test result after two full tests. Possible reported results include: “positive”, “negative”, “new specimen requested”. If sample results are discordant between samples on replicate plates (semi-automated method only), or if result is invalid or inconclusive the sample is requeued by the LIMS for a full retest starting in the accessioning phase.

Our resulting workflow is as follows: 1) call the physician that requisitioned the test with any critical (positive) results, 2) return all “positive,” “negative,” “inconclusive,” or “specimen insufficient” results as a PDF report per patient to the physician, then 3) after the test results have been officially reported, all “positive,” “negative,” and “inconclusive” results are transmitted as a .csv in a HIPAA-compliant manner (using UC Berkeley's Google email service with additional Virtru encryption) to CDPH to be uploaded to the California Reportable Disease Information Exchange (CalREDIE). CDPH uses CalREDIE to then report on state-level statistics to the CDC. In our ongoing phase two LIMS implementation, corresponding to our automated method, we have built a physician access portal that will provide results in PDF format to the requisitioning physician, integrate with provider electronic medical record systems like EPIC, and return CDPH reports to CalREDIE via an HL7 interface instead of .csv.

5.5.13 LIMS development

The Lockbox LIMS software v1.48 was used as the template for the LIMS, which is an OEM installed software package run on the Salesforce platform. Significant customizations were implemented to meet all the laboratory user requirements of the SARS-CoV-2 detection workflow. These customizations included: 1) automated sample record creation upon upload of sample barcodes, 2) the display of customized sample processing instructions for the laboratory user as samples are tested (e.g., sample testing SOPs), 3) automated record updating throughout the sample testing process (e.g., automated updating of sample or testing plates statuses), 4) creation of user-friendly data import tools (e.g., upload of raw Ct data upon completion of RT-qPCR), 5) automated analysis of all control and sample data to generate a final patient result instantaneously upon raw data upload, 6) customized and user-friendly workflow for the final review of sample data by appropriate laboratory personnel, 7) export of the result data for import into health partners systems (e.g. the UHS Tang Center), 8) automated sample record creation upon the re-testing of a sample, 9) the tracking of all reagents and control lots during testing, 10) customized dashboards to track all samples throughout the testing process, 11) creation of a provider portal (in the form of a partner community) for the ordering and reporting of results to non-UC Berkeley patients, and 12) the implementation of robust security measures to ensure that all PHI are encrypted (at rest) and that only appropriate personnel have access to appropriate records and data fields.

5.5.14 Institutional approval for human subject data

Descriptive statistics for patient samples used in this manuscript come from de-identified datasets in accordance with human subjects protections, as approved by UC Berkeley's Committee for Protection of Human Subjects (CPHS). While Institutional Review Board (IRB) approval is required for any research using human subjects, clinical laboratory activities exclusively supporting CLIA-certified clinical operations do not. These activities are governed by CMS and HIPAA legislation. In order to publish our data in this manuscript as results of developing a testing workstream, we sought IRB approval through UC Berkeley's CPHS. The UC Berkeley Committee for Protection of Human Subjects determined that all the analyses presented in this manuscript do not qualify as human subjects research as the data sets were de-identified to those analyzing them for these results (IRB submission # 2020-04-13177).

5.6 Accession codes

This work did not generate any data with accession codes.

5.7 Acknowledgments

We are indebted to the following organizations and individuals whose support and contributions made our work possible: UC Berkeley Chancellor Carol Christ, UC Berkeley Vice Chancellor for Research Randy Katz, UC Berkeley Dean for Biomedical Sciences Michael Botchan, lead UC Counsel Charles Robinson and UC President Janet Napolitano provided strategic support throughout the course of our effort. Robert Tjian of the MCB Department enabled our effort by consistent engagement and action at every level necessary. UC Berkeley Vice-Chancellor for Health Guy Nicolette, Holly Stern and the Tang Center chief medical officer, Anna Harte, enabled and guided our interface to University Health Services. Joseph DeRisi and Eric Chow (CZ Biohub) allowed our personnel to visit the CZ Biohub, where UCSF Clinical operates a SARS-CoV-2 diagnostic laboratory, and shared key details about their organizational setup and workflow that informed our own effort. Lisa Hernandez (City of Berkeley Public Health) and Chief David Brannigan (City of Berkeley Fire Department) envisioned and led the establishment of a partnership between the city, the IGI and LifeLong Medical. Inspector Dori Tieu (City of Berkeley Fire Department) provided key logistical and operational support. Michael Stacey, Kim Nguyen and Yui Nishiike (LifeLong Medical Care) provided key guidance and support on establishing workflows at the IGI–LifeLong interface; Breanne Grady and Noha Aboelata (Roots Community Clinic) lent their cognate enabling effort for the IGI–Roots interface. Commissioner Karen Douglas (California Energy Commission) and Barry Moline (California Municipal Utilities Association) initiated and guided the testing partnership between the IGI and State of California publicly owned utilities. Jeffrey Shapiro and Jacek Skarbinsky (Kaiser Permanente) provided de-identified patient samples. Alan Sachs, Elizabeth Kerr, Christopher Cowen and David Woo (Thermo Fisher) provided scientific and logistical support and ensured a sustained supply chain. Ray Turner and Nigel Mouncey (JGI) provided enabling personnel support. The Dillin, Karpen, Naar, Niyogi and Herr laboratories donated essential lab equipment. The Stanley laboratory made their BSL-3 space available to us. Pattern and UC Berkeley College of Engineering donated PPE. Thomson Reuters provided our staff with HIPAA training. Initial funding was provided by IGI Founders Fund, Anders Naar, Rick and Rachel Klausner, and the Shurl and Kay Curci Foundation. We also thank Jamie Cate, UC Berkeley CEND Director Julia Schaletsky, Hamilton, Bayer, Zebra, Salesforce, ThirdWave Analytics, Illumina, Active Motif, Riffyn, ACT Catering, USA Scientific, Sartorius, NanoCellect, Dovetail Genomics, Takara Bio, Immudex, Twist Bioscience, Fisher Scientific, 10X Genomics and Novogene for support, encouragement and sustenance.

5.8 Author contributions

All authors contributed equally to this work.

5.9 Supplementary information

Category	Nature of challenge	Solution
Regulatory (CLIA)	A CLIA certificate and license for a diagnostic testing facility is required	Extend CLIA license from UC Berkeley University Health Services to the IGI testing lab
	CLIA regulations require testing personnel to be licensed CLS	Regulations are temporarily revised, allowing non-CLS scientists to act as testing personnel with revised training requirements. IGI created an accelerated in-house CLIA training program.

	Testing personnel must be proficiency tested	Test blinded contrived specimens provided by the American Proficiency Institute
Regulatory (FDA/CLIA)	Validation study must be performed to determine LOD of LDT	Test contrived specimens with synthetic viral RNA in collection medium (semi-automated) and in clinical matrix (automated)
	Clinical validation of LDT must be performed	Test panel of specimens resulted as negative and positive by a CLIA-certified laboratory with an issued EUA for a SARS-CoV-2 test. If specimens unavailable, use contrived ones in clinical matrix.
	An initial batch of negative and positive specimens tested by the LDT must be tested by a third party	Submit the first 5 negative and 5 positive specimens tested under the LDT to a CLIA laboratory with an issued EUA for a SARS-CoV-2 LDT
Regulatory (HIPAA)	Privacy and security of PHI must be consistently maintained during data handling	Develop and implement a LIMS that meets HIPAA standards for PHI privacy and security; restrict access to PHI within the LIMS to authorized personnel
	Testing lab and other personnel must comply with HIPAA regulations	Ensure all personnel complete an online HIPAA training class and pass the end-of-class assessment, with certificate of training placed on permanent record
Biosafety	Specimens must be processed in a manner that complies with institutional biosafety regulations	Obtain a BUA from the university CLEB that defines how specimens are to be processed safely and what PPE the testing personnel must wear
	Measures must be taken to reduce the possibility of SARS-CoV-2 transmission between testing team members	Require all personnel to take a self-assessment questionnaire daily before entering the building where the diagnostic lab is located
	Concerns about bringing samples with live virus on site	Develop a customized sample collection kit that uses a deactivating sample transport medium (DNA/RNA Shield)
Healthcare partners	Challenges of serving non-campus patients	Build a physician portal into the LIMS that can be accessed by non-UC Berkeley physicians. Partner with local healthcare providers to perform swabbing of community members.
	Testing kits must be made available to UHS and to third-party clinicians	Establish a kit assembly SOP. Employ a dedicated kit assembly team to generate sufficient kits on a weekly basis. Use a professional courier service to deliver the kits to the testing sites. Establish a kit use SOP and provide it to the clinicians. Establish diagnostic specimen return SOPs for every partner site.
	The physicians requisitioning the test must receive the results in a CLIA- and HIPAA-compliant manner	As per State of California regulations, CLS report all positive results within 24 h to the requisitioning physician via a direct phone call. All test results are subsequently accessible through our CLIA- and HIPAA-compliant clinician portal.
Supply chain	A sustained supply of specimen collection components is necessary	Identify a tube and swab manufacturer with sufficient supply, and configure the LDT SOPs around it
	A sustained supply of diagnostic testing reagents and disposables is necessary	Identify a supplier (Thermo Fisher) with robust reagent production capacity; develop an LDT that uses half-reactions

Table 5.1: Summary of challenges for establishment of the IGI SARS-CoV-2 testing laboratory

BIBLIOGRAPHY

1. Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* **315**, 1709–1712 (2007).
2. Hille, F. *et al.* The Biology of CRISPR-Cas: Backward and Forward. *Cell* **172**, 1239–1259 (2018).
3. Knott, G. J. & Doudna, J. A. CRISPR-Cas guides the future of genetic engineering. *Science* **361**, 866–869 (2018).
4. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).
5. Koonin, E. V., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol* **37**, 67–78 (2017).
6. Makarova, K. S. *et al.* Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol* **18**, 67–83 (2020).
7. Wang, J. Y., Pausch, P. & Doudna, J. A. Structural biology of CRISPR–Cas immunity and genome editing enzymes. *Nat Rev Microbiol* **20**, 641–656 (2022).
8. Liu, T. Y. & Doudna, J. A. Chemistry of Class 1 CRISPR-Cas effectors: Binding, editing, and regulation. *J Biol Chem* **295**, 14473–14487 (2020).
9. Yoshimi, K. & Mashimo, T. Genome editing technology and applications with the type I CRISPR system. *Gene and Genome Editing* **3–4**, 100013 (2022).
10. Dolan, A. E. *et al.* Introducing a Spectrum of Long-Range Genomic Deletions in Human Embryonic Stem Cells Using Type I CRISPR-Cas. *Molecular Cell* **74**, 936-950.e5 (2019).
11. Cameron, P. *et al.* Harnessing type I CRISPR–Cas systems for genome engineering in human cells. *Nat Biotechnol* **37**, 1471–1477 (2019).
12. Kolesnik, M. V., Fedorova, I., Karneyeva, K. A., Artamonova, D. N. & Severinov, K. V. Type III CRISPR-Cas Systems: Deciphering the Most Complex Prokaryotic Immune System. *Biochemistry (Mosc)* **86**, 1301–1314 (2021).
13. Colognori, D., Trinidad, M. & Doudna, J. A. Precise transcript targeting by CRISPR-Csm complexes. *Nat Biotechnol* 1–9 (2023) doi:10.1038/s41587-022-01649-9.
14. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).
15. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *eLife* **2**, e00471 (2013).
16. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819–823 (2013).
17. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* **339**, 823–826 (2013).
18. Zetsche, B. *et al.* Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163**, 759–771 (2015).
19. Chen, J. S. *et al.* CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360**, 436–439 (2018).
20. Pausch, P. *et al.* CRISPR-CasΦ from huge phages is a hypercompact genome editor. *Science* **369**, 333 (2020).
21. Burstein, D. *et al.* New CRISPR–Cas systems from uncultivated microbes. *Nature* **542**, 237–241 (2017).
22. Harrington, L. B. *et al.* Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science* **362**, 839–842 (2018).
23. Yan, W. X. *et al.* Functionally diverse type V CRISPR-Cas systems. *Science* **363**, 88–91 (2019).
24. Abudayyeh, O. O. *et al.* C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573 (2016).
25. Konermann, S. *et al.* Transcriptome engineering with RNA-targeting Type VI-D CRISPR effectors. *Cell* **173**, 665-676.e14 (2018).
26. Abudayyeh, O. O. *et al.* RNA targeting with CRISPR–Cas13. *Nature* **550**, 280–284 (2017).
27. van Beljouw, S. P. B., Sanders, J., Rodríguez-Molina, A. & Brouns, S. J. J. RNA-targeting CRISPR–Cas systems. *Nat Rev Microbiol* **21**, 21–34 (2023).
28. Wienert, B. *et al.* Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. *Science* **364**, 286–289 (2019).

29. Liu, J.-J. *et al.* CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature* **566**, 218–223 (2019).
30. Yeh, C. D., Richardson, C. D. & Corn, J. E. Advances in genome editing through control of DNA repair pathways. *Nature Cell Biology* **21**, 1468–1478 (2019).
31. Xue, C. & Greene, E. C. DNA repair pathway choices in CRISPR-Cas9 mediated genome editing. *Trends Genet* **37**, 639–656 (2021).
32. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nature Biotechnology* **34**, 339–344 (2016).
33. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).
34. Knight, S. C. *et al.* Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science* **350**, 823–826 (2015).
35. Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat Rev Mol Cell Biol* **18**, 495–506 (2017).
36. Song, B., Yang, S., Hwang, G.-H., Yu, J. & Bae, S. Analysis of NHEJ-Based DNA Repair after CRISPR-Mediated DNA Cleavage. *Int J Mol Sci* **22**, 6397 (2021).
37. Nakade, S. *et al.* Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nat Commun* **5**, 5560 (2014).
38. Iyer, S. *et al.* Precise therapeutic gene correction by a simple nuclease-induced double-strand break. *Nature* **568**, 561–565 (2019).
39. Jasin, M. & Rothstein, R. Repair of Strand Breaks by Homologous Recombination. *Cold Spring Harb Perspect Biol* **5**, a012740 (2013).
40. Liu, M. *et al.* Methodologies for Improving HDR Efficiency. *Frontiers in Genetics* **9**, (2019).
41. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3**, e04766 (2014).
42. Maruyama, T. *et al.* Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat Biotechnol* **33**, 538–542 (2015).
43. Doudna, J. A. The promise and challenge of therapeutic genome editing. *Nature* **578**, 229–236 (2020).
44. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat Biotechnol* **38**, 824–844 (2020).
45. Porto, E. M., Komor, A. C., Slaymaker, I. M. & Yeo, G. W. Base editing: advances and therapeutic opportunities. *Nat Rev Drug Discov* **19**, 839–859 (2020).
46. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat Rev Genet* **19**, 770–788 (2018).
47. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
48. Gaudelli, N. M. *et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
49. Richter, M. F. *et al.* Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat Biotechnol* **38**, 883–891 (2020).
50. Gaudelli, N. M. *et al.* Directed evolution of adenine base editors with increased activity and therapeutic application. *Nat Biotechnol* **38**, 892–900 (2020).
51. Lam, D. K. *et al.* Improved cytosine base editors generated from TadA variants. *Nat Biotechnol* 1–12 (2023) doi:10.1038/s41587-022-01611-9.
52. Rees, H. A. *et al.* Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat Commun* **8**, 15790 (2017).
53. Clement, K. *et al.* CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol* **37**, 224–226 (2019).
54. Zuo, E. *et al.* Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **364**, 289–292 (2019).
55. Doman, J. L., Raguram, A., Newby, G. A. & Liu, D. R. Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nat Biotechnol* **38**, 620–628 (2020).
56. Yu, Y. *et al.* Cytosine base editors with minimized unguided DNA and RNA off-target events and high on-target activity. *Nat Commun* **11**, 2052 (2020).

57. Grünewald, J. *et al.* Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**, 433–437 (2019).
58. Grünewald, J. *et al.* CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nature Biotechnology* **37**, 1041–1048 (2019).
59. Rees, H. A., Wilson, C., Doman, J. L. & Liu, D. R. Analysis and minimization of cellular RNA editing by DNA adenine base editors. *Sci Adv* **5**, eaax5717 (2019).
60. Lapteva, L., Purohit-Sheth, T., Serabian, M. & Puri, R. K. Clinical Development of Gene Therapies: The First Three Decades and Counting. *Mol Ther Methods Clin Dev* **19**, 387–397 (2020).
61. Milone, M. C. & O'Doherty, U. Clinical use of lentiviral vectors. *Leukemia* **32**, 1529–1541 (2018).
62. Themis, M. *et al.* Oncogenesis Following Delivery of a Nonprimate Lentiviral Gene Therapy Vector to Fetal and Neonatal Mice. *Molecular Therapy* **12**, 763–771 (2005).
63. Hacein-Bey-Abina, S. *et al.* Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J Clin Invest* **118**, 3132–3142 (2008).
64. Roth, T. L. *et al.* Reprogramming human T cell function and specificity with non-viral genome targeting. *Nature* **559**, 405–409 (2018).
65. DeWitt, M. A. *et al.* Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells. *Sci. Transl. Med.* **8**, (2016).
66. Foy, S. P. *et al.* Non-viral precision T cell receptor replacement for personalized cell therapy. *Nature* 1–3 (2022) doi:10.1038/s41586-022-05531-1.
67. Zhang, J. *et al.* Non-viral, specifically targeted CAR-T cells achieve high safety and efficacy in B-NHL. *Nature* 1–6 (2022) doi:10.1038/s41586-022-05140-y.
68. Peters, J. E., Makarova, K. S., Shmakov, S. & Koonin, E. V. Recruitment of CRISPR-Cas systems by Tn7-like transposons. *Proc Natl Acad Sci U S A* **114**, E7358–E7366 (2017).
69. Faure, G. *et al.* CRISPR–Cas in mobile genetic elements: counter-defence and beyond. *Nat Rev Microbiol* **17**, 513–525 (2019).
70. Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S. & Sternberg, S. H. Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature* **571**, 219–225 (2019).
71. Strecker, J. *et al.* RNA-guided DNA insertion with CRISPR-associated transposases. *Science* **365**, 48–53 (2019).
72. Rubin, B. E. *et al.* Species- and site-specific genome editing in complex bacterial communities. *Nat Microbiol* **7**, 34–47 (2022).
73. Lampe, G. D. *et al.* Targeted DNA integration in human cells without double-strand breaks using CRISPR-associated transposases. *Nat Biotechnol* (2023) doi:10.1038/s41587-023-01748-1.
74. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
75. Anzalone, A. V. *et al.* Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. *Nat Biotechnol* **40**, 731–740 (2022).
76. Choi, J. *et al.* Precise genomic deletions using paired prime editing. *Nat Biotechnol* **40**, 218–226 (2022).
77. Böck, D. *et al.* In vivo prime editing of a metabolic liver disease in mice. *Sci Transl Med* **14**, eabl9238 (2022).
78. Everette, K. A. *et al.* Ex vivo prime editing of patient haematopoietic stem cells rescues sickle-cell disease phenotypes after engraftment in mice. *Nat. Biomed. Eng* 1–13 (2023) doi:10.1038/s41551-023-01026-0.
79. Chemello, F. *et al.* Precise correction of Duchenne muscular dystrophy exon deletion mutations by base and prime editing. *Sci Adv* **7**, eabg4910 (2021).
80. Xu, X. & Qi, L. S. A CRISPR–dCas Toolbox for Genetic Engineering and Synthetic Biology. *Journal of Molecular Biology* **431**, 34–47 (2019).
81. Nakamura, M., Gao, Y., Dominguez, A. A. & Qi, L. S. CRISPR technologies for precise epigenome editing. *Nat Cell Biol* **23**, 11–22 (2021).
82. Gilbert, L. A. *et al.* CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell* **154**, 442–451 (2013).
83. Thakore, P. I. *et al.* Highly Specific Epigenome Editing by CRISPR/Cas9 Repressors for Silencing of Distal Regulatory Elements. *Nat Methods* **12**, 1143–1149 (2015).
84. Amabile, A. *et al.* Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. *Cell* **167**, 219-232.e14 (2016).

85. Nuñez, J. K. *et al.* Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. *Cell* **184**, 2503-2519.e17 (2021).
86. Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nat Methods* **12**, 326–328 (2015).
87. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2015).
88. 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* **159**, 635–646 (2014).
89. Porteus, M. H. A New Class of Medicines through DNA Editing. *N Engl J Med* **380**, 947–959 (2019).
90. Copelan, E. A. Hematopoietic Stem-Cell Transplantation. *N Engl J Med* **354**, 1813–1826 (2006).
91. Dever, D. P. *et al.* CRISPR/Cas9 β-globin gene targeting in human haematopoietic stem cells. *Nature* **539**, 384–389 (2016).
92. Bauer, D. E. *et al.* An Erythroid Enhancer of BCL11A Subject to Genetic Variation Determines Fetal Hemoglobin Level. *Science* **342**, 253–257 (2013).
93. Wu, Y. *et al.* Highly efficient therapeutic gene editing of human hematopoietic stem cells. *Nat Med* **25**, 776–783 (2019).
94. Canver, M. C. *et al.* BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **527**, 192–197 (2015).
95. Frangoul, H. *et al.* CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. *New England Journal of Medicine* **384**, 252–260 (2021).
96. Fu, B. *et al.* CRISPR–Cas9-mediated gene editing of the BCL11A enhancer for pediatric β0/β0 transfusion-dependent β-thalassemia. *Nat Med* **28**, 1573–1580 (2022).
97. Kingwell, K. First CRISPR therapy seeks landmark approval. *Nature Reviews Drug Discovery* (2023) doi:10.1038/d41573-023-00050-8.
98. Melenhorst, J. J. *et al.* Decade-long leukaemia remissions with persistence of CD4+ CAR T cells. *Nature* **602**, 503–509 (2022).
99. Lim, W. A. & June, C. H. The Principles of Engineering Immune Cells to Treat Cancer. *Cell* **168**, 724–740 (2017).
100. Sterner, R. C. & Sterner, R. M. CAR-T cell therapy: current limitations and potential strategies. *Blood Cancer J.* **11**, 1–11 (2021).
101. Stadtmauer, E. A. *et al.* CRISPR-engineered T cells in patients with refractory cancer. *Science* **367**, eaba7365 (2020).
102. Diorio, C. *et al.* Cytosine base editing enables quadruple-edited allogeneic CART cells for T-ALL. *Blood* **140**, 619–629 (2022).
103. Webber, B. R. *et al.* Highly efficient multiplex human T cell engineering without double-strand breaks using Cas9 base editors. *Nat Commun* **10**, 5222 (2019).
104. Georgiadis, C. *et al.* Base-edited CAR T cells for combinational therapy against T cell malignancies. *Leukemia* **35**, 3466–3481 (2021).
105. Huang, Q., Zhang, X. & Liu, D. CAR T cell therapy for T cell leukemia and lymphoma: latest updates from 2022 ASH Annual Meeting. *J Hematol Oncol* **16**, 1–3 (2023).
106. Chiesa, R. *et al.* Tvt CAR7: Phase 1 Clinical Trial of Base-Edited Universal" CAR7 T Cells for Paediatric Relapsed/Refractory T-ALL. *Blood* **140**, 4579–4580 (2022).
107. Eyquem, J. *et al.* Targeting a CAR to the TRAC locus with CRISPR/Cas9 enhances tumour rejection. *Nature* **543**, 113–117 (2017).
108. Zhang, Y.-N., Poon, W., Tavares, A. J., McGilvray, I. D. & Chan, W. C. W. Nanoparticle-liver interactions: Cellular uptake and hepatobiliary elimination. *J Control Release* **240**, 332–348 (2016).
109. Gillmore, J. D. *et al.* CRISPR-Cas9 In Vivo Gene Editing for Transthyretin Amyloidosis. *New England Journal of Medicine* **385**, 493–502 (2021).
110. Sequence Variations in PCSK9, Low LDL, and Protection against Coronary Heart Disease | NEJM. https://www.nejm.org/doi/10.1056/NEJMoa054013?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%200www.ncbi.nlm.nih.gov.
111. Musunuru, K. *et al.* In vivo CRISPR base editing of PCSK9 durably lowers cholesterol in primates. *Nature* **593**, 429–434 (2021).

112. Lee, R. G. *et al.* Efficacy and Safety of an Investigational Single-Course CRISPR Base-Editing Therapy Targeting PCSK9 in Nonhuman Primate and Mouse Models. *Circulation* **147**, 242–253 (2023).
113. Thomas, C. E., Ehrhardt, A. & Kay, M. A. Progress and problems with the use of viral vectors for gene therapy. *Nat Rev Genet* **4**, 346–358 (2003).
114. Milone, M. C. & O'Doherty, U. Clinical use of lentiviral vectors. *Leukemia* **32**, 1529–1541 (2018).
115. Chew, W. L. *et al.* A multifunctional AAV–CRISPR–Cas9 and its host response. *Nat Methods* **13**, 868–874 (2016).
116. Charlesworth, C. T. *et al.* Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nat Med* **25**, 249–254 (2019).
117. Wagner, D. L. *et al.* High prevalence of Streptococcus pyogenes Cas9-reactive T cells within the adult human population. *Nat Med* **25**, 242–248 (2019).
118. Miller, D. G., Petek, L. M. & Russell, D. W. Adeno-associated virus vectors integrate at chromosome breakage sites. *Nat Genet* **36**, 767–773 (2004).
119. Hanlon, K. S. *et al.* High levels of AAV vector integration into CRISPR-induced DNA breaks. *Nat Commun* **10**, 4439 (2019).
120. Leader, B., Baca, Q. J. & Golan, D. E. Protein therapeutics: a summary and pharmacological classification. *Nat Rev Drug Discov* **7**, 21–39 (2008).
121. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J.-S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res* **24**, 1012–1019 (2014).
122. Wilson, R. C. & Carroll, D. The Daunting Economics of Therapeutic Genome Editing. *The CRISPR Journal* **2**, 280–284 (2019).
123. Mandal, P. K. *et al.* Efficient Ablation of Genes in Human Hematopoietic Stem and Effector Cells using CRISPR/Cas9. *Cell Stem Cell* **15**, 643–652 (2014).
124. Schumann, K. *et al.* Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. *Proceedings of the National Academy of Sciences* **112**, 10437–10442 (2015).
125. Nguyen, D. N. *et al.* Polymer-stabilized Cas9 nanoparticles and modified repair templates increase genome editing efficiency. *Nat Biotechnol* **38**, 44–49 (2020).
126. Hendel, A. *et al.* Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat Biotechnol* **33**, 985–989 (2015).
127. Fogarty, N. M. E. *et al.* Genome editing reveals a role for OCT4 in human embryogenesis. *Nature* **550**, 67–73 (2017).
128. Hashimoto, M., Yamashita, Y. & Takemoto, T. Electroporation of Cas9 protein/sgRNA into early pronuclear zygotes generates non-mosaic mutants in the mouse. *Developmental Biology* **418**, 1–9 (2016).
129. Chen, S., Lee, B., Lee, A. Y.-F., Modzelewski, A. J. & He, L. Highly Efficient Mouse Genome Editing by CRISPR Ribonucleoprotein Electroporation of Zygotes*. *Journal of Biological Chemistry* **291**, 14457–14467 (2016).
130. Wang, W. *et al.* Delivery of Cas9 Protein into Mouse Zygotes through a Series of Electroporation Dramatically Increases the Efficiency of Model Creation. *Journal of Genetics and Genomics* **43**, 319–327 (2016).
131. Remy, S. *et al.* Generation of gene-edited rats by delivery of CRISPR/Cas9 protein and donor DNA into intact zygotes using electroporation. *Sci Rep* **7**, 16554 (2017).
132. De Plano, L. M. *et al.* Applications of CRISPR-Cas9 in Alzheimer's Disease and Related Disorders. *Int J Mol Sci* **23**, 8714 (2022).
133. Terstappen, G. C., Meyer, A. H., Bell, R. D. & Zhang, W. Strategies for delivering therapeutics across the blood–brain barrier. *Nat Rev Drug Discov* **20**, 362–383 (2021).
134. Staahl, B. T. *et al.* Efficient genome editing in the mouse brain by local delivery of engineered Cas9 ribonucleoprotein complexes. *Nature Biotechnology* **35**, 431–434 (2017).
135. Sellers, D. L., Lee, K., Murthy, N. & Pun, S. H. TAxI-peptide targeted Cas12a ribonuclease protein nanoformulations increase genome editing in hippocampal neurons. *Journal of Controlled Release* **354**, 188–195 (2023).
136. He, M. *et al.* A traceless linker for aliphatic amines that rapidly and quantitatively fragments after reduction. *Chemical Science* **11**, 8973–8980 (2020).
137. Lee, K. *et al.* Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. *Nat Biomed Eng* **1**, 889–901 (2017).

138. Lee, B. *et al.* Nanoparticle delivery of CRISPR into the brain rescues a mouse model of fragile X syndrome from exaggerated repetitive behaviours. *Nat Biomed Eng* **2**, 497–507 (2018).
139. Chen, G. *et al.* A biodegradable nanocapsule delivers a Cas9 ribonucleoprotein complex for in vivo genome editing. *Nat Nanotechnol* **14**, 974–980 (2019).
140. Chae, S.-Y. *et al.* Rationally designed nanoparticle delivery of Cas9 ribonucleoprotein for effective gene editing. *Journal of Controlled Release* **345**, 108–119 (2022).
141. Lee, Y.-W. *et al.* In Vivo Editing of Macrophages through Systemic Delivery of CRISPR-Cas9-Ribonucleoprotein-Nanoparticle Nanoassemblies. *Adv Ther (Weinh)* **2**, 1900041 (2019).
142. Iyer, A. K., Khaled, G., Fang, J. & Maeda, H. Exploiting the enhanced permeability and retention effect for tumor targeting. *Drug Discovery Today* **11**, 812–818 (2006).
143. Liu, Q. *et al.* NanoRNP Overcomes Tumor Heterogeneity in Cancer Treatment. *Nano Lett.* **19**, 7662–7672 (2019).
144. Li, X. *et al.* Hypoxia-Responsive Gene Editing to Reduce Tumor Thermal Tolerance for Mild-Photothermal Therapy. *Angewandte Chemie International Edition* **60**, 21200–21204 (2021).
145. Wan, T. *et al.* Genome editing of mutant KRAS through supramolecular polymer-mediated delivery of Cas9 ribonucleoprotein for colorectal cancer therapy. *Journal of Controlled Release* **322**, 236–247 (2020).
146. Deng, S. *et al.* Codelivery of CRISPR-Cas9 and chlorin e6 for spatially controlled tumor-specific gene editing with synergistic drug effects. *Science Advances* **6**, eabb4005 (2020).
147. Kazemian, P. *et al.* Lipid-Nanoparticle-Based Delivery of CRISPR/Cas9 Genome-Editing Components. *Mol. Pharmaceutics* **19**, 1669–1686 (2022).
148. Dilliard, S. A. & Siegwart, D. J. Passive, active and endogenous organ-targeted lipid and polymer nanoparticles for delivery of genetic drugs. *Nat Rev Mater* **8**, 282–300 (2023).
149. Baden, L. R. *et al.* Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *N Engl J Med* **384**, 403–416 (2021).
150. Polack, F. P. *et al.* Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N Engl J Med* **383**, 2603–2615 (2020).
151. Warne, N. *et al.* Delivering 3 billion doses of Comirnaty in 2021. *Nat Biotechnol* **41**, 183–188 (2023).
152. Hafez, I. M., Maurer, N. & Cullis, P. R. On the mechanism whereby cationic lipids promote intracellular delivery of polynucleic acids. *Gene Ther* **8**, 1188–1196 (2001).
153. Qiu, M. *et al.* Lipid nanoparticle-mediated codelivery of Cas9 mRNA and single-guide RNA achieves liver-specific in vivo genome editing of Angptl3. *Proceedings of the National Academy of Sciences* **118**, e2020401118 (2021).
154. Han, J. P. *et al.* In vivo delivery of CRISPR-Cas9 using lipid nanoparticles enables antithrombin gene editing for sustainable hemophilia A and B therapy. *Sci Adv* **8**, eabj6901 (2022).
155. Rothgangl, T. *et al.* In vivo adenine base editing of PCSK9 in macaques reduces LDL cholesterol levels. *Nat Biotechnol* **39**, 949–957 (2021).
156. Finn, J. D. *et al.* A Single Administration of CRISPR/Cas9 Lipid Nanoparticles Achieves Robust and Persistent In Vivo Genome Editing. *Cell Reports* **22**, 2227–2235 (2018).
157. O'Sullivan, J. *et al.* Beyond GalNAc! Drug delivery systems comprising complex oligosaccharides for targeted use of nucleic acid therapeutics. *RSC Adv* **12**, 20432–20446.
158. Balwani, M. *et al.* Phase 3 Trial of RNAi Therapeutic Givosiran for Acute Intermittent Porphyria. *N Engl J Med* **382**, 2289–2301 (2020).
159. Kasiewicz, L. N. *et al.* Lipid nanoparticles incorporating a GalNAc ligand enable in vivo liver ANGPTL3 editing in wild-type and somatic LDLR knockout non-human primates. 2021.11.08.467731 Preprint at https://doi.org/10.1101/2021.11.08.467731 (2021).
160. Cheng, Q. *et al.* Selective organ targeting (SORT) nanoparticles for tissue-specific mRNA delivery and CRISPR–Cas gene editing. *Nat. Nanotechnol.* **15**, 313–320 (2020).
161. Liu, S. *et al.* Membrane-destabilizing ionizable phospholipids for organ-selective mRNA delivery and CRISPR–Cas gene editing. *Nat. Mater.* **20**, 701–710 (2021).
162. Wei, T., Cheng, Q., Min, Y.-L., Olson, E. N. & Siegwart, D. J. Systemic nanoparticle delivery of CRISPR-Cas9 ribonucleoproteins for effective tissue specific genome editing. *Nat Commun* **11**, 3232 (2020).

163. Dilliard, S. A., Cheng, Q. & Siegwart, D. J. On the mechanism of tissue-specific mRNA delivery by selective organ targeting nanoparticles. *Proceedings of the National Academy of Sciences* **118**, e2109256118 (2021).
164. Dahlman, J. E. *et al.* Barcoded nanoparticles for high throughput in vivo discovery of targeted therapeutics. *Proceedings of the National Academy of Sciences* **114**, 2060–2065 (2017).
165. Li, B. *et al.* Combinatorial design of nanoparticles for pulmonary mRNA delivery and genome editing. *Nat Biotechnol* 1–6 (2023) doi:10.1038/s41587-023-01679-x.
166. Choi, J. G. *et al.* Lentivirus pre-packed with Cas9 protein for safer gene editing. *Gene Ther* **23**, 627–633 (2016).
167. Lyu, P., Javidi-Parsijani, P., Atala, A. & Lu, B. Delivering Cas9/sgRNA ribonucleoprotein (RNP) by lentiviral capsid-based bionanoparticles for efficient ‘hit-and-run’ genome editing. *Nucleic Acids Res* **47**, e99 (2019).
168. Yadav, M., Atala, A. & Lu, B. Developing all-in-one virus-like particles for Cas9 mRNA/single guide RNA co-delivery and aptamer-containing lentiviral vectors for improved gene expression. *International Journal of Biological Macromolecules* **209**, 1260–1270 (2022).
169. Mangeot, P. E. *et al.* Genome editing in primary cells and in vivo using viral-derived Nanoblades loaded with Cas9-sgRNA ribonucleoproteins. *Nat Commun* **10**, 45 (2019).
170. Banskota, S. *et al.* Engineered virus-like particles for efficient in vivo delivery of therapeutic proteins. *Cell* **185**, 250-265.e16 (2022).
171. Hamilton, J. R. *et al.* Targeted delivery of CRISPR-Cas9 and transgenes enables complex immune cell engineering. *Cell Reports* **35**, (2021).
172. Dobson, C. S. *et al.* Antigen identification and high-throughput interaction mapping by reprogramming viral entry. *Nat Methods* **19**, 449–460 (2022).
173. Hamilton, J. R., Chen, E., Perez, B. S., Espinoza, C. R. S. & Doudna, J. A. Cell type-programmable genome editing with enveloped delivery vehicles. 2022.08.24.505004 Preprint at https://doi.org/10.1101/2022.08.24.505004 (2022).
174. Kalluri, R. & LeBleu, V. S. The biology, function, and biomedical applications of exosomes. *Science* **367**, eaau6977 (2020).
175. Osteikoetxea, X. *et al.* Engineered Cas9 extracellular vesicles as a novel gene editing tool. *Journal of Extracellular Vesicles* **11**, e12225 (2022).
176. Ye, Y. *et al.* An engineered exosome for delivering sgRNA:Cas9 ribonucleoprotein complex and genome editing in recipient cells. *Biomater. Sci.* **8**, 2966–2976 (2020).
177. Wan, T. *et al.* Exosome-mediated delivery of Cas9 ribonucleoprotein complexes for tissue-specific gene therapy of liver diseases. *Science Advances* **8**, eabp9435 (2022).
178. Alsaiari, S. K. *et al.* Endosomal Escape and Delivery of CRISPR/Cas9 Genome Editing Machinery Enabled by Nanoscale Zeolitic Imidazolate Framework. *J. Am. Chem. Soc.* **140**, 143–146 (2018).
179. Alyami, M. Z. *et al.* Cell-Type-Specific CRISPR/Cas9 Delivery by Biomimetic Metal Organic Frameworks. *J. Am. Chem. Soc.* **142**, 1715–1720 (2020).
180. Jiang, F. & Doudna, J. A. CRISPR–Cas9 Structures and Mechanisms. *Annual Review of Biophysics* **46**, 505–529 (2017).
181. Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S. & Yang, S.-H. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Molecular Therapy - Nucleic Acids* **4**, e264 (2015).
182. Marraffini, L. A. & Sontheimer, E. J. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463**, 568–571 (2010).
183. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology* **31**, 827–832 (2013).
184. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**, 822–826 (2013).
185. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* **31**, 839–843 (2013).
186. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184–191 (2016).
187. Listgarten, J. *et al.* Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. *Nat Biomed Eng* **2**, 38–47 (2018).

188. Haeussler, M. *et al.* Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biology* **17**, 148 (2016).

189. Tsai, S. Q. *et al.* CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR–Cas9 nuclease off-targets. *Nat Methods* **14**, 607–614 (2017).

190. Horlbeck, M. A. *et al.* Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *eLife* **5**, e12677 (2016).

191. Yarrington, R. M., Verma, S., Schwartz, S., Trautman, J. K. & Carroll, D. Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo. *Proceedings of the National Academy of Sciences* **115**, 9351–9358 (2018).

192. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature Biotechnology* **33**, 187–197 (2015).

193. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* **32**, 279–284 (2014).

194. Kocak, D. D. *et al.* Increasing the specificity of CRISPR systems with engineered RNA secondary structures. *Nat Biotechnol* **37**, 657–666 (2019).

195. Yin, H. *et al.* Partial DNA-guided Cas9 enables genome editing with reduced off-target activity. *Nat Chem Biol* **14**, 311–316 (2018).

196. Donohoue, P. D. *et al.* Conformational control of Cas9 by CRISPR hybrid RNA-DNA guides mitigates off-target activity in T cells. *Molecular Cell* **81**, 3637-3649.e5 (2021).

197. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).

198. Kleinstiver, B. P. *et al.* High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).

199. Chen, J. S. *et al.* Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature* **550**, 407–410 (2017).

200. Lee, J. K. *et al.* Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat Commun* **9**, 3048 (2018).

201. Casini, A. *et al.* A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat Biotechnol* **36**, 265–271 (2018).

202. Kim, N. *et al.* Prediction of the sequence-specific cleavage activity of Cas9 variants. *Nat Biotechnol* **38**, 1328–1336 (2020).

203. Shen, M. W. *et al.* Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* **563**, 646–651 (2018).

204. Chakrabarti, A. M. *et al.* Target-Specific Precision of CRISPR-Mediated Genome Editing. *Mol Cell* **73**, 699-713.e6 (2019).

205. Brunet, E. & Jasin, M. Induction of chromosomal translocations with CRISPR-Cas9 and other nucleases: Understanding the repair mechanisms that give rise to translocations. *Adv Exp Med Biol* **1044**, 15–25 (2018).

206. Richardson, C. & Jasin, M. Frequent chromosomal translocations induced by DNA double-strand breaks. *Nature* **405**, 697–700 (2000).

207. Torres, R. *et al.* Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR–Cas9 system. *Nat Commun* **5**, 3964 (2014).

208. Vanoli, F. *et al.* CRISPR-Cas9–guided oncogenic chromosomal translocations with conditional fusion protein expression in human mesenchymal cells. *Proceedings of the National Academy of Sciences* **114**, 3696–3701 (2017).

209. Samuelson, C. *et al.* Multiplex CRISPR/Cas9 genome editing in hematopoietic stem cells for fetal hemoglobin reinduction generates chromosomal translocations. *Molecular Therapy - Methods & Clinical Development* **23**, 507–523 (2021).

210. Foss, D. V. *et al.* Peptide-mediated delivery of CRISPR enzymes for the efficient editing of primary human lymphocytes. *Nat. Biomed. Eng* 1–14 (2023) doi:10.1038/s41551-023-01032-2.

211. Bothmer, A. *et al.* Detection and Modulation of DNA Translocations During Multi-Gene Genome Editing in T Cells. *CRISPR J* **3**, 177–187 (2020).

212. Leibowitz, M. L. *et al.* Chromothripsis as an on-target consequence of CRISPR–Cas9 genome editing. *Nat Genet* **53**, 895–905 (2021).

213. Adikusuma, F. *et al.* Large deletions induced by Cas9 cleavage. *Nature* **560**, E8–E9 (2018).

214. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat Biotechnol* **36**, 765–771 (2018).
215. Cullot, G. *et al.* CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. *Nat Commun* **10**, 1136 (2019).
216. Nahmad, A. D. *et al.* Frequent aneuploidy in primary human T cells after CRISPR-Cas9 cleavage. *Nat Biotechnol* (2022) doi:10.1038/s41587-022-01377-0.
217. Alanis-Lobato, G. *et al.* Frequent loss of heterozygosity in CRISPR-Cas9–edited early human embryos. *Proceedings of the National Academy of Sciences* **118**, e2004832117 (2021).
218. Papathanasiou, S. *et al.* Whole chromosome loss and genomic instability in mouse embryos after CRISPR-Cas9 genome editing. *Nat Commun* **12**, 5855 (2021).
219. Tsuchida, C. A. *et al.* Mitigation of chromosome loss in clinical CRISPR-Cas9-engineered T cells. *bioRxiv* 2023.03.22.533709 (2023) doi:10.1101/2023.03.22.533709.
220. Mojica, F. J. M. & Rodriguez-Valera, F. The discovery of CRISPR in archaea and bacteria. *The FEBS Journal* **283**, 3162–3169 (2016).
221. Le Rhun, A., Escalera-Maurer, A., Bratovič, M. & Charpentier, E. CRISPR-Cas in Streptococcus pyogenes. *RNA Biology* **16**, 380–389 (2019).
222. Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
223. Wright, A. V., Nuñez, J. K. & Doudna, J. A. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* **164**, 29–44 (2016).
224. Cao, C. *et al.* A CRISPR/dCasX-mediated transcriptional programming system for inhibiting the progression of bladder cancer cells by repressing c-MYC or activating TP53. *Clin Transl Med* **11**, e537 (2021).
225. Roberson, E. D. O. A catalog of CasX genome editing sites in common model organisms. *BMC Genomics* **20**, 528 (2019).
226. Yang, H. & Patel, D. J. CasX: a new and small CRISPR gene-editing protein. *Cell Res* **29**, 345–346 (2019).
227. Adamcik, J., Jeon, J.-H., Karczewski, K. J., Metzler, R. & Dietler, G. Quantifying supercoiling-induced denaturation bubbles in DNA. *Soft Matter* **8**, 8651–8658 (2012).
228. Yamano, T. *et al.* Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell* **165**, 949–962 (2016).
229. Yang, H., Gao, P., Rajashankar, K. R. & Patel, D. J. PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease. *Cell* **167**, 1814-1828.e12 (2016).
230. Holm, L. & Laakso, L. M. Dali server update. *Nucleic Acids Research* **44**, W351–W355 (2016).
231. Li, L. *et al.* HOLMESv2: A CRISPR-Cas12b-Assisted Platform for Nucleic Acid Detection and DNA Methylation Quantitation. *ACS Synth. Biol.* **8**, 2228–2237 (2019).
232. Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186–191 (2015).
233. Kim, D. Y. *et al.* Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by adeno-associated virus. *Nat Biotechnol* **40**, 94–102 (2022).
234. Kuzmin, D. A. *et al.* The clinical landscape for AAV gene therapies. *Nature Reviews Drug Discovery* **20**, 173–174 (2021).
235. Samulski, R. J. & Muzyczka, N. AAV-Mediated Gene Therapy for Research and Therapeutic Purposes. *Annual Review of Virology* **1**, 427–451 (2014).
236. Wang, D., Zhang, F. & Gao, G. CRISPR-Based Therapeutic Genome Editing: Strategies and In Vivo Delivery by AAV Vectors. *Cell* **181**, 136–150 (2020).
237. Kleinstiver, B. P. *et al.* Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat Biotechnol* **37**, 276–282 (2019).
238. Li, X. *et al.* Base editing with a Cpf1–cytidine deaminase fusion. *Nature Biotechnology* **36**, 324–327 (2018).
239. Crudele, J. M. & Chamberlain, J. S. Cas9 immunity creates challenges for CRISPR gene editing therapies. *Nat Commun* **9**, 3497 (2018).
240. Zhang, L. *et al.* Systematic in vitro profiling of off-target affinity, cleavage and efficiency for CRISPR enzymes. *Nucleic Acids Research* **48**, 5037–5053 (2020).
241. Wu, Z. *et al.* Programmed genome editing by a miniature CRISPR-Cas12f nuclease. *Nat Chem Biol* **17**, 1132–1138 (2021).

242. Xu, X. *et al.* Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. *Molecular Cell* **81**, 4333-4345.e4 (2021).
243. Takeda, S. N. *et al.* Structure of the miniature type V-F CRISPR-Cas effector enzyme. *Molecular Cell* **81**, 558-570.e3 (2021).
244. Xiao, R., Li, Z., Wang, S., Han, R. & Chang, L. Structural basis for substrate recognition and cleavage by the dimerization-dependent CRISPR–Cas12f nuclease. *Nucleic Acids Research* **49**, 4120–4128 (2021).
245. Patel, A., Toso, D., Litvak, A. & Nogales, E. Efficient graphene oxide coating improves cryo-EM sample preparation and data collection from tilted grids.
246. Mastronarde, D. N. SerialEM: A Program for Automated Tilt Series Acquisition on Tecnai Microscopes Using Prediction of Specimen Position. *Microsc Microanal* **9**, 1182–1183 (2003).
247. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331–332 (2017).
248. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290–296 (2017).
249. Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife* **5**, e18722 (2016).
250. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Crystallogr D Struct Biol* **75**, 861–877 (2019).
251. Chen, V. B. *et al. MolProbity*: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12–21 (2010).
252. Trabuco, L. G., Villa, E., Mitra, K., Frank, J. & Schulten, K. Flexible Fitting of Atomic Structures into Electron Microscopy Maps Using Molecular Dynamics. *Structure* **16**, 673–683 (2008).
253. Casañal, A., Lohkamp, B. & Emsley, P. Current developments in Coot for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. *Protein Science* **29**, 1055–1064 (2020).
254. Bae, S., Park, J. & Kim, J.-S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).
255. Lu, Y. *et al.* Safety and feasibility of CRISPR-edited T cells in patients with refractory non-small-cell lung cancer. *Nat Med* **26**, 732–740 (2020).
256. Mackensen, A. *et al.* Anti-CD19 CAR T cell therapy for refractory systemic lupus erythematosus. *Nat Med* **28**, 2124–2132 (2022).
257. Mougiakakos, D. *et al.* CD19-Targeted CAR T Cells in Refractory Systemic Lupus Erythematosus. *New England Journal of Medicine* **385**, 567–569 (2021).
258. Bendle, G. M. *et al.* Lethal graft-versus-host disease in mouse models of T cell receptor gene therapy. *Nat Med* **16**, 565–570, 1p following 570 (2010).
259. Depil, S., Duchateau, P., Grupp, S. A., Mufti, G. & Poirot, L. 'Off-the-shelf' allogeneic CAR T cells: development and challenges. *Nat Rev Drug Discov* **19**, 185–199 (2020).
260. McGinnis, C. S. *et al.* MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat Methods* **16**, 619–626 (2019).
261. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
262. Shibata, M. *et al.* Real-space and real-time dynamics of CRISPR-Cas9 visualized by high-speed atomic force microscopy. *Nat Commun* **8**, 1430 (2017).
263. Datlinger, P. *et al.* Pooled CRISPR screening with single-cell transcriptome readout. *Nature Methods* **14**, 297–301 (2017).
264. Hill, A. J. *et al.* On the design of CRISPR-based single-cell molecular screens. *Nat Methods* **15**, 271–274 (2018).
265. Shifrut, E. *et al.* Genome-wide CRISPR Screens in Primary Human T Cells Reveal Key Regulators of Immune Function. *Cell* (2018) doi:10.1016/j.cell.2018.10.024.
266. Hou, P. *et al.* Genome editing of CXCR4 by CRISPR/cas9 confers cells resistant to HIV-1 infection. *Sci Rep* **5**, 15577 (2015).
267. Deuse, T. *et al.* Hypoimmunogenic derivatives of induced pluripotent stem cells evade immune rejection in fully immunocompetent allogeneic recipients. *Nat Biotechnol* **37**, 252–258 (2019).
268. Dabrowska, M., Juzwa, W., Krzyzosiak, W. J. & Olejniczak, M. Precise Excision of the CAG Tract from the Huntingtin Gene by Cas9 Nickases. *Frontiers in Neuroscience* **12**, (2018).

269. Odate, S., Strapps, W. & Lescarbeau, R. M. Compositions and methods for treating alpha-1 antitrypsin deficiency. (2023).

270. Schiroli, G. *et al.* Precise Gene Editing Preserves Hematopoietic Stem Cell Function following Transient p53-Mediated DNA Damage Response. *Cell Stem Cell* **24**, 551-565.e8 (2019).

271. Mao, Z., Bozzella, M., Seluanov, A. & Gorbunova, V. Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair (Amst)* **7**, 1765–1771 (2008).

272. Rose, J. C. *et al.* Rapidly inducible Cas9 and DSB-ddPCR to probe editing kinetics. *Nat Methods* **14**, 891–896 (2017).

273. Shy, B. R. *et al.* High-yield genome engineering in primary cells using a hybrid ssDNA repair template and small-molecule cocktails. *Nat Biotechnol* 1–11 (2022) doi:10.1038/s41587-022-01418-8.

274. Eyquem, J. *et al.* Targeting a CAR to the *TRAC* locus with CRISPR/Cas9 enhances tumour rejection. *Nature* **543**, 113–117 (2017).

275. Brentjens, R. J. *et al.* CD19-Targeted T Cells Rapidly Induce Molecular Remissions in Adults with Chemotherapy-Refractory Acute Lymphoblastic Leukemia. *Science Translational Medicine* **5**, 177ra38-177ra38 (2013).

276. Watanabe, M., Moon, K. D., Vacchio, M. S., Hathcock, K. S. & Hodes, R. J. Downmodulation of Tumor Suppressor p53 by T Cell Receptor Signaling Is Critical for Antigen-Specific CD4+ T Cell Responses. *Immunity* **40**, 681–691 (2014).

277. Li, M. *et al.* The ATM–p53 pathway suppresses aneuploidy-induced tumorigenesis. *Proceedings of the National Academy of Sciences* **107**, 14188–14193 (2010).

278. Svoboda, J. *et al.* Interleukin-18 Secreting Autologous Anti-CD19 CAR T-Cells (huCART19-IL18) in Patients with Non-Hodgkin Lymphomas Relapsed or Refractory to Prior CAR T-Cell Therapy. *Blood* **140**, 4612–4614 (2022).

279. Gómez-González, B. & Aguilera, A. Transcription-mediated replication hindrance: a major driver of genome instability. *Genes Dev.* **33**, 1008–1026 (2019).

280. Xin, C. *et al.* Comprehensive assessment of miniature CRISPR-Cas12f nucleases for gene disruption. *Nat Commun* **13**, 5623 (2022).

281. Yin, J. *et al.* Cas9 exo-endonuclease eliminates chromosomal translocations during genome editing. *Nat Commun* **13**, 1204 (2022).

282. Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biology* **15**, 554 (2014).

283. Kim, M. C. *et al.* memento: Generalized differential expression analysis of single-cell RNA-seq with method of moments estimation and efficient resampling. 2022.11.09.515836 Preprint at https://doi.org/10.1101/2022.11.09.515836 (2022).

284. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **28**, 27–30 (2000).

285. Fang, Z., Liu, X. & Peltz, G. GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* **39**, btac757 (2023).

286. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).

287. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

288. Sloan, C. A. *et al.* ENCODE data at the ENCODE portal. *Nucleic Acids Res* **44**, D726–D732 (2016).

289. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).

290. Newick, K., O'Brien, S., Moon, E. & Albelda, S. M. CAR T Cell Therapy for Solid Tumors. *Annual Review of Medicine* **68**, 139–152 (2017).

291. Wagner, J., Wickman, E., DeRenzo, C. & Gottschalk, S. CAR T Cell Therapy for Solid Tumors: Bright Future or Dark Reality? *Mol Ther* **28**, 2320–2339 (2020).

292. Bailey, S. R. & Maus, M. V. Gene editing for immune cell therapies. *Nat Biotechnol* **37**, 1425–1434 (2019).

293. Hu, W. *et al.* CRISPR/Cas9-mediated PD-1 disruption enhances human mesothelin-targeted CAR T cell effector functions. *Cancer Immunol Immunother* **68**, 365–377 (2019).

294. Rupp, L. J. *et al.* CRISPR/Cas9-mediated PD-1 disruption enhances anti-tumor efficacy of human chimeric antigen receptor T cells. *Sci Rep* **7**, 737 (2017).

295. Tang, N. *et al.* TGF-β inhibition via CRISPR promotes the long-term efficacy of CAR T cells against solid tumors. *JCI Insight* **5**, e133977.

296. Weber, E. W., Maus, M. V. & Mackall, C. L. The Emerging Landscape of Immune Cell Therapies. *Cell* **181**, 46–62 (2020).

297. Wilson, R. C. & Gilbert, L. A. The promise and challenge of in vivo delivery for genome therapeutics. *ACS Chem Biol* **13**, 376–382 (2018).

298. van Haasteren, J., Li, J., Scheideler, O. J., Murthy, N. & Schaffer, D. V. The delivery challenge: fulfilling the promise of therapeutic genome editing. *Nat Biotechnol* **38**, 845–855 (2020).

299. Del'Guidice, T. *et al.* Membrane permeabilizing amphiphilic peptide delivers recombinant transcription factor and CRISPR-Cas9/Cpf1 ribonucleoproteins in hard-to-modify cells. *PLoS One* **13**, e0195558 (2018).

300. Ramakrishna, S. *et al.* Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. *Genome Res* **24**, 1020–1027 (2014).

301. Wang, H.-X. *et al.* Nonviral gene editing via CRISPR/Cas9 delivery by membrane-disruptive and endosomolytic helical polypeptide. *Proc Natl Acad Sci U S A* **115**, 4903–4908 (2018).

302. Fesnak, A. D., Levine, B. L. & June, C. H. Engineered T Cells: The Promise and Challenges of Cancer Immunotherapy. *Nat Rev Cancer* **16**, 566–581 (2016).

303. Sadelain, M., Rivière, I. & Riddell, S. Therapeutic T cell engineering. *Nature* **545**, 423–431 (2017).

304. Cronin, J., Zhang, X.-Y. & Reiser, J. Altering the Tropism of Lentiviral Vectors through Pseudotyping. *Curr Gene Ther* **5**, 387–398 (2005).

305. Clapham, P. R. & McKnight, Á. HIV-1 receptors and cell tropism. *British Medical Bulletin* **58**, 43–59 (2001).

306. Miyauchi, K., Urano, E., Takizawa, M., Ichikawa, R. & Komano, J. Therapeutic potential of HIV protease-activable CASP3. *Sci Rep* **2**, 359 (2012).

307. Aoki, T., Miyauchi, K., Urano, E., Ichikawa, R. & Komano, J. Protein transduction by pseudotyped lentivirus-like nanoparticles. *Gene Ther* **18**, 936–941 (2011).

308. Cai, Y. *et al.* DNA transposition by protein transduction of the piggyBac transposase from lentiviral Gag precursors. *Nucleic Acids Res* **42**, e28 (2014).

309. Cai, Y., Bak, R. O. & Mikkelsen, J. G. Targeted genome editing by lentiviral protein transduction of zinc-finger and TAL-effector nucleases. *eLife* **3**, e01911 (2014).

310. Izmiryan, A., Basmaciogullari, S., Henry, A., Paques, F. & Danos, O. Efficient gene targeting mediated by a lentiviral vector-associated meganuclease. *Nucleic Acids Res* **39**, 7610–7619 (2011).

311. Michel, G., Yu, Y., Chang, T. & Yee, J.-K. Site-specific Gene Insertion Mediated by a Cre-loxP-carrying Lentiviral Vector. *Mol Ther* **18**, 1814–1821 (2010).

312. Schenkwein, D., Turkki, V., Kärkkäinen, H.-R., Airenne, K. & Ylä-Herttuala, S. Production of HIV-1 Integrase Fusion Protein-Carrying Lentiviral Vectors for Gene Therapy and Protein Transduction. *Human Gene Therapy* **21**, 589–602 (2010).

313. Wu, X. *et al.* Targeting foreign proteins to human immunodeficiency virus particles via fusion with Vpr and Vpx. *J Virol* **69**, 3389–3398 (1995).

314. Indikova, I. & Indik, S. Highly efficient 'hit-and-run' genome editing with unconcentrated lentivectors carrying Vpr.Prot.Cas9 protein produced from RRE-containing transcripts. *Nucleic Acids Res* **48**, 8178–8187 (2020).

315. Rouet, R. *et al.* Receptor-Mediated Delivery of CRISPR-Cas9 Endonuclease for Cell-Type-Specific Gene Editing. *J. Am. Chem. Soc.* **140**, 6596–6603 (2018).

316. Hill, Z. B., Martinko, A. J., Nguyen, D. P. & Wells, J. A. Human Antibody-Based Chemically Induced Dimerizers for Cell Therapeutic Applications. *Nat Chem Biol* **14**, 112–117 (2018).

317. Muller, Y. D. *et al.* The CD28-Transmembrane Domain Mediates Chimeric Antigen Receptor Heterodimerization With CD28. *Front Immunol* **12**, 639818 (2021).

318. Ghebreyesus, T. A. WHO Director-General's opening remarks at the media briefing on COVID-19. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 (2020).

319. De Savi, C., Hughes, D. L. & Kvaerno, L. Quest for a COVID-19 Cure by Repurposing Small-Molecule Drugs: Mechanism of Action, Clinical Development, Synthesis at Scale, and Outlook for Supply. *Org. Process Res. Dev.* **24**, 940–976 (2020).

320. Centers for Disease Control and Prevention. Implementation of mitigation strategies for communities with local COVID-19 transmission. (2020).
321. Li, R. *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493 (2020).
322. Guan, W. *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *New England Journal of Medicine* **382**, 1708–1720 (2020).
323. An antibody test for the novel coronavirus will soon be available. (2020).
324. Bennhold, K. A. German exception? Why the country's coronavirus death rate is low. (2020).
325. Şimşek, A. COVID-19: Germany deaths hit 815, infection ratesslows. (2020).
326. TaqPath COVID-19 Combo Kit Instructions for Use. (2020).
327. TaqPath COVID-19 Multiplex Diagnostic Solution. (2020).
328. Guidance Document. Policy for Diagnostic Tests for CoronavirusDisease-2019 during the Public Health Emergency. (2020).
329. Dorn, A. van, Cooney, R. E. & Sabin, M. L. COVID-19 exacerbating inequalities in the US. *The Lancet* **395**, 1243–1244 (2020).
330. Kim, E. J., Marrast, L. & Conigliaro, J. COVID-19: Magnifying the Effect of Health Disparities. *J GEN INTERN MED* **35**, 2441–2442 (2020).
331. Abrams, E. M. & Szefler, S. J. COVID-19 and the impact of social determinants of health. *The Lancet Respiratory Medicine* **8**, 659–661 (2020).
332. Godoy, M. & Wood, D. What do coronavirus racial disparities look like state by state? (2020).