Engineering minimally immunogenic cargos and delivery modalities for gene therapy

by
Rumya S. Raghavan

B.S.
Massachusetts Institute of Technology, 2017

Submitted to the Harvard-MIT Program in Health Sciences and Technology in partial fulfillment of the requirement for the degree of

DOCTOR OF PHILOSOPHY IN MEDICAL ENGINEERING AND MEDICAL PHYSICS

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2024

Authored by: Rumya S. Raghavan
Harvard-MIT Program in Health Sciences and Technology
January 30, 2024

Certified by: Feng Zhang PhD
Professor of Biological Engineering, Thesis Supervisor

Accepted by: Collin M. Stultz MD, PhD
Director, Harvard-MIT Program in Health Sciences and Technology
Nina T. and Robert H Rubin Professor in Medical Engineering and Science
Professor of Electrical Engineering and Computer Science

Abstract

Since the discovery of CRISPR-Cas9 systems, gene therapies have revolutionized the field of molecular biology by introducing functional genes into cells to correct genetic defects or diseases. To date, several gene therapies are pending approval for use in the clinic and have shown promise in the treatment of a variety of genetic disorders including retinal dystrophy, hemophilia, lysosomal storage disorders and certain types of cancer. However, there are several challenges to using CRISPR-Cas9 in the clinic, including the efficiency and specificity of the gene editing process, the potential for off-target effects, and the immunogenicity of the CRISPR-Cas9 system. One of the main challenges of gene therapies is the immunogenicity of the (1) therapeutic vector and (2) cargo. Existing delivery systems trigger immune responses, rendering therapies ineffective and pose considerable risks to the patient population. Even the cargos, Cas nucleases, have been shown to generate humoral and cellular immunity in the general population. Thus, there is a need for minimally immunogenic cargos and delivery modalities to advance gene therapy to the clinic.

The goal of this thesis is to design and optimize minimally immunogenic (1) vehicles and (2) cargos for translational gene therapy delivery. (1) For the development of gene therapy delivery vectors, previous work has identified endogenous proteins that can form capsids and package nucleic acid. In this work, I focus on the PNMA or Paraneoplastic MA-containing protein family to engineer a delivery system that can form capsids, package nucleic acid, and deliver functional, minimally immunogenic cargo to target cells. (2) For the development of non-immunogenic gene therapy cargos, I engineer existing gene therapy cargos, such as SaCas9 and AsCas12a, to be minimally immunogenic while retaining native functionality.

This work overall highlights the promise of protein engineering to minimize immunogenicity of delivery systems and gene editing nucleases while optimizing for their functionality *in vivo*. I hope this work will be expanded and grow to serve as a foundation for personalized gene therapy medicine.

Table of Contents:

Acknowledgements:

This PhD work would not have been possible without the unyielding support of the many friends, mentors, and colleagues I have had the pleasure to work with and learn from.

First and foremost, I would like to express my gratitude to Feng Zhang. As my PhD advisor, the time and attention Feng dedicated to me always felt like he was right there on a team with me, troubleshooting and brainstorming together. His excitement and enthusiasm for science are truly unparalleled. Feng possesses a burning curiosity that compels him to be in the lab answering questions, and it is positively infectious. Thank you, Feng, for a truly transformative five years in your lab, for all the opportunities you've given me, and for the wisdom you've imparted. I also extend my thanks to Michael Birnbaum for serving on my committee. I met Michael in 2016 in an undergraduate course, and since then, I have looked up to him and sought his advice on everything from scientific minutiae to career guidance. It was Michael who advised me to pursue my PhD at HST, for which I am eternally grateful. Michael embodies everything I aspire to be as a scientist—purely curious, deeply inquisitive, and, most importantly, compassionately wise. Thanks also to Jim Collins for demonstrating what it means to pursue science with passion. Every project he undertakes holds deep personal significance, and his care for human health resonates deeply with my own motivations. To all of you on my committee, thank you for your insightful comments and questions.

I extend my appreciation to the Zhang Office. Thank you to Rhiannon Macrae for her fearless leadership and relentless work ethic. Erin Blackwell, thank you for being a compassionate listener while ensuring the lab runs smoothly. Sarah Mahoney, your efficiency and camaraderie are greatly appreciated. And a special thank you to the wonderful lab managers I've had—Rich Belliveau and Joannie Mok. Both of you exemplify efficiency and warmth, and I feel fortunate to have shared such wonderful memories with you.

To my collaborators within and outside the lab, for the PNMA project, I express my gratitude to Yugang Zhang, Guilhem Faure, Max Wilkinson, Daniel Strebinger, Victoria Madigan, Elena Puccio, Blake Lash, Michael Segel, and Katie Delong. For the RediCas9 project, I extend my thanks to Mirco Friedrich, Indigo King, Lucas Nivon, and Sam Vo. To the mentors within the lab, thank you to Soumya Kannan for her unlimited patience, Daniel Strebinger for his scientific guidance, training, and friendship, Serge Gueroussov for his life advice, training, and honesty, Jonathan Strecker for his scientific advice and lab antics, and Kalli Kappel, Grace Edmonds, and James Briggs for their career guidance. Finally, thank you to Makoto Saito for keeping the lab running and always extending a helping hand.

To my past mentors outside the lab who shaped me into the person I am today, thank you to Scott Manalis, Doug Lauffenburger, Genevieve Boland, Andrea White, Gary Nelson, and Linda Griffith for always believing in me. Special thanks to Darrell Irvine for introducing me to scientific research as an undergraduate, and to Kavya Rakhra for taking the time to teach and nurture me even when I had no idea what I was doing.

To the friendships I have built within the lab. Everyone told me that doing a PhD was emotionally grueling but I could have never predicted how deep the friendships I would form as a product of this journey would be. Thank you to Soumya Kannan whom I revered as a role model for her scientific acumen but also for her sharp wit. Thank you to Chris Frangieh whose work ethic and fearless ambition inspired and drove me to become a better scientist. Thank you to Suchita Nety who would always take a walk or a 'little treat' trip with me whenever I needed it and would impart her sage wisdom onto me (knowing very well that I rarely took it). Thank you to Kepler Mears who would always be down for an adventure, or an antic to lift my spirits, whether it was scooter trips or matcha breaks. Thank you to Michelle Walsh who has repaired my soul with her smile and sense of humor more times than I could count. Thank you to Elena Puccio who I was lucky enough to get to mentor but even luckier to get to call a friend for life. Thank you to Blake Lash and Daniel Strebinger because getting to work alongside you both for my first year in the lab will forever remain my favorite era of the lab and I will carry the memories we made together for the rest of my scientific career. Thank you to the members of the Chen lab, Sandeep Kambhampati, Ruth Raichur, Andy Russell, Hattie Chung, Jackson Weir and Yifan Zhang who were always there for me whenever I needed advice, or a hug, or a laugh. You all were a second family to me, and that beanbag chair, a second home.

To the friendships cultivated during my PhD, thank you to my HST cohort, Julie Greenberg, Traci Anderson, Laurie Ward, Joe Stein, and academic advisor Brett Bouma. Special thanks to Jon Arizti, Connor Verheyen, Neha Kapate, Vincent Miao, Daphne Schlesinger, Carmen Martin, and Amy Stoddard for inspiring me to be a better classmate, scientist, and person. My extended HST family—Kaitavjeet Chowdhary, Rahul Gupta, Carla Winter, Leonard Nettey, Dan Rubin, Jenny Walsh, Sarah Chang, Elliot Akama-Garren, Preston Ge, and Deborah Plana—thank you for making me feel loved and accepted even though I was not in your MD-PhD cohort. Many of you are my inspiration and role models to do an MD myself, so thank you. I would also like to thank my Harvard 2017 friends who have been brilliant beams of light in my darkest hours. I'd like to thank Ankit Gupta for his wisdom, fiercely loyal companionship and 8am tennis/breakfast sessions and Jonah Kallenbach for his unrelenting support, guitar lessons and oven s'mores. Thank you to Lucy Nam, Dipal Nagda, Kavya Pathak, Avni Nahar, Karl Aspelund, Kruti Vora, Avinash Saraf, Raj Vatsa, Anika Gupta, Curren Iyer, Shivangi Goel and Jen Hao for always lending a listening ear whether it was at a party or at my apartment sipping chai together. I have gotten to be your cheerleader through major milestones of your lives and it means so much to me that you have all been there to cheer me on too.

Thank you to the people with whom I had the opportunity to share a home during my PhD, who quickly transformed from roommates to family. To Nalini, you will always be a role model and one of my dearest friends. To Tally Portnoi, you are a once-in-a-lifetime kind of friend. You have changed my life and have been with me, learned with me, and laughed with me every step (and misstep) of my journey as an adult. I never had to wonder what it would be like to have a sister because I know I will always have you. To Regina Apodaca, I cannot begin to comprehend how a heart so big can fit inside a person who insists she's the same height as me, but meeting you in Israel in 2013 was one of the greatest things that could have possibly happened to me. You are more than a roommate; you are family for life. To Kai Xiao, I truly could not imagine going

through a pandemic with anyone else. There is not a single person with whom I've spent more hours discussing random things, and no one who knows the inside of my mind quite like you do. Thank you for always being there to catch me when I fall. To Lena Downes, Sam Wuu, Sarah Cen, and Sarah Wu, even though I only got to live with you for a year, that year was magical. You all carry a warmth within you that made me feel so loved and cared for and transformed our house into a home. I felt so lucky that I got to come home to you all every day, and I'll cherish the memories we've made together forever.

There is a collection of beautiful souls who have never 'officially' shared a home with me but have opened their hearts and homes to me so much that I waltz into their homes as if I actually live there. To Karina Smoylar, thank you for embracing me as a fellow cousin and for feeding my soul and my stomach with your endless care and compassion. To Claire Luo, Taylor Tang, and Sadie Swift, thank you for weaving me so seamlessly into the fabric of your lives. Many of the happiest memories of my Ph.D. have been just sitting in your apartment, playing games, dressing up like various characters, or simply living life alongside one another. To Adam Pissaris, thank you for always providing me with a fun reprieve from my stress, whether it be a run, a bike ride, a movie, a game, or a guitar lesson. To Daniel Zuo and Rebca van de Ven, thank you for accepting me and embracing me as I am, even in moments where I couldn't even accept myself. And finally, to Tanya Talkar, thank you for your unyielding support and being by my side through my darkest moments and my brightest days. You're the older sister I always wished I had, and I feel so lucky to have someone like you in my life who, despite all my flaws, loves me fiercely and unconditionally.

The person I became in graduate school would not have been possible without the community that supported my dreams and goals from near and far throughout undergrad and beyond. To Jitesh Maiyuran, thank you for celebrating all of my victories as if they were your own and anchoring me when I felt like I was drowning. To Advaith Anand and Rajeev Parvathala, thank you for all of the all-nighters, goofy adventures, deep talks, and relentless teasing. There's truly no one else I'd rather play lawn tennis on Killian Court after an all-nighter with. To Stuart and Laura Finney, thank you for being my home away from home in New York. I can't thank you both enough for being constant pillars of support and for always being there to talk through quite literally anything. To my undergraduate friends, David Doan, Margaret Tian, Tony Zeng, Aneesh Anand, Deepti Raghavan, Nikita Kodali, Divya Shanmugam, Apurva Shrivastava, Dheevesh Arulmani, Shraman Ray Chaudhuri, Benji Lin, Parker Zhao, Julia Guo, Tara Lee, Surya Bhupatiraju, Grace Li, Jackie Xu, and Michael Shum, thank you for the support from near and far. It has meant so much to me to know that in times of need, you'll all only be a phone call or Facebook message away. To my BE classmates, Lyla Atta, Christian Richardson, Jessie Blumenfeld, Juan Hurtado, Deborah Plana, and Vincent Tjeng, thank you for igniting and sustaining my passion and curiosity for science. I truly do not know if I would have done a PhD without the inspirational support of each and every one of you. To my college roommates, Kiran Wattamwar, Niki Tubacki, Christina Sun, and Erica Green, I am so grateful for the Aromatics. I have the comfort of knowing that no matter how far apart we move, we will always be there for each other to help one another through life's challenges and celebrate each other's milestones. Finally, I have the Camp Kesem community to thank for who I have grown to be over my last 10

years in Cambridge. Camp Kesem has altered the course of my life and is my constant source of inspiration in everything I do. I am so grateful for the fellow counselors I have met through Camp Kesem (Jasmin Joseph, Alicia Lai, Nick Schwartz), my campers, many of whom have now grown up and inspire me every day, and the camper parents whose inner strength and endless compassion remind me why I want to be a physician scientist every day of my life.

I'd like to thank my friends from my years before undergrad who continued to support me well into graduate school. I'd like to thank Matthew Nguyen and Sam Ong for always cheering me up with a video call. To my high school friends Michael Mankbadi, Oisin Harrington, Rosemary Kelley, and Jack Beck who always carried a bit of home with them and reminded me of who I was whenever I felt lost. To my science competition friends Rohan Batra, Pratheek Nagaraj, and Miguel Paredes who continue to constantly inspire me, console me, and cheer me on from near and far. To my Orlando friends Divya Dhulipala, Anjali Cherukuri, and Preeti Sarangarajan who have stood by my side through all of the adventures and antics I've put them through no matter how many of them were bad ideas. And finally, a huge thank you to Priya Sathyanarayan, Deepak Sathyanarayan, Priya Gurjar, and Shreya Singireddy who were such big parts of my childhood that spending time with them will always remind me of what it is like to be a kid again.

Finally, I'd like to thank my family for instilling deep educational values in me, always pushing me to greater heights, and for showering me with unconditional love and support. My late grandfather was the kindest, most loving soul I have ever known and believed in me far before I believed in myself. Before I even applied to MIT for undergrad, he had a vision that I would go there for college. Sadly, he never got to see me graduate, but I like to imagine that with every degree I earn from MIT, I am in some way fulfilling his dreams for me. I'd also like to thank my grandmothers for serving as my biggest inspirations and staunchest supporters. My late grandmother instilled in me a love for learning. Every day on our walk home from school, I would have a captive audience asking me about everything I had learned with intent curiosity. Her passing served as one of the core inspirations for my career in medicine and medical research, and I owe everything that I am now to her memory. My maternal grandmother is easily one of the brightest and mentally strong people I have ever met. She takes an interest in any subject from history to biology to geography, has such a deep appreciation for the arts, and is still an avid gardener. She never shies away from giving accurate feedback, but in doing so, has shown me how much she believes in my potential for growth. I hope every day to grow up to be just like her.

I'd like to give the biggest thank you of all to my parents. Growing up, I idolized my parents and every single quality about them, and as an adult, they have been my guiding light for the type of person I want to be. My mother is the most selfless and supportive person I have ever met. In college, I would routinely make decisions with my mother's voice in the back of my head, and it transformed me into a kinder, more empathetic person. Any problem I had, my mother had a solution, and her unrelenting belief in me is the core reason I have achieved everything that I have. If my mother had the solution to every problem, then my dad had an answer to every question (even if it wasn't necessarily the right answer). My dad's fierce ambition, coupled with the lightness and optimism with which he approaches life, has formed the foundation for my

entire personality and outlook towards life. My dad taught me that no goal is unattainable and spent hours reviewing every essay and every email I ever sent to any college. Everything I was stressed about or worried about, my dad's carefree attitude was always that "we" would handle it together. Above all else, my parents' sense of humor and playfulness will always be my favorite thing about my family. There is nothing we cannot laugh through, and I hope to carry that with me for the rest of my life. Just like all parents, my parents did also scold me from time to time. Their go-to admonishment was “I hope you end up with a daughter just like you one day” and honestly, it’s a credit to my own self-awareness because I knew that whenever they said “I hope you end up with a daughter like you” meant I had definitely done something wrong. I can confidently say that if someone tells me I turned out to be a parent just like my parents, I would have definitely done something right. My parents are so much more than just parents to me; they are my role models, my confidants, and my best friends in the whole world. I am so grateful that I get to be their daughter.

Thank you to everyone who has made my journey possible. You have been so generous with your time, enthusiasm, and support, and I would not be here today without you. You have made me a better person, professionally and personally, just by being a part of my life. This PhD is the product of the beautiful community that I am fortunate to have had for the past five years. Thank you all.

1. Introduction:

1.1 CRISPR-Cas9 gene editing technology

The advent of CRISPR-Cas9 systems have had broad implications for the field of genome editing, biotechnology and medicine. In this section, I will explore CRISPR-Cas9 mechanisms, delving into the key functional domains. Additionally, I will provide a comparative analysis of prominent Cas9 variants, including SpCas9, SaCas9 and AsCas12a, shedding light on their structural distinctions, functional nuances, and diverse applications.

1.1.1 Mechanism of CRISPR-Cas9

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) originate from the adaptive immune system in eubacteria and archaebacteria that function to counteract foreign nucleic acids. They integrate fragments of viral DNA as spaces within the CRISPR array and this acquired memory enables cells to recognize and neutralize future invasions by targeting complementary sequences. The CRISPR array encodes guide RNAs (gRNAs) that complex with the Cas (CRISPR-associated) proteins. The gRNA forms base pairs with the target DNA and directs Cas9 to introduce a double-stranded DNA break (DSB). The gRNA sequence specifies the cleavage target and the co-expression of a Cas protein and its cognate gRNA introduces DSB at specific positions in the genome. Repair mechanisms like non-homologous end joining (NHEJ) or homology-directed repair (HDR) then enable the desired genetic modifications.

The Cas-gRNA complex recognizes its target sequences through a protospacer adjacent motif (PAM) sequences defined by the gRNA[1]. The nucleotide sequence of each PAM varies based on the CRISPR-Cas system[2]. For example, SpCas9 from *Streptococcus pyogenes* has a G-rich PAM (NGG) that follows the 3′ end of the target sequence. AsCas12a from *Acidaminococcus* sp. has a T-rich PAM (TTTV) that precedes the 5′ end of the target sequence[3]. SaCas9 from *Staphylococcus aureus* has a PAM with an intermediate GC content (NNGRRT) that flanks the 3′ end of the target sequence[4].

1.1.2 Structural Insights into Functional Domains

The recognition domain (REC) comprises the conserved RuvC-like motif and the alpha-helical lobe, interacting with the gRNA scaffold. Structural studies reveal that the REC is pivotal for accurate DNA recognition and binding. The two non-coding RNAs, the crRNA and the transactivating crRNA hybridize with one another to form a hairpin loop duplex that is loaded into the REC domain. The REC lobe is poorly conserved and research has demonstrated that a Cas9 mutant lacking the REC1 domain, which is duplex-interacting region abolished DNA cleavage activity, indicating that the recognition of the repeat-antirepeat duplex by the REC1 domain is critical for Cas9 function. The PAM-interacting (PI) domain confers specificity for the Cas9 protein and is positioned to recognize the PAM sequence on the non-complementary DNA strand. Researchers found that the deletion of the PI domain abolished cleavage activity,

demonstrating that the PI domain was critical for Cas9 function. Additionally if the PI domain of one Cas nuclease was swapped with the PI domain of a different Cas nuclease, researchers demonstrated that the Cas nuclease was no longer able to recognize its native PAM sequence indicating that the PI domain is essential for PAM recognition. The nuclease domains (HNH and RuvC domains) are responsible for cleaving the target DNA strands, generating the DSB. The HNH domain targets the strand complementary to the sgRNA sequence while the RuvC domain cleaves the non-complementary strand.

1.1.3 Comparative Analysis of Cas9 Variants

SpCas9 (Streptococcus pyogenes Cas9) has multifaceted applications across research and medicine. Its broad targeting range is attributed to its intricate interplay between recognition and nuclease domains. Ongoing research focuses on enhancing its specificity and minimizing off-target effects[5]. SaCas9 (Staphylococcus aureus Cas9) is characterized by its compact size, showcases a truncated REC domain. Its structural variation does not compromise its DNA targeting capabilities, rendering it amenable for applications with size constraints, such as viral vector delivery[4]. AsCas12a (Cpf1) is an alternative to Cas9 which introduces staggered DNA cuts with single-stranded overhangs. Its compact size, coupled with this unique cleavage pattern, holds promise for targeting challenging genomic regions and enabling precise genetic modifications[3].

1.2 Clinical landscape of gene therapies with CRISPR-Cas9

In this section, we will investigate the use of gene therapy, specifically CRISPR-Cas9 systems, to treat genetic disorders. I will explore the present clinical landscape of gene therapies with ongoing clinical trials, delivery challenges, ethical considerations, and the future prospects of these technologies in medical treatments.

1.2.1 Clinical Applications of CRISPR-Cas9 to Genetic Disorders:

CRISPR-Cas9 based gene therapy has allowed for rapid progress in the treatment of transfusion dependent beta-thalassemia (TDT) and sickle cell anemia (SCD), two prevalent monogenic disorders. TDT entails mutations in the hemoglobin β subunit gene (HBB), causing reduced or absent β-globin synthesis and leading to issues in hemoglobin chain balance and ineffective erythropoiesis. SCD arises from a point mutation in HBB, transforming glutamic acid to valine and inducing hemoglobin polymerization upon deoxygenation. Current treatments include transfusions, iron chelation, pain management, and hydroxyurea[6]. Novel therapies like luspatercept and crizanlizumab show promise but don't address the root cause comprehensively. Bone marrow transplants offer a cure but face donor availability limitations. Betibeglogene autotemcel and gene-editing techniques are under investigation as potential treatments. Elevated fetal hemoglobin levels are linked to improved outcomes in TDT and SCD patients. Fetal hemoglobin's presence diminishes postnatally, leading to symptoms emerging within the first year of life. Those with hereditary persistence of fetal hemoglobin exhibit milder disease. The BCL11A gene is a repressor of fetal hemoglobin and has been identified as a

target for CRISPR-Cas9 gene editing whereby targeting its enhancer region will reduce its repression, restore fetal hemoglobin synthesis, and alleviate disease symptoms.

Another disease being targeted with CRISPR-Cas9 gene editing is Leber congenital amaurosis (LCA). LCA is the most common cause of inherited childhood blindness, caused by a mutation in the photoreceptor gene CEP290 that causes loss of the CEP290 protein and gradual vision loss or blindness within the first few months of life. CRISPR-Cas9 has been used to correct the defective CEP290 gene through a knock-in strategy and is currently being used in phase III clinical trials to cure LCA[7].

Hereditary transthyretin amyloidosis (hATTR) is a rare genetic disorder characterized by the accumulation of abnormal amyloid protein deposits in various organs, leading to progressive organ dysfunction and damage. It is caused by mutations in the TTR gene, which encodes the transthyretin protein. Transthyretin is mainly produced in the liver and plays a role in transporting thyroid hormone and retinoid binding protein. Mutations in the TTR gene lead to the production of abnormal transthyretin protein that misfolds and aggregates into amyloid fibrils. These amyloid deposits accumulate in tissues, particularly in nerves and the heart, disrupting their normal function. hATTR manifests clinically as peripheral neuropathy, cardiomyopathy, and gastrointestinal symptoms. CRISPR-Cas9 is being used to knock down the TTR gene, preventing misfolded protein from forming disease-causing amyloid aggregates[8].

CRISPR-Cas9 gene editing has also allowed for advances in the treatment of hereditary angioedema (HAE). HAE is a rare genetic disorder characterized by recurrent episodes of swelling all over the body, including the skin, GI tract, and upper respiratory tract. HAE is caused by mutations in the SERPING1 gene which encodes the C1 inhibitor protein that regulates the complement and contact systems. The absence of C1 allows for excessive activation and inflammation through the release of bradykinin. CRISPR-Cas9 editing works by suppressing the production of bradykinin through kallikrein (KLKB1) and thereby restoring balance with the C1 inhibitor protein and preventing inflammation[9].

Another disease, cystic fibrosis, is a complex genetic disorder affecting the respiratory and digestive systems and has emerged as a candidate for CRISPR-based gene therapy. Ongoing clinical trials are assessing the safety and efficacy of correcting mutations in the disease driving CFTR gene[10]. Early preclinical success has paved the way for exploring CRISPR-Cas9 platforms' potential for curing cystic fibrosis. Finally, Duchenne muscular dystrophy (DMD) is a challenging genetic disorder that CRISPR-Cas9 based platforms are being used to target mutations in the dystrophin gene[11,12]. For this clinical application, researchers are exploring advanced delivery strategies such as nanoparticle-mediated delivery and AAV vectors to efficiently target muscle tissue and achieve therapeutic effects.

1.2.2 Ongoing Clinical Trials

There are several ongoing clinical trials to use CRISPR-Cas9 gene editing for the treatment and management of the diseases discussed in the previous section. Presently there are two modes

of delivery for CRISPR-Cas9 gene editing in the clinic, ex vivo gene editing or in vivo gene editing.

For ex-vivo gene editing, cells are removed from the patient, edited outside of the body with CRISPR-Cas9, and then re-injected back into the patient. An example of clinical trials using ex vivo editing is the use of CRISPR-Cas9 editing of hematopoietic stem cells for TDT and SCD. Two clinical trials, CLIMB THAL-111 and CLIMB SCD-121, have outlined the first instances of patients with TDT and SCD being infused with genetically edited autologous hematopoietic stem and progenitor cells and have yielded promising results in restoring normal hemoglobin levels[13]. The recent completion of Phase I/II clinical trials underscore the promise of CRISPR-based therapies to cure these hemoglobinopathies as soon as the end of 2023. Another clinical trial, conducted by Editas Medicine is using a CRISPR system with a Cas12a protein rather than Cas9 to create edits that also turn on HbF. Thus far, the clinical trial has appeared to be effective without serious side effects but only reflect less than six months of follow-up[14]. Finally, Beam Therapeutics began enrolling patients for a phase I/II clinical trial using base editing to turn on HbF with single-nucleotide changes to the DNA without creating double stranded DNA breaks. While all of these approaches open new frontiers for CRISPR-Cas9 to cure genetic diseases, they are limited by their need for inpatient procedures such as bone marrow transplants that can be expensive, painful and time-intensive for patients. Unfortunately, the only way to circumvent these shortcomings is to develop in-vivo gene editing tools.

For in-vivo gene editing, cells are edited within the body with CRISPR-Cas9 being delivered to target cells either using adeno-associated viruses (AAVs) or lipid nanoparticles (LNPs). An example of in-vivo gene editing using AAVs is the treatment of inherited retinal diseases. Clinical trials to treat inherited retinal diseases are currently using CRISPR-Cas9 based systems to target mutations in genes such as CEP290 and RPE65 [15]. Initial results from Phase I trials highlight potential benefits but further refinement of delivery mechanisms remains critical to ensure long-lasting success in the clinic. LCA10 was the target for the first ever in vivo CRISPR therapy trial, sponsored by Editas Medicine. Patients were dosed in a single eye with the other eye serving as a control against which to test vision. Editas revealed that three patients had clinically meaningful changes to their vision and also found that the treatment was most effective for patients who had mutations in both copies of the relevant gene. This treatment, however, was largely limited by the inability to deliver AAV systemically due to immunogenicity concerns and has only worked well in the eye as the eye is an immune-privileged organ.

In contrast to AAVs, LNPs can be used systemically without triggering a strong immune response but are still filtered out by the liver. Therefore, most of the clinical trials involving LNPs deliver CRISPR-Cas9 mRNA targeting genes specifically expressed in the liver. For example, for hATTR, since TRR is primarily made in the liver, LNPs deliver the gRNA targeting the TTR as well as the CRISPR Cas9 mRNA directly to hepatocytes. This is the first trial to deliver genome-editing components systemically and is sponsored by Intellia, with sites in the EU, UK and New Zealand. One arm of the study is monitoring patients with neuropathy symptoms and the other arm is focused on cardiomyopathy symptoms. Between the two arms, even at the lowest dose, there is a reported 85% reduction in the amount of toxic protein in the participant's

circulation with a greater than 90% reduction in the patients receiving the highest dose ([8]). All patients demonstrate sustained reduction in the TTR protein over time, correlating with improved disease outcomes. Intellia is continuing to collect efficacy data for treatment approval by the FDA and other regulatory agencies.

In addition to hATTR, LNPs are also being used to treat HAE as the gene target for CRISPR-Cas9 therapies, KLKB1, is highly expressed in the liver. Researchers are testing patient serum to assess if the delivery of CRISPR-Cas9 mRNA is successfully reducing the levels of proteins that cause inflammation as well as the frequency of inflammatory attacks after treatment ([9]). Early stage trials are assessing safety and dosage of the treatments and have found that participants have been free from attacks as long as 10 months from the first treatment. Just four months after the first treatment, researchers have noted a 64% reduction in the amount of bradykinin in the lowest dose group and over 90% in the highest dose group. There have been no significant adverse events and these early results suggest that one-time treatment may serve as a functional cure for minor forms of HAE. Based on this early success, Intellia expects to begin Phase 2 dose-controlled trials in 2023 ([9,16,17]). These two disease examples present exciting possibilities for the use of LNPs for CRISPR-Cas9 use for genes expressed in the liver. However, LNPs are still less efficient than viral vectors and presently only target the liver successfully.

The scope of CRISPR-Cas9 extends beyond monogenic disorders to encompass cancer immunotherapy. Engineered T cells which have been edited to target specific antigens, offer immense potential to enhance immune responses against malignancies[18]. Ongoing clinical trials investigating CAR-T cells in the clinic demonstrate avenues that CRISPR-Cas9 could be adapted for complex disease treatments.

1.3 Immunogenicity of CRISPR-Cas9 nucleases

In this section I will discuss the cellular and humoral immunogenicity of SaCas9, SpCas9, and AsCas12a. SaCas9 and SpCas9 are the two most commonly used Cas9 orthologs, both of which are prevalent human commensals. Approximately 40% of the human population is colonized by S. aureus and 12% of children under 18 have an asymptomatic colonization with S. pyogenes [19,20]. Researchers have documented both humoral, antibody mediated, and cellular, T-cell mediated, immunity against *S. aureus* and *S. pyogenes* in 80% of healthy individuals [21,22].

The majority of immune responses target accessible secreted proteins and membrane surface proteins of bacteria. Since Cas9 is an intracellular protein that is primarily used in therapies that temporarily express or deliver recombinant to target cells, it was assumed that anti-Cas9 antibodies would have minimal effect on therapeutic efficacy[16]. However, a study conducted by Wang et al contradicted this by demonstrating that specific SpCas9 antibodies existed 14 days after adenoviral Cas9 delivery. They were able to achieve successful genome editing but IgG1. IgG2a, and IgG2b antibodies indicated an immune response to adenoviral Cas9[23]. Chew et al also found immune responses independent of Cas9 delivery. They noted an enrichment of CD45+ leukocytes and identified four T cell receptor ß-chain (TCR-ß) clonotypes of which one

clonotype recognized Cas9. Variable Cas9-specific antibody levels post exposure also indicated a humoral immune response. Collectively, these results highlight that there are distinct cellular and humoral responses to Cas9[17].

1.3.1 Humoral Immunity to Cas9

In addition to the aforementioned work in mice, several groups have reported the existence of pre-existing SpCas9, and SaCas9 antibodies in healthy human donors. One group in 2018 reported 2.5% of human donors with pre-existing anti-SpCas9 antibodies and 10% of human donors with pre-existing SaCas9 antibodies in a total of 200 individuals tested. In 2019, a study conducted by Charlesworth and colleagues found that in the population of 125 human donors profiled, 58% were seropositive for SpCas9 antibodies and 78% were seropositive for SaCas9 antibodies[24].

1.3.2 Cellular Immunity to Cas9

Several studies have also detected pre-existing cellular immunity against SpCas9 and SaCas9 in healthy human donors. Charlesworth et al. and Stadtmauer et al. report that over 67% of healthy human donors have CD8 T cells reactive against SpCas9 and 78% of healthy human donors have CD8 T cells reactive against SaCas9[25,26]. Wagner et al. reports even higher percentages, reporting that 95% of healthy human donors were CD8 T cell positive against SpCas9, 100% against SaCas9, and 100% against AsCas12a[27,28]. Interestingly this report demonstrates pre-existing T cell responses to Cas12a at comparable frequencies to SpCas9 and SaCas9. This is hypothesized to be due to sequence similarity between Cas9 orthologs and other CRISPR-related bacterial proteins that contribute to widespread pre-existing adaptive immune responses[24].

1.3.3 Strategies to mitigate immunogenicity

Strategies to mitigate pre-existing immunity to Cas9 encompass multiple approaches. One approach involves modifying the structure of Cas9 to conceal immunogenic epitopes. Epitope masking can modify potentially immunogenic peptide sequences in Cas9, thereby reducing immune recognition while preserving function. This has been shown with limited success in SpCas9 by eliminating the immunodominant T cell epitopes and then demonstrating that the nuclease can preserve function. Additionally, the strategy of the Epstein Barr Virus to inhibit antigen presentation through inhibitory signals such as the introduction of Gly-Ala repeats into Cas9 to prevent proteasomal degradation and immune recognition. Another approach is the use of Cas9 orthologs from non-pathogenic bacteria, like Geobacillus stearothermophilus, which could reduce immunogenicity due to limited human exposure. However, this strategy is limited because it has been demonstrated that the sequence similarity between CRISPR systems may be sufficient to generate immunogenicity as with AsCas12a. A way to circumvent immunogenicity of Cas9 altogether is to only use Cas9 systems in immune privileged organs such as the eye, brain and placenta which are microenvironments amenable to foreign proteins. However this limits the scope of potential gene therapy solutions.

1.4 Delivery strategies for gene therapy

In this section, I will discuss ways in which the efficacy of CRISPR-Cas9 gene editing hinges on effective delivery strategies. I will focus on the delivery modalities as they apply to gene editing with CRISPR-Cas9 based systems with a specific focus on strategies such as the use of AAVs, lipid nanoparticles (LNPs), lentiviral vectors, adenoviral vectors, physical methods and chemical approaches. I will discuss the mechanisms, advantages, challenges and immunogenic implications of each delivery strategy. I hope to underscore the critical role of delivery strategies in translation of CRISPR-Cas9 based gene editing in the clinic.

1.4.1 Adeno-Associated Virus (AAV) Vectors

AAV vectors are the main delivery modality used for CRISPR-Cas9 applications as they are able to efficiently ferry Cas9 and gRNAs to target cells. AAVs transduce cells and enable precise gene editing through Cas9-induced DNA cleavage. AAV vectors offer enduring gene expression, and tissue specific targeting [29]. Stable, long-term expression of the therapeutic gene in target cells makes AAV vectors particularly beneficial for treating chronic conditions. However, AAVs struggle with several limitations including limited cargo capacity, pre-existing immunity in patients and immune responses that hamper therapeutic outcomes[30]. Additionally, since AAVs are non-integrating, their potential to provide permanent cures is limited. Unfortunately, repeat doses of AAV gene therapy is also intractable given that many patients build up immunity to AAV vectors and the production of scalable and cost-effective AAV vectors is challenging to accomplish in vitro. Therefore, until there are solutions to reducing immunogenicity of AAV and improving scalability, AAV vectors are not sustainable long-term solutions for gene therapy delivery.

1.4.2 Lipid Nanoparticles (LNPs)

LNPs are nanoscale vesicles composed of lipids and amphiphilic molecules, engineered to encapsulate and protect nucleic acids, such as small interfering RNA (siRNA), messenger RNA (mRNA), or plasmid DNA, during their delivery into target cells. LNPs are attractive delivery vectors as they overcome the inherent challenges associated with nucleic acid therapeutics, including instability, limited cellular uptake, and susceptibility to degradation. LNPs have demonstrated the ability to encapsulate Cas9 mRNA or gRNA, facilitating intracellular delivery. The LNPs fuse with the cell membranes, deposit the mRNA cargo where it is translated from mRNA into protein, and the gRNA associates with the translated Cas9 protein to initiate the gene editing process. LNPs have several advantages including versatility in nucleic acid delivery, transient Cas9 expression and broad translatability[31]. Clinical trials exploring the use of LNPs in gene therapy are rapidly advancing, demonstrating the translational potential of this technology. For instance, mRNA-based COVID-19 vaccines, such as Pfizer-BioNTech's BNT162b2 and Moderna's mRNA-1273. Several other clinical trials, such as those described in **Section 1.2**, are investigating LNPs for various gene therapy applications, including treatments for genetic disorders, cancer immunotherapy, and infectious diseases. However, LNPs have their own unique set of challenges including immune activation, cargo stability and potential transient Cas9 off-target effects[32]. LNPs, unlike viral carriers, do not have built in mechanisms

for quickly translating their cargo and therefore are less efficient of a delivery system than lentiviral vectors or AAV. Additionally, LNPs are less specific than viral vectors since many LNPs traffic to the liver where they are filtered out by the body before they can reach target tissues.

1.4.3 Lentiviral vectors

Lentiviral vectors act as vehicles for Cas9 and gRNA delivery into dividing and non-dividing cells. These viral vectors enable genome integration which ensures sustained gene expression. This is both an advantage as well as a disadvantage since sustained gene expression can also lead to deleterious effects if there are off-target sites.

1.4.4 Adenoviral vectors

Adenoviral vectors serve for transient Cas9 expression. However, many patients experience immune responses due to their immunogenicity and therefore in order for them to become widely used in the clinic there is a need for immune-modulatory strategies.

1.4.5 Physical methods

Electroporation and biolistic particle delivery directly introduces CRISPR components into cells which enable precise editing. However, these techniques are only useful for ex vivo gene editing and re-entry back into the human body since physical methods are challenging to translate for in vivo editing.

1.5 Immunogenicity of Delivery Vectors

The success of gene therapy is closely linked to the efficiency of gene delivery vectors, but their immunogenicity can significantly impact therapeutic outcomes. In this section, I provide a comprehensive analysis of the interplay between the immunogenicity of two prominent gene delivery vectors, adeno-associated virus (AAV) and lipid nanoparticles (LNPs), and the resultant humoral and cellular immune responses. By dissecting the specific mechanisms through which these vectors interact with the immune system, this section sheds light on their effects on gene therapy efficacy and offers insights into strategies to navigate the complex landscape of vector immunogenicity.

1.5.1 Humoral Immune Responses

Humoral immune responses are largely driven by antibody formation. Upon administration of delivery vectors such as AAVs and LNPs, the host immune system can generate neutralizing antibodies that impede vector-cell interactions and diminish therapeutic efficacy[33]. Additionally activation of the complement system by AAV and LNPs can lead to opsonization, a process that enhances phagocytosis and clearance. This immune cascade critically influences the distribution of vectors within the body and their cellular uptake[34]. Immune complexes can form as a result of interactions between circulating antibodies and the delivery vector. These

complexes can contribute to inflammation and potentially exacerbate cytotoxic responses, impacting the vectors' performance[35].

1.5.2 Cellular Immune Responses

Both AAV and LNPs can activate T-cell responses, leading to cytotoxic activity. These immune reactions have the potential to limit transduction efficiency and induce unwanted adverse effects, thereby affecting therapeutic outcomes[36]. Dendritic cells play a pivotal role in presenting antigens derived from AAV and LNPs to T cells. The nature of this presentation profoundly influences the immune response, and the persistence of these vectors within the body[37]. Finally, exposure to AAV and LNPs can trigger the release of proinflammatory cytokines, creating an inflammatory milieu. This cytokine release has implications for vector biodistribution, cellular interactions, and immune responses [38].

1.5.3 Strategies to Mitigate Immunogenicity

Modifications to AAV and LNP components can mitigate immunogenic epitopes, potentially enhancing vector persistence and overall therapeutic efficacy[39]. The application of immunosuppressive agents can modulate immune responses triggered by AAV and LNPs. This approach aims to prolong vector presence and enhance gene transfer efficiency while minimizing adverse reactions[40]. Choosing vectors with inherently lower immunogenic profiles, such as LNPs, presents an avenue to minimize unwanted immune reactions and improve the success of gene therapy. Understanding the immunogenicity of AAV and LNPs is pivotal in designing effective clinical trials. Monitoring and adjusting therapeutic strategies based on vector immunogenicity profiles can enhance both the safety and efficacy of gene therapy interventions.

1.6 The PNMA family of Proteins

In this section, I will provide background about a family of endogenous retroviral gag-homolog proteins that we will explore as a delivery vehicle for gene therapy. This section will provide an in-depth exploration of PNMA proteins, their phylogeny, parallels to retroviral Gag homolog proteins in capsid formation, broader biological functions, involvement in cancer and potential as a delivery vector in biomedicine.

1.6.1 Phylogeny, Evolution and Gag Homology

The PNMA family of proteins are endogenous mammalian retroelements, remnants of ancient invasions into the mammalian genome by retroviruses and mobile genetic elements. Although many of these retroelements are silenced, some have been co-opted for function in mammalian biology [41]. One class of these elements are the long terminal repeat (LTR) retrotransposons, some of which encode the capsid-forming protein Gag. The domesticated Gag homolog Arc, which is involved in cognition and memory, has been shown to form capsids [42]. Another Gag homolog, PEG10, also forms capsids [43]. Our group recently showed that PEG10 could be engineered to programmably package and deliver a cargo RNA, demonstrating the potential of

endogenous retroviral-derived proteins as the starting point for development of new modalities for delivery of nucleic acid cargoes[44]. Both Arc and PEG10 are derived from the Ty3/Gypsy LTR retrotransposons. Another set of domesticated genes derived from this class of retrotransposons comprise the paraneoplastic antigen MA (PNMA) family [45]. *PNMA1-3* genes were initially identified as encoding auto-antigens in patients with paraneoplastic neurological disease [46–48]. Following this initial identification of three PNMA genes, a number of additional genes were reported based on bioinformatic searches[49,50]. PNMA genes have been implicated in a range of biological functions, including cell proliferation and apoptosis [49,51–54]. Like Arc and PEG10, PNMA genes share homology with Gag, but it is unknown if these proteins retain the ability to form capsids. Phylogenetic analysis of PNMA proteins reveals an intriguing convergence with retroviral Gag proteins—an unexpected parallel that suggests shared evolutionary ancestry. This resemblance suggests potential functional overlaps and adaptive pathways. Human PNMAs are spread across 4 chromosomes, but some are clustered together – PNMA8a/b/c and CCDC8 share the same locus in chromosome 19, and PNMA3/5/6a/e/f share the same locus in chromosome X.

1.6.2 Structural Insights and Capsid formation

The PNMA family likely emerged from the domestication of a Gypsy retrotransposon by the loss of the polymerase (POL) region[45]. Multiple duplications of the ancestral PNMA occurred in Eutherians giving rise to a large family of PNMAs in some mammalian species, including humans. Ty3/Gypsy is a family of LTR retrotransposons widespread in Eukaryotes genomes, able to form a capsid and interact with RNAs via a zinc finger domain located downstream the capsid domain within the nucleocapsid domain [55–58]. turGypsy harbors an N-terminal domain similar to an RNA Recognition Motif-like domain (RRM), a capsid domain formed by a tandem of helical domains, a region between the RRM and capsid domains that folds into 4 helices, and a C-terminal disordered region that contains a zinc finger at the distal end. mePNMA has the same domain architecture as Gypsy, whereas human PNMAs feature partial or complete loss of some of these components as well as insertion of additional regions. Most of the human PNMAs contain the capsid domain, although PNMA8a, b, and c and CCDC8, which form a single branch in the phylogenetic tree, lack the capsid domain. The RRM domain is also highly conserved, with only PNMA7a and 7b lacking it. The zinc finger domain, which may be involved in interaction with nucleic acids[59], is found in some PNMAs, but other PNMAs lack this domain. Of the proteins that lack the zinc finger, some alternatively contain a K-R rich domain, which could similarly function to interact with nucleic acids.

Given the presence of the capsid domain in most PNMAs, we sought to predict if multimers of PNMA proteins formed capsomers. PNMA1-5 are predicted to assemble as capsomers maintained by the interaction of the capsid domains. We did not observe capsomers for PNMA6-8 and CCDC8 suggesting additional components might be required for capsomer assembly, they might not be involved in capsid function, or alphafold-multimer failed to predict the assembly. PNMA1-5 capsomers harbor their N-terminal region including the RRM domain and the downstream helical domain sticking on the outer side of the capsid. This spatial location and absence of charge in the RRM domain might support a function related to target recognition or regulation/transport of the capsid. Such a function would likely involve potential protein

protein interaction which is compatible with protein interaction function reported in some RRM [60]. The capsid domain includes 2 domains that are involved in the formation of the capsomer and the complex assembly of capsomer. Between the RRM and capsid domain, the helical domain (conserved across most of PNMA) folds into 4 helices that partially form a globular domain and is predicted to form a homodimer in PNMA1, PNMA2, PNMA3, and PNMA5 resulting in a compact globular domain maintained by a hydrophobic core and salt bridges – for this reason we named this domain dimerization domain. Such dimerization constrains RRM to be aligned by pair and creates a subassembly within the capsomer whereby the capsid domain forms a pentamer, the dimerization domain forms 2 homodimers RRM plus one monomer of RRM sticking outside of the capsid.

Single-cell RNA sequencing data and histology data from the Human Protein Atlas and GTEx were used to map PNMA expression throughout tissues in the human body. These data illustrate that RNA transcript levels of all PNMA family proteins are highest in the brain, with some PNMAs, namely PNMA1, PNMA4, PNMA7a and PNMA8a, demonstrating elevated transcript levels in both male and female tissues. Protein expression for all the PNMAs also are highest in the brain with some PNMAs, namely PNMA1, PNMA3, PNMA5, PNMA7A, and PNMA8A, demonstrating elevated protein expression in both male and female tissues[49,61].

1.6.3 Endogenous Biological Functions

3.1 Neurological Implications and Beyond: Beyond their initial association with paraneoplastic neurological disorders, PNMA proteins play roles in neural development, synaptic transmission, and neuronal plasticity—interconnections that resonate with Gag-like attributes.
3.2 Immunomodulation and Beyond: Growing evidence implicates PNMA proteins in immune regulation and signaling, suggesting their broader involvement in physiological processes, mirroring the multifunctionality observed in Gag proteins.

1.6.4 PNMA proteins in Cancer

Anti-PNMA antibodies primarily coincide with the presence of a paraneoplastic syndrome secondary to a tumor elsewhere in the body. Current research suggests that the immune system could naturally defend against tumor cells through paraneoplastic neurological degenerations (PNDs). PNDs are a group of neurodegenerative diseases associated with cancer and antitumor immunity. Patients with PNDs may experience various neurological symptoms and are often unaware that they have cancer until it is discovered later. The link between neurological degeneration and cancer in PNDs is believed to be immunological[62]. The immune response targets neuronal proteins expressed by tumor cells (onconeural antigens), suppressing tumor growth. However, this immune response can lead to autoimmunity with neurological symptoms when it starts recognizing neurons expressing the same protein. Patients with PNDs generally have a more favorable cancer prognosis compared to patients with histologically identical tumors not associated with PNDs. Some PND patients experience tumor suppression or spontaneous tumor regression, even to the point where no primary tumor can be identified[63]. This suggests that the onset of neurodegenerative symptoms and immune responses may correlate with tumor suppression[64]. Histological analysis of PND-associated tumors reveals the

presence of inflammatory plasma cells and T cells, indicating an immune response within the tumor microenvironment. Lymphocytic infiltrates have also been observed in some cases. Finally, the discovery of antigen-specific killer T cells in the blood of PND patients provides strong evidence for naturally occurring antitumor immunity in humans.

Each type of PND is associated with specific tumor types, such as breast, ovarian, or small-cell lung cancer. In the case of PNMA2, it mostly co-occurs with testicular cancer. In men younger than 50 years, Anti-Ma2 encephalitis is strongly associated with testicular germ-cell tumors[46,47]. These tumors can be microscopic and challenging to detect. Some researchers suggest that Anti-Ma2 antibodies may be part of an effective antitumor immune response, possibly explaining the small size of the tumors detected in these cases. In older individuals, the most common associated tumors are non-small-cell lung cancer and breast cancer. We see that PNMA2 is highly expressed in male tissues such as the testes and prostate so this immunity could develop as a response to overexpression in the setting of cancer dysregulation. This cancer antigen is then expressed to the immune system and the subsequent immune response generates an autoimmune response against a previously inert self protein.

1.6.5 Potential as a Delivery Vector

The capacity of specific PNMA proteins to form capsid-like structures raises prospects for their use as delivery vectors in biomedicine. Their structural kinship with Gag proteins imbues them with the potential for efficient cargo encapsulation and targeted delivery. Harnessing PNMA proteins as delivery vectors has transformative implications for drug delivery, gene therapy, and vaccine development.

1.7 Overall Thesis Introduction:

CRISPR-Cas9 is a powerful tool for gene editing that consists of two components: a guide RNA (gRNA) that recognizes a specific sequence in the genome, and the Cas nuclease, which cuts the DNA at the targeted site[26,28]. Since the discovery of CRISPR-Cas9 systems, gene therapies have revolutionized the field of molecular biology by introducing functional genes into cells to correct genetic defects or diseases[24]. To date, several gene therapies are pending approval for use in the clinic and have shown promise in the treatment of a variety of genetic disorders including retinal dystrophy, hemophilia, lysosomal storage disorders and certain types of cancer[25,33]. However, there are several challenges to using CRISPR-Cas9 in the clinic, including the efficiency and specificity of the gene editing process, the potential for off-target effects, and the immunogenicity of the CRISPR-Cas9 system[33]. One of the main challenges of gene therapies is the immunogenicity of the (1) therapeutic vector and (2) cargo.

1.7.1 Vectors

Many gene therapies use viral vectors, which can trigger an immune response in the host and limit the effectiveness of the therapy[14]. Non-viral vectors, on the other hand, are generally considered to be less immunogenic but are also less efficient at delivering genes to the target cells[2,3]. There are three primary categories of RNA delivery systems: liposomal carriers, viral

vectors, and virus-like particles (VLPs). Liposomal carriers, such as those used in the Moderna vaccine, have had success in vivo but have limitations, including a selective accumulation in the liver and challenges in engineering them to deliver mRNA and bypass the endosome. Viral vectors, including lentiviral vectors, are efficient at delivering genetic material into host cells but carry the risk of compromising the genome and inducing tumorigenesis by integrating into the host genome[14]. VLPs, which are composed of endogenous proteins, are a newer class of vectors that require further evaluation for their ability to avoid the endosome and minimize immunogenicity compared to other delivery methods.

Natural delivery systems from the human genome might provide the basis for new engineered gene transfer modalities that can address some of these limitations. Recent work has uncovered a diverse array of endogenous *gag*-like genes within the human genome, which resemble retroviral structural proteins and therefore could potentially be engineered for gene transfer (2). Retroviral genomes encode three major proteins: Gag, Pol, and Env. Gag or Group Antigen, is a polyprotein that forms the viral core structure and serves as a RNA genome binding protein, making up the major proteins in the nucleoprotein core particle[1,2,3]. The gag proteins include the matrix protein (MA), the capsid protein (CA), and the nucleocapsid protein (NC)[2,4]. These proteins work together to form the capsid, which has a spherical shape with icosahedral symmetry[2,3,4]. The capsid protein forms the outer shell of the capsid and is responsible for the shape and stability of the capsid[3,4,5]. The nucleocapsid protein is located inside the capsid and helps to package the viral genome. Some endogenous gag-containing proteins, such as the PNMA (para neoplastic antigen Ma) family, only contain the CA and NC subdomains of Gag, while others, like RLT1 and PEG10 (also known as RTL2), contain the additional subdomains of Pol, including a PRO domain and a predicted RT-like domain[6,8,9,10]. A number of these, including *Arc* and *Peg10,* have been domesticated and serve vital roles in normal mammalian physiology [42,65,66]. Previous work has also shown that the ARC and PEG10 *gag*-like proteins have the ability to form capsid structures that can package their cognate mRNAs [42–44,66]. Extending this natural ability, PEG10 was recently engineered to programmably package and deliver an exogenous cargo mRNA into human cells, demonstrating the potential of these endogenous retrotransposon-derived proteins as a new nucleic acid delivery modality [44]. To further explore the potential of endogenous *gag*-like proteins for therapeutic RNA delivery, we sought to systematically characterize the paraneoplastic Ma antigen (PNMA) protein family [45]. The PNMA family, which is predominantly expressed in the brain and also found in the testes and ovaries, appears to have originated from a Ty3 LTR retrotransposon[1,9,10,11]. The PNMAs all appear to contain an RNA recognition motif at their N-terminus, followed by the N-term and C-term capsid regions. Some of the PNMAs, namely PNMA3 contain a zinc finger domain and an arginine rich region, which may additionally be useful for binding nucleic acids[1,8,10]. Their PNMA family's abilities to form capsid and potentially bind to and protect RNA make them ideal endogenous vehicles for use in gene therapy delivery.

Here we explore the potential for human PNMA proteins to form capsids and package RNA. We found that PNMA2 is robustly secreted as an icosahedral capsid from human cells and can self-assemble *in vitro* from recombinant protein. We used cryo-electron microscopy (cryo-EM) to resolve the structure of the PNMA2 capsid and structure-guided engineering to modify the

PNMA2 protein capsid to package mRNA. We show that these engineered PNMA2 capsids can functionally deliver mRNA into recipient cells, demonstrating promise as a therapeutic mRNA delivery vehicle. In addition to PNMA2, we found that other PNMA family members are capable of forming virus-like capsids, suggesting they may also be suitable for delivery and raising the possibility that these proteins are involved in intercellular communication.

1.7.2 Cargos:

To date, most CRISPR-based therapies rely on one of three Cas nucleases: *Streptococcus pyogenes* Cas9 (SpCas9), *Staphylococcus aureus* Cas9 (SaCas9), and *Acidaminococcus species* Cas12a (AsCas12a)[4,16,67,68]. Among these three, SaCas9 and AsCas12a are the focus of most *in vivo* therapeutic strategies because their size allows them to be packaged into adeno-associated viral (AAV) vectors, the leading delivery modality for *in vivo* gene therapies[4,16,67,69].

In addition to efficient delivery of these tools, a second challenge to their *in vivo* use is their potential immunogenicity, particularly due to their bacterial origin, as many patients have pre-existing exposures and immune responses to microbially-derived molecules[21,26,70]. It has been reported that 80% of healthy individuals have both humoral immunity, mediated by antibodies, and cellular immunity, mediated by T-cells, against proteins derived from *Staphylococcus aureus* and *Streptococcus pyogenes*[27,71]. This pre-existing immunity extends to Cas nucleases as well: profiling blood from healthy human donors revealed that 78% had class-switched to immunoglobulin (IgG) antibodies against SaCas9 and 58% had antibodies against SpCas9, and all donors that were positive for cellular immunity against Cas9 also had antibody activity, indicating a high concordance between adaptive and humoral immunity[16,17,26,72,73]. This pre-existing immune response to Cas9 could be a significant barrier to the development of in vivo therapies, as the immune system may recognize and target cells expressing Cas9, rendering the therapy ineffective[27,29,33]. It is therefore important to develop strategies to avoid a Cas9-directed immune response in order to improve the effectiveness of gene therapies.

To address this challenge, we profiled SaCas9 and AsCas12a to identify putative immunogenic epitopes and then computationally designed variants of these two nucleases predicted to evade immune detection. We experimentally validated these variants, showing that they eliminated CD8 T cell reactivity *in vitro* while retaining native levels of nuclease activity and specificity. Additionally, in an immunocompetent MHC class I/II humanized mouse model, we demonstrated that the SaCas9 variants effectively reduce humoral and cellular immune reactions as compared to the wild-type nucleases. These results provide a framework for engineering therapeutic proteins to reduce immunogenicity and provide a starting point for development of safer CRISPR-based therapeutics.

2. Rational engineering of minimally immunogenic nucleases for gene therapy

Rumya Raghavan[1-5], Mirco J. Friedrich[1-5], Indigo King[6], Michael Kilian[7], Michael Platten[7], Rhiannon K. Macrae[1-5], Yifan Song[6] , Lucas Nivon[6] & Feng Zhang[1-5,†]

2.1 Abstract

Genome editing using CRISPR-Cas systems is a promising avenue for the treatment of genetic diseases. However, cellular and humoral immunogenicity of genome editing tools, which originate from bacteria, complicates their clinical use. Here we report reduced immunogenicity (Red)(i)-variants of two clinically-relevant nucleases, SaCas9 and AsCas12a. Through MHC-associated peptide proteomics (MAPPs) analysis, we identified putative immunogenic epitopes on each nuclease. Then, we used computational modeling to rationally design these proteins to evade the immune response. SaCas9 and AsCas12a Redi variants were substantially less recognized by adaptive immune components, including reduced binding affinity to MHC molecules and attenuated generation of cytotoxic T cell responses, while maintaining wild-type levels of activity and specificity. *In vivo* editing of *PCSK9* with SaCas9.Redi.1 was comparable in efficiency to wild-type SaCas9, but significantly reduced undesired immune responses. This demonstrates the utility of this approach in engineering proteins to evade immune detection.

2.2 Background

CRISPR-based genome editing therapies are currently being tested in the clinic to treat a variety of genetic disorders including retinal dystrophy, hemophilia, lysosomal storage disorders and certain types of cancer[11,12,27,74–76]. These powerful therapies consist of a Cas nuclease and a guide RNA (gRNA) that dictates the target site to be edited[77]. To date, most CRISPR-based therapies rely on one of three Cas nucleases: *Streptococcus pyogenes* Cas9 (SpCas9), *Staphylococcus aureus* Cas9 (SaCas9), and *Acidaminococcus species* Cas12a (AsCas12a)[4,16,67,68]. Among these three, SaCas9 and AsCas12a are the focus of most *in vivo* therapeutic strategies because their size allows them to be packaged into adeno-associated viral (AAV) vectors, the leading delivery modality for *in vivo* gene therapies[4,16,67,69].

In addition to efficient delivery of these tools, a second challenge to their *in vivo* use is their potential immunogenicity, particularly due to their bacterial origin, as many patients have pre-existing exposures and immune responses to microbially-derived molecules[21,26,70]. It has been reported that 80% of healthy individuals have both humoral immunity, mediated by antibodies, and cellular immunity, mediated by T-cells, against proteins derived from *Staphylococcus aureus* and *Streptococcus pyogenes*[27,71]. This pre-existing immunity extends to

Cas nucleases as well: profiling blood from healthy human donors revealed that 78% had class-switched to immunoglobulin (IgG) antibodies against SaCas9 and 58% had antibodies against SpCas9, and all donors that were positive for cellular immunity against Cas9 also had antibody activity, indicating a high concordance between adaptive and humoral immunity[16,17,26,72,73].

To address this challenge, we profiled SaCas9 and AsCas12a to identify putative immunogenic epitopes and then computationally designed variants of these two nucleases predicted to evade immune detection. We experimentally validated these variants, showing that they eliminated CD8 T cell reactivity *in vitro* while retaining native levels of nuclease activity and specificity. Additionally, in an immunocompetent MHC class I/II humanized mouse model, we demonstrated that the SaCas9 variants effectively reduce humoral and cellular immune reactions as compared to the wild-type nucleases. These results provide a framework for engineering therapeutic proteins to reduce immunogenicity and provide a starting point for development of safer CRISPR-based therapeutics.

2.3 Results

2.3.1 Computational design of non-immunogenic epitopes in SaCas9 and AsCas12a

Although pre-existing immunity to SaCas9 has been reported[21,22,70], the specific epitopes recognized by the immune system are not known. Therefore, we profiled the peptide sequences mediating cellular immunogenicity to both SaCas9 and AsCas12a by performing MHC-associated peptide proteomics (MAPPs) analysis on HLA-A*02:01-expressing MDA-MB-231 cells transfected with a plasmid expressing either SaCas9 or AsCas12a[78]. Peptides bound to MHC class I molecules were then identified by mass spectrometry (Fig. 1a). The MAPPs analysis nominated three immunodominant epitopes for each nuclease; for SaCas9, the predicted epitopes were (1) 8-GLDIGITSV-16, (2) 926-VTVKNLDVI-934, and (3) 1034-ILGNLYEVK-1050, and for AsCas12a, the predicted epitopes were (1) 210-RLITAVPSL-218, (2) 277-LNEVLNLAI-285, and (3) 971-YLSQVIHEI-979. We then took a structure-guided computational approach to design mutants that would ablate or reduce MHC class I binding to a representative set of 14 HLA alleles but preserve nuclease activity (Extended Data Table 1). Briefly, we used the Rosetta protein design package to introduce mutations to the nuclease model to reduce MHC-binding propensity of all peptide subsequences around the epitope for a representative allele from every HLA-A, HLA-B, and HLA-C supertype cluster as defined in Rasmussen et al[79]. This approach aimed to eliminate known MHC-binding epitopes without creating new predicted epitopes while also maintaining protein stability. Mutations were evaluated *in silico* for protein stability and peptide-MHC binding (Fig. 1b). From this process, we found that that of all of the tested MHCs, the peptide-MHC binding was most pronounced for HLA-A*0201 and we chose to focus on this allele for *in vitro* and *in vivo* work moving forward (Extended Data Table 1). For each immunogenic epitope, we designed three or four variants containing single point mutations, which we mapped onto the structure and domain

architecture of the wild-type proteins (Figs. 1c,d). The locations of the proposed mutations do not overlap with either the DNA or RNA binding regions or the catalytic sites. Furthermore, all mutations were modeled *in silico* to ensure adequate shape complementarity to avoid clashes, such that no mutations were predicted to disrupt native nuclease function outside the tolerable range (Fig. 1d).

2.3.2 SaCas9 and AsCas12a peptide variants demonstrate reduced CD8 T cell reactivity

We used NetMHCpan 4.1, a tool that predicts the likelihood of peptide-MHC class I binding, to verify the findings of the MAPPs analysis and predict whether our proposed point mutant peptides would be presented on MHC class I molecules to CD8 T cells[80]. Consistent with our MAPPs analysis, the wild-type peptides for SaCas9 epitope 1 as well as AsCas12a epitopes 1 and 3 had a NetMHCpan predicted rank score below 0.5, indicating strong binding (shown as inverted rank score in Fig. 2a). Wild-type peptides for SaCas9 epitopes 2 and 3 as well as AsCas12a epitope 2 were not predicted to be strongly immunogenic by NetMHCpan 4.1. All peptide variants containing single point mutations were predicted to decrease the binding strength between peptide and MHC. Next, we experimentally assessed the degree of CD8 T cell mediated immunogenicity in PBMCs derived from healthy human donors to each identified epitope and their peptide variants using an ELISpot assay, which measures T-cell recognition following peptide binding to MHC class I molecules (Extended Data Table 2, Fig. 2b). We synthesized peptides of either wild-type or point mutant-containing variants of each epitope and then mixed with peripheral blood mononuclear cells (PBMCs) from healthy donors (HLA-A*0201). For SaCas9, we observed a robust immune response to wild-type peptide epitopes 1 and 2 and a more muted immune response to the epitope 3 peptide. Compared to the wild-type peptide, we observed that the single-mutant peptide variants for epitopes 1 and 2 had produced significantly fewer spots, indicating a reduced immune reaction to these variants. The epitope 3 peptide variants showed comparable levels of spot formation as wild-type epitope 3 (Figs. 2c,d). For AsCas12a, we observed robust spot formation in the presence of wild-type peptide epitopes 1 and 3 but less robust spot formation in the presence of wild-type peptide epitope 2 which was also consistent with our NetMHCpan computational predictions. As with SaCas9, we found that all AsCas12a single-mutant peptide variants significantly reduced spot formation relative to wild-type (Figs. 2c,d). Together, these results validate the MAPPs analysis and indicate that the mutant peptide sequences reduce recognition by CD8 T cells as compared to the native sequences.

2.3.3 Evaluation of nuclease efficiency and specificity of SaCas9 and AsCas12a variants

We generated full-length proteins containing the single point mutants and tested their editing efficiency in human cells by transfecting them with nuclease-containing plasmids. For SaCas9, we found that most single mutants exhibited indel activity at or above 60% of the wild-type

activity, although certain amino acid substitutions, such as L9F and L1035T, completely ablated nuclease activity (Fig. 3a and Extended Data Table 3). To reduce immune reactivity to any of the three immunodominant epitopes identified in SaCas9, we generated double and triple mutants combining the single mutations from across epitopes that maintained high indel activity (see Methods and Extended Data Tables 3, 4). We selected the eight of the highest performing triple mutants and profiled their activity across a panel of different target sites (Extended Data Table 3). Of these eight mutants, three showed comparable levels of activity to wild-type SaCas9 across the panel of targets: L9A/I934T/L1035A, L9S/I934K/ L1035V and V16A/I934K/L1035V, which we refer to collectively as reduced immunogenicity (Redi) variants and specifically as SaCas9.Redi.1, .2, and .3, respectively (Fig. 3b).

We repeated this process for AsCas12a and again observed that most single point mutants maintained nuclease activity, except L211V, L211A, and I285V (Fig. 3c and Extended Data Table 4). After testing double mutants (Extended Data Table 4), we generated triple mutants and profiled their indel activity at a single target (Extended Data Table 4). We expanded our analysis of the top seven triple mutants by testing their indel activity at six target sites. Based on these results, we selected three triple mutants for further characterization: L218S/I285S/L972A, L218S/I285T/L972A, and L218T/I285A/L972A, which we refer to as AsCas12a.Redi.1, .2, and .3, respectively (Fig. 3d). We then confirmed that the SaCas9.Redi and AsCas12a.Redi variants maintained specificity by assessing their genome-wide off-target effects, and we found no significant differences from wild-type (Fig. 3e,f)[81].

2.3.4 Engineered SaCas9 variants retain in vivo *editing efficacy with minimal host immune response*

We next investigated the efficiency and immunogenicity profiles of the SaCas9 Redi variants *in vivo*. To do so, we administered AAV8 vectors, which target hepatocytes, encoding either wild-type SaCas9 or the SaCas9.Redi variants, along with a gRNA targeting *Pcsk9*, to HLA-A*0201 HLA-DRA*0101 HLA-DRB1*0101 transgenic mice (A2.DR1)[82]. These mice lack mouse MHC and express human MHC class I and II complexes. The mice were re-treated after 14 days to allow for the development of memory T cells. At day 21, we evaluated the editing efficiency, as well as the innate and adaptive immune response (Fig. 4a). The indel rates at *Pcsk9* in hepatocytes were comparable between wild-type SaCas9 and SaCas9.Redi.1. Both SaCas9.Redi.2 and SaCas9.Redi.3 showed lower editing than WT but still detectable editing levels) (Fig. 4b). Genome editing of *Pcsk99* translated to significantly reduced PCSK9 serum levels in treated mice, again with SaCas9.Redi.1 achieving comparable reductions to wild-type SaCas9 (Fig. 4c). We next isolated splenocytes from each of the AAV8-treated mice and treated them *ex vivo* with the respective immunogenic wild-type or single-mutant epitopes for 24 hours to evaluate the development of an adaptive T cell response following exposure to these nucleases (Figs. 4d). For SaCas9 epitope 1, the L9A (SaCas9.Redi.1) and V16A mutations (SaCas9.Redi.3) resulted in a decreased T cell response in treated animals, while the L9S (SaCas9.Redi.2) mutation exhibited comparable immunogenicity to the wild-type epitope. For epitopes 2 and 3, all engineered mutations induced a less pronounced T cell response upon

recall, mirroring our *in vitro* ELISpot findings in healthy human donors (Fig. 4d; Extended Data Fig. 3a). We next reasoned that if these single epitopes were less immunogenic in SaCas9-exposed mice, it would lead to an overall mitigated immunogenicity profile of the engineered nuclease variants compared to wild-type SaCas9. Therefore, we integrated the immunogenicity testing results of all immunodominant epitopes and single-point mutant variants for each of the nucleases used *in vivo* by averaging the ELISpot recall data for each epitope per assembled nuclease (Fig 4e). We found that repeated treatment with wild-type SaCas9 resulted in the generation of T cell memory in A2.DR1 mice, which was mitigated when using SaCas9.Redi.1, Redi.2, and Redi.3 variants (Fig. 4e). Additionally, we assessed the humoral immune response by multiplex cytokine detection in AAV8-WT-SaCas9 and AAV8-SaCas9.Redi-treated animals. Consistent with the observed reduced adaptive immunogenicity profile, treatment with AAV8-SaCas9.Redi.1 resulted in decreased serum levels of all tested pro-inflammatory cytokines compared to AAV8-WT-SaCas9, while treatment with both AAV8-SaCas9.Redi.1 and -2 resulted in decreased serum levels of specifically Interleukin-1β (Il-1β), which is typically produced by monocytes and macrophages in response to a various stimuli, including tissue damage[83–85] (Fig. 4f, Extended Data Fig. 3c). However, treatment with AAV-GFP alone would induce nonspecific inflammatory cytokine activity. Lastly, all tested variants did not result in any short-term liver toxicity (Extended Data Fig. 3b). Together, these data demonstrate that SaCas9.Redi.1 exhibited reduced immunogenicity without compromising its efficacy when compared to wild-type SaCas9 *in vivo*, highlighting its potential as a minimally immunogenic genome editing tool.

2.4 Discussion

Recent studies have highlighted important considerations regarding the immune response to genome editing tools. For example, observational studies have reported a high prevalence of SpCas9 reactive T cells in the adult human population. Additionally, mice immunized against SaCas9 with Freund's adjuvant prior to intravenous delivery of AAV8-SaCas9-sgRNA exhibited CD8+ T cell accumulation in the liver and subsequent lysis of edited hepatocytes. These findings underscore the potential influence of pre-existing immunity due to exposure to *S. aureus* under inflammatory conditions on the host response to Cas9 in humans[26,28,86]. Here, we establish a workflow to identify and remove immunogenic T cell epitopes while preserving wild-type levels of activity and specificity to address this challenge, generating a suite of reduced immunogenicity (Redi) variants of two clinically relevant nucleases, SaCas9 and AsCas12a.

To validate these Redi variants *in vivo*, we replicated the development of adaptive immunity to AAV8-SaCas9 by repetitive dosing of mice with AAV8-SaCas9-sgRNA at a 14-day interval. Consistent with previous studies, we observed substantial T cell immunity against SaCas9 wild-type epitopes, which was significantly reduced with the SaCas9.Redi.1 variant. Although it remains unclear how well immunized mouse models recapitulate pre-existing immunity in humans, we did also observe a reduced CD8+ T cell response to this variant from human PBMCs, suggesting the Redi variants may be safer for *in vivo* clinical use. Currently most *in vivo*

investigational genome editing therapies are delivered using lipid nanoparticles or AAV vectors which are both immunogenic[87–89]. With our engineered minimally immunogenic variants, we hope that delivery with less immunogenic delivery vehicles will lead to robust, long term clinical benefit for *in vivo* gene editing therapy. To further enhance these variants, future work should focus on engineering nuclease variants that are minimally immunogenic across multiple HLA types, extending our work here with the HLA-A*0201 haplotype. Development of these measures is essential for improving the safety and efficacy of genome editing therapies in diverse clinical settings, especially if multiple such treatments might be performed in a patient's lifetime and across different patient populations. We hope that with further advancements in engineering safer genome engineering platforms, clinical applications of these systems can extend from treatment of genetic disorders to *in vivo* engineering cellular immunotherapies[90]. The findings of this study offer valuable insights into reducing immunogenicity of genome editing tools through protein engineering and, with continued investigation, will facilitate the translation of these tools for long term clinical use.

2.5 Materials and Methods

2.5.1 MHC-Associated Peptide Proteomics (MAPPs)

MDA-MB-231 cells (#HTB-26, ATCC) were maintained in RPMI Medium supplemented with 10% Fetal bovine serum and 100 U/mL penicillin-streptomycin. Cells were expanded to 1.8e7 cells in a T225 flask and transfected with 60 µg of SaCas9 (#61591, Addgene) or AsCas12a (#69982, Addgene) using Lipofectamine 3000 (ThermoFisher, L3000001) at 80% confluence. Media was changed 4 hours post transfection to reduce toxicity and cells were harvested, washed 3x, and pelleted 48 hours after transfection. Nuclease expression was confirmed by anti-HA (#3724S, Cell Signaling Technologies) western blot and surface MHC expression was confirmed by flow cytometry (#343330, Biolegend). MAPPs analysis was performed by Abzena.

2.5.2 Computational mutation analysis

Rosetta[91,92] was used to design and evaluate mutations in epitope regions identified in MAPPs. Crystal structures (PDB accessions 5AXW, 5B43) were downloaded from the PDB and refined in Rosetta. Mutations were scored with a modified Ref2015 score function[93] with additional score terms for evaluating peptide-MHC class I binding[94,95]. PSSMs for a panel of 14 MHC class I alleles were created by calculating 9mer amino acid frequencies for all peptides in the IEDB [96] with a binding affinity less than 500 nM. These PSSMs were used for protein design using Rosetta Scripts. Final models were curated by manual inspection.

2.5.3 NetMHCpan 4.1 predictions

NetMHCpan 4.1[97] was used to iteratively evaluate the effect of mutations on the binding affinity of epitope sequences to a panel of 14 MHC alleles (HLA-A31:01, HLA-A02:01, HLA-B08:01, HLA-B39:01, HLA-B27:05, HLA-A24:02, HLA-B15:01, HLA-A01:01, HLA-B58:01, HLA-A26:01, HLA-B07:02, HLA-B41:01, HLA-C17:01, HLA-C02:02). NetMHCpan 4.1 was used to predict

binding affinities for all 9-mer peptides in the protein sequence in order to ensure that new predicted epitopes were not created by introduction of point mutations. The resulting predicted 9-mer peptides were run through NetMHCpan 4.1 to ensure that the new mutants would decrease predicted binding to all MHC alleles, including HLA-A*0201.

2.5.4 Cloning and selection of single, double, and triple mutant nuclease variants

Single, double, and triple mutants were cloned for SaCas9 into construct pX601 (Addgene #61591) and for AsCas12a into construct pY010 (Addgene #69982). All mutations were cloned using site-directed mutagenesis using PhusionFlash for amplification and KLD for digestion and ligation. Guides were cloned into gRNA scaffold-containing constructs for SaCas9 (Addgene #70709) and AsCas12a (Addgene #pY020) respectively. Following assessment of indel activity for each single mutant, all mutants that retained activity that was 50% or greater than that of wild-type activity were selected as “top-performing” and double mutants were made with combinations of the single mutants. This was repeated for the creation and evaluation of the triple mutants. Specifically, for SaCas9, double mutants were made by creating the ‘top-performing’ mutations in epitope 1 and 2 and triple mutants were made by adding mutations in epitope 3 to the top-performing double mutants. For AsCas12a, double mutants were made as two-mutation combinations of all top-performing single mutations from unique epitopes and triple mutants were created as three-mutation combinations at unique epitopes of the top-performing double mutations.

2.5.5 HEK293FT cell line transfection

Human embryonic kidney 293FT cells (Life Technologies) were maintained in Dulbecco’s modified Eagle’s Medium (DMEM) supplemented with 10% FBS (HyClone), 2 mM GlutaMAX (Life Technologies), 100 U/ml penicillin, and 100 µg/ml streptomycin at 37°C with 5% CO_2 incubation. Cells were seeded into 96-well plates (Corning) one day prior to transfection at a density of 150,000 cells per well, and transfected at 70–80% confluency using Lipofectamine 3000 (Life Technologies) following the manufacturer’s recommended protocol. For each well of a 96-well plate, a total of 200 ng DNA was used.

2.5.6 DNA isolation from liver tissue

Genomic DNA was extracted using the QIAmp DNA mini kit (#51304, Qiagen) according to the manufacturer’s instructions. Tissue samples were cut into smaller fragments from bulk tissue, resuspended in tissue lysis buffer with Proteinase K, and digested overnight at 56°C. Following this, DNA was extracted using column-based purification and quantified on a NanoDrop.

2.5.7 Sequencing and indel analysis

Following DNA extraction, 20 ng of gDNA per sample was used for the first PCR. The PCR products from the first reaction were then normalized and 20 ng of the purified product was annealed with the Illumina adapter and barcode sequences for the second PCR. The resulting

product was isolated, purified, and analyzed using MiSeq. The primers used for PCR are in Supplementary Information. CRISPRESSOv2[98] was used to analyze the frequency of indels within each amplicon.

2.5.8 In vitro ELISpot assay

We used HLA-A*0201 specific patient peripheral blood mononuclear cells (PBMCs) (LP_159, Cellular Technology Limited) to profile HLA-A2 patient specific CD8 T cell reactivity for all ELISpot assays. Predicted SaCas9 and AsCas12a peptides, as listed in Figure 2a, were synthesized from Genscript with >98% purity. Each peptide was reconstituted according to manufacturer's solubility instructions and made up to 1 mg/ml in H_2O. A human IFNγ precoated ELISpot kit was used to detect antigen-reactive T cells (EL285, R&D Systems). A total of 5e5 PBMCs were plated per well of a 96-well plate in ImmunoCult-XF T-cell Expansion media (#10981, STEMCELL Technologies) with recombinant IL-2 (#78036.3, STEMCELL technologies) and stimulated with 10 µg/ml of peptide in IFNγ precoated ELISpot plates for 48 hours at 37°C. Plates were developed according to the manufacturer's instructions. DMSO was used as a negative control and HLA Class I Control peptide CEF pools at 10 µg/ml (#100-0675, STEMCELL Technologies) were used as a positive control.

2.5.9 AAV production

For production of AAV virus, HEK293FT cells (Life Technologies) were maintained as described above in 150-mm plates. For each transfection, 10 µg of pAAV2/8 serotype packaging plasmid (#112864, Addgene), 12 µg of pAdDeltaF6 helper plasmid (#112867, Addgene), and 6 µg of ITR flanked plasmid carrying the nuclease construct of interest were added to 1 mL of serum-free DMEM. 125 µL of PEI "Max" solution (1 mg/mL, pH = 7.1) was then added to the mixture and incubated at room temperature for 10 minutes. After incubation, the mixture was added to 25 mL of warm maintenance media and applied to each dish to replace the old growth media. Cells were harvested between 48 hours and 72 hours post transfection by scraping and pelleting by centrifugation. The AAV2/8 (AAV2 ITR vectors pseudotyped with AAV8 capsid) viral particles were then purified from the pellet according to a previously published protocol [99].

2.5.10 Tagmentation based tag-integration site sequencing

The target site specificity of SaCas9 and AsCas12a in the human genome was tested with Tagmentation-based tag integration site sequencing (TTISS[81]). 2e5 HEK293FT cells in 12-well plates were co-transfected with combinations of SaCas9/AsCas12a expression plasmid (800 ng) and target gRNA expression plasmid (200 ng). After 3 days of incubation at 37°C, the supernatant was removed and total DNA of the cells was extracted by DNeasy Blood & Tissue Kit (Qiagen 69506). The 2 µg of extracted DNA was tagmented by house-made loaded Tn5 in TAPS buffer (50 mM TAPS-NaOH pH8.5, 25 mM $MgCl_2$) in 62.5-µL reactions. Reactions were mixed with 312.5 µL PB buffer (Qiagen) and purified on a silica spin column. Tagmented DNA was eluted in 25 µL water, and amplified in 100 µL PCR reactions using KOD Hot Start Master Mix (Millipore 71842) under the following thermal cycling conditions: 1 cycle, 94°C, 2 minutes; 12 cycles, 98°C, 10 seconds, 60°C, 30 seconds, 68°C, 1 minute; 1 cycle, 68°C, 3 minutes; 4°C

hold. 3 µL of this first PCR product was used as the template for each 50 µl-second PCR reaction: 1 cycle, 94°C, 2 minutes; 20 cycles, 98°C, 10 seconds, 65°C, 30 seconds, 68°C, 1 minute; 1 cycle, 68°C, 3 minutes; 4°C hold (total 32 cycles for first and second PCR reactions). PCR products from 6 different experimental conditions were pooled together, purified, and 300–1000-bp fragments were enriched using a 2% agarose gel. After column purifications, the resulting libraries were sequenced using a NextSeq v2 kit (Illumina), 75 cycle kit with 59 forward cycles, 25 reverse cycles and Index1 8 cycles. Reads were mapped to human genome version hg38.2 using http://BrowserGenome.org [100] following the procedure previously described[81].

2.5.11 In vivo *AAV treatment and processing*

HLA-A*0201 HLA-DRA*0101 HLA-DRB1*0101 transgenic mice devoid of mouse MHC [A2.DR1 mice, B6-Tg(HLA-DRA*0101, HLA-DRB1*0101)^{1Dmz} Tg(HLA-A/H2-D/B2M)^{1Bpe} H2-Ab1^{tm1Doi} B2m^{tm1Unc} H2-D1^{tm1Bpe}] were provided by M. Berard and bred at the DKFZ animal facility[82]. Mice were housed under Specific and Opportunistic Pathogen Free (SOPF) conditions and 12-hour day/night cycles. All animal procedures followed the institutional laboratory animal research guidelines and were approved by the governmental authorities (Regional Administrative Authority Karlsruhe, Germany. Animal approval protocol G256/18). AAV vectors were delivered to 12-14 week old male A2.DR1 mice intravenously via lateral tail vein injection. An absolute dose of 2e11 vg of the respective AAV vector per treatment was administered to each animal. Fourteen days after the first treatment, mice were re-treated with 2e11 vg of the respective AAV vector. All dosages of AAV were adjusted to 100 µL volume with sterile phosphate buffered saline (PBS), pH 7.4 (Gibco) before the injection. Mice received 50 µg agonistic αCD40 antibody (FGK4.5, Bioxcell) and 50,000 iU hIL2 (Roche) at the first and second immunization as well as 24 h after the first immunization. Animals were not immunosuppressed or otherwise handled differently prior to injection or during the course of the experiment. The animals were randomized to the different experimental conditions, with the investigator not blinded to the assignments.

For terminal procedures to organs and larger serum volumes for chemistry panels, mice were euthanized by terminal Ketamine/Xylazine injection. Subsequently, blood was collected via cardiac puncture, followed by collection of the spleen. Transcardial perfusion with 30 mL PBS removed the remaining blood, after which liver and lymph node samples were collected. The median lobe of liver was removed and snap-frozen on dry-ice for subsequent DNA extraction.

2.5.12 T cell recall from AAV-treated mice

Spleens were mashed through a 70-µm strainer. Contaminating erythrocytes were lysed using ACK lysis buffer (Gibco). ELISpot was performed as previously described[101]. Briefly, wetted ELISpot plates (MAIPSWU10, Millipore) were incubated with 100 µL 15 µg/mL IFNγ coating antibody (AN-18, Mabtech) and incubated overnight at 4°C. Cells were resuspended in RPMI-1640 supplemented with 10% FBS, P/S as above, 50 µmol/L beta-mercaptoethanol (Sigma), 2 mmol/L L-glutamine, 25 mmol/L Hepes, 1 mmol/L sodium pyruvate (all Invitrogen), and 0.1 mmol/L nonessential amino acids (NEAA, Lonza; called T-cell medium, TCM). IFNγ coating antibody was removed, and ELISpot plates were blocked with TCM. 500,000 cells were

plated, and peptides were added at 10 µg/mL. As positive control, 20 ng/mL phorbol 12-myristate 13-acetate (PMA) and 1 µg/mL ionomycin were used. Plates were incubated for 40 hours. Cells were removed and plates were incubated with 1 µg/mL biotinylated IFNγ detection antibody (R4–6A2, Mabtech) in PBS with 0.5% FBS for 2 hours at room temperature. The detection antibody was removed, and wells were incubated with 1 µg/mL streptavidin–alkaline phosphatase (ALP; Mabtech) in PBS with 0.5% FBS for 1 hour. Streptavidin-ALP was removed, and the plate was incubated with ALP development buffer (Bio-Rad) until distinct spots emerged. Spots were quantified with an ImmunoSpot Analyzer (Cellular Technology Ltd).

2.5.13 Multiplex cytokine detection from mouse sera

Heart blood was aspirated from sacrificed mice using EDTA-precoated syringes and serum was isolated by centrifugation at 1,000–2,000 x g for 10 minutes in a refrigerated centrifuge. Serum samples were then diluted 1:200 with PBS and tested using the Milliplex MAP Mouse Cytokine/Chemokine Magnetic Bead Panel (Millipore, Cat No. MCYTOMAG-70K-PMX) according to the manufacturer‘s protocol. Data was collected on a Luminex detection system (Luminex) and data analysis was completed using the BELYSA® 1.1.0 software. The data collected by the instrument software are expressed as Median Fluorescence Intensity (MFI). MFI values for each analyte were collected per each individual sample well. Analyte standards, quality controls, and sample MFI values were adjusted for background. Calibrator data were fit to either a five-parameter logistic (5PL) or four parameter logistic (4PL) model depending on best fit to produce accurate standard curves for each analyte. Quality control and sample data were interpolated from the standard curves and then adjusted according to dilution factor to provide calculated final concentrations of each analyte present in the sample.

2.5.14 PCSK9 serum ELISA

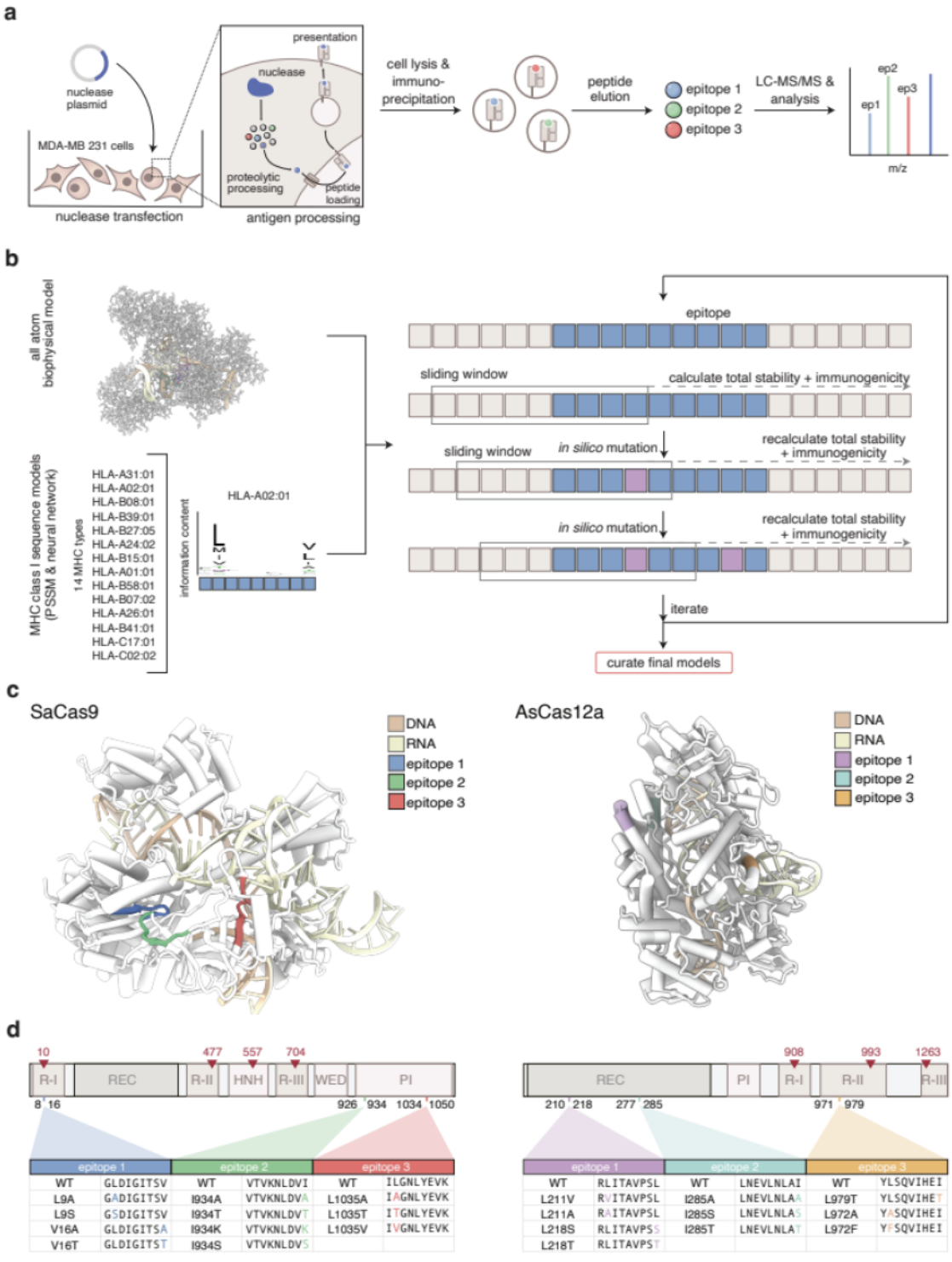
Serum samples were diluted 1:200 with PBS and tested using the Mouse Proprotein Convertase 9/PCSK9 Quantikine ELISA Kit (R&D Systems, Cat No. MPC900) according to the manufacturer’s protocol. Data were collected by $OD_{450nm-570nm}$ measurement. PCSK9 concentrations were calculated corresponding to diluted standards and control samples and were fit to a linear model to interpolate serum concentrations.

2.6 Author Contributions

I.K., L.N., and F.Z. conceived the study. R.R., M.J.F., I.K., and F.Z. designed the experiments. R.R., M.J.F., I.K., M.K. and M.P. performed experiments and analyzed data. F.Z., R.K.M., Y.S., and L.N. supervised the work. R.R., M.J.F., R.K.M., and F.Z. wrote the manuscript with input from all authors.

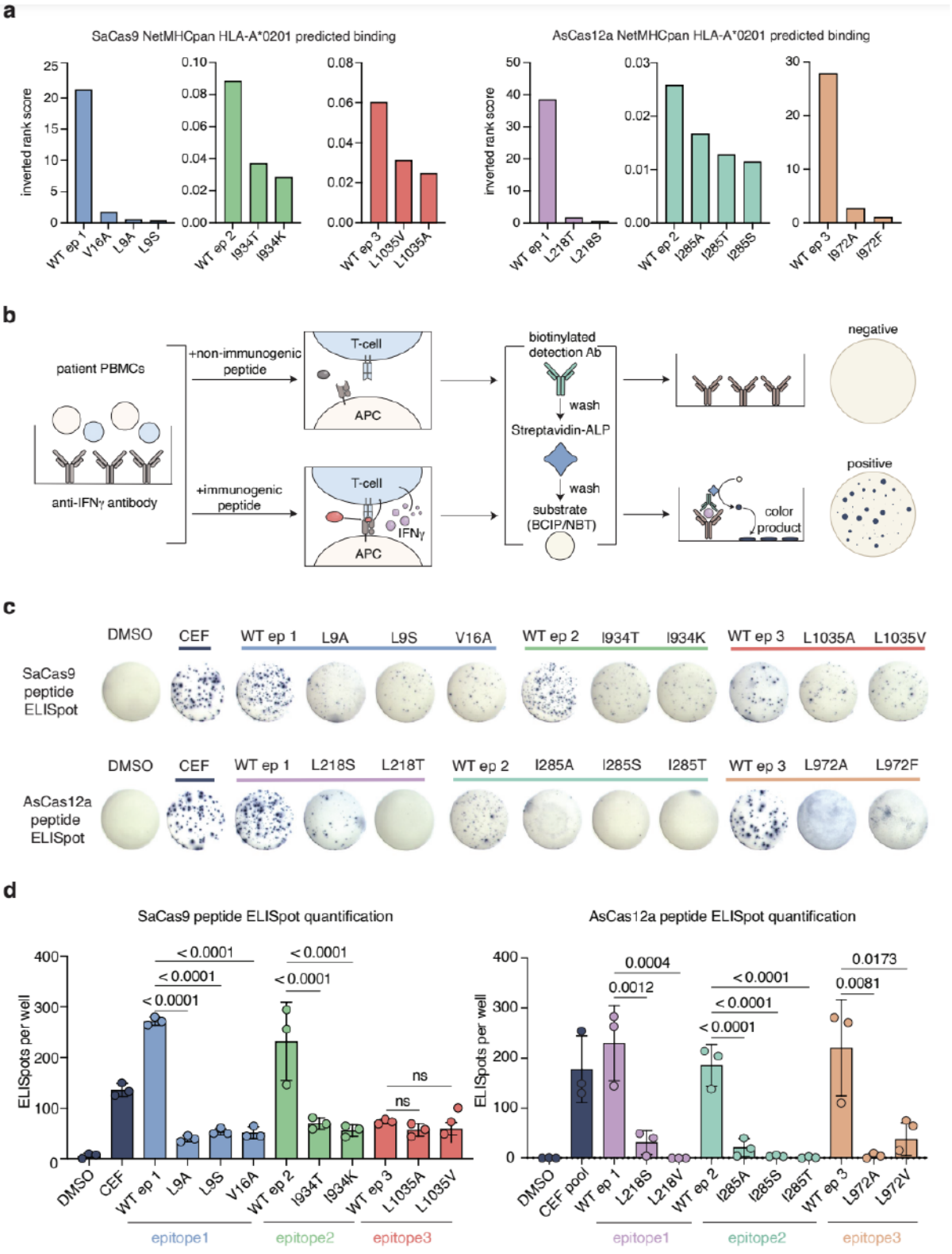
2.7 Main Figures and Figure Legends

2.7.1 Figure 1 | Prediction of SaCas9 and AsCas12a epitopes with reduced binding to MHC I molecules



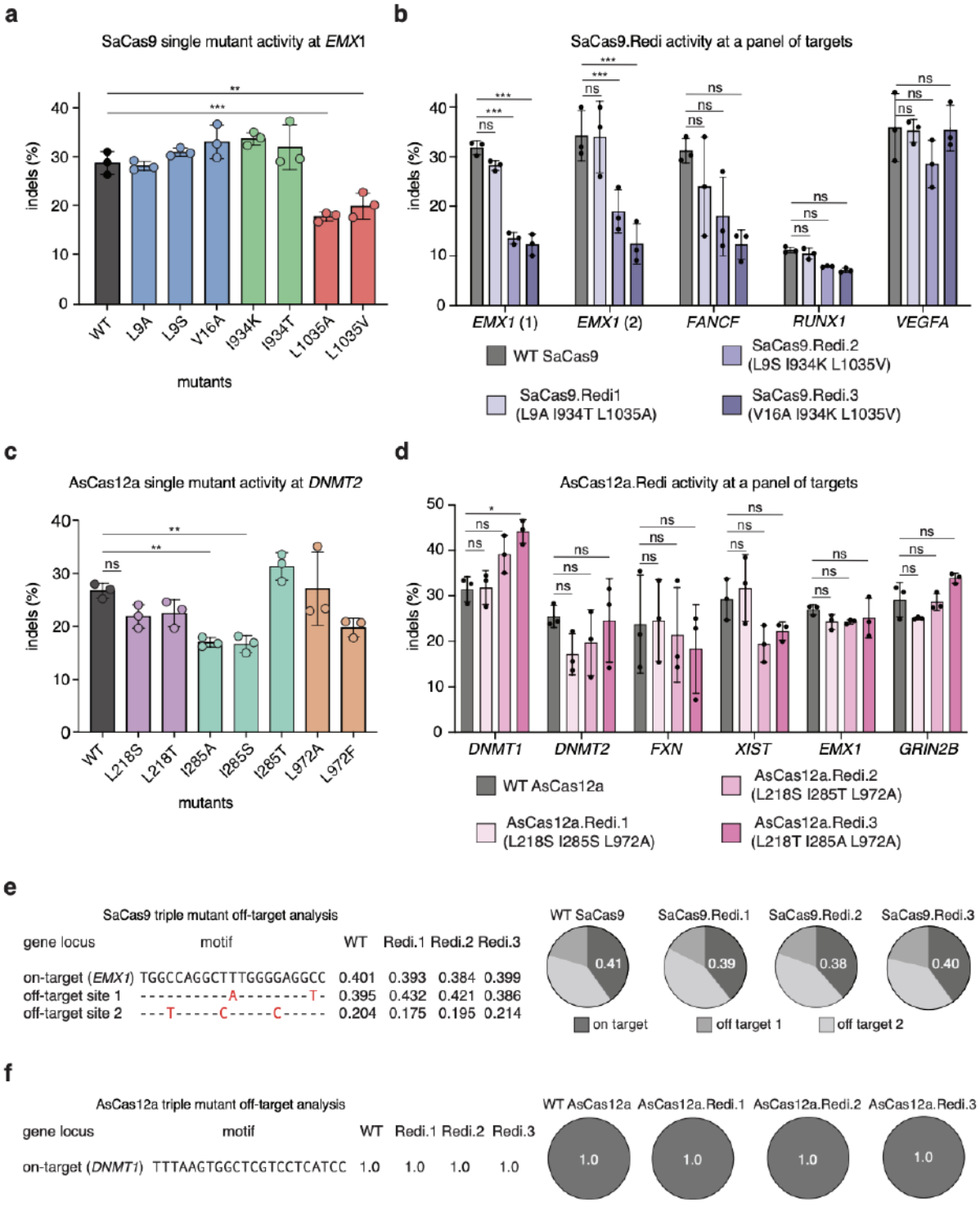
a. Schematic of MAPPs analysis to identify epitopes from SaCas9 and AsCas12a that bind to MHC I molecules.
b. Computational workflow to nominate mutations predicted to abrogate epitope binding to MHC I molecules while maintaining nuclease function. Crystal structures were used to create all-atom protein models in Rosetta. Epitope regions identified in MAPPs were targeted for mutational analysis, along with adjacent n-terminal and c-terminal subsequence frames to ensure that new epitopes were not created for any overlapping peptide subsequences. A computational protein design method utilized 14 MHC Class I PSSM models to introduce mutations predicted to eliminate MHC binding of epitope peptides while avoiding the creation of new predicted epitopes and maintaining predicted protein stability. Final models were evaluated using NetMHCpan and Rosetta.
c. Location of immunogenic epitopes on SaCas9 (left) and AsCas12a (right).
d. Sequences of immunogenic epitopes. Domain architecture of SaCas9 (left) and AsCas12a (right) with catalytic sites shown in red above and location of immunogenic epitopes indicated below. Sequences of immunogenic epitopes and proposed single amino acid mutations for each epitope are listed below R-I, RuvC-I; REC, recognition domain; R-II, RuvC-II; HNH, HNH nuclease; R-III, RuvC-III; WED, wedge domain; PI, PAM-interacting domain.

2.7.2 Figure 2 | SaCas9 and AsCas12a peptides with single point mutations are less immunogenic *in vitro*



A. Inverted rank scores for predicted binding between HLA-A*0201and SaCas9 (left) and AsCas12a (right) wild-type and predicted low-immunogenic peptides based on NetMHCpan 4.1 predictions. An inverted rank score > 2 indicates strong binding and an inverted rank score < 2 but > 0.5 indicates weak binding.
B. Schematic of ELISpot assay.
C. Representative ELISpot images from peptide-treated PBMCs from HLA-A*0201 healthy donors (see Extended Data Fig. 1A and 1B for additional images).
D. Quantification of ELISpot images for SaCas9 (left) and AsCas12a (right). Bars indicate average ELISpot spots for three technical replicates for each peptide condition. Significance comparisons were assessed with an ANOVA and are shown with an asterisk.

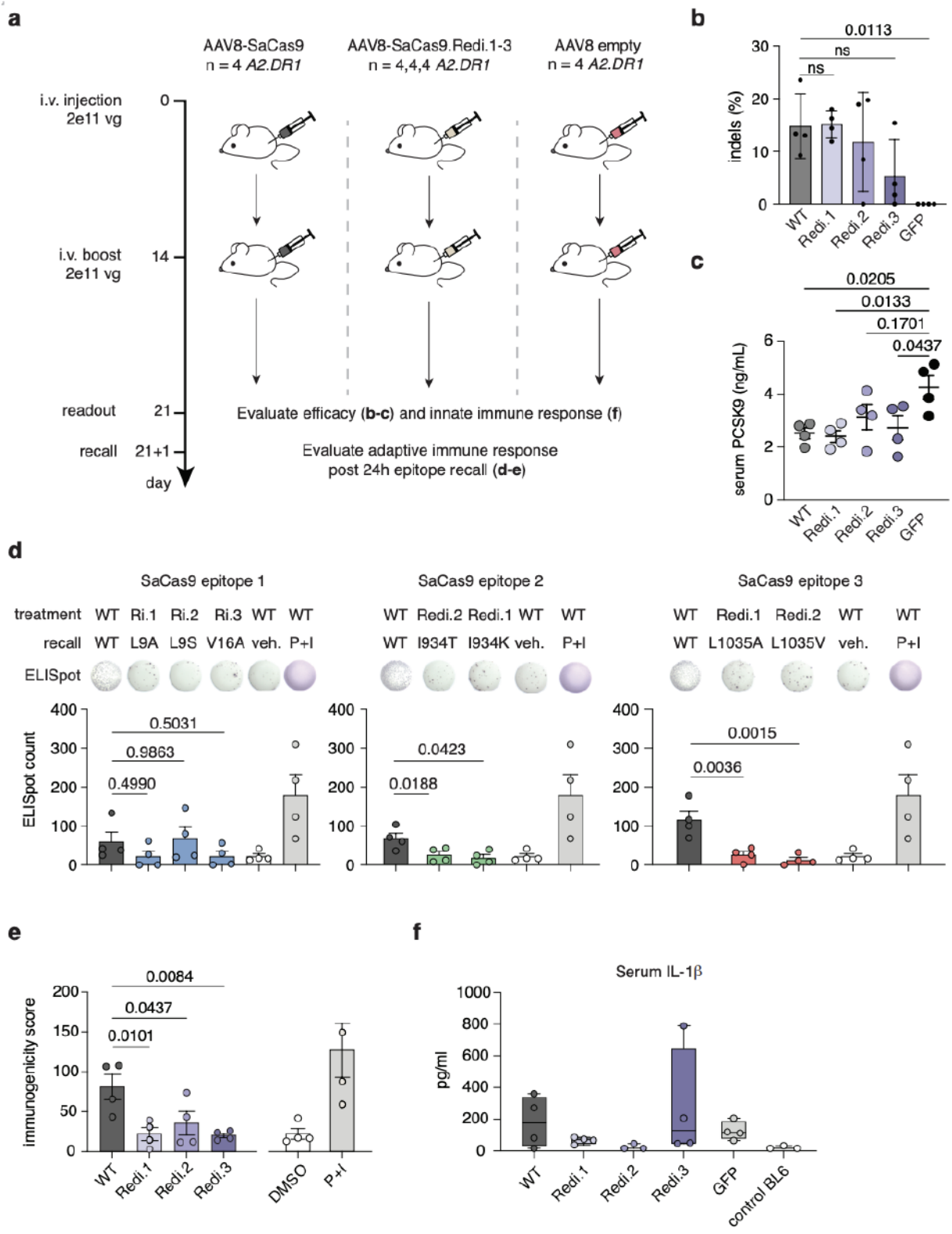
2.7.3 Figure 3 | SaCas9 and AsCas12a reduced immunogenicity (Redi) variants retain activity and specificity



a. Indel rates for wild-type (WT) SaCas9 and single-point mutant variants at *EMX1*. Error bars represent three biological replicates.

b. Indel rates for WT SaCas9 and Redi variants at a panel of targets. Error bars represent three biological replicates. SaCas9.Redi1 contains mutations L9A, I934T, L1035A. SaCas9.Redi.2. contains mutations L9S, I934K, and L1035V and SaCas9.Redi.3 contains mutations V16A, I934K, L1035V.
c. Indel rates for WT AsCas12a and single-point mutant variants at *DNMT2*. Error bars represent three biological replicates.
d. Indel rates for WT AsCas12a and Redi variants at a panel of targets. Error bars represent three biological replicates. AsCas12a.Redi.1 contains mutations L218S, I285S, L972A. AsCas12.Redi.2 contains mutations L218S, I285T and L972A. AsCas12a.Redi.3 contains mutations L218T, I285A, and L972A.
e. Off-target analysis for WT SaCas9 and Redi variants using an *EMX1*-targeting guide
f. Off-target analysis for WT AsCas12 and Redi variants using a *DNMT1*-targeting guide.

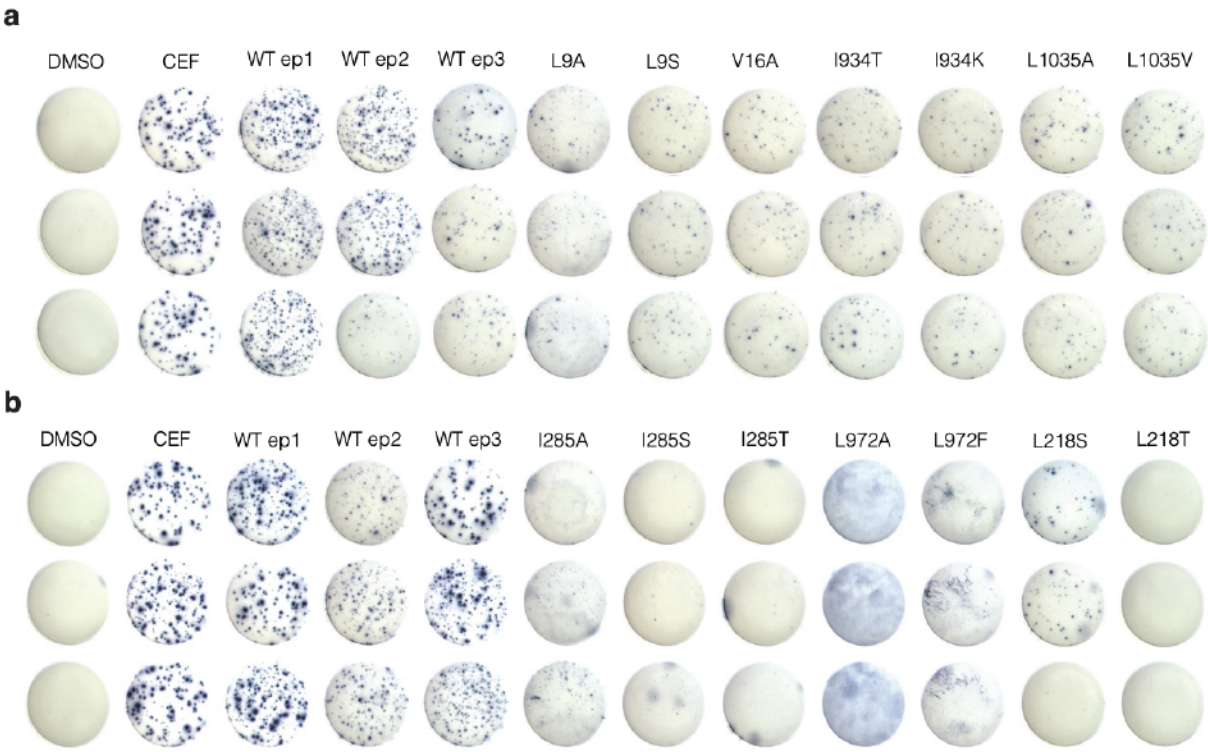
2.7.4 Figure 4. | SaCas9.Redi variants show reduced immunogenicity *in vivo*



a. Schematic of experiment to assay editing efficiency of *Pcsk9* and innate and adaptive immune responses to wild-type (WT) SaCas9 or SaCas9.Redi variants. A2.DR1 mice were treated with 2e11 vg of AAV8-encoded nuclease variants at d0 and d14. At d21, splenocytes were isolated and T cell memory recall was performed by exposing splenocytes with 3 immunodominant epitopes of each nuclease variant. Concentrations of PCSK9, inflammatory cytokines and liver function parameters were measured from heart blood-derived mouse sera.
b. *Pcsk9* indel rates measured in liver tissue after 21 days post-treatment. N=4 biological replicates per condition. Mean with standard deviation is shown. Statistical significance was determined by one-way ANOVA followed by Dunnett test for multiple hypothesis correction.
c. PCSK9 measurements in sera of AAV8-treated animals. Statistical significance was determined by one-way ANOVA followed by Dunnett test for multiple hypothesis correction.
d. Quantification of SaCas9 T cell recalls split by epitope and respective variant mutation. IFN-γ ELISpot counts shown. Full counts in Extended Data Table 5. N=4 animals per epitope.
e. Quantification of overall nuclease immunogenicity. Immunogenicity score summarizes the autologous T cell response to each included epitope. DMSO as vehicle control, PMA+Ionomycin (P+I) as positive control. Statistical significance was determined by one-way ANOVA followed by Sidak post hoc test for multiple hypothesis correction.
f. Multiplex cytokine measurement in sera of AAV8-treated animals. Heatmap depicting median concentrations of indicated inflammatory cytokines measured in sera of animals treated with AAV-SaCas9-WT (n=3) and AAV-SaCas9.Redi1-3 (n=3 each).

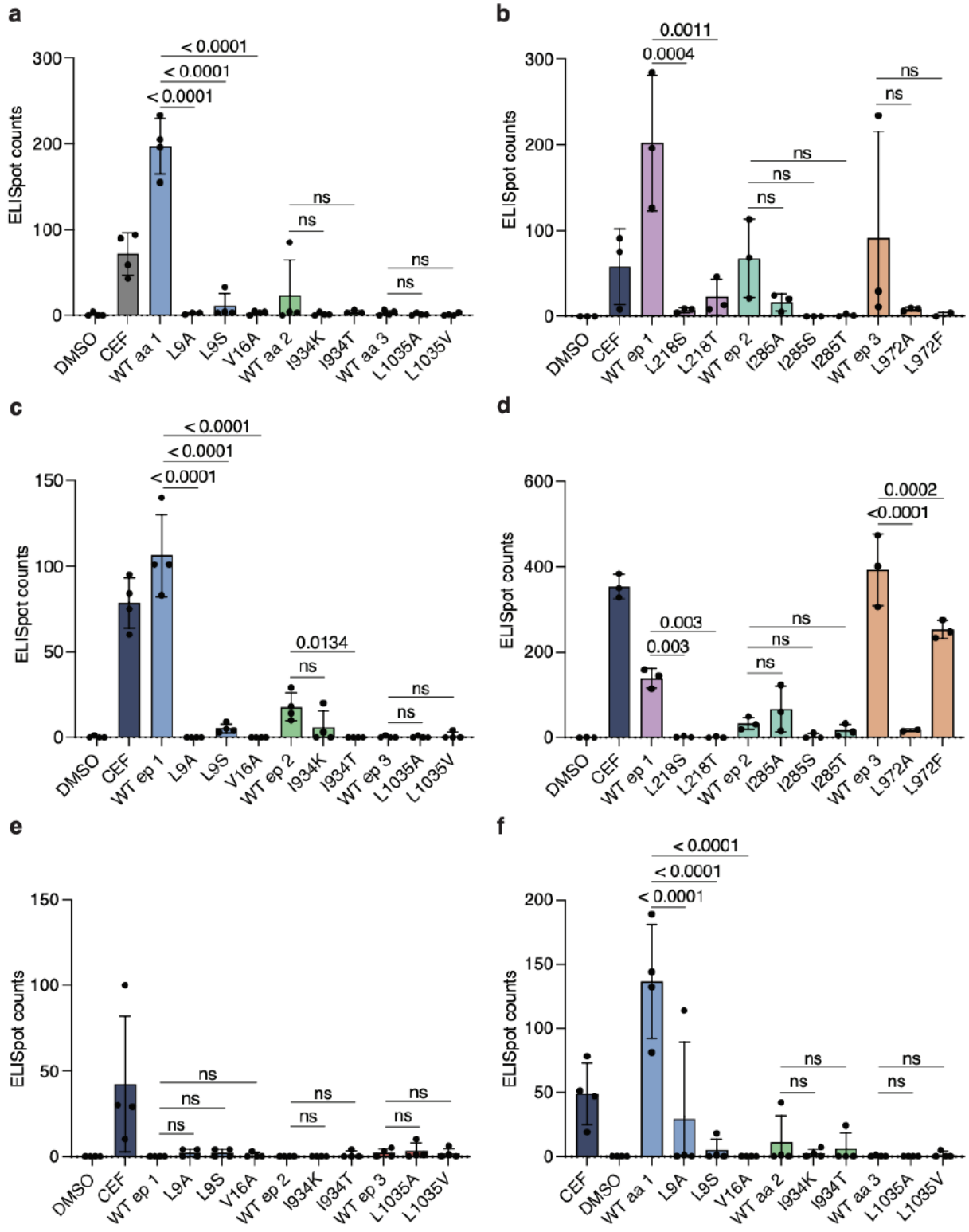
2.8 Extended figure legends

2.8.1 Extended Data Fig. 1. | Expanded ELISpot images of SaCas9 and AsCas12a peptides



a. Full ELISpot images for patient 1 treated with SaCas9 peptides and variants
b. Full ELISpot images for patient 1 treated with AsCas12a peptides and variants

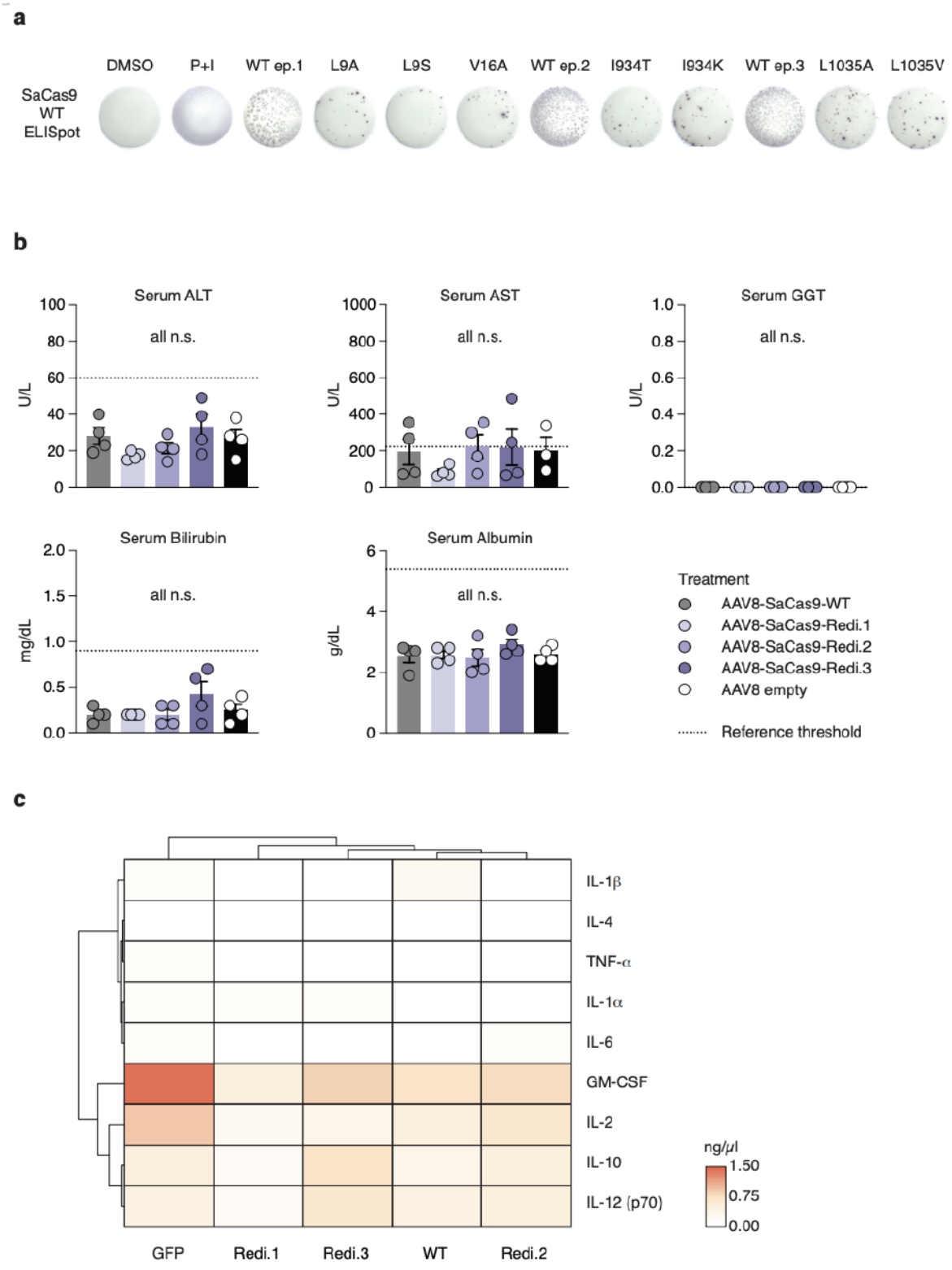
2.8.2 Extended Data Fig. 2. | ELISpot quantification with HLA-A*02:01 patient samples



a. ELISpot quantification for patient 2 treated with SaCas9 peptides and variants

b. ELISpot quantification for patient 2 treated with AsCas12a peptides and variants
c. ELISpot quantification for patient 3 treated with SaCas9 peptides and variants
d. ELISpot quantification for patient 3 treated with AsCas12a peptides and variants
e. ELISpot quantification for patient 4 treated with SaCas9 peptides and variants
f. ELISpot quantification for patient 5 treated with SaCas9 peptides and variants

2.8.3 Extended Data Fig. 3. | Immunogenicity and liver toxicity profiles of wild-type SaCas9 and SaCas9.Redi variants



a. Representative IFN-γ ELISpot images for wild-type SaCas9 exposed splenocytes after T cell recall with immunodominant wild-type epitopes 1-3 and mutant peptides.
b. Bar charts indicating biochemical measurements of serum ALT, AST, albumin and total bilirubin. Reference thresholds for adult C57BL/6J mice indicated as dotted lines. Sera from n=4 animals per condition as biological replicates. Statistical significance was determined by one-way ANOVA followed by Sidak post hoc test for multiple hypothesis correction.
c. Multiplex cytokine measurement in sera of AAV8-treated animals. Heatmap depicting median concentrations of indicated inflammatory cytokines measured in sera of animals treated with AAV-SaCas9-WT (n=3) and AAV-SaCas9.Redi1-3 (n=3 each).

2.8.4 Extended Data Table 1. | Computationally predicted binding affinities for SaCas9 and AsCas12a

SaCas9 peptide	seq	Mutation	Nat Binders	Mut Binders	Nat Strong Binders	Mut Strong Binders	Nat Global Binders	Mut Global Binders
8	GLDIGITSV	L9A	5	7	2	1	1599	1596
		L9F	5	6	2	0	1599	1601
		L9S	5	5	2	1	1599	1594
		V16A	5	4	2	1	1599	1597
		V16T	5	2	2	0	1599	1595
926	VTVKNLDVI	I934A	4	0	0	0	1599	1594
		I934S	4	0	0	0	1599	1593
		I934T	4	0	0	0	1599	1591
		I934K	4	2	0	0	1599	1595
1034	ILGNLYEVK	L1035A	1	0	0	0	1599	1597
		L1035T	1	1	0	0	1599	1597
		L1035V	1	1	0	0	1599	1597

AsCas12a peptide	seq	Mutation	Nat Binders	Mut Binders	Nat Strong Binders	Mut Strong Binders	Nat Global Binders	Mut Global Binders
210	RLITAVPSL	L2111V,I212A	13	11	4	3	2231	2230
		L218T	13	4	4	1	2231	2224
		L211A,L218A	13	7	4	0	2231	2223
		L211V,L218A	13	9	4	0	2231	2227
		L218S	13	2	4	0	2231	2217
277	LNEVLNLAI	I285V	0	0	0	0	2231	2232
		I285T	0	0	0	0	2231	2230
		I285A	0	0	0	0	2231	2231
		I285S	0	0	0	0	2231	2230
972	YLSQVIHEI	L972F	10	9	3	3	2231	2233
		I979A	10	9	3	3	2231	2229
		I979T	10	3	3	1	2231	2221
		L972A,I979V	10	9	3	3	2231	2228
		I979A,I907F	10	9	3	3	2231	2223
		I979T,I907F	10	3	3	1	2231	2225

2.8.5 Extended Data Table 2. | ELISpot quantification for SaCas9 and AsCas12a peptide treatment of PBMCs *in vitro*

SaCas9 peptide treated patient 1				
DMSO	1	10	8	
CEF	150	133	125	
WT ep 1	271	264	280	
L9A	34	39	46	
L9S	50	50	61	
V16A	45	65	45	
WT ep 2	294	256	146	
I934T	69	81	59	
I934K	65	59	43	
WT ep 3	78	69	73	
L1035A	69	58	44	
L1035V	73	49	56	
SaCas9 peptide treated patient 2				
DMSO	0	4	0	0
CEF	90	94	59	43
WT aa 1	205	233	196	155
L9A	0	1	3	3
L9S	3	33	3	4
V16A	5	0	4	3
WT aa 2	0	4	85	3
I934K	1	1	0	4
I934T	6	3	3	N/A
WT aa 3	1	4	0	6
L1035A	1	0	1	3
L1035V	0	0	1	1
SaCas9 peptide treated patient 3				
DMSO	0	0	1	0
CEF	75	60	95	84
WT ep 1	83	101	140	101
L9A	0	0	0	0
L9S	9	5	4	3
V16A	0	0	0	0
WT ep 2	29	10	18	14
I934K	20	0	0	3
I934T	0	0	0	0
WT ep 3	1	0	0	0
L1035A	1	0	0	0
L1035V	0	4	0	0

SaCas9 peptide treated patient 4

DMSO	0	0	0	0
CEF	10	30	100	29
WT ep 1	0	0	0	0
L9A	4	0	0	4
L9S	4	0	0	4
V16A	3	0	0	0
WT ep 2	0	0	0	0
I934K	0	0	0	0
I934T	0	0	0	4
WT ep 3	0	5	3	0
L1035A	1	0	10	1
L1035V	1	0	6	0

SaCas9 peptide treated patient 5

CEF	19	47	78	51
DMSO	0	0	0	0
WT aa 1	81	144	189	132
L9A	1	118	0	0
L9S	0	1	18	0
V16A	0	0	0	0
WT aa 2	0	0	1	42
I934K	2	7	0	0
I934T	0	0	0	24
WT aa 3	1	0	0	0
L1035A	0	0	0	0
L1035V	0	0	5	0

AsCas12a peptide treated patient 1

DMSO	0	1	0
CEF pool	130	254	149
WT ep 1	263	144	283
L218S	40	49	4
L218V	0	0	0
WT ep 2	138	214	204
I285A	4	19	40
I285S	3	3	6
I285T	0	3	1
WT ep 3	110	271	281
L972A	0	10	4
L972V	23	75	15

AsCas12a peptide treated patient 2

DMSO	0	0	0
CEF	8	91	75
WT ep 1	196	284	126
L218S	8	9	4
L218T	8	13	46
WT ep 2	68	113	21
I285A	24	5	20
I285S	0	0	0
I285T	1	3	0
WT ep 3	29	11	234
L972A	6	10	9
L972F	4	N/A	0

AsCas12a peptide treated patient 3

DMSO	0	0	1
CEF	350	328	384
WT ep 1	159	145	114
L218S	1	1	4
L218T	0	3	0
WT ep 2	49	31	20
I285A	61	123	16
I285S	0	11	0
I285T	13	33	4
WT ep 3	401	474	306
L972A	N/A	19	15
L972F	234	248	276

2.8.6 Extended Data Table 3. | Indel rates of SaCas9 mutants

Single Mutants

Nuclease	Non-Targeting			Average	St Dev
WT	0.80	1.04	0.59	0.81	0.22
L9A	0.78	0.80	0.82	0.80	0.02
L9F	0.46	0.92	0.42	0.60	0.28
L9S	0.61	0.47	0.34	0.47	0.13
V16A	0.80	0.79	0.44	0.68	0.21
V16T	0.73	0.77	0.48	0.66	0.15
I934A	0.56	0.57	0.31	0.48	0.15
I934K	0.80	0.58	0.53	0.64	0.14
I934S	0.58	0.69	0.52	0.59	0.09
I934T	0.55	0.68	0.55	0.59	0.07
L1035A	0.80	0.78	0.57	0.72	0.13
L1035T	0.46	0.92	0.42	0.60	0.28
L1035V	0.61	0.47	0.34	0.47	0.13

Single Mutants

Nuclease	Targeting (EMX1)			Average	St Dev
WT	31.37	26.68	29.21	29.09	2.35
L9A	28.19	29.58	27.73	28.50	0.96
L9F	3.40	3.42	3.40	3.41	0.01
L9S	31.13	32.25	30.61	31.33	0.84
V16A	37.10	33.01	30.19	33.43	3.48
V16T	34.07	27.85	27.14	29.69	3.81
I934A	12.06	5.27	12.82	10.05	4.16
I934K	35.21	32.76	34.23	34.06	1.23
I934S	24.09	29.10	29.48	27.56	3.01
I934T	29.37	29.95	37.68	32.33	4.64
L1035A	18.45	18.78	17.02	18.08	0.93
L1035T	0.74	0.45	0.54	0.58	0.15
L1035V	19.63	17.89	23.07	20.20	2.64

Double Mutants

Nuclease	Non-Targeting			Average	St Dev
WT	0.13	0.26	0.09	0.16	0.09
L9A V16A	0.11	0.11	0.17	0.13	0.03
L9A V16T	0.11	0.11	0.17	0.13	0.03
L9S V16A	0.05	0.10	0.25	0.13	0.11
L9S V16T	0.07	0.56	0.07	0.23	0.29
L9A I934K	0.09	0.27	0.21	0.19	0.09
L9A I934T	0.06	0.16	0.26	0.16	0.10
L9S I934K	0.18	0.18	0.05	0.14	0.08
L9S I934T	0.13	0.02	0.15	0.10	0.07
V16A I934K	0.09	0.17	0.20	0.15	0.06
V16A I934T	0.12	1.08	0.23	0.48	0.52
V16T I934K	0.09	0.17	0.20	0.15	0.06
V16T I934T	0.12	1.08	0.23	0.48	0.52

Double Mutants

Nuclease	Targeting (EMX1)			Average	St Dev
WT	25.59	25.65	25.32	25.52	0.18
L9A V16A	0.26	0.04	0.23	0.18	0.12
L9A V16T	0.26	0.04	0.23	0.18	0.12
L9S V16A	0.11	0.09	0.21	0.14	0.07
L9S V16T	0.11	0.09	0.21	0.14	0.07
L9A I934K	21.51	23.46	22.42	22.46	0.97
L9A I934T	23.07	25.20	22.96	23.74	1.26
L9S I934K	20.68	20.64	19.02	20.12	0.94
L9S I934T	23.51	26.49	25.39	25.13	1.51
V16A I934K	21.50	22.01	22.25	21.92	0.38
V16A I934T	19.91	22.61	20.92	21.15	1.37
V16T I934K	21.50	22.01	22.25	21.92	0.38
V16T I934T	19.91	22.61	20.92	21.15	1.37

Triple Mutants

Nuclease	Non-Targeting			Average	StDev
WT	0.66	0.80	0.23	0.56	0.29
L9A I934K L1035A	0.00	0.12	0.17	0.09	0.09
L9A I934T L1035A	0.44	0.26	0.20	0.30	0.13
L9S I934K L1035A	0.53	0.36	0.18	0.36	0.17
L9S I934T L1035A	0.32	0.49	0.14	0.32	0.18
V16A I934K L1035A	0.37	0.61	0.23	0.40	0.20
V16A I934T L1035A	0.06	0.48	0.23	0.26	0.21
V16T I934K L1035A	0.54	0.39	0.23	0.39	0.15
V16T I934T L1035A	0.37	0.48	0.23	0.36	0.12
L9A I934K L1035V	0.17	0.56	0.23	0.32	0.21
L9A I934T L1035V	1.92	1.11	0.10	1.04	0.92
L9S I934K L1035V	1.26	0.40	0.90	0.85	0.43
L9S I934T L1035V	0.19	0.29	0.23	0.24	0.05
V16A I934K L1035V	0.73	0.45	0.38	0.52	0.18
V16A I934T L1035V	0.79	0.62	0.26	0.55	0.27
V16T I934K L1035V	0.62	0.48	0.59	0.56	0.08
V16T I934T L1035V	0.94	0.72	0.88	0.84	0.11

Triple Mutants

Nuclease	Targeting (EMX1)			Average	Stdev
WT	29.06	25.53	27.09	27.23	1.77
L9A I934K L1035A	0.60	0.76	0.41	0.59	0.18
L9A I934T L1035A	17.06	15.05	15.78	15.97	1.02
L9S I934K L1035A	8.41	8.36	8.11	8.29	0.16
L9S I934T L1035A	0.24	0.46	0.44	0.38	0.12
V16A I934K L1035A	2.67	2.70	1.24	2.20	0.83
V16A I934T L1035A	0.35	0.50	0.31	0.39	0.10
V16T I934K L1035A	0.03	0.12	0.20	0.12	0.08
V16T I934T L1035A	2.70	0.03	0.50	1.08	1.43
L9A I934K L1035V	0.50	0.33	1.07	0.63	0.39
L9A I934T L1035V	6.80	2.96	4.02	4.59	1.98
L9S I934K L1035V	26.31	27.42	29.35	27.69	1.54
L9S I934T L1035V	0.60	0.79	0.24	0.54	0.28
V16A I934K L1035V	19.98	17.29	19.57	18.95	1.45
V16A I934T L1035V	9.03	8.32	11.27	9.54	1.54
V16T I934K L1035V	6.73	4.44	4.73	5.30	1.25
V16T I934T L1035V	10.09	6.65	7.09	7.95	1.87

Panel of Targets for Top performing triple mutants

	pX601 (WT)			L9A I934T L1035A (SaCas9.Redi.1)			L9S I934K L1035V (SaCas9.Redi.2)			V16A I934K L1035V (SaCas9.Redi.3)		
EMX1 site 6	32.73	30.19	32.49	28.41	27.11	29.10	13.63	14.65	10.03	12.46	14.31	12.32
EMX1 site 7	30.68	32.00	40.00	26.00	40.00	36.00	23.81	15.38	11.11	9.09	17.00	17.60
FANCF site 13	29.30	30.30	34.00	14.00	34.00	24.00	12.00	15.00	13.95	8.79	14.00	27.00
RUNX1 site 14	11.84	10.94	10.74	9.30	11.50	10.53	7.70	7.84	6.96	7.57	6.75	7.99
VEGFA site 8	29.20	35.40	43.00	32.60	37.00	36.00	23.80	28.48	33.40	40.60	32.00	33.36

Off-Target TTISS analysis

		Raw Reads			
gene locus	motif	WT	Redi.1	Redi.2	Redi.3
on-target (EMX1)	TGGCCAGGCTTTGGGGAGGCC	1,096.00	2,011.00	1,318.00	2,503.00
off-target site 1	TGGCCAGGCTATGGGGAGGCC	1,077.00	2,214.00	1,442.00	2,420.00
off-target site 2	TGGTCAGGCCTTGGGCAGGCC	557.00	896.00	669.00	1,343.00

2.8.7 Extended Data Table 4. | Indel rates of AsCas12a mutants

Single Mutants

Nuclease	Non-Targeting			Average	StDev
WT	0.00	0.00	0.00	0.00	0.00
L211V	0.00	0.00	0.00	0.00	0.00
L211A	0.00	0.00	0.00	0.00	0.00
L218S	0.00	0.00	0.00	0.00	0.00
L218T	0.00	0.00	0.00	0.00	0.00
I285A	0.00	0.00	0.00	0.00	0.00
I285S	0.00	0.00	0.00	0.00	0.00
I285T	0.00	0.00	0.00	0.00	0.00
I285V	0.00	0.00	0.00	0.00	0.00
L972A	0.00	0.00	0.00	0.00	0.00
L972F	0.00	0.00	0.00	0.00	0.00
I907F	0.00	0.00	0.00	0.00	0.00
I979T	0.00	0.00	0.00	0.00	0.00

Single Mutants

Nuclease	Targeting (DNMT2)			Average	StDev
WT	29.82	28.80	26.85	28.49	1.51
L211V	4.95	6.24	4.28	5.16	1.00
L211A	0.84	0.45	0.41	0.57	0.24
L218S	20.97	23.42	25.65	23.35	2.34
L218T	20.96	24.75	26.22	23.98	2.72
I285A	17.30	17.90	19.14	18.11	0.94
I285S	15.96	18.05	19.35	17.79	1.71
I285T	35.85	30.39	33.98	33.41	2.77
I285V	0.23	0.60	0.41	0.41	0.19
L972A	24.93	37.43	24.45	28.94	7.36
L972F	22.07	22.26	18.99	21.11	1.83
I907F	1.55	1.46	1.32	1.44	0.11
I979T	1.47	1.65	1.80	1.64	0.17

Double Mutants

Nuclease	Non-Targeting			Average	StDev
WT (pY010)	0.00	0.00	0.00	0.00	0.00
L218S I285A	0.00	0.00	0.00	0.00	0.00
L218S I285S	0.00	0.00	0.00	0.00	0.00
L218S I285T	0.00	0.00	0.00	0.00	0.00
L218S L972A	0.00	0.00	0.00	0.00	0.00
L218S L972F	0.00	0.00	0.00	0.00	0.00
L218T I285A	0.00	0.00	0.00	0.00	0.00
L218T I285S	0.00	0.00	0.00	0.00	0.00
L218T I285T	0.00	0.00	0.00	0.00	0.00
L218T L972A	0.00	0.00	0.00	0.00	0.00
L218T L972F	0.00	0.00	0.00	0.00	0.00
I907F L972A	0.00	0.00	0.00	0.00	0.00
I907F L972F	0.00	0.00	0.00	0.00	0.00

Double Mutants

Nuclease	Targeting (DNMT2)			Average	StDev
WT (pY010)	15.63	14.03	16.27	15.31	1.15
L218S I285A	13.23	11.00	13.65	12.63	1.43
L218S I285S	12.55	9.35	12.15	11.35	1.74
L218S I285T	11.29	9.74	13.12	11.39	1.69
L218S L972A	14.06	11.87	15.28	13.74	1.73
L218S L972F	11.73	11.49	13.04	12.08	0.84
L218T I285A	9.68	8.56	13.38	10.54	2.52
L218T I285S	13.54	13.91	12.53	13.32	0.71
L218T I285T	0.00	0.07	0.08	0.05	0.04
L218T L972A	14.51	13.45	13.89	13.95	0.53
L218T L972F	11.32	10.88	7.26	9.82	2.23
I907F L972A	0.00	0.00	0.07	0.02	0.04
I907F L972F	0.05	0.09	0.30	0.14	0.13

Triple Mutants

Nuclease	Non-Targeting			Average	StDev
WT (pY010)	0.00	0.00	0.00	0.00	0.00
L218S I285A L972A	0.00	0.00	0.00	0.00	0.00
L218S I285S L972A	0.00	0.00	0.00	0.00	0.00
L218S I285T L972A	0.00	0.00	0.00	0.00	0.00
L218T I285A L972A	0.00	0.00	0.00	0.00	0.00
L218T I285S L972A	0.00	0.00	0.00	0.00	0.00
L218T I285T L972A	0.00	0.00	0.00	0.00	0.00
L218S I285A L972F	0.00	0.00	0.00	0.00	0.00
L218S I285S L972F	0.00	0.00	0.00	0.00	0.00
L218S I285T L972F	0.00	0.00	0.00	0.00	0.00
L218T I285A L972F	0.00	0.00	0.00	0.00	0.00
L218T I285S L972F	0.00	0.00	0.00	0.00	0.00
L218T I285T L972F	0.00	0.00	0.00	0.00	0.00

Triple Mutants

Nuclease	Targeting (DNMT2)			Average	StDev
WT (pY010)	7.28	7.69	4.76	6.58	1.59
L218S I285A L972A	8.83	7.47	6.70	7.67	1.08
L218S I285S L972A	8.91	4.31	7.85	7.02	2.41
L218S I285T L972A	9.06	6.52	10.82	8.80	2.16
L218T I285A L972A	7.22	7.22	7.41	7.28	0.11
L218T I285S L972A	1.30	0.07	0.17	0.51	0.68
L218T I285T L972A	1.37	0.05	0.62	0.68	0.66
L218S I285A L972F	9.13	9.00	8.03	8.72	0.60
L218S I285S L972F	7.52	3.66	7.64	6.27	2.27
L218S I285T L972F	5.28	6.42	5.45	5.72	0.62
L218T I285A L972F	6.10	5.23	8.74	6.69	1.83
L218T I285S L972F	10.59	5.79	6.42	7.60	2.61
L218T I285T L972F	0.72	0.48	0.69	0.63	0.13

Off-Target TTISS analysis

		Raw Reads			
gene locus	motif	WT	Redi.1	Redi.2	Redi.3
on-target (DNMT1)	TTTAAGTGGCTCGTCCTCATCC	1,096.00	2,011.00	1,318.00	2,503.00

	WT AsCas12a			L218S I285A L972A			L218S I285S L972A (AsCas12.Redi.1)			L218S I285T L972A (AsCas12.Redi.2)			L218T I285A L972A (AsCas12.Redi.3)		
DNMT1	28.23	33.36	32.66	7.98	11.42	10.07	27.59	34.48	33.33	34.99	39.05	43.23	41.31	46.42	44.55
DNMT2	24.39	23.81	28.29	16.97	14.80	21.60	22.22	13.64	15.66	20.49	12.15	26.52	28.06	14.17	31.50
FXN	21.54	14.29	35.45	12.50	6.67	8.00	25.00	33.33	15.38	14.29	16.67	33.33	7.14	22.86	25.00
XIST	24.64	29.38	33.70	11.38	13.27	11.73	32.81	38.39	23.85	19.33	23.52	15.48	19.95	23.55	23.25
EMX1	27.53	25.70	27.63	23.96	26.79	25.38	24.65	25.71	22.53	24.48	23.75	24.63	21.18	24.31	29.78
GRIN2B	24.83	30.80	31.76	6.26	7.02	6.38	25.04	25.39	24.84	27.72	27.41	30.78	34.15	32.66	34.79

2.8.8 Extended Data Table 5. | ELISpot quantification for SaCas9.Redi variants *in vivo*

SaCas9 epitope 1					

WT	L9A	L9S	V16A	vehicle (DMSO)	P+I
42	61	146	3	42	68
38	27	24	30	17	123
133	0	80	0	17	309
25	0	20	56	13	217
SaCas9 epitope 2					
WT	I934T	I934K	vehicle (DMSO)	P+I	
104	43	24	42	68	
60	13	40	17	123	
71	39	0	17	309	
36	8	8	13	217	
SaCas9 epitope 3					
WT	L1035A	L1035V	vehicle (DMSO)	P+I	
178	32	32	42	68	
100	25	0	17	123	
116	44	11	17	309	
69	1	7	13	217	

2.9 Supplementary Information

SaCas9.Redi.1-DNA sequence

gccaagcggaactacatcctgggc**gcg**gacatcggcatcaccagcgtgggctacggcatcatcgactacgagacacgggacgtgatcgatgccggc
gtgcggctgttcaaagaggccaacgtggaaaacaacgagggcaggcggagcaagagaggcgccagaaggctgaagcggcggaggcggcataga
atccagagagtgaagaagctgctgttcgactacaacctgctgaccgaccacagcgagctgagcggcatcaacccctacgaggccagagtgaagggcc
tgagccagaagctgagcgaggaagagttctctgccgccctgctgcacctggccaagagaagaggcgtgcacaacgtgaacgaggtggaagaggaca
ccggcaacgagctgtccaccaaagagcagatcagccggaacagcaaggccctggaagagaaatacgtggccgaactgcagctggaacggctgaag
aaagacggcgaagtgcggggcagcatcaacagattcaagaccagcgactacgtgaaagaagccaaacagctgctgaaggtgcagaaggcctaccac
cagctggaccagagcttcatcgacacctacatcgacctgctggaaacccggcggacctactatgagggacctggcgagggcagccccttcggctgga
aggacatcaaagaatggtacgagatgctgatgggccactgcacctacttccccgaggaactgcggagcgtgaagtacgcctacaacgccgacctgtac
aacgccctgaacgacctgaacaatctcgtgatcaccagggacgagaacgagaagctggaatattacgagaagttccagatcatcgagaacgtgttcaag
cagaagaagaagcccaccctgaagcagatcgccaaagaaatcctcgtgaacgaagaggatattaagggctacagagtgaccagcaccggcaagccc
gagttcaccaacctgaaggtgtaccacgacatcaaggacattaccgcccggaaagagattattgagaacgccgagctgctggatcagattgccaagatc
ctgaccatctaccagagcagcgaggacatccaggaagaactgaccaatctgaactccgagctgacccaggaagagatcgagcagatctctaatctgaa
gggctataccggcacccacaacctgagcctgaaggccatcaacctgatcctggacgagctgtggcacaccaacgacaaccagatcgctatcttcaacc
ggctgaagctggtgcccaagaaggtggacctgtcccagcagaaagagatccccaccaccctggtggacgacttcatcctgagccccgtcgtgaagag
aagcttcatccagagcatcaaagtgatcaacgccatcatcaagaagtacggcctgcccaacgacatcattatcgagctggcccgcgagaagaactccaa

ggacgcccagaaaatgatcaacgagatgcagaagcggaaccggcagaccaacgagcggatcgaggaaatcatccggaccaccggcaaagagaac
gccaagtacctgatcgagaagatcaagctgcacgacatgcaggaaggcaagtgcctgtacagcctggaagccatccctctggaagatctgctgaacaa
ccccttcaactatgaggtggaccacatcatccccagaagcgtgtccttcgacaacagcttcaacaacaaggtgctcgtgaagcaggaagaaaacagcaa
gaagggcaaccggaccccattccagtacctgagcagcagcgacagcaagatcagctacgaaaccttcaagaagcacatcctgaatctggccaagggc
aagggcagaatcagcaagaccaagaaagagtatctgctggaagaacgggacatcaacaggttctccgtgcagaaagacttcatcaaccggaacctggt
ggataccagatacgccaccagaggcctgatgaacctgctgcggagctacttcagagtgaacaacctggacgtgaaagtgaagtccatcaatggcggctt
caccagctttctgcggcggaagtggaagtttaagaaagagcggaacaaggggtacaagcaccacgccgaggacgccctgatcattgccaacgccgat
ttcatcttcaaagagtggaagaaactggacaaggccaaaaaagtgatggaaaaccagatgttcgaggaaaagcaggccgagagcatgcccgagatcg
aaaccgagcaggagtacaaagagatcttcatcacccccaccagatcaagcacattaaggacttcaaggactacaagtacagccaccgggtggacaag
aagcctaatagagagctgattaacgacaccctgtactccacccggaaggacgacaagggcaacaccctgatcgtgaacaatctgaacggcctgtacga
caaggacaatgacaagctgaaaaagctgatcaacaagagccccgaaaagctgctgatgtaccaccacgacccccagacctaccagaaactgaagctg
attatggaacagtacggcgacgagaagaatcccctgtacaagtactacgaggaaaccgggaactacctgaccaagtactccaaaaaggacaacggcc
ccgtgatcaagaagattaagtattacggcaacaaactgaacgcccatctggacatcaccgacgactaccccaacagcagaaacaaggtcgtgaagctgt
ccctgaagccctacagattcgacgtgtacctggacaatggcgtgtacaagttcgtgaccgtgaagaatctggatgtg**acc**aaaaaagaaaactactacga
agtgaatagcaagtgctatgaggaagctaagaagctgaagaagatcagcaaccaggccgagtttatcgcctccttctacaacaacgatctgatcaagatc
aacggcgagctgtatagagtgatcggcgtgaacaacgacctgctgaaccggatcgaagtgaacatgatcgacatcacctaccgcgagtacctggaaaa
catgaacgacaagaggcccccccaggatcattaagacaatcgcctccaagacccagagcattaagaagtacagcacagacatt**gcc**ggcaacctgtatg
aagtgaaatctaagaagcaccctcagatcatcaaaaagggcaaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaagggatcc
tacccatacgatgttccagattacgcttacccatacgatgttccagattacgcttacccatacgatgttccagattacgcttaa

SaCas9.Redi.1-protein sequence

AKRNYILG**A**DIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQR
VKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNE
LSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQ
SFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN
DLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLK
VYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLS
LKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAII
KKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQE
GKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSK
ISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYF
RVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKV
MENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRK
DDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYK
YYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDN
GVYKFVTVKNLDV**T**KKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVN
NDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDI**A**GNLYEVKSKKHPQIIKK
GKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA*

SaCas9.Redi.2-DNA sequence

gccaagcggaactacatcctgggc**TCT**gacatcggcatcaccagcgtgggctacggcatcatcgactacgagacacgggacgtgatcgatgccgg
cgtgcggctgttcaaagaggccaacgtggaaaacaacgagggcaggcggagcaagagaggcgccagaaggctgaagcggcggaggcggcatag

aatccagagagtgaagaagctgctgttcgactacaacctgctgaccgaccacagcgagctgagcggcatcaacccctacgaggccagagtgaagggc
ctgagccagaagctgagcgaggaagagttctctgccgccctgctgcacctggccaagagaagaggcgtgcacaacgtgaacgaggtggaagaggac
accggcaacgagctgtccaccaaagagcagatcagccggaacagcaaggccctggaagagaaatacgtggccgaactgcagctggaacggctgaa
gaaagacggcgaagtgcggggcagcatcaacagattcaagaccagcgactacgtgaaagaagccaaacagctgctgaaggtgcagaaggcctacca
ccagctggaccagagcttcatcgacacctacatcgacctgctggaaacccggcggacctactatgagggacctggcgagggcagccccttcggctgg
aaggacatcaaagaatggtacgagatgctgatgggccactgcacctacttccccgaggaactgcggagcgtgaagtacgcctacaacgccgacctgta
caacgccctgaacgacctgaacaatctcgtgatcaccagggacgagaacgagaagctggaatattacgagaagttccagatcatcgagaacgtgttcaa
gcagaagaagaagcccaccctgaagcagatcgccaaagaaatcctcgtgaacgaagaggatattaagggctacagagtgaccagcaccggcaagcc
cgagttcaccaacctgaaggtgtaccacgacatcaaggacattaccgcccggaaagagattattgagaacgccgagctgctggatcagattgccaagat
cctgaccatctaccagagcagcgaggacatccaggaagaactgaccaatctgaactccgagctgacccaggaagagatcgagcagatctctaatctga
agggctataccggcacccacaacctgagcctgaaggccatcaacctgatcctggacgagctgtggcacaccaacgacaaccagatcgctatcttcaac
cggctgaagctggtgcccaagaaggtggacctgtcccagcagaaagagatccccaccaccctggtggacgacttcatcctgagccccgtcgtgaaga
gaagcttcatccagagcatcaaagtgatcaacgccatcatcaagaagtacggcctgcccaacgacatcattatcgagctggcccgcgagaagaactcca
aggacgcccagaaaatgatcaacgagatgcagaagcggaaccggcagaccaacgagcggatcgaggaaatcatccggaccaccggcaaagagaa
cgccaagtacctgatcgagaagatcaagctgcacgacatgcaggaaggcaagtgcctgtacagcctggaagccatccctctggaagatctgctgaaca
acccttcaactatgaggtggaccacatcatccccagaagcgtgtccttcgacaacagcttcaacaacaaggtgctcgtgaagcaggaagaaaacagca
agaagggcaaccggaccccattccagtacctgagcagcagcgacagcaagatcagctacgaaaccttcaagaagcacatcctgaatctggccaaggg
caagggcagaatcagcaagaccaagaaagagtatctgctggaagaacgggacatcaacaggttctccgtgcagaaagacttcatcaaccggaacctg
gtggataccagatacgccaccagaggcctgatgaacctgctgcggagctacttcagagtgaacaacctggacgtgaaagtgaagtccatcaatggcgg
cttcaccagctttctgcggcggaagtggaagtttaagaaagagcggaacaaggggtacaagcaccacgccgaggacgccctgatcattgccaacgcc
gatttcatcttcaaagagtggaagaaactggacaaggccaaaaaagtgatggaaaaccagatgttcgaggaaaagcaggccgagagcatgcccgagat
cgaaaccgagcaggagtacaaagagatcttcatcacccccaccagatcaagcacattaaggacttcaaggactacaagtacagccaccgggtggaca
agaagcctaatagagagctgattaacgacaccctgtactccacccggaaggacgacaagggcaacaccctgatcgtgaacaatctgaacggcctgtac
gacaaggacaatgacaagctgaaaaagctgatcaacaagagccccgaaaagctgctgatgtaccaccacgacccccagacctaccagaaactgaagc
tgattatggaacagtacggcgacgagaagaatcccctgtacaagtactacgaggaaaccgggaactacctgaccaagtactccaaaaaggacaacggc
cccgtgatcaagaagattaagtattacggcaacaaactgaacgcccatctggacatcaccgacgactaccccaacagcagaaacaaggtcgtgaagct
gtccctgaagccctacagattcgacgtgtacctggacaatggcgtgtacaagttcgtgaccgtgaagaatctggatgtg**aag**aaaaaagaaaactactac
gaagtgaatagcaagtgctatgaggaagctaagaagctgaagaagatcagcaaccaggccgagtttatcgcctccttctacaacaacgatctgatcaaga
tcaacggcgagctgtatagagtgatcggcgtgaacaacgacctgctgaaccggatcgaagtgaacatgatcgacatcacctaccgcgagtacctggaa
aacatgaacgacaagaggccccccaggatcattaagacaatcgcctccaagacccagagcattaagaagtacagcacagacatt**gtg**ggcaacctgta
tgaagtgaaatctaagaagcaccctcagatcatcaaaaagggc

SaCas9.Redi.2 protein sequence

AKRNYILG**S**DIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQR
VKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNE
LSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQ
SFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN
DLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLK
VYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLS
LKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAII
KKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQE
GKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSK

ISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYF
RVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKV
MENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRK
DDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYK
YYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDN
GVYKFVTVKNLDV**K**KKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVN
NDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDI**V**GNLYEVKSKKHPQIIKK
G*

SaCas9.Redi.3 -DNA sequence

gccaagcggaactacatcctgggcctggacatcggcatcaccagc**gcc**ggctacggcatcatcgactacgagacacgggacgtgatcgatgccggcg
tgcggctgttcaaagaggccaacgtggaaaacaacgagggcaggcggagcaagagaggcgccagaaggctgaagcggcggaggcggcatagaat
ccagagagtgaagaagctgctgttcgactacaacctgctgaccgaccacagcgagctgagcggcatcaacccctacgaggccagagtgaagggcctg
agccagaagctgagcgaggaagagttctctgccgccctgctgcacctggccaagagaagaggcgtgcacaacgtgaacgaggtggaagaggacac
cggcaacgagctgtccaccaaagagcagatcagccggaacagcaaggccctggaagagaaatacgtggccgaactgcagctggaacggctgaaga
aagacggcgaagtgcggggcagcatcaacagattcaagaccagcgactacgtgaaagaagccaaacagctgctgaaggtgcagaaggcctaccacc
agctggaccagagcttcatcgacacctacatcgacctgctggaaacccggcggacctactatgagggacctggcgagggcagccccttcggctggaa
ggacatcaaagaatggtacgagatgctgatgggccactgcacctacttccccgaggaactgcggagcgtgaagtacgcctacaacgccgacctgtaca
acgccctgaacgacctgaacaatctcgtgatcaccagggacgagaacgagaagctggaatattacgagaagttccagatcatcgagaacgtgttcaagc
agaagaagaagcccaccctgaagcagatcgccaaagaaatcctcgtgaacgaagaggatattaagggctacagagtgaccagcaccggcaagcccg
agttcaccaacctgaaggtgtaccacgacatcaaggacattaccgcccggaaagagattattgagaacgccgagctgctggatcagattgccaagatcct
gaccatctaccagagcagcgaggacatccaggaagaactgaccaatctgaactccgagctgacccaggaagagatcgagcagatctctaatctgaagg
gctataccggcacccacaacctgagcctgaaggccatcaacctgatcctggacgagctgtggcacaccaacgacaaccagatcgctatcttcaaccgg
ctgaagctggtgcccaagaaggtggacctgtcccagcagaaagagatccccaccaccctggtggacgacttcatcctgagccccgtcgtgaagagaag
cttcatccagagcatcaaagtgatcaacgccatcatcaagaagtacggcctgcccaacgacatcattatcgagctggcccgcgagaagaactccaagga
cgcccagaaaatgatcaacgagatgcagaagcggaaccggcagaccaacgagcggatcgaggaaatcatccggaccaccggcaaagagaacgcc
aagtacctgatcgagaagatcaagctgcacgacatgcaggaaggcaagtgcctgtacagcctggaagccatccctctggaagatctgctgaacaaccc
cttcaactatgaggtggaccacatcatccccagaagcgtgtccttcgacaacagcttcaacaacaaggtgctcgtgaagcaggaagaaaacagcaagaa
gggcaaccggaccccattccagtacctgagcagcagcgacagcaagatcagctacgaaaccttcaagaagcacatcctgaatctggccaagggcaag
ggcagaatcagcaagaccaagaaagagtatctgctggaagaacgggacatcaacaggttctccgtgcagaaagacttcatcaaccggaacctggtgga
taccagatacgccaccagaggcctgatgaacctgctgcggagctacttcagagtgaacaacctggacgtgaaagtgaagtccatcaatggcggcttcac
cagctttctgcggcggaagtggaagtttaagaaagagcggaacaaggggtacaagcaccacgccgaggacgccctgatcattgccaacgccgatttca
tcttcaaagagtggaagaaactggacaaggccaaaaaagtgatggaaaaccagatgttcgaggaaaagcaggccgagagcatgcccgagatcgaaa
ccgagcaggagtacaaagagatcttcatcacccccaccagatcaagcacattaaggacttcaaggactacaagtacagccaccgggtggacaagaag
cctaatagagagctgattaacgacaccctgtactccacccggaaggacgacaagggcaacaccctgatcgtgaacaatctgaacggcctgtacgacaa
ggacaatgacaagctgaaaaagctgatcaacaagagccccgaaaagctgctgatgtaccaccacgacccccagacctaccagaaactgaagctgatta
tggaacagtacggcgacgagaagaatcccctgtacaagtactacgaggaaaccgggaactacctgaccaagtactccaaaaaggacaacggccccgt
gatcaagaagattaagtattacggcaacaaactgaacgcccatctggacatcaccgacgactaccccaacagcagaaacaaggtcgtgaagctgtccct
gaagccctacagattcgacgtgtacctggacaatggcgtgtacaagttcgtgaccgtgaagaatctggatgtg**aag**aaaaaagaaaactactacgaagt
gaatagcaagtgctatgaggaagctaagaagctgaagaagatcagcaaccaggccgagtttatcgcctccttctacaacaacgatctgatcaagatcaac
ggcgagctgtatagagtgatcggcgtgaacaacgacctgctgaaccggatcgaagtgaacatgatcgacatcacctaccgcgagtacctggaaaacat

gaacgacaagaggccccccaggatcattaagacaatcgcctccaagacccagagcattaagaagtacagcacagacattgtggcaacctgtatgaag
tgaaatctaagaagcaccctcagatcatcaaaaagggc

SaCas9.Redi.3 protein sequence

AKRNYILGLDIGITSAGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQR
VKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNE
LSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQ
SFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN
DLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLK
VYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLS
LKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAII
KKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQE
GKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSK
ISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYF
RVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKV
MENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRK
DDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYK
YYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDN
GVYKFVTVKNLDVKKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVN
NDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDIVGNLYEVKSKKHPQIIKK
G*

AsCas12a.Redi.1 DNA sequence

Atgacacagttcgagggctttaccaacctgtatcaggtgagcaagacactgcggtttgagctgatcccacagggcaagaccctgaagcacatccagga
gcagggcttcatcgaggaggacaaggcccgcaatgatcactacaaggagctgaagcccatcatcgatcggatctacaagacctatgccgaccagtgcc
tgcagctggtgcagctggattgggagaacctgagcgccgccatcgactcctatagaaaggagaaaaccgaggagacaaggaacgccctgatcgagg
agcaggccacatatcgcaatgccatccacgactacttcatcggccggacagacaacctgaccgatgccatcaataagagacacgccgagatctacaag
ggcctgttcaaggccgagctgtttaatggcaaggtgctgaagcagctgggcaccgtgaccacaaccgagcacgagaacgccctgctgcggagcttcg
acaagtttacaacctacttctccggcttttatgagaacaggaagaacgtgttcagcgccgaggatatcagcacagccatcccacaccgcatcgtgcagga
caacttccccaagtttaaggagaattgtcacatcttcacacgcctgatcaccgccgtgcccagctccgggagcactttgagaacgtgaagaaggccatc
ggcatcttcgtgagcacctccatcgaggaggtgttttccttccctttttataaccagctgctgacacagacccagatcgacctgtataaccagctgctgggag
gaatctctcgggaggcaggcaccgagaagatcaagggcctgaacgaggtgctgaatctggcctccagaagaatgatgagacagcccacatcatcgc
ctccctgccacacagattcatccccctgtttaagcagatcctgtccgataggaacaccctgtctttcatcctggaggagtttaagagcgacgaggaagtgat
ccagtccttctgcaagtacaagacactgctgagaaacgagaacgtgctggagacagccgaggccctgtttaacgagctgaacagcatcgacctgacac
acatcttcatcagccacaagaagctggagacaatcagcagcgccctgtgcgaccactgggatacactgaggaatgccctgtatgagcggagaatctccg
agctgacaggcaagatcaccaagtctgccaaggagaaggtgcagcgcagcctgaagcacgaggatatcaacctgcaggagatcatctctgccgcagg
caaggagctgagcgaggccttcaagcagaaaaccagcgagatcctgtcccacgcacacgccgccctggatcagccactgcctacaaccctgaagaag
caggaggagaaggagatcctgaagtctcagctggacagcctgctgggcctgtaccacctgctggactggtttgccgtggatgagtccaacgaggtgga
ccccgagttctctgcccggctgaccggcatcaagctggagatggagccttctctgagcttctacaacaaggccagaaattatgccaccaagaagccctac
tccgtggagaagttcaagctgaactttcagatgcctacactggcctctggctgggacgtgaataaggagaagaacaatggcgccatcctgtttgtgaaga
acggcctgtactatctgggcatcatgccaaagcagaagggcaggtataaggccctgagcttcgagcccacagagaaaaccagcgagggctttgataag
atgtactatgactacttccctgatgccgccaagatgatcccaaagtgcagcacccagctgaaggccgtgacagcccactttcagacccacacaaccccca

tcctgctgtccaacaatttcatcgagcctctggagatcacaaaggagatctacgacctgaacaatcctgagaaggagccaaagaagtttcagacagccta
cgccaagaaaaccggcgaccagaagggctacagagaggccctgtgcaagtggatcgacttcacaagggattttctgtccaagtataccaagacaacct
ctatcgatctgtctagcctgcggccatcctctcagtataaggacctgggcgagtactatgccgagctgaatcccctgctgtaccacatcagcttccagagaa
tcgccgagaaggagatcatggatgccgtggagacaggcaagctgtacctgttccagatctataacaaggactttgccaagggccaccacggcaagcct
aatctgcacacactgtattggaccggcctgttttctccagagaacctggccaagacaagcatcaagctgaatggccaggccgagctgttctaccgcccta
agtccaggatgaagaggatggcacaccggctgggagagaagatgctgaacaagaagctgaaggatcagaaaaccccaatccccgacaccctgtacc
aggagctgtacgactatgtgaatcacagactgtcccacgacctgtctgatgaggccagggccctgctgcccaacgtgatcaccaaggaggtgtctcacg
agatcatcaaggataggcgctttaccagcgacaagttctttttccacgtgcctatcacactgaactatcaggccgccaattccccatctaagttcaaccagag
ggtgaatgcctacctgaaggagcaccccgagacacctatcatcggcatcgatcggggcgagagaaacctgatctatatcacagtgatcgactccaccgg
caagatcctggagcagcggagcctgaacaccatccagcagtttgattaccagaagaagctggacaacagggagaaggagagggtggcagcaaggca
ggcctggtctgtggtgggcacaatcaaggatctgaagcagggctat**gcc**agccaggtcatccacgagatcgtggacctgatgatccactaccaggccgt
ggtggtgctggagaacctgaatttcggctttaagagcaagaggaccggcatcgccgagaaggccgtgtaccagcagttcgagaagatgctgatcgata
agctgaattgcctggtgctgaaggactatccagcagagaaagtgggaggcgtgctgaacccataccagctgacagaccagttcacctcctttgccaagat
gggcacccagtctggcttcctgttttacgtgcctgccccatatacatctaagatcgatcccctgaccggcttcgtggacccccttcgtgtggaaaaccatcaa
gaatcacgagagccgcaagcacttcctggagggcttcgactttctgcactacgacgtgaaaaccggcgacttcatcctgcactttaagatgaacagaaat
ctgtccttccagaggggcctgcccggctttatgcctgcatgggatatcgtgttcgagaagaacgagacacagtttgacgccaagggcacccctttcatcgc
cggcaagagaatcgtgccagtgatcgagaatcacagattcaccggcagataccgggacctgtatcctgccaacgagctgatcgccctgctggaggaga
agggcatcgtgttcagggatggctccaacatcctgccaaagctgctggagaatgacgattctcacgccatcgacaccatggtggccctgatccgcagcg
tgctgcagatgcggaactccaatgccgccacaggcgaggactatatcaacagccccgtgcgcgatctgaatggcgtgtgcttcgactcccggtttcaga
acccagagtggcccatggacgccgatgccaatggcgcctaccacatcgccctgaagggccagctgctgctgaatcacctgaaggagagcaaggatct
gaagctgcagaacggcatctccaatcaggactggctggcctacatccaggagctgcgcaac

AsCas12a.Redi.1 protein sequence

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCL
QLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGL
FKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNF
PKFKENCHIFTRLITAVPS**S**REHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISR
EAGTEKIKGLNEVLNLA**S**QKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKY
KTLLRNENVLETAEALFNELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKS
AKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQL
DSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQ
MPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFP
DAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQ
KGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMD
AVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKR
MAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDR
RFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRS
LNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGY**A**SQVIHEIVDLMIHYQAVVVLENL
NFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQS
GFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSF
QRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIV

FRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPE
WPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRN*

AsCas12a.Redi.2 DNA sequence

atgacacagttcgagggctttaccaacctgtatcaggtgagcaagacactgcggtttgagctgatcccacagggcaagaccctgaagcacatccaggag
cagggcttcatcgaggaggacaaggcccgcaatgatcactacaaggagctgaagcccatcatcgatcggatctacaagacctatgccgaccagtgcct
gcagctggtgcagctggattgggagaacctgagcgccgccatcgactcctatagaaaggagaaaaccgaggagacaaggaacgccctgatcgagga
gcaggccacatatcgcaatgccatccacgactacttcatcggccggacagacaacctgaccgatgccatcaataagagacacgccgagatctacaagg
gcctgttcaaggccgagctgtttaatggcaaggtgctgaagcagctgggcaccgtgaccacaaccgagcacgagaacgccctgctgcggagcttcgac
aagtttacaacctacttctccggcttttatgagaacaggaagaacgtgttcagcgccgaggatatcagcacagccatcccacaccgcatcgtgcaggaca
acttccccaagtttaaggagaattgtcacatcttcacacgcctgatcaccgccgtgcccagc**tcc**cgggagcactttgagaacgtgaagaaggccatcgg
catcttcgtgagcacctccatcgaggaggtgttttccttccctttttataaccagctgctgacacagacccagatcgacctgtataaccagctgctgggagga
atctctcgggaggcaggcaccgagaagatcaagggcctgaacgaggtgctgaatctggcc**acc**cagaagaatgatgagacagcccacatcatcgcct
ccctgccacacagattcatccccctgtttaagcagatcctgtccgataggaacaccctgtctttcatcctggaggagtttaagagcgacgaggaagtgatcc
agtccttctgcaagtacaagacactgctgagaaacgagaacgtgctggagacagccgaggccctgtttaacgagctgaacagcatcgacctgacacac
atcttcatcagccacaagaagctggagacaatcagcagcgccctgtgcgaccactgggatacactgaggaatgccctgtatgagcggagaatctccga
gctgacaggcaagatcaccaagtctgccaaggagaaggtgcagcgcagcctgaagcacgaggatatcaacctgcaggagatcatctctgccgcaggc
aaggagctgagcgaggccttcaagcagaaaaccagcgagatcctgtcccacgcacacgccgccctggatcagccactgcctacaaccctgaagaagc
aggaggagaaggagatcctgaagtctcagctggacagcctgctgggcctgtaccacctgctggactggtttgccgtggatgagtccaacgaggtggac
cccgagttctctgcccggctgaccggcatcaagctggagatggagccttctctgagcttctacaacaaggccagaaattatgccaccaagaagccctact
ccgtggagaagttcaagctgaactttcagatgcctacactggcctctggctgggacgtgaataaggagaagaacaatggcgccatcctgtttgtgaagaa
cggcctgtactatctgggcatcatgccaaagcagaagggcaggtataaggccctgagcttcgagcccacagagaaaaccagcgagggctttgataaga
tgtactatgactacttccctgatgccgccaagatgatcccaaagtgcagcacccagctgaaggccgtgacagcccactttcagacccacacaaccccat
cctgctgtccaacaatttcatcgagcctctggagatcacaaaggagatctacgacctgaacaatcctgagaaggagccaaagaagtttcagacagcctac
gccaagaaaaccggcgaccagaagggctacagagaggccctgtgcaagtggatcgacttcacaagggattttctgtccaagtataccaagacaacctct
atcgatctgtctagcctgcggccatcctctcagtataaggacctgggcgagtactatgccgagctgaatcccctgctgtaccacatcagcttccagagaatc
gccgagaaggagatcatggatgccgtggagacaggcaagctgtacctgttccagatctataacaaggactttgccaagggccaccacggcaagcctaa
tctgcacacactgtattggaccggcctgttttctccagagaacctggccaagacaagcatcaagctgaatggccaggccgagctgttctaccgccctaagt
ccaggatgaagaggatggcacaccggctgggagagaagatgctgaacaagaagctgaaggatcagaaaaccccaatccccgacaccctgtaccagg
agctgtacgactatgtgaatcacagactgtcccacgacctgtctgatgaggccagggccctgctgcccaacgtgatcaccaaggaggtgtctcacgagat
catcaaggataggcgctttaccagcgacaagttcttttttccacgtgcctatcacactgaactatcaggccgccaattccccatctaagttcaaccagagggt
gaatgcctacctgaaggagcaccccgagacacctatcatcggcatcgatcggggcgagagaaacctgatctatatcacagtgatcgactccaccggca
agatcctggagcagcggagcctgaacaccatccagcagtttgattaccagaagaagctggacaacagggagaaggagagggtggcagcaaggcag
gcctggtctgtggtgggcacaatcaaggatctgaagcagggctat**gcc**agccaggtcatccacgagatcgtggacctgatgatccactaccaggccgtg
gtggtgctggagaacctgaatttcggctttaagagcaagaggaccggcatcgccgagaaggccgtgtaccagcagttcgagaagatgctgatcgataa
gctgaattgcctggtgctgaaggactatccagcagagaaagtgggaggcgtgctgaacccataccagctgacagaccagttcacctcctttgccaagat
gggcacccagtctggcttcctgttttacgtgcctgccccatatacatctaagatcgatcccctgaccggcttcgtggacccccttcgtgtggaaaaccatcaa
gaatcacgagagccgcaagcacttcctggagggcttcgactttctgcactacgacgtgaaaaccggcgacttcatcctgcactttaagatgaacagaaat
ctgtccttccagaggggcctgcccggctttatgcctgcatgggatatcgtgttcgagaagaacgagacacagtttgacgccaagggcacccctttcatcgc
cggcaagagaatcgtgccagtgatcgagaatcacagattcaccggcagataccgggacctgtatcctgccaacgagctgatcgccctgctggaggaga
agggcatcgtgttcagggatggctccaacatcctgccaaagctgctggagaatgacgattctcacgccatcgacaccatggtggccctgatccgcagcg
tgctgcagatgcggaactccaatgccgccacaggcgaggactatatcaacagccccgtgcgcgatctgaatggcgtgtgcttcgactcccggtttcaga

acccagagtggcccatggacgccgatgccaatggcgcctaccacatcgccctgaagggccagctgctgctgaatcacctgaaggagagcaaggatct
gaagctgcagaacggcatctccaatcaggactggctggcctacatccaggagctgcgcaac

AsCas12a.Redi.2 protein sequence

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCL
QLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGL
FKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNF
PKFKENCHIFTRLITAVPS**S**REHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISR
EAGTEKIKGLNEVLNLA**T**QKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCK
YKTLLRNENVLETAEALFNELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITK
SAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQ
LDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQ
MPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFP
DAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQ
KGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMD
AVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKR
MAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDR
RFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRS
LNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGY**A**SQVIHEIVDLMIHYQAVVVLENL
NFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQS
GFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSF
QRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIV
FRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPE
WPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRN*

AsCas12a.Redi.3 DNA sequence

atgacacagttcgagggctttaccaacctgtatcaggtgagcaagacactgcggtttgagctgatcccacagggcaagaccctgaagcacatccaggag
cagggcttcatcgaggaggacaaggcccgcaatgatcactacaaggagctgaagcccatcatcgatcggatctacaagacctatgccgaccagtgcct
gcagctggtgcagctggattgggagaacctgagcgccgccatcgactcctatagaaaggagaaaaccgaggagacaaggaacgccctgatcgagga
gcaggccacatatcgcaatgccatccacgactacttcatcggccggacagacaacctgaccgatgccatcaataagagacacgccgagatctacaagg
gcctgttcaaggccgagctgtttaatggcaaggtgctgaagcagctgggcaccgtgaccacaaccgagcacgagaacgccctgctgcggagcttcgac
aagtttacaacctacttctccggcttttatgagaacaggaagaacgtgttcagcgccgaggatatcagcacagccatcccacaccgcatcgtgcaggaca
acttccccaagtttaaggagaattgtcacatcttcacacgcctgatcaccgccgtgcccagc**acc**cgggagcactttgagaacgtgaagaaggccatcg
gcatcttcgtgagcacctccatcgaggaggtgttttccttccctttttataaccagctgctgacacagacccagatcgacctgtataaccagctgctgggagg
aatctctcgggaggcaggcaccgagaagatcaagggcctgaacgaggtgctgaatctggcc**gcc**cagaagaatgatgagacagcccacatcatcgcc
tccctgccacacagattcatccccctgtttaagcagatcctgtccgataggaacaccctgtctttcatcctggaggagtttaagagcgacgaggaagtgatc
cagtccttctgcaagtacaagacactgctgagaaacgagaacgtgctggagacagccgaggccctgtttaacgagctgaacagcatcgacctgacaca
catcttcatcagccacaagaagctggagacaatcagcagcgccctgtgcgaccactgggatacactgaggaatgccctgtatgagcggagaatctccga
gctgacaggcaagatcaccaagtctgccaaggagaaggtgcagcgcagcctgaagcacgaggatatcaacctgcaggagatcatctctgccgcaggc
aaggagctgagcgaggccttcaagcagaaaaccagcgagatcctgtcccacgcacacgccgccctggatcagccactgcctacaaccctgaagaagc
aggaggagaaggagatcctgaagtctcagctggacagcctgctgggcctgtaccacctgctggactggtttgccgtggatgagtccaacgaggtggac
cccgagttctctgcccggctgaccggcatcaagctggagatggagccttctctgagcttctacaacaaggccagaaattatgccaccaagaagccctact

ccgtggagaagttcaagctgaactttcagatgcctacactggcctctggctgggacgtgaataaggagaagaacaatggcgccatcctgtttgtgaagaa
cggcctgtactatctgggcatcatgccaaagcagaagggcaggtataaggccctgagcttcgagcccacagagaaaaccagcgagggctttgataaaga
tgtactatgactacttccctgatgccgccaagatgatcccaaagtgcagcacccagctgaaggccgtgacagcccactttcagacccacacaacccccat
cctgctgtccaacaatttcatcgagcctctggagatcacaaaggagatctacgacctgaacaatcctgagaaggagccaaagaagtttcagacagcctac
gccaagaaaaccggcgaccagaagggctacagagaggccctgtgcaagtggatcgacttcacaagggattttctgtccaagtataccaagacaacctct
atcgatctgtctagcctgcggccatcctctcagtataaggacctgggcgagtactatgccgagctgaatcccctgctgtaccacatcagcttccagagaatc
gccgagaaggagatcatggatgccgtggagacaggcaagctgtacctgttccagatctataacaaggactttgccaagggccaccacggcaagcctaa
tctgcacacactgtattggaccggcctgttttctccagagaacctggccaagacaagcatcaagctgaatggccaggccgagctgttctaccgccctaagt
ccaggatgaagaggatggcacaccggctgggagagaagatgctgaacaagaagctgaaggatcagaaaaccccaatccccgacaccctgtaccagg
agctgtacgactatgtgaatcacagactgtcccacgacctgtctgatgaggccagggccctgctgcccaacgtgatcaccaaggaggtgtctcacgagat
catcaaggataggcgctttaccagcgacaagttcttttccacgtgcctatcacactgaactatcaggccgccaattccccatctaagttcaaccagagggt
gaatgcctacctgaaggagcaccccgagacacctatcatcggcatcgatcggggcgagagaaacctgatctatatcacagtgatcgactccaccggca
agatcctggagcagcggagcctgaacaccatccagcagtttgattaccagaagaagctggacaacagggagaaggagagggtggcagcaaggcag
gcctggtctgtggtgggcacaatcaaggatctgaagcagggctat**gcc**agccaggtcatccacgagatcgtggacctgatgatccactaccaggccgtg
gtggtgctggagaacctgaatttcggctttaagagcaagaggaccggcatcgccgagaaggccgtgtaccagcagttcgagaagatgctgatcgataa
gctgaattgcctggtgctgaaggactatccagcagagaaagtgggaggcgtgctgaacccataccagctgacagaccagttcacctcctttgccaagat
gggcacccagtctggcttcctgttttacgtgcctgccccatatacatctaagatcgatcccctgaccggcttcgtggaccccttcgtgtggaaaaccatcaa
gaatcacgagagccgcaagcacttcctggagggcttcgactttctgcactacgacgtgaaaaccggcgacttcatcctgcactttaagatgaacagaaat
ctgtccttccagaggggcctgcccggctttatgcctgcatgggatatcgtgttcgagaagaacgagacacagtttgacgccaagggcacccctttcatcgc
cggcaagagaatcgtgccagtgatcgagaatcacagattcaccggcagataccgggacctgtatcctgccaacgagctgatcgccctgctggaggaga
agggcatcgtgttcagggatggctccaacatcctgccaaagctgctggagaatgacgattctcacgccatcgacaccatggtggccctgatccgcagcg
tgctgcagatgcggaactccaatgccgccacaggcgaggactatatcaacagccccgtgcgcgatctgaatggcgtgtgcttcgactcccggtttcaga
acccagagtggcccatggacgccgatgccaatggcgcctaccacatcgccctgaagggccagctgctgctgaatcacctgaaggagagcaaggatct
gaagctgcagaacggcatctccaatcaggactggctggcctacatccaggagctgcgcaac

AsCas12a.Redi.3 protein sequence

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCL
QLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGL
FKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNF
PKFKENCHIFTRLITAVPS**T**REHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISR
EAGTEKIKGLNEVLNLA**A**QKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCK
YKTLLRNENVLETAEALFNELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITK
SAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQ
LDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQ
MPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFP
DAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQ
KGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMD
AVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKR
MAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDR
RFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRS
LNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGY**A**SQVIHEIVDLMIHYQAVVVLENL
NFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQS

GFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSF
QRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIV
FRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPE
WPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRN*

2.10 References

[1] Hakim CH, Kumar SRP, Pérez-López DO, Wasala NB, Zhang D, Yue Y, et al. Cas9-specific immune responses compromise local and systemic AAV CRISPR therapy in multiple dystrophic canine models. *Nat Commun*. 2021;12(1):6769.

[2] Wagner DL, Peter L, Schmueck-Henneresse M. Cas9-directed immune tolerance in humans-a model to evaluate regulatory T cells in gene therapy? *Gene Ther*. 2021;28(9):549-559.

[3] Nelson CE, Hakim CH, Ousterout DG, Thakore PI, Moreb EA, Castellanos Rivera RM, et al. In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science*. 2016;351(6271):403-407.

[4] Tabebordbar M, Zhu K, Cheng JKW, Chew WL, Widrick JJ, Yan WX, et al. In vivo gene editing in dystrophic mouse muscle and muscle stem cells. *Science*. 2016;351(6271):407-411.

[5] Suzuki K, Tsunekawa Y, Hernandez-Benitez R, Wu J, Zhu J, Kim EJ, et al. In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature*. 2016;540(7631):144-149.

[6] Long C, McAnally JR, Shelton JM, Mireault AA, Bassel-Duby R, Olson EN. Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA. *Science*. 2014;345(6201):1184-1188.

[7] Wang JY, Doudna JA. CRISPR technology: A decade of genome editing is only the beginning. *Science*. 2023;379(6629):eadd8643.

[8] Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 2013;339(6121):819-823.

[9] Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, et al. In vivo genome editing using Staphylococcus aureus Cas9. *Nature*. 2015;520(7546):186-191.

[10] Chew WL. Immunity to CRISPR Cas9 and Cas12a therapeutics. *Wiley Interdiscip Rev Syst Biol Med*. 2018;10(1). doi:10.1002/wsbm.1408

[11] Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*. 2013;31(9):827-832.

[12] Wang D, Tai PWL, Gao G. Adeno-associated virus vector as a platform for gene therapy delivery. *Nat Rev Drug Discov*. 2019;18(5):358-378.

[13] Colque-Navarro Patricia, Jacobsson Gunnar, Andersson Rune, Flock Jan-Ingmar, Möllby Roland. Levels of Antibody against 11 Staphylococcus aureus Antigens in a Healthy Population. *Clin Vaccine*

Immunol. 2010;17(7):1117-1123.

[14] Kolata JB, Kühbandner I, Link C, Normann N, Vu CH, Steil L, et al. The Fall of a Dogma? Unexpected High T-Cell Memory Response to Staphylococcus aureus in Humans. *J Infect Dis*. 2015;212(5):830-838.

[15] Charlesworth CT, Deshpande PS, Dever DP, Camarena J, Lemgart VT, Cromer MK, et al. Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nat Med*. 2019;25(2):249-254.

[16] Chew WL, Tabebordbar M, Cheng JKW, Mali P, Wu EY, Ng AHM, et al. A multifunctional AAV-CRISPR-Cas9 and its host response. *Nat Methods*. 2016;13(10):868-874.

[17] Simhadri VL, Hopkins L, McGill JR, Duke BR, Mukherjee S, Zhang K, et al. Cas9-derived peptides presented by MHC Class II that elicit proliferation of CD4+ T-cells. *Nat Commun*. 2021;12(1):5090.

[18] Crudele JM, Chamberlain JS. Cas9 immunity creates challenges for CRISPR gene editing therapies. *Nat Commun*. 2018;9(1):3497.

[19] Mehta A, Merkel OM. Immunogenicity of Cas9 Protein. *J Pharm Sci*. 2020;109(1):62-67.

[20] Mortensen R, Nissen TN, Blauenfeldt T, Christensen JP, Andersen P, Dietrich J. Adaptive Immunity against Streptococcus pyogenes in Adults Involves Increased IFN-γ and IgG3 Responses Compared with Children. *J Immunol*. 2015;195(4):1657-1664.

[21] Adams S, Robbins FM, Chen D, Wagage D, Holbeck SL, Morse HC 3rd, et al. HLA class I and II genotype of the NCI-60 cell lines. *J Transl Med*. 2005;3(1):11.

[22] Rasmussen M, Harndahl M, Stryhn A, Boucherma R, Nielsen LL, Lemonnier FA, et al. Uncovering the peptide-binding specificities of HLA-C: a general strategy to determine the specificity of any MHC class I molecule. *J Immunol*. 2014;193(10):4790-4802.

[23] Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res*. 2020;48(W1):W449-W454.

[24] Schmid-Burgk JL, Gao L, Li D, Gardner Z, Strecker J, Lash B, et al. Highly Parallel Profiling of Cas9 Variant Specificity. *Mol Cell*. 2020;78(4):794-800.e8.

[25] Pajot A, Michel ML, Fazilleau N, Pancré V, Auriault C, Ojcius DM, et al. A mouse model of human adaptive immune functions: HLA-A2.1-/HLA-DR1-transgenic H-2 class I-/class II-knockout mice. *Eur J Immunol*. 2004;34(11):3060-3069.

[26] Ott LW, Resing KA, Sizemore AW, Heyen JW, Cocklin RR, Pedrick NM, et al. Tumor Necrosis Factor-alpha- and interleukin-1-induced cellular responses: coupling proteomic and genomic information. *J Proteome Res*. 2007;6(6):2176-2185.

[27] Katsura H, Kobayashi Y, Tata PR, Hogan BLM. IL-1 and TNFα Contribute to the Inflammatory Niche to Enhance Alveolar Regeneration. *Stem Cell Reports*. 2019;12(4):657-666.

[28] Morales RA, Allende ML. Peripheral Macrophages Promote Tissue Regeneration in Zebrafish by

Fine-Tuning the Inflammatory Response. *Front Immunol*. 2019;10:253.

[29] Wagner DL, Amini L, Wendering DJ, Burkhardt LM, Akyüz L, Reinke P, et al. High prevalence of Streptococcus pyogenes Cas9-reactive T cells within the adult human population. *Nat Med*. 2019;25(2):242-248.

[30] Li A, Tanner MR, Lee CM, Hurley AE, De Giorgi M, Jarrett KE, et al. AAV-CRISPR Gene Editing Is Negated by Pre-existing Immunity to Cas9. *Mol Ther*. 2020;28(6):1432-1441.

[31] Verdera HC, Kuranda K, Mingozzi F. AAV Vector Immunogenicity in Humans: A Long Journey to Successful Gene Transfer. *Mol Ther*. 2020;28(3):723-746.

[32] Huysmans H, Zhong Z, De Temmerman J, Mui BL, Tam YK, Mc Cafferty S, et al. Expression Kinetics and Innate Immune Response after Electroporation and LNP-Mediated Delivery of a Self-Amplifying mRNA in the Skin. *Mol Ther Nucleic Acids*. 2019;17:867-878.

[33] Ju Y, Carreño JM, Simon V, Dawson K, Krammer F, Kent SJ. Impact of anti-PEG antibodies induced by SARS-CoV-2 mRNA vaccines. *Nat Rev Immunol*. 2023;23(3):135-136.

[34] Roth TL, Li PJ, Blaeschke F, Nies JF, Apathy R, Mowery C, et al. Pooled Knockin Targeting for Genome Engineering of Cellular Immunotherapies. *Cell*. 2020;181(3):728-744.e21.

[35] King CA, Bradley P. Structure-based prediction of protein-peptide specificity in Rosetta. *Proteins*. 2010;78(16):3437-3449.

[36] Leman JK, Weitzner BD, Lewis SM, Adolf-Bryfogle J, Alam N, Alford RF, et al. Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat Methods*. 2020;17(7):665-680.

[37] Alford RF, Leaver-Fay A, Jeliazkov JR, O'Meara MJ, DiMaio FP, Park H, et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J Chem Theory Comput*. 2017;13(6):3031-3048.

[38] King C, Garza EN, Mazor R, Linehan JL, Pastan I, Pepper M, et al. Removing T-cell epitopes with computational protein design. *Proc Natl Acad Sci U S A*. 2014;111(23):8577-8582.

[39] Yachnin BJ, Mulligan VK, Khare SD, Bailey-Kellogg C. MHCEpitopeEnergy, a Flexible Rosetta-Based Biotherapeutic Deimmunization Platform. *J Chem Inf Model*. 2021;61(5):2368-2382.

[40] Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res*. 2019;47(D1):D339-D343.

[41] Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J Immunol*. 2017;199(9):3360-3368.

[42] Clement K, Rees H, Canver MC, Gehrke JM, Farouni R, Hsu JY, et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol*. 2019;37(3):224-226.

[43] Grieger JC, Choi VW, Samulski RJ. Production and characterization of adeno-associated viral vectors. *Nat Protoc*. 2006;1(3):1412-1428.

[44] Schmid-Burgk JL, Hornung V. BrowserGenome.org: web-based RNA-seq data analysis and visualization. *Nat Methods*. 2015;12(11):1001.

[45] Kilian M, Friedrich M, Sanghvi K, Green E, Pusch S, Kawauchi D, et al. T-cell Receptor Therapy Targeting Mutant Capicua Transcriptional Repressor in Experimental Gliomas. *Clin Cancer Res*. 2022;28(2):378-389.

3. Human Paraneoplastic antigen Ma2 (PNMA2) forms icosahedral capsids that can be engineered for mRNA delivery

3.1 Abstract

A number of endogenous genes in the human genome encode retroviral *gag*-like proteins, which were domesticated from ancient retroelements. The Paraneoplastic Ma antigen (PNMA) family members encode a *gag*-like capsid domain, but their ability to assemble as capsids and traffic between cells remains mostly uncharacterized. Here, we systematically investigate human PNMA proteins and find that a number of PNMAs are secreted by human cells. We determine that PNMA2 forms icosahedral capsids efficiently but does not naturally encapsidate nucleic acids. We resolve the cryo-electron microscopy (cryo-EM) structure of PNMA2 and leverage the structure to design engineered PNMA2 (ePNMA2) particles with RNA packaging abilities. Recombinantly purified ePNMA2 proteins package mRNA molecules into icosahedral capsids and can function as delivery vehicles in mammalian cell lines, demonstrating the potential for engineered endogenous capsids as a nucleic acid therapy delivery modality.

3.2 Significance statement

mRNA-based therapeutics have the potential to treat a wide range of diseases, but the treatments are limited by the repertoire of available mRNA delivery methods. In this work, we developed a new mRNA delivery modality based on the human protein PNMA2. We show that PNMA2 naturally forms capsids in human cells. To engineer PNMA2 for delivery, we established a method to produce a variant of PNMA2 (ePNMA2) that can encapsidate mRNA *in vitro* and deliver the cargo to recipient cells, forming the basis for an all-protein delivery vehicle that can be assembled and loaded *in vitro*. ePNMA2 expands the delivery toolbox for future mRNA-based therapeutics.

3.3 Background

RNA-based therapeutics, including mRNA-based vaccines, have the potential to be deployed in a wide range of disease contexts. To achieve this potential, we need a suite of delivery vehicles that can efficiently package and safely deliver therapeutic RNA cargoes to specific tissues. Several delivery modalities have already been developed, including non-viral approaches such as lipid nanoparticles (LNPs), which have successfully been used to deliver oligonucleotide and mRNA therapeutics, and viral vectors such as adeno-associated virus (AAV) [102]. However, the broad applicability of these approaches is limited due to a combination of factors including cargo

size constraints, immunogenicity, difficulty in achieving tissue-specific targeting, and scalable production.

Natural delivery systems from the human genome might provide the basis for new engineered gene transfer modalities that can address some of these limitations. Recent work has uncovered a diverse array of endogenous *gag*-like genes within the human genome, which resemble retroviral structural proteins and therefore could potentially be engineered for gene transfer (2). A number of these, including *Arc* and *Peg10,* have been domesticated and serve vital roles in normal mammalian physiology [42,65,66]. Previous work has also shown that the ARC and PEG10 *gag*-like proteins have the ability to form capsid structures that can package their cognate mRNAs [42–44,66]. Extending this natural ability, PEG10 was recently engineered to programmably package and deliver an exogenous cargo mRNA into human cells, demonstrating the potential of these endogenous retrotransposon-derived proteins as a new nucleic acid delivery modality [44]. To further explore the potential of endogenous *gag*-like proteins for therapeutic RNA delivery, we sought to systematically characterize the paraneoplastic Ma antigen (PNMA) protein family [45]. The PNMA family, which in humans contains over a dozen proteins, was initially identified due to some PNMAs encoding auto-antigens in patients with paraneoplastic neurological disease [46–49]. Although several PNMA family members have been previously reported to have roles in apoptosis [51,103–106], most remain poorly characterized. Some PNMAs, however, including mouse PNMA2, have been shown to form capsid structures, suggesting they may be suitable for development as delivery vehicles [44,107].

Here we explore the potential for human PNMA proteins to form capsids and package RNA. We found that PNMA2 is robustly secreted as an icosahedral capsid from human cells and can self-assemble *in vitro* from recombinant protein. We used cryo-electron microscopy (cryo-EM) to resolve the structure of the PNMA2 capsid and structure-guided engineering to modify the PNMA2 protein capsid to package mRNA. We show that these engineered PNMA2 capsids can functionally deliver mRNA into recipient cells, demonstrating promise as a therapeutic mRNA delivery vehicle. In addition to PNMA2, we found that other PNMA family members are capable of forming virus-like capsids, suggesting they may also be suitable for delivery and raising the possibility that these proteins are involved in intercellular communication.

3.3 Results

3.3.1 Domain architectures and origins of the PNMA family of domesticated retroelements

The PNMA family likely emerged from the domestication of a *Ty3/mdg4* (also known as *gypsy*) retrotransposon by the loss of the polymerase (POL) region [45,108,109]. Multiple duplications of the ancestral PNMA in Eutherians gave rise to a large family of PNMAs in some mammalian species, including humans. In the human genome, PNMAs are spread across 4 chromosomes, with some clustered together – PNMA8a/b/c and CCDC8 share the same locus on chromosome 19, and PNMA3/5/6a/6e/6f share the same locus in chromosome X (Figure 1A).

We used structural modeling to compare the domain architectures of 14 human PNMA proteins (1, 2, 3, 4, 5, 6a, 6e, 6f, 7a, 7b, 8a, 8b, 8c, and CCDC8), as well as the marsupial PNMA from the tammar wallaby, *Macropus eugenii* (MePNMA), which is a close relative to *Ty3/mdg4* (see Methods) (Figure 1B). Most of the human PNMAs contain the N and C terminal capsid domains, except PNMA8a/b/c and CCDC8, and AlphaFold-based oligomeric prediction suggests that PNMAs 1, 2, 3, 4, and 5 may form multimers (Figure S1). The RRM domain is also highly conserved, absent only in PNMA7a and 7b. The zinc finger domain, which is typically involved in interaction with nucleic acids [59], is found in only some PNMAs. Of the proteins that lack the zinc finger, some alternatively contain a K-R rich domain, which could similarly function to interact with nucleic acids.

3.3.2 PNMA2 is secreted by human cells as a non-enveloped capsid

Given that ARC and PEG10 are secreted from mammalian cells, we tested PNMA secretion by transfecting each human *PNMA* into HEK293FT cells with a C-terminal HA tag (Figures 2A, B). Although all PNMAs expressed robustly, only a subset of PNMAs were secreted into the virus-like particle (VLP) fraction, with PNMA2 and PNMA6ev1 demonstrating especially robust secretion (Figure 2B). We chose to further investigate PNMA2 given its predicted multimeric assembly (Figure S1) and confirmed that PNMA2 is highly secreted across multiple cell lines (Figures 2B, S2). We observed that PNMA2 expression in cells is primarily localized to the cytosol (Figure S3), and PNMA2 is secreted by cells as non-enveloped capsid-like structures (Figure 2C). This is consistent with similar findings in mouse cells [107].

We investigated whether PNMA2 capsids package their own mRNA by comparing the levels of PNMA2 mRNA in cellular and VLP fractions from HEK293FT cells overexpressing either PNMA2 or a start codon deficient version (Figure 2D). We found no significant difference in PNMA2 mRNA levels between the VLP fraction with PNMA2 capsid and the start codon deficient version, indicating that PNMA2 does not package its own mRNA (Figure 2D, bottom). We also sequenced mRNA from the VLP fractions of U20S cells transfected with either a PNMA2 CRISPRa cassette or a non-targeting control [44] and found that, while whole cells demonstrated increased transcript abundance of PNMA2 (and a number of other transcripts (Supplemental Dataset 1)), there were no transcripts with strongly significant increased abundance in the VLP fraction, suggesting that PNMA2 capsids do not specifically package PNMA2 mRNA or any other cellular mRNA (Figure 2E, Supplemental Dataset 1).

Although our results indicate PNMA2 does not package mRNA in cultured cells, we sought to determine if purified PNMA2 could assemble around an mRNA *in vitro.* We first tested whether PNMA2 capsids can self-assemble *in vitro* from recombinantly produced protein. We found that PNMA2 purified from *E. coli* readily assembles into capsid structures similar to those secreted from mammalian cells (compare Figure 2C with Figure 3A). The PNMA2 size-exclusion chromatogram (SEC) that demonstrates elution before the void volume is also consistent with efficient capsid assembly (Figure 3B). We hypothesized that orchestrating PNMA2 disassembly and reassembly around RNA might facilitate RNA packaging (Figure 3C). To test this, we introduced *in vitro* transcribed Cre mRNA into the assembly reaction and then assayed for

nuclease resistance. However, self-assembled PNMA2 did not protect Cre mRNA from nuclease degradation.

3.3.3 PNMA2 forms an icosahedral capsid structure with a negatively charged lumen

To guide our engineering efforts to package RNA within PNMA2, we used cryo-EM to resolve the structure of recombinant human PNMA2 expressed in *E. coli* at 3.1 Å resolution (Figures 4A, S4). Sixty identical copies of the PNMA2 monomer assemble to form a capsid with icosahedral symmetry and a triangulation number T=1 (Figure 4B). The capsid has a mean diameter of 210 Å and encloses a volume of 1,400,000 cubic Ångstrom – approximately 60% of the volume of AAV-2 [110]. Cryo-EM density was only resolvable for residues 158 to 340 of PNMA2, corresponding to the N- and C-terminal capsid domains, which fold into α-helical domains similar to other *gag* proteins. The interfaces at the 2- and 3-fold symmetry axes are composed of the C-terminal capsid domains, while the 5-fold symmetry axis is composed of the N-terminal capsid domains (Figure 4C). The first ordered residue of PNMA2, Leu158, is found at the 5-fold axis on the exterior side of the capsid, suggesting that the diffuse cryo-EM density forming "spikes" on the 5-fold axis (Figure 4A) is attributable to the N-terminal dimerization and RRM-like domains of PNMA2. The interior of the capsid is rich in acidic residues, including the last resolvable residues, which form a poly-Glu tract (333-EEEEEEAS-340). The interior of the capsid is therefore predicted to have a negative charge, likely accounting for the capsid's lack of RNA cargo (Figure 4D). We were unable to resolve the final 24 residues of PNMA2 (ten of which are also acidic), but they likely account for the cloud of cryo-EM density inside the capsid (Figure 4E).

3.3.4 Engineering PNMA2 for functional mRNA delivery

To enhance mRNA packaging efficiency, we used structure-guided engineering to modify the capsid lumen by replacing the C-terminal disordered region with an RNA-binding motif, cowpea chlorotic mottle virus N-terminal 30 residues (CCMV1-30), which is known to efficiently bind single stranded RNA without obvious sequence preference [111,112] (Figure 5A). We purified the resulting PNMA2(340)-CCMV(30) (referred to as engineered PNMA2 (ePNMA2)) from *E. coli* and confirmed capsid formation similar to wild-type PNMA2 (Figure 5A). In contrast to wild-type PNMA2 capsids, ePNMA2 capsids were more stable at low ionic strengths and required 6M urea for disassembly (Figure S5). We tested various conditions for reassembly in the presence of cargo RNA and found that 500 mM NaCl and 10 mM $CaCl_2$ led to the most efficient packaging and protection of cargo RNA from nuclease degradation (Figures 5B, S5C, S6).

We examined whether ePNMA2 capsids (which are non-enveloped) can enter cells via endocytosis. Confocal microscopy of Neuro2A cells 6 hours after treatment showed ePNMA2 capsids at the cell periphery (Figure 5C). Given previous data showing that the amphipathic peptide LAH4 can aid proteins in both cellular entry and endosomal escape [113], we assessed

whether treating ePNMA2 with LAH4 before addition to cells could enhance entry beyond the cell periphery. We found that LAH4 treatment of ePNMA2 before addition to cells increased the cytosolic localization of ePNMA2 (Figure 5C).

Finally, we tested if LAH4-treated ePNMA2 capsids could deliver a Cre mRNA cargo to Neuro2A-*loxP*-GFP reporter cells (Figure 5D). We RNase A treated ePNMA2(Cre) to degrade unpackaged mRNA and prepared equivalent naked Cre mRNA diluted in the same assembly buffer with and without RNase treatment, all of which were combined with LAH4 peptide before being added to Neuro2A-*loxP*-GFP reporter cells (Figures 5E,F, Figure S7, S8B). We also tested naked Cre mRNA with and without RNase A treatment with Lipofectamine3000 transfection (Figure S8A). RNase treatment completely degraded mRNA in the absence of ePNMA2, as confirmed by the absence of GFP expression (Figure 5E,F, Figure S8). At 100 ng RNA per 2.5e4 cells, ePNMA2(Cre) delivery produced higher levels of GFP reporter expression relative to an equivalent amount of undigested and unpackaged RNA(Figures 5E,F). Even a low dose (3.125 ng RNA per 2.5e4 cells) of ePNMA2(Cre) was sufficient to induce GFP expression in roughly 7% of Neuro2A-*loxP*-GFP reporter cells (Figure S8). These data show that ePNMA2 can protect a functional RNA cargo from nuclease degradation, a key characteristic for nucleic acid delivery vehicles due to abundant nuclease activity in the extracellular milieu [114,115]. These results demonstrate the potential of ePNMA2 as a gene transfer tool in mammalian cell lines.

3.4 Discussion

In the current study, we demonstrated that human PNMA2 is robustly secreted from cells as an icosahedral, non-enveloped capsid. We showed that although PNMA2 does not package RNA in human cells, an engineered variant with an RNA-binding domain grafted on to the C-terminus can package cargo RNA *in vitro*. Combining these self-assembled, packaged ePNMA2 capsids with the cell-penetrating peptide LAH4 led to efficient functional delivery of mRNA.

Our demonstration of an all protein, *in vitro* produced delivery vehicle offers a starting point for further bioengineering. For example, increasing positive charges in the ePNMA2 capsid lumen could enhance RNA packaging efficiency. Further engineering of the ePNMA2 capsid surface residues may allow robust cell entry without LAH4, or targeted cell-type or tissue tropism. Engineering strategies applied to AAVs, which bear a similar T=1 icosahedral capsid structure, could be used to modify the ePNMA2 capsid surface with integrin binding motifs or nanobodies and thus modulate ePNMA2 tropism [116,117]. The tropism and immunogenicity of these vectors *in vivo* merits further investigation. Finally, our work with PNMA2 may be extended to other PNMA family members, some of which also form capsids (Figure S9A). This, together with the fact that many PNMA family members are most highly expressed in the central nervous system (CNS) (Figures S9B,C), raises the question of whether secreted PNMA capsids may enter specific recipient cells in the CNS. Further investigation of PNMA cell tropism and engineering of PNMA capsids may allow these vehicles to be harnessed for delivery of genetic cargoes to the brain, a long-standing goal in the delivery field.

3.5 Materials and Methods

3.5.1 Determination and comparison of domain architecture of the PNMA family

A structural model was built for each member of the PNMA family using AlphaFold2 [118] under the colabfold framework [119] using default parameters. Models with plddt >=70 were selected for analysis, and additional AlphaFold2 cycles were performed until plddt was greater than 70. Structures were analyzed and compared usingPyMOL (The PyMOL Molecular Graphic System Version 1.2, Schrodinger, LLC) to annotate protein domain architecture. Hydrophobic Cluster Analysis was used to compare local structure and patterns across all PNMAs [120,121]. The RRM-like domain was identified from structural mining using the Dali server [122,123]. Domain architectures were compared across all human PNMAs, the marsupial PNMA (NCBI accession number: BAK55632.1), and the turtle Ty3/mdg4 (NCBI accession number: XP_048704523), which was the closest non-PNMA relative we identified from a preliminary phylogenetic analysis from homologs of PNMAs. A final tree was built using PhyML [124] on the MPI Bioinformatics Toolkit website [125] with LG model and 200 replicates. The final tree was visualized with the interactive tree of life (itol) webserver [126] (Figure 1B).

3.5.2 Prediction and analysis of capsomer assemblies

Pentamer assembly of PNMA2 was predicted using Alphafold2 multimer [127] under the colabfold framework using 40 cycles and 5 replicas. All replicas formed a capsomer in which the capsid domain forms a ring pentamer, and the N-terminal region forms dimers leading to two dimers and one monomer in the pentamer assembly. The interaction region was evaluated and analyzed using PyMOL software.

3.5.3 Plasmid Cloning

PNMA open reading frames (ORFs) were human codon optimized and ordered as gblocks from Twist (Supplemental Table 1). Gblocks were also cloned into a CMV promoter driven mammalian expression backbone (Addgene #11153) with the WPRE and SV40 polyadenylation signal (Addgene #83281) via Gibson Assembly (NEB E2611S) (Supplemental Table 2 primer set 1), after which C-terminal HA tags were added via site-directed mutagenesis (NEB M0554S) (Supplemental Table 2 primer sets 2-18). A plasmid encoding the PNMA2 transcript driven by was generated by nested PCRs of the human PNMA2 exons from HeLa genomic DNA (New England Biolabs N4006S). PNMA2 sequence specific primers were designed using PrimerBlast (NCBI), and PCR fragments were subject to nested PCRs and then joined via Gibson Assembly (Supplemental Table 2 primer sets 19-25), after which they were subcloned into a CMV expression backbone by Gibson Assembly (Supplemental Table 2 primer sets 26-27). Site-directed mutagenesis was used to ablate start codons (Supplemental Table 2 primer sets 28-30) such that PNMA2 protein was no longer expressed as measured via western blot. CRISPRa guide RNAs and a non-targeting control phosphorylated, annealed, and then cloned

into the PB-Unisam CRISPRa backbone (Addgene #99866) using Golden Gate assembly (NEB R0539S) (Supplemental Table 2 primer sets 31-33).

For *in vitro* expression constructs, hsPNMA ORFs were cloned from HeLa genomic DNA using primers designed using PrimerBlast (NCBI) into an *E. coli* expression backbone (Addgene #104129) with an N-terminal Maltose Binding Protein (MBP) tag and bdSUMO for purification via Gibson Assembly (Supplemental Table 2 primer sets 34-40). MePNMA was cloned into the same backbone from the gblock described above (Supplemental Table 2 primer sets 34, 41). The c-terminal disordered region of PNMA2 was replaced with the CCMV RNA binding peptide via nested PCRs to produce fragments that were subject to Gibson Assembly (Supplemental Table 2, primer sets 42-44). For *in vitro* transcription, a plasmid was generated with Cre RNA downstream of T7 promoter (Supplemental Table 2 primer set 45). One hundred A's were inserted at the 3' end of the Cre RNA to serve as a poly A sequence. The PsiI digestion site was inserted downstream of the poly A sequence (Supplemental Table 2 primer set 46).

3.5.4 In vitro production and purification of PNMA proteins

In vitro PNMA expression plasmids described above were transformed into Rosetta 2 (DE3) pLyseS cells. A single colony was inoculated in Terrific Broth (TB) media overnight at 37°C with 100 ug/mL ampicillin and 25 ug/mL chloramphenicol. When optical density 600 (OD600) reached 0.6, the culture was cooled to 4°C for 30 minutes. IPTG was added to a final concentration of 0.5 mM, and the culture was incubated at 21°C for 20 hours. Bacteria were centrifuged at 4000 rpm for 15 minutes, media supernatant was decanted, and the bacterial pellet was then resuspended in a lysis buffer containing 50 mM Tris pH 8, 250 mM NaCl and 0.5 mM TCEP. Lysis was achieved with two passes through the LM20 Microfluidizer system at 27,000 p.s.i. The lysis was cleared with centrifugation at 9000 rpm for 30 minutes. The lysis was incubated with 2 mL amylose beads for 2 hours at 4°C. The amylose beads were washed, and the bound PNMA2 was cleaved overnight with lysis buffer with 1.5% NP-40 and 1ug/mL bdSENP1. The elution was collected and used for SEC analysis using an AKTA pure system with Superdex200 increase 10/300 GL column with an isocratic run using lysis buffer at 0.4 mL/min.

3.5.5 Negative staining and transmission electron microscopy

For sample preparation of TEM imaging grids, 5 ul of sample at a protein concentration of approximately 0.3 mg/ml was loaded onto glow-discharged, carbon-coated 300-mesh copper grids (Electron Microscopy Sciences #FCF300-CU-50). Sample was adhered to the grid for one minute at room temperature and stained in five sequential droplets for a total of one minute in freshly filtered 2% uranyl formate. Following the staining procedure, excess uranyl formate was carefully blotted off with Whatman filter paper (Cytiva, #1001-032). The grid was dried at room temperature for 1 minute before placement into a grid holder. All TEM images were acquired using the FEI Tecnai (G2 Spirit TWIN) 120 kV multipurpose TEM at the MIT MRL facility. The grid was mounted on a JEOL single tilt holder equipped in the TEM column and cooled down with liquid nitrogen. The microscope was operated at 200 kV and with a magnification in the

range of 10,000~60,000x, and all images were recorded on a Gatan 2kx2k UltraScan CCD camera.

3.5.6 Cell Culture

U20S cells (ATCC HTB-96) were maintained in McCoy's 5A (Modified) Medium supplemented with 10% fetal bovine serum and 100 U/mL penicillin-streptomycin.
HEK293FT (Thermo Fisher R700-07), HeLa (ATCC CCL-2), U87 (ATCC HTB-14) and Neuro2A (ATCC CCL-131) cells were maintained in Dulbecco's Modified Eagle Medium supplemented with 10% fetal bovine serum and 100 U/mL penicillin-streptomycin. U20S, HeLa, U87 and Neuro2A cells were transfected with Lipofectamine 3000 (ThermoFisher, L3000001) at 80% confluence, and media was changed 4 hours post transfection to reduce toxicity. HEK293FT were transfected at 70% confluence with PEI HCl MAX (Polysciences 24765-1).

Neuro2A-*loxP*-GFP stable reporter cells were generated by subcloning the *loxP*-GFP cassette from RV-Cag-Dio-GFP (Addgene #87662) into a lentiviral transfer plasmid encoding a Blasticidin resistance gene for stable integration. To produce virus, HEK293FT cells were seeded at 1e7 cells per 15-cm dish. 16 hours later, cells were co-transfected with 5 ug psPAX2 (Addgene #12260), 4.7 ug pMD2.G (Addgene #12259), and 7.7 ug of the Cre reporter plasmid using PEI HCl MAX (Polysciences 24765-1), and media was changed 4 hours post transfection.
Forty-eight hours later, viral supernatant was harvested, spun at 2000 g for 10 mins to remove cell debris, filtered through a 0.45-μm filter, and stored at -80°C degrees. Neuro2A reporter cell lines were created by lentiviral transduction with 8 ug/mL polybrene(TR1003G). Media was changed one day later, and cells were selected for two weeks starting on day 3 with 10 ug/mL Blasticidin-HCl (Thermo Fisher Scientific A1113903). Single clones of Blasticidin-resistant cells were isolated by serial dilution, expanded, and then screened for successful reporter expression by transfection of a Cre encoding plasmid.

3.5.7 Isolation of VLPs from human cells

Forty-eight hours post transfection, media supernatant for sucrose cushion purification was filtered through a 0.45-μm filter, added to conical ultracentrifuge tubes (Beckman Coulter 358126), and underlaid with 4 mL of 20% sucrose in 1X PBS. Tubes were then spun at 120,000g for 2 hours in a Beckman Coulter SW28 rotor, after which supernatant was decanted and the pellet was resuspended in 100 uL of 1X PBS.

3.5.8 Western blot analysis

For cellular lysate, cells were washed in 1X PBS and lysed in RIPA buffer (ThermoFisher 89901) with Halt protease inhibitor (ThermoFisher 87786) for 30 minutes at 4°C. Lysate was then spun at 20,000g for 10 minutes at 4°C to pellet insoluble protein. Cellular lysate supernatant and resuspended VLP were combined with 1X Bolt LDS Sample Buffer (Life Technologies B0007) and 100 mM DTT and boiled at 95°C for 10 minutes. Samples were loaded into Bolt 4-12% Bis-Tris Plus gels (ThermoFisher) and run at 200V for 30 minutes, before being transferred onto PVDF membrane with the iBlot2 system (ThermoFisher).

Membranes were blocked in 5% milk in 1X TBS Tween 20 (TBST) Buffer, and incubated at 4°C overnight with primary antibody in 2% milk in 1X TBST. Following three 1X TBST washes, samples were incubated with secondary antibody for 1 hour and then imaged. Antibodies and dilutions used are listed in Supplemental Table 3.

3.5.9 Immunofluorescence and confocal microscopy

Cells were seeded at 5e4 cells/well on Poly-D-Lysine/Laminin coated glass coverslips (VWR 354087). The following day, cells were washed with 1X PBS, fixed in 4% PFA in PBS for 30 minutes, permeabilized in 0.1% Triton X100 for 30 minutes, and then blocked in 1% BSA for 30 minutes. Cells were stained with anti-PNMA2 primary antibody diluted 1:200 in 1% BSA for one hour, washed and then stained with AlexaFluor488 conjugated secondary antibody for 1 hour in the dark (as detailed in Supplemental Table 3). Cells were then stained with Alexa-Fluor 647 Phalloidin (Cell Signaling Technologies 8940S) and DAPI at 0.01 mg/mL for 5 minutes, washed three times with 1X PBS, and mounted in Diamond ProLong mounting media on glass slides. Mounted specimens were imaged on a Leica Stellaris 5 confocal microscope with the 63X oil objective.

3.5.10 HA Immunoprecipitation of HEK secreted HA-PNMA2

ePNMA2 VLPs were isolated from the supernatant of HEK293FT cells transfected with PNMA2 with an HA-tag at the N-terminus as described previously. Following ultracentrifugation, HA-tag pulldown was performed using the HA tagged protein purification kit from MBL (#3320) using the manufacturer's instructions. Successful isolation was confirmed by Coommassie staining and protein capsids were imaged using TEM as described previously.

3.5.11 RNA isolation and RT-qPCR

Cells or VLPs were resuspended in Trizol (Thermo Fisher 15596026), vortexed, and incubated at room temperature for 5 minutes. Total RNA was then prepared via phenol chloroform extraction with glycoblue coprecipitate (Thermo Fisher AM9515). DNA contaminants were removed using the Ambion Turbo DNA-free kit (Thermo Fisher AM1907), after which DNAsed RNA was reverse transcribed using random hexamer priming and the SmartScribe Reverse Transcriptase Kit (Takara Bio 639537). cDNA was then input into qPCR reactions with Fast Sybr Green Master Mix (Life Technologies 4385612) and signal was quantified with the BioRad CFX Opus system. qPCR primers (sets 47 and 48) are listed in Supplemental Table 2.

3.5.12 RNAseq of cells and VLPs

U2OS cells were selected for CRISPRa due to the cell line expressing PNMA2 protein (measured via western blot) after CRISPRa treatment. U2OS cells were seeded at 6e6 cells per plate in 15-cm tissue culture dishes and transfected the following day with CRISPRa cassettes containing non-targeting guides or guides targeted against the transcriptional start site of PNMA2 with Lipofectamine 3000 (ThermoFisher L3000001) per the manufacturer's protocol. Forty-eight hours post transfection, media was harvested and centrifuged on a 20% sucrose cushion as described above. VLP pellets were resuspended in 1X PBS with 2 mM $MgCl_2$ and

250 units of Benzonase (Sigma-Aldrich E1014), incubated at 37°C for 1 hour to degrade non-encapsidated RNAs, and then resuspended in Trizol (ThermoFisher 15596026). Cells were washed in 1X PBS and then split into two aliquots. One aliquot of 6e5 cells were resuspended in lysis buffer and subject to western blot as described above to confirm CRISPRa efficacy, while another aliquot of 6e5 cells were separately resuspended in Trizol for mRNA isolation via phenol chloroform extraction. Samples were prepared in biological triplicate. Following DNAse treatment, RNA concentrations were normalized, and RNAseq libraries were prepared with the NEBNext Ultra II Directional RNA Library Prep Kit (New England Biolabs E7765S) per the manufacturer's directions. RNAseq libraries were quantified and normalized with the KAPA library quantification kit (Roche 07960204001) and loaded onto an Illumina NextSeq 550 with 50 cycles for read 1 and 25 cycles for read 2. Raw reads were trimmed using Trimmomatic [128] and quality control was performed using fastqc [129] to eliminate low quality reads and adaptors. Resulting reads were mapped to a reference of the human genome (GRCh38) using STAR [130], and full read alignments were converted to indexed BAM files with SAMtools [131]. A counts table was generated using htseq [132] and used to perform differential gene expression analysis using DESeq2 [133] in R.

3.5.13 Cryo-electron microscopy

Assembled PNMA2 capsids were diluted to 1.5 mg/mL in PBS, and 4 µL was applied to a freshly glow-discharged (60 s at 25 mA) Cu300 R1.2/1.3 holey carbon grid (Quantifoil) mounted in the chamber of a Vitrobot Mark IV (Thermo Fisher Scientific) maintained at 4°C and 100% humidity. The grid was blotted with Ø55 grade 595 filter paper (Ted Pella) for 4 seconds after a wait time of 0 seconds at a blot force of +10, and after a drain time of 1 second was plunged into liquid ethane. Cryo-EM data were collected using the Thermo Scientific Titan Krios G3i at MIT.nano using a K3 detector (Gatan) operating in super-resolution mode with 2-fold binning and an energy filter with slit width of 20 eV. Micrographs were collected using EPU in AFIS mode, yielding 17,600 movies at 130,000x magnification with a real pixel size of 0.6788 Å, a defocus range from -1 to -2.6 µm, an exposure time of 0.6 seconds fractionated into 24 frames, a flux of 23.6 e^-/pix/s and a total fluence per micrograph of 30.7 $e^-/Å^2$. Cryo-EM data were processed using RELION 4.0 [134]. Movies were corrected for motion using the RELION implementation of MotionCor2, with 4x4 patches and dose-weighting, and CTF parameters were estimated using CTFFIND-4.1 [135]. Particles were picked using Topaz and a general model [136], yielding 722,571 particles which were extracted with a 512 pixel box, binned to 128 pixels, and classified using the VDAM 2D classification algorithm (Figure S4A,B). 229,149 particles with high-quality 2D averages were re-extracted with a 512 pixel box binned to 360 pixels. 3D refinement with I4 symmetry, using an initial model generated by RELION from screening data on a Talos Arctica microscope, gave a 3.4 Å reconstruction, however the map showed radial blurring suggesting individual capsids had slightly different radii (Figure S4C). 3D classification with regularization parameter T = 15 allowed isolation of 88,320 capsids that were slightly smaller than average and had more well-defined density; these refined to 3.3 Å resolution after CTF refinement and Bayesian particle polishing but still showed some radial blurring. To improve the density, individual capsid particle images were converted to 12 sub-particles corresponding to individual pentons (Figure S4C). This was done by I4 symmetry expansion in RELION to convert each particle to 60 subparticles, then only keeping the 1, 2, 3, 4, 5, 6, 10, 12,

20, 28, 29 and 38th subparticles, then performing particle subtraction with a mask around one of the z-axis-aligned pentons, and finally correcting for the local defocus of the subparticle based on its projected distance to the capsid center. Subparticles were then refined with C5 symmetry and 0.9° local angular sampling, producing a 3.1 Å reconstruction of an individual PNMA2 penton. Resolution is reported using the gold-standard Fourier Shell Correlation with 0.143 cutoff. The AlphaFold2 model of PNMA2 was docked into the penton cryo-EM density and adjusted using Coot [137]. The model was duplicated around the 2-, 3- and 5-fold axes to produce all interfaces, then refined using ISOLDE [138]. The extra monomers were then deleted, and the original monomer was duplicated with I4 symmetry and refined using PHENIX real_space_refine [139] into the I4-symmetric overall map using the starting model as a reference (sigma = 0.1), one macrocycle of global minimization and ADP refinement, and a nonbonded weight of 2000. Structural figures were generated using UCSF ChimeraX [140].

3.5.14 In vitro assembly and disassembly of PNMA2 and ePNMA2 capsids

Purified PNMA2 protein was pH adjusted to 5, 6, 7, 8, 9, 10, 11 and 12 and NaCl concentration adjusted to 25 mM and 1M. Divalent ions were screened with addition of 10 mM MgCl2, 10 mM CaCl2 or 100 uM ZnCl2 into 50 mM Tris pH 8 with varying concentrations of NaCl. Co-addition of 10 mM CaCl2 and 100 uM ZnCl2 were tested with addition of 10 mM CaCl2, 100 uM ZnCl2 or 10 mM CaCl2 and 100 uM ZnCl2 into 50 mM Tris pH 8 with varying concentrations of NaCl. PNMA2 protein was mixed with purified RNA at a molar ratio of 10:1 in a buffer containing 50 mM HEPES pH 8 on ice for 2 hrs. The mixture was dialyzed overnight in a buffer containing 50 mM HEPES pH 8, 500 mM NaCl, 10 mM CaCl2, and 0.5 mM TCEP and varying NaCl concentration. ePNMA2 was treated with 0 M NaCl or 1 M NaCl in addition to 0 M Urea, 1 M Urea or 6 M Urea. ePNMA2 concentration was kept the same for all conditions.

3.5.15 In vitro transcription of Cre mRNA

A plasmid encoding Cre mRNA was digested with PsiI for 1 hour at 37°C, then purified with the QIAquick PCR clean up kit (Qiagen). mRNA was synthesized using Hiscribe T7 ARCA mRNA kit (NEB E2060S) according to the manufacturer's protocol for mRNA synthesis with 50% modified nucleotides 5mCTP and Pseudo-UTP (TriLink Biotechnologies N10145 and N10195). mRNA synthesis was performed without enzymatic polyadenylation due to the polyadenylation sequence being present in our plasmid template. After 30 minutes incubation at 37 °C, 28 µl of water and 2 µl of Dnase I were added to the reaction for 15 minutes at 37 °C. 25 µl of 3M LiCl solution was added and incubated at -20 °C for 30 minutes to precipitate RNA, and RNA precipitate was collected by centrifugation for 15 minutes at 21000 gat 4 °C. The pellet was washed with 500 µl cold 70% ethanol, then dissolved in 20 µl nuclease-free water. mRNA quality was checked by running a 1% E-gel. mRNA concentration was measured by nanodrop.

3.5.16 In vitro packaging of ePNMA2 capsid

ePNMA2 was purified with the same protocol as other PNMAs with the following modifications. Bacterial pellet was lysed in 50 mM Tris pH 8, 1 M NaCl and 0.5 mM TCEP. After overnight bdSENP1 cleavage, ePNMA2 was collected, concentrated into 1 mL, injected into a 2 mL-loop and passed through Superose 6 increase 10/30 GL column using AKTA pure system. The retention volume 9 ml to 13.5 ml was collected.
Purified ePNMA2 was diluted 10 times into a buffer containing 50 mM HEPES pH 8, 6.6 M Urea. The solution was left at 4 °C for at least 30 min. The solution was purified using Hitrap Heparin 5 mL column using the following protocol:

Set program	Volume
Sample application	Continuous application until entire volume is run through column
Column wash: Buffer A	5 column volume
Gradient elution: 0% Buffer B to 100% Buffer B	10 column volume
Column wash: 100% Buffer B	5 column volume
Equilibration: Buffer A	5 column volume

Buffer A: 50 mM HEPES pH 8, 100 mM NaCl, 6M Urea and 0.5 mM TCEP
Buffer B: 50 mM HEPES pH 8, 2 M NaCl, 6M Urea and 0.5 mM TCEP

ePNMA2 protein was mixed with purified RNA at a molar ratio of 10:1 in a buffer containing 50 mM HEPES pH 8 and 6 M Urea on ice for 2 hrs. The mixture was dialyzed overnight in a buffer containing 50 mM HEPES pH 8, 10 mM CaCl2,0.5 mM TCEP and varying NaCl concentration. ePNMA2 capsid was treated with 10ng/uL RNaseA at room temperature for 10 min. The encapsulated mRNA was extracted using Trizol reagent from 21.3 ug of input protein, as quantified by running an SDS PAGE gel and quantifying against a BSA standard using FIJI. RT-PCR was performed to quantify RNA concentration. Full-length mRNA concentration was further quantified by running 1% agarose gel against Cre mRNA standard.

3.5.17 Transduction assays

VLP samples or a naked mRNA control were normalized to the same mRNA concentration, and treated with 10 ug/mL RNAse A (Qiagen) for 30 minutes at room temperature to degrade un-encapsidated RNAs. A second mRNA sample was prepared at the same concentration without RNAse A treatment as a positive control. Neuro2a-*loxP*-GFP reporter cells were seeded at 1.5e4 cells/well in a 96-well format. The following day, positive control mRNA, RNAse treated mRNA, and VLP samples were mixed with LAH4 cell penetrating peptide (Genscript RP20096) in Optimem (Gibco) and added to cells at a final concentration of 10 ug/mL LAH4. Media was

changed the day after VLP and mRNA treatment. Ninety-six hours post treatment, cells were imaged on an EVOS M5000 fluorescence microscope (Thermo Fisher) at 20X using the GFP and transmitted light channels. Cells were then prepared for flow cytometry as follows. Cells were washed with 1X PBS, trypsinized, resuspended in full media, and spun at 1000g for 3 minutes in a 96-well V-bottom plate. Cells were washed with FACS buffer (1x PBS with 2% FBS and 2 mM EDTA), spun at 1000g for 3 minutes, and then resuspended in FACS buffer with DAPI at .01 mg/mL. Following an additional FACS buffer wash, cell fluorescence was read out on a CytoFlex S Flow Cytometer (Beckman Coulter).

3.6 References

1. J. A. Kulkarni, *et al.*, The current landscape of nucleic acid therapeutics. *Nat. Nanotechnol.* **16**, 630–643 (2021).

2. M. P. Hantak, J. Einstein, R. B. Kearns, J. D. Shepherd, Intercellular Communication in the Nervous System Goes Viral. *Trends Neurosci.* **44**, 248–259 (2021).

3. M. Segel, *et al.*, Mammalian retrovirus-like protein PEG10 packages its own mRNA and can be pseudotyped for mRNA delivery. *Science* **373**, 882–889 (2021).

4. R. Ono, *et al.*, Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. *Nat. Genet.* **38**, 101–106 (2006).

5. E. D. Pastuzyn, *et al.*, The Neuronal Gene Arc Encodes a Repurposed Retrotransposon Gag Protein that Mediates Intercellular RNA Transfer. *Cell* **172**, 275–288.e18 (2018).

6. J. Ashley, *et al.*, Retrovirus-like Gag Protein Arc1 Binds RNA and Traffics across Synaptic Boutons. *Cell* **172**, 262–274.e11 (2018).

7. M. Abed, *et al.*, The Gag protein PEG10 binds to RNA and regulates trophoblast stem cell lineage specification. *PLoS One* **14**, e0214110 (2019).

8. S. Iwasaki, *et al.*, Identification of a novel PNMA-MS1 gene in marsupials suggests the LTR retrotransposon-derived PNMA genes evolved differently in marsupials and eutherians. *DNA Res.* **20**, 425–436 (2013).

9. R. Voltz, *et al.*, A serologic marker of paraneoplastic limbic and brain-stem encephalitis in patients with testicular cancer. *N. Engl. J. Med.* **340**, 1788–1795 (1999).

10. J. Dalmau, *et al.*, Ma1, a novel neuron- and testis-specific protein, is recognized by the serum of patients with paraneoplastic neurological disorders. *Brain* **122 (Pt 1)**, 27–39 (1999).

11. M. Schüller, D. Jenne, R. Voltz, The human PNMA family: novel neuronal proteins implicated in paraneoplastic neurological disease. *J. Neuroimmunol.* **169**, 172–176 (2005).

12. M. R. Rosenfeld, J. G. Eichen, D. F. Wade, J. B. Posner, J. Dalmau, Molecular and clinical diversity in paraneoplastic immunity to Ma proteins. *Ann. Neurol.* **50**, 339–348 (2001).

13. Y. H. Lee, S. W. Pang, C. L. Poh, K. O. Tan, Distinct functional domains of PNMA5 mediate protein-protein interaction, nuclear localization, and apoptosis signaling in human cancer cells. *J. Cancer Res. Clin. Oncol.* **142**, 1967–1977 (2016).

14. Y. H. Lee, S. W. Pang, K. O. Tan, PNMA2 mediates heterodimeric interactions and antagonizes chemo-sensitizing activities mediated by members of PNMA family. *Biochem. Biophys. Res. Commun.* **473**, 224–229 (2016).

15. S. Baksh, *et al.*, The tumor suppressor RASSF1A and MAP-1 link death receptor signaling to Bax conformational change and cell death. *Mol. Cell* **18**, 637–650 (2005).

16. K. O. Tan, *et al.*, MAP-1 is a mitochondrial effector of Bax. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 14623–14628 (2005).

17. K. O. Tan, *et al.*, MAP-1, a novel proapoptotic protein containing a BH3-like motif that associates with Bax through its Bcl-2 homology domains. *J. Biol. Chem.* **276**, 2802–2807 (2001).

18. J. Xu, *et al.*, PNMA2 forms non-enveloped virus-like capsids that trigger paraneoplastic neurological syndrome. *bioRxiv* (2023) https:/doi.org/10.1101/2023.02.09.527862.

19. D. Chalopin, D. Galiana, J.-N. Volff, Genetic innovation in vertebrates: gypsy integrase genes and other genes derived from transposable elements. *Int. J. Evol. Biol.* **2012**, 724519 (2012).

20. A. A. Bayev Jr, *et al.*, Structural organization of transposable element mdg4 from Drosophila melanogaster and a nucleotide sequence of its long terminal repeats. *Nucleic Acids Res.* **12**, 3707–3723 (1984).

21. S. Sandmeyer, K. Patterson, V. Bilanchone, Ty3, a Position-specific Retrotransposon in Budding Yeast. *Microbiol Spectr* **3**, MDNA3–0057–2014 (2015).

22. Q. Xie, *et al.*, The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 10405–10410 (2002).

23. R. F. Garmann, C. M. Knobler, W. M. Gelbart, Protocol for Efficient Cell-Free Synthesis of Cowpea Chlorotic Mottle Virus-Like Particles Containing Heterologous RNAs. *Methods Mol. Biol.* **1776**, 249–265 (2018).

24. M. Comas-Garcia, R. D. Cadena-Nava, A. L. N. Rao, C. M. Knobler, W. M. Gelbart, In vitro quantification of the relative packaging efficiencies of single-stranded RNA molecules by viral capsid protein. *J. Virol.* **86**, 12271–12282 (2012).

25. A. Kichler, A. J. Mason, B. Bechinger, Cationic amphipathic histidine-rich peptides for

gene delivery. *Biochim. Biophys. Acta* **1758**, 301–307 (2006).

26. N. B. Y. Tsui, E. K. O. Ng, Y. M. D. Lo, Stability of endogenous and added RNA in blood specimens, serum, and plasma. *Clin. Chem.* **48**, 1647–1653 (2002).

27. T. C. Roberts, R. Langer, M. J. A. Wood, Advances in oligonucleotide drug delivery. *Nat. Rev. Drug Discov.* **19**, 673–694 (2020).

28. P. Crespo-Escobar, *et al.*, The role of gluten consumption at an early age in celiac disease development: a further analysis of the prospective PreventCD cohort study. *Am. J. Clin. Nutr.* **105**, 890–896 (2017).

29. W. Shi, J. S. Bartlett, RGD inclusion in VP3 provides adeno-associated virus type 2 (AAV2)-based vectors with a heparan sulfate-independent cell entry mechanism. *Mol. Ther.* **7**, 515–525 (2003).

30. L. Mendonça, *et al.*, CryoET structures of immature HIV Gag reveal six-helix bundle. *Commun Biol* **4**, 481 (2021).

31. N. Renner, *et al.*, HIV-1 is dependent on its immature lattice to recruit IP6 for mature capsid assembly. *Nat. Struct. Mol. Biol.* **30**, 370–382 (2023).

32. S. Erlendsson, *et al.*, Structures of virus-like capsids formed by the Drosophila neuronal Arc proteins. *Nat. Neurosci.* **23**, 172–175 (2020).

33. J. Jumper, *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

34. M. Mirdita, *et al.*, ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).

35. G. Faure, I. Callebaut, Comprehensive repertoire of foldable regions within whole genomes. *PLoS Comput. Biol.* **9**, e1003280 (2013).

36. I. Callebaut, *et al.*, Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cell. Mol. Life Sci.* **53**, 621–645 (1997).

37. L. Holm, Dali server: structural unification of protein families. *Nucleic Acids Res.* **50**, W210–W215 (2022).

38. L. Holm, Using Dali for Protein Structure Comparison. *Methods Mol. Biol.* **2112**, 29–42 (2020).

39. S. Guindon, *et al.*, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).

40. F. Gabler, *et al.*, Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Curr.*

Protoc. Bioinformatics **72**, e108 (2020).

41. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).

42. R. Evans, *et al.*, Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, 2021.10.04.463034 (2022).

43. O. Lenz, J. ter Meulen, H. D. Klenk, N. G. Seidah, W. Garten, The Lassa virus glycoprotein precursor GP-C is proteolytically processed by subtilase SKI-1/S1P. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 12701–12705 (2001).

44. J. Schindelin, *et al.*, Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

45. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

46. , Babraham bioinformatics - FastQC A quality control tool for high throughput sequence data (March 28, 2023).

47. A. Dobin, *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

48. H. Li, *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

49. S. Anders, P. T. Pyl, W. Huber, HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

50. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

51. D. Kimanius, L. Dong, G. Sharov, T. Nakane, S. H. W. Scheres, New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochem. J* **478**, 4169–4185 (2021).

52. A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

53. T. Bepler, *et al.*, Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019).

54. A. Casañal, B. Lohkamp, P. Emsley, Current developments in Coot for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. *Protein Sci.* **29**, 1069–1078 (2020).

55. T. I. Croll, ISOLDE: a physically realistic environment for model building into

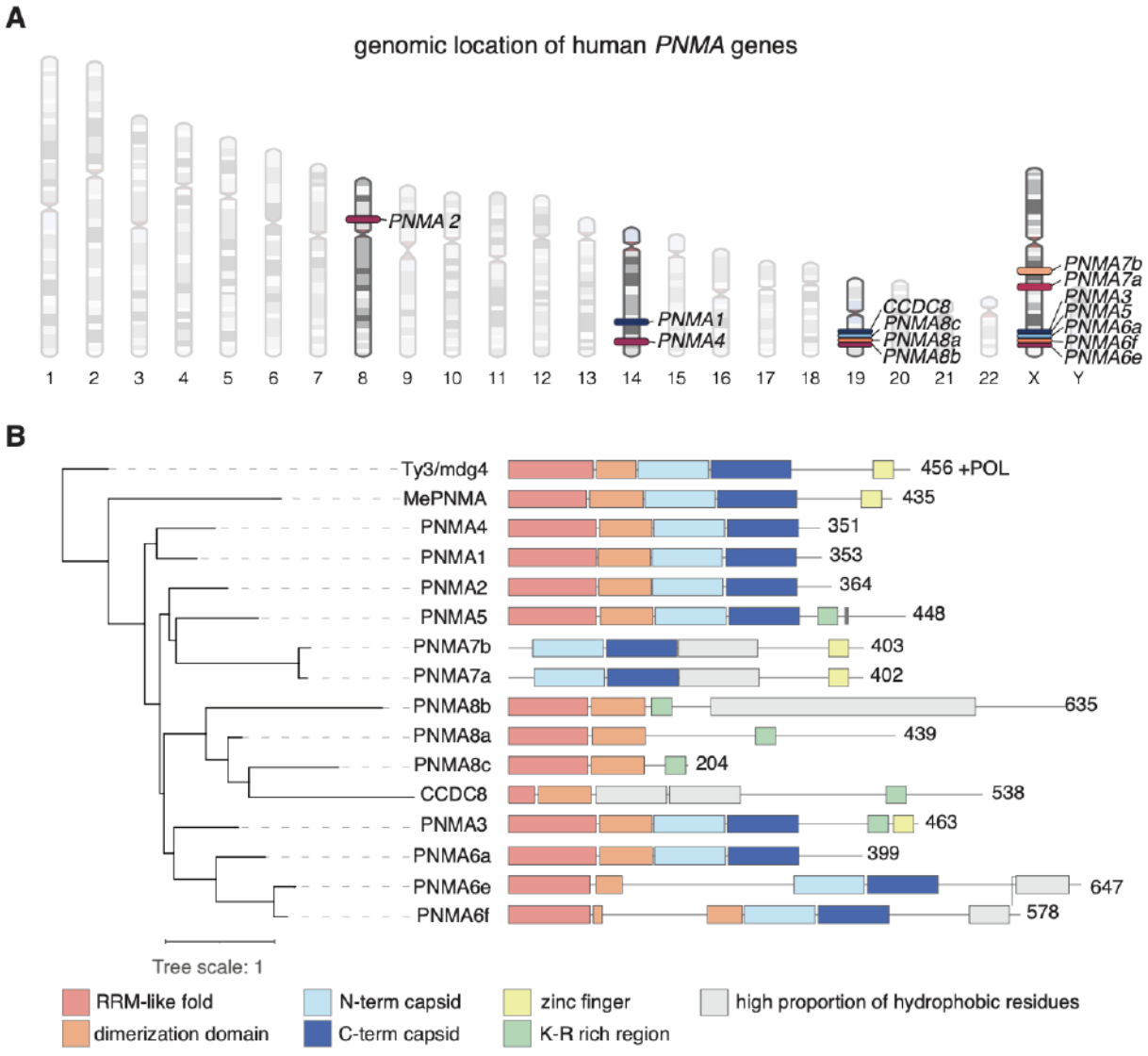
low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* **74**, 519–530 (2018).

56. P. V. Afonine, *et al.*, Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531–544 (2018).

57. E. F. Pettersen, *et al.*, UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).

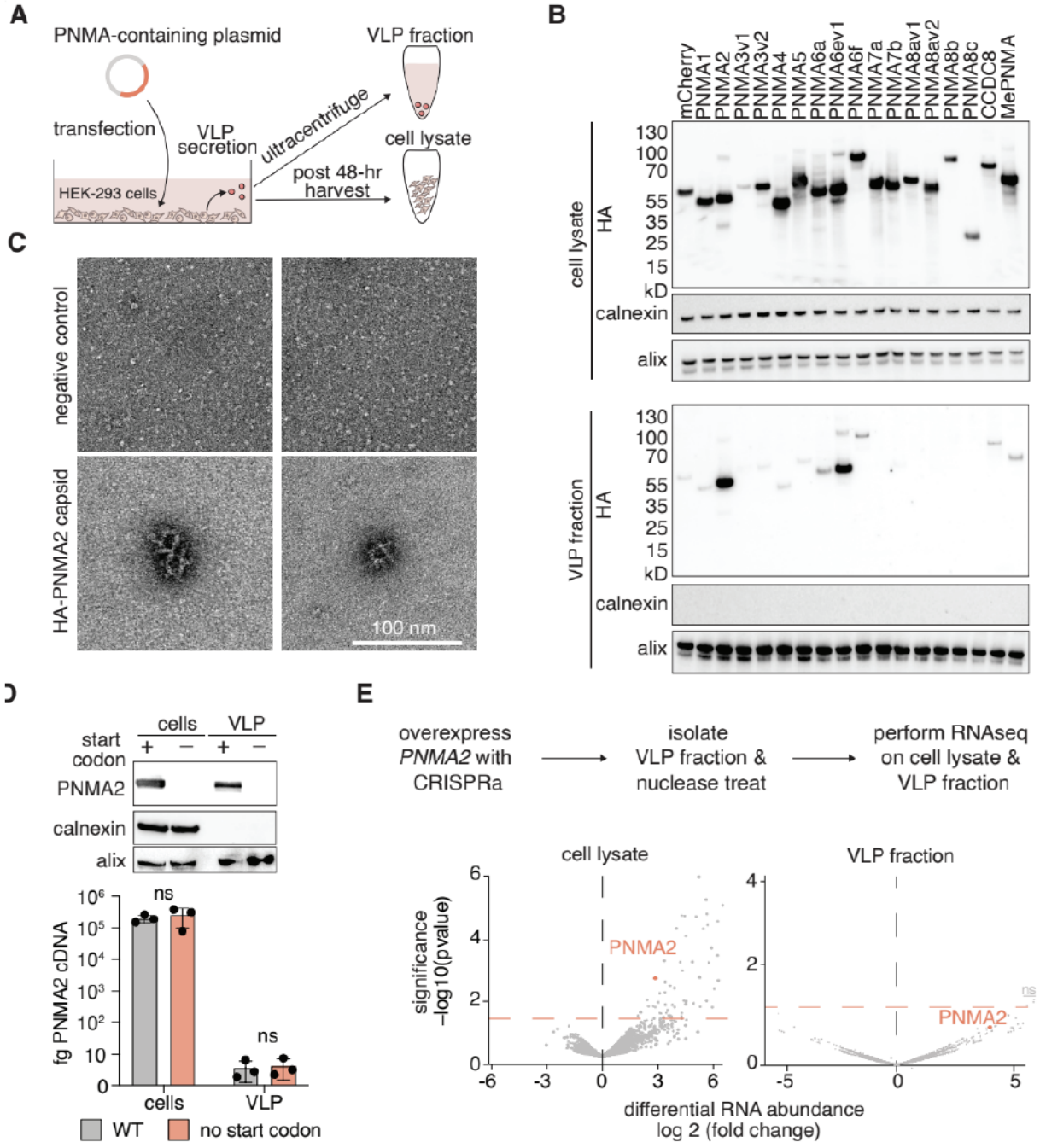
3.7 Figures and Tables

3.7.1 Figure 1. Genomic location and domain architecture of human *PNMA* genes.



A. Genomic location of human *PNMA* genes.
B. Phylogenetic tree of PNMA family including all human PNMAs (PNMA1-8 and CCDC8), the marsupial PNMA (MePNMA), and a turtle Ty3/mdg4 (see **Methods**) from which the tree is rooted. Domain architecture of each protein is deduced from the structural models (see **Methods**). Domain architecture encompasses an RRM-like fold domain (pink), dimerization domain that forms only upon interaction (in orange), and capsid domain in light blue for N-terminal capsid domain and dark blue for C-terminal capsid domain. Additional domains predicted to fold are shown in gray. Zinc fingers are shown in yellow, and regions with a high concentration of K-R are shown in green. RRM, RNA recognition motif.

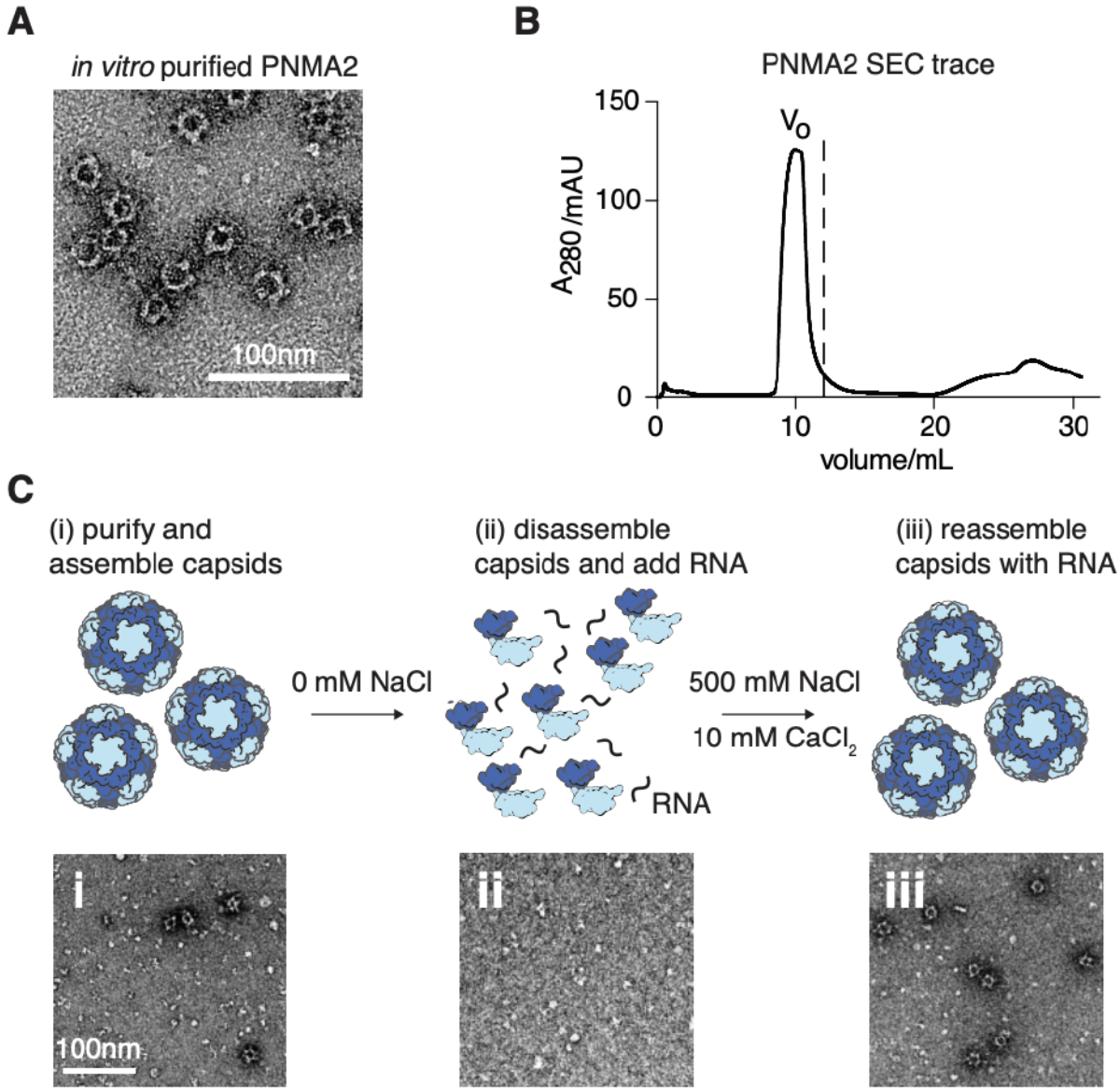
3.7.2 Figure 2. PNMA2 capsids are secreted from human cells without encapsidated RNA.



A. Schematic of isolation of cell lysate and viral-like particle (VLP) fraction.
B. Western blot showing PNMA protein expression in HEK293FT cells in either cell lysate (top) or the VLP fraction (bottom).

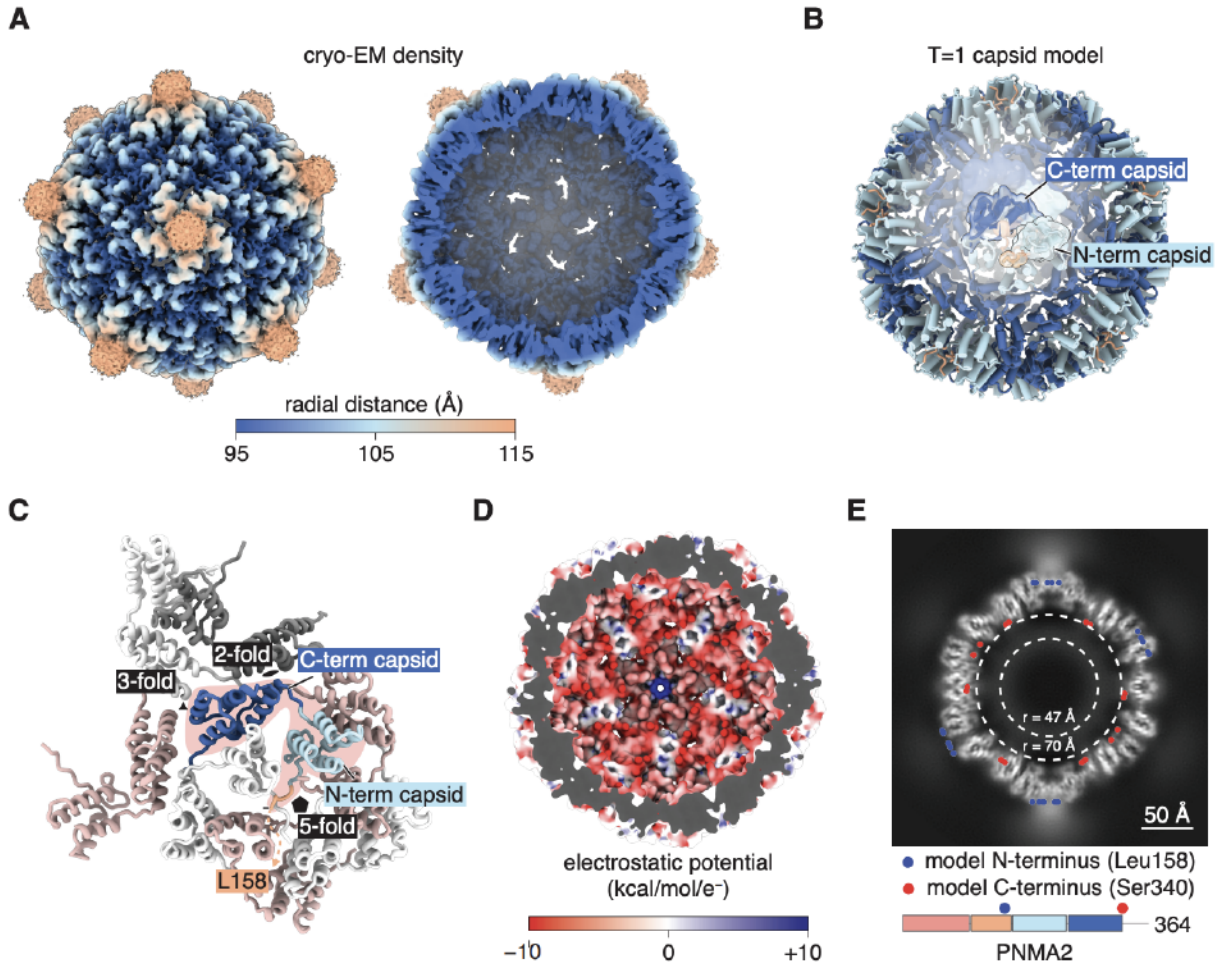
C. TEM micrographs of HA-immunoprecipitated VLP fraction from negative control cells (top panels) or cells overexpressing HA-tagged PNMA2 (bottom panels). Scale bar is 100 nm.
D. (Top) Expression of PNMA2 or a mutant of PNMA2 lacking the start codon in whole-cell lysate (cells) or the VLP fraction. (Bottom) Quantification of *PNMA2* mRNA (cDNA) in either the whole-cell lysate or VLP fraction in cells expressing either wild-type PNMA2 (black) or a mutant lacking the start codon (gray). Three transcripts (ZNF142, GOLGA2P9, and TTC25) were above the significance threshold but were found to be insignificant after adjusting for multiple hypothesis testing.
E. (Top) Schematic of experimental procedure to identify mRNA packaged in PNMA2 capsids. (Bottom) Volcano plots showing differential mRNA expression of PNMA2 CRISPRa samples versus a non-targeting guide control in either the cell lysate (left) or VLP fraction (right).

3.7.3 Figure 3. *In vitro* assembly and RNA packaging of PNMA2 capsids.



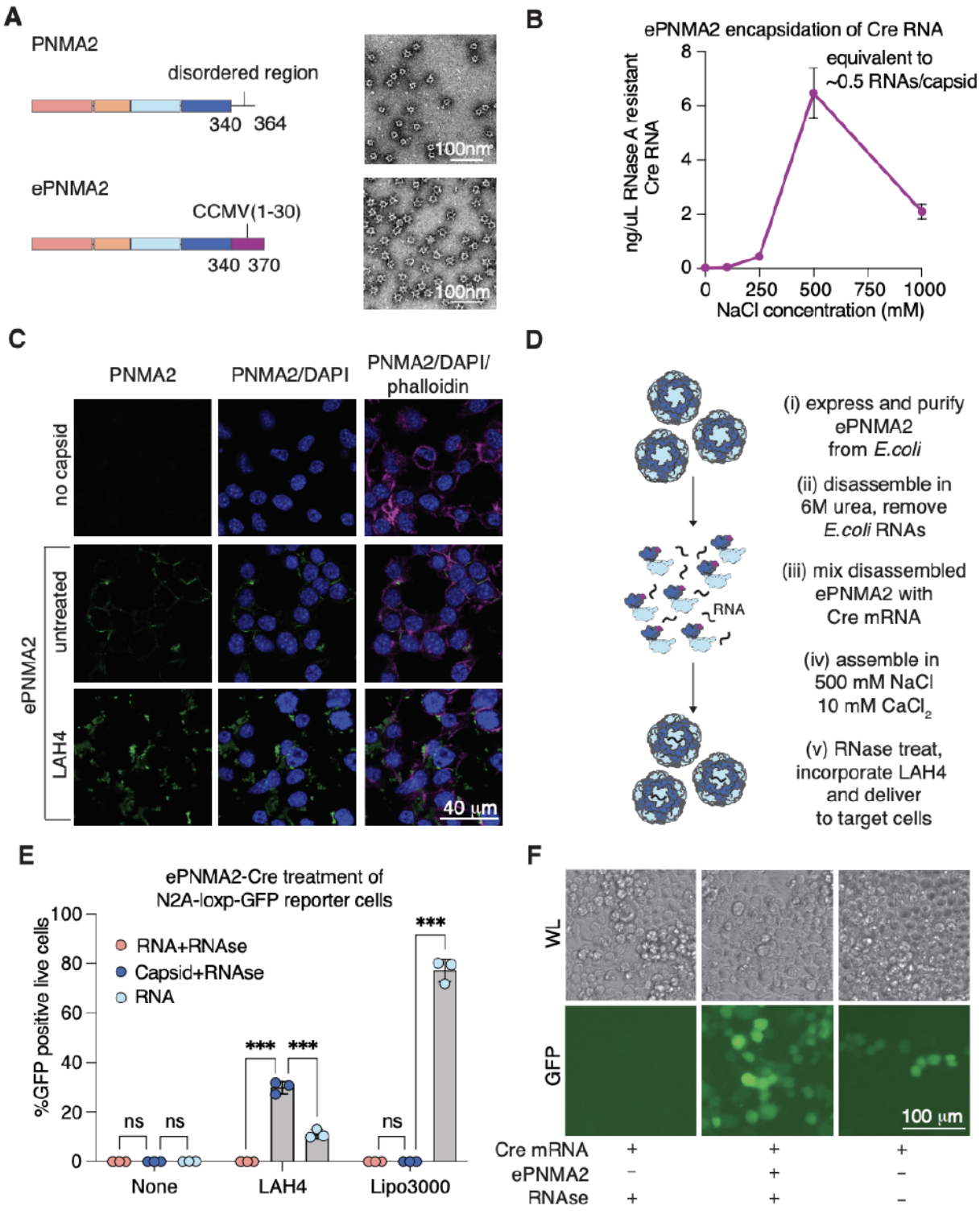
A. TEM micrograph of PNMA2 purified from *E. coli*. Scale bar is 100 nm.
B. Size-exclusion chromatography (SEC) trace of PNMA2 particles purified from *E. coli*, where V_o indicates void volume.
C. Schematic of workflow for *in vitro* production of PNMA2 capsids with representative TEM images below.

3.7.4 Figure 4. Cryo-EM structure of human PNMA2 capsids.



A. Cryo-EM density of PNMA2 with I4 symmetry imposed, colored by radial distance from the center of the capsid, with the exterior of the PNMA2 capsid pictured on the left and a cross section of the capsid shown on the right.
B. Model of the PNMA2 capsid with a protein monomer outlined.
C. Details of the interactions of a PNMA2 monomer (pink highlight) with adjacent monomers, with symmetry axes indicated.
D. Electrostatic potential of the inside of the PNMA2 capsid. Red indicates negative charge.
E. Central slice of the PNMA2 cryo-EM density. The projected positions of the modeled N- and C-termini are shown as blue and red circles.

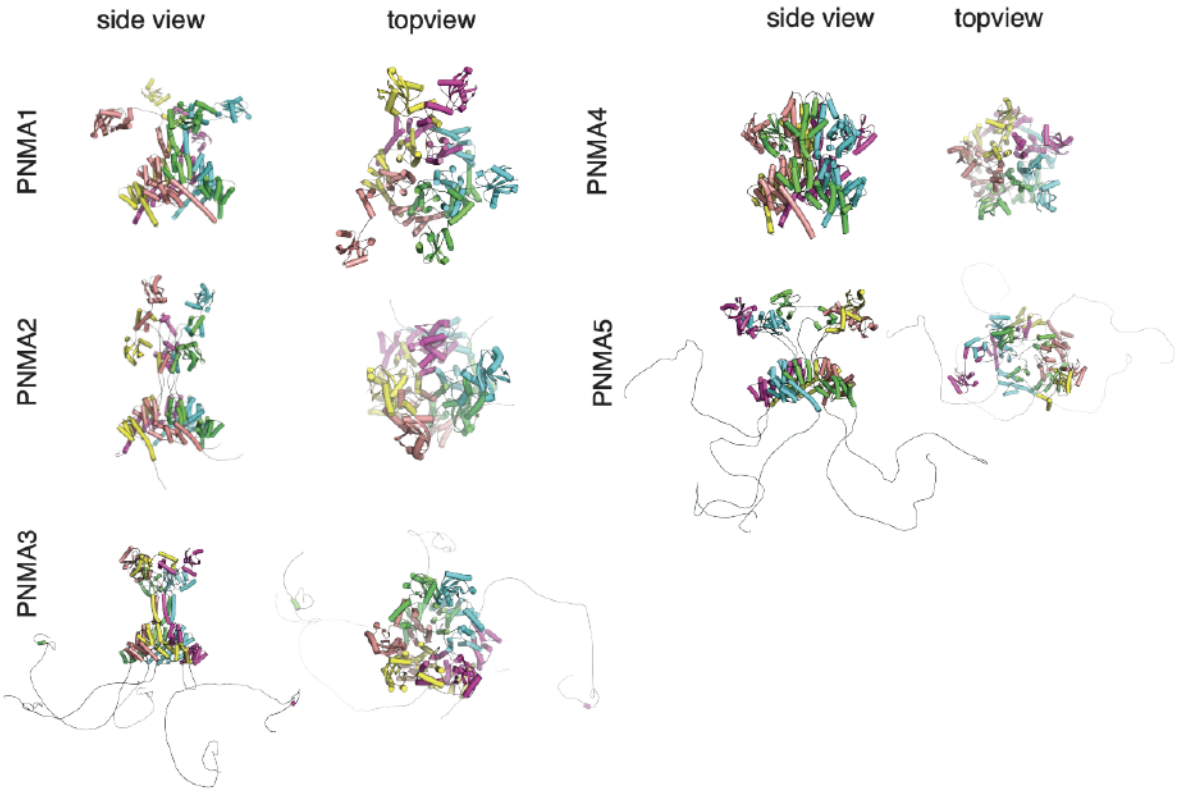
3.7.5 Figure 5. Engineering of ePNMA2 for RNA encapsidation and delivery.



A. Schematic of ePNMA2 protein and TEM micrograph of ePNMA2 capsids. Scale bar is 100 nm.
B. Quantification of RNAs per capsid packaged by ePNMA2 during reassembly across a range of salt conditions following RNaseA treatment.
C. Immunofluorescence showing ePNMA2 entry into Neuro2A cells with either no co-treatment or LAH4. Scale bar is 40 um.
D. Schematic of workflow for *in vitro* production of ePNMA2 capsids
E. Quantification of live (DAPI negative) GFP positive cells by flow cytometry following delivery by ePNMA2 of Cre mRNA to Neuro2A-*loxP-GFP* recipient cells.
F. White light (WL) and GFP fluorescence images of Neuro2A-*loxP-GFP* recipient cells 96 hours after ePNMA2-mediated delivery of Cre mRNA. Scale bar is 100 um.

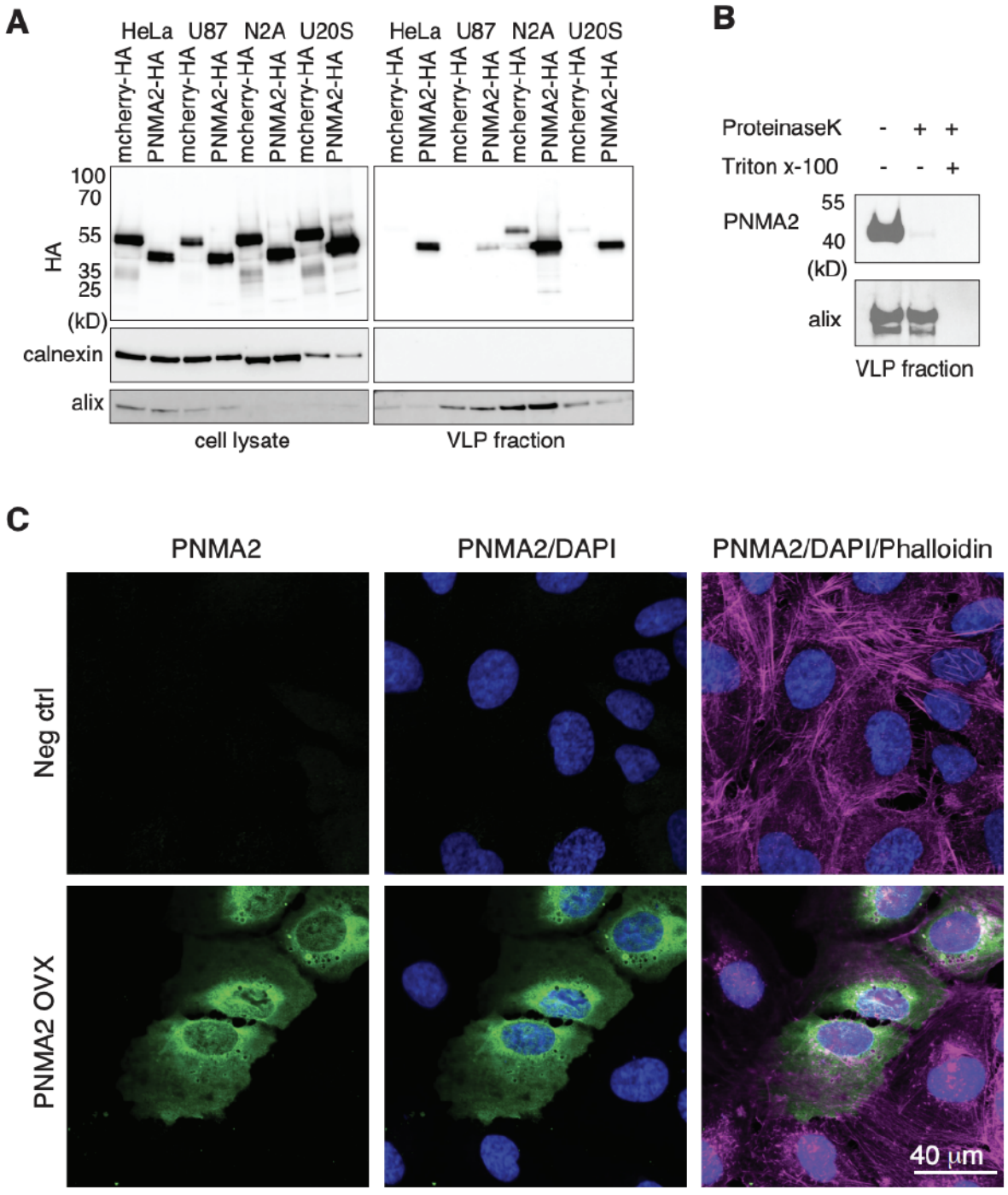
3.8 Supplementary Figure Captions

3.8.1 Figure S1. Pentamer assembly of human PNMA proteins using AlphaFold2 multimer.



3.8.2 Figure S2. Secretion of human PNMA2 from mammalian cell lines.

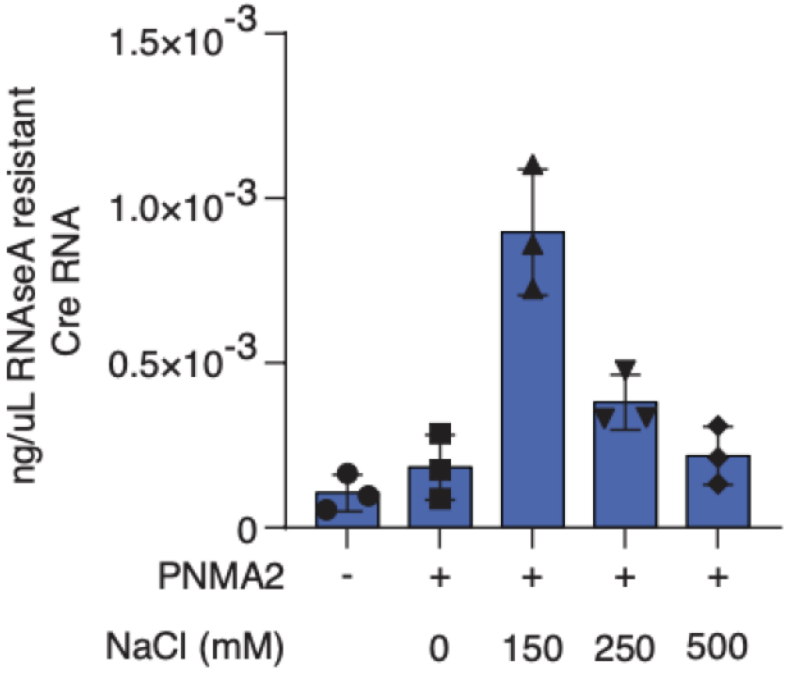
Western blot of cell lysate (left) and VLP fractions (right) from HeLa, U87, Neuro2A, and U20S cells transfected with either PNMA2-HA or mCherry-HA control plasmid.



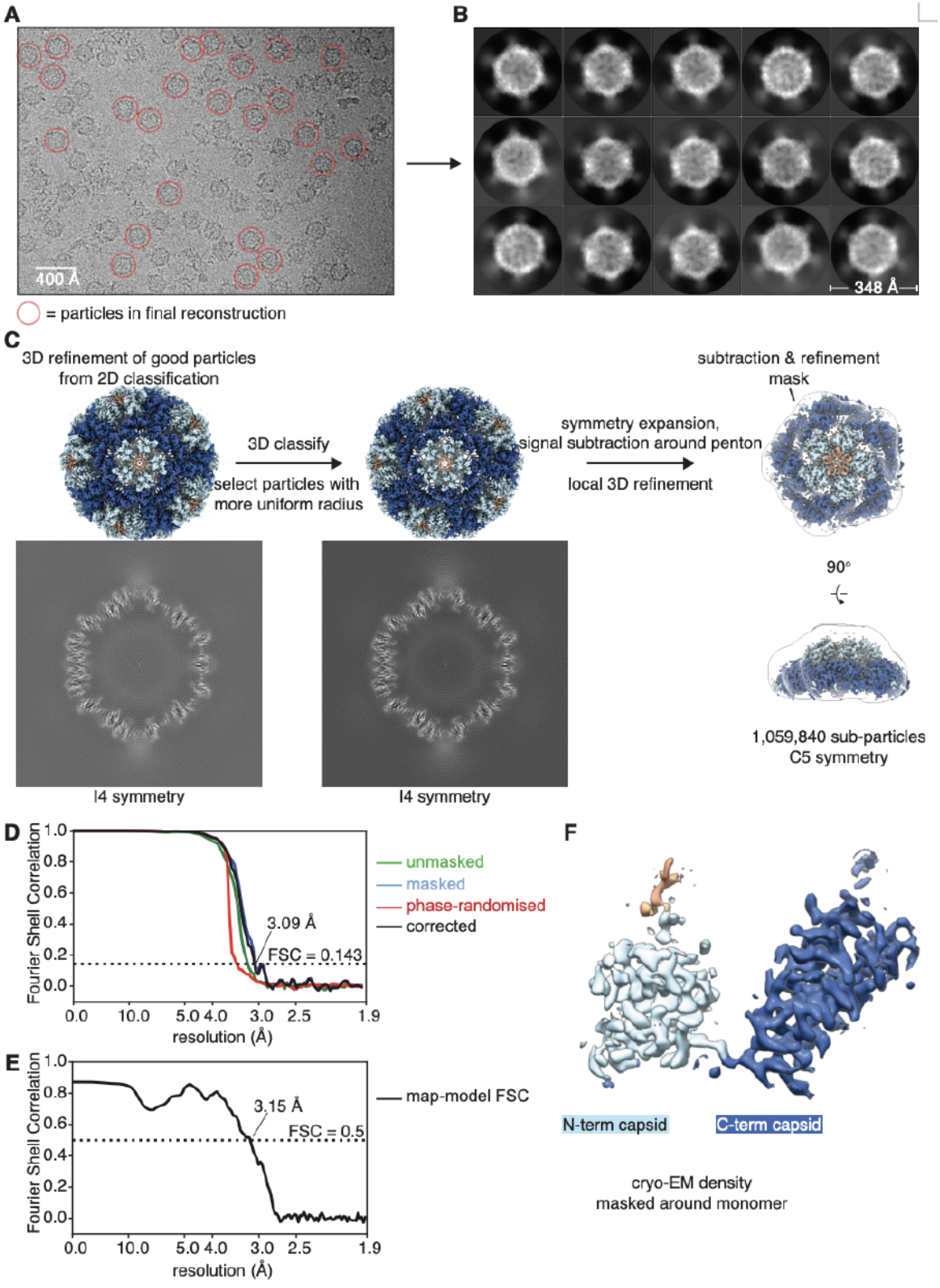
3.8.3 Figure S3. Intracellular localization of human PNMA2.

Immunofluorescence images of U20S cells transfected with a CMV-driven PNMA2 overexpression plasmid or negative control empty vector plasmid, where PNMA2 is labeled in green, DAPI in blue, and phalloidin in pink. Scale bar is 40 um.

PNMA2 VLP encapsidation of Cre RNA

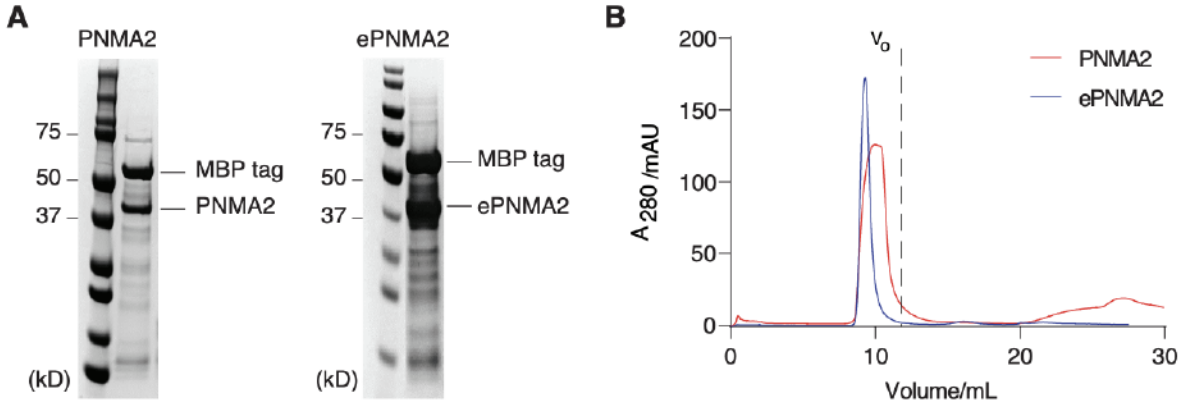


3.8.4 Figure S4. Structural analysis of recombinantly expressed human PNMA2.



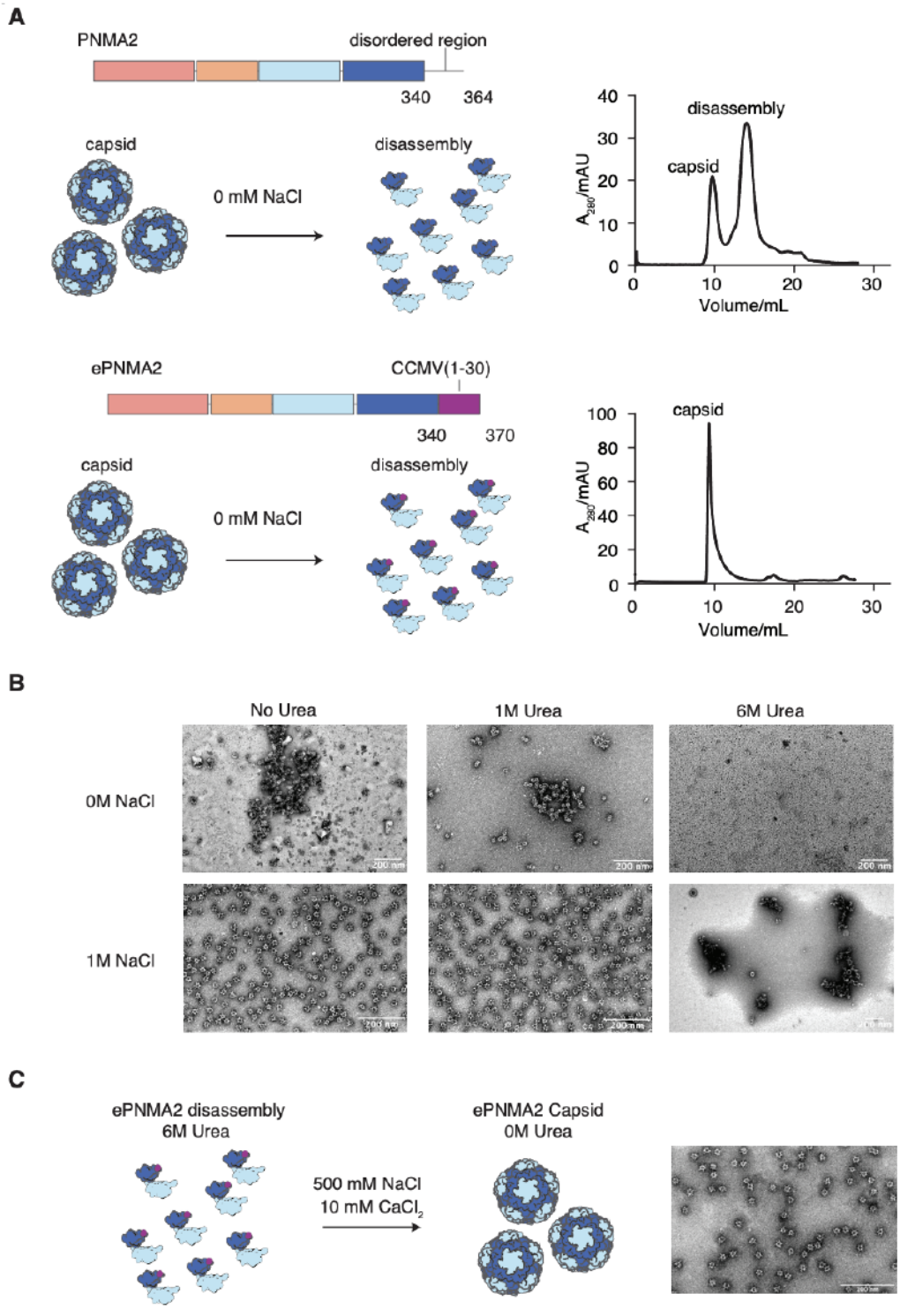
A. Example cryo-EM micrograph of PNMA2.
B. Representative 2D class averages.
C. Cryo-EM data processing workflow.
D. Gold-standard half-map Fourier Shell Correlation (FSC) curves.
E. FSC between the final map and the built model.
F. Cryo-EM density (C5 reconstruction) masked around a single PNMA2 monomer.

3.8.5 Figure S5. Optimization of ePNMA2 assembly and disassembly conditions.



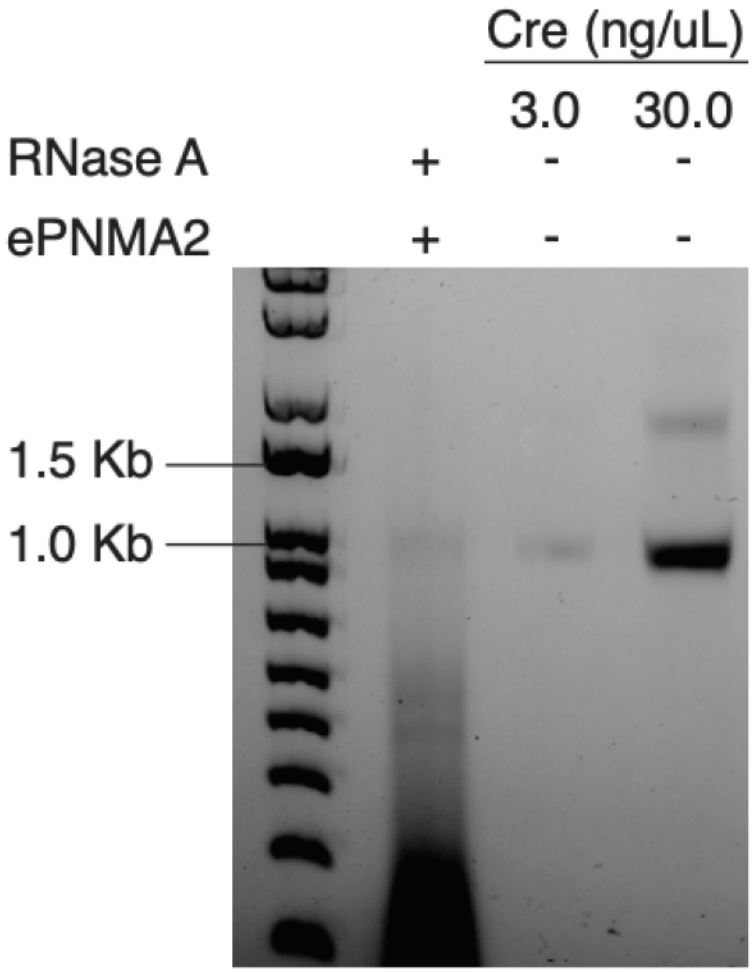
A. Schematic of disassembly conditions for the PNMA2 capsid (top) and resulting size exclusion chromatography trace, compared to treatment of ePNMA2 with the same disassembly conditions (bottom).
B. Transmission electron micrographs of ePNMA2 under different screening conditions for disassembly. Scale bars are 200 nm.
C. Schematic of ePNMA2 reassembly conditions and electron micrograph of resulting capsids. Scale bar is 200 nm.

3.8.6 Figure S6. ePNMA2(Cre) RNA encapsidation and protection from RNase.



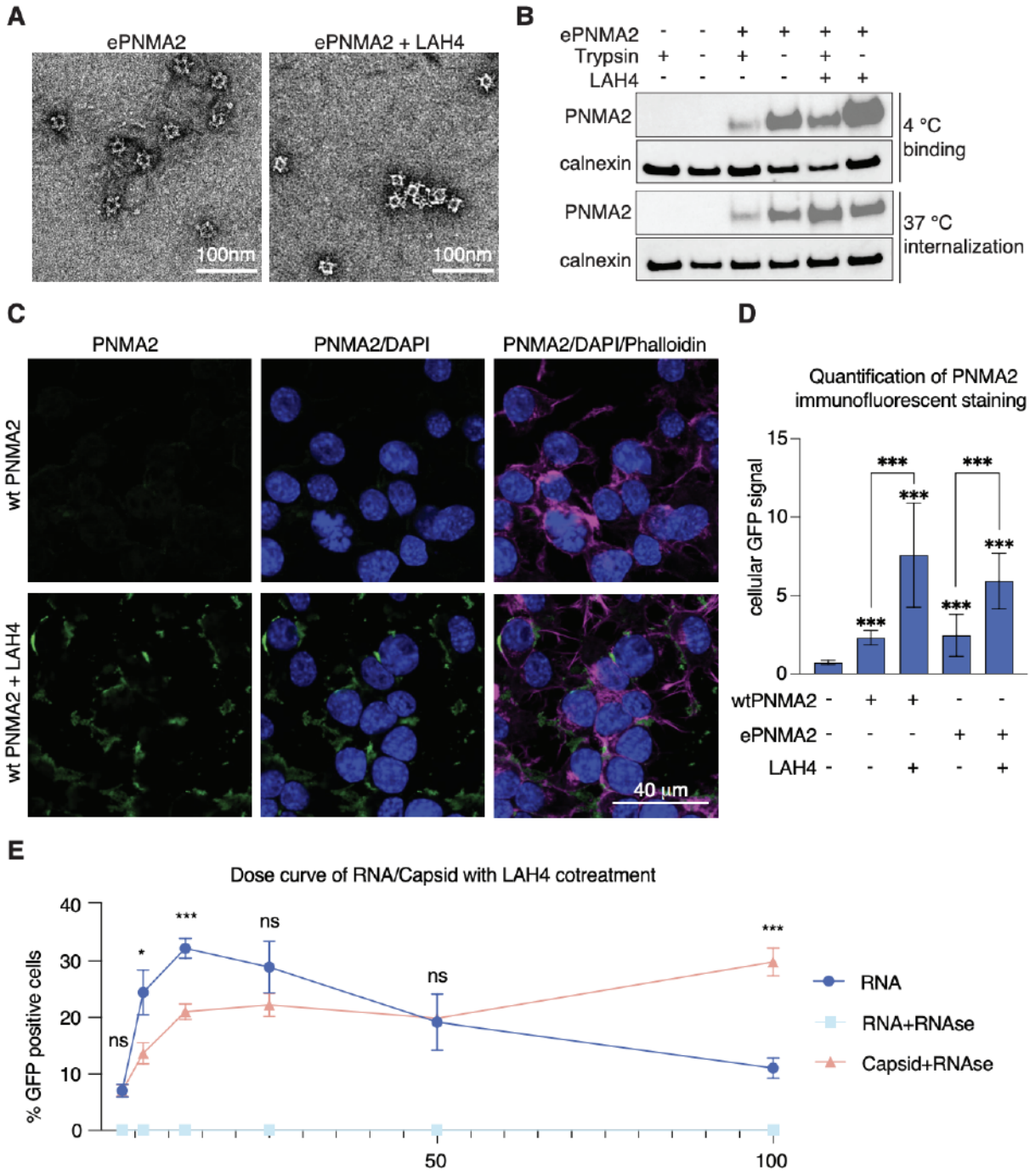
A. Quantification of full-length Cre mRNA by 1% agarose gel against Cre mRNA standard.

3.8.7 Figure S7. ePNMA2(Cre) RNA encapsidation and protection from RNAse



Quantification of full-length Cre mRNA by 1% agarose gel against Cre mRNA standard.

3.8.8 Figure S8. Transduction of Neuro2A cells treated with ePNMA2 with LAH4



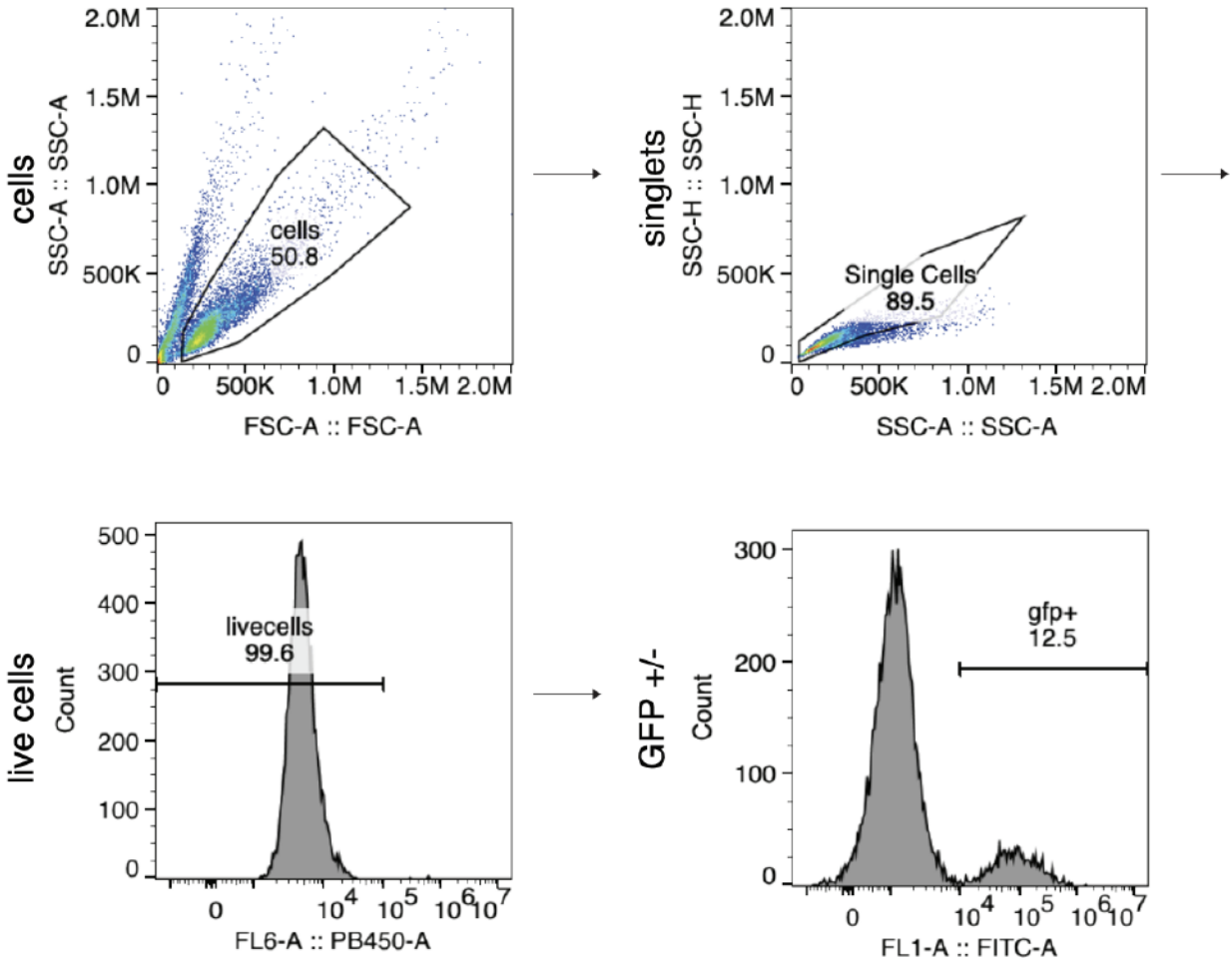
A. TEM micrographs of ePNMA2 capsids with or without LAH4 incorporation. Scale bar is 100 nm.

B. Western blot of naive Neuro2A cells treated with ePNMA2 capsids with and without LAH4 incorporation. Cells were pre chilled to 4 °C, then treated with ePNMA2 with or

without LAH4 for 1 hr at 4 °C. Trypsin digestion was used to disassociate bound but not internalized virions

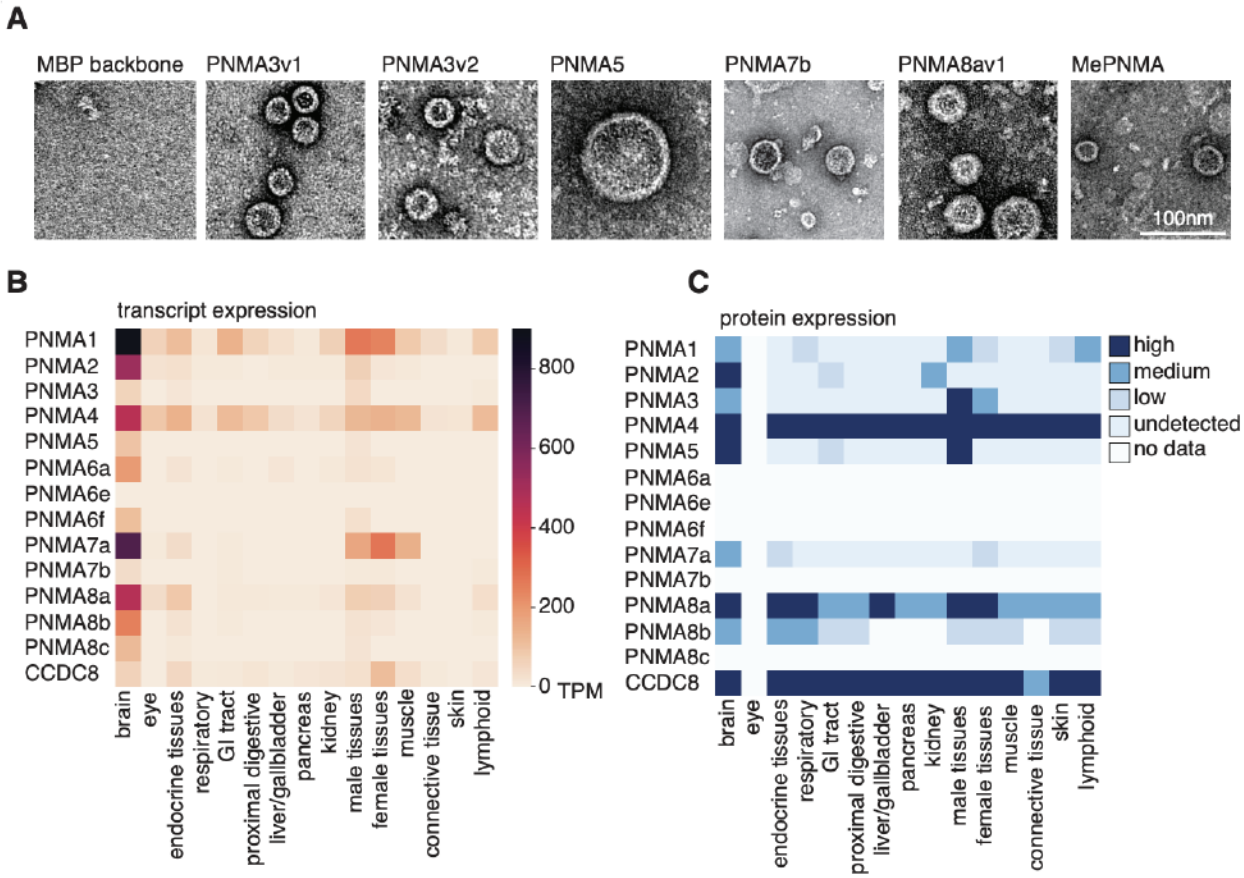
C. Immunofluorescence showing PNMA2 and ePNMA2 entry into Neuro2A cells with either no co-treatment or LAH4. Scale bar is 40 um.
D. Quantification of GFP signal from immunofluorescent staining of Neuro2A cells, n=30 cells. Significance of GFP signal relative to no capsid control is represented above columns, while brackets between columns compare untreated capsid to LAH4 treatedconjugated capsids. Samples were compared via unpaired t test where ns represents a $p > 0.05$, * represents $p \leq 0.05$, ** represents $p \leq 0.01$, and *** represents $p \leq 0.001$.
E. Flow cytometric quantification of GFP fluorescence of Neuro2A-*loxP*-GFP reporter cells transduced with various doses of RNAse treated ePNMA2(Cre) and an equal dose of Cre mRNA with and without RNAse treatment. Capsids and RNAs were treated with LAH4 before being added to Neuro2A cells. Error bars represent STD of technical triplicate, n=3, and statistical significance compares RNase treated capsid to RNA conditions, where ns represents a $p > 0.05$, * represents $p \leq 0.05$, ** represents $p \leq 0.01$, and *** represents $p \leq 0.001$.

3.8.9 Figure S9. Representative flow cytometry gating scheme for ePNMA2(Cre) gene transfer experiments.



Cells were initially gated on FSC and SSC to remove debris, then singlets were gated on SSC. Next, dead, DAPI positive cells were gated out (PB450-A). Finally, GFP positive cells were gated based on untreated controls.

3.8.9 Figure S10. I*n vitro* capsid formation and *in vivo* tissue expression of human PNMAs.



A. TEM micrographs of a mock purification (MBP backbone), human PNMA3v1, PNMA3v2, PNMA5, PNMA7b, and PNMA8av1, as well as MePNMA following purification from *E. coli*. Scale bar is 100 nm.
B. Expression of human *PNMA* transcripts across human tissues from Human Protein Atlas
C. Expression of human PNMA proteins across human tissues from Human Protein Atlas

4. Conclusions and Future Work

In this thesis, we have investigated the development of delivery vectors and cargos to evade immune detection. In **Chapter 2**, we underwent a thorough interrogation of the immunogenic epitopes in SaCas9 and AsCas12a. We established a workflow to identify and remove immunogenic T cell epitopes while preserving wild-type levels of activity and specificity to generate a suite of reduced immunogenicity (Redi) variants of two clinically relevant nucleases. To validate these Redi variants *in vivo*, we replicated the development of adaptive immunity to AAV8-SaCas9 by repetitive dosing of mice with AAV8-SaCas9-sgRNA at a 14-day interval. Consistent with previous studies, we observed substantial T cell immunity against SaCas9 wild-type epitopes, which was significantly reduced with the SaCas9.Redi.1 variant. Although it remains unclear how well immunized mouse models recapitulate pre-existing immunity in humans, we did also observe a reduced CD8+ T cell response to this variant from human PBMCs, suggesting the Redi variants may be safer for *in vivo* clinical use. Currently most *in vivo* investigational genome editing therapies are delivered using lipid nanoparticles or AAV vectors which are both immunogenic. With our engineered minimally immunogenic variants, we hope that delivery with less immunogenic delivery vehicles will lead to robust, long term clinical benefit for *in vivo* gene editing therapy.

In **Chapter 3**, we demonstrated that human PNMA2 is robustly secreted from cells as an icosahedral, non-enveloped capsid. We showed that although PNMA2 does not package RNA in human cells, an engineered variant with an RNA-binding domain grafted on to the C-terminus can package cargo RNA *in vitro*. Combining these self-assembled, packaged ePNMA2 capsids with the cell-penetrating peptide LAH4 led to efficient functional delivery of mRNA. Our demonstration of an all protein, *in vitro* produced delivery vehicle offers a starting point for further bioengineering. For example, increasing positive charges in the ePNMA2 capsid lumen could enhance RNA packaging efficiency. Further engineering of the ePNMA2 capsid surface residues may allow robust cell entry without LAH4, or targeted cell-type or tissue tropism. Engineering strategies applied to AAVs, which bear a similar T=1 icosahedral capsid structure, could be used to modify the ePNMA2 capsid surface with integrin binding motifs or nanobodies and thus modulate ePNMA2 tropism. The tropism and immunogenicity of these vectors *in vivo* merits further investigation. Finally, our work with PNMA2 may be extended to other PNMA family members, some of which also form capsids. This, together with the fact that many PNMA family members are most highly expressed in the central nervous system (CNS), raises the question of whether secreted PNMA capsids may enter specific recipient cells in the CNS. Further investigation of PNMA cell tropism and engineering of PNMA capsids may allow these vehicles to be harnessed for delivery of genetic cargoes to the brain, a long-standing goal in the delivery field.

4.1 Can we engineer a broadly non-immunogenic Cas9?

From the work in **Chapter 2**, future work should focus on engineering nuclease variants that are minimally immunogenic across multiple HLA types, extending our work here with the HLA-A*0201 haplotype. Development of these measures is essential for improving the safety

and efficacy of genome editing therapies in diverse clinical settings, especially if multiple such treatments might be performed in a patient's lifetime and across different patient populations. We hope that with further advancements in engineering safer genome engineering platforms, clinical applications of these systems can extend from treatment of genetic disorders to *in vivo* engineering cellular immunotherapies. The findings of this study offer valuable insights into reducing immunogenicity of genome editing tools through protein engineering and, with continued investigation, will facilitate the translation of these tools for long term clinical use.

4.2 What is the biological function of PNMA2?

PNMA2 is biologically associated with paraneoplastic neurological disorders, which occur when the body's immune system mistakenly attacks normal cells in the nervous system in response to cancer. PNMA2 has been identified as one of the antigens that is a target of the immune response. Researchers have demonstrated that the N-terminus of the PNMA2 protein, contained within the first 150 amino acids, serves as a spike protein and the majority of the antibodies in paraneoplastic syndrome patients are directed against this spike protein. The biological function of PNMA2 in both diseased and healthy individuals is largely unknown. If I were to speculate, my hypothesis would be that PNMA2's capsid formation serves as a mechanism for antigen presentation. In the same way increasing valency of antigens on vaccines promotes increased antibody production against that antigen, the presentation of the spike protein of PNMA2 on its N-terminus is presented 60 times on the surface of the PNMA2 capsid and vastly increases the likelihood of recognition by the immune system. We have also seen that PNMA2, if too dilute, does not assemble into capsids and therefore it may be that when there is increase in PNMA2 production, either due to cancer or as a response to a cancer elsewhere in the body, the immune system attempts to launch a response against it. However, since PNMA2 is a self-protein that forms capsids, the immune response is heightened and autoimmune in nature. When antibodies produced by the immune system specifically target PNMA2, they can cross-react with neurons in the nervous system, leading to a range of neurological symptoms. These symptoms can include muscle weakness, difficulty walking, loss of coordination, and other neurological impairments. Therefore, in disease settings, PNMA proteins may have immune-regulatory functions or even play a role in tumor suppression.

This naturally brings us to speculate about the biological function of the PNMA proteins in normal physiological contexts. While the exact function of PNMA2 in the nervous system and in normal physiological processes is still an area of ongoing research, we can look at their structural features to hypothesize about their function. For example, some PNMA proteins have a DNA-binding domain and may be involved in gene expression control. The N-terminus of all of the PNMA proteins contain a domain with structural similarity to RNA-binding motifs and point to the role of the PNMA proteins in post-transcriptional regulation of gene expression. Given that most of the PNMA proteins are found in neurons, we could hypothesize that they may have a role in neuronal development, maintenance, or function. Since some PNMAs are secreted from cells, they could be involved in processes such as axonal transport, synaptic function or neurotransmitter regulation. Intracellular signaling is essential for neuronal function and survival

and may be mediated by the PNMAs that are secreted from cells. More research needs to be done to definitively establish what the native role of PNMA is in normal physiological contexts.

4.3 Do there exist non-immunogenic capsid-forming proteins?

Capsid-forming proteins can increase valency, the number of antigenic epitopes displayed on the surface of a vaccine particle, and promote immunogenicity when used in vaccine design or as vaccine platforms. Capsid-like structures, such as virus-like particles (VLPs) or protein nanocages, are advantageous for vaccine development because they can enhance both valency and immunogenicity through multiple epitope presentation, and enhanced antigen presentation. The increased exposure leads to efficient antigen uptake and presentation to T cells, promoting a robust immune response. The repetitive antigen presentation can also lead to cross-linking of BCRs on B-cells and cross-linking multiple BCRs by multivalent antigens triggers strong B cell activation resulting in the production of high-affinity antibodies. Capsid-like structures can also persist in the body for longer than soluble antigens, providing sustained exposure to the immune system. Finally, multivalent antigen presentation on capsid-like structures can promote higher affinity antibodies which are then selectively expanded and lead to more potent and specific antibodies over time. This then begs the question, are we naturally always inviting an immune response when using any protein-capsid as a delivery system?

There are some capsid forming proteins that are not inherently immunogenic. Some cellular structures such as organelles, have capsid-like protein shells that are not recognized as foreign. For example the nuclear pore complex is a highly structured protein coat that resembles a capsid and shuttles molecules in and out of the cell’s nucleus. However many self-assembling protein nanoparticles and protein nanocages that have been made synthetically do trigger some mild form of immune response. Researchers have used ferritin proteins for their ability to form nanocage structures in research and medicine as it can encapsulate molecules without triggering a significant immune response. However the degree of immune response is still significant such that the most widespread use of ferritin has still been for vaccine applications.

Therefore there still exists a need for engineering truly minimally immunogenic protein capsids. From working with the PNMA proteins, we have learned that the amino acids near the end of the N-term, specifically amino acid residues 110-120 are somehow involved in assembly and secretion of the capsid. One interesting engineering angle could be if we could take inert proteins, engineer in amino acid sequences that are known to assist with assembly, and create synthetic capsid architectures. We could then engineer minimally immunogenic proteins, with adjustable valency and strength of interaction. Ideally capsids should have larger internal volume to maximize packaging capacity but require a low N number of monomers so as to reduce valency. Another approach that we have investigated within this thesis is to truncate the N-term of PNMA2 such that there is no antigen presented from the surface, thereby rendering the delivery vehicle less immunogenic. While PNMA2 capsids are not inherently minimally immunogenic, there are many lessons we can take from PNMA2 capsid assembly to inform the design and development of less immunogenic delivery strategies.

4.4 References

1. 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).
2. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
3. Zetsche, B. *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).
4. Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186–191 (2015).
5. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).
6. Frangoul, H. *et al.* CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. *N. Engl. J. Med.* **384**, 252–260 (2021).
7. Ahmad, I. CRISPR/Cas9-A Promising Therapeutic Tool to Cure Blindness: Current Scenario and Future Prospects. *Int. J. Mol. Sci.* **23**, (2022).
8. Shahzad, F. *et al.* CRISPR/Cas9 gene editing: A new hope for transthyretin amyloidosis treatment. *Ann Med Surg (Lond)* **83**, 104784 (2022).
9. Seitzer, J. NTLA-2002: CRISPR/Cas9-mediated gene knockout of KLKB1 to treat hereditary angioedema. *J. Allergy Clin. Immunol.* **147**, AB147 (2021).
10. Ensinck, M., Mottais, A., Detry, C., Leal, T. & Carlon, M. S. On the Corner of Models and Cure: Gene Editing in Cystic Fibrosis. *Front. Pharmacol.* **12**, 662110 (2021).
11. Long, C. *et al.* Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA. *Science* **345**, 1184–1188 (2014).
12. Tabebordbar, M. *et al.* In vivo gene editing in dystrophic mouse muscle and muscle stem cells. *Science* **351**, 407–411 (2016).

13. Yu, K.-R., Natanson, H. & Dunbar, C. E. Gene Editing of Human Hematopoietic Stem and Progenitor Cells: Promise and Potential Hurdles. *Hum. Gene Ther.* **27**, 729–740 (2016).
14. De Dreuzy, E. *et al.* EDIT-301: An Experimental Autologous Cell Therapy Comprising Cas12a-RNP Modified mPB-CD34+ Cells for the Potential Treatment of SCD. *Blood* **134**, 4636 (2019).
15. Maeder, M. L. *et al.* Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. *Nat. Med.* **25**, 229–233 (2019).
16. Chew, W. L. Immunity to CRISPR Cas9 and Cas12a therapeutics. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **10**, (2018).
17. Mehta, A. & Merkel, O. M. Immunogenicity of Cas9 Protein. *J. Pharm. Sci.* **109**, 62–67 (2020).
18. Ren, J. *et al.* Multiplex Genome Editing to Generate Universal CAR T Cells Resistant to PD1 Inhibition. *Clin. Cancer Res.* **23**, 2255–2266 (2017).
19. Lowy, F. D. Staphylococcus aureus infections. *N. Engl. J. Med.* **339**, 520–532 (1998).
20. Shaikh, N., Leonard, E. & Martin, J. M. Prevalence of streptococcal pharyngitis and streptococcal carriage in children: a meta-analysis. *Pediatrics* **126**, e557–64 (2010).
21. Kolata, J. B. *et al.* The Fall of a Dogma? Unexpected High T-Cell Memory Response to Staphylococcus aureus in Humans. *J. Infect. Dis.* **212**, 830–838 (2015).
22. Mortensen, R. *et al.* Adaptive Immunity against Streptococcus pyogenes in Adults Involves Increased IFN-γ and IgG3 Responses Compared with Children. *J. Immunol.* **195**, 1657–1664 (2015).
23. Wang, D. *et al.* Adenovirus-Mediated Somatic Genome Editing of Pten by CRISPR/Cas9 in Mouse Liver in Spite of Cas9-Specific Immune Responses. *Hum. Gene Ther.* **26**, 432–442 (2015).
24. Ewaisha, R. & Anderson, K. S. Immunogenicity of CRISPR therapeutics-Critical considerations for clinical translation. *Front Bioeng Biotechnol* **11**, 1138596 (2023).

25. Stadtmauer, E. A. *et al.* CRISPR-engineered T cells in patients with refractory cancer. *Science* **367**, (2020).
26. Charlesworth, C. T. *et al.* Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nat. Med.* **25**, 249–254 (2019).
27. Wagner, D. L., Peter, L. & Schmueck-Henneresse, M. Cas9-directed immune tolerance in humans-a model to evaluate regulatory T cells in gene therapy? *Gene Ther.* **28**, 549–559 (2021).
28. 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).
29. Yin, H. *et al.* Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system components in vivo. *Nat. Biotechnol.* **34**, 328–333 (2016).
30. Mingozzi, F. & High, K. A. Immune responses to AAV vectors: overcoming barriers to successful gene therapy. *Blood* **122**, 23–36 (2013).
31. Warashina, S. *et al.* A lipid nanoparticle for the efficient delivery of siRNA to dendritic cells. *J. Control. Release* **225**, 183–191 (2016).
32. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
33. Li, C. & Samulski, R. J. Engineering adeno-associated virus vectors for gene therapy. *Nat. Rev. Genet.* **21**, 255–272 (2020).
34. Oldham, R. J. *et al.* FcγRII (CD32) modulates antibody clearance in NOD SCID mice leading to impaired antibody-mediated tumor cell deletion. *J Immunother Cancer* **8**, (2020).
35. Mays, L. E. & Wilson, J. M. The complex and evolving story of T cell activation to AAV vector-encoded transgene products. *Mol. Ther.* **19**, 16–27 (2011).
36. Albert, M. L. *et al.* Immature dendritic cells phagocytose apoptotic cells via alphavbeta5 and CD36, and cross-present antigens to cytotoxic T lymphocytes. *J. Exp. Med.* **188**, 1359–1368 (1998).

37. Iwasaki, A. & Medzhitov, R. Regulation of adaptive immunity by the innate immune system. *Science* **327**, 291–295 (2010).
38. Mingozzi, F. *et al.* CD8(+) T-cell responses to adeno-associated virus capsid in humans. *Nat. Med.* **13**, 419–422 (2007).
39. Amrani, S. & Tabrizian, M. Characterization of Nanoscale Loaded Liposomes Produced by 2D Hydrodynamic Flow Focusing. *ACS Biomater Sci Eng* **4**, 502–513 (2018).
40. Pardi, N. *et al.* Expression kinetics of nucleoside-modified mRNA delivered in lipid nanoparticles to mice by various routes. *J. Control. Release* **217**, 345–351 (2015).
41. Modzelewski, A. J., Gan Chong, J., Wang, T. & He, L. Mammalian genome innovation through transposon domestication. *Nat. Cell Biol.* **24**, 1332–1340 (2022).
42. Pastuzyn, E. D. *et al.* The Neuronal Gene Arc Encodes a Repurposed Retrotransposon Gag Protein that Mediates Intercellular RNA Transfer. *Cell* **172**, 275–288.e18 (2018).
43. Abed, M. *et al.* The Gag protein PEG10 binds to RNA and regulates trophoblast stem cell lineage specification. *PLoS One* **14**, e0214110 (2019).
44. Segel, M. *et al.* Mammalian retrovirus-like protein PEG10 packages its own mRNA and can be pseudotyped for mRNA delivery. *Science* **373**, 882–889 (2021).
45. Iwasaki, S. *et al.* Identification of a novel PNMA-MS1 gene in marsupials suggests the LTR retrotransposon-derived PNMA genes evolved differently in marsupials and eutherians. *DNA Res.* **20**, 425–436 (2013).
46. Voltz, R. *et al.* A serologic marker of paraneoplastic limbic and brain-stem encephalitis in patients with testicular cancer. *N. Engl. J. Med.* **340**, 1788–1795 (1999).
47. Dalmau, J. *et al.* Ma1, a novel neuron- and testis-specific protein, is recognized by the serum of patients with paraneoplastic neurological disorders. *Brain* **122 (Pt 1)**, 27–39 (1999).
48. Rosenfeld, M. R., Eichen, J. G., Wade, D. F., Posner, J. B. & Dalmau, J. Molecular and clinical diversity in paraneoplastic immunity to Ma proteins. *Ann. Neurol.* **50**, 339–348

(2001).

49. Schüller, M., Jenne, D. & Voltz, R. The human PNMA family: novel neuronal proteins implicated in paraneoplastic neurological disease. *J. Neuroimmunol.* **169**, 172–176 (2005).
50. Campillos, M., Doerks, T., Shah, P. K. & Bork, P. Computational characterization of multiple Gag-like human proteins. *Trends Genet.* **22**, 585–589 (2006).
51. Lee, Y. H., Pang, S. W. & Tan, K. O. PNMA2 mediates heterodimeric interactions and antagonizes chemo-sensitizing activities mediated by members of PNMA family. *Biochem. Biophys. Res. Commun.* **473**, 224–229 (2016).
52. Zhang, X.-L. *et al.* Pnma5 is essential to the progression of meiosis in mouse oocytes through a chain of phosphorylation. *Oncotarget* **8**, 96809–96825 (2017).
53. Chen, H.-L. & D'Mello, S. R. Induction of neuronal cell death by paraneoplastic Ma1 antigen. *J. Neurosci. Res.* **88**, 3508–3519 (2010).
54. Pang, S. W., Lahiri, C., Poh, C. L. & Tan, K. O. PNMA family: Protein interaction network and cell signalling pathways implicated in cancer and apoptosis. *Cell. Signal.* **45**, 54–62 (2018).
55. Dodonova, S. O., Prinz, S., Bilanchone, V., Sandmeyer, S. & Briggs, J. A. G. Structure of the Ty3/Gypsy retrotransposon capsid and the evolution of retroviruses. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 10048–10057 (2019).
56. D'Souza, V. & Summers, M. F. How retroviruses select their genomes. *Nat. Rev. Microbiol.* **3**, 643–655 (2005).
57. Orlinsky, K. J. & Sandmeyer, S. B. The Cys-His motif of Ty3 NC can be contributed by Gag3 or Gag3-Pol3 polyproteins. *J. Virol.* **68**, 4152–4166 (1994).
58. Larsen, L. S. Z. *et al.* Ty3 nucleocapsid controls localization of particle assembly. *J. Virol.* **82**, 2501–2514 (2008).
59. Sandmeyer, S., Patterson, K. & Bilanchone, V. Ty3, a Position-specific Retrotransposon in Budding Yeast. *Microbiol Spectr* **3**, MDNA3–0057–2014 (2015).

60. Cléry, A., Blatter, M. & Allain, F. H.-T. RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.* **18**, 290–298 (2008).
61. Takaji, M., Komatsu, Y., Watakabe, A., Hashikawa, T. & Yamamori, T. Paraneoplastic antigen-like 5 gene (PNMA5) is preferentially expressed in the association areas in a primate specific manner. *Cereb. Cortex* **19**, 2865–2879 (2009).
62. Albert, M. L. & Darnell, R. B. Paraneoplastic neurological degenerations: keys to tumour immunity. *Nat. Rev. Cancer* **4**, 36–44 (2004).
63. Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J. & Schreiber, R. D. Cancer immunoediting: from immunosurveillance to tumor escape. *Nat. Immunol.* **3**, 991–998 (2002).
64. Darnell, R. B. & DeAngelis, L. M. Regression of small-cell lung carcinoma in patients with paraneoplastic neuronal antibodies. *Lancet* **341**, 21–22 (1993).
65. Ono, R. *et al.* Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. *Nat. Genet.* **38**, 101–106 (2006).
66. Ashley, J. *et al.* Retrovirus-like Gag Protein Arc1 Binds RNA and Traffics across Synaptic Boutons. *Cell* **172**, 262–274.e11 (2018).
67. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
68. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
69. Wang, D., Tai, P. W. L. & Gao, G. Adeno-associated virus vector as a platform for gene therapy delivery. *Nat. Rev. Drug Discov.* **18**, 358–378 (2019).
70. Colque-Navarro Patricia, Jacobsson Gunnar, Andersson Rune, Flock Jan-Ingmar & Möllby Roland. Levels of Antibody against 11 Staphylococcus aureus Antigens in a Healthy Population. *Clin. Vaccine Immunol.* **17**, 1117–1123 (2010).
71. Chew, W. L. *et al.* A multifunctional AAV-CRISPR-Cas9 and its host response. *Nat. Methods* **13**, 868–874 (2016).

72. Simhadri, V. L. *et al.* Cas9-derived peptides presented by MHC Class II that elicit proliferation of CD4+ T-cells. *Nat. Commun.* **12**, 5090 (2021).

73. Crudele, J. M. & Chamberlain, J. S. Cas9 immunity creates challenges for CRISPR gene editing therapies. *Nat. Commun.* **9**, 3497 (2018).

74. Hakim, C. H. *et al.* Cas9-specific immune responses compromise local and systemic AAV CRISPR therapy in multiple dystrophic canine models. *Nat. Commun.* **12**, 6769 (2021).

75. Nelson, C. E. *et al.* In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science* **351**, 403–407 (2016).

76. Suzuki, K. *et al.* In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540**, 144–149 (2016).

77. Wang, J. Y. & Doudna, J. A. CRISPR technology: A decade of genome editing is only the beginning. *Science* **379**, eadd8643 (2023).

78. Adams, S. *et al.* HLA class I and II genotype of the NCI-60 cell lines. *J. Transl. Med.* **3**, 11 (2005).

79. Rasmussen, M. *et al.* Uncovering the peptide-binding specificities of HLA-C: a general strategy to determine the specificity of any MHC class I molecule. *J. Immunol.* **193**, 4790–4802 (2014).

80. Reynisson, B., Alvarez, B., Paul, S., Peters, B. & Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* **48**, W449–W454 (2020).

81. Schmid-Burgk, J. L. *et al.* Highly Parallel Profiling of Cas9 Variant Specificity. *Mol. Cell* **78**, 794–800.e8 (2020).

82. Pajot, A. *et al.* A mouse model of human adaptive immune functions: HLA-A2.1-/HLA-DR1-transgenic H-2 class I-/class II-knockout mice. *Eur. J. Immunol.* **34**, 3060–3069 (2004).

83. Ott, L. W. *et al.* Tumor Necrosis Factor-alpha- and interleukin-1-induced cellular responses: coupling proteomic and genomic information. *J. Proteome Res.* **6**, 2176–2185 (2007).

84. Katsura, H., Kobayashi, Y., Tata, P. R. & Hogan, B. L. M. IL-1 and TNFα Contribute to the Inflammatory Niche to Enhance Alveolar Regeneration. *Stem Cell Reports* **12**, 657–666 (2019).

85. Morales, R. A. & Allende, M. L. Peripheral Macrophages Promote Tissue Regeneration in Zebrafish by Fine-Tuning the Inflammatory Response. *Front. Immunol.* **10**, 253 (2019).

86. Li, A. *et al.* AAV-CRISPR Gene Editing Is Negated by Pre-existing Immunity to Cas9. *Mol. Ther.* **28**, 1432–1441 (2020).

87. Verdera, H. C., Kuranda, K. & Mingozzi, F. AAV Vector Immunogenicity in Humans: A Long Journey to Successful Gene Transfer. *Mol. Ther.* **28**, 723–746 (2020).

88. Huysmans, H. *et al.* Expression Kinetics and Innate Immune Response after Electroporation and LNP-Mediated Delivery of a Self-Amplifying mRNA in the Skin. *Mol. Ther. Nucleic Acids* **17**, 867–878 (2019).

89. Ju, Y. *et al.* Impact of anti-PEG antibodies induced by SARS-CoV-2 mRNA vaccines. *Nat. Rev. Immunol.* **23**, 135–136 (2023).

90. Roth, T. L. *et al.* Pooled Knockin Targeting for Genome Engineering of Cellular Immunotherapies. *Cell* **181**, 728–744.e21 (2020).

91. King, C. A. & Bradley, P. Structure-based prediction of protein-peptide specificity in Rosetta. *Proteins* **78**, 3437–3449 (2010).

92. Leman, J. K. *et al.* Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat. Methods* **17**, 665–680 (2020).

93. Alford, R. F. *et al.* The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).

94. King, C. *et al.* Removing T-cell epitopes with computational protein design. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8577–8582 (2014).

95. Yachnin, B. J., Mulligan, V. K., Khare, S. D. & Bailey-Kellogg, C. MHCEpitopeEnergy, a Flexible Rosetta-Based Biotherapeutic Deimmunization Platform. *J. Chem. Inf. Model.* **61**, 2368–2382 (2021).

96. Vita, R. *et al.* The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* **47**, D339–D343 (2019).

97. Jurtz, V. *et al.* NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J. Immunol.* **199**, 3360–3368 (2017).

98. Clement, K. *et al.* CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).

99. Grieger, J. C., Choi, V. W. & Samulski, R. J. Production and characterization of adeno-associated viral vectors. *Nat. Protoc.* **1**, 1412–1428 (2006).

100. Schmid-Burgk, J. L. & Hornung, V. BrowserGenome.org: web-based RNA-seq data analysis and visualization. *Nat. Methods* **12**, 1001 (2015).

101. Kilian, M. *et al.* T-cell Receptor Therapy Targeting Mutant Capicua Transcriptional Repressor in Experimental Gliomas. *Clin. Cancer Res.* **28**, 378–389 (2022).

102. Kulkarni, J. A. *et al.* The current landscape of nucleic acid therapeutics. *Nat. Nanotechnol.* **16**, 630–643 (2021).

103. Lee, Y. H., Pang, S. W., Poh, C. L. & Tan, K. O. Distinct functional domains of PNMA5 mediate protein-protein interaction, nuclear localization, and apoptosis signaling in human cancer cells. *J. Cancer Res. Clin. Oncol.* **142**, 1967–1977 (2016).

104. Baksh, S. *et al.* The tumor suppressor RASSF1A and MAP-1 link death receptor signaling to Bax conformational change and cell death. *Mol. Cell* **18**, 637–650 (2005).

105. Tan, K. O. *et al.* MAP-1 is a mitochondrial effector of Bax. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 14623–14628 (2005).

106. Tan, K. O. *et al.* MAP-1, a novel proapoptotic protein containing a BH3-like motif that

associates with Bax through its Bcl-2 homology domains. *J. Biol. Chem.* **276**, 2802–2807 (2001).

107. Xu, J. *et al.* PNMA2 forms non-enveloped virus-like capsids that trigger paraneoplastic neurological syndrome. *bioRxiv* (2023) doi:10.1101/2023.02.09.527862.
108. Chalopin, D., Galiana, D. & Volff, J.-N. Genetic innovation in vertebrates: gypsy integrase genes and other genes derived from transposable elements. *Int. J. Evol. Biol.* **2012**, 724519 (2012).
109. Bayev, A. A., Jr *et al.* Structural organization of transposable element mdg4 from Drosophila melanogaster and a nucleotide sequence of its long terminal repeats. *Nucleic Acids Res.* **12**, 3707–3723 (1984).
110. Xie, Q. *et al.* The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 10405–10410 (2002).
111. Garmann, R. F., Knobler, C. M. & Gelbart, W. M. Protocol for Efficient Cell-Free Synthesis of Cowpea Chlorotic Mottle Virus-Like Particles Containing Heterologous RNAs. *Methods Mol. Biol.* **1776**, 249–265 (2018).
112. Comas-Garcia, M., Cadena-Nava, R. D., Rao, A. L. N., Knobler, C. M. & Gelbart, W. M. In vitro quantification of the relative packaging efficiencies of single-stranded RNA molecules by viral capsid protein. *J. Virol.* **86**, 12271–12282 (2012).
113. Kichler, A., Mason, A. J. & Bechinger, B. Cationic amphipathic histidine-rich peptides for gene delivery. *Biochim. Biophys. Acta* **1758**, 301–307 (2006).
114. Tsui, N. B. Y., Ng, E. K. O. & Lo, Y. M. D. Stability of endogenous and added RNA in blood specimens, serum, and plasma. *Clin. Chem.* **48**, 1647–1653 (2002).
115. Roberts, T. C., Langer, R. & Wood, M. J. A. Advances in oligonucleotide drug delivery. *Nat. Rev. Drug Discov.* **19**, 673–694 (2020).
116. Crespo-Escobar, P. *et al.* The role of gluten consumption at an early age in celiac disease development: a further analysis of the prospective PreventCD cohort study. *Am. J. Clin.*

Nutr. **105**, 890–896 (2017).

117. Shi, W. & Bartlett, J. S. RGD inclusion in VP3 provides adeno-associated virus type 2 (AAV2)-based vectors with a heparan sulfate-independent cell entry mechanism. *Molecular therapy: the journal of the American Society of Gene Therapy* vol. 7 515–525 (2003).
118. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
119. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
120. Faure, G. & Callebaut, I. Comprehensive repertoire of foldable regions within whole genomes. *PLoS Comput. Biol.* **9**, e1003280 (2013).
121. Callebaut, I. *et al.* Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cell. Mol. Life Sci.* **53**, 621–645 (1997).
122. Holm, L. Dali server: structural unification of protein families. *Nucleic Acids Res.* **50**, W210–W215 (2022).
123. Holm, L. Using Dali for Protein Structure Comparison. *Methods Mol. Biol.* **2112**, 29–42 (2020).
124. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
125. Gabler, F. *et al.* Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Curr. Protoc. Bioinformatics* **72**, e108 (2020).
126. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
127. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 2021.10.04.463034 (2022) doi:10.1101/2021.10.04.463034.
128. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

129. Babraham bioinformatics - FastQC A quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

130. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

131. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

132. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

133. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

134. Kimanius, D., Dong, L., Sharov, G., Nakane, T. & Scheres, S. H. W. New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochem. J* **478**, 4169–4185 (2021).

135. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

136. Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019).

137. Casañal, A., Lohkamp, B. & Emsley, P. Current developments in Coot for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. *Protein Sci.* **29**, 1069–1078 (2020).

138. Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* **74**, 519–530 (2018).

139. Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531–544 (2018).

140. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).