UNIVERSITY OF CALIFORNIA

Los Angeles

Advancing Gene Therapy Technologies for the Treatment of Genetic Disorders

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Molecular and Medical Pharmacology

by

Paul George Ayoub

2023

ABSTRACT OF THE DISSERTATION

Developing Gene Therapies and Tools for the Treatment of Genetic Disorders

by

Paul George Ayoub

Doctor of Philosophy in Molecular and Medical Pharmacology

University of California, Los Angeles, 2023

Professor Donald Barry Kohn, Chair

This work is a culmination of the development of various gene therapy technologies for the treatment of genetic disorders. Chapter 1 highlights the generation of a modified S glycoprotein of SARS-Cov-2 to pseudotype lentiviral vectors which efficiently transduced ACE2-expressing cells with high specificity and contain minimal off-target transduction of ACE2 negative cells. Chapter 2 illustrates the generation of a series of lentiviral vectors using regulatory elements from the *SH2D1A* locus. By iterative process, we developed two candidate vectors that express SAP protein in the native pattern at levels similar to those of the endogenous gene. These vectors may support reconstitution of the immunologic deficiencies of XLP1 in a safe manner. Chapter 3 delineates the optimization of a gene editing approach for site-specific insertion of a *CFTR* cDNA cassette for the treatment of Cystic Fibrosis. Utilizing various techniques practiced throughout the laboratory, we designed various improvements to gene editing cargo to increase the cutting efficiency and integration of gene editing cargo including the inclusion of electroporation enhancers, double stranded DNA donor chemical modifications and small molecule inhibitors. This work is an ongoing effort, in collaboration with colleagues in the labs of Dr. Steven Jonas and Dr. Brigitte Gomperts to generate a gene editing strategy for the treatment of Cystic Fibrosis.

The dissertation of Paul George Ayoub is approved.

Gay M. Crooks

Yvonne Y. Chen

Ting-Ting Wu

Donald Barry Kohn, Committee Chair

University of California Los Angeles

2023

DEDICATION

This work is dedicated to all those who have helped guide me throughout my career and doctoral program:

- To my friends and colleagues, thank you for providing endless laughter, fun, spontaneity, and excitement through this all. I am an extension of each one of you and I can't wait what's in store for us next.
- To my family, I am standing on the shoulders of giants. This work is dedicated to all the hard work you've done to help me get to this point. Thank you for the support when I was down, the encouragement when I was lost, and the celebration when I was thriving. I know you always have my back and I appreciate all the guidance you've given me along the way.
- To the gene therapy and life science community, for their patience, trust, and contribution towards advancing our research.

TABLE OF CONTENTS

Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

ACKNOWLEDGMENTS

There is no quantity of words that can express how grateful I am for you, Dr. Donald Kohn. Over the past 8 years – through interning as a summer student, working as a technician, and completing my graduate degree – you have been nothing short of amazing. You have provided unparalleled mentorship as I evolved from my training bike. A walking encyclopedia, a pop culture savant, a laker fanatic, and a low-profile comedian; Don, you consistently find ways to impress me with your breadth of expertise. It goes without saying that I learned valuable lessons from you within the life sciences, but more importantly, I continue to learn life lessons from you, emulating your heart of gold and your humility. Thank you from the bottom of my heart for paving this path and supporting me across it; you provide meaning and excitement for researchers like me to participate in and develop research in gene therapy.

Hey Rodge. I genuinely could not have done this without you, Dr. Roger Hollis – seriously. Some of the biggest laughs, deepest conversations, and most meaningful lessons were had with you in Bay 5. You have and will continue to be a best friend and a mentor; you have taught me to learn and grow while enjoying life, and I couldn't thank you more for it. I am going to miss our daily breakfast chats, caffeine surpluses, and candy rushes. I will certainly miss making fun of you and then laughing from my belly when you jab back. Across my 8 years, the Kohn Lab continued to evolve as technicians and students came and went, but nothing made me happier than knowing you stayed through it all. Thank you for going above and beyond to make my experience here one I'll never forget.

Thank you to my committee, Dr. Ting-Ting Wu, Dr. Yvonne Chen, and Dr. Gay Crooks. Never in my life would I have imagined learning from such a powerhouse of

individuals. You are all exceptionally brilliant, determined, and kind. You say it like it is, but what you say is filled with wisdom and encouragement. I look up to each one of you and I can't thank you enough for your assistance and direction.

Thank you to everyone in the Kohn Lab, staff, graduates, undergraduates, post docs, and more; what a tremendous group of people to be a part of. The Kohn Lab is truly one of kind. Thank you for the friendship, laughter, and support throughout my time here. I truly feel like I have made a family. You've helped me through thick and thin and given me endless memories to look back on together. To the graduate students, you guys are the best humans. So smart, so funny, and so kind. I couldn't imagine a world without you all in it. A family of hooligans; I can't wait for what's next for us all and I am excited for the celebrations. I will miss working with you all more than you know.

To LL, CJ, and JR: my knights in shining armor. I can't describe how thankful I am for the three of you. You have not only helped me conduct most of my PhD research, but you have kept me on my toes to learn and remain engaged through it all. You three are incredibly brilliant humans and en route to become incredible researchers/physicians. I am excited to see what each of you has in store. LL, I wanted to give you a special thank you for joining me on day 1 of my journey. You truly were there for all the ups and the downs, and I felt as though we both were learning together. Thank you for putting up with me and being the best boost to help me reach my goal.

Thank you to my collaborators for their guidance and comradery, from when all is going poorly to when we can't miss at all. I wanted to give a special shoutout to those that helped with the Cystic Fibrosis work. The jolliest of groups, with knowledge and humility that I could only wish to have. You three have been such a large part of my PhD and

incredible pillars of support along the way. Thank you dearly for making me feel welcome and wanted in our journey of developing molecular therapies together.

To the UCLA BSCRC Flow Cytometry Core, the UCLA/CFAR Virology Core Laboratory, and the UCLA Technology Center for Genomics & Bioinformatics (TCGB), thank you for your continued support and assistance in all my work at UCLA. I learned so much from each one of you, and I couldn't have done this without you. Thank you as well to my funding sources, the California Institute for Regenerative Medicine (CIRM) and UCLA's Broad Stem Cell Research Center.

Thank you to all my friends and family, nobody gets me like you all do. You all bring excitement to life outside the lab. You have shown me more love and encouragement than I could ever dream of. You help me get back up when spirits are low, and you keep me up top when spirits are high. Love you all.

To the most supportive family in the game. At this point, I am unsure if I am or you all are more excited for me receiving my doctorate. But that's all understandable given the effort each of you put into me getting here. It's not every day my family is just as engaged in my PhD research and journey as I am. I am fortunate beyond belief. I said it before, and I'll say it again: I stand on the shoulders of giants. My OG idols and mentors and the ones who I continuously turn to when I am unsure what to do next. Thank you for carrying me to the start and then cheering me on towards the finish line.

To my 3aroos lahtah. You were the one who tipped the first domino; thank you for encouraging me to get my PhD and for sticking with me through it all. You supported me in every possible way: showing up to my presentations, offering advice, listening to my practice talks, pretending to understand what the hell I am saying, helping me take it one step further, and boosting my confidence; I've talked your ear off so much that you might

be reaching for an honorary PhD at this point yourself. You push me to be the best person I can be, and you do so while leading by example. There's no one else I'd rather be climbing this hill with than you.

VITA

Education

2017 B.A., Molecular and Cellular Biology – Immunology and Pathogenesis
University of California, Berkeley, Berkeley, CA

Experience:

Graduate Student Researcher: September 2018 – Present

Laboratory of Donald B. Kohn, M.D., University of California, Los Angeles – Los Angeles, CA

- Generate lentiviral vectors (LVs) using bioinformatics to detect enhancers that modulate lineage specific expression of genes regulating X-Linked Lymphoproliferative Disease.
- Utilize site-specific cDNA insertion of the cystic fibrosis transmembrane conductance regulator (CFTR) into the 5'UTR of the CFTR gene with CRISPR/Cas9 reagents to correct for any of the 2000 disease causing mutations of cystic fibrosis.
- Improve Pseudotyping Efficiency of SARS-CoV-2 Spike glycoprotein for LVs via amino acid substitutions and cytoplasmic tail modifications.

Master of Business Administration: September 2022 – Present

- Fully Employed, Anderson School of Management, UCLA – Los Angeles, CA

Student Researcher – 2015 – 2017

Laboratory of Jacob Corn, M.D., University of California, Berkeley – Berkeley, CA

- Conducted various CRIPSR/Cas9 experiments to induce a V617F mutation within the JAK2 locus to test for correlation with myeloproliferative neoplasms. Tested various modified sgRNAs to see their comparative efficiency and life within a cell compared to lentiviral guides. Tested whether BTBD3 was a novel substrate adapter to the CUL5-RNF7 ubiquitin ligase scaffold.

Research Associate – 2015 – 2016

Laboratory of Donald B. Kohn, M.D., UCLA – Los Angeles, CA

- Examined various methods in manipulating upstream and downstream elements of a ß-globin lentiviral cDNA vector in hopes of improving polyadenylation as well as titer and expression of the vector. Optimized lentiviral vectors to decrease backbone size and increase transcriptase processivity and fidelity to see effects on the vector's titer and expression.

Research Associate – 2014

Laboratory of Victoria Sork, University of California, Los Angeles – Los Angeles, CA

- Examined *Quercus lobata* interactions with temperature and stress and their effects on the plants genetic composition. Conducted numerous TruSeq DNA and RNA tests after weeks of sampling oaks after different temperature and stress manipulations. Archived 903 *Quercus lobata* samples onto herbarium specimen.

Awards:

Dean's Merit Scholarship: 2022 – 2025

UCLA Graduate Division Fellowships & Financial Services

BSCRC Training Program Fellowship: 2021 – 2022

Eli & Edythe Broad Center of Regenerative Medicine & Stem Cell Research, UCLA

Graduate Dean's Scholar Award: 2018 – 2020
UCLA Graduate Division Fellowships & Financial Services

Publications and Presentations:

- **Ayoub PG**, Kohn DB. β-Thalassemia: all about that base, no cutting. *Blood*. 2023 Mar 9;141(10):1098-1099. doi: 10.1182/blood.2022019350. PMID: 36893006.
- Segura EER, **Ayoub PG**, Hart KL, Kohn DB. Gene Therapy for β-Hemoglobinopathies: From Discovery to Clinical Trials. Viruses. 2023 Mar 9;15(3):713. doi: 10.3390/v15030713. PMID: 36992422; PMCID: PMC10054523.
- Cuvelier GDE, *et al*. Outcomes following treatment for ADA-deficient severe combined immunodeficiency: a report from the PIDTC. Blood. 2022 Aug 18;140(7):685-705.
- Ibarrondo, F. J., Hofmann, C., Ali, A., **Ayoub, P.**, Kohn, D. B., & Yang, O. O. (2021). Previous Infection Combined with Vaccination Produces Neutralizing Antibodies with Potency against SARS-CoV-2 Variants. In A. Pekosz & D. E. Griffin (Eds.), mBio (Vol. 12, Issue 6). American Society for Microbiology.
- **Ayoub, P.G.**, Purkayastha, A., Quintos, J., Tam, C., Lathrop, L., Tam, K., Ruiz, M., Hollis, R.P., Gomperts, B.N., & Kohn, D.B. (2021). Improved SARS-CoV-2 Spike Glycoproteins for Pseudotyping Lentiviral Vectors. Frontiers in Virology.
- Benitez EK, Lomova Kaufman A, Cervantes L, Clark DN, **Ayoub PG**, Senadheera S, Osborne K, Sanchez JM, Crisostomo RV, Wang X, Reuven N, Shaul Y, Hollis RP, Romero Z and Kohn DB (2020) Global and Local Manipulation of DNA Repair Mechanisms to Alter Site-Specific Gene Editing Outcomes in Hematopoietic Stem Cells. Front. Genome Ed. 2:601541.
- Morgan RA, Unti MJ, Aleshe B, Brown D, Osborne KS, Koziol C, **Ayoub PG**, Smith OB, O'Brien R, Tam C, et al. Improved Titer and Gene Transfer by Lentiviral Vectors Using Novel, Small β-Globin Locus Control Region Elements. Molecular Therapy. 2020;28(1):328–340.
- Morgan RA, Ma F, Unti MJ, Brown D, **Ayoub PG**, Tam C, Lathrop L, Aleshe B, Kurita R, Nakamura Y, et al. Creating New β-Globin-Expressing Lentiviral Vectors by High-Resolution Mapping of Locus Control Region Enhancer Sequences. Molecular Therapy - Methods & Clinical Development.
- **Ayoub, P.G.** American Society of Gene & Cell Therapy Annual Meeting. Poster Presentation. 2023. Regulated Lentiviral Vectors for the Treatment of X-Linked Lymphoproliferative Disease
- **Ayoub, P.G.** CF Foundation Research Conference. June 2022. Poster Presentation. Optimizing CRISPR/Cas9 Editing Strategy for site-specific knock-in into the *CFTR* locus.
- **Ayoub, P.G.** Broad Stem Cell Research Center Annual Symposium. January 2022. Poster Presentation. Rationally Designed Lentiviral Vector for Regulated Expression of SH2D1A to Treat X-Linked Lymphoproliferative Disease.
- **Ayoub, P.G.** American Society of Gene & Cell Therapy COVID-19 Symposium. September 2020. Oral Presentation. Improved SARS-CoV-2 Spike Glycoprotein for Pseudotyping Lentiviral Vectors
- **Ayoub, P.G.** UCLA COVID-19 Basic, Translational and Clinical Research Task Forces. September 2020. Oral Presentation. Improved SARS-CoV-2 Spike Glycoprotein for Pseudotyping Lentiviral Vector

INTRODUCTION

Transplantation of allogeneic hematopoietic stem cells (HSCs) offers a potential cure for various inherited blood cell disorders, including primary immune deficiencies, hemoglobinopathies, and other metabolic, storage, and stem cell defects.[7,8] Hematopoietic stem cells give rise to all other blood cells and thus, healthy, genetically normal HSCs from a donor can provide a continuous supply of all blood cell types, potentially eradicating these conditions with a single, life-long beneficial treatment.

Success rates are generally higher when the donor is an HLA-identical sibling.[7,8] However, the effectiveness diminishes when the donor is a less closely matched allogeneic source, such as haplo-identical relatives or unrelated individuals, leading to increased risks of graft rejection and graft-versus-host disease (GVHD). If an HSC graft is rejected, it can leave the recipient vulnerable, necessitating immediate action to reinstate blood cell production to avoid prolonged pancytopenia, which includes anemia, infections, and bleeding risks. Sometimes, the primary donor may not be accessible, for instance, when cord blood units can't be traced back to the source, making it challenging to find a second suitable match.

Given the severe immune-related complications often seen with allogeneic HSC transplantation, HSC gene therapy (a form of autologous HSC transplantation), uses the patient's own HSCs that have been genetically modified to circumvent the allogeneic transplantation complications. This could potentially lead to improved outcomes for those with genetic blood cell disorders. Since autologous cells are not immunogenic, they permit a less intense preconditioning regimen before transplant, which is necessary to create space in the bone marrow for new HSCs, as compared to allogeneic transplants.[7,8]

Clinical benefits comparable to those from allogeneic transplants have been achieved through autologous transplants combined with gene therapy techniques like lentiviral vectors (LVs) and gene editing. These methods have shown reliable success in clinical trials, maintaining stable levels of gene-corrected blood cells across all lineages, demonstrating the engraftment, durability, and ongoing productive capacity of the gene-modified HSCs without significant reduction over time in patients.

Gene therapy is the process of genetically modifying cells and, in conjunction with autologous transplantation, may produce a therapeutic effect or treat a disease. Over the past few decades, the field has continued to advance from treating hematopoietic disorders and inborn errors of immunity to treating infectious, cardiovascular, hepatic, and retinal diseases. With several gene therapies receiving FDA approval for treatment (Zolgensma, Zynteglo, Luxturna, and Skysona), a path to clinic and commercialization has been built.

Now, a new wave of gene therapies is progressing from bench to bedside, many of which are designed and developed here, at the laboratory of Dr. Donald B. Kohn. These include lentiviral mediated gene transfer to hematopoietic stem and progenitor cells for the treatment of inborn errors of immunity such as ADA-SCID, Wiskott Aldrich Syndrome, X-Chronic Granulomatuous Disease, IPEX Syndrome or even hemoglobinopathies such as Sickle Cell Disease, and β/α-thalassemia.[1,2,3] Furthermore, the lab has evolved to develop therapies in the form of gene editing from a decade of studies for gene correction in hematopoietic stem cells of the β-globin gene mutation that causes sickle cell disease, as well as for site-specific insertion and physiologically-regulated expression of the *CD40L* and *BTK* genes for gene therapy of X-linked Hyper-IgM Syndrome and X-linked

Agammaglobulinemia.[4,5,6] This progress is the culmination of prior advancements in gene therapy, and each iterative advancement continues to propel gene therapy technologies and gene-based treatments towards clinical application, furthering the evolution of other stem cell therapies.

The first of these technologies, described in Chapter 1 of this work describes the generation of a pseudotyped lentiviral vector to not only study the development of vaccines and therapeutics to treat SARS-CoV-2 but to also expand the capacity of research to help investigators study the effectiveness of current vaccine candidates, establish new treatments via high-throughput drug screens, examine lung pathology via infection of animal models or air-liquid interface cultures, and even explore the possible applications to gene therapy for treatment of lung diseases such as cystic fibrosis.

Throughout this work, we modified the SARS-CoV-2 spike glycoprotein to improve the transduction efficiency of spike pseudotyped lentiviral vectors by 1000-fold, which avoids the need for BSL-3 laboratories when studying the virus. We have demonstrated that our pseudotyped lentivirus contains high-on target specificity to ACE2 expressing cells and minimal off-target specificity to non-ACE2 expressing cells. Our pseudotyped virus also depends on TMPRSS2 protease for increased transduction, thereby indicating similar protein priming and entry as the SARS-CoV-2 virus. While previous attempts to pseudotype lentiviral vectors with the spike glycoprotein have demonstrated sufficient viral infectivity for downstream assays, their delta-19 pseudotype suffers from off-target infectivity to non-ACE2 expressing cells. The modified spike pseudotyped lentiviral vector generated in this report (HA) maintains the high-infectivity of the d19 pseudotype while offering greater specificity to ACE-2 expressing cells.

The second chapter of this work aims to develop the first lineage specific LV for the treatment of XLP1. Additionally, also it describes a general method for the quick identification and incorporation of enhancers into LV that achieve temporal and lineage specific expression of any target gene. Dr. Ryan Wong, a previous graduate student within the laboratory of Dr. Donald B. Kohn had recently published work describing this pipeline;[2] however, excluding that work, the only gene therapy LVs to contain fully regulated endogenous elements are the Locus Control Region driven beta-globin vectors and the conserved noncoding sequence driven FoxP3 vectors.[1-3] However, the beta-globin vectors are supported by decades of research into regulation of the globin genes and the FoxP3 vector was designed solely focused on promoter proximal CNS enhancer elements, detected by utilizing mouse knockout models and the UCSC genome browser. In contrast, the explication of this multiplexed-bioinformatics pipeline will drastically improve the speed, strength, and specificity of future developed stem cell therapies, especially those requiring tight regulation such as JAK3-SCID, RAG1-SCID, cyclic neutropenia, XLA, and others.

While gene transfer to multipotent long-term HSC is needed for an enduring effect, expression of SAP is best limited to its normal cell range, including cytotoxic T lymphocytes, NK cells and iNKT cells. To achieve precise expression pattern, we developed a series of lentiviral vectors using regulatory elements from the SH2D1A locus. By iterative process, we developed two candidate vectors that express SAP protein in the native pattern at levels similar to those of the endogenous gene. These vectors may support reconstitution of the immunologic deficiencies of XLP1 in a safe manner.

The final chapter describes an ongoing effort of a multi-disciplinary and collaborative project that leverages expertise in stem cell biology & regenerative medicine, bioengineering & nanotechnology, and gene therapy to develop gene delivery strategies targeting correction of the Cystic Fibrosis (CF) gene in the long-lived stem cells of the airway, leading to lasting and durable cures for all patients with CF. The work presented in this thesis aims to develop will generate reporters and gene-editing tools to successfully correct >99% of all CFTR mutations in *in vitro, ex vivo,* and *in vivo* CF disease models. The work presented describes the first leg of this research: optimizing the gene editing cargo to maximize integration and expression of a CFTR cDNA donor cassette after CRIPSR/Cas9 induction of a double stranded break. We designed various improvements to gene editing cargo to increase the cutting efficiency and integration of gene editing cargo including the inclusion of electroporation enhancers, double stranded DNA donor chemical modifications and small molecule inhibitors. These improvements have led to over 2-fold increases in sgRNA cutting efficiency and over 2-fold increases in donor integration at the *CFTR* 5'UTR locus. Efforts are still underway to assess the effects of *CFTR* cDNA codon optimization on expression.

The overarching goal is to develop a gene therapeutic strategy that results in effective delivery to long-lived airway basal stem cells (ABSCs) that will lead to durable and curative gene therapy approaches for patients with CF. The future of this project therefore aims to induce transient injury to the mucosal layer of the airway epithelium to enable inhaled aerosol-based delivery of nucleic acids to the ABSCs. As we develop the gene editing tools for the correction of the CF phenotype, the laboratories of Dr. Steven Jonas and Dr. Brigitte Gomperts will develop the delivery and validation tools necessary to reach

our goals. Dr. Steven Jonas's group, experts in bioengineering and nanotechnology, will evaluate strategies that incorporate surfactants to break up mucus and disrupt cell-cell junctions between differentiated apical cells, exposing the ABSCs to the gene-editing constructs. The Gomperts group, experts in airway stem cell biology, will test the efficacy and safety of the tools utilized by Jonas and Kohn labs in human in vitro, ex vivo, and in vivo airway models. This will include effects of surfactants on airway injury, repair, and transport of gene-editing cargo.

Chapter 1: Improved SARS-CoV-2 Spike Glycoproteins for Pseudotyping Lentiviral Vectors

ABSTRACT

The spike (S) glycoprotein of SARS-Cov-2 facilitates viral entry into target cells via the cell surface receptor angiotensin-converting enzyme 2 (ACE2). Third generation HIV-1 lentiviral vectors can be pseudotyped to replace the native CD4 tropic envelope protein of the virus and thereby either limit or expand the target cell population. We generated a modified S glycoprotein of SARS-Cov-2 to pseudotype lentiviral vectors which efficiently transduced ACE2-expressing cells with high specificity and contain minimal off-target transduction of ACE2 negative cells. By utilizing optimized codons, modifying the S cytoplasmic tail domain, and including a mutant form of the spike protein, we generated an expression plasmid encoding an optimized protein that produces S-pseudotyped lentiviral vectors at an infectious titer (TU/mL) 1000-fold higher than the unmodified S protein and 4 to 10-fold more specific than the widely used delta-19 S-pseudotyped lentiviral vectors. S-pseudotyped replication-defective lentiviral vectors eliminate the need for biosafety-level-3 laboratories required when developing therapeutics against SARS-CoV-2 with live infectious virus. Furthermore, S-pseudotyped vectors with high activity and specificity may be used as tools to understand the development of immunity against SARS-CoV-2, to develop assays of neutralizing antibodies and other agents that block viral binding, and to allow *in vivo* imaging studies of ACE2-expressing cells.

INTRODUCTION

Human coronaviruses (CoV) are enveloped, positive stranded RNA viruses of the family of *Coronaviridae* (36). Three coronaviruses within the past two decades were transmitted from animals to humans to cause severe respiratory diseases in afflicted individuals: the 2002 severe acute respiratory syndrome coronavirus (SARS-CoV), the 2012 Middle East respiratory syndrome coronavirus (MERS-CoV), and most recently, the 2019 SARS-CoV-2 (36). Although researchers have raced to develop safe and efficacious vaccines to prevent SARS-CoV-2 spread (26), the virus has resulted in over 4.25 million deaths worldwide since the inception of its pandemic spread.

The spike (S) glycoprotein mediates viral entry into target cells after engaging with the cell surface receptor angiotensin converting enzyme 2 (ACE2) (20). In addition to the prevalence of ACE2 receptors, various factors affect the potency of SARS-CoV-2 entry and transmission. One such factor includes the availability of proteases in target cells, such as the transmembrane protease serine 2 (TMPRSS2) or furin (2,4,20). TMPRSS2 facilitates S protein priming of SARS-Cov-2 to promote fusion of viral and cellular membranes to the target cell. Similarly, furin, a ubiquitous protease that activates a variety of viruses such as influenza A, HIV, Ebola, or measles (4), cleaves SARS-CoV-2 S protein at the S1/S2 site to facilitate infection. A viral factor affecting potency includes the D614G substitution in the spike glycoprotein, which has quickly become the most prevalent form of SARS-Cov-2. The dominance of the D614G variant is attributed to its increased viral infectivity and transmission (38).

Despite the progressive development of vaccines and therapeutics to treat SARS-CoV-2, there is a need for a safer alternative to infectious SARS-CoV-2 virus for research studies quantifying neutralizing antibody activity, developing high-throughput drug screens, or performing animal studies. By “pseudotyping” a virus, one can effectively replace the envelope protein of a virus with that of another virus to limit or broaden its targeting capabilities (37). Recent studies demonstrated that the spike glycoprotein from SARS-CoV-2 can be “pseudotyped” onto replication defective viral particles such as HIV-based lentiviral particles, murine leukemia virus (MLV)-based retroviral particles, or Vesicular Stomatitis Virus (VSV) (7,9,14,15,20,25,29). As a result, generating S glycoprotein pseudotyped non-replicative viral particles capable of specifically infecting ACE2-expressing cells will avoid the need for a biosafety-level-3 facility and expand the capabilities for studies of SARS-CoV-2.

A major hurdle with this approach, however, stems from the low titer of S-pseudotyped vectors (24,32). Previous groups have shown that modifications to the cytoplasmic tail (CT) domain of viruses can increase their lentiviral vector pseudotype efficiency (5,12,27,28,30). For example, researchers have attempted to increase the infectious titers of lentiviral vectors pseudotyped by the gibbon ape leukemia virus (GaLV) by adjusting its CT (29,34). The GaLV envelopes were modified to harbor the CT from MLV-A thereby increasing infectious titers by 25-fold. In the case of SARS-CoV-2, various researchers have demonstrated that simply deleting the final 19 or 21 amino acids of the cytoplasmic tail of SARS-CoV-2 (d19/d21) can substantially increase titers of SARS-CoV-2 pseudotyped lentiviral vectors (14,15,25,29). We describe a strategy to generate an S glycoprotein, with a CT replaced with that of the influenza hemagglutinin protein, capable

of pseudotyping a non-replicative 3rd generation HIV-1 lentiviral vector with a titer 3-logs greater than its unmodified counterpart. Furthermore, the HA-tailed pseudotype not only maintains high levels of infectivity but also demonstrates a 4 to 10-fold greater specificity to ACE2-expressing cells than the d19 pseudotyped viruses.

MATERIALS AND METHODS

Human Tissue

Large airways and bronchial tissues were acquired from de-identified normal human donors after lung transplantations at the Ronald Reagan UCLA Medical Center. Tissues were procured under an Institutional Review Board-approved protocol at the David Geffen School of Medicine at UCLA, protocol no.16-000742. ABSCs were used from two biological replicates. One experiment was done using normal human bronchial epithelial cells (NHBE) from non-smokers obtained from Lonza and all samples were de-identified.

ABSC Isolation

Human ABSCs were isolated following a previously published method (43-48). Briefly, airways were dissected, cleaned, and incubated in 16U/mL dispase for 30 min at room temperature. Tissues were then incubated in 0.5mg/mL DNase for another 30 min at room temperature. Epithelium was stripped and incubated in 0.1% Trypsin-EDTA for 30 min shaking at 37°C to generate a single cell suspension. Isolated cells were passed through a 40 μm strainer and plated for Air Liquid Interface cultures.

Air Liquid Interface Cultures and Transduction

24-well 6.5mm transwells with 0.4 μm pore polyester membrane inserts were coated with collagen type I dissolved in cell culture grade water at a ratio of 1:10. 100 μl was added to each transwell and allowed to air dry. ABSCs were seeded at 100,000 cells per well directly onto collagen-coated transwells and allowed to grow in the submerged phase of culture for 4-5 days with 500 μL media in the basal chamber and 200 μL media in the apical chamber. ALI cultures were then established and transduced with equal amounts

of p24 protein content. 72 hours post transduction, cultures were harvested for IF studies and vector copy number analysis. Human ABSCs were grown in Pneumacult Ex serum-free media during the submerged and Pneumacult ALI media during the ALI phases of culture, respectively. Media was changed every other day and cultures were maintained at 37°C and 5% CO2.

Immunocytochemistry, Confocal Imaging and Cell Counting

ALI cultures were fixed in 4% paraformaldehyde for 15 min followed by permeabilization with 0.5% Triton-X for 10 min. Cells were then blocked using serum-free protein block (Dako X090930) for one hour at room temperature and overnight for primary antibody incubation. After several washes of Tris-Buffered Saline and Tween-20 (TBST), secondary antibodies were incubated on samples for 1 hour in darkness, washed, and mounted using Vectashield hardest mounting medium with DAPI (Vector Labs H-1500). The following antibodies were used for staining: Mouse Acetylated α-Tubulin (Cell Signaling Technology) and GFP antibody (Rockland Inc). IF images were obtained using an LSM700 or LSM880 Zeiss confocal microscope and composite images generated using ImageJ. Approximately equal numbers of cells (around 1,000 cells) were counted for each experimental group. All immunofluorescence images used for scoring cells consisted of a z-series of optical sections captured on the Zeiss LSM 700 or 880 confocal microscopes.

Cell lines and culture

HEK-293T (#CRL-3216; American Type Culture Collection (ATCC), Manassass, VA), PKR -/- HEK 293T (in house)13, ACE-293T (ACE2-expressing 293T cells generously provided by Dr. Lili Yang (UCLA) and Dr. Pin Wang (USC)), VeroE6 (generously provided

by Dr. Jocelyn Kim (UCLA)), and VeroE6/TMPRSS2 (#JCRB1819; JCRB Cell Bank (41)) cells were cultured in DMEM (#10-017-CV; Corning Inc., Corning, NY) supplemented with 10% fetal bovine serum (#100-106; GeminiBio, Calabasas, CA) and 1% penicillin/streptomycin/L-glutamine (#400-110; GeminiBio). Cell counts were measured with a Vi-CELL XR automated cell counter (Beckman Coulter, Indianapolis, IN).

Generation of Spike Glycoprotein Envelope Plasmids

The spike pseudotype backbone was generated by PCR amplification using the pMD.G-VSVG (22) and primers oPAF147 and oPAR43 (see Supplemental materials). Ou (25) and IDT (7) codon optimized full-length spike glycoproteins were ordered as double stranded DNA fragments (Integrated DNA Technologies (IDT), Coralville, IA) with the inclusion of wildtype, HA (7), ALAYT (21), and MLV (28) cytoplasmic tails for downstream cloning. The IDT codon optimization was generated by inserting the Wuhan-Hu-1 SARS-CoV-2 spike glycoprotein sequence (Reference Sequence: NC_045512.2) into the IDT Codon Optimization Tool. The Ou codon optimization was taken directly from supplemental materials from a paper by Ou et al. and codon optimized from the SARS-CoV-2 S glycoprotein (QHU36824.1) (25). The codon optimized fragments were synthesized as double-stranded DNA fragments from IDT (See Supplementary Materials). The following constructs were generated using the New England Biosciences (NEB) Gibson Assembly workflow with double-stranded DNA fragments that replaced the VSV-G open reading frame from the pMD2.G plasmid (Addgene #12259): IDT (oPAG40 and oPAG41), IDT-ALAYT (oPAG40 and oPAG42), IDT-HA (oPAG40 and oPAG43), IDT-MLV (oPAG40 and oPAG44), Ou (oPAG45 and oPAG46), and Ou-MLV (oPAG45 and oPAG47). The cytoplasmic tail of the spike glycoprotein – the final 34 amino acids – were

replaced with either the HA or the MLV tail. The HA tail (7) consists of the final 10 amino acids of the influenza A hemagglutinin protein (NGSLQCRICI) whereas the MLV tail consists of the final 30 amino acids of the murine leukemia virus (4070A) cytoplasmic tail (28). The Q5 Site-Directed Mutagenesis Kit (#E0554S, NEB, Ipswich, MA) – with primers oPAF176 and oPAR56 or oPAF177 and oPAR57 – was used to generate the D614G mutants from each respective plasmid. Ou-D614G-HA spike glycoprotein was PCR amplified and cloned from the Ou-D614G plasmid using primers oPAF184 and oPAR63. Primers oPAF228 and oPAR103 were used to remove the final 19 amino acids of the Ou-D614G cytoplasmic tail to generate Ou-D614G-d19 spike glycoprotein. The United Kingdom Alpha variant, the South African Beta variant, and the Indian Kappa and Delta variants were generated using the NEB Gibson Assembly workflow and the following double-stranded DNA fragments: Alpha (oPAG59 and oPAG60), Beta (oPAG61 and oPAG62), Kappa (oPAG67) and Delta (oPAG68). Any further mutations for Alpha, Beta, Kappa, and Delta were generated via the Q5 Site-Directed Mutagenisis Kit with primers (oPAF222, oPAF232-234, oPAR97, oPAR108-110, oPAF282-284, and oPAR127-129). Alpha pseudotypes contained the following mutations: ∆69-70, ∆144Y, N501Y, A570D, D614G, P681H, T716I, S982A, and D1118H. Beta pseudotypes contained the following mutations: L18F, D80A, D215G, ∆241-243, R246I, K416N, E484K, N501Y, D614G, and A701V. Kappa pseudotypes contained the following mutations: L452R, E484Q, D614G, P681R, Q1071H. Delta pseudotypes contained the following mutations: T19R, delta157-158, L452R, T478K, D614G, P681R, D950N. The addition of the HA tail and the d19 tail were completed as described above. All plasmids were mini prepped using the PureLink™ Quick Plasmid Miniprep Kit (#K210010; Invitrogen, Carlsbad, CA). All

plasmids were maxi prepped using NucleoBond Xtra Maxi Kit (#740414; Machery-Nagel Inc., Düren, Germany).

Vector Packaging and Titration

Spike-pseudotyped lentiviruses were packaged by transient transfection of PKR -/- 293T cells with fixed amounts of HIV Gag/Pol, Rev, and Lentiviral envelope (VSV-G or Spike) expression plasmids and equimolar amounts of either MNDU3-eGFP or roUBC-mCitrine transfer plasmid using TransIT-293 (Mirus Bio, Madison, WI) as described in the Supplemental Methods and Cooper et al. (2011) (6,13). Viral supernatants were then directly used for titer determination or concentrated by tangential flow filtration, as described by Cooper et al., 2011. Briefly, ACE-293T, HEK-293T, VeroE6, or VeroE6/TMPRSS2 (41) cells were transduced at equal amounts of p24 protein content with either a 1:10 dilution of raw or a 1:10,000 dilution of concentrated vector. To calculate titers, we harvested cells and determined VCNs by ddPCR approximately 72 h post-transduction.

Transduction of cell lines with spike-pseudotyped lentivirus

ACE-293T, HEK-293T, VeroE6, or VeroE6/TMPRSS2 cells, 1 x 105 per sample, were collected by trypsinization, centrifuged at 90g for 10 minutes and resuspended in 2mL of culture medium for plating in a 6-well plate (#3516; Corning Inc.). 24 hours after plating, cells were transduced with equal amounts of p24 protein content; culture medium was replaced with a 1:10 dilution of viral supernatant in 1mL of culture medium. 24 hours after transduction, culture medium was refreshed on all wells. 72 hours after transduction, cells were harvested for downstream analyses. Cell counts were measured with a Vi-CELL XR automated cell counter. Cells were assayed for GFP expression with a BD LSRFortessa

or BD LSRII flow cytometer (BD Biosciences, San Jose, CA) and analyzed with FlowJo (Tree Star, Ashland, OR).

Digital Droplet PCR for VCN and Titer (TU/mL) Quantification

Genomic DNA from transduced cells was extracted using PureLink Genomic DNA Mini Kit (K182002; Invitrogen). VCN was calculated by using the vector GFP gene (primers eGFP616F and eGFP705R; probe eGFP653Pr) and an endogenous human diploid gene control (SCD4 (Human Syndecan 4) primers oPAF-SDC4 and oPAR-SDC4; probe oPAP-SDC4) as a reference. Reaction mixtures of 22uL volume, comprising 1 × Digital droplet (dd)PCR Master Mix (#1863010; BioRad, Hercules, CA), 400 nmol/L primers and 100 nmol/L probe for each set, 40 U DraI (R0129S; NEB) and 30– 100 g of the gDNA to study, were prepared and incubated at 37°C for 1 h. Droplet generation was performed as described in Hindson et al (9). with 20 L of each reaction mixture. The droplet emulsion was then transferred with a multichannel pipet to a 96-well twin.tec® real-time PCR Plates (Eppendorf, Hamburg, Germany), heat sealed with foil, and amplified in a conventional thermal cycler (T100 Thermal Cycler, Bio-Rad). Thermal cycling conditions consisted of 95°C 10 min, (94°C 30 s and 60°C 1 min) (55 cycles), 98°C 10 min (1 cycle) and 12°C hold. After PCR, the 96-well plate was transferred to a droplet reader (Bio-Rad). Acquisition and analysis of the ddPCR data was performed with the QuantaSoft software (Bio-Rad), provided with the droplet reader. Vector Titer (TU/mL) was calculated as TU=VCN x (cell count at day of transduction) x virus dilution.

P24 Assay

p24 antigen concentration in vector supernatants were measured by the UCLA/CFAR (Center for AIDS Research) Virology Core using the Alliance HIV-1 p24 Antigen ELISA Kit (#NEK050, PerkinElmer, Waltham, MA), following the manufacturer's manual.

RESULTS

Generation of Modified S-Pseudotyped Lentiviral Vectors

To achieve the most effective S-pseudotyped lentiviral vector, we modified the full-length SARS-CoV-2 S glycoprotein gene to generate a series of envelope expression plasmids for transfection experiments. The modifications included various codon optimizations, an amino acid substitution, and CT modifications (Figure 1.1 A). As such, the S glycoprotein was codon optimized using the Integrated DNA Technologies algorithm, denoted “IDT,” or using the codon optimization described by a Ou et al., denoted “Ou” [25]. We also modified the constructs to include an aspartic acid to glycine substitution at the 614th amino acid position of the S protein. This mutation reproduces the D614G SARS-CoV-2 variant that has rapidly become the dominant form around the world [38]. Recent research has demonstrated that this strain exhibits increased competitive fitness and infectivity thereby increasing transduction of ACE2-expressing cells.

We modified the spike CT to introduce amino acid substitutions into its C terminus or replace it with the CT of the influenza A or murine leukemia envelope glycoprotein (Figure 1.1 A). Various groups have shown that mutating the five most C-terminal residues of the S glycoprotein eliminates the endoplasmic reticulum retention signal and improves surface expression of S-pseudovirions [7,21]. This mutant, termed “ALAYT,” replaces K1269 and H1271 with alanines (A). Additionally, we replaced the spike CT with the final 10 amino acids of the influenza A virus hemagglutinin (HA) glycoprotein [7]. Given that SARS-Cov-2, influenza HA, and HIV-1 are homotrimeric class I fusion glycoproteins [10,36,38], we reasoned that the influenza CT may pseudotype HIV-1 lentiviral vectors more effectively than that of SARS-CoV-2. Lentiviral vectors pseudotyped with GALV,

RD114, or baboon envelopes have also shown increased stability and transduction when their native CTs are replaced with the MLV cytoplasmic tail [12,28]. As a result, we also replaced the spike CT with that from the murine leukemia virus (MLV) glycoprotein.

We generated the SARS-CoV-2 pseudotyped virus using a 3rd generation HIV-1 packaging system (Figure 1.1 B). We transfected PKR -/- 293T cells with plasmids encoding for HIV-1 proteins, Gag- Pol and Rev; a plasmid encoding for the full-length spike glycoprotein envelope; and a lentiviral transfer plasmid containing an eGFP reporter cassette [13]. The S glycoprotein on the produced lentivirus should improve specificity towards ACE2-expressing cells.

Functionality and Expression of S-pseudotyped Lentiviral Vectors

We first examined the effect of glycoprotein modifications on viral expression and functionality. ACE2-expressing 293T cells (ACE-293T) or parental 293T cells that do not express ACE2 (HEK-293T) were transduced with the series of lentiviral vectors. Three days post-transduction cells were harvested to quantify vector expression and infectivity. All spike-pseudotyped lentiviral vectors exhibited specificity to ACE2-expressing cells with essentially no off-target transduction of the parental HEK-293T cells that lack expression of ACE2 (Figure 1.2 A). Furthermore, Ou codon optimized variants presented greater infectivity than their IDT counterparts. The inclusion of either the D614G mutant or the HA tail increased vector copy number (VCN) by 10-fold (Figure 1.2B), whereas the ALAYT amino acid modifications had no impact on pseudotyping efficiency. When combined, the D614G mutant and HA tail achieved a 50-fold synergistic effect of the percentage of cells expressing GFP (Figure 1.2A), a 1000-fold increase of vector copy number (Figure 1.2B), and a mid-10^7 titer (Figure 1.2C) – 3 logs greater than its unmodified spike counterpart.

In contrast, the MLV tail impeded pseudotyping efficiency. When combined with the D614G mutant, the MLV tail reduced expression and infectivity of the pseudotyped virus in comparison to the D614G mutant alone (Figure 1.2A, 2C). The S-pseudotyped vector also supports large scale concentration (Figure 1.5) greater than 1000-fold by means of tangential flow filtration [6]. Ultimately, these data demonstrate that the combination of the Ou codon optimization, the D614G mutation, and the HA cytoplasmic tail drastically increased pseudotyping efficiency of lentiviruses with the spike glycoprotein.

Infected cells can be measured by means of the eGFP reporter driven by the MNDU3 promoter, an enhancer/synthetic promoter that contains the U3 region of the Myeloproliferative Sarcoma Virus long terminal repeat [17]. It is also worth noting the reduction in transduction activity assessed by GFP expression across all spike-pseudotyped variants compared to the VSV-G control (Figure 1.2A). Even with a VCN of 0.1, various pseudotyped lentiviruses exhibited negligible GFP percentage (<5%) measured by flow cytometry. This reduced expression may be attributed to the MNDU3 promoter driving the GFP reporter, as it is a relatively weak expressing promoter within 293T cell lines. This is evidenced by the increase in expression when we packaged the spike-pseudotyped vectors with a GFP reporter cassette driven by the ubiquitin C promoter (Figure 1.6). These UBC-mCitrine spike-pseudotyped vectors had greater GFP expression in ACE-293T cells in comparison to the MNDU3-eGFP pseudotyped variants after transduction at equal p24.

Enhanced Transduction of S Pseudotypes in the Presence of TMPRSS2 Protease

We also determined whether increasing levels of TMPRSS2 protease in target cells could affect infectivity by the S-pseudotyped lentiviral vectors. VeroE6 cells, which lack

TMPRSS2 expression, or VeroE6 cells transfected to express TMPRSS2 (VeroE6/TMPRSS2 [41]) were used to titer the vectors. VeroE6 and VeroE6/TMPRSS2 cells were transduced with IDT, IDT-D614G-HA, Ou, and Ou-D614G-HA pseudotyped lentiviral vectors. Three days post-transduction, cells were harvested to quantify vector expression and infectivity. For both the IDT-D614G-HA and Ou-D614H-HA pseudotyped vectors, transduction was 5-fold higher on the VeroE6/TMPRSS2 cells, demonstrating their dependence on this protease for enhanced transduction (Figure 1.3A-3B; Figure 1.7).

Comparison of Influenza Hemagluttinin and d19 Cytoplasmic Tail on S Pseudotype Infectivity

We next set out to compare the infectivity and specificity of the S-pseudotyped HA variant to the d19 cytoplasmic tail, the current standard for SARS-CoV-2 pseudotyping vectors [14,15,25,29]. The d19 plasmid encodes for a spike envelope protein with the final 19 most C-terminal amino acid residues removed from the cytoplasmic tail. We generated two new pseudotype sets containing the Alpha and Beta SARS-CoV-2 mutations. These included the following glycoprotein envelopes: wildtype spike containing the HA or d19 tail (Ou-HA, Ou-d19), the United Kingdom spike variant (Ou-Alpha, Ou-Alpha-HA, Ou-Alpha-d19) or the South African spike variant (Ou-Beta, Ou-Beta-HA, and Ou-Beta-d19). We packaged each S-pseudotyped vector head-to-head with a GFP reporter cassette driven by a reverse oriented ubiquitin C promoter (roUBC-mCitrine) [13]. ACE-293T or HEK-293T cells were transduced with the S-pseudotyped virus at equal amounts of p24 particles. Three days post-transduction, the cells were measured for their GFP percentage by flow cytometry (Figure 1.4A, 4B, S4D, S4E). When transducing ACE2-

expressing 293Ts, Ou-Alpha-d19 and Ou-Beta-d19 variants contained a 2.5 and 1.25-fold greater GFP percentage than their HA-tail counterparts, respectively (Figure 1.4A, S4D). Both the HA and d19 tails produced at least 10-fold greater GFP percentages than their unmodified counterparts. When transducing non-ACE2-expressing cells, Ou-Alpha-d19 and Ou-Beta-d19 variants contained a 10-fold greater GFP percentage than their HA-tail counterparts (Figure 1.4B, S4E; 0.4% vs 4% for Ou-Alpha and 0.25% vs 3% for Ou-Beta). The infectious capability of Kappa and Delta spike-pseudotyped lentiviral vectors were also tested in ACE2 expressing and non-expressing cells (Figure 1.9). Kappa-HA contained a 4-fold greater GFP percentage than Kappa-d19 in ACE2 expressing 293T cells. Further, Kappa and Delta spike-pseudotyped lentiviral vectors bearing a d19 cytoplasmic tail had 2 to 5-fold more GFP positive cells when transducing non-ACE2 expressing 293Ts (Figure 1.9). These data further suggest that the HA-tail spike pseudotypes exhibit greater specificity to ACE2 expressing cells than the d19-tail spike pseudotypes.

Assess the Ability for S-pseudotypes with an HA Cytoplasmic Tail to Infect Air Liquid Interface Airway Cultures

Finally, we investigated the ability of S-pseudotypes bearing an HA cytoplasmic tail to infect air-liquid interface (ALI) cultures derived from primary human airway basal stem cells (ABSCs). We utilized primary human ABSCs from three healthy lung transplant donors to generate the ALI cultures [43-48]. The cultures were transduced with HA-tailed SARS-CoV-2 pseudotypes at equal amounts of p24 particles. Three days after infection, we examined the cultures for evidence of transduction. We validated the differentiation of the ALI cultures by immunofluorescent (IF) staining with primary antibodies for Acetylated

α-Tubulin for ciliated cells for ABSCs (Figure 1.10). ALI cultures were immunostained with an anti-GFP antibody and images were obtained by confocal microscopy (Figure 1.4C, S6). Confocal imaging demonstrated GFP expression in cells of the ALI cultures that were transduced with SARS-CoV-2 HA tail pseudotypes. However, as expected, VSV-G pseudotyped vectors failed to transduce the ALI cultures. The number of infected cells in ALI cultures across the pseudotyped lentiviral vectors was quantified by counting GFP positive cells per total number of nuclei (Figure 1.4D). The %GFP cells in the ALI cultures ranged from 1% (Alpha) to 7% (Kappa). No GFP positive cells were detectable in VSV-G or Mock samples. These data demonstrate that HA-tailed SARS-Cov-2 pseudotyped lentiviral vectors are capable of transduction of primary human airway cells in an epithelial barrier culture system.

DISCUSSION

Despite the continued progression of vaccines and therapeutics for the treatment of SARS-CoV-2, the methods to support high-throughput neutralization studies and drug screens remain sub-optimal. It has been shown for several envelope glycoproteins, such as those of HIV-1, GaLV, RD114, and the baboon envelope retroviral glycoprotein, that the CT domain of the envelope determines pseudotyping constraints [5,12,27,30,31] Ultimately, the CT recruits the envelope proteins for incorporation into viral particles through interactions with the Gag protein [19]. As such, previous groups have attempted to pseudotype viral vectors with the S glycoprotein [7,14,15,20,24,25,29], by replacing the S cytoplasmic tail with the influenza HA cytoplasmic tail or by modifying its amino acids composition. Those pseudotyped vectors, however, failed to achieve infectious titers high enough for use in downstream assays such as ***in vivo*** studies. In this report, we incorporated these methods to generate modified S-pseudotyped lentiviral vectors capable of infecting ACE2-expressing cells at greater levels than the unmodified counterpart.

We investigated whether various CT modifications to the spike glycoprotein could affect pseudotype efficiency. We utilized previously published codon optimizations and CT modifications as well as the D614G mutant of SARS-Cov-2 [7,12,21,25,27]. We showed that the influenza HA cytoplasmic tail and the D614G mutant could synergistically increase infectivity over the unmodified S counterpart and maintain the high infectivity of the widely utilized d19 pseudotype. This alternative pseudotyped lentiviral vector can provide a means for studying both the neutralizing antibodies in recovered or symptomatic patients and the potency of antibody responses from current vaccine candidates.

Furthermore, the ability to concentrate the spike-pseudotyped vectors by means of tangential flow filtration (Figure 1.5) should support ***in vivo*** studies and the infection of difficult to transduce cell lines and primary cells [6].

In contrast to the HA cytoplasmic tail, the murine leukemia virus (MLV) CT hampered pseudotyping efficiency. It has been shown that within the MLV cytoplasmic tail, several elements regulate the envelope's incorporation into the virion and aid with fusogenicity into host cell membranes. The C-terminus of the MLV cytoplasmic tail, known as the R peptide, includes a conserved leucine-valine dipeptide cleavage site and a tyrosine (YXXL) motif that has been implicated in promoting endocytosis of the envelope glycoprotein [1,16,18,19]. These motifs contained within the MLV tail – but not the HA tail – may hinder particle formation or membrane fusion of spike-pseudotyped vectors.

Various groups have indicated the importance of the TMPRSS2 protease for S protein priming and subsequent infectivity [20,25,38]. We demonstrated that TMPRSS2 increased the transduction of the HA tail pseudotypes suggesting that the S-pseudotyped lentiviral vectors depend on TMPRSS2 for successful transduction. As a result, they can be utilized for drug screens against SARS-CoV-2 or as models for ***in vivo*** imaging studies of ACE2 and TMPRSS2-expressing cells.

Many groups have also established that the removal of the final 19 amino acids of the spike cytoplasmic tail (d19) can significantly increase titer and infectivity of S-pseudotyped lentiviral vectors. We compared the HA tail to the d19 tail (harboring the D614G, Alpha, or Beta mutations [42-45]) and demonstrated that S-pseudotyped lentiviral vectors with an HA cytoplasmic tail exhibit not only high levels of infectivity but also increased specificity over the widely used d19 S-pseudotyped vector. We speculate that

the increase in specificity of the HA tail could be a result of increased binding affinity to ACE2 receptors. Previous literature indicated that although the spike proteins with the HA tail and wildtype S cytoplasmic tail are neutralized by patient sera at similar IC50s, the HA tail renders S pseudotypes more sensitive to neutralization by soluble ACE2 than the wildtype S cytoplasmic tail [7]. Although it has not been further explored, an increase in binding affinity of HA-tailed S-pseudotyped vectors to ACE2 may explain both its increase in specificity and its increase in sensitivity to neutralization by soluble ACE2.

The increase in specificity of the HA S-pseudotyped vector will be important when studying the pathology of SARS-CoV-2 infection in animal models or air liquid interface cultures – including the newly dominant Delta variant [49-50] – given it will more accurately represent the specificity to ACE2 of the SARS-CoV-2 virus. Furthermore, there is an unmet need for a lentiviral vector gene therapy to combat cystic fibrosis and other lung disorders. However, the receptors for the widely used VSV-G glycoprotein are located on the basolateral surface of epithelial cells, preventing efficient transduction of basal stem cells without disruption of the airway junctions [39]. Our results demonstrated the lack of transduction of VSV-G pseudotyped lentiviral vectors in the differentiated ALI culture system. The HA tail S-pseudotyped lentiviral vectors demonstrated successful transduction of human epithelial cells within ALI cultures suggesting that ALI cultures can be utilized as tools for SARS-CoV-2 research. The lack of GFP positive cells in ALI cultures may be attributed to both donor variation and viral titer (TU/mL). Increasing the viral titer through ultracentrifugation or tangential flow filtration (Figure 1.5) can increase the transduction capability of each lentivirus. It is also worth noting that the GFP positive cells in the transduced ALI cultures do not accurately depict the transmission capability

of each variant, given the most transmissible variant (Delta) contained 3-fold less GFP positive cells than the less transmissible wildtype counterpart (Ou-HA). This may be due to multiple factors such as donor to donor variability, differences in cell entry mechanisms, kinetics of viral entry, expression of TMPRSS2 and furin proteases, or the requirement of other viral proteins to facilitate increased infection [51-52] (i.e. membrane protein, or ORF1a/b accessory proteins) – all of which may be studied with the HA-pseudotypes described in this manuscript.

The HA tail S-pseudotyped lentiviral vector may provide a potential alternative to target epithelial cells expressing ACE2 and TMPRSS2 without the aberrant expression of the transgene cassette in off-target populations caused by transduction of the d19 tail. These experiments include deducing the transduction capability of single mutations within different variants (i.e. D614G, N501Y, E484K, L452R, K417N, T478K etc.) or even testing therapeutics to combat SARS-CoV-2 infection. The non-physiological expression of the truncated d19 cytoplasmic tail may also prevent the appropriate understanding of how each mutation affects the kinetics of new variants. A great example is that of the Kappa variant, where the HA-tailed pseudotype demonstrates a 3-fold greater transduction of ACE-2 expressing cells over the d19 tail (Figure 1.9) and at least a 3-fold greater transduction in ALI cultures compared to other variants. This superiority in infection, however, is only a snapshot in time illustrated by the number of GFP expressing cells and not representative of viral transmission. To better understand transmission capabilities of SARS-CoV-2 variants, the HA tail S-pseudotyped vectors can be employed to study kinetics of infection without the compromised specificity of infection of the d19 tail S-pseudotypes. The SARS-CoV-2 model system described in this manuscript can help

determine the physiological alterations in kinetics after mutating the receptor binding domain or even the furin cleavage site. These kinetic alterations include changes in receptor binding, enhanced cleavage of the S protein, accelerated fusion, and increased cell to cell infection. Understanding these changes, may be a causal factor for the Delta variant outcompeted previous SARS-CoV-2 variants such as D614G, Alpha, Beta, Gamma, or Kappa.

Given the non-replicative nature of pseudotyped lentiviral vectors, they have significant value in studying the biology of pathogenic viruses, such as SARS-CoV-2, due to their lower biosafety requirements. The highly infectious nature of SARS-CoV-2 requires biosafety level 3 (BSL-3) equipment within laboratories to appropriately handle and study the pathogenesis or treatment of the virus [3]. By designing an efficient spike-pseudotyped HIV-1 lentiviral vector with greater potency than the unmodified S spike pseudotype, the need for BSL-3 laboratories can be avoided for many studies of SARS-CoV-2. Given the rampant, global spread of the virus, the shift from BSL-3 to BSL-2 laboratories will facilitate screening of patients' serum for neutralizing antibodies in a high-throughput fashion without risk of infection. Furthermore, this pseudotyped vector expands the capacity of research to help investigators study the effectiveness of current vaccine candidates, establish new treatments via high-throughput drug screens, examine lung pathology via infection of animal models or air-liquid interface cultures, and even explore the possible applications to gene therapy for treatment of lung diseases such as cystic fibrosis.

FIGURES

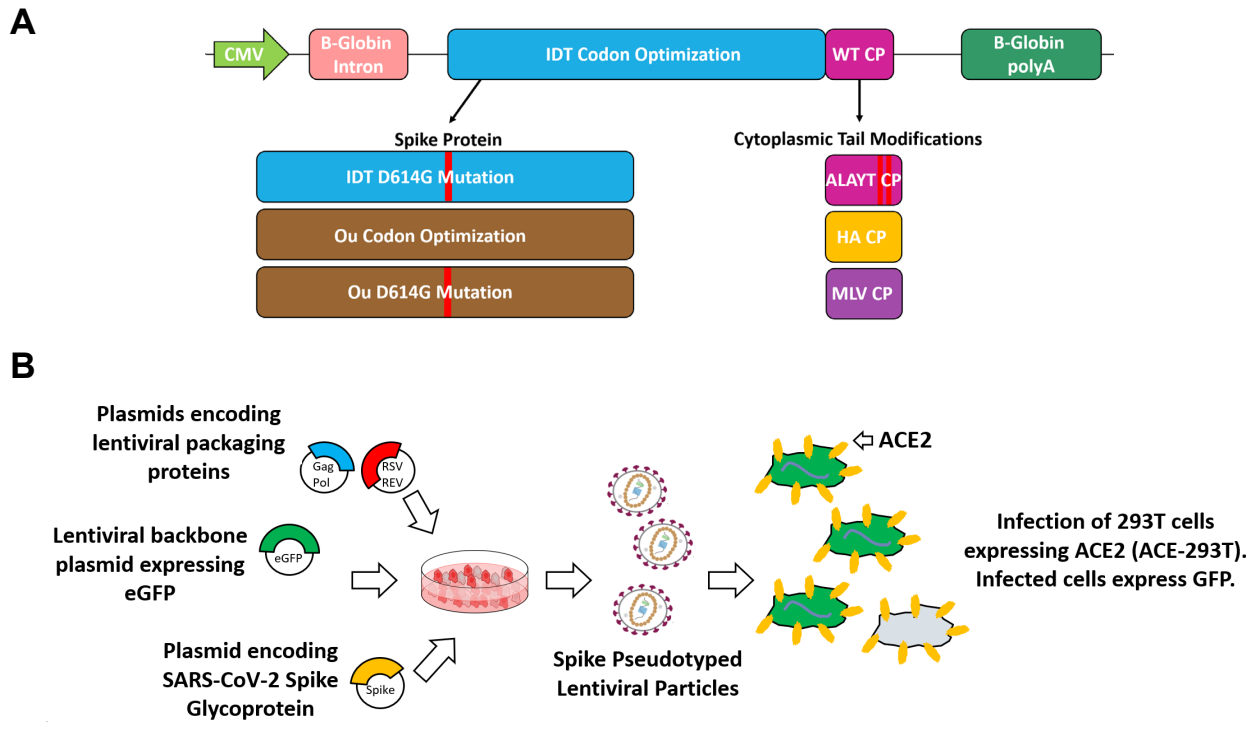


Figure 1. 1

Schematic of the spike glycoprotein variants (A), vector packaging (B), and transduction (C-E) experiments. (A) The S glycoproteins were cloned into the pMD2.G plasmid backbone (Addgene plasmid #12259) which is driven by a CMV promoter and includes a beta-globin intron and beta-globin polyA. The three variants include codon optimization (IDT or Ou), an amino acid substitution (D614G), and cytoplasmic tail modifications (wildtype tail, mutated ALAYT, influenza HA tail, and murine leukemia virus MLV tail). (B) The spike (S) glycoprotein variants were packaged in PKR-/- HEK-293T cells alongside plasmids encoding for a 3rd generation HIV-1 lentivirus and a plasmid encoding for an eGFP transgene (pCCL-MNDU3-eGFP (11, 17)).

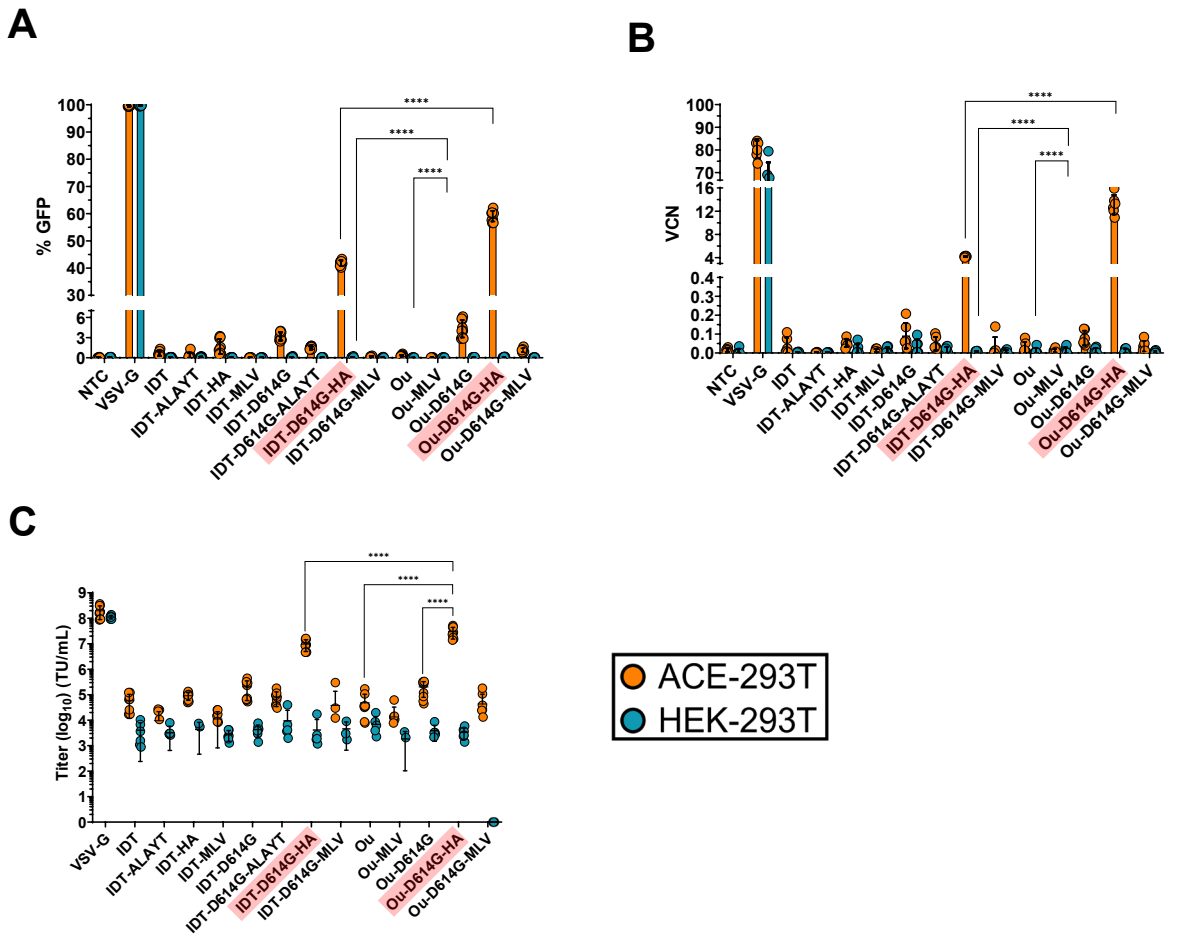


Figure 1. 2

GFP percentage (A), vector copy number (B) and titer (C) of spike pseudotyped lentiviral vectors. Raw viral supernatants were harvested 3 days post transfection and stored at -80C for future use. (A-C) 100,000 ACE2-expressing 293T cells (ACE-293T) and HEK-293T cells were transduced with raw virus at equal amounts of p24 protein content from a series of S-pseudotyped lentiviral vectors containing an MNDU3-eGFP reporter cassette. Three days post-transduction, cells were harvested and measured for %GFP (A) via flow cytometry to assess pseudotype functionality and expression. Panels B and C illustrate the infectivity of spike pseudotype vectors. To measure infectivity, genomic

DNA was extracted, and vector copy number (VCN) was measured (B) by quantifying GFP integrants via QX200 Droplet Digital PCR System. The infectious titer (transducing units/mL) of each vector was determined (C) from the VCN shown in panel B. All data sets were compared to a non-transduced control (NTC) and a VSV-G pseudotyped lentiviral vector containing an MNDU3-eGFP reporter cassette. Data are represented as mean ± SD of biological triplicates from n=3 experiments. Statistical significance was analyzed using one-way ANOVA followed by multiple paired comparisons for normally distributed data (Tukey test). All statistical tests were two-tailed and a p value of < 0.05 was deemed significant (ns non-significant, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.)

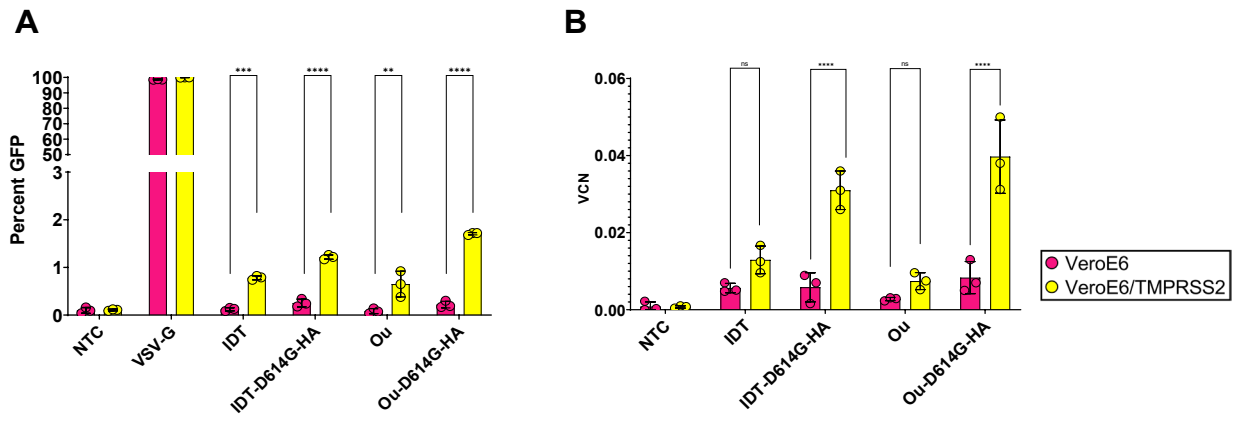


Figure 1. 3

100,000 VeroE6, and VeroE6/TMPRSS2 cells were transduced with raw viral supernatant containing equal amounts of p24 protein content from a series of pseudotyped lentiviral vectors containing an MNDU3-eGFP reporter cassette. Panels A and B illustrate the effect of TMPRSS2 expression on spike pseudotype functionality. Three days post-transduction, cells were harvested and measured for (A) GFP percentage and (B) vector copy number (VCN). GFP percentage was measured via flow cytometry. To measure infectivity, genomic DNA was extracted to measure VCN by quantifying GFP integrants via QX200 Droplet Digital PCR System. All data sets were compared to a non-transduced control (NTC) and a VSV-G pseudotyped lentiviral vector containing an MNDU3-eGFP reporter cassette. Data are represented as mean ± SD of biological triplicates from two experiments. We analyzed statistical significance using a two-way ANOVA followed by multiple paired comparisons for normally distributed data (Tukey test). All statistical tests were two-tailed and a p value of < 0.05 was deemed significant (ns non-significant, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.)

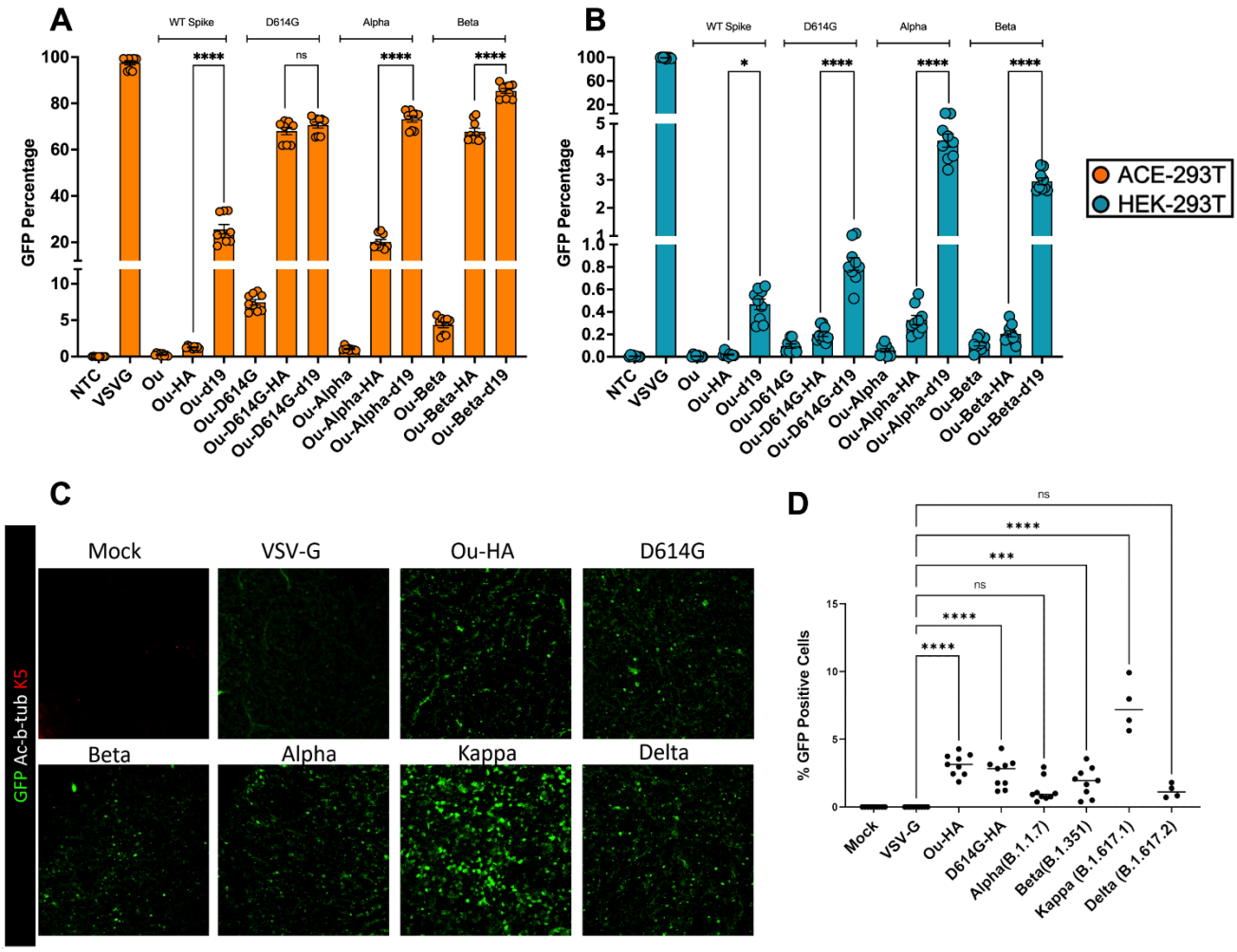


Figure 1. 4

100,000 ACE-293T or HEK-293T cells were transduced with raw viral supernatant containing equal amounts of p24 protein content from a series of pseudotyped lentiviral vectors containing an roUBC-mCitrine reporter cassette. Three days post-transduction, cells were harvested and measured for GFP percentage in (A) ACE2-expressing ACE-293T cells or (B) non-ACE2 expressing HEK-293T cells via flow cytometry. (C) ALI cultures were transduced with raw viral supernatant containing equal amounts of p24 protein content from a series of pseudotyped lentiviral vectors containing an roUBC-mCitrine reporter cassette. Three days post-transduction, cells were harvested and

measured for immunofluorescence of GFP using confocal imaging. Nuclei were determined by DAPI staining and basal stem cells were determined by Keratin5 staining. (D) The number of infected cells in ALI cultures across the pseudotyped lentiviral vectors was quantified by counting GFP positive cells per total number of nuclei. All data sets were compared to a non-transduced control (NTC) and a VSV-G pseudotyped lentiviral vector containing an roUBC-mCitrine reporter cassette. With exception to Kappa and Delta in panels C and D (n=1), data are represented as mean ± SEM of biological triplicates from three experiments. We analyzed statistical significance using a one-way ANOVA followed by multiple paired comparisons for normally distributed data (Tukey test). All statistical tests were two-tailed and a p value of < 0.05 was deemed significant (ns non-significant, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.)

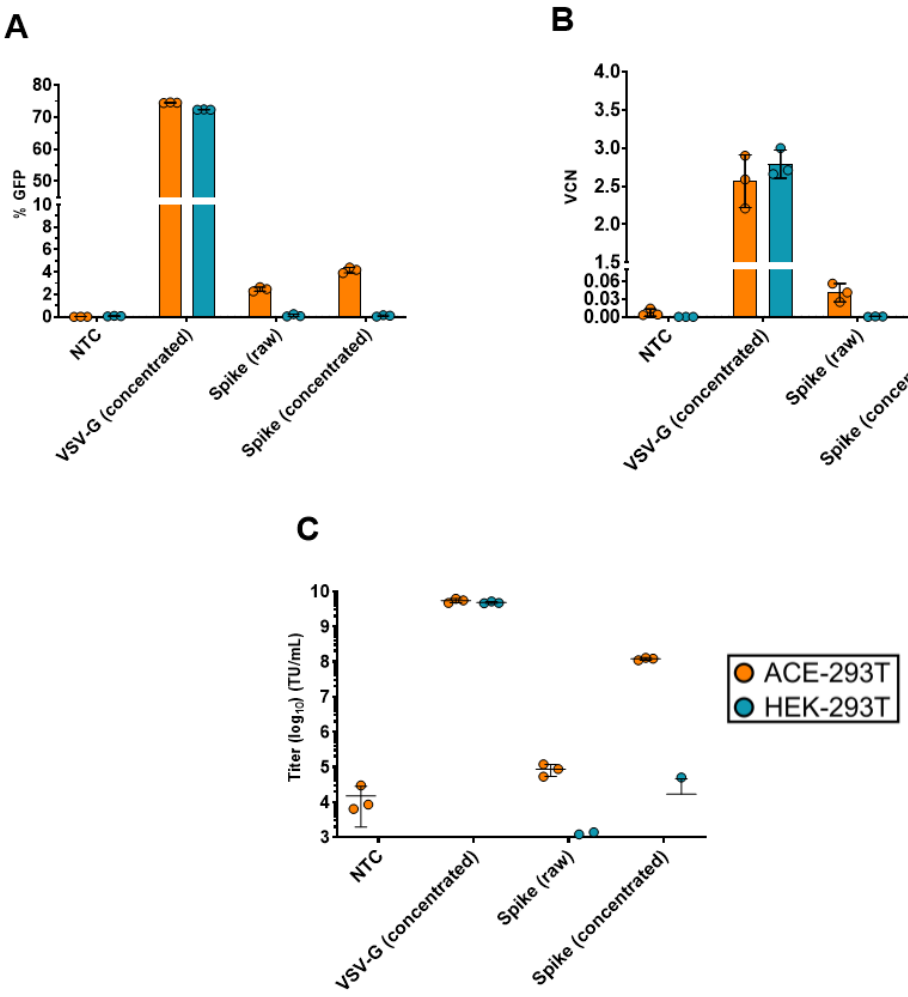


Figure 1. 5

Spike Pseudotype Concentration by Tangential Flow Filtration: (A-C) 100,000 ACE-293T and HEK-293T cells were transduced with a 1:10 dilution of raw virus or a 1:10,000 dilution of TFF concentrated virus from a VSV-G or Spike (Ou-D614G-HA) pseudotyped lentiviral vector containing an MNDU-3-eGFP reporter cassette. Three days post transduction, cells were harvested and measured for %GFP (A) via flow cytometry to assess pseudotype functionality and expression. Panels B and C illustrate the infectivity of spike pseudotype vectors. To measure infectivity, genomic DNA was extracted to measure vector copy number (VCN) (B) by quantifying GFP integrants via QX200 Droplet Digital PCR System. The infectious titer (transducing units/mL) of each vector was

determined (C) from the VCN shown in panel B. All data sets were compared to a non-transduced control (NTC) and a VSV-G pseudotyped lentiviral vector containing an MNDU3-eGFP reporter cassette. Data are represented as mean ± SD of biological triplicates from one experiment.

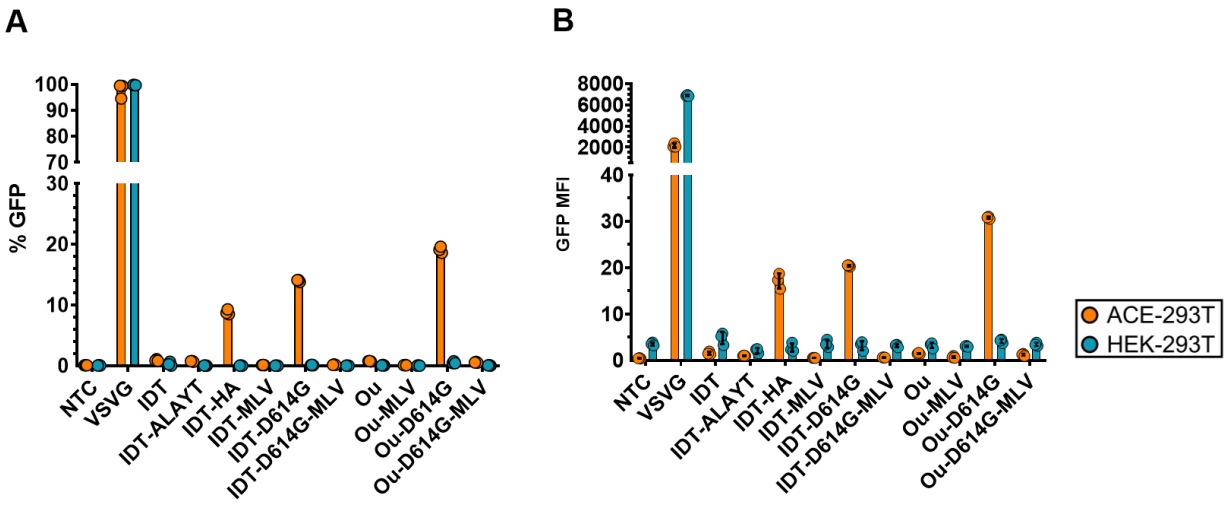
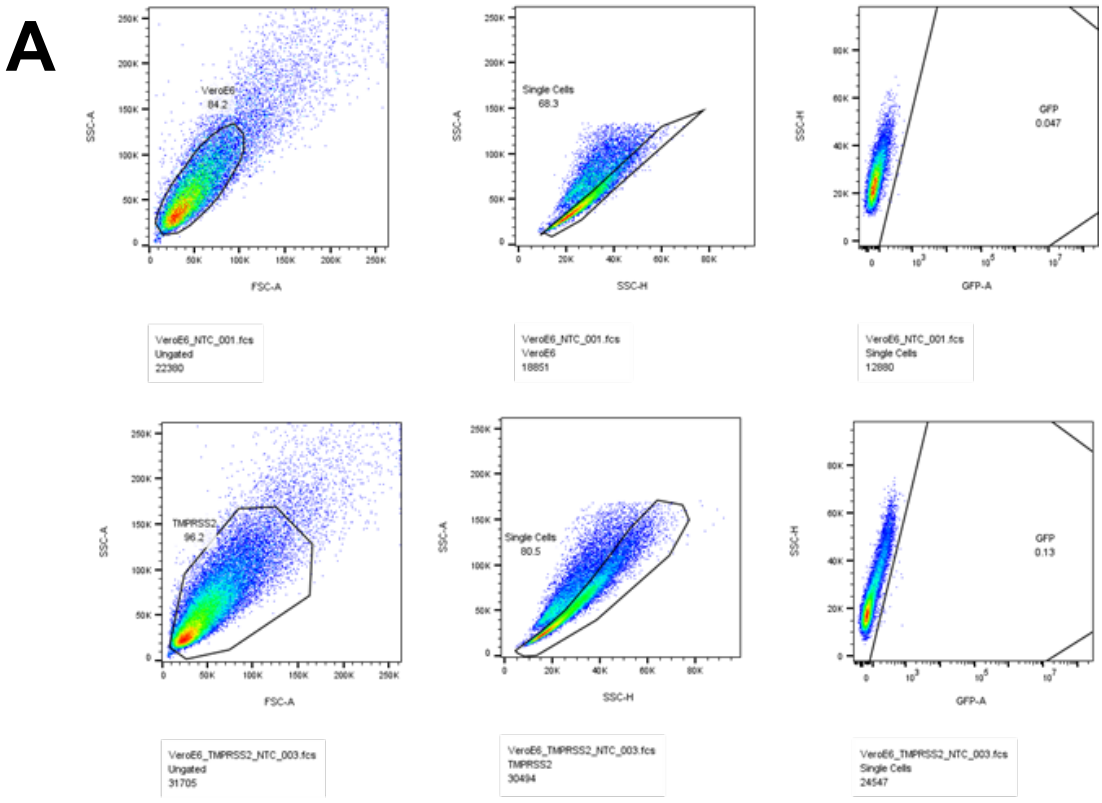


Figure 1. 6

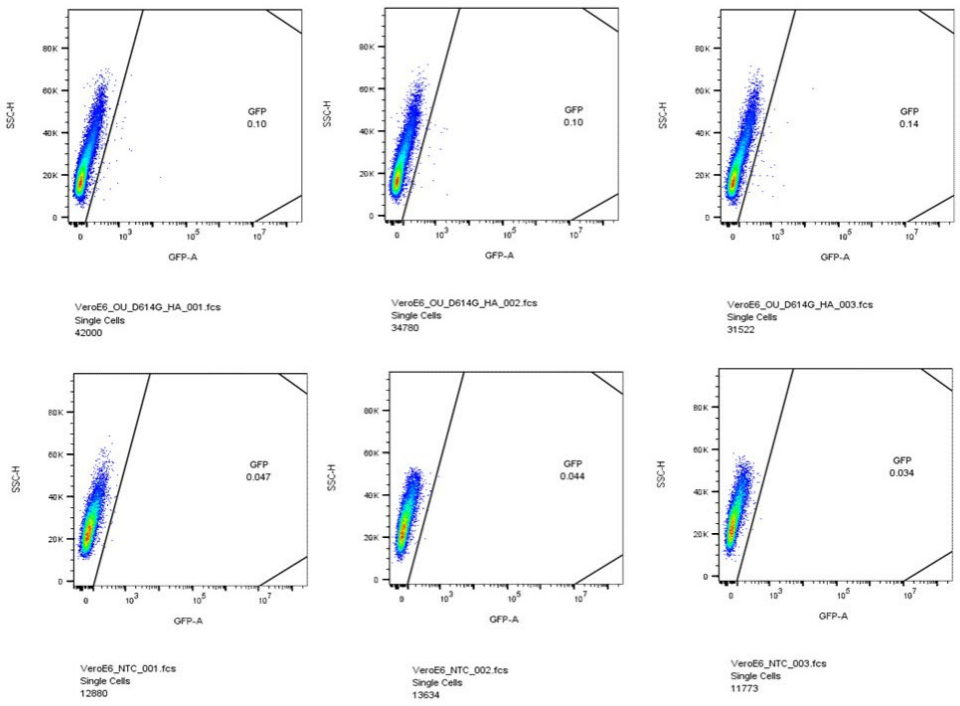
Pseudotype Expression with a UBC-mCitrine Reporter: (A-B) 100,000 ACE2-expressing 293T cells (ACE-293T) and HEK-293T cells were transduced with raw virus at equal amounts of p24 lentiviral particles from a series of pseudotyped lentiviral vectors containing a UBC-mCitrine reporter cassette. Three days post transduction, cells were harvested and measured for (A) %GFP and (B) GFP mean fluorescent intensity (MFI) via flow cytometry to assess pseudotype functionality and expression. All data sets were compared to a non-transduced control (NTC) and a VSV-G pseudotyped lentiviral vector containing an UBC-GFP reporter cassette. Data are represented as mean ± SD of biological triplicates from one experiment.

Figure 1. 7

Representative VeroE6 and VeroE6/TMPRSS2 Flow Plots: 100,000 VeroE6 and VeroE6/TMPRSS2 cells were transduced with raw viral supernatant containing equal amounts of p24 protein content from a series of pseudotyped lentiviral vectors containing an MNDU3-eGFP reporter cassette. Three days post-transduction, cells were harvested and measured for GFP percentage via flow cytometry with a BD LSRII flow cytometer. Cells were analyzed with FlowJo by gating for GFP positive cells against a non-transduced control (A). Panels B and C detail GFP percentage of VeroE6 (B) and VeroE6/TMPRSS2 (C) cells pertaining to Figure 1.3 of the manuscript.



B


SSC-H
GFP-A
GFP 0.10
VeroE6_OU_D614G_HA_001.fcs
Single Cells
42000
GFP 0.10
VeroE6_OU_D614G_HA_002.fcs
Single Cells
34780
GFP 0.14
VeroE6_OU_D614G_HA_003.fcs
Single Cells
31522
GFP 0.047
VeroE6_NTC_001.fcs
Single Cells
12880
GFP 0.044
VeroE6_NTC_002.fcs
Single Cells
13634
GFP 0.034
VeroE6_NTC_003.fcs
Single Cells
11773

C

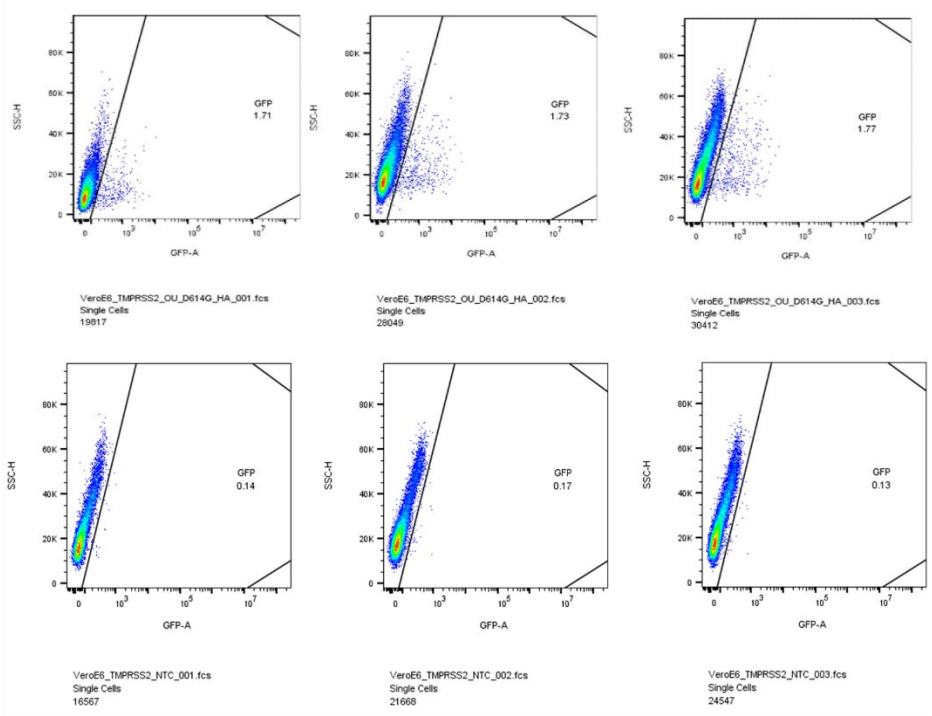
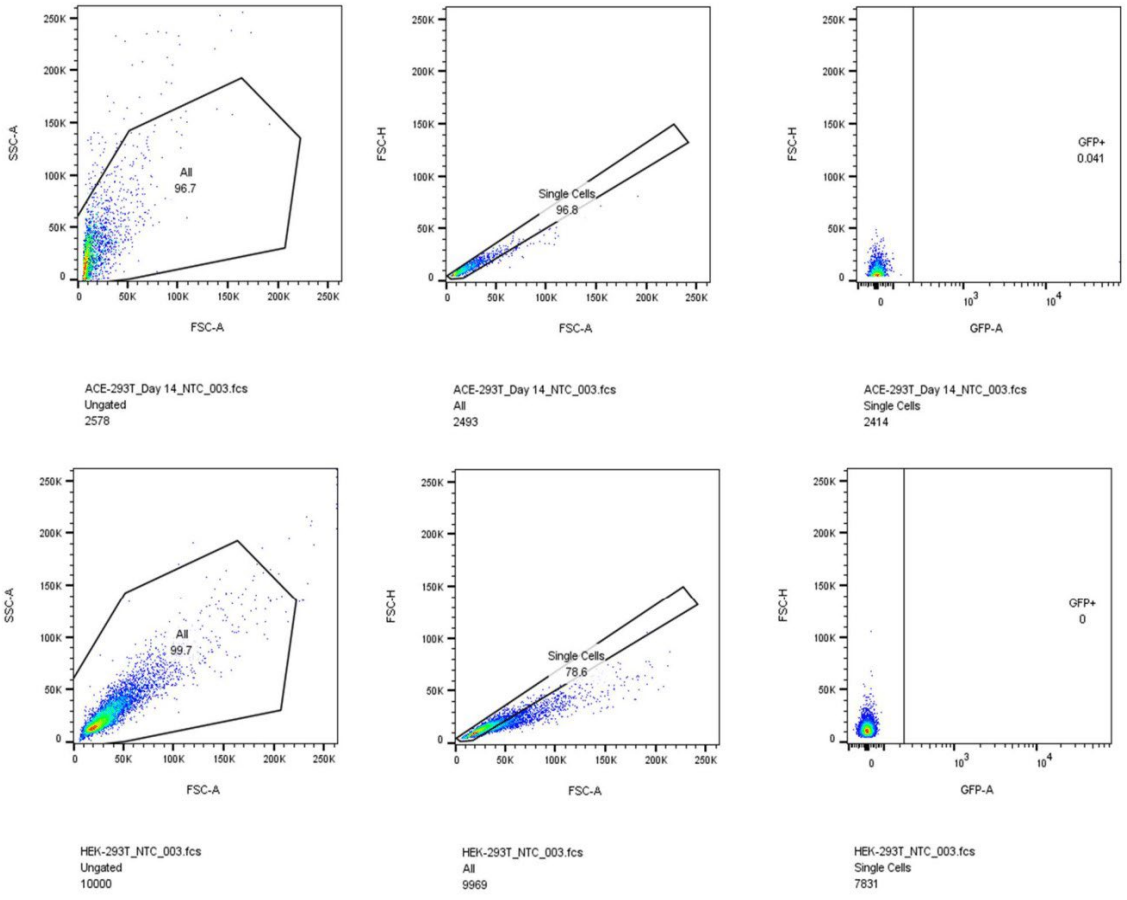

SSC-H
GFP-A
GFP 1.71
VeroE6_TMPRSS2_OU_D614G_HA_001.fcs
Single Cells
19817
GFP 1.73
VeroE6_TMPRSS2_OU_D614G_HA_002.fcs
Single Cells
28049
GFP 1.77
VeroE6_TMPRSS2_OU_D614G_HA_003.fcs
Single Cells
30412
GFP 0.14
VeroE6_TMPRSS2_NTC_001.fcs
Single Cells
16567
GFP 0.17
VeroE6_TMPRSS2_NTC_002.fcs
Single Cells
21668
GFP 0.13
VeroE6_TMPRSS2_NTC_003.fcs
Single Cells
24547

Figure 1. 8

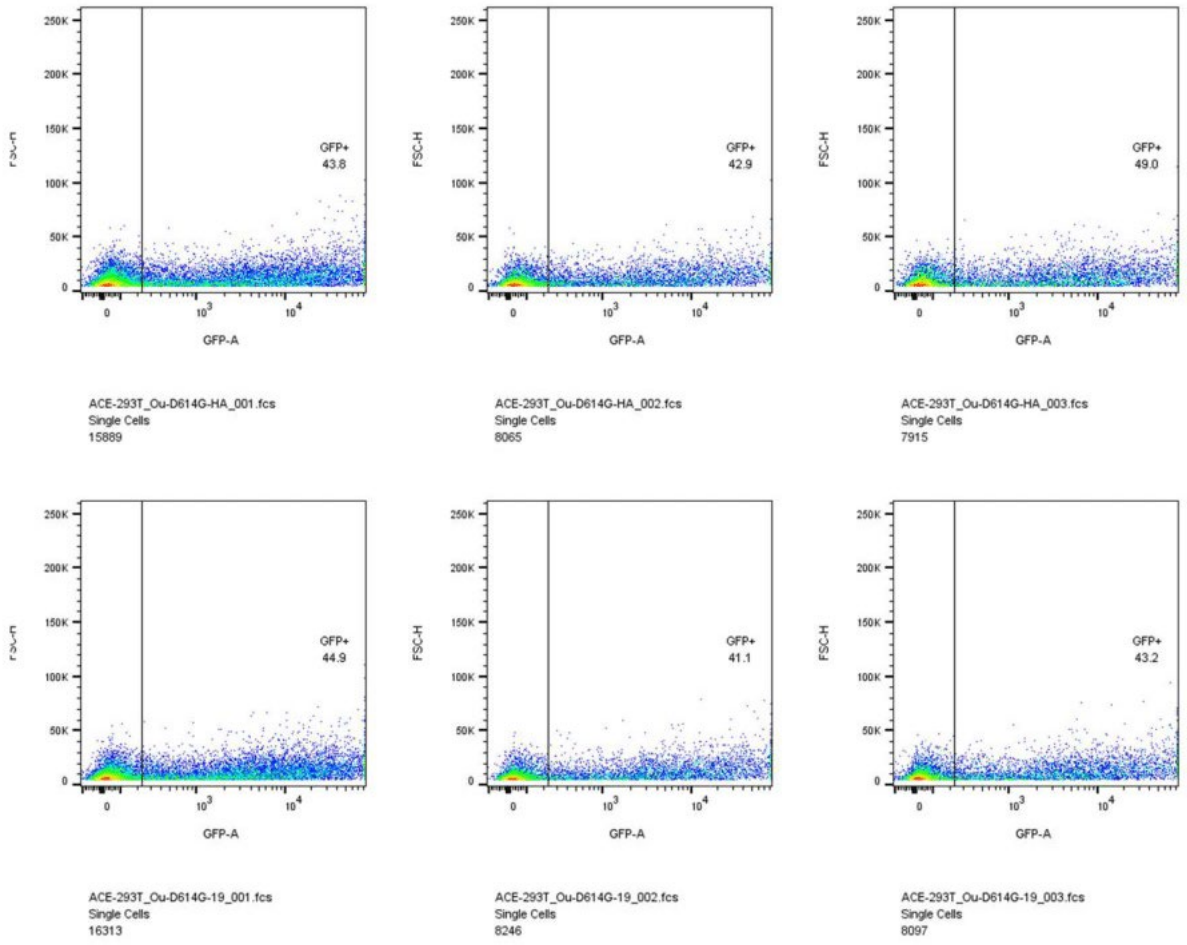
Representative ACE-293T and HEK-293T Flow Plots: 100,000 ACE-293T and HEK-293T cells were transduced with raw viral supernatant containing equal amounts of p24 protein content from a series of pseudotyped lentiviral vectors containing an roUBC-mCitrine reporter cassette. Seven days post-transduction (A-C) or three days post-transduction (D and E), cells were harvested and measured for GFP percentage via flow cytometry with a BD LSRII flow cytometer. Cells were analyzed with FlowJo by gating for GFP positive cells against a non-transduced control (A). Panels B and C detail GFP percentage of ACE-293T (B) and HEK-293T (C) cells pertaining to Figure 1.4 of the manuscript. Panels D and E detail GFP percentage of ACE-293T (D) and HEK-293T (E) cells pertaining to Figure 5 of the manuscript.

A

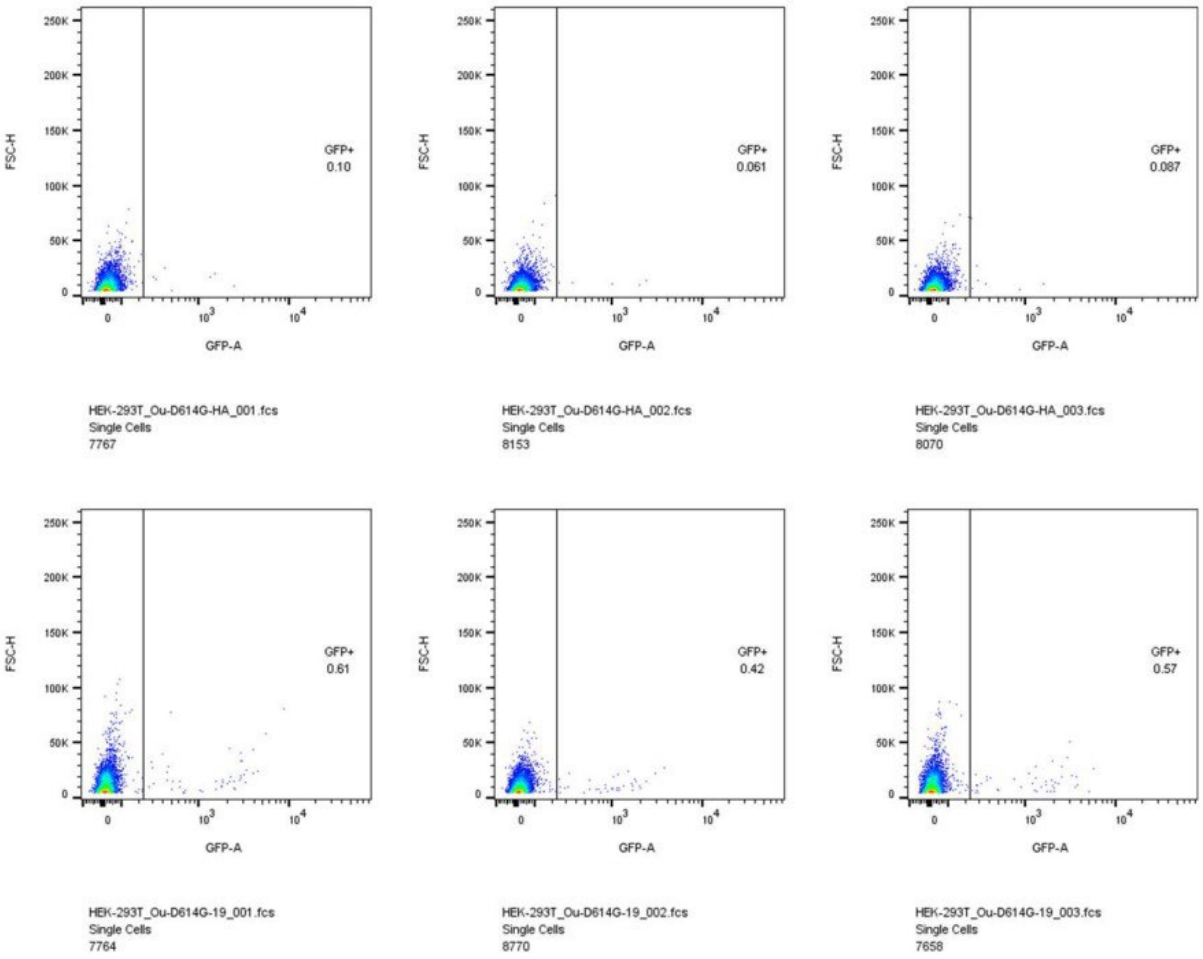
.



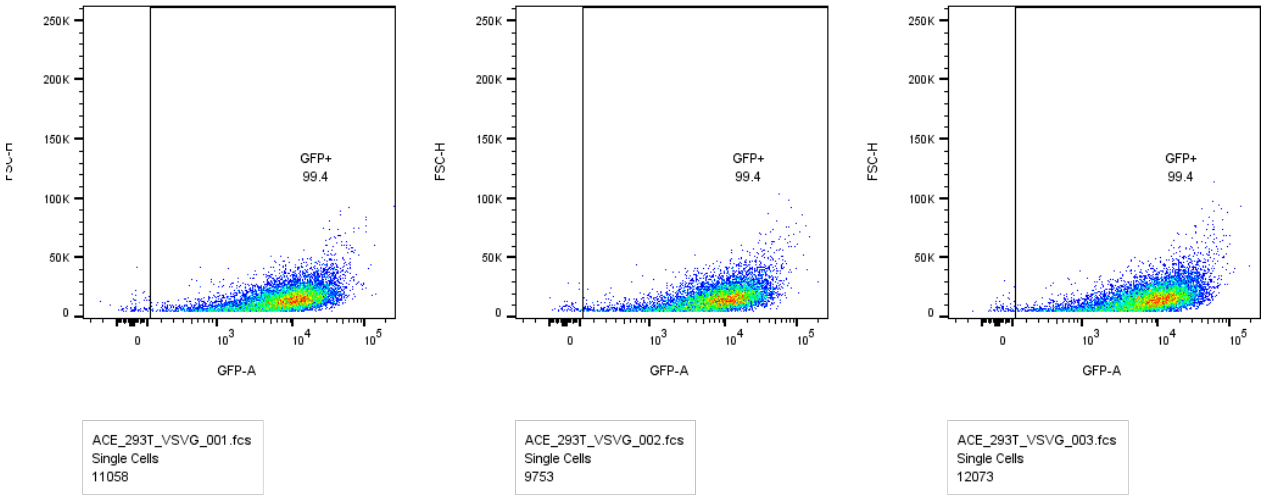
B


GFP+
43.8
GFP-A
ACE-293T_Ou-D614G-HA_001.fcs
Single Cells
15889
GFP+
42.9
GFP-A
ACE-293T_Ou-D614G-HA_002.fcs
Single Cells
8065
GFP+
49.0
GFP-A
ACE-293T_Ou-D614G-HA_003.fcs
Single Cells
7915
GFP+
44.9
GFP-A
ACE-293T_Ou-D614G-19_001.fcs
Single Cells
16313
GFP+
41.1
GFP-A
ACE-293T_Ou-D614G-19_002.fcs
Single Cells
8246
GFP+
43.2
GFP-A
ACE-293T_Ou-D614G-19_003.fcs
Single Cells
8097

C


FSC-H
GFP-A
GFP+
0.10
HEK-293T_Ou-D614G-HA_001.fcs
Single Cells
7767
GFP+
0.061
HEK-293T_Ou-D614G-HA_002.fcs
Single Cells
8153
GFP+
0.087
HEK-293T_Ou-D614G-HA_003.fcs
Single Cells
8070
GFP+
0.61
HEK-293T_Ou-D614G-19_001.fcs
Single Cells
7764
GFP+
0.42
HEK-293T_Ou-D614G-19_002.fcs
Single Cells
8770
GFP+
0.57
HEK-293T_Ou-D614G-19_003.fcs
Single Cells
7658

D


FSC-H
GFP-A
GFP+
99.4
ACE_293T_VSVG_001.fcs
Single Cells
11058
GFP+
99.4
ACE_293T_VSVG_002.fcs
Single Cells
9753
GFP+
99.4
ACE_293T_VSVG_003.fcs
Single Cells
12073

E

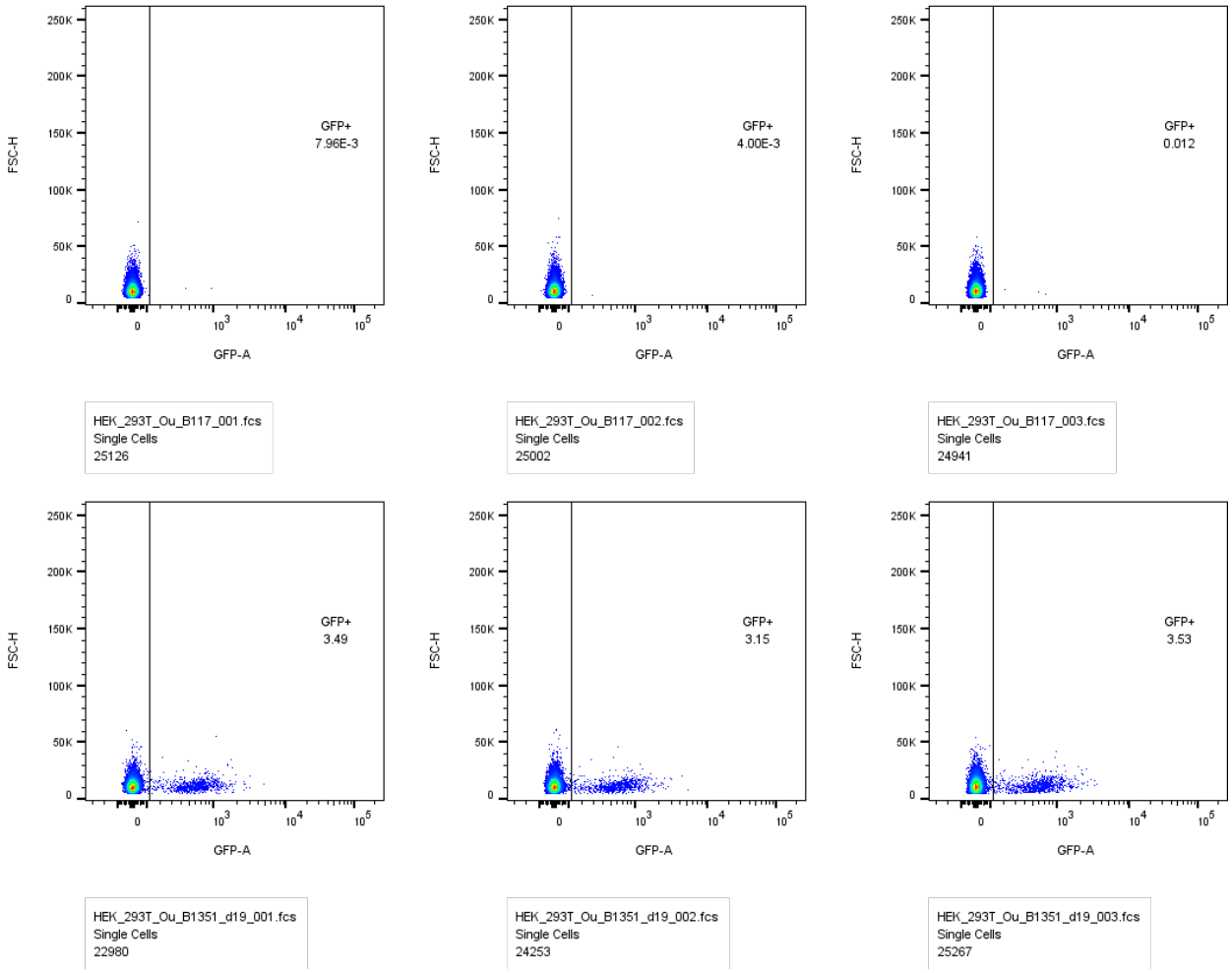

FSC-H
GFP-A
GFP+
7.96E-3
HEK_293T_Ou_B117_001.fcs
Single Cells
25126
GFP+
4.00E-3
HEK_293T_Ou_B117_002.fcs
Single Cells
25002
GFP+
0.012
HEK_293T_Ou_B117_003.fcs
Single Cells
24941
GFP+
3.49
HEK_293T_Ou_B1351_d19_001.fcs
Single Cells
22980
GFP+
3.15
HEK_293T_Ou_B1351_d19_002.fcs
Single Cells
24253
GFP+
3.53
HEK_293T_Ou_B1351_d19_003.fcs
Single Cells
25267

Figure 1. 9

Kappa and Delta Comparison of Influenza Hemagglutinin and d19 Cytoplasmic Tail on S Pseudotype Infectivity: 100,000 ACE-293T or HEK-293T cells were transduced with raw viral supernatant containing equal amounts of p24 protein content from a series of pseudotyped lentiviral vectors containing an roUBC-mCitrine reporter cassette. Three days post-transduction, cells were harvested and measured for GFP percentage in (A) ACE2-expressing ACE-293T cells or non-ACE2 expressing HEK-293T cells via flow cytometry. Fata are represented as mean ± SEM of biological triplicates from one experiment. We analyzed statistical significance using a one-way ANOVA followed by multiple paired comparisons for normally distributed data (Tukey test). All statistical tests were two-tailed and a p value of < 0.05 was deemed significant (ns non-significant, *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.)

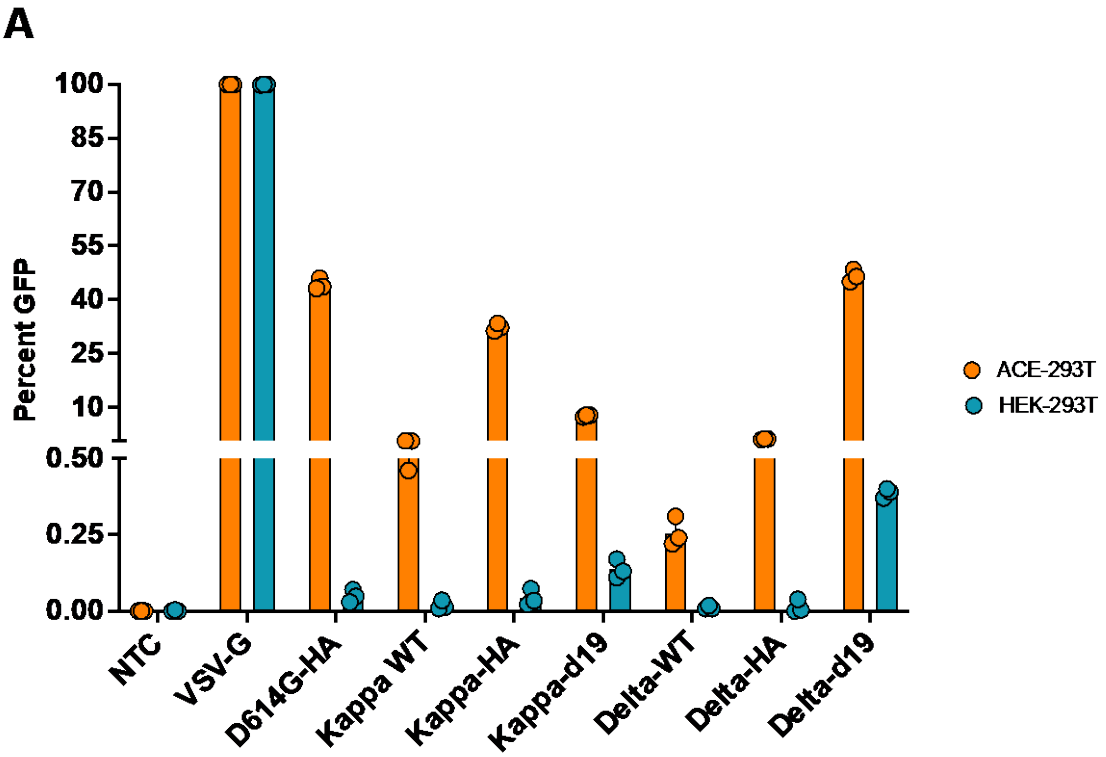
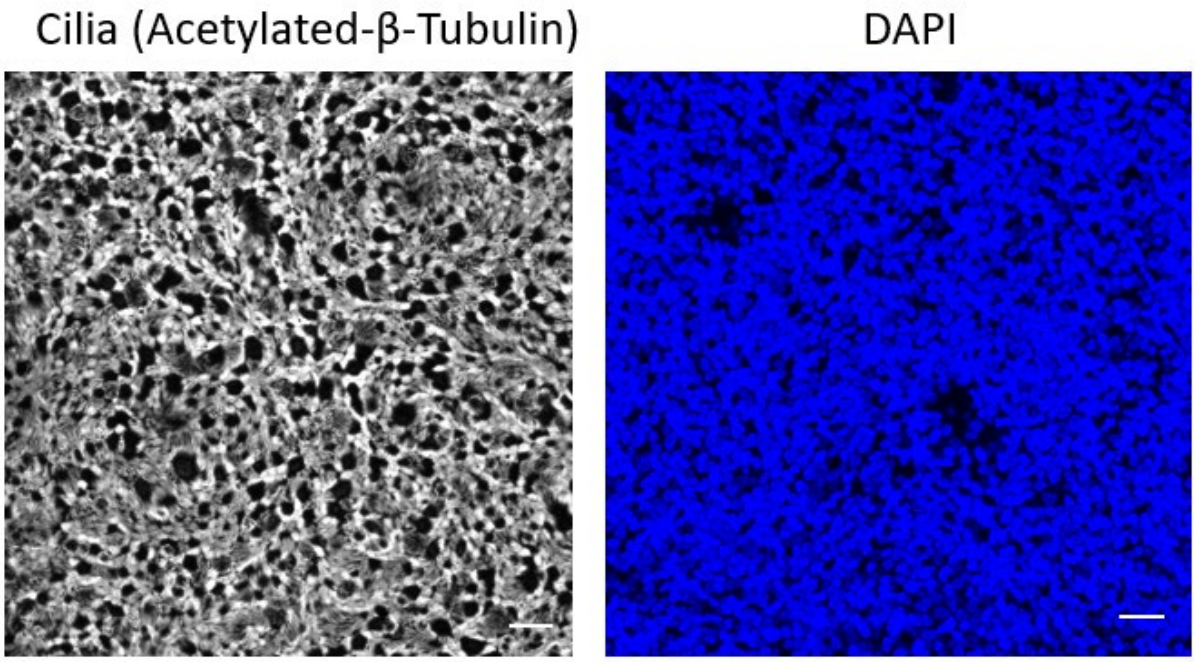


Figure 1. 10

Cilia and DAPI Confocal Imaging of ABSCs: ABSCs three healthy lung transplant donor were seeded at 100,000 cells and cultured for 5 days prior to establish air-liquid interface cultures (ALIs). After ALI establishment, ciliated cells were measured by staining for Acetylated α-Tubulin and nuclei were stained with DAPI (blue).



REFERENCES

1. Blot, V., Lopez-Vergès, S., Breton, M., Pique, C., Berlioz-Torrent, C., & Grange, M. P. (2006). The conserved dileucine- and tyrosine-based motifs in MLV and MPMV envelope glycoproteins are both important to regulate a common Env intracellular trafficking. Retrovirology, 3, 62. https://doi.org/10.1186/1742-4690-3-62
2. B. Coutard, C. Valle, X. de Lamballerie, B. Canard, N. G. Seidah, E. Decroly, The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. Antiviral Res. 176, 104742 (2020). doi:10.1016/j.antiviral.2020.104742pmid:32057769
3. Bain W, Lee JS, Watson AM, Stitt-Fischer MS. Practical Guidelines for Collection, Manipulation and Inactivation of SARS-CoV-2 and COVID-19 Clinical Specimens. Curr Protoc Cytom. 2020 Jun;93(1):e77. doi: 10.1002/cpcy.77. PMID: 32502333; PMCID: PMC7300551.
4. Bestle, Dorothea et al. “TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells.” *Life science alliance* vol. 3,9 e202000786. 23 Jul. 2020, doi:10.26508/lsa.202000786
5. Christodoulopoulos I, Cannon PM. Sequences in the cytoplasmic tail of the gibbon ape leukemia virus envelope protein that prevent its incorporation into lentivirus vectors. J Virol. 2001 May;75(9):4129-38. doi: 10.1128/JVI.75.9.4129-4138.2001. PMID: 11287562; PMCID: PMC114158.
6. Cooper, Aaron R. et al. “Highly Efficient Large-Scale Lentiviral Vector Concentration by Tandem Tangential Flow Filtration.” Journal of Virological Methods 177.1 (2011): 1–9. doi.org/10.1016/j.jviromet.2011.06.019.

7. Crawford, K. H. D. et al. (2020). Protocol and Reagents for Pseudotyping Lentiviral Particles with SARS-CoV-2 Spike Protein for Neutralization Assays. Viruses, 12(5), 513.
8. Dai B, Xiao L, Bryson PD, Fang J, Wang P. PD-1/PD-L1 blockade can enhance HIV- 1 Gag-specific T cell immunity elicited by dendritic cell-directed lentiviral vaccines. Mol Ther **2012**;20(9):1800-9.
9. Dieterle ME, Haslwanter D, Bortz RH 3rd, Wirchnianski AS, Lasso G, Vergnolle O, Abbasi SA, Fels JM, Laudermilch E, Florez C, Mengotto A, Kimmel D, Malonis RJ, Georgiev G, Quiroz J, Barnhill J, Pirofski LA, Daily JP, Dye JM, Lai JR, Herbert AS, Chandran K, Jangra RK. A Replication-Competent Vesicular Stomatitis Virus for Studies of SARS-CoV-2 Spike-Mediated Cell Entry and Its Inhibition. Cell Host Microbe. 2020 Sep 9;28(3):486-496.e6. doi: 10.1016/j.chom.2020.06.020. Epub 2020 Jul 3. PMID: 32738193; PMCID: PMC7332447.
10. Du, R.; Cui, Q.; Rong, L. Competitive Cooperation of Hemagglutinin and Neuraminidase during Influenza A Virus Entry. Viruses 2019, 11, 458.
11. Dull, Tom, et al. “A Third-Generation Lentivirus Vector with a Conditional Packaging System.” Journal of Virology, vol. 72, no. 11, American Society for Microbiology, Nov. 1998, pp. 8463–71. doi:10.1128/jvi.72.11.8463-8471.1998.
12. Girard-Gagnepain A, Amirache F, Costa C, Lévy C, Frecha C, Fusil F, Nègre D, Lavillette D, Cosset FL, Verhoeyen E. Baboon envelope pseudotyped LVs outperform VSV-G-LVs for gene transfer into early-cytokine-stimulated and resting HSCs. Blood. 2014 Aug 21;124(8):1221-31. doi: 10.1182/blood-2014-02-558163. Epub 2014 Jun 20. PMID: 24951430.

13. Han, J., Tam, K., Ma, F., Tam, C., Aleshe, B., Wang, X., Quintos, J. P., Morselli, M., Pellegrini, M., Hollis, R. P., & Kohn, D. B. (2021). β-Globin Lentiviral Vectors Have Reduced Titers due to Incomplete Vector RNA Genomes and Lowered Virion Production. Stem Cell Reports, 16(1), 198–211. https://doi.org/10.1016/j.stemcr.2020.10.007
14. Havranek KE, Jimenez AR, Acciani MD, Lay Mendoza MF, Reyes Ballista JM, Diaz DA, Brindley MA. SARS-CoV-2 Spike Alterations Enhance Pseudoparticle Titers and Replication-Competent VSV-SARS-CoV-2 Virus. Viruses. 2020 Dec 18;12(12):1465. doi: 10.3390/v12121465. PMID: 33353101; PMCID: PMC7767099.
15. Johnson, Marc C., et al. “Optimized Pseudotyping Conditions for the SARS-COV-2 Spike Glycoprotein.” Journal of Virology, edited by Viviana Simon, vol. 94, no. 21, American Society for Microbiology, Oct. 2020. doi:10.1128/jvi.01062-20.
16. Kubo Y, Tominaga C, Yoshii H, Kamiyama H, Mitani C, Amanuma H, Yamamoto N. Characterization of R peptide of murine leukemia virus envelope glycoproteins in syncytium formation and entry. Arch Virol. 2007;152(12):2169-82. doi: 10.1007/s00705-007-1054-6. Epub 2007 Sep 14. PMID: 17851730.
17. Logan AC, Nightingale SJ, Haas DL, Cho GJ, Pepper KA, Kohn DB. Factors influencing the titer and infectivity of lentiviral vectors. Hum Gene Ther. 2004 Oct;15(10):976-88. doi: 10.1089/hum.2004.15.976. PMID: 15585113.
18. Löving R, Li K, Wallin M, Sjöberg M, Garoff H. R-Peptide cleavage potentiates fusion-controlling isomerization of the intersubunit disulfide in Moloney murine

leukemia virus Env. J Virol. 2008 Mar;82(5):2594-7. doi: 10.1128/JVI.02039-07. Epub 2007 Dec 19. PMID: 18094170; PMCID: PMC2258935.

19. Duvergé A, Negroni M. Pseudotyping Lentiviral Vectors: When the Clothes Make the Virus. Viruses. 2020 Nov 16;12(11):1311. doi: 10.3390/v12111311. PMID: 33207797; PMCID: PMC7697029.
20. M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T. S. Schiergens, G. Herrler, N.-H. Wu, A. Nitsche, M. A. Müller, C. Drosten, S. Pöhlmann, SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell 181, 271–280.e8 (2020). doi:10.1016/j.cell.2020.02.052pmid:32142651
21. McBride, C. E., Li, J., & Machamer, C. E. (2006). The Cytoplasmic Tail of the Severe Acute Respiratory Syndrome Coronavirus Spike Protein Contains a Novel Endoplasmic Reticulum Retrieval Signal That Binds COPI and Promotes Interaction with Membrane Protein. Journal of Virology, 81(5), 2418–2428.
22. Naldini L, Blömer U, Gallay P, Ory D, Mulligan R, Gage FH, Verma IM, Trono D. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science. 1996 Apr 12;272(5259):263-7. doi: 10.1126/science.272.5259.263. PMID: 8602510.
23. N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G. F. Gao, W. Tan; China Novel Coronavirus Investigating and Research Team, A Novel Coronavirus from Patients with Pneumonia in China, 2019. N. Engl. J. Med. 382, 727–733 (2020). doi:10.1056/NEJMoa2001017pmid:31978945

24. Nie, J., Li, Q., Wu, J., Zhao, C., Hao, H., Liu, H., Zhang, L., Nie, L., Qin, H., Wang, M., Lu, Q., Li, X., Sun, Q., Liu, J., Fan, C., Huang, W., Xu, M., & Wang, Y. (2020). Quantification of SARS-CoV-2 neutralizing antibody by a pseudotyped virus-based assay. Nature Protocols, 15(11), 3699–3715. https://doi.org/10.1038/s41596-020-0394-5
25. Ou, X., Liu, Y., Lei, X. et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nat Commun 11, 1620 (2020). https://doi.org/10.1038/s41467-020-15562-9
26. Poland, Gregory A., et al. “SARS-CoV-2 Immunity: Review and Applications to Phase 3 Vaccine Candidates.” The Lancet, vol. 396, no. 10262, Elsevier BV, Nov. 2020, pp. 1595–606. Crossref, doi:10.1016/s0140-6736(20)32137-1.
27. Sandrin V, Cosset FL. Intracellular versus cell surface assembly of retroviral pseudotypes is determined by the cellular localization of the viral glycoprotein, its capacity to interact with Gag, and the expression of the Nef protein. J Biol Chem. 2006 Jan 6;281(1):528-42. doi: 10.1074/jbc.M506070200. Epub 2005 Sep 28. PMID: 16195228.
28. Sandrin V, Boson B, Salmon P, Gay W, Nègre D, Le Grand R, Trono D, Cosset FL. Lentiviral vectorS-pseudotyped with a modified RD114 envelope glycoprotein show increased stability in sera and augmented transduction of primary lymphocytes and CD34+ cells derived from human and nonhuman primates. Blood. 2002 Aug 1;100(3):823-32. doi: 10.1182/blood-2001-11-0042. PMID: 12130492.

29. Schmidt F, Weisblum Y, Muecksch F, Hoffmann HH, Michailidis E, Lorenzi JCC, Mendoza P, Rutkowska M, Bednarski E, Gaebler C, Agudelo M, Cho A, Wang Z, Gazumyan A, Cipolla M, Caskey M, Robbiani DF, Nussenzweig MC, Rice CM, Hatziioannou T, Bieniasz PD. Measuring SARS-CoV-2 neutralizing antibody activity using pseudotyped and chimeric viruses. J Exp Med. 2020 Nov 2;217(11):e20201181. doi: 10.1084/jem.20201181. PMID: 32692348; PMCID: PMC7372514.
30. Schnierle BS, Stitz J, Bosch V, Nocken F, Merget-Millitzer H, Engelstädter M, Kurth R, Groner B, Cichutek K. Pseudotyping of murine leukemia virus with the envelope glycoproteins of HIV generates a retroviral vector with specificity of infection for CD4-expressing cells. Proc Natl Acad Sci U S A. 1997 Aug 5;94(16):8640-5. doi: 10.1073/pnas.94.16.8640. PMID: 9238030; PMCID: PMC23056.
31. Stitz J, Buchholz CJ, Engelstadter M, Uckert W, Bloemer U, Schmitt I, Cichutek K. Lentiviral vectorS-pseudotyped with envelope glycoproteins derived from gibbon ape leukemia virus and murine leukemia virus 10A1. Virology. 2000;273:16–20.
32. Tandon, R., Mitra, D., Sharma, P. *et al.* Effective screening of SARS-CoV-2 neutralizing antibodies in patient serum using lentivirus particleS-pseudotyped with SARS-CoV-2 spike glycoprotein. *Sci Rep* **10,** 19076 (2020). https://doi.org/10.1038/s41598-020-76135-w
33. Tomás, Hélio A et al. “Improved GaLV-TR Glycoproteins to Pseudotype Lentiviral Vectors: Impact of Viral Protease Activity in the Production of LV Pseudotypes.” *Molecular therapy. Methods & clinical development* vol. 15 1-8. 14 Aug. 2019, doi:10.1016/j.omtm.2019.08.001

34. Tortorici, M. A., & Veesler, D. (2019). Structural insights into coronavirus entry. In Advances in Virus Research (pp. 93–116). Elsevier. https://doi.org/10.1016/bs.aivir.2019.08.002
35. Turoňová, B., Sikora, M., Schürmann, C., Hagen, W. J. H., Welsch, S., Blanc, F. E. C., von Bülow, S., Gecht, M., Bagola, K., Hörner, C., van Zandbergen, G., Landry, J., de Azevedo, N. T. D., Mosalaganti, S., Schwarz, A., Covino, R., Mühlebach, M. D., Hummer, G., Krijnse Locker, J., & Beck, M. (2020). In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. Science, 370(6513), 203–208. https://doi.org/10.1126/science.abd5223
36. V'kovski, Philip, et al. "Coronavirus Biology and Replication: Implications for SARS-CoV-2." Nature Reviews Microbiology, Springer Science and Business Media LLC, Oct. 2020. doi:10.1038/s41579-020-00468-6.
37. Verhoeyen E, Cosset FL. Surface-engineering of lentiviral vectors. J Gene Med. 2004 Feb;6 Suppl 1:S83-94. doi: 10.1002/jgm.494. PMID: 14978753.
38. Yurkovetskiy, Leonid, et al. "Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant." Cell, vol. 183, no. 3, Elsevier BV, Oct. 2020, pp. 739-751.e8. doi:10.1016/j.cell.2020.09.032.
39. Marquez Loza LI, Cooney AL, Dong Q, Randak CO, Rivella S, Sinn PL, McCray PB Jr. Increased CFTR expression and function from an optimized lentiviral vector for cystic fibrosis gene therapy. Mol Ther Methods Clin Dev. 2021 Feb 27;21:94-106. doi: 10.1016/j.omtm.2021.02.020. PMID: 33768133; PMCID: PMC7973238.
40. Mauro VP, Chappell SA. A critical analysis of codon optimization in human therapeutics. Trends Mol Med. 2014 Nov;20(11):604-13. doi:

10.1016/j.molmed.2014.09.003. Epub 2014 Sep 25. PMID: 25263172; PMCID: PMC4253638.

41. Matsuyama S, Nao N, Shirato K, Kawase M, Saito S, Takayama I, Nagata N, Sekizuka T, Katoh H, Kato F, Sakata M, Tahara M, Kutsuna S, Ohmagari N, Kuroda M, Suzuki T, Kageyama T, Takeda M. Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. Proc Natl Acad Sci U S A. 2020 Mar 31;117(13):7001-7003. doi: 10.1073/pnas.2002589117. Epub 2020 Mar 12. PMID: 32165541; PMCID: PMC7132130.
42. Wibmer, C.K., Ayres, F., Hermanus, T. et al. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. Nat Med (2021). https://doi.org/10.1038/s41591-021-01285-x

Chapter 2: Lentiviral Vectors for Precise Expression to Treat X-Linked Lymphoproliferative Disease

ABSTRACT

X-Linked Lymphoproliferative Disease (XLP1) is caused by loss of function mutations in the SH2 domain–containing protein 1A (*SH2D1A*) gene, which encodes for SLAM family signaling adaptor, SLAM-associated protein (SAP). A regulated lentiviral vector (LV), expressing at therapeutic levels within target T, NK, and NKT cell populations, is necessary for an autologous hematopoietic stem cell gene therapy. We thus experimentally elucidated genomic elements that regulate the *SH2D1A* gene and rationally designed LVs (XLP1-SMART-LVs) to emulate the lineage and temporal specific control of SAP. We identified 34 regulatory elements of *SH2D1A* and screened them for their on-target enhancer activity in T, NK, and NKT cells and their off-target enhancer activity in B cell and myeloid populations. In combination, three enhancer elements increased SAP promoter expression up to 4-fold in on-target populations *in vitro*. NSG-Tg(Hu-IL15) xenograft studies with XLP1-SMART-LVs demonstrated up to 7-fold greater expression in on-target cells over a control EFS-LV, with no off-target expression. The XLP1-SMART-LVs exhibited stage-specific T- and NK-cell expression in peripheral blood, bone marrow, spleen, and thymic tissues (mimicking temporal expression patterns of SAP). XLP1 patient CD8+ T cells or BM CD34+ cells transduced with the XLP1-SMART-LVs restored restimulation-induced cell death and NK cytotoxicity to wildtype levels, respectively. These data demonstrate that it is feasible to create a lineage and temporal specific lentiviral vector for the purpose of restoring the XLP1 phenotype by gene therapy.

INTRODUCTION

X-linked lymphoproliferative disease (**XLP1**), also known as Duncan Disease, is an inborn error of immunity caused by mutations in the *SH2D1A* gene, affecting 1 in 1 million males.[1-5] The *SH2D1A* gene encodes the SLAM-associated protein (SAP), an adaptor molecule involved in the signaling of immune cell receptors of the SLAM family.[1-5] SAP binds to the intracellular domain of SLAM family signaling receptors, and supports the activation/inhibition of immune cell signaling.[2,4,6-8] SAP mRNA and protein expression are predominately expressed in human thymocytes, T cells, NK cells, and NKT cells.[2,4,6-12] Patients with XLP1 suffer from impairments in CD4+ T cell function, CD8+ T cell cytotoxicity, NK cell cytotoxicity, plasma cell and memory B cell generation, and NKT cell development.[13-16]

In over 90% of XLP1 cases, Epstein-Barr virus (EBV) primary infection is the major cause for clinical presentations of the disease.[1,2,17] After EBV infection, XLP1 patients mount a dysregulated immune response, with nearly 60% of patients developing hemophagocytic lymphohistiocytosis (HLH).[1,8] HLH treatment is highly immune suppressive and can be significantly toxic. As such, the mortality associated with HLH presentation is greater than 60%.[1] Those that survive the EBV infection may develop malignant lymphoma, hypogammaglobulinemia, and lymphoproliferation, and are thus treated with continuous immunoglobulin replacement therapy (IRT) and immune suppression.[9] The expense and inconvenience of life-long IRT, with administration requiring IV or SQ injections, are drawbacks, but are accepted due to the important clinical benefits IRT confers. However, a more effective treatment for XLP1

remains an unmet need, and – given the severe nature of the disease – an autologous hematopoietic stem cell (HSC) transplantation is a viable approach.[11] An autologous HSC transplant treats a patient's own HSCs to integrate a stable copy of the *SH2D1A* *ex-vivo* with a lentiviral vector (LV).[11] This method of gene therapy may provide the same benefit as an allogeneic transplant, while eliminating any risks of graft rejection or graft versus host disease, since each patient serves as their own donor. A successful treatment by autologous HSC transplantation requires an LV capable of transducing a functional copy of *SH2D1A* into long-lived multipotent HSCs but expressing at therapeutic levels within specific leukocyte lineages.[1,11]

A lentiviral vector was developed previously for gene therapy of XLP1 through LV-mediated gene transfer into autologous HSCs.[11] However, this LV did not utilize the endogenous regulatory elements of the encoded transgene, like many of the current clinical lentiviral vectors, thereby it did not recapitulate the precise expression pattern of the native *SH2D1A* gene.[11,18] Instead, a ubiquitously expressed promoter, such as the elongation factor 1 α short promoter (EFS), was used. For most gene targets, the endogenous promoter *alone* is insufficient to drive proper physiological expression and regulation of the gene, as additional regulatory elements such as distant enhancer and silencers are also required. Since SAP expression is tightly regulated within T, NK, and NKT cells, gene expression in off-target HSC populations may pose safety concerns including skewing of hematopoietic potential and improper signaling of factors associated with lymphocyte development and function. [11,19,20]

To overcome the issues of ectopic and dysregulated expression, other groups have utilized non-endogenous regulatory elements to mimic endogenous patterns of expression.[21,22] In contrast, our group has advanced an approach utilizing bioinformatics-assisted design to construct lentiviral vectors driven by the endogenous promoter and regulatory elements necessary to treat inborn errors of immunity such as X-linked Chronic Granulomatous Disease (X-CGD), and Wiskott-Aldrich Syndrome (WAS). [23] The approach leveraged the publicly available bioinformatic tool GeneHancer database, which pools data from ENCODE, Ensembl, FANTOM and VISTA. By integrating data including histone modification, chromatin accessibility, bound transcription factors and Hi-C interactions across multiple datasets, we identified various putative enhancer regions of each endogenous gene which assist with regulating their target gene. Using this methodology, we recently described a lentiviral vector with a superior lineage-specific expression pattern compared to a LV currently in clinical trials for X-CGD that uses non-endogenous regulatory elements from other myeloid lineage genes.[23]

We have since optimized this methodology for the efficient identification and testing of endogenous regulatory elements to generate rationally designed bioinformatics-assisted (SMART) lentiviruses for the treatment of XLP1. These LVs contain regulatory elements of the *SH2D1A* gene to achieve lineage- and temporal-specific expression of SAP after transduction of autologous CD34+ HSPCs in cell types needed for maximal therapeutic benefit, with minimal off-target expression in inappropriate cell types. These LV may achieve functional restoration of humoral and cytotoxic defects in XLP1 patients through gene therapy.

MATERIALS AND METHODS

Elucidation of putative enhancer elements: Genomic regions containing putative regulatory elements of the *SH2D1A* gene were compiled using data from ENCODE, Ensembl, FANTOM, and VISTA. Functional boundaries of the putative enhancer elements were defined using lineage-specific DNase I accessibility, transcription factor binding, epigenetic histone modification, and vertebrate sequence conservation. Primers were then designed to amplify these putative enhancer regions from human genomic DNA for downstream cloning into plasmid lentiviral transgene cassettes.

Vector Packaging and Titration: Lentiviruses were packaged by transient transfection of PKR -/- 293T cells with fixed amounts of HIV Gag/Pol, Rev, and VSV-G envelope expression plasmids and equimolar amounts of transfer plasmid using TransIT-293 (Mirus Bio; Madison, WI) as described in Cooper et al. (2011).[25,37] Viral supernatants were then directly used for titer determination or concentrated by tangential flow filtration. To titer the lentivirus, 1 x 10^5 HT-29 cells per sample were plated in 2mL of culture medium in 6-well plates (#3516; Corning Inc.; Corning, NY). 24 hours after plating, cells were transduced with a 1:10 dilution of viral supernatant in 1mL of culture medium. 24 hours after transduction, culture medium was refreshed on all wells. 72 hours after transduction, cells were harvested to determine VCNs by ddPCR. Vector titer (TU/mL) was calculated as TU=VCN x (cell count at day of transduction) x virus dilution. Cell counts were measured with a Vi-CELL XR automated cell counter (Beckman Coulter; Brea, CA).

Lentiviral Vector Transduction: Primary human T, NK, and NKT cells (5 x 10^4 per sample) isolated from peripheral blood mononuclear cells were plated on retronectin (20 ug/ml) (Takara Bio Inc.; Kusatsu, Shiga, Japan)-coated plates with XVIVO15 medium (Lonza Biosciences; Basel, Switzerland) supplemented with 1mg/mL poloxamer synperonic F108 (Kolliphor P338; BASF Pharma; Ludwigshafen, Germany) added for 24 hours during lentiviral transduction. Cell counts were measured with a Vi-CELL XR automated cell counter. For transduction of CD34+ cells, cells were prestimulated with 50 ng/mL each of human stem cell factor (hSCF), human thrombopoietin (hTPO), and human FMS-like tyrosine kinase 3 ligand (hFlt3-L) (Peprotech; Rocky Hill, NJ) for 24 hours before lentiviral transduction. During LV transduction of CD34+ cells, 10uM prostaglandin E2 (PGE2) was also added for 24 hours. LV supernatant (raw or concentrated) was added to the culture medium for 24 hours at various concentrations to achieve the necessary vector copy number (VCN).

Droplet Digital PCR for VCN and Titer (TU/mL) Quantification: Genomic DNA from transduced cells was extracted using PureLink Genomic DNA Mini Kit (K182002; Invitrogen, Waltham, MA). VCN was calculated by using the vector HIV-1 PSI gene primers (oPAF-PSI, oPAR-PSI) and probes (oPAP-PSI) and an endogenous human diploid gene control (SCD4 (Human Syndecan 4) primers (oPAF-SDC4 and oPAR-SDC4) and probe (oPAP-SDC4) as a reference [See Supplemental Table]. Reaction mixtures of 22uL volume, comprising 1 × droplet digital (dd)PCR Master Mix (#1863010; Bio-Rad; Hercules, CA), 400 nmol/L primers and 100 nmol/L probe for each set, 40 U DraI (R0129S; New England Biolabs; Ipswich, MA) and 30– 100 g of the gDNA to study, were prepared and incubated at 37°C for 1 hour. Droplet generation was

performed as described in Hindson et al. with 20 uL of each reaction mixture.[38] The droplet emulsion was then transferred with a multichannel pipet to a 96-well twin.tec® real-time PCR Plates (Eppendorf; Hamburg, Germany), heat sealed with foil, and amplified in a conventional thermal cycler (T100 Thermal Cycler; Bio-Rad; Hercules, CA). Thermal cycling conditions consisted of 95°C 10 min, (94°C 30 s and 60°C 1 min) (55 cycles), 98°C 10 min (1 cycle) and 12°C hold. After PCR, the 96-well plate was transferred to a droplet reader (Bio-Rad; Hercules, CA). Acquisition and analysis of the ddPCR data was performed with the QuantaSoft software (Bio-Rad; Hercules, CA), provided with the droplet reader. Vector Titer (TU/mL) was calculated as TU=VCN *x* (cell count at day of transduction) *x* virus dilution.

NSG-Tg(hu-IL15) Xenografts: Transduced human cord blood (CB) CD34+ cells were washed and incubated with 1ug/100uL of OKT3 (Tonbo Biosciences; San Diego, CA) for 30 minutes at 4◦C to prevent contaminating T cell–derived graft-versus-host disease. Immediately before transplant, 1- to 3-day-old neonatal NSG-Tg(hu-IL-15) mice (NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl} Tg(IL15)1Sz/SzJ, Strain #030890, The Jackson Laboratory; Bar Harbor, ME) were irradiated at a dose of 150 rads with a Cesium-137 source. Each mouse was injected intra-hepatically with 1×10^5 to 5×10^5 cells. 16 weeks post-transplant, mouse bone marrow, spleen, thymus, and peripheral blood were harvested into single cell suspensions for downstream flow cytometry analysis. The mice were maintained at UCLA under an approved protocol by the UCLA Animal Research Committee under the Division of Laboratory Medicine.

Plasmid Generation: All LVs were cloned into an empty pCCL backbone.[39] Fragments of transcriptional regulatory elements were synthesized as gBlocks (Integrated DNA Technologies; Coralville, IA) or amplified from genomic DNA by polymerase chain reaction with compatible ends to be cloned using NEBuilder HiFi DNA Assembly Kit (New England Biolabs; Ipswich, MA). Enhancer elements were inserted upstream of a 600bp *SH2D1A* promoter to drive expression of the SLAM-associated protein or an mCitrine (mCit) reporter cassette. Furthermore, each vector contains the Woodchuck Hepatitis Virus post-transcriptional regulatory element (WPRE) in replacement of the endogenous *SH2D1A* 3' UTR.

Cell Culture: CD34+ cells were cultured in XVIVO15 medium (Lonza Biosciences; Basel, Switzerland) supplemented with 1X Penicillin-Streptomycin-Glutamine (P/S/G), with 50 ng/mL each of human stem cell factor (hSCF), human thrombopoietin (hTPO), and human FMS-like tyrosine kinase 3 ligand (hFlt3-L) for 24 hours before LV transduction. CD3+ T cells were isolated from peripheral blood mononuclear cells (PBMCs) through CD3+ positive magnetic selection (Miltenyi Biotec; Bergisch Gladbach, Germany). Isolated cells were then activated using anti-CD3/CD28 Immunocult (StemCell Technologies; Vancouver, Canada) and cultured in X-VIVO15 medium (Lonza Biosciences; Basel, Switzerland), 5% human serum, and 100 U/mL of human recombinant IL-2. NKT cells were isolated from PBMCs through iNKT Vα24+ positive magnetic selection (Miltenyi Biotec; Bergisch Gladbach, Germany). NKT cells were cultured in RPMI 1640, 1% Pen/Strep (P/S), 10% fetal bovine serum(FBS), 1% MEM non-essential amino acids, 10mM HEPES, 1mM Sodium Pyruvate, and 50μM 2-mercaptoethanol. NKT cells were stimulated at a ratio of 1:1 with autologous PBMCs

loaded with 5ng/mL of α-GalCer (Cayman Chemical, Ann Arbor, Michigan; Catalog 158021-47-7) and irradiated at 6000rpm. The cocultured cells were subsequently cultured with 10ng/mL of IL-15 and IL-7. NK cells were enriched from PBMCs using the CD56+ positive magnetic selection (Miltenyi Biotec; Bergisch Gladbach, Germany). Isolated NK cells were cultured in NK MACS medium (Miltenyi Biotech; Bergisch Gladbach, Germany) supplemented with 5% human AB serum, 1% P/S, 10ng/mL hIL-15 and 500U/mL of hIL-2 at 5% CO2 and 37°C humidified atmosphere.

Artificial Thymic Organoid (ATO) Generation: ATOs were generated from human mobilized peripheral blood stem cells as previously described.[29,30] Previously frozen MS5-hDLL4 cells were thawed and resuspended in serum free ATO culture medium ("RB27") composed of RPMI 1640 (Corning, Manassas, VA), 4% B27 supplement (ThermoFisher Scientific; Grand Island, NY), 30 μM L-ascorbic acid 2-phosphate sesquimagnesium salt hydrate (Sigma-Aldrich; St. Louis, MO) reconstituted in PBS, 1% penicillin/streptomycin (Gemini Bio-Products, West Sacramento, CA), and 2% Glutamax (ThermoFisher Scientific; Grand Island, NY). RB27 lasts 3 weeks at 4C. MS5-hDLL4 cells and CD34+ mPB cells edited with XLP1-SMART LVs and mock control were combined in an Eppendorf tube at a concentration of 150K MS5-hDLL4 and 5000 mPB per ATO to make 28 ATOs per group. Cells were centrifuged at 300 g for 5 min at 4˚C in a swinging bucket centrifuge. Supernatants were carefully removed, and the cell pellet was resuspended in 5 μl RB27 per ATO and mixed by brief vortexing. ATOs were plated on a 0.4 μm Millicell transwell insert (EMD Millipore; Billerica, MA; Cat. PICM0RG50) placed in a 6-well plate containing 1 ml complete RB27 (RB27 with the addition of 5 ng/ml rhFLT3L, 2.5 ng/ml rhIL-7, and 5ng/ml hSCF (Peprotech; Rocky Hill, NJ) per well.

Medium was changed completely every 3–4 days by aspiration from around the cell insert followed by replacement with 1 ml with fresh RB27/cytokines. SCF is only added for the first week of culture. ATOs are kept in an incubator at 37C with 5% CO2. At weeks 3, 7, and 12, ATO cells were harvested by adding FACS buffer (PBS/0.5% bovine serum albumin/2mM EDTA) to each well and briefly disaggregating the ATO by pipetting, followed by passage through a 50 μm nylon cell strainer. Cells were stained in 96 well plates. 200K cells per ATO were stained without fixation to determine mCitrine levels and T cell differentiation kinetics, and the remaining cells were fixed and permeabilized to assess SAP expression and T cell differentiation.

Colony Forming Unit (CFU) Assay: 100, 300, and 900 BM CD34+ HSPCs per replicate were plated in MethoCult (Cat. # 04445, StemCell Technologies; Vancouver, Canada) 24 hours after LV transduction. After 14 days of culture at 5% CO2, 37°C and humidified atmosphere, the number of mature colonies were counted and scored under the microscope based on their specific morphology after 14 days in culture.

Generation of SH2D1A knockout cell lines: To generate an XLP1 model cell line, Jurkat T cells were modified to knockout *SH2D1A* by electroporation of SpCas9 recombinant protein (QB3 Macrolab, UC Berkeley; Berkeley, CA) complexed to sgRNA (5'AACAGGTTCTTGGAGTGCTG3') (Synthego; Redwood City, CA) and FACS single-cell sorted and cultured in R20 (RPMI 1640 [Gibco; Grand Island, NY]/20% FBS [Gibco; Grand Island, NY]/1x Penicillin/Streptomycin/Glutamine [PSG, Gemini Bio Products]). Primers for amplification of the *SH2D1A* locus were used to confirm knock-out (oPAF605 [TCCTATGAATGCAATGACACCA] and oPAR340

[TGTGGCAATTTTCAGGAGTTCAC]) by Synthego ICE. Absence of SAP expression was confirmed by Western blot analysis. Cells were cultured in R10 at 37ºC with 5% CO_2.

Western Blot: For immunoblots, cells were lysed in RIPA Lysis and Extraction Buffer (Cat: 89901; ThermoFisher Scientific; Grand Island, NY) with added HALT protease inhibitor (Cat: 87786; ThermoFisher Scientific; Grand Island, NY) at a 1 × concentration following the manufacturer's protocols. Lysate concentrations were determined using the Pierce BCA protein assay (Cat: 23227; ThermoFisher Scientific, Grand Island, NY) following the manufacturer's protocol. Samples were treated for sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) with NuPAGE LDS Sample Buffer (Cat: NP0007; ThermoFisher Scientific; Grand Island, NY) and NuPAGE Sample Reducing Agent (Cat: NP0009; ThermoFisher Scientific; Grand Island, NY), each to a 1 × concentration. Lysates were diluted to contain equivalent amounts of total protein for immunoblot gel loading, using lysate from the SAP-deficient Jurkat cells to keep the total amount of protein loaded per lane constant to allow for valid loading controls. Wild-type Jurkat lysate was used as a control to indicate the relative expression levels of the *SH2D1A* codon optimized XLP1 lentiviral vectors. SAP levels were detected using Abnova monoclonal antibody, Clone 1C9 (Catalog H00004068-M01).

Flow Cytometry: Intracellular staining of SAP was performed using the eBioscience™ Foxp3 / Transcription Factor Staining Buffer Set (Invitrogen; Waltham, MA) using the manufacturers protocol. The primary antibody was rat anti-human SAP antibody, PE (Catalog# 12-9787-42, Invitrogen; Waltham, MA). To discern the various HSPC

lineages, cell populations were gated for the following; NK Cell: CD56+,CD16-; NKT: CD3+, CD56+; iNKT: CD3+, TCR Va24+, CD56+; B Cell: CD19+, CD3-; T Cell: CD3+, CD19- (CD4+/CD8+); Monocytes: CD33-, CD16-, CD14+; T cells: To discern the various T cell subpopulations and thymocyte populations in ATOs, we used the following monoclonal antibodies for staining: hCD45, hCD14, hCD56, hCD19, hCD34, hCD5, hCD7, hTCRab, hCD4, hCD8, hCD3, hCD45RA. To discern the various NK cell subpopulations, we used the following monoclonal antibodies for staining: hCD122, hCD16, hCD117, hCD45, hNKP80, hCD56, hCD34, hCD94, hCD57. See Supplemental Table for more information (sheet label: Human PBMC NK Dev. Flow).

Next Generation Sequencing Library Preparation: 14 days after transduction, genomic DNA and mRNA were harvested from the primary T cells, primary NK cells, primary NKT cells, and the B-LCLs. For library preparation, an initial PCR was completed to amplify the barcodes using primers oPAF255 - XLP Barcode Amp for NGS (See Table 1) and oPAR119 - XLP Barcode Amp for NGS (See Table 1). A second PCR was completed to add Illumina adapters and indexes. Following Illumina barcoding, PCR products were pooled at equal concentrations, purified twice using AMPure XP beads (Beckman Coulter; Brea, CA), and then quantified by ddPCR (QX 200; Bio-Rad Laboratories Inc.; Hercules, CA). The high-throughput sequencing was performed at the UCLA Technology Center for Genomics & Bioinformatics (TCGB) using an Illumina MiSeq instrument 2 × 150 paired-end reads (Illumina; San Diego, CA).

Restimulation Induced Cell Death (*RICD)*: RICD was assessed using an established protocol.[33,40] LV modified or unmodified T cells were cultured for 10 days, before plating

at 5×10^4 cells/well in a 96 well plate in 100 µl media. Dilutions of OKT3 antibody (Tonbo Biosciences; San Diego, CA) were prepared at 2000 ng/ml, 200 ng/ml, and 20 ng/ml, and 100 µl added to the cells to make final concentrations of 1000 ng/ml, 100 ng/ml and 10 ng/ml OKT3 in the wells. After 24 h, a final concentration of 1ug/mL of Propodium Iodide (PI) was added before running a fixed volume of cell suspension from each well for flow cytometry. The number of live cells (PI-) in stimulated controls were compared to unstimulated controls to measure the % cell loss = [1-(# PI- restimulated cells/# PI- untreated cells)]x100.

NK Differentiation: Untransduced or transduced healthy donor and XLP1 patient BM CD34+ cells were differentiated into CD56+ NK cells over 28 days of culture using the StemSpan™ NK Cell Generation Kit (Catalog # 09960, StemCell Technologies; Vancouver, Canada) following manufacturers protocol. Briefly, $1x10^4$ BM CD34+ cells were plated on non-tissue culture treated plates coated with StemSpan™ Lymphoid Differentiation Coating Material. Half-volume media changes were performed using StemSpan™ Lymphoid Progenitor Expansion Medium until day 14. The cells were then replated at $1x10^5$ cells/mL in StemSpan™ NK Cell Differentiation Medium and cultured for another 14 days. On day 28, the cells were harvested and enriched using CD56+ positive magnetic selection (Miltenyi Biotec; Bergisch Gladbach, Germany). The differentiated NK cells were then assessed for cytotoxic capabilities in downstream analysis.

NK Killing Assay: BM CD34+ cells differentiated into CD56+ NK cells were counted and resuspended in NK medium at $5x10^4$ cells per 200uL and added to row B on a 96 well

plate. 100uL of NK medium was added to row C to G. Six serial dilutions were performed of the resuspended NK cells using 100uL multichannel pipette from rows B to G (from $2.5x10^4$ cells per well to 800 cells per well). Target cells were either K562 cells expressing one copy of a GFP reporter cassette or Raji cells labeled with CellTrace™ CFSE (C34554; Invitrogen, Waltham, MA) resuspended in R10 (RPMI 1640, 10% FBS, 1% P/S/G) at $1x10^4$ cells per 100uL. 100uL of the target cells were added to rows A through G to generate the following effector to target ratios – 2.5:1, 1.25:1, 0.625:1, 1:3, 1:6, 1:12.5. The plate was incubated for 18 hours at 5% CO2, 37°C. After incubation, GFP+ tumor cells were counted via FACS using a BD FACSCelesta™ Cell Analyzer (BD Biosciences; Franklin Lakes, NJ) and normalized to target only control wells (rows A and G) to normalize the killing percentage.[34]

Monocyte Differentiation: CB CD34+ HSCs were cultured in X-VIVO15 medium (Lonza Biosciences; Basel, Switzerland), 4% FBS (Gibco; Grand Island, NY), 1X Pen/Strep/Glut (PSG, Gemini Bio Products), 50ng hSCF, 15ng/mL hTPO, 30ng/mL hIL3, and 30ng/mL hFlt3-L (all cytokines: Peprotech; Rocky Hill, NJ) for 9 days. After 9 days of culture, the CB CD34+ cells were differentiated in the following medium for 7 days: STEMSpan II (StemCell Technologies; Vancouver, Canada), 20% FBS (Gibco; Grand Island, NY), 1X Pen/Strep/Glut (PSG, Gemini Bio Products), 25ng hSCF, 30ng/mL M-CSF, 30ng/mL hIL3, and 30ng/mL hFlt3-L (all cytokines: Peprotech; Rocky Hill, NJ). On day 16 of culture, cells were confirmed for NK differentiation and purity by flow cytometry, using anti-CD14 and anti-CD16 antibodies.

T Cell Electroporation to Knock-Out SH2D1A: After isolation from PBMCs, CD3+ primary T cells were counted by hemacytometer via trypan blue exclusion prior to electroporation. 1 x 10^6 cells per condition were centrifuged at 300xg for 10 min at RT, resuspended in 20uL of P3 electroporation buffer (Lonza Biosciences; Basel, Switzerland). 100pmol of SpCas9 recombinant protein (QB3 Macrolab, UC Berkeley; Berkeley, CA) was combined with 120pmol of each sgRNA to *SH2D1A* (sgRNA4: 5'-GACGCAGTGGCTGTGTATCA-3'; sgRNA7: 5'- AACAGGTTCTTGGAGTGCTG-3' both from Synthego; Redwood City CA) for 15 minutes at RT for RNP complex formation. The cell and RNP mixtures were combined and electroporated using the EH-100 setting on the Amaxa 4D Nucleofector X Unit (Lonza Biosciences; Basel, Switzerland). Cells were rested in 16-well electroporation strips (Lonza Biosciences; Basel, Switzerland) for 10 min at RT and then recovered with 480uL of T cell medium. 24 hours after electroporation, *SH2D1A* knockout T cells were transduced with XLP1-SMART-LVs and used for an RICD assay 10 days post electroporation.

Statistical Analysis: All data are reported as mean ± SD unless otherwise stated. All statistical analyses were carried out using GraphPad Prism version 10.0.0 (GraphPad Software; San Diego, CA). The statistical significance between two averages was established using unpaired t tests. When the statistical significance between three or more averages was evaluated, a one-way ANOVA was applied, followed by multiple paired comparisons for normally distributed data (Tukey's test). When the statistical significance between two or more categorical variables was evaluated, a two-way ANOVA was applied, followed by multiple paired comparisons for normally distributed data (Tukey's test). Linear regression analyses were used to determine the correlation

between Titer and proviral size. All statistical tests were two-tailed and a p value of < 0.05 was deemed significant (ns non-significant, *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.) Details of statistical tests used, including all p values, are indicated in the relevant figure legend.

RESULTS

To elucidate the putative elements responsible for the lineage- and temporal-specific expression of the endogenous *SH2D1A* locus, we employed a bioinformatics-guided approach, utilizing GeneHancer, a bioinformatic tool that links over 285,000 candidate enhancer elements across the human genome to their respective target genes.[24] GeneHancer integrates four genome-wide enhancer databases (ENCODE, Ensembl, FANTOM and VISTA), to generate a comprehensive list of putative regulatory elements for each gene. By integrating data including histone modification, chromatin accessibility, bound transcription factors and Hi-C interactions, we identified 34 potential regulatory elements of *SH2D1A*, located within a 200 kb window of the *SH2D1A* transcription start site. We entered each genomic coordinate into the UCSC Genome Browser to examine the corresponding regulatory elements (**Figure 2.1A**).

We designed a series of lentiviral vectors (XLP1-SMART LVs) to test each element's enhancer capability independently (**Figure 2.1A**). The elements were inserted upstream of a 600 bp *SH2D1A* promoter – identified via the The Eukaryotic Promoter Database (EPD) – to drive expression of an mCitrine (mCit) reporter cassette. Furthermore, each vector was designed with the Woodchuck Hepatitis Virus post-transcriptional regulatory element (WPRE) replacing the endogenous 3' UTR of *SH2D1A* to enhance expression of the transgene cassette. To multiplex the series of vectors, unique 20 bp barcodes were cloned directly upstream of the WPRE (**Figure 2.1A**). At this location, the barcode remains within the transcript and vector provirus genome for identification and normalization by RNA or DNA, respectively, but it will not be translated to affect any

protein function. The barcodes were designed, each with a Hamming Distance of 10, to tolerate accidental mutations. Each of the 34 vectors was cloned twice—each with its own unique barcode—to further assess *any* potential biases from transduction, recombination, or PCR amplification. A lentiviral vector used in pre-clinical XLP1 gene therapy studies[1], EFS-SAP, was cloned in duplicate as a control. The *SH2D1A* gene was replaced with an mCit reporter and termed EFS-mCit-WPRE.

To mimic the endogenous expression of SAP, we tested each lentiviral vector construct for its ability to drive high-level expression in primary T cells, NKT cells, and NK cells, each isolated from healthy donor peripheral blood mononuclear cells. B-lymphoblastoid cell lines (B-LCLs) were utilized to assess off-target B lymphoid expression of each regulatory element. We transduced each cell population with the 34 candidate vectors and the EFS control. 14 days post-transduction, cells were harvested to isolate their gDNA and mRNA fractions. Bulk VCN of the transduced populations was assessed via droplet-digital (dd)PCR. The barcode-containing DNA regions were PCR amplified from the gDNA fraction and assessed via next-generation sequencing. Furthermore, the barcode-containing regions were PCR amplified from the RNA fraction after cDNA conversion and assessed via next-generation sequencing. The RNA barcode counts are proportionate to the transcriptional activity of the enhancers within each lineage, whereas each genomic barcode count is used to normalize each barcode in the transcript to the genomic copy number to account for potential differences in transduction efficiency of the vectors. The number of RNA barcode reads, normalized to frequency of gDNA barcodes, within each cell type, determined the relative expression directed by each element (**Figure 2.1A**).

Of the 34 regulatory elements tested, elements 3 and 5R demonstrated T cell-specific enhancer expression (**Figure 2.1B; Figure 2.6C**). Elements 3, 4, 5, 7, 20, and 5R demonstrated NK and NKT cell-specific enhancer expression (**Figure 2.1B; Figure 2.6D-1E**). All 34 regulatory elements tested demonstrated minimal to no off-target expression in B-LCLs (**Figure 2.1B; Figure 2.6F**).

Enhancers 3, 4, 5, 7, 20, and 5R were refined to decrease their size (thereby increasing the vector titers and gene transfer) with the goal of retaining high enhancer-driven function (**Figure 2.1C-1E**). XLP1-SMART LV plasmids were similarly designed with the refined enhancer fragments and the 600 bp *SH2D1A* promoter driving expression of an mCit cassette. XLP1-SMART LV plasmids were packaged and titered head-to-head using methods previously described by our laboratory (**Figure 2.1C**).[23,25,26] Primary T and NK cells were transduced with the refined LVs to achieve equivalent vector copy numbers (VCNs) ranging from ~0.10 to 0.20 to increase the probability of each transduced cell containing a single integrant. The MFI of the mCit+ populations were assessed via flow cytometry to compare the ability of each refined enhancer element to drive expression in T cells (**Figure 2.1D**) and NK cells (**Figure 2.1E**). Enhancers 3M and 5RL retained 90% and 100% of the full-length enhancer activity in T cells, while reducing the enhancer sizes by 3.5 kb and 4 kb, respectively from the original elements (**Figure 2.1D**). Enhancers 3M, 3GP, 20R, and 5RL retained 120%, 110%, 130%, and 100% of the full-length enhancer activity in NK cells, while reducing the enhancer sizes by 3.5 kb. 4.2 kb, 1 kb, and 4 kb, respectively (**Figure 2.1E**). Enhancers 4, 5, and 7 and their shortened iterations did not significantly increase expression in T cells or NK cells when compared to the promoter only control (Pro) and were thus removed from

subsequent analyses (**Figures 2.1D and 2.1E**). After refining the enhancers, the elements were combined to assess additive and/or synergistic effects on expression.

Enhancer 3M was combined with either 20R, 5RL or both 20R and 5RL and a 600bp *SH2D1A* promoter to drive expression of an mCit cassette (**Figure 2.2A**). Similarly, XLP1-SMART LV plasmids were packaged and titered head-to-head using methods previously described by our lab (**Figure 2.2B**).[23,25,26] We transduced primary T and NK cells for evaluating on-target expression (**Figures 2.2C and 2.2D**). Transduced CB CD34+ cells were differentiated into monocytes and along with B-LCLs, were used to determine off-target expression levels (**Figures 2.2E and 2.2F**).

Enhancers E3M/GP, E20R, and 5RL demonstrated additive expression levels when combined and evaluated in primary T and NK cells (**Figures 2.2C and 2.2D**). The combination vectors (E3M-20R-5RL and E3GP-20R-5RL) increased expression in T cells up to 4-fold compared to the promoter only control and up to 1.8-fold greater than the control EFS vector (**Figure 2.2C**). Further, the combination vectors increased expression in NK cells 2-fold greater than the promoter only and 1.5-fold greater than the EFS preclinical vector (**Figure 2D**). XLP1-SMART LVs that included E3M showed minor levels of off-target expression, with the combination XLP1-SMART vector (E3M-20R-5RL) harboring a 1.25 or 1.67-fold increase in expression levels monocytes and B cells over the promoter only control; however, the E3GP combination vector (E3GP-20R-5RL) showed no significant levels of off-target expression (**Figures 2.2E and 2.2F**).

To determine the lineage and temporal-specific expression of the XLP1-SMART LVs, we tested the ability of each LV to drive mCit reporter expression *in vivo* in NSG-Tg(Hu-

IL15) immunodeficient mice. Healthy human donor human CB CD34+ HSCs were pre-stimulated for 24 hours in 50 ng/mL of human hSCF, hTPO, and hFT3L before transduction with 2e7 TU/mL of our XLP1-SMART-LVs. The E3M-20R-5RL and E3GP-20R-5RL LVs were compared to both mock transduced cells and to CB CD34+ HSCs transduced with 2e7 TU/mL of the control EFS-mCit vector. We transplanted the transduced cells via intra-hepatic injection into sub-lethally irradiated NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl} Tg(IL15)1Sz/SzJ (NSG-Tg[HuIL15]) neonatal pups receiving 150 rads of irradiation. These NSG-Tg(HuIL15) mice express human IL-15 to help enhance the development of human NK cells within the immunodeficient mice engrafted with human CD34+ cells.[27] Furthermore, the transplantation of NSG pups supports more efficient human T cell development after HSC injection than adult NSG mice.[28]

Mice were harvested 16 weeks post-transplantation to allow for the assessment of engraftment and vector-mediated gene expression in multiple human hematopoietic cell lineages. The bone marrow, thymus, and spleen were processed into single cell suspensions. The vector copy number (VCN) was determined for the transplanted cells within the bone marrow of each transplanted mouse (**Figure 2.3A**) and the bone marrow engraftment (**Figure 2.3B**) of each mouse was quantified via the percent hCD45+ human cells via flow cytometry. Average VCNs for Promoter Only (Pro), EFS-mCit (EFS), E3M-20R-5RL-mCit (3M), and E3GP-20R-5RL-mCit (3GP) in the bone marrow compartment of NSG-Tg(Hu-IL15) mice were 2.17, 1.72, 0.81, and 1.57, respectively. With exception of one mouse that did not engraft in the E3GP-20R-5RL-mCit group, average human cell engraftment was 75% for all groups.

We evaluated **lineage-specific** expression of the LVs by analyzing the level of expression measured as the mean fluorescence intensity (MFI) of mCit expression in the hCD45+ cells across the different cell lineages in on-target and off-target populations. Mice transplanted with CB CD34+ HSCs transduced with the E3M-20R-5RL-mCit LV contained 7.1-fold, 2.8-fold, and 4.3-fold brighter mCit+ expression than the EFS-mCit vector in T cells, NK cells, and NKT cells, respectively (**Figure 2.3C**). Furthermore, mice transplanted with CB CD34+ HSCs transduced with E3M-20R-5RL-mCit or E3GP-20R-5RL-mCit LVs contained no off-target expression in myeloid cells or B cells, unlike the EFS-mCit counterpart.

To evaluate the **stage-specific** expression of our rationally designed XLP1-SMART LVs, we stained the bone marrow, thymus, and spleen cell suspensions from the mice transplanted and harvested with a NK cell or T cell differentiation antibody panel and evaluated their mCitrine+ expression via flow cytometry (**Figure 3.3E and 3.3F; Figure 2.9B-2.9D**). Mice transplanted with CB CD34+ HSCs transduced with the E3M-20R-5RL-mCit LV contained 1.7-fold to 3.7-fold brighter mCit+ expression than the EFS-mCit vector in mature NK lineages (Stage 4a to Stage 6; **Figure 2.3E**). The mCit+ expression measured across mature NK lineages in peripheral blood of mice transplanted with CB CD34+ HSCs – transduced with the E3M-20R-5RL-mCit LV (in contrast to EFS-mCit) – demonstrated similar patterns of endogenous SAP expression across NK development in the peripheral blood of a healthy donor (**Figure 2.10B and 2.10D**). Mice transplanted with CB CD34+ HSCs transduced with the E3M-20R-5RL-mCit LV contained 4.7-fold to 5.0-fold brighter mCit+ expression than cells in mice receiving the EFS-mCit vector in mature T lineages. (**Figure 2.3F).** The mCit+ expression measured across mature T

lineages in the thymus of NSG-Tg(Hu-IL15) mice transplanted with CB CD34+ HSCs – transduced with the E3M-20R-5RL-mCit LV (in contrast to EFS-mCit) – demonstrated similar patterns of expression as SAP expression across T development from a heathy donor thymus (**Figure 2.10A and 2.10C**).

To further test the temporal specificity of the XLP1 SMART-LVs, we utilized the Artificial Thymic Organoid (ATO) system, which fully recapitulates thymopoiesis from multiple stem cell sources.[29,30] Human CD34+ mobilized peripheral blood (mPB) cells were transduced with E3M-20R-5RL-mCit LV, E3GP-20R-5RL-mCit, and EFS-mCit at equivalent vector copy number. Edited mPB cells and a mock control were combined with the MS5-hDLL4 stromal cell line which constitutively expresses human Notch delta-like ligand 4 (DLL4). Cell count and T cell differentiation kinetics were measured at week 3, 7, and 10 using flow cytometry, and the percentage of mCitrine was compared to endogenous SAP percentage to compare the expressio across multiple stages of T cell differentiation. The mCitrine+ percentage of the thymocytes from ATOs with XLP1 SMART-LV mPB cells is expressed across all stages of T-cell development, consistent with SAP expression throughout T-cell differentiation from uncommitted to mature thymocytes. (**Figure 2.7A-2.7B**).

To assess both the clonogenic potential of human hematopoietic progenitor cells after transduction with the LV and the functional restoration of the XLP1 phenotype, the mCitrine open reading frame of the XLP1 SMART-LVs was replaced with the coding region of human *SH2D1A*, codon optimized using the JCAT codon optimization algorithm (**Figure 2.8A-2.8B**).[31] Codon optimized XLP1-SMART LV plasmids were

packaged and titered head-to-head using methods previously described[23,25,26] BM CD34+ were obtained from XLP1 patients after their informed consent (UCLA IRB# 10-001399). XLP1 patient CD34+ cells were pre-stimulated for 24 hours in 50 ng/mL of human hSCF, hTPO, and hFT3L before transduction with the XLP1 SMART-LVs. Cells were transduced to achieve equivalent vector copy numbers of ~1.1 and ~2.4 (EFS = 1.1 and 3.0; E3M = 0.9 and 1.7; E3GP = 1.2 and 2.52) (based on prior dose response testing of VCN produced by each LV across a range of concentrations). Transduced CD34+ cells were used to conduct a colony-forming unit (CFU) assay for progenitor cell clonogenic potential and were also differentiated to NK cells to assess XLP1 functional restoration via an NK cell cytotoxicity assay.

Transduced cells were seeded in semi-solid methylcellulose medium and cultured for 14 days before quantifying vector effect on clonogenicity and generation of hematopoietic progenitor colonies (**Figure 2.4A-2.4C**). XLP1-SMART-LVs demonstrated no changes to clonogenicity or hematopoietic lineage skewing in comparison to a healthy donor control (**Figure 2.4A-2.4CC**) at a VCN of 1. Conversely, patient cells transduced with the EFS-SAP vector demonstrated significant skewing ($p<0.05$) into the myeloid lineage when compared to a healthy donor control at a vector copy number of 1.1 (**Figure 2.4B-2.4C**). E3M-20R-5RL-SAP (3M-SAP) XLP1-SMART-LVs demonstrated no changes to clonogenicity or hematopoietic lineage skewing in comparison to a healthy donor control (**Figure 2.4D-2.4E**) at a VCN of 1.7. Conversely, patient cells transduced with the EFS-SAP vector (VCN=3) demonstrated significant skewing ($p<0.05$) into the myeloid lineage, with a 1.3-fold increase in GM (granulocyte–macrophage) colonies compared to a healthy donor control (**Figure 2.4F-2.4G**). At a VCN of 2.52, E3GP-20R-5RL-SAP

(3GP-SAP) also demonstrated significant skewing into the myeloid lineage with a 1.15-fold increase in GM colonies compared to a healthy donor control (**Figure 2.4F-2.4G**).

XLP1 patient peripheral blood CD8+ T cells were transduced at equivalent vector copy numbers to assess functional restoration by means of a restimulation induced cell death (RICD) assay.[32,33] 14 days post transduction, transduced patient CD8+ T cells were quantified for recovery of RICD by flow cytometry. Transduced XLP1 patient T cells with the EFS, E3M, and E3GP vectors restored T cell RICD activity to healthy donor levels at an equal vector copy number of 1 (**Figure 2.5B)**.

Healthy donor human CD8+ cells were also transfected using the CRISPR/Cas9 system to knockout the *SH2D1A* gene. Knockout T cells were transduced with XLP1-SMART LVs to assess their RICD activity. The E3M-20R-5RL-SAP LV at a VCN of 2 was capable of restoring RICD activity to healthy donor levels. Conversely, EFS-SAP transduced *SH2D1A* knockout T cells did not restore T cell RICD activity to healthy donor levels at an equivalent vector copy number of 2 (**Figure 2.11**).

Transduced XLP1 patient BM CD34+ were also differentiated into NK cells and tested for functional recovery of NK cell cytotoxicity in a tumor killing assay with NK sensitive target cells that express SLAM family receptors (Raji) and those that do not (K562).[34,35] As expected, MHC-independent killing was observed in K562 cells, illustrating equal levels of killing of K562 for healthy donor and non-transduced/transduced XLP patient cells (VCNs of EFS = 6, E3M = 0.9, and E3GP = 1.2 [**Figure 2.5A**]). In a parallel assay with SLAM expressing Raji cells, E3M-20R-5RL-SAP transduced XLP1 patient BM CD34+ cells restored NK cell cytotoxicity to near healthy donor levels (VCNs of EFS =

1, E3M = 1.7, E3GP = 2.52 [**Figure 2.5B**]). On the other hand, E3GP-20R-5RL-SAP and EFS-SAP transduced XLP1 patient BM CD34+ cells did not fully restore NK cell cytotoxicity to healthy donor levels (**Figure 5B**) at an average vector copy number of ~1.7.

DISCUSSION

Autologous Hematopoietic Stem Cell Therapy, or Gene Therapy (GT), for XLP1 could provide a safe and effective curative treatment if gene therapy methods can achieve precise, regulated SAP expression in the correct lineages, at appropriate levels, at the right developmental stage. There are far fewer risks for unwanted immune responses with autologous GT that hamper allogeneic HSCT, with complete absence of GVHD and unlikely graft rejection. However, since SAP expression is tightly regulated in T, NK, and NKT cells, a gene therapy candidate is required that can restore appropriate *SH2D1A* gene expression after an autologous HSCT in the appropriate hematopoietic cell lineages.

A previous lentiviral vector gene therapy was produced to combat XLP1.[11] However, the proposed method utilized a ubiquitously active elongation factor 1 α short promoter to drive expression of the SAP protein (EFS-SAP). One concern with this approach lies in the non-physiological expression of SAP in hematopoietic cell populations. Since SAP expression is tightly regulated within T, NK, and NKT cells, gene expression in off-target cell populations may pose safety concerns, including improper signaling of factors associated with autoimmune diseases and an elevated apoptotic response to DNA damage.[20,36] Previous researchers have demonstrated significant skewing induced by the EFS-SAP vector of hematopoietic potential in the bone marrow compartment via a colony forming unit (CFU) assay.[8,11] As a result, we rationalized that an XLP1 LV therapy requires the incorporation of endogenous enhancer elements to regulate the expression of the *SH2D1A* transgene.

We implemented a bioinformatics-guided approach to develop a highly regulated LV driven by endogenous regulatory elements of the *SH2D1A* gene. Analysis of the topologically associated domain of *SH2D1A* revealed 3 genomic elements responsible for the physiological expression pattern of the *SH2D1A* gene. Element 5R, located directly 3kb downstream of the *SH2D1A* promoter, was shown to be a T, NK, and NKT-specific enhancer, lacking activity in B cell and myeloid cell lineages. Chromatin immunoprecipitation sequencing (ChIP-seq) data from ENCODE revealed various lymphoid-associated transcription factor binding sites such as MEIS2, TAL1, and SPI1 within element 5R. Element 3, located 125 kb upstream of the *SH2D1A* promoter, presented as a T, NK, and NKT-specific enhancer, lacking activity in B cell and myeloid cell lineages. Chromatin immunoprecipitation sequencing transcription factor analysis revealed various lymphoid-associated transcription factor binding sites within element 3, such as FOXA1, RUNX3, GATA2, and JUND. Finally, element 20, located within intron 1 of the *SH2D1A* gene, presented as an NK and NKT-specific enhancer lacking activity in T cell, B cell, and myeloid cell lineages. Taken together, these three enhancers can recapitulate the expression of SAP in all physiologically expressed lineages.

The therapeutic potential of an LV is greatly influenced by both gene transfer and expression. Enhancing gene transfer not only increases the percentage of transduced cells, but also results in more integrated LV copies within each cell, amplifying the overall cellular expression. Conversely, boosting expression directly results in a higher production of the therapeutic protein for every integrated copy. Previous studies have also demonstrated a negative correlation between titer, gene transfer, and proviral length.[26] To address this, we made systematic deletions in each of the genomic

elements to reduce the proviral length of our vector. As a result, we observed a 5- to 10-fold enhancement in titer and gene transfer.[23,26] Enhancer 3M was further reduced in size from 1.9 kb to 1 kb to generate the 3GP enhancer. These reductions in enhancer element lengths increased titer by 2-fold and gene transfer by 1.5-fold. Such improvements in titer and gene transfer are of significant value for the clinical application for clinical gene therapy because they substantially decrease the cost of vector production and reduce the volume of LV production lots needed for each patient's treatment. With these modifications, there is a potential concern about the regulation and expression becoming compromised. However, even after these adjustments, over 90% of the intended expression was maintained, and any unintended activity remained minimal. Further studies to refine these enhancer elements and retain vector expression can be made. While enhancer 3 demonstrated minimal off-target expression in B-LCLs and monocytes, when combined with enhancer 5RL or 5RL and 20R, off-target expression was abrogated, most likely a result of repressor elements within the 5R enhancer region.

Using reporter gene experiments in transduced human CB CD34+ cells that were then transplanted into NSG-Tg(Hu-IL15) neonates, we observed that the XLP1-SMART-LVs were predominantly expressed in the human T, NK, and NKT lineages. Importantly, there was no detectable expression in the B- and myeloid-cell lineages. Furthermore, XLP1-SMART-LVs exhibited stage-specific expression that increased through T cell and NK cell differentiation.

The reduction in expression of the smaller 3GP enhancer compared to 3M is more drastically observed in the functional restoration of the XLP1 phenotype. Unlike XLP1-SMART-LVs that harbored the E3M enhancer, the vector harboring the 3GP enhancer was not capable of fully restoring NK-cytotoxicity to wildtype levels at equivalent vector copy numbers. However, further studies in SH2D1A-/- knockout mice are necessary to fully elucidate the effects of this decreased expression. Ultimately, XLP1-SMART-LV expression recapitulated the expression patterns seen with the endogenous *SH2D1A* gene.

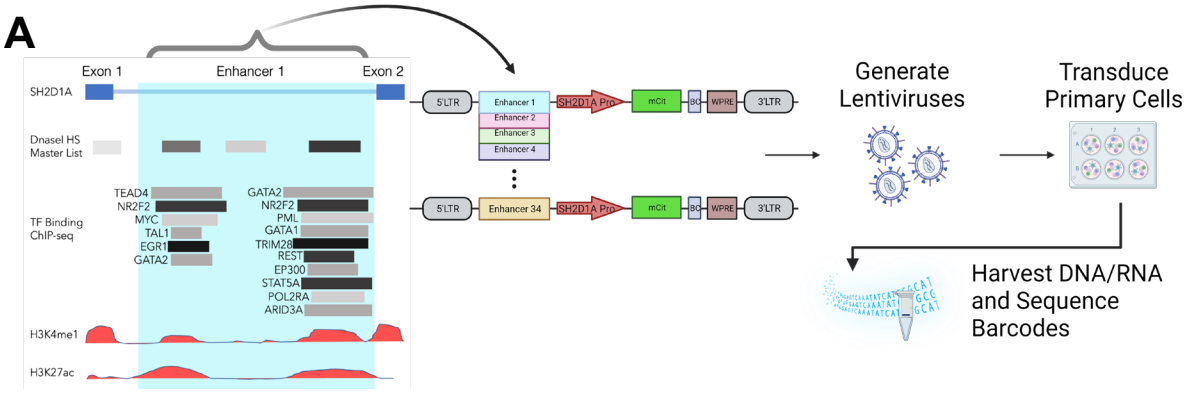
To further assess the effects of ectopic expression demonstrated by the EFS promoter, XLP1 patient BM CD34+ HSPCs were transduced and assessed for hematopoietic skewing via a CFU assay. While the E3M-20R-5RL XLP1-SMART-LV demonstrated no lineage skewing at VCNs of 1 and 1.7, the EFS-SAP LVs illustrated significant HSC skewing into the myeloid lineage, further indicating the potential adverse effects of ectopic SAP expression within hematopoietic cells. The lineage skewing demonstrated in the CFU assay may be attributed to the increased SAP expression within off-target lineages. The skewing into the myeloid lineage of the E3GP-20R-5RL XLP1-SMART LV may be attributed to the differences in enhancer sequences of enhancer 3. Consequently, the lack of skewing in the E3M-20R-5RL may be due to vector copy number differences between that sample and the EFS and E3GP counterparts (EFS = 3, E3M = 1.7, E3GP 2.52). However, further studies are necessary to determine the cause of this skewing and its effects on the XLP1 phenotype.

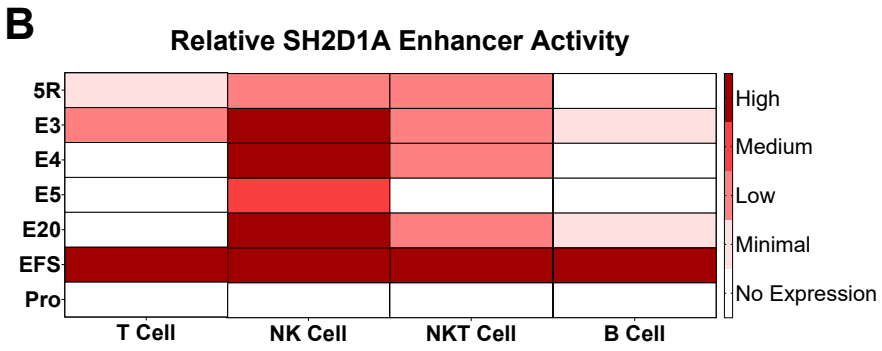
The dysregulated immune response seen in XLP1 patients can be attributed to their reductions in T cell function and NK cell cytotoxicity. As a result, we assessed the ability of XLP1 patient cells transduced with XLP1-SMART-LVs to have T cell and NK cell function restored to healthy donor (HD) levels. E3M-20R-5RL-SAP (E3M) and EFS-SAP transduced XLP1 patient CD8+ T cells restored RICD activity to HD levels at an average VCN of 1; in contrast, studies conducted in healthy donor samples – in which the *SH2D1A* was knocked out with CRISPR/Cas9 technology – demonstrated larger discrepancies of RICD restoration between the EFS, E3M, and E3GP vectors at equal vector copy number. While these variations may be due to donor-to-donor variability, further studies in XLP1 patient T cells or in a viable XLP1 *in vivo* model are necessary to confirm the capability of either the E3M or EFS vector's ability to restore T cell RICD function. In transduced XLP1 patient BM CD34+ cells differentiated into CD56+ NK cells, the E3M and E3GP vectors demonstrated greater restoration of cytolytic activity to SLAM expressing Raji cells than the EFS-SAP vector, a potential indicator of increased SAP expression in NK cells due to the E3M/GP, E5RL, and E20R enhancers. The reduction of NK expression in the E3GP sample compared to E3M may be due to core transcription factor binding sites and epigenetic modifications that were removed in the interest of decreasing vector size for an increased titer and gene transfer. The lack of EFS-SAP vector NK cytolytic activity is potentially a result of reduced SAP expression profile within NK cells or, more likely, the reduced vector copy number seen in the EFS transduced populations compared to the E3M and E3GP counterparts (VCNs were EFS=1.1, E3M =1.7, E3GP = 2.52). The limited access to more XLP1 patient CD34+ cells and PBMCs prevented repeated studies at equal vector copy numbers. Future *in*

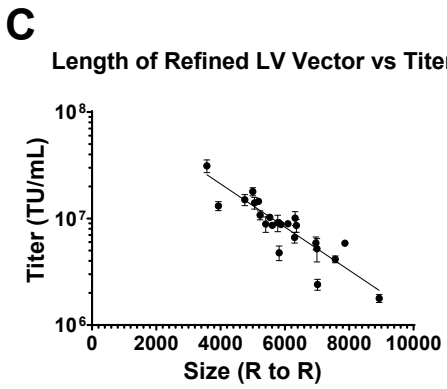
vivo studies utilizing the C57BL/6 *SH2D1A*-/- mouse model will help elucidate the ability for XLP1-SMART-LVs to restore the humoral defects and NKT developmental block seen in XLP1 patients.

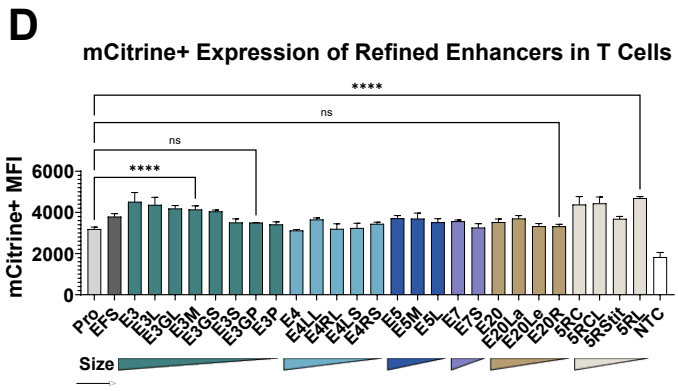
In conclusion, the outcomes of this study have elucidated a streamlined approach for the identification and incorporation of key elements into a vector cassette that can properly regulate a target gene (SMART-LVs). The rational generation of regulated LVs offers a novel, effective, and efficient approach for enhanced expression and regulation of transgenes required to correct various inborn errors of immunity, as evidenced by these highly specific XLP1-SMART-LVs for the treatment of XLP. Future studies will need to be conducted to assess the safety and efficacy of the XLP1-SMART-LVs (e.g., In Vitro Immortalization assay). In conjunction with these studies, the regulated expression by enhancers 3, 20, and 5R within the T, NK, and NKT lineages may provide a useful approach towards a lentiviral gene therapy for the treatment of Hereditary Lymphocytic Histiocytosis (HLH) disorders, such as Perforin deficiency, which share similar expression profiles and regulation to that of XLP1 and for cancer immunotherapy.

FIGURES


A
Exon 1
Enhancer 1
Exon 2
SH2D1A
Dnasel HS Master List
TF Binding ChIP-seq
TEAD4
NR2F2
MYC
TAL1
EGR1
GATA2
GATA2
NR2F2
PML
GATA1
TRIM28
REST
EP300
STAT5A
POL2RA
ARID3A
H3K4me1
H3K27ac
5'LTR
Enhancer 1
Enhancer 2
Enhancer 3
Enhancer 4
SH2D1A Pro
mCit
WPRE
3'LTR
Enhancer 34
Generate Lentiviruses
Transduce Primary Cells
Harvest DNA/RNA and Sequence Barcodes


B
Relative SH2D1A Enhancer Activity
5R
E3
E4
E5
E20
EFS
Pro
T Cell
NK Cell
NKT Cell
B Cell
High
Medium
Low
Minimal
No Expression


C
Length of Refined LV Vector vs Titer
Titer (TU/mL)
Size (R to R)


D
mCitrine+ Expression of Refined Enhancers in T Cells
mCitrine+ MFI
Size

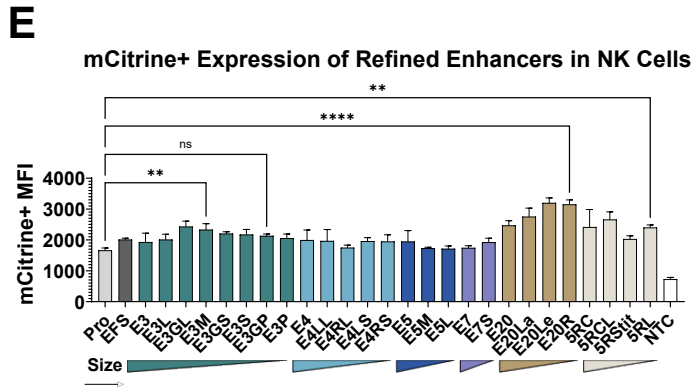

E
mCitrine+ Expression of Refined Enhancers in NK Cells
mCitrine+ MFI
Size

Figure 2. 1

A: UCSC Genome Browser interface of the *SH2D1A* locus encoding the SAP protein. Blue shaded columns indicate putative regulatory regions. Distinctive DNase I hypersensitive sites (DHSs) are shown across different cell lineages. TF binding, along with peaks of bound H3K4me3, H3K27ac, and DHSs can be used to define the presence and boundaries of putative enhancer elements. Proviral maps of a series of XLP1 “SMART” lentiviral vector constructs each containing a putative regulatory element upstream of the endogenous *SH2D1A* promoter (*SH2D1A* Pro) driving expression of an mCitrine (mCit) reporter cassette. A unique barcode (BC) is located upstream of the Woodchuck Hepatitis Virus post-transcriptional regulatory element (WPRE) to identify the ability for each element to drive lineage and temporal specific expression of the mCit reporter**.** We transduced primary T, NKT, and NK cells with a pool of raw viral supernatant containing each of the 34 candidate XLP1 SMART LVs, in duplicate with different BC, and the EFS-SAP vector as control. B-LCLs were transduced to measure off-target expression in B lymphocyte lineage. 14 days post-transduction, cells were harvested for their gDNA and RNA fractions to measure barcode expression and presence in gDNA via next-generation sequencing.

B: Relative *SH2D1A* Enhancer Activity of the Highest Expressing Elements. The relative enhancer expression measured by next generation sequencing of vector barcodes is shown. The highest expressing five enhancers are depicted (see supplemental for the remaining 29), with the lowest level of expression (white) denoted as that of the *SH2DA1* promoter only (Pro). Elements of interest contain increased

expression over the *SH2DA1* Pro (red). Minimal = 1-1.25-fold increase; Low = 1.25-1.5-fold increase; Medium = 1.5-2-fold increase; High = >2-fold increase.

C: Proviral size of Refined XLP1-SMART LV vs Titer. Putative enhancers were cloned into the plasmid backbone of a lentiviral vector (pCCL-c-MNDU3-X [Addgene Plasmid #81071]) with the MNDU3 promoter first removed, and the vectors were packaged and titered head-to-head. The quantities of infectious particles were plotted as a function of proviral length (bp). Each point in the plot represents an average of three individual 10-cm plate of virus titered on HT-29 cells. Proviral length is defined as sequence length from the beginning of the 5′ long terminal repeat (LTR) U3 through the end of the 3′ LTR U5. n = 3 per arm. Linear regression analyses were used to determine the correlation between titer and proviral size (R^2=0.78).

D-E: Expression from LV with refined enhancer elements (via GFP Mean Fluorescence Intensity (MFI)) in T and NK Cells. Healthy Donor CD3+ T Cells or CD56+ NK cells were isolated from peripheral blood mononuclear cells (PBMCs). CD3+ T Cells and CD56+ NK cells were transduced with each XLP1-SMART LV to achieve a VCN ranging from 0.1-0.2. 14 days post transduction (based on pre-determined titers); T cells were assessed for relative expression driven by each enhancer via mCitrine+ MFI using flow cytometry. Each enhancer was compared to basal *SH2D1A* promoter-driven expression (Pro) and the control LV (EFS). Data are represented as mean ± SD of biological triplicates from two experiments. Statistical significance was analyzed using a one-way ANOVA followed by multiple paired comparisons for normally distributed data (Tukey test). Statistical analysis was performed on all arms, but selected arms are shown.

All statistical tests were two-tailed and a p value of < 0.05 was deemed significant (ns = non-significant, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.).

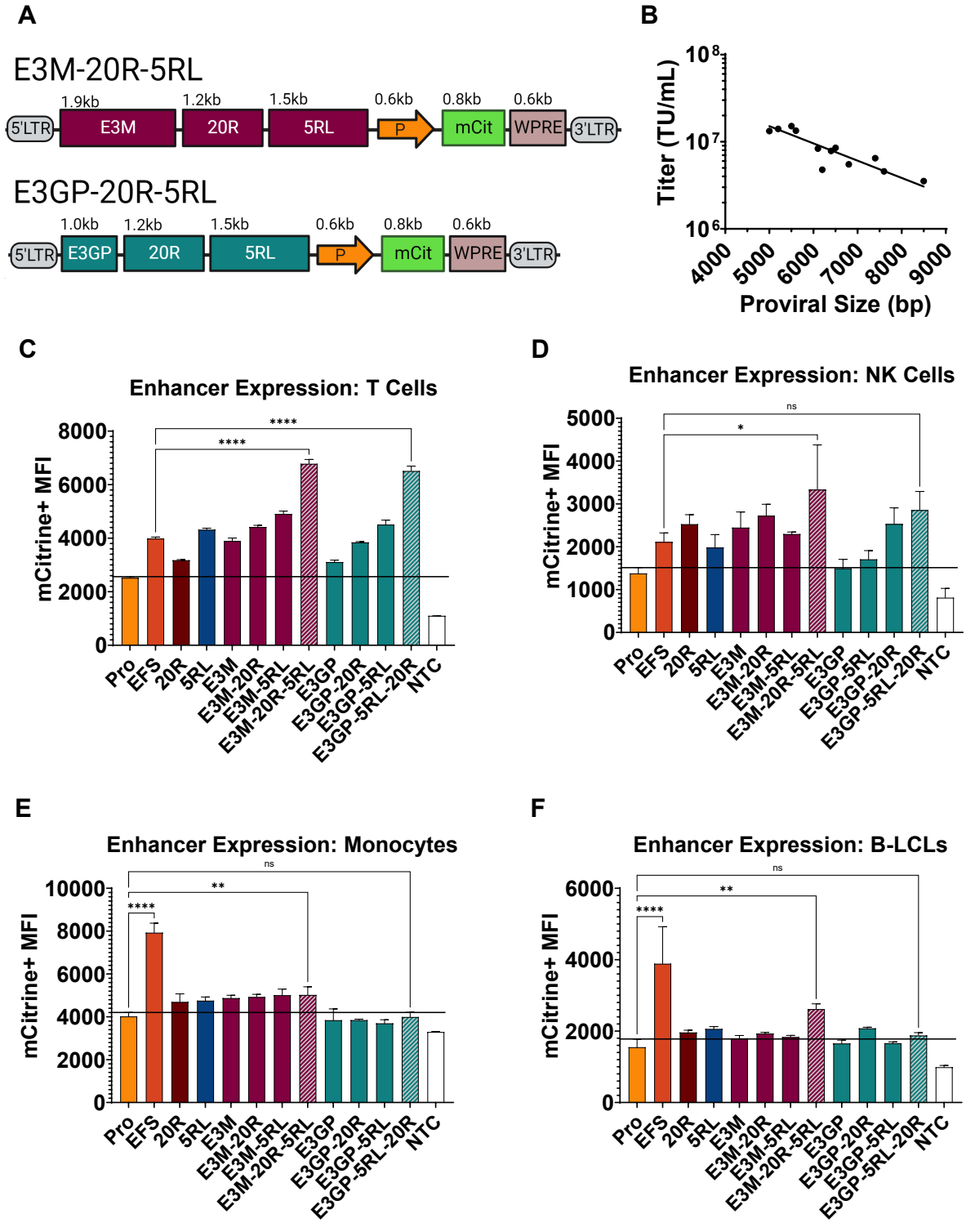

A
E3M-20R-5RL
1.9kb
1.2kb
1.5kb
0.6kb
0.8kb
0.6kb
5'LTR
E3M
20R
5RL
P
mCit
WPRE
3'LTR
E3GP-20R-5RL
1.0kb
1.2kb
1.5kb
0.6kb
0.8kb
0.6kb
5'LTR
E3GP
20R
5RL
P
mCit
WPRE
3'LTR
B
Titer (TU/mL)
Proviral Size (bp)
C
Enhancer Expression: T Cells
mCitrine+ MFI
D
Enhancer Expression: NK Cells
mCitrine+ MFI
E
Enhancer Expression: Monocytes
mCitrine+ MFI
F
Enhancer Expression: B-LCLs
mCitrine+ MFI
Pro
EFS
20R
5RL
E3M
E3M-20R
E3M-5RL
E3M-20R-5RL
E3GP
E3GP-20R
E3GP-5RL
E3GP-5RL-20R
NTC

Figure 2. 2

A: Schematic of composite XLP1-SMART lentiviral vectors. Diagrams of the two XLP1-SMART-LV composite constructs are shown with their sequence lengths (kb). 5'LTR and 3' LTR designate the 5' and 3' viral long terminal repeats (LTR), respectively; E3M, E3GP, 20R, and 5RL are *SH2D1A* enhancer elements; P designates the *SH2D1A* promoter; mCit, mCitrine reporter cassette; WPRE, Woodchuck Hepatitis Virus post-transcriptional regulatory element.

B: **Proviral size of composite XLP1-SMART LVs vs Titer.** Enhancers were cloned into the plasmid backbone of a therapeutic lentiviral vector (pCCL-c-MNDU3-X [Addgene Plasmid #81071]) (with the MNDU3 promoter removed), packaged, and titered head-to-head. The quantities of infectious particles were plotted as a function of proviral length (bp). Each point in the plot represents an average of three individual 10-cm plate of virus titered on HT-29 cells. Proviral length is defined as sequence length from the beginning of the 5′ long terminal repeat (LTR) U3 through the end of the 3′ LTR U5. n = 3 per arm. Linear regression analyses were used to determine the correlation between Titer and proviral size (R^2=0.79).

C: On-Target expression in T Cells *in vitro* of composite XLP1-SMART LVs. Healthy Donor CD3+ T cells were isolated from peripheral blood mononuclear cells (PBMCs). CD3+ T cells were transduced with each XLP1-SMART LV to achieve a VCN ranging from 0.1-0.2. at 14 days post transduction; T cells were assessed for the relative expression driven by each enhancer via mCitrine+ MFI using flow cytometry. Each enhancer was compared to basal *SH2D1A* promoter expression (Pro) and the

control LV (EFS). Data are represented as mean ± SD of biological triplicates from three experiments. Statistical significance was analyzed using a one-way ANOVA followed by multiple paired comparisons for normally distributed data (Tukey test). Statistical analysis was performed on all arms, but selected arms are shown. All statistical tests were two-tailed and a p value of < 0.05 was deemed significant (ns non-significant, *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.).

D: On-Target expression in NK cells *in vitro* of composite XLP1-SMART LVs. Healthy Donor CD56+ NK cells were isolated from peripheral blood mononuclear cells (PBMCs). CD56+ NK Cells were transduced with each XLP1-SMART LV to achieve a VCN ranging from 0.1-0.2. at 14 days post transduction, NK cells were assessed for the relative expression driven by each enhancer via mCitrine+ MFI using flow cytometry. Each enhancer was compared to basal *SH2D1A* promoter expression (Pro) and the control LV (EFS). Data are represented as mean ± SD of biological triplicates from three experiments. Statistical significance was analyzed using a one-way ANOVA followed by multiple paired comparisons for normally distributed data (Tukey test). Statistical analysis was performed on all arms, but selected arms are shown. All statistical tests were two-tailed and a p value of < 0.05 was deemed significant (ns non-significant, *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.).

E. Off-Target expression in CB CD34+ differentiated monocytes cells by composite XLP1-SMART LVs. Healthy Donor CB CD34+ cells were differentiated into monocytes as described.[41] Prior to differentiation, CB CD34+ cells were transduced with each XLP1-SMART LV to achieve a VCN ranging from 0.1-0.2. at 14 days post

transduction and differentiation, CD14+CD16+ monocytes were assessed for the relative expression driven by each enhancer via mCitrine+ MFI using flow cytometry. Each enhancer was compared to basal *SH2D1A* promoter expression (Pro) and the control LV (EFS). Data are represented as mean ± SD of biological triplicates from one experiment. Statistical significance was analyzed using a one-way ANOVA followed by multiple paired comparisons for normally distributed data (Tukey test). Statistical analysis was performed on all arms, but selected arms are shown. All statistical tests were two-tailed and a p value of < 0.05 was deemed significant (ns non-significant, *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.).

F. Off-Target expression in B-LCLs by composite XLP1-SMART LVs. B-LCLs, cultured in R10, were transduced with each XLP1-SMART LV to achieve a VCN ranging from 0.1-0.2. at 14 days post transduction, B-LCLs were assessed for the relative expression driven by each enhancer via mCitrine+ MFI using flow cytometry. Each enhancer was compared to basal *SH2D1A* promoter expression (Pro) and the control LV (EFS). Data are represented as mean ± SD of biological triplicates from three experiments. Statistical significance was analyzed using a one-way ANOVA followed by multiple paired comparisons for normally distributed data (Tukey test). Statistical analysis was performed on all arms, but selected arms are shown. All statistical tests were two-tailed and a p value of < 0.05 was deemed significant (ns non-significant, *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.).

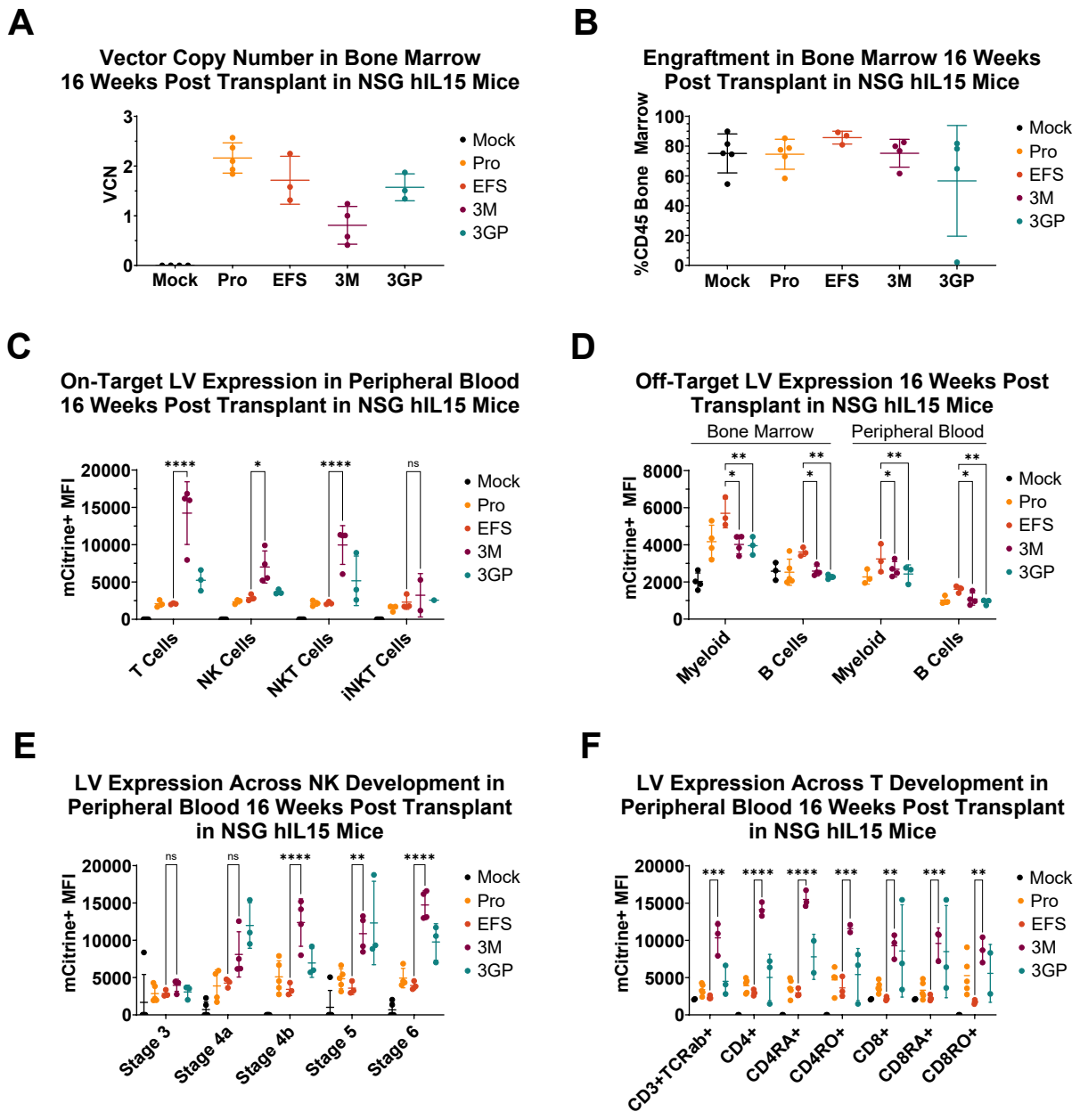


Figure 2. 3

A: Vector Copy Number 16 weeks post-transplant from NSG-Tg(Hu-IL15) mouse bone marrow (BM). Whole BM was taken from each mouse at time of euthanasia and processed into single cell suspension. Genomic DNA was extracted from the BM suspension and analyzed for vector copy number by ddPCR. n = 4, Mock; n = 5,

promoter only (Pro); n = 3, EFS-mCitrine (EFS); n = 4, E3M-20R-5RL-mCitrine (E3M); n = 3, E3GP-20R-5RL-mCitrine (E3GP).

B: Engraftment 16 weeks post-transplant from NSG-Tg(Hu-IL15) mouse bone marrow (BM). Whole BM was taken from each mouse at time of euthanasia and analyzed for engraftment by flow cytometry using an anti-hCD45 antibody. n = 5, Mock; n = 5, promoter only (Pro); n = 3, EFS-mCitrine (EFS); n = 4, E3M-20R-5RL-mCitrine (E3M); n = 4, E3GP-20R-5RL-mCitrine (E3GP).

C: On-target XLP1-SMART-LV expression in peripheral blood 16 weeks post-transplant in NSG-Tg(Hu-IL15) mice. Mice were bled at 16 weeks post-transplant to analyze peripheral blood for XLP1-SMART-LV expression. Lysed red blood cells were stained for various on-target lineages within the hCD45+ gate (T cells: hCD33-, hCD19-, hCD3+; NK Cells: hCD33-, hCD3-, hCD19-, hCD56+; NKT Cells: hCD33-, hCD19-, hCD3+, hCD56+; and iNKT cells: hCD33-, hCD19-, hCD3+, hCD56+, hVα24+). Each LV's relative expression was measured in on-target lineages via mCitrine+ MFI using flow cytometry. Each enhancer was compared to basal *SH2D1A* promoter-driven expression (Pro) and the control LV (EFS). Data are represented as mean ± SD of biological triplicates from one experiment. Statistical significance was analyzed using a two-way ANOVA followed by multiple paired comparisons for normally distributed data (Tukey test). Statistical analysis was performed on all arms, but selected arms are shown. All statistical tests were two-tailed and a p value of < 0.05 was deemed significant (ns non-significant, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$).

D: Off-target XLP1-SMART-LV expression in bone marrow (BM) and peripheral blood (PB) 16 weeks post-transplant in NSG-Tg(Hu-IL15) mice. Whole BM was taken from each mouse at time of euthanasia and mice were bled at 16 weeks post-transplant to analyze PB for XLP1-SMART-LV expression. Whole BM, processed into single cell suspension, and RBC-lysed PB were stained for various off-target lineages within the hCD45+ gate (Myeloid Cells: hCD33+; B Cells: hCD33-, hCD19+, hCD3-). Each LV's relative expression was measured in off-target lineages via mCitrine+ MFI using flow cytometry. Each enhancer was compared to basal *SH2D1A* promoter expression (Pro) and the control LV (EFS). Data are represented as mean ± SD of biological triplicates from one experiment. Statistical significance was analyzed using a two-way ANOVA followed by multiple paired comparisons for normally distributed data (Tukey test). Statistical analysis was performed on all arms, but selected arms are shown. All statistical tests were two-tailed and a p value of < 0.05 was deemed significant (ns non-significant, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.).

E: XLP1-SMART-LV expression across NK cell development in peripheral blood (PB) 16 weeks post-transplant in NSG-Tg(Hu-IL15) mice. Mice were bled at 16 weeks to analyze PB for XLP1-SMART-LV expression after red blood cell lysis. Cells were stained for various stages of NK cell differentiation within the hCD45+CD33- gate (Stage 1 [data not shown]: hCD34+; Stage 2a [data not shown]: hCD34+, hCD117+, hCD122-; Stage 2b [data not shown]: hCD34+, hCD117+, hCD122+; Stage 3: hCD34-, hCD117+, hCD122+, hCD56-; Stage 4a: hCD34-, hCD117+, hCD122+, hCD56+, hCD94+; Stage 5: hCD34-, hCD117-, hCD122+, hCD56+, hCD94+, hNKp80+; Stage 5: hCD34-, hCD117-, hCD122+, hCD56+, hCD94+, hNKp80+, hCD16+; and Stage 6:

hCD34-, hCD117-, CD122+, hCD56+, hCD94+, hNKp80+, hCD16+, hCD57). Each LV's relative expression was measured across NK-cell subpopulations via mCitrine+ MFI using flow cytometry. Each enhancer was compared to basal *SH2D1A* promoter expression (Pro) and the control LV (EFS). Data are represented as mean ± SD of biological triplicates from one experiment. Statistical significance was analyzed using a two-way ANOVA followed by multiple paired comparisons for normally distributed data (Tukey test). Statistical analysis was performed on all arms, but selected arms are shown. All statistical tests were two-tailed and a p value of < 0.05 was deemed significant (ns non-significant, *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.).

F: XLP1-SMART-LV expression across T cell development in peripheral blood 16 weeks post-transplant in NSG-Tg(Hu-IL15) mice. Mice were bled at 16 weeks to analyze PB for XLP1-SMART-LV expression after red blood cell lysis. Cells were stained for various stages of mature T-cell populations within the hCD45+ hCD34- hCD14- hCD19- hCD56- hCD5+ hCD7+ TCRab+ CD3+ gate (mature CD4: hCD4+, hCD8-; mature CD4RA: hCD4+, hCD8-, hCD45RA+, hCD45RO-; mature CD4RO: hCD4+, hCD8-, hCD45RA-, hCD45RO+; mature CD8: hCD4-, hCD8+; mature CD8RA: hCD4-, hCD8+, hCD45RA+, hCD45RO-; mature CD8RO: hCD4-, hCD8+, hCD45RA-, hCD45RO+). Each LV's relative expression was measured in mature T-cell subsets via mCitrine+ MFI using flow cytometry. Each enhancer was compared to basal *SH2D1A* promoter expression (Pro) and the control LV (EFS), both harboring an mCitrine reporter cassette. Data are represented as mean ± SD of biological triplicates from one experiment. Statistical significance was analyzed using a two-way ANOVA followed by multiple paired comparisons for normally distributed data (Tukey test). Statistical analysis

was performed on all arms, but selected arms are shown. All statistical tests were two-tailed and a p value of < 0.05 was deemed significant (ns non-significant, *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.).

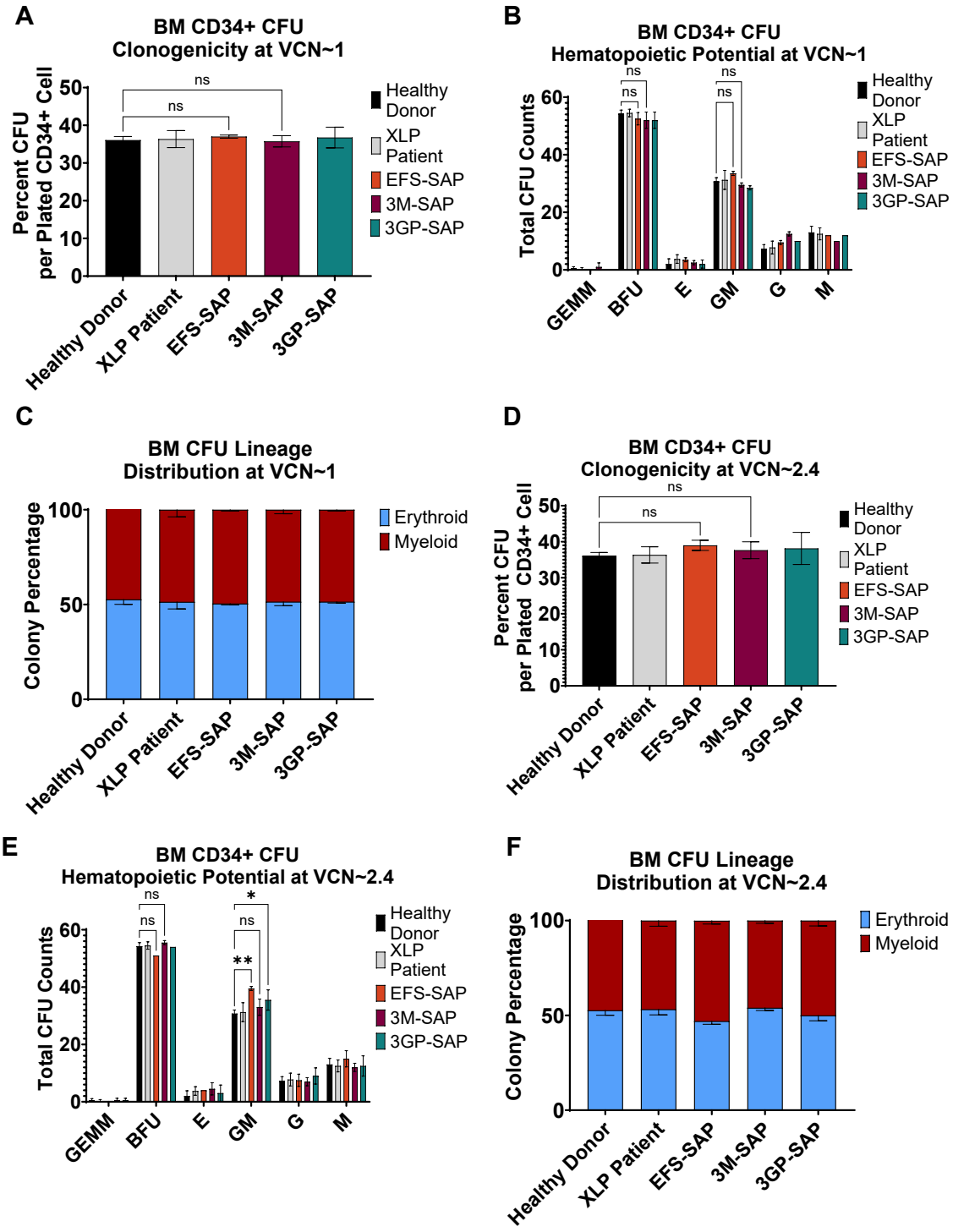

A
BM CD34+ CFU
Clonogenicity at VCN~1
Percent CFU per Plated CD34+ Cell
ns
ns
Healthy Donor
XLP Patient
EFS-SAP
3M-SAP
3GP-SAP
B
BM CD34+ CFU
Hematopoietic Potential at VCN~1
Total CFU Counts
GEMM
BFU
E
GM
G
M
C
BM CFU Lineage
Distribution at VCN~1
Colony Percentage
Erythroid
Myeloid
D
BM CD34+ CFU
Clonogenicity at VCN~2.4
E
BM CD34+ CFU
Hematopoietic Potential at VCN~2.4
*
**
F
BM CFU Lineage
Distribution at VCN~2.4

Figure 2. 4

Colony forming unit (CFU) Assay in XLP1 patient bone marrow (BM) CD34+ cells transduced with XLP1-SMART-LVs. BM CD34+ cells from a healthy donor (HD) and an XLP1 patient were prestimulated for 24 hours with 50 ng/mL each of human stem cell factor (hSCF), human thrombopoietin (hTPO), and human FMS-like tyrosine kinase 3 ligand (hFlt3-L) before transduction with XLP1-SMART LVs. 24-hours after transduction, 100, 300, and 900 BM CD34+ HSPCs per replicate were plated in MethoCult. After 14 days of culture at 5% CO2, 37°C and humidified atmosphere, the number of mature colonies were scored under the microscope for total colony forming units (CFU) at a VCN of 1.1 (A) or 2.4 (D); total hematopoietic progenitor cell counts at a VCN of 1.1 (B) or 2.4 (E), denoted as Myeloid (CFU-G/M/GM), Erythroid (BFU-E), or Mixed (CFU-GEMM); and finally, percentage of total Myeloid or Erythroid lineage distribution for cells at a VCN of 1.1 (C) or 2.4 (F). Data are represented as mean ± SD of biological duplicates from one experiment. Clonogenicity was analyzed for statistical significance using a one-way ANOVA followed by multiple paired comparisons for normally distributed data (Tukey test). CFU hematopoietic potential was analyzed statistical significance using a two-way ANOVA followed by multiple paired comparisons for normally distributed data (Tukey test). Statistical analysis was performed on all arms, but selected arms are shown. All statistical tests were two-tailed and a p value of < 0.05 was deemed significant (ns non-significant, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.).

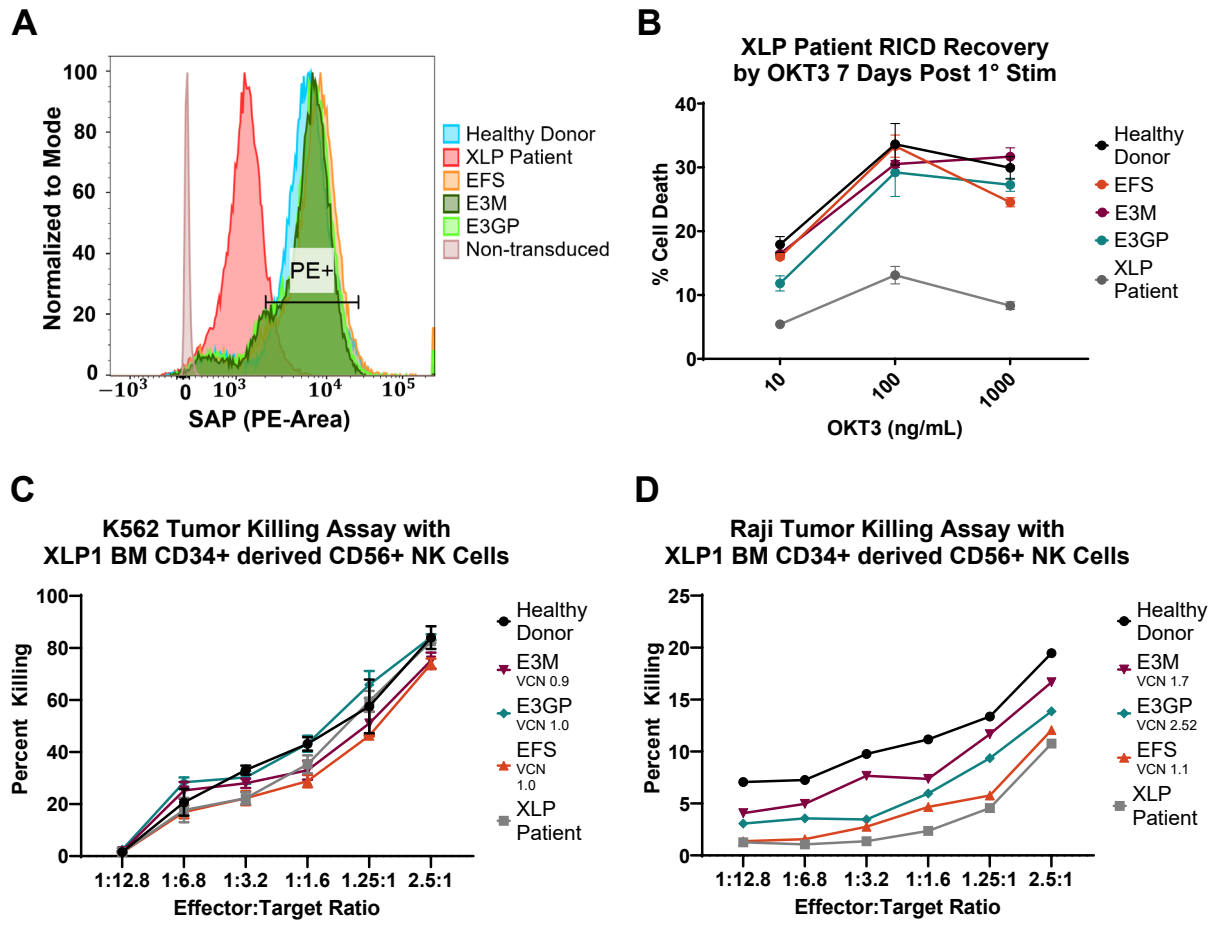


Figure 2. 5

A: FACS representation of SAP protein restoration in XLP1 patient CD8+ T cells. CD8+ T cells from a healthy donor (HD) and an XLP1 patient were isolated from PBMCs and transduced with XLP1-SMART-LVs. 10 days after transduction, cells were fixed, permeabilized and stained for SAP protein using an anti-SAP monoclonal antibody. Stained cells were then assessed for their SAP expression via total SAP MFI within each target subpopulation using flow cytometry.

B: T Cell Restimulation Induced Cell Death (RICD) assay of XLP1 patient CD8+ T cells transduced with XLP1-SMART-LVs. CD8+ T cells from a healthy donor (HD) and

an XLP1 patient were isolated from PBMCs and transduced with XLP1-SMART-LVs. 10 days after transduction, cells were plated for RICD assay in OKT3 at final concentrations of 1000 ng/ml, 100 ng/ml and 10 ng/ml. After 24 hours, XLP1-SMART-LV transduced cells were taken to measure the recovery of RICD in comparison to an EFS-SAP transduced condition and a HD control. The number of live cells (PI-) in stimulated controls were compared to unstimulated controls to measure the % cell loss = [1-(# PI- restimulated cells/# PI- untreated cells)]x100. Data are represented as mean ± SD of biological triplicates from one experiment. Statistical significance was analyzed using a two-way ANOVA followed by multiple paired comparisons for normally distributed data (Tukey test). All statistical tests were two-tailed and a p value of < 0.05 was deemed significant (ns non-significant, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.). Compared to SH2D1A knockout T cells (KO), the EFS, E3M, and E3GP conditions were deemed significant with a p value <0.0001.

C: K562 NK cell cytotoxicity assay of XLP1 patient bone marrow (BM) CD34+ cells transduced with XLP1-SMART-LVs. BM CD34+ cells from a healthy donor (HD) and an XLP1 patient were transduced with XLP1-SMART LVs and differentiated into CD56+ NK cells using the StemSpan™ NK Cell Generation Kit. On day 28 of differentiation, CD56+ NK cells were enriched using magnetic bead isolation and serially diluted with target cells at various effector to target (K562) ratios – 2.5:1, 1.25:1, 0.1:1.6, 1:3.2, 1:6.8, 1:12.8. After 18 hours of incubation, GFP+ tumor cells were counted via FACS to assess NK cytotoxicity and normalized to target only control wells. Due to limited XLP1 patient cells, data are represented as single replicates from one experiment. Statistical significance was analyzed using a two-way ANOVA followed by multiple paired

comparisons for normally distributed data (Tukey test). All statistical tests were two-tailed and a p value of < 0.05 was deemed significant (ns non-significant, *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.). Compared to XLP Patient samples, Healthy Donor was deemed significant at a p value <0.05; EFS was deemed significant with a p value <0.001; E3M was deemed not significant; and E3GP was deemed significant with a p value <0.0001.

D: Raji NK cell cytotoxicity assay of XLP1 patient bone marrow (BM) CD34+ cells transduced with XLP1-SMART-LVs. BM CD34+ cells from a healthy donor (HD) and an XLP1 patient were transduced with XLP1-SMART LVs and differentiated into CD56+ NK cells using the StemSpan™ NK Cell Generation Kit. On day 28 of differentiation, CD56+ NK cells were enriched using magnetic bead isolation and serially diluted with target cells at various effector to target (Raji) ratios – 2.5:1, 1.25:1, 0.1:1.6, 1:3.2, 1:6.8, 1:12.8. After 18 hours of incubation, 7AAD+ CFSE+ tumor cells were counted via FACS to assess NK cytotoxicity and normalized to target only control wells. Due to insufficient patient cells, data are represented as single replicates from one experiment. Statistical significance was analyzed using a two-way ANOVA followed by multiple paired comparisons for normally distributed data (Tukey test). Compared to XLP Patient samples, EFS was deemed not significant; E3M was deemed significant with a p value <0.0001; and E3GP was deemed significant with a p value <0.0001.

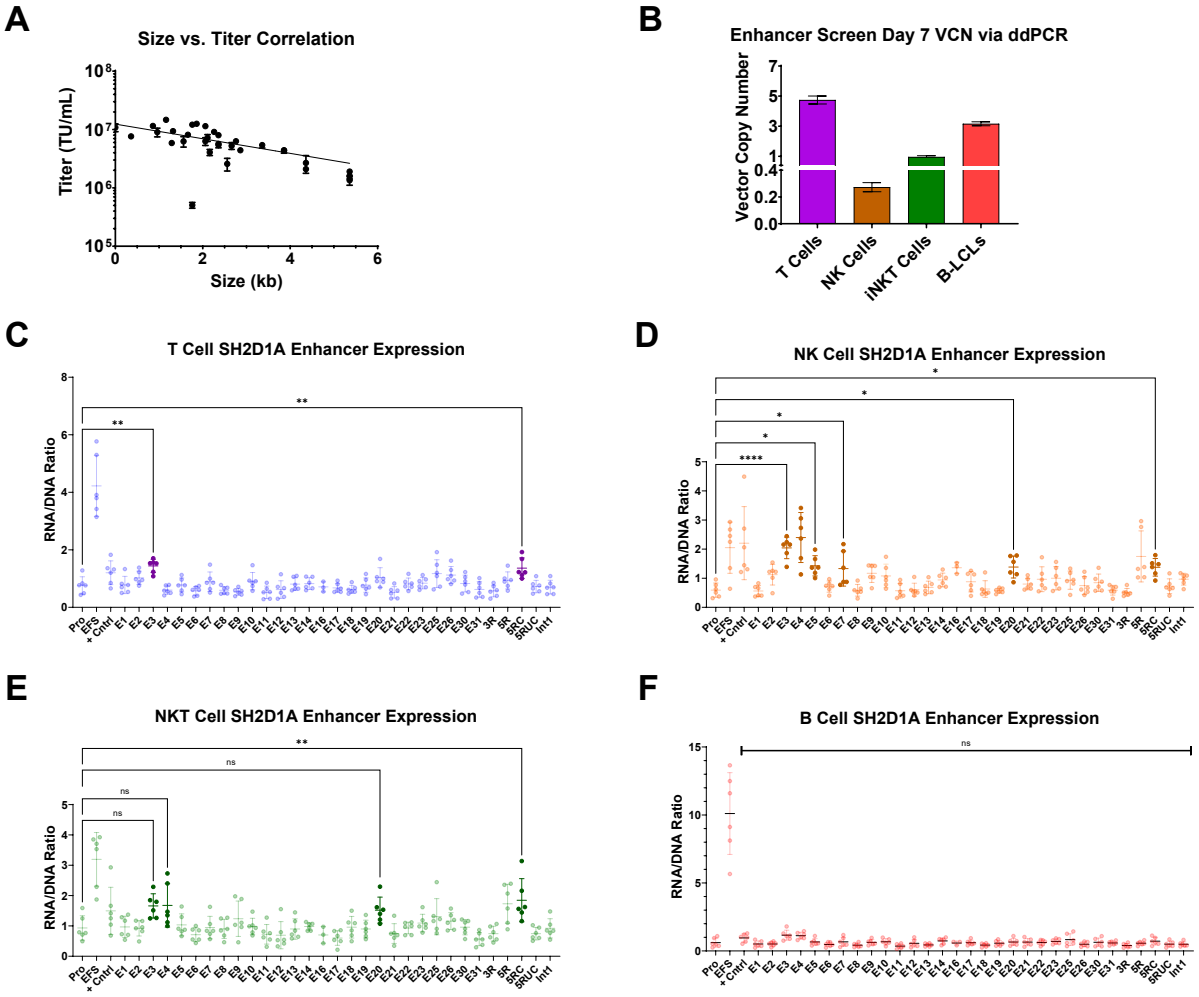


Figure 2. 6

A: Effect of Lentivirus Proviral Size on Titer. Putative Enhancers were cloned into the plasmid backbone of a therapeutic lentiviral vector (pCCL-c-MNDU3-X [Addgene Plasmid #81071]), packaged, and titered head-to-head, and the quantity of infectious particles were plotted as a function of proviral length (bp). Each point in the plot represents an average of three individual 10-cm plate of virus titered on HT-29 cells. Proviral length is defined as sequence length from the beginning of the 5′ long terminal repeat (LTR) U3 through the end of the 3′ LTR U5. n = 3 per arm. Linear regression

analyses were used to determine the correlation between titer and proviral size (R^2=0.51).

B: Enhancer Screen Bulk Vector Copy Number. We transduced primary T, NKT, and NK cells with a pool of raw viral supernatant containing $5x10^4$ TU/mL of the 34 candidate XLP1 SMART LVs and the EFS-mCitrine vector. B-LCLs were transduced to measure off-target expression. 14 days post-transduction, cells were harvested for their gDNA fraction to measure vector copy number using digital droplet PCR. Data are represented as mean ± SD of biological triplicates from one experiment.

C-F: Relative SH2D1A Enhancer Activity in T, NK, NKT, and B-LCL Cells. We transduced primary T, NKT, and NK cells with a pool of raw viral supernatant containing each of the 34 candidate XLP1 SMART LVs, in duplicate, and the EFS-SAP vector. Each candidate LV was cloned in with two unique barcodes. B-LCLs were transduced to measure off-target expression. 14 days post-transduction, cells were harvested for their gDNA and RNA fractions to measure barcode expression via next-generation sequencing. The RNA barcode counts identify active enhancers within each lineage whereas each genomic barcode count is used normalize each barcode in the transcript to the genome. The number of RNA barcode reads, normalized to frequency of gDNA barcodes within each cell type, determined the relative expression of each element. Data are represented as mean ± SD of biological triplicates from two experiments. We analyzed statistical significance using a one-way ANOVA followed by multiple paired comparisons for normally distributed data (Tukey test). All statistical tests were two-

tailed and a p value of < 0.05 was deemed significant (ns non-significant, *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.).

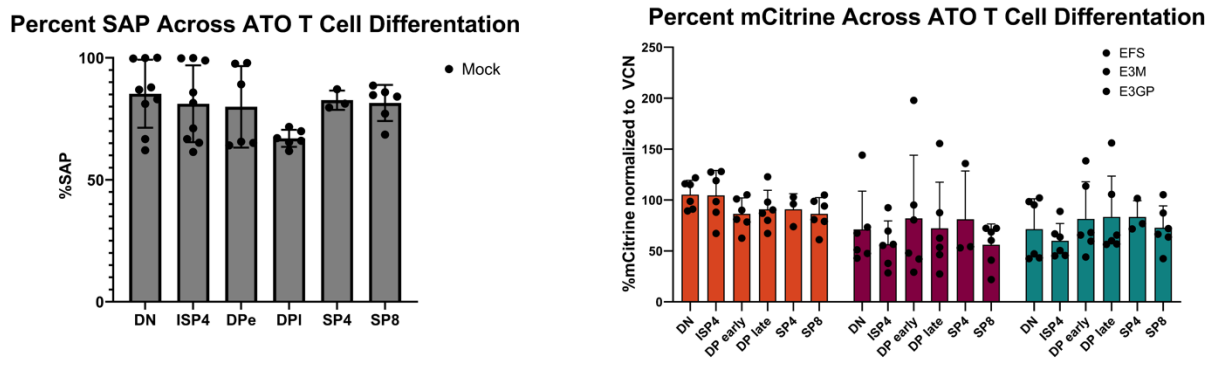


Figure 2. 7

A-B: Representative SAP+ and mCitrine+ Expression within Healthy Donor PBSC CD34+ Derived Artificial Thymic Organoid Cultures. To determine SAP expression and SMART LV expression across T cell maturation, healthy donor mobilized peripheral blood CD34+ cells were transduced with XLP1-SMART LVs harboring an mCitrine reporter cassette and differentiated into T cells using the 3D artificial thymic organoid (ATO) system. At weeks 3, 7, and 12 of differentiation, ATOs were harvested and stained to measure their relative (A) SAP expression (after fixation and permeabilization) or (B) mCitrine expression using flow cytometry. ATOs were stained with various antibodies to differentiate the following stages of T-cell development: CD5-CD7-, CD5-CD7+, CD5+CD7+, double negative (DN), immature single positive 4 (ISP4), double positive (DP) early, DP late, single positive (SP) 4, and SP8. A representative plot of the SAP+ percentage (A) and the mCitrine+ percentage (B) are across weeks 3, 7, and 12, in which data were normalized to the VCN of the bulk ATO populations. VCNs were as follows: EFS = 0.39, E3M = 0.184, E3GP = 0.25. Data are represented as mean ± SD of biological triplicates from one experiment across three timepoints (weeks 3, 7, and 12).

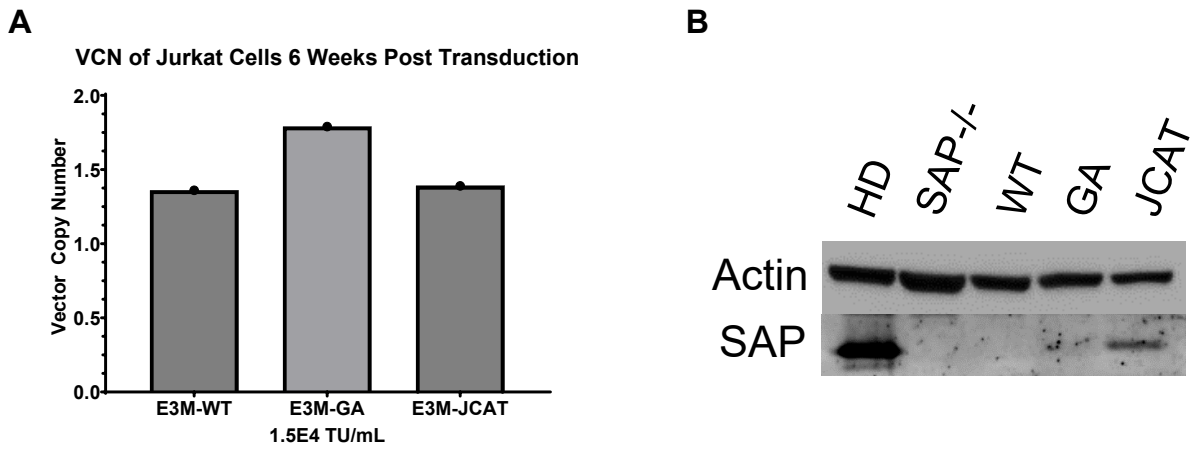


Figure 2. 8

A: Vector Copy Number in SH2D1A Knockout Jurkat Cells Transduced with Codon Optimized Vectors. SH2D1A -/- Jurkat cells were transduced at an equivalent vector copy number with the E3M-E20R-5RL lentiviral vector harboring an SH2D1A cDNA cassette containing wildtype codons or GeneArt (GA) or Java Codon Adaptation Tool (JCAT) codon optimizations. 14 days after transduction, cells were harvested for VCN measurement by ddPCR.

B: Western Blot Expression of SH2D1A Codon Optimized Vectors in Jurkat SH2D1A Knockout Cells. Protein was extracted from the transduced populations shown in panel A. The reconstitution of SAP protein after transduction was measured by western blot using an anti-SAP antibody (clone 1C9; Abnova) using a 1:1000 dilution of primary antibody.

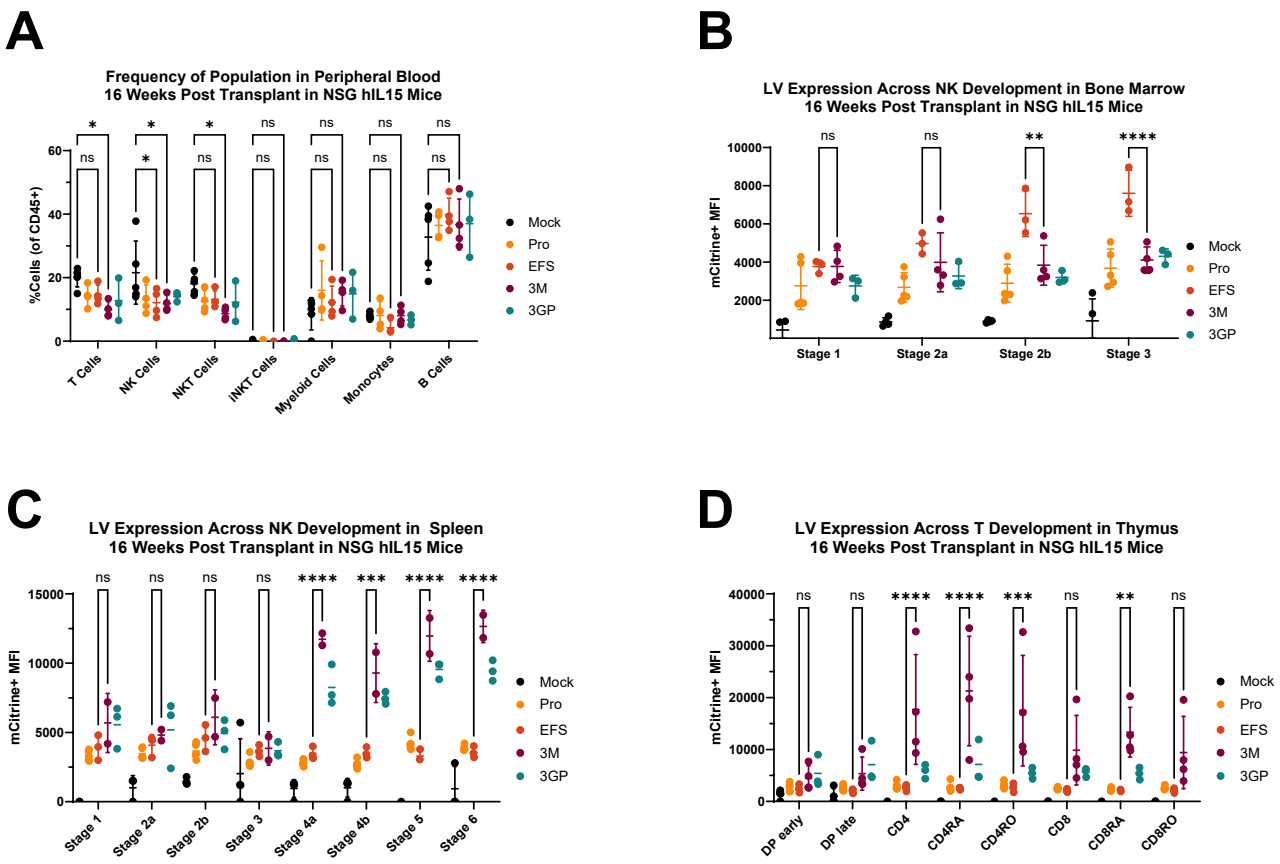


Figure 2. 9

A: Frequency of Populations in Peripheral Blood 16 Weeks Post Transplant in hIL15 NSG Mice. Mice were bled at 16 weeks to analyze peripheral blood for XLP1-SMART-LV expression. Lysed red blood cells were stained for various lineages within the hCD45+ gate (T Cells: hCD33- hCD19- hCD3+; NK Cells: hCD33- hCD3- hCD19- hCD56+; NKT Cells: hCD33- hCD19- hCD3+ hCD56+; and iNKT cells: hCD33- hCD19- hCD3+ hCD56+ hVα24+; Myeloid Cells: hCD33+; Monocytes: hCD33+, CD14+, CD16-; B Cells: hCD33- hCD19+ hCD3-). Frequency of each population is plotted as a percentage of total hCD45+ cells using flow cytometry. Data are represented as mean ± SD of biological triplicates from one experiment. We analyzed statistical significance

using a two-way ANOVA followed by multiple paired comparisons for normally distributed data (Tukey test). All statistical tests were two-tailed and a p value of < 0.05 was deemed significant (ns non-significant, *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.).

B: XLP1-SMART-LV Expression across NK-cell Development in Bone Marrow 16 Weeks Post Transplant in hIL15 NSG Mice. Whole bone marrow (BM) was taken from each mouse at time of euthanasia and processed into single cell suspension. Single cells were stained for various stages of NK-cell differentiation within the hCD45+CD33- gate (Stage 1: hCD34+; Stage 2a: hCD34+ hCD117+ hCD122-; Stage 2b: hCD34+h CD117+h CD122+; Stage 3: hCD34- hCD117+ hCD122+ hCD56-) Stage 4a: hCD34- hCD117+ hCD122+ hCD56+ hCD94+; Stage 4b: hCD34- hCD117- hCD122+ hCD56+ hCD94+ hNKp80+; Stage 5: hCD34- hCD117- hCD122+ hCD56+ hCD94+ hNKp80+ hCD16+; and Stage 6: hCD34- hCD117- CD122+ hCD56+ hCD94+ hNKp80+ hCD16+ hCD57). Each LV's relative expression was measured via mCitrine+ MFI using flow cytometry. Each enhancer was compared to basal SH2D1A promoter expression (Pro) and the preclinical vector (EFS). Data are represented as mean ± SD of biological triplicates from one experiment. We analyzed statistical significance using a two-way ANOVA followed by multiple paired comparisons for normally distributed data (Tukey test). All statistical tests were two-tailed and a p value of < 0.05 was deemed significant (ns non-significant, *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.).

C: XLP1-SMART-LV Expression across NK-cell Development in Spleen 16 Weeks Post Transplant in hIL15 NSG Mice. The spleen was taken from each mouse at time

of euthanasia, processed into a single cell suspension, and lysed. Lysed spleen single cells were stained for various stages of NK-cell differentiation within the hCD45+CD33- gate (Stage 1: hCD34+; Stage 2a: hCD34+ hCD117+ hCD122-; Stage 2b: hCD34+h CD117+h CD122+; Stage 3: hCD34- hCD117+ hCD122+ hCD56-; Stage 4a: hCD34- hCD117+ hCD122+ hCD56+ hCD94+; Stage 4b: hCD34- hCD117- hCD122+ hCD56+ hCD94+ hNKp80+; Stage 5: hCD34- hCD117- hCD122+ hCD56+ hCD94+ hNKp80+ hCD16+; and Stage 6: hCD34- hCD117- CD122+ hCD56+ hCD94+ hNKp80+ hCD16+ hCD57). Each LV's relative expression was measured via mCitrine+ MFI using flow cytometry. Each enhancer was compared to basal SH2D1A promoter expression (Pro) and the preclinical vector (EFS). Data are represented as mean ± SD of biological triplicates from one experiment. We analyzed statistical significance using a two-way ANOVA followed by multiple paired comparisons for normally distributed data (Tukey test). All statistical tests were two-tailed and a p value of < 0.05 was deemed significant (ns non-significant, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.).

D: XLP1-SMART-LV Expression across T-cell Development in the Thymus 16 Weeks Post Transplant in hIL15 NSG Mice. The thymus was taken from each mouse at time of euthanasia and processed into a single cell suspension. The single cell suspension was stained for various stages of T-cell differentiation within the hCD45+ hCD34- hCD14- hCD19- hCD56- hCD5+ hCD7+ TCRab+ CD3+ gate (mature CD4: hCD4+ hCD8-; mature CD4RA: hCD4+ hCD8- hCD45RA+ hCD45RO-; mature CD4RO: hCD4+ hCD8- hCD45RA- hCD45RO+; mature CD8: hCD4- hCD8+ mature CD8RA: hCD4- hCD8+ hCD45RA+ hCD45RO-; mature CD8RO: hCD4- hCD8+ hCD45RA- hCD45RO+). Each LV's relative expression was measured via mCitrine+ MFI using flow

cytometry. Each enhancer was compared to basal SH2D1A promoter expression (Pro) and the preclinical vector (EFS), both harboring an mCitrine reporter cassette. Data are represented as mean ± SD of biological triplicates from one experiment. We analyzed statistical significance using a two-way ANOVA followed by multiple paired comparisons for normally distributed data (Tukey test). All statistical tests were two-tailed and a p value of < 0.05 was deemed significant (ns non-significant, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.).

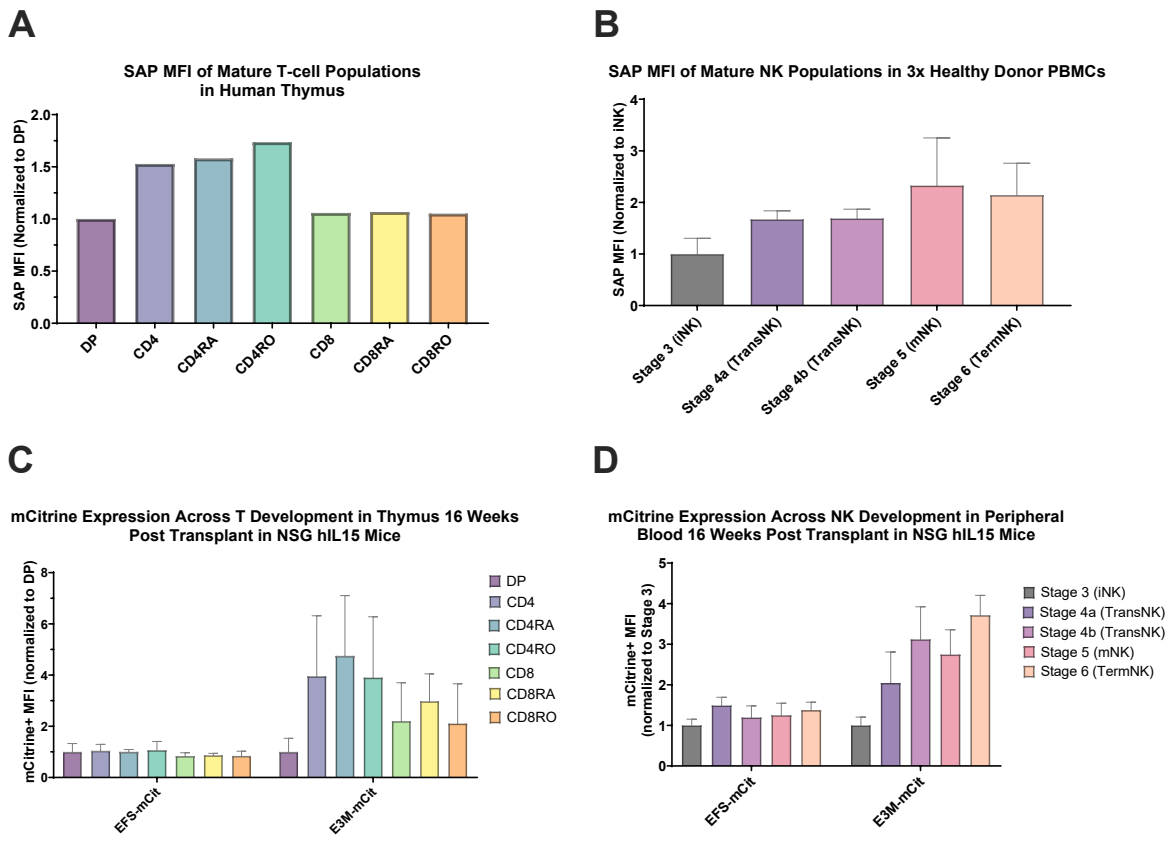


Figure 2. 10

A: SAP Mean Fluorescent Intensity in T-cell Populations in a Human Thymus: Healthy Donor human thymic tissue was processed into single cell suspensions. The single cell suspension was stained for various stages of T-cell differentiation within the hCD45+ hCD34- hCD14- hCD19- hCD56- hCD5+ hCD7+ TCRab+ CD3+ gate (mature CD4: hCD4+ hCD8-; mature CD4RA: hCD4+ hCD8- hCD45RA+ hCD45RO-; mature CD4RO: hCD4+ hCD8- hCD45RA- hCD45RO+; mature CD8: hCD4- hCD8+ mature CD8RA: hCD4- hCD8+ hCD45RA+ hCD45RO-; mature CD8RO: hCD4- hCD8+ hCD45RA- hCD45RO+). Stained cells were then fixed, permeabilized and stained for SAP using an anti-SAP monoclonal antibody. Stained cells were then assessed for their

SAP expression via total SAP MFI within each target subpopulation using flow cytometry.

B: SAP Mean Fluorescent Intensity in Mature NK-cell Populations in Three Healthy Donor PBMC Samples: CD56+ NK cells from a healthy donor (HD) and an XLP1 patient were isolated from PBMCs. Cells were stained with monoclonal antibodies for various stages of NK development (Stage 3: hCD34- hCD117+ hCD122+ hCD56-; Stage 4a: hCD34- hCD117+ hCD122+ hCD56+ hCD94+; Stage 4b: hCD34- hCD117- hCD122+ hCD56+ hCD94+ hNKp80+; Stage 5: hCD34- hCD117- hCD122+ hCD56+ hCD94+ hNKp80+ hCD16+; and Stage 6: hCD34- hCD117- CD122+ hCD56+ hCD94+ hNKp80+ hCD16+ hCD57). Stained cells were then fixed, permeabilized and stained for SAP using an anti-SAP monoclonal antibody. Stained cells were then assessed for their SAP expression via total SAP MFI within each target subpopulation using flow cytometry.

C-D: XLP1-SMART-LV Expression across T-cell and NK-cell Development 16 Weeks Post Transplant in hIL15 NSG mice. The following graphs are representative data shown in Figure 2.3, panels E and F. They depict XLP1-SMART-LV Expression across T-cell Development (C) and NK-cell development (D) 16 Weeks Post Transplant in hIL15 NSG mice. mCitrine+ MFI was detected in T and NK cell subpopulations to assess relative XLP1-SMART-LV pattern of expression in comparison to patterns of SAP expression shown in Supplemental Figure 5 panels A and B.

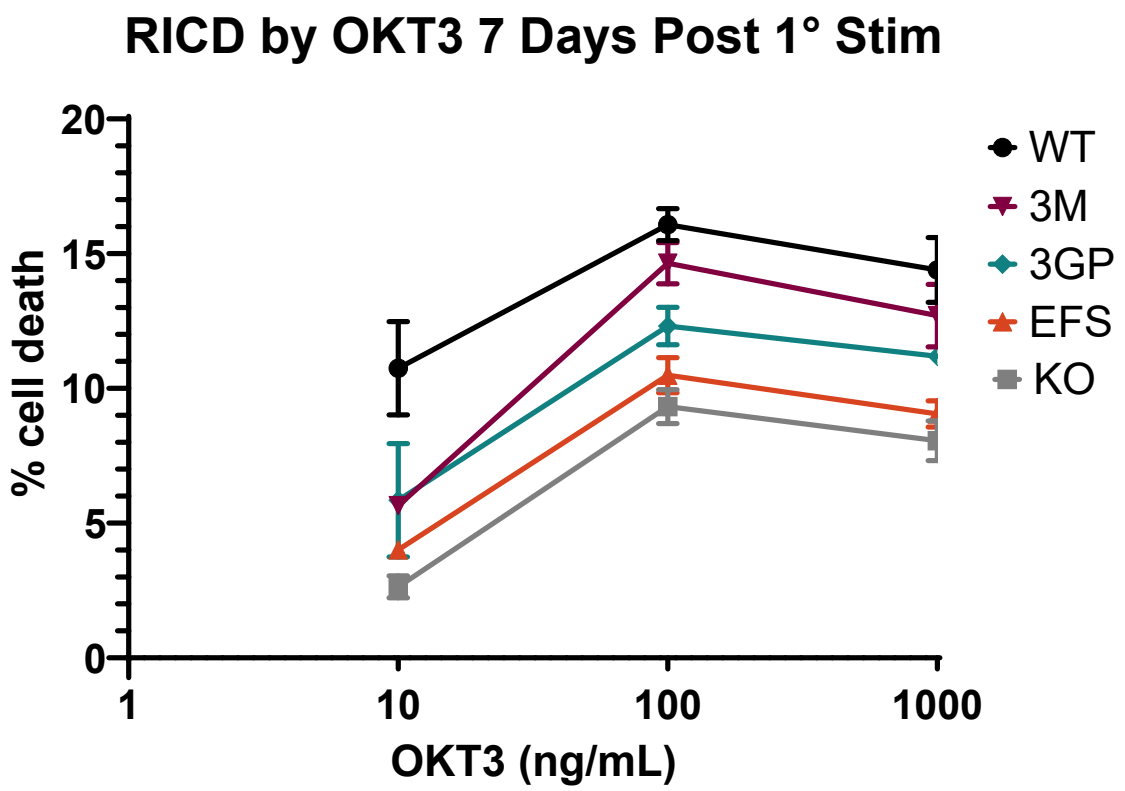


Figure 2.11

A: T Cell Restimulation Induced Cell Death (RICD) Assay of SH2D1A-/- CD8+ T-Cells Transduced with XLP1-SMART-LVs. CD8+ T cells from a healthy donor (HD) and were isolated from PBMCs and transfected with sgRNAs targeting Exon 2 of the SH2D1A gene to knockout SAP expression. SH2D1A -/- T cells were transduced with XLP1-SMART-LVs to achieve an equivalent VCN of 2.5 (VCNs are EFS = 2.79, E3M = 2.5, E3GP = 2.46). 10 days after transduction, cells were plated for RICD in OKT3 at final concentrations of 1000 ng/ml, 100 ng/ml and 10 ng/ml. After 24 hours, XLP1-SMART-LV transduced cells were taken to measure the recovery of RICD in comparison to a HD control. The number of live cells (PI-) in stimulated controls were compared to unstimulated controls to measure the % cell loss = [1-(# PI- restimulated cells/# PI-

untreated cells)]x100. Data are represented as mean ± SD of biological triplicates from one experiment. We analyzed statistical significance using a two-way ANOVA followed by multiple paired comparisons for normally distributed data (Tukey test). All statistical tests were two-tailed and a p value of < 0.05 was deemed significant (ns non-significant, *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.). Compared to XLP Patient samples, EFS was deemed significant with a p value <0.05; E3M was deemed significant with a p value <0.0001; and E3GP was deemed significant with a p value <0.0001.

REFERENCES

1 Booth, C. *et al.* X-linked lymphoproliferative disease due to SAP/SH2D1A deficiency: a multicenter study on the manifestations, management and outcome of the disease. *Blood* **117**, 53-62 (2011). https://doi.org:10.1182/blood-2010-06-284935

2 Ma, C. S., Nichols, K. E. & Tangye, S. G. Regulation of cellular and humoral immune responses by the SLAM and SAP families of molecules. *Annu Rev Immunol* **25**, 337-379 (2007). https://doi.org:10.1146/annurev.immunol.25.022106.141651

3 Coffey, A. J. *et al.* Host response to EBV infection in X-linked lymphoproliferative disease results from mutations in an SH2-domain encoding gene. *Nature Genetics* **20**, 129-135 (1998). https://doi.org:10.1038/2424

4 Latour, S. *et al.* Regulation of SLAM-mediated signal transduction by SAP, the X-linked lymphoproliferative gene product. *Nature Immunology* **2**, 681-690 (2001). https://doi.org:10.1038/90615

5 Sayos, J. *et al.* The X-linked lymphoproliferative-disease gene product SAP regulates signals induced through the co-receptor SLAM. *Nature* **395**, 462-469 (1998). https://doi.org:10.1038/26683

6 Dupré, L. C. *et al.* SAP controls the cytolytic activity of CD8+ T cells against EBV-infected cells. *Blood* **105**, 4383-4389 (2005). https://doi.org:10.1182/blood-2004-08-3269

7 Crotty, S., Kersh, E. N., Cannons, J., Schwartzberg, P. L. & Ahmed, R. SAP is required for generating long-term humoral immunity. *Nature* **421**, 282-287 (2003). https://doi.org:10.1038/nature01318

8 Panchal, N. *et al.* Transfer of gene-corrected T cells corrects humoral and cytotoxic defects in patients with X-linked lymphoproliferative disease. *Journal of Allergy and Clinical Immunology* **142**, 235-245.e236 (2018). https://doi.org:10.1016/j.jaci.2018.02.053

9 Veillette, A. *et al.* SAP expression in T cells, not in B cells, is required for humoral immunity. *Proceedings of the National Academy of Sciences* **105**, 1273-1278 (2008). https://doi.org:10.1073/pnas.0710698105

10 Sullivan, J. L., Byron, K. S., Brewster, F. E., Baker, S. M. & Ochs, H. D. X-linked lymphoproliferative syndrome. Natural history of the immunodeficiency. *Journal of Clinical Investigation* **71**, 1765-1778 (1983). https://doi.org:10.1172/jci110932

11 Rivat, C. *et al.* SAP gene transfer restores cellular and humoral immune function in a murine model of X-linked lymphoproliferative disease. *Blood* **121**, 1073-1076 (2013). https://doi.org:10.1182/blood-2012-07-445858

12 Tangye, S. G., Cherwinski, H., Lanier, L. L. & Phillips, J. H. 2B4-mediated activation of human natural killer cells. *Molecular Immunology* **37**, 493-501 (2000). https://doi.org:https://doi.org/10.1016/S0161-5890(00)00076-6

13 Sharifi, R. *et al.* SAP mediates specific cytotoxic T-cell functions in X-linked lymphoproliferative disease. *Blood* **103**, 3821-3827 (2004). https://doi.org:10.1182/blood-2003-09-3359

14 Parolini, S. *et al.* X-Linked Lymphoproliferative Disease. *Journal of Experimental Medicine* **192**, 337-346 (2000). https://doi.org:10.1084/jem.192.3.337

15 Ma, C. S. *et al.* Impaired humoral immunity in X-linked lymphoproliferative disease is associated with defective IL-10 production by CD4+ T cells. **115**, 1049-1059 (2005). https://doi.org:10.1172/jci23139

16 Qi, H., Cannons, J. L., Klauschen, F., Schwartzberg, P. L. & Germain, R. N. SAP-controlled T–B cell interactions underlie germinal centre formation. *Nature* **455**, 764-769 (2008). https://doi.org:10.1038/nature07345

17 Priatel, J. J., Chung, B. K., Tsai, K. & Tan, R. Natural killer T cell strategies to combat Epstein–Barr virus infection. *OncoImmunology* **3**, e28329 (2014). https://doi.org:10.4161/onci.28329

18 Poletti, V. & Mavilio, F. Designing Lentiviral Vectors for Gene Therapy of Genetic Diseases. *Viruses* **13** (2021). https://doi.org:10.3390/v13081526

19 Nagy, N. *et al.* The proapoptotic function of SAP provides a clue to the clinical picture of X-linked lymphoproliferative disease. *Proc Natl Acad Sci U S A* **106**, 11966-11971 (2009). https://doi.org:10.1073/pnas.0905691106

20 Gartshteyn, Y., Askanase, A. D. & Mor, A. SLAM Associated Protein Signaling in T Cells: Tilting the Balance Toward Autoimmunity. *Front Immunol* **12**, 654839 (2021). https://doi.org:10.3389/fimmu.2021.654839

21 Passerini, L., Santoni de Sio, F. R., Roncarolo, M. G. & Bacchetta, R. Forkhead box P3: the peacekeeper of the immune system. *Int Rev Immunol* **33**, 129-145 (2014). https://doi.org:10.3109/08830185.2013.863303

22 Santilli, G. *et al.* Biochemical correction of X-CGD by a novel chimeric promoter regulating high levels of transgene expression in myeloid cells. *Mol Ther* **19**, 122-132 (2011). https://doi.org:10.1038/mt.2010.226

23 Wong, R. L. *et al.* Lentiviral gene therapy for X-linked chronic granulomatous disease recapitulates endogenous CYBB regulation and expression. *Blood* **141**, 1007-1022 (2023). https://doi.org:10.1182/blood.2022016074

24 Fishilevich, S. *et al.* GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)* **2017** (2017). https://doi.org:10.1093/database/bax028

25 Cooper, A. R. *et al.* Highly efficient large-scale lentiviral vector concentration by tandem tangential flow filtration. *J Virol Methods* **177**, 1-9 (2011). https://doi.org:10.1016/j.jviromet.2011.06.019

26 Morgan, R. A. *et al.* Improved Titer and Gene Transfer by Lentiviral Vectors Using Novel, Small β-Globin Locus Control Region Elements. *Molecular Therapy* **28**, 328-340 (2020). https://doi.org:https://doi.org/10.1016/j.ymthe.2019.09.020

27 Aryee, K. E. *et al.* Enhanced development of functional human NK cells in NOD-scid-IL2rg(null) mice expressing human IL15. *Faseb j* **36**, e22476 (2022). https://doi.org:10.1096/fj.202200045R

28 Brehm, M. A. *et al.* Parameters for establishing humanized mouse models to study human immunity: analysis of human hematopoietic stem cell engraftment in three immunodeficient strains of mice bearing the IL2rgamma(null) mutation. *Clin Immunol* **135**, 84-98 (2010). https://doi.org:10.1016/j.clim.2009.12.008

29 Montel-Hagen, A. *et al.* Organoid-Induced Differentiation of Conventional T Cells from Human Pluripotent Stem Cells. *Cell Stem Cell* **24**, 376-389.e378 (2019). https://doi.org:10.1016/j.stem.2018.12.011

30 Seet, C. S. *et al.* Generation of mature T cells from human hematopoietic stem and progenitor cells in artificial thymic organoids. *Nature Methods* **14**, 521-530 (2017). https://doi.org:10.1038/nmeth.4237

31 Grote, A. *et al.* JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Research* **33**, W526-W531 (2005). https://doi.org:10.1093/nar/gki376

32 Houghton, B. C. *et al.* Genome Editing With TALEN, CRISPR-Cas9 and CRISPR-Cas12a in Combination With AAV6 Homology Donor Restores T Cell Function for XLP. *Front Genome Ed* **4**, 828489 (2022). https://doi.org:10.3389/fgeed.2022.828489

33 Pohida, K., Lake, C. M., Yee, D. & Snow, A. L. Restimulation-Induced Cell Death (RICD): Methods for Modeling, Investigating, and Quantifying RICD Sensitivity in Primary Human T Cells via Flow Cytometric Analysis. *Bio Protoc* **12**, e4326 (2022). https://doi.org:10.21769/BioProtoc.4326

34 Kandarian, F., Sunga, G. M., Arango-Saenz, D. & Rossetti, M. A Flow Cytometry-Based Cytotoxicity Assay for the Assessment of Human NK Cell Activity. *J Vis Exp* (2017). https://doi.org:10.3791/56191

35 Aoukaty, A. & Tan, R. Association of the X-linked lymphoproliferative disease gene product SAP/SH2D1A with 2B4, a natural killer cell-activating molecule, is

dependent on phosphoinositide 3-kinase. *J Biol Chem* **277**, 13331-13337 (2002). https://doi.org:10.1074/jbc.M112029200

36 Xiang, Q. M. *et al.* Overexpression of SH2D1A promotes cancer progression and is associated with immune cell infiltration in hepatocellular carcinoma via bioinformatics and in vitro study. *BMC Cancer* **23**, 1005 (2023). https://doi.org:10.1186/s12885-023-11315-1

37 Han, J., Tam, K., Tam, C., Hollis, R. P. & Kohn, D. B. Improved lentiviral vector titers from a multi-gene knockout packaging line. *Mol Ther Oncolytics* **23**, 582-592 (2021). https://doi.org:10.1016/j.omto.2021.11.012

38 Hindson, B. J. *et al.* High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal Chem* **83**, 8604-8610 (2011). https://doi.org:10.1021/ac202028g

39 Zufferey, R. *et al.* Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. *J Virol* **72**, 9873-9880 (1998). https://doi.org:10.1128/jvi.72.12.9873-9880.1998

40 Katz, G. *et al.* FOXP3 renders activated human regulatory T cells resistant to restimulation-induced cell death by suppressing SAP expression. *Cellular Immunology* **327**, 54-61 (2018). https://doi.org:10.1016/j.cellimm.2018.02.007

41 Stec, M. *et al.* Expansion and differentiation of CD14+CD16(-) and CD14+ +CD16+ human monocyte subsets from cord blood CD34+ hematopoietic progenitors. *J Leukoc Biol* **82**, 594-602 (2007). https://doi.org:10.1189/jlb.0207117

Chapter 3: Optimizing a Gene Editing Approach for the Treatment of Cystic Fibrosis

ABSTRACT

Cystic Fibrosis (CF) is a monogenic hereditary disorder caused by mutations in the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) gene. Despite many new successes in treatments that alleviate some symptoms and complications, patients who are eligible still require a lifetime of treatment and a reduced lifespan. Further depending on the disease-causing mutation of the patient, these treatments may not provide any therapeutic benefits. Since 1990, researchers have worked to develop a therapeutic gene therapy strategy for patients suffering from CF, which could serve to help patients independent of their mutation type. Currently new developments in gene delivery, gene editing, and animal models are being explored. These new technologies along with new understandings of CF and the airway epithelium highlight a growing possibility for clinically therapeutic gene therapy options for CF. The following chapter describes the utilization of CRISPR/Cas9 machinery for site-specific insertion of a normal CFTR gene cassette into the endogenous CFTR locus to correct most CF-causing mutations and achieve appropriate expression to reduce the disease phenotype.

INTRODUCTION

Cystic fibrosis (CF) is an autosomal recessive monogenic disorder that causes persistent lung infections and progressive pulmonary failure. CF is caused by mutations within the cystic fibrosis transmembrane conductance regulator (CFTR), a gene encoding an ion transport channel.[4] A dysfunctional CFTR protein fails to transport chloride across the apical membrane, leading to infections, inflammation, lung destruction, and ultimately respiratory failure.[1] Despite progressive advances in CF management, including the introduction of CFTR-potentiators, not all patients benefit from these medications and it is likely that patients with very low CFTR function or null mutations (about 10-18% of CF patients) will still develop lung failure. [5, 6] Furthermore, recent studies have demonstrated that abrupt interruption of CFTR modulator therapy may cause severe clinical consequences and patient compliance is still an issue. Therefore, gene therapeutic strategies enabling restoration of sufficient CFTR activity to rehydrate the mucus layer will likely be curative for almost all CF patients. It is estimated that restoration of CFTR function in 6-10% of non-differentiated airway epithelial cells will be effective in correcting the transport defect.[7, 8]

Gene therapy has made great advances in the last decade, and, as CF is a single gene defect, it is an attractive disease target for new therapeutic interventions.[4] Despite the introduction of new CFTR potentiating therapies, CF patients continue to have a greatly reduced lifespan and suffer from significant morbidities, including recurrent lung infections and progressive respiratory failure. While gene therapy strategies have been explored previously for CF there remain several critical barriers that have precluded robust gene correction in airway epithelial tissue.[4] First, the airway mucus and immune response

provides a barrier to standard approaches for stable transgene expression. Moreover, when targeting the airway epithelium, vector must be administered repeatedly because of continuous epithelial cell replacement.[4, 8] Unfortunately, the host adaptive immune response renders this strategy ineffective; however, these limitations may be overcome by permanently correcting a mutated *CFTR* gene in long-lived airway basal stem cells (ABSCs).

CRISPR/Cas9 system can be utilized to overcome the potential limitations of CF gene therapies. The CRISPR/Cas9 nuclease system provides the tools for effective genome modification of the *CFTR* gene. The system is composed of a programmable single-guide RNA (sgRNA) that binds and directs the Cas9 nuclease to a specific location in the genome.[1, 2] The protospacer adjacent motif (PAM) sequence, adjoining the sgRNA DNA binding site, aids in target recognition for the Cas9 nuclease. The Cas9 nuclease induces a blunt, double-stranded break (DSB) three base pairs upstream of the PAM sequence.[1, 2] Gene knockouts can be generated if the host cell utilizes non-homologous end joining (NHEJ) to repair the DSB *via* ligation of the broken ends. Conversely, homology directed repair (HDR) utilizes DNA that contains homology to sequences adjoining the DSB – designated as homology arms (HA) – as a template for repair.[1, 2] This “donor” template can be used to repair the DSB with a DNA sequence of choice. Thus, HDR machinery may be configured to integrate a sequence of interest (*e.g., CFTR*) into a specific site of the genomic locus by utilizing donor DNA that contains both the HAs flanking the target site.

Various groups have demonstrated the use of site-specific complementary DNA (cDNA) insertion to ameliorate disease phenotypes**.** [1, 2] One common approach utilizes the

CRISPR/Cas9 system to correct multiple mutated sites in a gene with the site-specific insertion of the corrected target gene's cDNA. For example, we (*Kuo et al.*) utilized the CRISPR/Cas9 nuclease system to efficiently integrate a normal copy of CD40L cDNA downstream of the endogenous promoter.[1] By utilizing site specific cDNA insertion downstream of the endogenous promoter, the corrected cDNA template can be transcribed from the endogenous promoter to correct all downstream mutations. **Given that there are over 2000 mutations for CF that have been characterized as disease causing, we hypothesize that the site-specific insertion of a normal CFTR gene cassette into the endogenous CFTR locus will correct most CF-causing mutations and achieve appropriate expression to reduce the disease phenotype.**

RESULTS

We designed and analyzed various sgRNAs within the 5' untranslated region (UTR) of the CFTR gene (sgRNA3, 19, and 20) as well as various sgRNAs within intron 1 of the CFTR gene (sgRNA13-18) for site-specific insertion of the *CFTR* gene for regulation by the endogenous *CFTR* promoter. In a pilot study, we utilized T84 epithelial cells, derived from a human colon carcinoma, and 16HBE14o- (HBE) cells, derived from human bronchial epithelial cells, to assess sgRNA cutting efficiency and *CFTR* gene correction. These cell lines, that endogenously expresses CFTR, was selected based on guidance from CFF and serves as a representative model for CFTR gene repair. For these studies, we transfected T84 and HBE cells with a sgRNA-Cas9 ribonucleoprotein (RNP) complex using the Lonza 4D- Nucleofector to improve toxicity and gene insertion frequency. 72 h post-transfection, we harvested cells for analysis of sgRNA cutting efficiency at the *CFTR* locus and any off-target locations. We designed primers to amplify DNA flanking the sgRNA cut site; amplified regions were Sanger sequenced and analyzed for indel formation by using Tracking of Indels by Decomposition (TIDE) analysis. The percentage of on-target indel formation serves as an indication of each sgRNA at directing site-specific DSB formation (**Figure 3.1A and Figure 3.1B**).

To optimize CRISPR/Cas9 targeting of model T84 cells, we measured various forms of reagent delivery, including the Lonza 4D-Nucleofector, the BTX Electro Square Porator ECM 830, and Lipofectamine CRISPR/Max. CRISPR/Cas9 gene-editing reagents targeting the *CFTR* 5'UTR were delivered to T84 cells as described by manufacturer protocol. In short, 120 pmol of Cas9 and 100 pmol of sgRNA were delivered to T84s using DS-138 program of the Lonza electroporator. To increase editing efficiency, we

incorporated Integrated DNA Technologies' (IDT) Alt-R electroporation enhancer (EE) into the transfection mix. In the context of Lonza electroporation, the EE increased editing efficiency greater than two-fold (**Figure 3.2A**). Furthermore, T84 cells were electroporated with 200 pmol of Cas9 and 240 pmol of sgRNA with/without IDT's EE. Finally, 1250 ng of Cas9 and 240ng of sgRNA were transfected into T84s using the Lipofectamine CRISPR/Max reagents. Lonza's electroporation conditions with EE provided with greatest editing efficiency with fewest reagent utilized (**Figure 3.2B**). At 72 h post-transfection, we harvested cells for analysis of sgRNA cutting efficiency at the *CFTR* locus using TIDE analysis.

We also set out to profile the safety of these sgRNAs. We performed a preliminary off-target analysis of these guides in K562 and T84 cells using the Genome-wide, Unbiased Identification of DSBs Enabled by Sequencing (GUIDE-Seq) analysis (**Figure 3.3**). GUIDE-seq detects off-target effects of CRISPR/Cas9 in the genome at the nucleotide level with unbiased amplification and next-generation sequencing (NGS).[13] No off-target sites were determined in sgRNA3 targeting the 5'UTR of the *CFTR* gene (**Figure 3.3A**); however, efforts are currently underway to repeat these studies in T84 and HBE cells. While our analysis revealed off-target cutting within the intron 1 guides (**Figure 3.3B)**, no off-targets measured were within 2 megabases of known oncogenic or tumor-suppressing genes. Efforts are currently underway to repeat these studies in HBE cells.

After generating optimal sgRNAs for *CFTR* gene insertion (**Fig. 3.7**), our next step in knocking-in a therapeutic *CFTR* cDNA cassette is to test efficient site-specific integration using a green fluorescent protein (GFP) reporter construct. The GFP reporter will be designed with homology arms flanking the cut site of the sgRNA in *CFTR* as well as a

strong poly adenylation signal, bovine growth hormone polyadenylation (BGH polyA) signal. Given the large size of the CFTR gene (~4.4 kb coding region), the DNA donor reporter constructs that we design will be delivered in multiple formats, including expression plasmids, linear double-stranded DNAs, and Nanoplasmid™ (Aldeveron and Nature Technologies). We designed a series of homologous donors to perform efficient site-specific integration of a donor cassette just downstream from the endogenous *CFTR* gene promoter. Donors were designed with flanking homology arms to the sgRNA cut site, located at the 5'UTR of *CFTR.* At 24 h post-transfection, we measured nucleofected cells for analysis of relative cell survival by means of MTT assay (**Figure 3.4A**). In short, transfected cells will be incubated with MTT for 4 h at 37°C before solubilization by an organic solvent such as dimethyl sulfoxide (DMSO). Consequently, the solubilized formazan is measured by a spectrophotometer. Relative cell survival is calculated as the difference of background absorbance at 630 nm to the signal absorbance measured at 570 nm. Furthermore, we quantified site-specific reporter cassette integration *via* in-out digital-droplet (dd) PCR, using Taqman probes designed to the GFP amplicon cassette. After testing each donor type, the dsDNA donor faired the best in both integration and toxicity when compared to plasmid and nanoplasmid donors (**Figure 3.4B**). These efficiencies of gene insertion are relatively low for clinical efficacy (3-7% observed vs. 6-10% CFTR for clinical benefit)[7, 8] and efforts will be made to improve the gene insertion frequencies in this cell line before studying integration efficiencies in ABSC cultures.

Due to the proximity of our donor cassette to the endogenous *CFTR* promoter, we investigated DNA donor template modifications to remove excess promoter sequence. Specifically, we investigated shortened donor homology arms (HAs) from 500 bp to 50 bp

to both remove any promoter sequence within our donor DNA and minimize template size. By capping the 50 bp HA dsDNA donor with a 5' end chemical modification (*i.e.,* an amine group with a C6 linker, termed AmC6), we observed no reduction in toxicity (**Figure 3.5A**) and maintained equivalent integration efficiency compared to a model plasmid donor (**Figure 3.5B**).[12] Furthermore, AmC6 modified 500 bp HA dsDNA donors displayed a two-fold increase in integration efficiency over plasmid donors. The absolute quantity of donor cassette knock-in was measured *via* in-out digital-droplet (dd)PCR against the inserted reporter cassette.

Researchers from Stanford and the ICR London have demonstrated that two small molecule inhibitors, AZD7648 and ART558, boost HDR-mediated editing by inhibiting NHEJ and MMEJ, respectively.[14,15] As such, we tested our AmC6 modified 50bp HA dsDNA donors for their ability to integrate into the 5'UTR *CFTR* locus in the presence or absence of AZD7648 and ART558 (**Figure 3.6**). HBE cells were transfected with 120pmol of sgRNA, 100pmol of Cas9, and 1ug of dsDNA donor. While the inclusion of small molecule inhibitors AZD7648 and ART558 did not drastically affect relative cell survival as measured by MTT (**Figure 3.6A**), the two inhibitors increased the rate of donor integration by nearly 2-fold over no-drug conditions.

We then designed and tested a series of homologous donors to perform efficient site-specific integration of a codon optimized *CFTR* cDNA into the 5' UTR of the *CFTR* gene. Efficiency of insertion and the level of expression of *CFTR* will be assessed to define optimal sgRNA and homologous donor cassette designs. We codon optimized the *CFTR* cDNA cassette using JCAT, GeneArt, or IDT codon optimizations and compare each cassette to the expression levels of wildtype *CFTR* cDNA.

Using a Lonza 4D- Nucleofector, we will transfected the *CFTR* cDNA donor and RNP complexes into the *CFTR* deficient HBE cell line provided to us by the Cystic Fibrosis Foundation (**Figure 3.7**)**.** At 24 h post transfection, relative cell survival was measured using an MTT assay (**Figure 3.7A**). Furthermore, 72 h post transfection, relative sgRNA cutting efficiency was confirmed in edited cells using TIDE analysis. We also measured *CFTR* gene insertion by in-out ddPCR 14 days post transfection (**Figure 3.7B**). We are currently quantifying CFTR protein expression, by lysing and immunoblotting transfected cells with the monoclonal antibody M3A7 to compare their expression to that of the level expressed by the parental, non-disrupted cells. Furthermore, mRNA specific primers will be designed to quantify the levels of transgene CFTR mRNA expression *via* RT-qPCR. Low levels of integration were observed, ranging from 1-4% integration of CFTR cDNA cassettes.

DISCUSSION

Previous groups have described that nearly 6-10% of corrected CFTR is sufficient to ameliorate the CF phenotype.[19,20] However, our efforts have demonstrated low levels of *CFTR* cDNA insertion, ranging from 1-4%. While we are currently undergoing functional studies to determine whether the codon optimized cDNA is sufficient to overcome decreased expression resulting from minimal integration, further studies are being conducted to increase expression potential. For example, we are currently assessing the inclusion of the hyperactive W493R ENaC variant and the gain of function K978C variant into our *CFTR* cDNA cassettes.[16,17,18] W493R has previously been demonstrated to double the sodium absorption through the amiloride sensitive epithelial sodium channel (ENaC), a key defect in CF patients. The K978C mutation has elicited greater than 10-fold more CFTR protein and displayed ~4-fold greater activity than wild-type (WT) CFTR. The inclusion of W493R and K978C mutations may be able to increase and restore CFTR function in cases with lower integration rates of our donor cassette.

More importantly, while therapeutic strategies that enable permanent correction of the underlying genetic mutation would be transformative for CF patient, previous efforts to establish gene therapeutic strategies for CF have been met by significant challenges. Targeting airway basal stem cell (ABSC) populations for long-lived gene correction in the CF lung has been especially difficult because of their protected location. These multipotent stem cells are capable of self-renewal and differentiation into the mucociliary cell types of the cartilaginous airway epithelium. However, successful transduction of these stem cells will require delivery of nucleic acid-based gene-editing cargoes to the deeper regions of the airway epithelium. Thus, vehicles for nucleic acid-based treatments

will need to be designed and packaged for safe, targeted delivery to, and subsequent gene editing of, these stem cell populations.

As such, our current studies are also incorporating an integrated, multi-investigator collaborative research program to overcome these barriers of gene delivery to the airway epithelium and to develop broadly applicable methods and technologies for gene delivery and correction in ABSCs to lead to a durable and curative gene therapy approach for all CF patients.

Our overarching goal is to develop a gene therapeutic strategy that results in effective delivery to long-lived ABSCs with successful correction of >99% of all CFTR mutations. Novel reporters and CRISPR/Cas9-based gene-editing tools for *CFTR* correction will be generated and packaged into nanoparticle-based carriers for delivery to ABSCs. The efficacy and safety of the proposed nanocarriers will be evaluated in human *in vitro* and *ex vivo* airway models that use CF samples to recapitulate the excessive mucus in CF. To further enable this work, we will design and apply new research tools to distribute aerosol suspensions of nanocarriers to *in vitro* air liquid interface (ALI) and *ex vivo* airway explant cultures so that changes in nanocarrier design and packaging can be evaluated systematically. We will evaluate strategies that incorporate surfactant-functionalized nanoparticles, which break up mucus and disrupt cell-cell junctions between differentiated apical cells, to expose the ABSCs and promote delivery of gene-editing constructs while minimizing airway injury.

The strength of this interdisciplinary scope allows the expertise of three separate UCLA groups, including that of Dr. Donald Kohn, to offer new delivery methods (such as the design of nanoparticle delivery vehicles) to integrate *CFTR* cDNA into the endogenous

CFTR locus. The utilization of microfluidic technologies would not only leverage new and improved delivery systems into CF model lines and stem cells but also would integrate surfactant functionalities to circumvent the thick mucus barrier in CF lungs and the tight junctions of the epithelium.

FIGURES

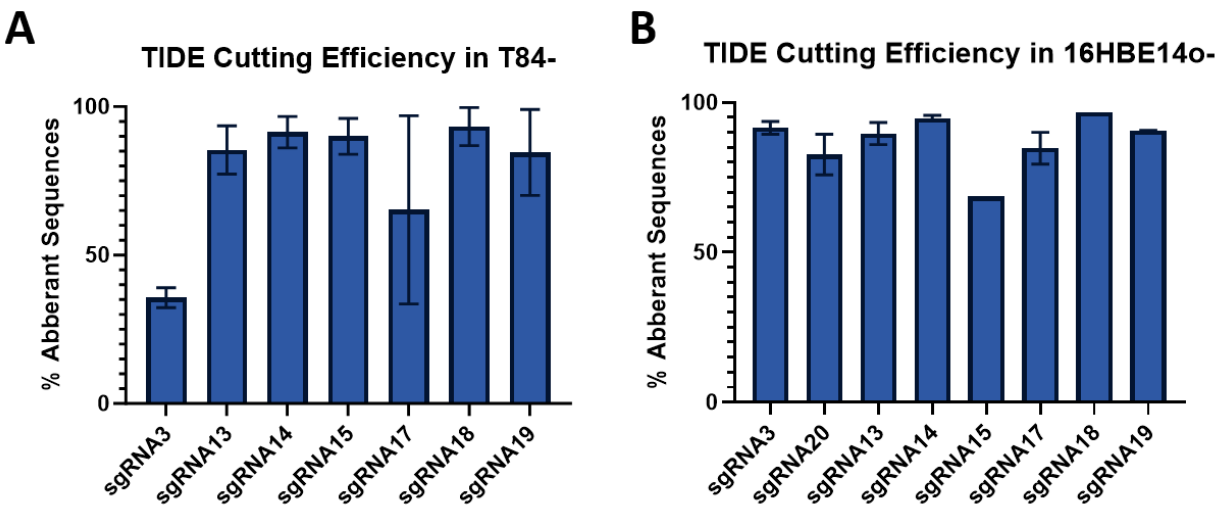


Figure 3. 1

Assessment of sgRNA cutting efficiency of sgRNAs targeting the *CFTR* 5'UTR (sgRNA 3, 19, and 20) and sgRNAs targeting the *CFTR* intron 1 (sgRNAs 13-18) in (A) T84 cells and (B) 16HBE14o- cells. Cutting Efficiency is measured by TIDE sequencing.

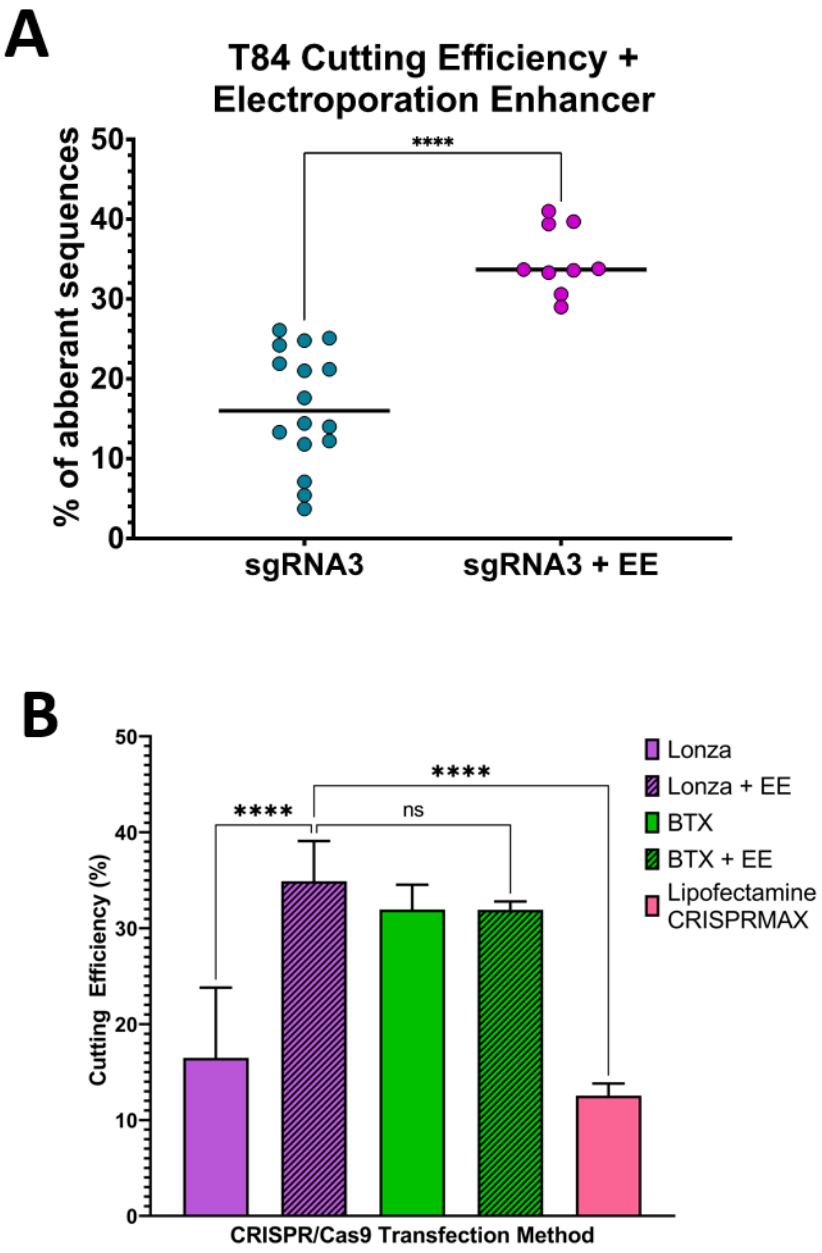


Figure 3. 2

(A) T84 cells were electoporated with Cas9 RNPs with or without IDT's Alt-R electroporation enhancer (EE) to measure potential effects on cutting efficiency as measured by TIDE **(B)** Lonza, BTX, and Lipofectamine transfection methods were compared in T84 cells targeting the CFTR 5'UTR.

A

B

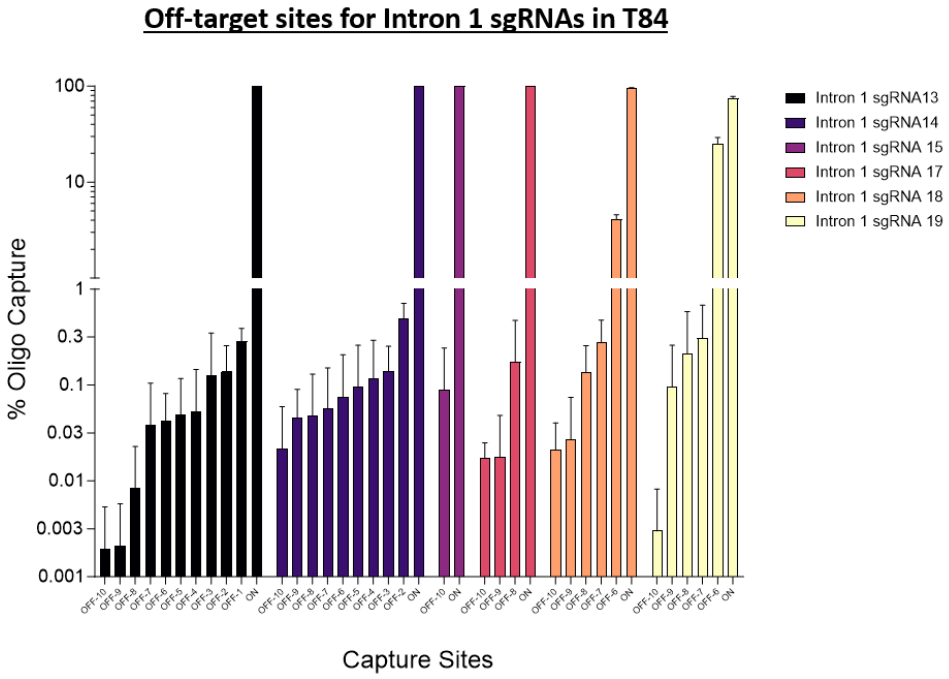


Figure 3. 3

We investigated the off-target cutting efficiency using Guide-Seq of (A) sgRNAs targeting the *CFTR* 5'UTR in K562 cells and of (B) sgRNAs targeting the *CFTR* intron 1 in T84 cells.

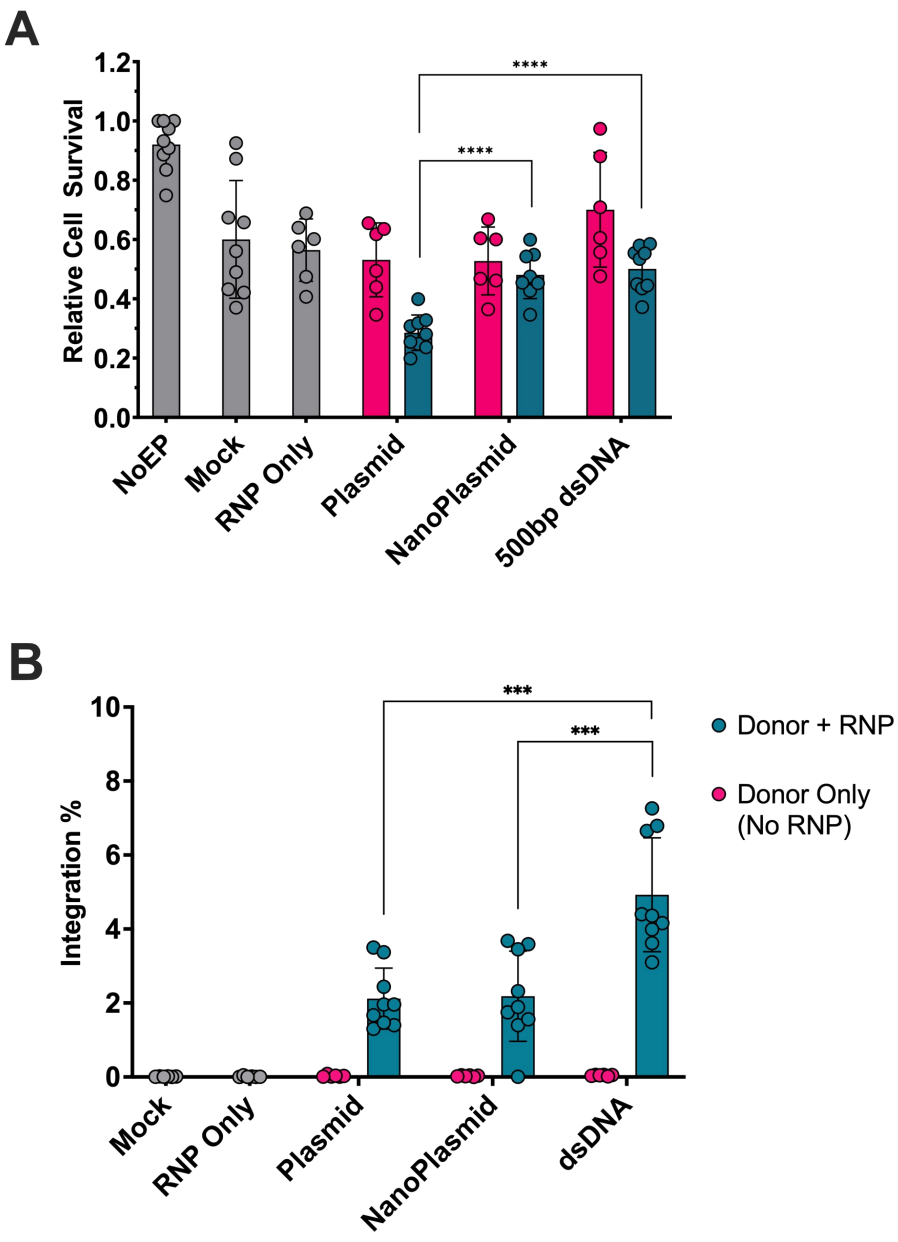


Figure 3. 4

T84s were electroporated with 820 pmol of a conventional 500 bp homology arm (HA) plasmid, 500 bp HA Nanoplasmid™, or a 500 bp HA linear dsDNA all targeting the *CFTR* 5'UTR. Relative toxicity was measured *via* MTT assay 24 h post transfection (**A**) and integration was measured 14 days post transfection *via* in-out digital droplet (dd)PCR (**B**).

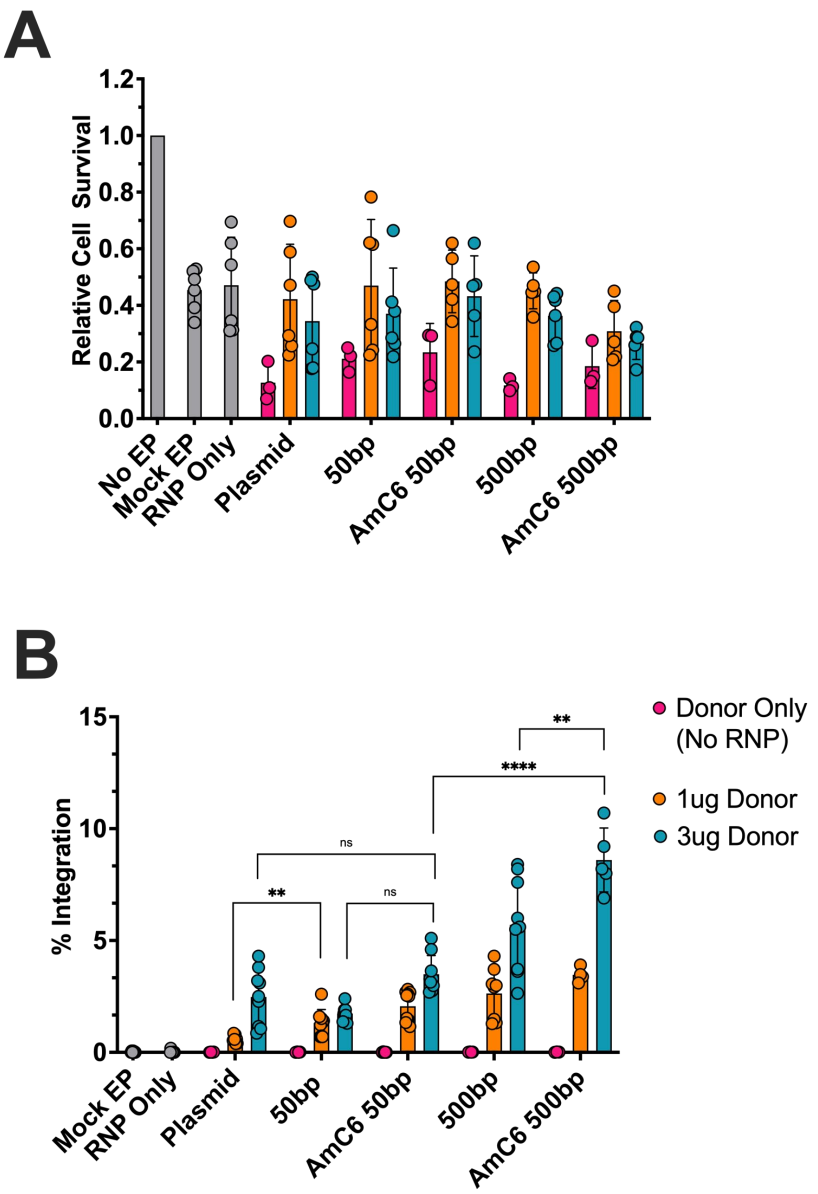


Figure 3. 5

We investigated methods to increase HDR efficiency on donor templates with short HAs to remove promoter sequence in our donor template and minimize potential off-target effects. T84s were electroporated with both reagents targeting the *CFTR* 5'UTR and 1 μg or 3 μg of donor DNA. 50 bp dsDNA donors were capped with an AmC6 5’ end chemical modification and tested for toxicity by MTT (**A**) and integration by ddPCR (**B**).

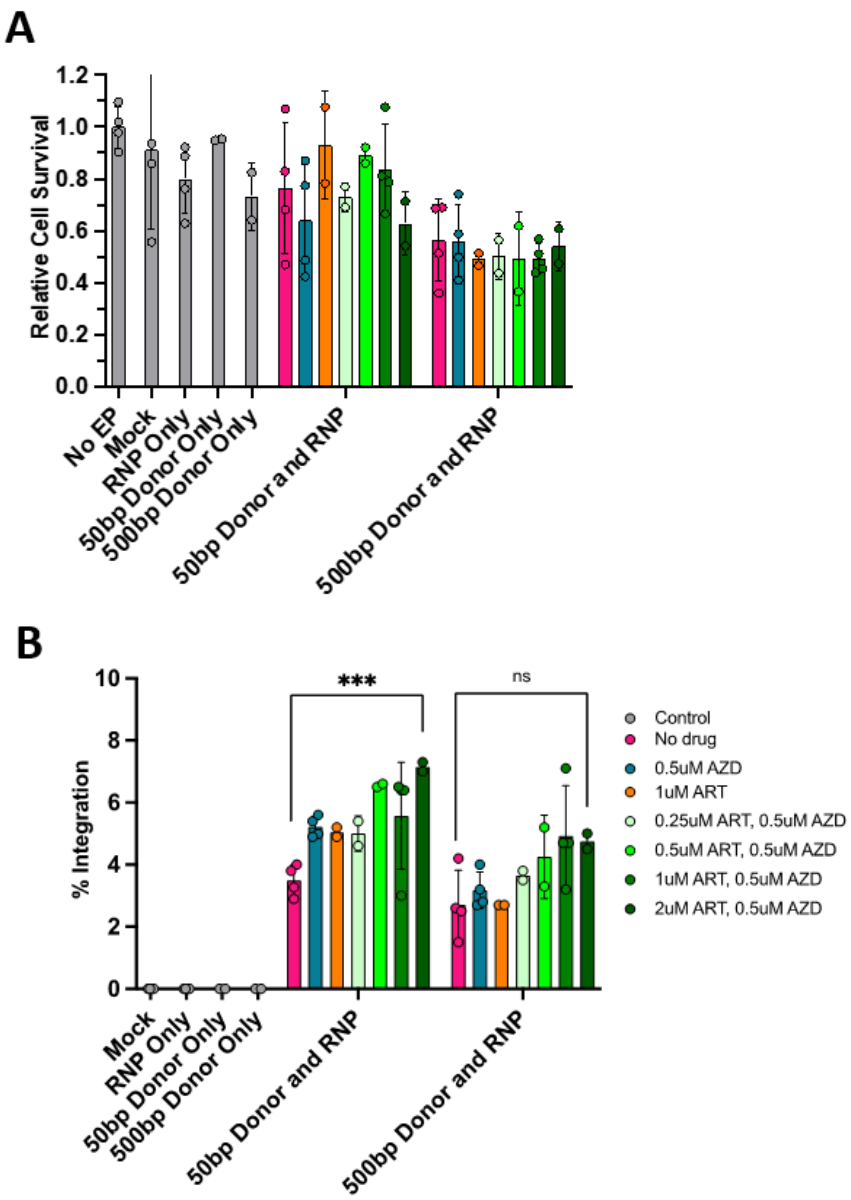


Figure 3. 6

16HBEo14- cells were electroporated with 1ug of a AmC6 modified 50 bp homology arm (HA) linear dsDNA targeting the *CFTR* 5'UTR. Relative toxicity was measured *via* MTT assay 24 h post transfection (**A**) and integration was measured 14 days post transfection *via* in-out digital droplet (dd)PCR (**B**). Various concentrations of a DNA PK inhibitor AZD 7648 (AZD) and an Pol0 inhibitor ART558 (ART) were used to measure effect on integration frequency.

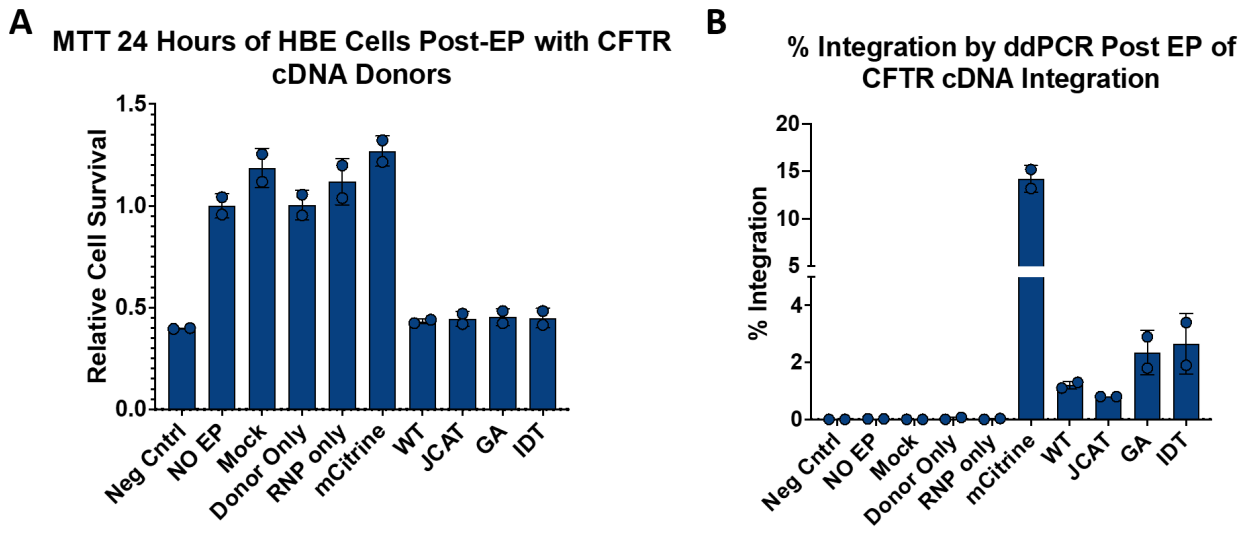


Figure 3. 7

16HBEo14- cells were electroporated with 1ug of a AmC6 modified 50 bp homology arm (HA) linear dsDNA targeting the *CFTR* 5'UTR in the presence of 0.5uM AZD7648 and 2uM of ART558. Various CFTR cDNA codon optiomized donors were tested, including wildtype sequence (WT), JCAT, GA, and IDT codon optimized variants. Relative toxicity was measured *via* MTT assay 24 h post transfection (**A**) and integration was measured 14 days post transfection *via* in-out digital droplet (dd)PCR (**B**).

REFERENCES

1. Kuo, C.Y., et al., *Site-Specific Gene Editing of Human Hematopoietic Stem Cells for X-Linked Hyper-IgM Syndrome.* Cell Rep, 2018. **23**(9): p. 2606-2616.
2. Gray, D.H., et al., *Optimizing Integration and Expression of Transgenic Bruton's Tyrosine Kinase for CRISPR-Cas9-Mediated Gene Editing of X-Linked Agammaglobulinemia.* CRISPR J, 2021. **4**(2): p. 191-206.
3. Hoban, M.D., et al., *Correction of the sickle cell disease mutation in human hematopoietic stem/progenitor cells.* Blood, 2015. **125**(17): p. 2597-604.
4. Cooney, A.L., P.B. McCray, Jr., and P.L. Sinn, *Cystic Fibrosis Gene Therapy: Looking Back, Looking Forward.* Genes (Basel), 2018. **9**(11).
5. Hegab, A.E., et al., *Isolation of basal cells and submucosal gland duct cells from mouse trachea.* J Vis Exp, 2012(67): p. e3731.
6. Hegab, A.E., et al., *Isolation and in vitro characterization of basal and submucosal gland duct stem/progenitor cells from human proximal airways.* Stem Cells Transl Med, 2012. **1**(10): p. 719-24.
7. Chadeuf, G., et al., *Evidence for encapsidation of prokaryotic sequences during recombinant adeno-associated virus production and their in vivo persistence after vector delivery.* Mol Ther, 2005. **12**(4): p. 744-53.
8. Johnson, L.G., et al., *Efficiency of gene transfer for restoration of normal airway epithelial function in cystic fibrosis.* Nature Genetics, 1992. **2**.
9. Long, J., et al., *Characterization of Gene Alterations following Editing of the beta-Globin Gene Locus in Hematopoietic Stem/Progenitor Cells.* Mol Ther, 2018. **26**(2): p. 468-479.

10. Romero, Z., et al., *Editing the Sickle Cell Disease Mutation in Human Hematopoietic Stem Cells: Comparison of Endonucleases and Homologous Donor Templates.* Mol Ther, 2019. **27**(8): p. 1389-1406.
11. Joglekar, A.V., et al., *Integrase-defective lentiviral vectors as a delivery platform for targeted modification of adenosine deaminase locus.* Mol Ther, 2013. **21**(9): p. 1705-17.
12. Nguyen, D.N., et al., *Polymer-stabilized Cas9 nanoparticles and modified repair templates increase genome editing efficiency.* Nat Biotechnol, 2020. **38**(1): p. 44-49.
13. Tsai, S.Q., et al., *GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases.* Nat Biotechnol, 2015. **33**(2): p. 187-197.
14. Fok, J.H.L., Ramos-Montoya, A., Vazquez-Chantada, M. et al. AZD7648 is a potent and selective DNA-PK inhibitor that enhances radiation, chemotherapy and olaparib activity. Nat Commun 10, 5065 (2019).
15. Zatreanu, D., Robinson, H.M.R., Alkhatib, O. *et al.* Polθ inhibitors elicit *BRCA*-gene synthetic lethality and target PARP inhibitor resistance. *Nat Commun* **12**, 3636 (2021).
16. Woodall M, Tarran R, Lee R, Anfishi H, Prins S, Counsell J, Vergani P, Hart S, Baines D. Expression of gain-of-function CFTR in cystic fibrosis airway cells restores epithelial function better than wild-type or codon-optimized CFTR. *Mol Ther Methods Clin Dev*. 2023 Aug 12;30:593-605. doi: 10.1016/j.omtm.2023.08.006. PMID: 37701179; PMCID: PMC10494266.

17. Handschick M, Hedtfeld S, Tümmler B. Frequency of the hyperactive W493R ENaC variant in carriers of a CFTR mutation. *J Cyst Fibros*. 2012 Jan;11(1):53-5. doi: 10.1016/j.jcf.2011.08.008. Epub 2011 Sep 13. PMID: 21917531.
18. Shobair M, Dagliyan O, Kota P, Dang YL, He H, Stutts MJ, Dokholyan NV. Gain-of-Function Mutation W493R in the Epithelial Sodium Channel Allosterically Reconfigures Intersubunit Coupling. J Biol Chem. 2016 Feb 19;291(8):3682-92. doi: 10.1074/jbc.M115.678052. Epub 2015 Dec 14. PMID: 26668308; PMCID: PMC4759151.
19. Johnson, Larry G., *et al*. “Efficiency of Gene Transfer for Restoration of Normal Airway Epithelial Function in Cystic Fibrosis.” *Nature Genetics*, vol. 2, no. 1, 1992, pp. 21–25, doi:10.1038/ng0992-21.
20. Chadeuf, Gilliane, *et al*. “Evidence for Encapsidation of Prokaryotic Sequences during Recombinant Adeno-Associated Virus Production and Their in Vivo Persistence after Vector Delivery.” *Molecular Therapy*, vol. 12, no. 4, The American Society of Gene Therapy, 2005, pp. 744–53.

CONCLUSIONS AND FUTURE DIRECTIONS

The work described in this thesis delineates various tools and techniques that are a result of advances in gene therapy: pseudotyping to enable vaccine and drug design, the identification and refinement of enhancer sequences to design lentiviral vectors that enable precise regulation of target genes, and the optimization of large gene editing cargoes to facilitate site-specific integration of difficult templates. The future of each of these projects lies in paving a path towards the clinic for the treatment of their respective genetic diseases. But in parallel to these studies, a multitude of new gene therapy advancements have begun changing how gene therapies are designed and delivered.

The field of gene therapy is still quickly evolving. The most common approach relies on autologous transplantation, reprogramming patient blood cells (hematopoietic or primary) and reinjecting them back into a patient's body to give rise to healthy, functioning cells. This can be through the process of viral delivery (such as lentiviruses or adeno-associated viruses) or non-viral approaches (e.g. chemical or physical delivery). In the recent decade, the field has begun advancing from blood cells to other forms of gene delivery, including eye, liver, and heart.[10-13]

These advancements have been accelerated by technological improvements in gene therapy delivery and repair, including the use of nanoparticles/nanocarriers or virus-like-particles as delivery vehicles and technological gene therapy advancements such as prime editing and base editing.[10-13] Nanoparticles and virus-like particles are non-infectious particles that are synthesized to assemble gene therapy reagents for delivery into specific cell types.[14] Base editing is a type of gene editing that leverages CRISPR technology and a deaminase for single nucleotide conversions, without the creation of a

double stranded break. And finally, prime editing, another form of gene editing, CRISPR technology and reverse transcriptase to insert or delete larger fragments of DNA without the need for DSBs.

Technological advancements in the realm of gene therapy are not in short supply; continuous improvements, new and emerging therapies, and streamlined approaches for design and development have made the field of gene therapy an attractive avenue to research. **But problems still hinder these gene therapy products from progressing to clinic or commercialization**: nanoparticles have difficulty with off-target effects in the liver; virus-like-particles (while in their infancy) struggle with efficiency to levels of current electroporation-based or lentiviral vector-based delivery; base editors are limited to single nucleotide changes (and currently, transitions but not transversions), preventing the generation of suitable therapies for a variety of disorders targetable by HDR based gene-editing strategies or lentiviral vectors; prime editors, a newer technology, shows promise, but it currently lacks the efficiency necessary for therapeutic benefit of hematopoietic disorders.[10-13,15]

While the pace of research is flourishing, the field is still riddled with hurdles in manufacturing, funding, and regulation.[15-17] Finding therapies for extremely rare diseases has proven difficult because the existing market conditions create obstacles for manufacturers to commercialize these products. Advancements in this area will necessitate innovative incentives and collaborative strategies to encourage the investment necessary for drug development for rare and ultra-rare diseases. These may be solved by streamlining the development process – for example using *in vivo* delivery – to decrease costs of manufacturing, funding the development of patient registries,

centralize reimbursement models, implement policies to accelerate time-to-clinic/commercialization for rare and ultra-rare diseases.[18] The NIH and FDA's Bespoke Gene Therapy Consortium is one such evolution necessary to overcome this hurdle, paving the way for a potential platform regulatory review (i.e. same base editor, different gRNA) that relies on previous knowledge to streamline development and regulation of rare therapies.[19] Further, Operation Warp Speed, which began to combat the COVID-19 epidemic, has been since expanded for rare diseases to accelerate their time to market without compromising the products safety and effectiveness.[20] Implementing these changes will bring about their own risks and rewards, but continued product innovation will only be a sustainable if the costs, policies, and management of these innovations are also evolving in parallel.

BIBLIOGRAPHY

1. Masiuk KE, Laborada J, Roncarolo MG, Hollis RP, Kohn DB. Lentiviral Gene Therapy in HSCs Restores Lineage-Specific Foxp3 Expression and Suppresses Autoimmunity in a Mouse Model of IPEX Syndrome. Cell Stem Cell. 2019 Feb 7;24(2):309-317.e7. doi: 10.1016/j.stem.2018.12.003. Epub 2019 Jan 10. PMID: 30639036.
2. Wong, R. L. *et al.* Lentiviral gene therapy for X-linked chronic granulomatous disease recapitulates endogenous CYBB regulation and expression. *Blood* **141**, 1007-1022, doi:10.1182/blood.2022016074 (2023).
3. Morgan, R. A. *et al.* Improved Titer and Gene Transfer by Lentiviral Vectors Using Novel, Small β-Globin Locus Control Region Elements. *Molecular Therapy* **28**, 328-340, doi:https://doi.org/10.1016/j.ymthe.2019.09.020 (2020).
4. Kohn DB, Booth C, Shaw KL, Xu-Bayford J, Garabedian E, Trevisan V, Carbonaro-Sarracino DA, Soni K, Terrazas D, Snell K, Ikeda A, Leon-Rico D, Moore TB, Buckland KF, Shah AJ, Gilmour KC, De Oliveira S, Rivat C, Crooks GM, Izotova N, Tse J, Adams S, Shupien S, Ricketts H, Davila A, Uzowuru C, Icreverzi A, Barman P, Campo Fernandez B, Hollis RP, Coronel M, Yu A, Chun KM, Casas CE, Zhang R, Arduini S, Lynn F, Kudari M, Spezzi A, Zahn M, Heimke R, Labik I, Parrott R, Buckley RH, Reeves L, Cornetta K, Sokolic R, Hershfield M, Schmidt M, Candotti F, Malech HL, Thrasher AJ, Gaspar HB. Autologous Ex Vivo Lentiviral Gene Therapy for Adenosine Deaminase Deficiency. N Engl J Med. 2021 May 27;384(21):2002-2013. doi: 10.1056/NEJMoa2027675. Epub 2021 May 11. PMID: 33974366; PMCID: PMC8240285.

5. Gray DH *et al.* Optimizing Integration and Expression of Transgenic Bruton's Tyrosine Kinase for CRISPR-Cas9-Mediated Gene Editing of X-Linked Agammaglobulinemia. CRISPR J. 2021 Apr;4(2):191-206.
6. Kuo, Caroline Y., *et al*. "Site-Specific Gene Editing of Human Hematopoietic Stem Cells for X-Linked Hyper-IgM Syndrome." *Cell Reports*, vol. 23, no. 9, ElsevierCompany., 2018, pp. 2606–16.
7. Long, Joseph, *et al*. "Characterization of Gene Alterations Following Editing of the β-Globin Gene Locus in Hematopoietic Stem Cells." *Molecular Therapy*, vol. 26, no. 2, Feb. 2018, pp. 468–79.
8. Boelens JJ, Aldenhoven M, Purtill D, Ruggeri A, DeFor T, Wynn R, Wraith E, Cavazzana-Calvo M, Rovelli A, Fischer A, Tolar J, Prasad VK, Escolar M, Gluckman E, O'Meara A, Orchard PJ, Veys P, Eapen M, Kurtzberg J, Rocha V on behalf of all participating centers from Eurocord, Inborn Errors Working Party of European Blood and Marrow Transplant group, Duke University, and the Centre for International Blood and Marrow Research. Outcomes of transplantation using various hematopoietic cell sources in children with Hurler syndrome after myeloablative conditioning. *Blood.* 2013;121:3981–3987. doi: 10.1182/blood-2012-09-455238.
9. Walters MC. Update of hematopoietic cell transplantation for sickle cell disease. *Current Opinion in Hematology.* 2015;22:227–233. doi: 10.1097/MOH.0000000000000136.
10. Lu X, Zhang M, Li G, Zhang S, Zhang J, Fu X, Sun F. Applications and Research Advances in the Delivery of CRISPR/Cas9 Systems for the Treatment of

Inherited Diseases. Int J Mol Sci. 2023 Aug 25;24(17):13202. doi: 10.3390/ijms241713202. PMID: 37686009; PMCID: PMC10487642.

11. Cantore A, Fraldi A, Meneghini V and Gritti A (2022) In vivo Gene Therapy to the Liver and Nervous System: Promises and Challenges. Front. Med. 8:774618. doi: 10.3389/fmed.2021.774618
12. Mendell JR, Al-Zaidy SA, Rodino-Klapac LR, Goodspeed K, Gray SJ, Kay CN, Boye SL, Boye SE, George LA, Salabarria S, Corti M, Byrne BJ, Tremblay JP. Current Clinical Applications of In Vivo Gene Therapy with AAVs. Mol Ther. 2021 Feb 3;29(2):464-488. doi: 10.1016/j.ymthe.2020.12.007. Epub 2020 Dec 10. PMID: 33309881; PMCID: PMC7854298.
13. Chavez, M., Chen, X., Finn, P.B. et al. Advances in CRISPR therapeutics. Nat Rev Nephrol 19, 9–22 (2023). https://doi.org/10.1038/s41581-022-00636-2
14. Nooraei, S., Bahrulolum, H., Hoseini, Z.S. et al. Virus-like particles: preparation, immunogenicity and their roles as nanovaccines and drug nanocarriers. J Nanobiotechnol 19, 59 (2021). https://doi.org/10.1186/s12951-021-00806-7
15. Anzalone AV, Randolph PB, Davis JR, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature. 2019;576(7785):149-157.
16. van Overbeeke E, Michelsen S, Toumi M, Stevens H, Trusheim M, Huys I, Simoens S. Market access of gene therapies across Europe, USA, and Canada: challenges, trends, and solutions. Drug Discov Today. 2021 Feb;26(2):399-415. doi: 10.1016/j.drudis.2020.11.024. Epub 2020 Nov 24. PMID: 33242695.
17. Cornetta K, Bonamino M, Mahlangu J, Mingozzi F, Rangarajan S, Rao J. Gene therapy access: Global challenges, opportunities, and views from Brazil, South

Africa, and India. Mol Ther. 2022 Jun 1;30(6):2122-2129. doi: 10.1016/j.ymthe.2022.04.002. Epub 2022 Apr 4. PMID: 35390542; PMCID: PMC9171243.

18. Pearson, C., Schapiro, L., & Pearson, S. D. (2022). The next generation of rare disease drug policy: ensuring both innovation and affordability. In Journal of Comparative Effectiveness Research (Vol. 11, Issue 14, pp. 999–1010). Becaris Publishing Limited. https://doi.org/10.2217/cer-2022-0120
19. Bueren JA, Auricchio A. Advances and Challenges in the Development of Gene Therapy Medicinal Products for Rare Diseases. Hum Gene Ther. 2023 Sep;34(17-18):763-775. doi: 10.1089/hum.2023.152. PMID: 37694572.
20. U.S. Department of Health and Human Services. (2023, June 13). *Bespoke gene therapy consortium*. National Institutes of Health. https://www.nih.gov/research-training/accelerating-medicines-partnership-amp/bespoke-gene-therapy-consortium