

**UCLA**

**UCLA Electronic Theses and Dissertations**

**Title**

Advancements in Viral Vector-Mediated Gene Therapy: Development and Assessment of Lentiviral and  $\gamma$ -Retroviral Approaches for Genetic Hematologic Disorders

**Permalink**

<https://escholarship.org/uc/item/3mc766sp>

**Author**

Hart, Kevyn Lopez

**Publication Date**

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Advancements in Viral Vector-Mediated Gene Therapy: Development and Assessment  
of Lentiviral and  $\gamma$ -Retroviral Approaches for Genetic Hematologic Disorders

A dissertation submitted in partial satisfaction  
of the requirements for the degree Doctor of Philosophy  
in Human Genetics

by

Kevyn Lopez Hart

2024

© Copyright by

Kevyn Lopez Hart

2024

## ABSTRACT OF THE DISSERTATION

Advancements in Viral Vector-Mediated Gene Therapy: Development and Assessment  
of Lentiviral and  $\gamma$ -Retroviral Approaches for Genetic Hematologic Disorders

by

Kevyn Lopez Hart

Doctor of Philosophy in Human Genetics

University of California, Los Angeles, 2024

Professor Donald Barry Kohn, Co-Chair

Professor Aldons Jake Lusic, Co-Chair

This work encompasses the development of lentiviral technology, and the assessment of viral vector-mediated gene therapy approaches for genetic disorders. Chapter 1 illustrates the development of a high-titer bifunctional lentiviral vector in a vector backbone that has reduced size, high vector yields, and efficient gene transfer to human CD34<sup>+</sup> hematopoietic stem and progenitor cells for sickle cell disease (SCD). This lentiviral vector induces high levels of anti-sickling hemoglobins, while concurrently reducing sickle hemoglobin in transduced SCD patient CD34<sup>+</sup> cells differentiated into erythrocytes and in the SCD Berkeley mouse model *in vivo*. Chapter 2 highlights DNA sequencing of peripheral blood cells from Adenosine Deaminase Severe Combined Immunodeficiency (ADA-SCID) patients treated with autologous CD34<sup>+</sup> cells

transduced with either a  $\gamma$ -retroviral or lentiviral *ADA* gene vector to assess transgene mutational profiles. We observed reverse transcriptase mutations as well as APOBEC3 guanine (G) to adenosine (A) mutational signatures.

The dissertation of Kevyn Lopez Hart is approved.

Gay Crooks

Yvonne Chen

Donald Barry Kohn, Committee Co-Chair

Aldons Jake Lulis, Committee Co-Chair

University of California, Los Angeles

2024

## DEDICATION

This work is dedicated to my family who have always supported my passion for science from a young age. This work reflects the time, dedication, love, and encouragement you have given me throughout my life. I am truly grateful.

# TABLE OF CONTENTS

## Contents

UNIVERSITY OF CALIFORNIA.....	i
ABSTRACT OF THE DISSERTATION.....	ii
COMMITTEE.....	iv
DEDICATION.....	v
TABLE OF CONTENTS.....	vi
ACKNOWLEDGMENTS.....	ix
VITA.....	xii
INTRODUCTION.....	1
Chapter 1: A Novel, High-Titer, Bifunctional Vector for Autologous Hematopoietic Stem Cell Gene Therapy of Sickle Cell Disease.....	5
ABSTRACT.....	5
INTRODUCTION.....	6
RESULTS.....	11
DISCUSSION.....	19
MATERIALS AND METHODS.....	25
FIGURES.....	32
REFERENCES.....	49
Chapter 2: Quantifying the Mutational Landscape of Retroviral and Lentiviral Vectors in Gene Therapy Patients.....	56
ABSTRACT.....	56
INTRODUCTION.....	57
RESULTS.....	60
DISCUSSION.....	68
MATERIALS AND METHODS.....	73
FIGURES.....	77
TABLES.....	94
REFERENCES.....	109
CONCLUSIONS AND FUTURE DIRECTIONS.....	112

BIBLIOGRAPHY..... 114

List of Figures

**Chapter 1**

Figure 1.1..... 32  
Figure 1.2..... 34  
Figure 1.3..... 36  
Figure 1.4..... 38  
Figure 1.5..... 39  
Figure 1.6..... 40  
Figure 1.7..... 42  
Figure 1.8..... 43  
Figure 1.9..... 44  
Figure 1.10..... 45  
Figure 1.11..... 47  
Figure 1.12..... 48

**Chapter 2**

Figure 2.1..... 77  
Figure 2.2..... 79  
Figure 2.3..... 80  
Figure 2.4..... 81  
Figure 2.5..... 83  
Figure 2.6..... 84  
Figure 2.7..... 85  
Figure 2.8..... 86  
Figure 2.9..... 87  
Figure 2.10..... 88  
Figure 2.11..... 89  
Figure 2.12..... 90  
Figure 2.13..... 91

Figure 2.14.....	92
Figure 2.15.....	93

## List of Tables

### **Chapter 2**

Table 2.1.....	94
Table 2.2.....	95
Table 2.3.....	96
Table 2.4.....	97
Table 2.5.....	98
Table 2.6.....	99
Table 2.7.....	101
Table 2.8.....	102
Table 2.9.....	103
Table 2.10.....	104
Table 2.11.....	105
Table 2.12.....	107
Table 2.13.....	108

## ACKNOWLEDGMENTS

I am so thankful to have gotten the opportunity to be mentored by you, Dr. Donald Kohn, for my graduate studies. You have truly been an inspiration to me being involved in so many projects, collaborations, clinical trials, and yet still found so much time to make sure I felt supported in my projects and career. I have learned so much from you over the past 5 years, and I have enjoyed watching you touch the lives of so many people with the work that has come from your lab. You are a hero to so many, including myself, and I feel very prepared for the next step in my scientific career thanks to your guidance and support.

To Dr. Roger Hollis, I wanted to thank you for not only your mentorship but for our amazing friendship. You've not only provided invaluable scientific support for my projects, but I also want to thank you for all the life advice you've shared through the ups and downs of graduate school. I am going to miss our caffeine runs, our daily lunches, and your silly humor. You made every day in the lab so much fun, and I am excited for the lifelong friendship we have developed.

Thank you to my committee, Dr. Jake Lusic, Dr. Yvonne Chen, and Dr. Gay Crooks. Thank you for your direction on my projects and the questions that kept me on toes. To have such kind, brilliant, and inspiring individuals on my committee made such an impact on my growth as a scientist.

Thank you to everyone in the Kohn Lab. It is evident that the Kohn lab's distinction goes far beyond its scientific achievements, it is the people who make this lab extraordinary. I have had the privilege of working alongside some of the brightest,

most creative, and fun individuals. You all have made coming to lab everyday such an enjoyable experience, I felt like I truly found my home in graduate school. I will miss the daily laughter shared, but I am excited to see all your successes in the future. Thank you all for the never-ending support.

To my friends, you all have provided me so much love and excitement throughout my graduate studies. Thank you all for always having my back.

To my family, you have been such a source of joy throughout this PhD journey. I could not have done this without you and the support you have given me in my career and life. I have been so thankful to have been able to live so close to you all throughout my PhD and spend so much time together as a family. Thank you for always believing in me.

**Chapter 1:** Chapter 1 is a version of: Hart, K. L., Liu, B., Brown, D., Campo-Fernandez, B., Tam, K., Orr, K., Hollis, R. P., Brendel, C., Williams, D. A., Kohn, D. B. A novel high-titer, bifunctional lentiviral vector for autologous hematopoietic stem cell gene therapy of sickle cell disease. *Molecular Therapy methods & Clinical Development* 32, 2 (2024). This work was supported by the Bill and Melinda Gates Foundation (INV-050202). Training grants provided support to Kevyn Hart (NIH TL1 DK132768 and U2C DK129496).

**Chapter 2:** Chapter 2 is in preparation for publication. These studies were supported by endowment funding from the UCLA Eli & Edythe Broad Center of Regenerative Medicine and Stem Cell Research. Training grants provided support to Kevyn Hart (NIH TL1 DK132768 and U2C DK129496). The UCLA Clinical Translational

Science Institute provided funding for sequencing through the T1/T2 Accelerator Program Core Voucher Award (NIH/National Center for Advancing Translational Science UCLA CTSI Grant UL1TR001881). The UCLA/CFAR Virology Core Lab (5P30 AI028697) provided healthy donor PBMCs.

## VITA

### EDUCATION

2016 B.S., Biomolecular Engineering  
University of California, Santa Cruz, Santa Cruz, CA

### EXPERIENCE

**Graduate Researcher** August 2019-Present  
UCLA, Laboratory of Donald B. Kohn M.D.

- Developed a novel high-titer bifunctional lentiviral vector for autologous hematopoietic stem cell gene therapy of sickle cell disease (SCD)
- DNA sequenced leukocyte cells from Adenosine Deaminase Severe Combined Immunodeficiency (ADA-SCID) patients treated with autologous CD34+ cells transduced with either a  $\gamma$ -retroviral or lentiviral *ADA* gene vector to assess transgene mutational profiles

**Research Associate I** August 2017-July 2019  
Beckman Research Institute, City of Hope, Laboratory of Lili Wang M.D., Ph.D.

- Utilized CRISPR-Cas9 technology to generate isogenic cell lines to understand the molecular mechanisms and functional impacts of the K700E mutation in the splicing factor, *SF3B1*, in hematologic malignancies
- Organized and executed weekly facial vein bleedings, blood processing, and performed 8-color flow cytometry panels to monitor lymphocyte populations and leukemia development in engrafted mice
- Investigated the role of *SF3B1* K700E mutation in the regulation of circular RNA expression in cell lines and primary patient samples

**Junior Specialist** January 2017 – July 2017  
UCSC, Laboratory of Angela Brooks, Ph.D.

- Assessed spliced alignment tools to determine the optimal tool to incorporate into our genomic pipeline. Wrote python scripts to analyze the alignment tools' ability to accurately identify transcript start position, end position, and splice junctions which led to the generation of a correction algorithm.

**Undergraduate Research Assistant** July 2015 – December 2016  
UCSC, Laboratory of Angele Brooks, Ph.D.

- Analyzed *RBFOX1* copy number variation data to study the relationship between deletion frequency and the colorectal adenocarcinoma phenotype. Examined RNA-Sequencing data from the Cancer Genome Atlas with the use of the alternative splicing tools to determine differentially expressed isoforms.

### AWARDS

1. UCLA Advanced Research Training Program in Benign Kidney, Urologic and Hematologic Disorders Fellowship Award, 2021-2024.
2. Eugene V. Cota-Robles Fellowship, UCLA, 2019.
3. Research Mentoring Institute and Diversity Fellowship, UCSC, 2016.

## PUBLICATIONS

- **Hart, K. L.**, Crisostomo, R., Zhan, L., Kononov, N., Mittelhauser, A., Bradford, K., Kohn, D.B. Quantifying the Mutational Landscape of Retroviral and Lentiviral Vectors in Gene Therapy Patients. (In Review)
- Fernandez, M. M., Yu, L., Jia, Q., Wang, X., **Hart, K. L.**, Jia, Z., Lin, R., Wang, L. Engineering Oncogenic Hotspot Mutations on SF3B1 via CRISPR-Directed PRECIS Mutagenesis. *Cancer Research Communications* 4, 9 (2024).
- Iyer, P., Zhang, B., Liu T., Jin, M., **Hart, K.**, et al. MGA deletion leads to Richter's transformation by modulating mitochondrial OXPHOS. *Science Translational Medicine* 16, 758 (2024).
- **Hart, K. L.**, Liu, B., Brown, D., Campo-Fernandez, B., Tam, K., Orr, K., Hollis, R. P., Brendel, C., Williams, D. A., Kohn, D. B. A novel high-titer, bifunctional lentiviral vector for autologous hematopoietic stem cell gene therapy of sickle cell disease. *Molecular Therapy methods & Clinical Development* 32, 2 (2024).
- Wu, Y., Jin, M., Fernandez, M., **Hart, K. L.**, Liao, A., et al. METTL3-Mediated m<sup>6</sup>A Modification Controls Splicing Factor Abundance and Contributes to Aggressive CLL. *Blood Cancer Discovery* 4, 3 (2023).
- Segura, E. E. R., Ayoub, P. G., **Hart, K. L.**, Kohn, D. B. Gene therapy for  $\beta$ -hemoglobinopathies: From discovery to clinical trials. *Viruses* 15, 3 (2023).
- White, S. L., Hart, K., Kohn, D. B. Diverse approaches to gene therapy of sickle cell disease. *Annual Review of Medicine* 74, 1 (2023).
- Cusan, M., Shen, H., Zhang, B., Liao, A., Yang, L. et al. [including **Hart, K.**]. SF3B1 mutation and ATM deletion codrive leukemogenesis via centromeric R-loop dysregulation. *The Journal of Clinical Investigation* 133, 17 (2023).
- Tang, A., Soulette, C., Baren., **Hart, K.**, Hrabeta-Robinson, E., Wu, C., Brooks, A. Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. *Nature Communications*. 11, 1438 (2020).

## INTRODUCTION

Gene therapy utilizing hematopoietic stem cells (HSCs) has emerged as a promising approach for treating a range of genetic disorders.<sup>1</sup> HSCs, which are responsible for generating all types of blood cells are a key target for gene therapy. Modifying these stem cells to correct genetic defects and reintroducing them into the patient can potentially provide long-term therapeutic benefits.

The process typically involves collecting hematopoietic stem and progenitor cells (HSPCs) from a patient's bone marrow or peripheral blood, followed by genetic modification *ex-vivo*. Many techniques including gene-editing techniques (i.e. CRISPR/Cas9) or viral vector-based delivery systems, are used to insert or correct genes within the HSPCs. After modification, the HSPCs are transplanted back into the patient. Upon transplantation, these genetically corrected HSPCs engraft into the bone marrow, where they begin to produce healthy, functional blood cells. This can provide a permanent, self-sustaining source of healthy cells, aiming to alleviate the symptoms of genetic disease.

Viral vector-mediated gene therapy has shown curative potential in treating many genetic disorders. Retroviruses and lentiviruses, are engineered to deliver therapeutic genes into target cells, leveraging their ability to integrate into the host genome.<sup>2,3</sup> This allows for stable, long-term expression of the introduced gene, making viral vectors particularly attractive in gene therapies for conditions including but not limited to sickle cell disease and severe combined immunodeficiencies (which are being investigated in this work).

Sickle cell disease (SCD) is a hemoglobinopathy, caused by a single point mutation in the  $\beta$ -globin chain of hemoglobin in all affected persons, leading to an abnormal hemoglobin protein. SCD can leave the patient with many complications including severe pain episodes (vaso-occlusive crises), strokes, and organ damage.<sup>4</sup> While curative treatment does exist through an allogeneic hematopoietic stem cell transplant, not all patients are eligible for this transplant due to associated toxicity.<sup>5</sup> If eligible, finding a matched donor can be difficult and the procedure comes with risks including graft versus host disease.<sup>5</sup> Also, allogeneic bone marrow transplants are typically only performed if the patient is experiencing severe complications. Gene therapy has shown potential in becoming a curative option for SCD.

A major limitation of gene therapy for SCD is the availability and access to a one-time potentially curative therapy for patients. While there has been major success in the SCD field with two gene therapies being FDA approved in 2023<sup>6</sup> (Casgevy, Vertex and CRISPR Therapeutics; Lyfgenia, bluebird bio), the costs of these treatments are high (\$2-\$3 million per treatment). With the future reimbursement strategy for these expensive therapies not being clear, many individuals will face challenges in accessing approved therapies. To be able to benefit the greater SCD patient population it is essential to focus on strategies to decrease the cost of producing complex autologous stem cell/gene therapies.

Currently, the approaches using lentiviral vectors (LVs) have large vector genomes, which significantly limits titer and efficiency of CD34+ HSPC transduction, increasing cost and limiting efficacy. **Chapter 1** of this work targets this limitation through the development of a high-titer bifunctional lentiviral vector, UV1-DS, that surmounts these

limits by being produced at 10 to 20-fold higher titer and needing less vector to effectively transduced CD34+ cells, which will decrease the costs of goods per patient. This vector also uses two different mechanisms to inhibit sickling – expression of a fetal-globin like anti-sickling  $\beta^{\text{AS3}}$ -globin<sup>7</sup> and inducing expression of endogenous gamma-globin by two complementary shmiRs<sup>8,9</sup> that block transcriptional factors that repress gamma-globin expression. This work and collaboration were inspired by the clinical success of each of these anti-sickling technologies individually. Clinical trials at UCLA (Dr. Donald Kohn,  $\beta^{\text{AS3}}$ -globin) and clinical trials at Boston Children’s Hospital (David Williams, BCH-BB694, shmiR to BCL11A to de-repress gamma-globin). Viral vector-based gene therapy has also shown curative results in treating adenosine deaminase severe combined immunodeficiency (ADA-SCID). ADA-SCID is a monogenic disorder caused by mutations in the *ADA* gene, leading to severely impaired immune function. Gene therapy has been a successful approach for ADA-SCID since the first patient received viral-vector gene therapy in 1990.<sup>10</sup> The first type of viral vector used a gamma-retrovirus, murine leukemia virus (MLV), but over the years lentivirus-based vector approaches were developed based on the human immunodeficiency virus (HIV-1) to minimize the risk of insertional oncogenesis.<sup>1</sup> The safety of viral vector-based gene therapy remains a concern, and developing strategies and conducting studies to elucidate its safety is a top priority.

**Chapter 2** of this work utilizes peripheral blood leukocyte samples from trials of gene therapy for ADA-SCID performed at the University of California, Los Angeles between 2009-2016.<sup>11,12,13</sup> In this study, we perform targeted DNA sequencing on the *ADA* transgene from the patients’ genome to assess mutational profiles. DNA copies of the

*ADA* transgene with imperfect sequences are expected to integrate into the patient's genome during gene therapy due to reverse transcriptase (RT), a highly error-prone enzyme involved in generating a functional DNA copy through reverse transcription. It has also been suggested, that Apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like (APOBEC) genes play key roles in mutagenesis during viral transduction.<sup>14</sup> This work presents mutational profiles of RT and APOBEC3 in ADA-SCID gene therapy treated patient cells and highlights the potential need to develop strategies to reduce these sequence errors in lentiviral vectors used for clinical gene therapy to enhance safety.

# Chapter 1: A Novel, High-Titer, Bifunctional Vector for Autologous Hematopoietic Stem Cell Gene Therapy of Sickle Cell Disease

## ABSTRACT

A major limitation of gene therapy for sickle cell disease (SCD) is the availability and access to a potentially curative one-time treatment, due to high treatment costs. We have developed a high-titer bifunctional lentiviral vector (LVV) in a vector backbone that has reduced size, high vector yields, and efficient gene transfer to human CD34+ hematopoietic stem and progenitor cells (HSPC). This LVV contains locus control region cores expressing an anti-sickling  $\beta^{\text{AS3}}$ -globin gene and two microRNA adapted short hairpin RNA simultaneously targeting *BCL11A* and *ZNF410* transcripts to maximally induce fetal hemoglobin (HbF) expression. This LVV induces high levels of anti-sickling hemoglobins (HbA<sup>AS3</sup> + HbF), while concurrently reducing sickle hemoglobin (HbS). The reduction in HbS and increased anti-sickling hemoglobin impedes deoxygenated HbS polymerization and red blood cell sickling at low vector copy per cell in transduced SCD patient CD34+ cells differentiated into erythrocytes. The dual alterations in red cell hemoglobins ameliorated the SCD phenotype in the SCD Berkeley mouse model *in vivo*. With high titer and enhanced transduction of HSPC at low multiplicity of infection, this LVV will increase the number of patient doses of vector from production lots to reduce costs and help improve accessibility to gene therapy for SCD.

## INTRODUCTION

Sickle cell disease (SCD) is a condition characterized by the production of abnormal hemoglobin (HbS), caused by a specific genetic mutation in the  $\beta$ -globin gene. This mutation results in the substitution of glutamine for valine at position 6 (E6V) and triggers the polymerization of sickle hemoglobin upon deoxygenation. SCD is associated with multiple complications including chronic hemolytic anemia, severe pain episodes (vaso-occlusive events or VOE), strokes, and organ damage.<sup>1</sup> SCD affects approximately 300,000 to 400,000 newborns annually and around 20 million people worldwide, with an estimated 100,000 individuals affected in the United States.<sup>2</sup> Current treatments include blood transfusions and medications aimed at reducing VOE and hemolysis which include Hydroxyurea<sup>3</sup>, L-Glutamine<sup>4</sup>, Crizanlizumab<sup>5</sup>, and Voxelotor<sup>6</sup>, but these options are non-curative. The only standard of care curative approach is allogeneic hematopoietic stem cell transplantation (HSCT) with a suitably-matched donor. Adverse side effects of allogeneic HSCT include acute and chronic conditioning toxicities, graft failure, and graft-vs-host disease. In addition, many patients lack appropriate donors. In recent years, autologous HSCT/gene therapy for SCD has moved from an attractive concept to clinical reality, with a variety of approaches currently in clinical testing appearing to provide sustained clinical benefits to SCD patients.

The development and manufacturing costs of gene therapy products are significant factors that may pose barriers to the accessibility of curative treatment for patients. The cost of allotransplantation or gene therapy treatment for SCD can vary depending on several factors including the type of therapy used and complications from the treatment,

the location of the treatment center, and the cost of the cell product used. With the cost of recently approved gene therapy products being ~\$1-3 million per treatment including the manufacturing of the Medicinal Drug Product as well as the administration of therapy in the setting of a myeloablative transplant with associated medical costs, many individuals will face challenges in accessing approved therapies. To help address these accessibility issues and provide benefit to a larger patient population, it is crucial to develop strategies aimed at reducing the production costs of complex autologous stem cell gene therapies. Research and development are critical to enhance vector production, increase vector yields, and develop more cost-effective approaches to scale up manufacturing.

One strategy to reduce costs of manufacturing includes engineering smaller lentiviral vectors (LVV) and enhancing both the efficiency of HSPC transduction at lower vector multiplicity of infection (MOI) and enhancing the biological effect of the payload at low MOI in the target cell population. LVVs in current clinical use for the treatment of  $\beta$ -hemoglobinopathies including SCD have relatively large genomes (e.g. 8-9 kb) that include essential enhancer elements from the  $\beta$ -globin Locus Control Region (LCR), to obtain high-level erythroid-specific expression.<sup>7</sup> These vectors are costly to produce and are relatively inefficient in transducing human CD34+ HSPC, in part due to high percentages of incomplete virion genomes and may require a high MOI due to low expression/integrated vector genome.<sup>8</sup> They have required the use of transduction enhancer compounds to improve their infectivity at clinical scale. Development of refined  $\beta$ -globin LVV with reduced sizes of the LCR based on bioinformatics design has

led to significantly improved titers and CD34+ cell infectivity compared with current clinical vectors.<sup>9</sup>

Two successful approaches to genetic therapy for SCD include expressing a modified  $\beta$ -globin gene with anti-sickling characteristics (e.g., T87Q,  $\beta^{AS3}$ ), and inducing fetal hemoglobin (HbF) by increasing expression of  $\gamma$ -globin. Fetal hemoglobin has potent anti-sickling characteristics and induction via reversing the fetal-adult hemoglobin switch has the additional benefit of concurrently and coordinately reducing  $\beta^S$ -globin production. Intracellular polymerization of deoxygenated hemoglobin is exquisitely sensitive to the concentration of HbS in the red cell.<sup>10</sup> Thus, a strategy to both increase the level of anti-sickling hemoglobin and reducing the concentration of HbS may prove most efficient in reducing HbS polymerization and thus cellular sickling phenotypes.

We developed an optimized anti-sickling  $\beta$ -globin LVV (UV1)<sup>9</sup> of minimal size using a bifunctional approach to treating SCD to help address the barriers to accessibility imposed by high vector costs. The first mechanism in this approach incorporates a modified  $\beta$ -globin gene ( $\beta^{AS3}$ -globin)<sup>11</sup>.  $\beta^{AS3}$ -globin contains three amino acid substitutions (G16D, E22A, T87Q) that give this  $\beta$ -globin variant anti-sickling properties similar to fetal  $\gamma$ -globin. These amino acid changes incorporated into the  $\beta^{AS3}$ -globin polypeptide reduce sickle polymerization through disruption of axial and lateral contact with the canonical valine 6 (Val6) of sickle  $\beta$ -globin and also confer a competitive advantage over the sickle  $\beta$ -globin chain for binding to  $\alpha$ -globin chains to form hemoglobin tetramers. Vectors carrying the  $\beta^{AS3}$ -globin transgene corrected hematologic and clinical findings in the Townes Sickle Cell mouse model, and were also

shown to transduce SCD patient BM CD34+ cells and induce therapeutic levels of HbA<sup>AS3</sup>-globin to correct red blood cell (RBC) physiology.<sup>11,12</sup>

The second approach incorporated in this vector utilizes microRNA adapted short hairpin RNAs (shmiRs)<sup>13</sup> to simultaneously target *BCL11A* and *ZNF410*, two independent repressors of  $\gamma$ -globin expression, to induce HbF. HbF induction is a strategy in current clinical testing to ameliorate SCD phenotypes based on the observation that elevated levels of fetal hemoglobin attenuate clinical severity of SCD<sup>14,15,16</sup>. The prime example is co-inheritance of mutations causing hereditary persistence of fetal hemoglobin (HPFH) with SCD leads to marked attenuation of SCD phenotypes in comparison to individuals without HPFH.<sup>17</sup> *BCL11A* was identified as an important repressor of fetal globin expression based upon GWAS mapping.<sup>18, 19</sup> Further studies showed that generating a *BCL11A* knockout in SCD mouse models corrected the pathogenic defects associated with SCD through increased HbF expression.<sup>20</sup> A LVV expressing a *BCL11A* shmiR (BCH-BB694) only in the erythroid lineage under the control of the  $\beta$ -globin promoter and regulatory elements derived from HS2 and HS3 of the LCR ameliorated the sickle phenotype in mice and induced up to 40% HbF induction in erythroid differentiated SCD CD34+ cells.<sup>21</sup> A Phase I clinical trial with BCH-BB694 showed a sustained increase of HbF levels with a median of 30.5% of all hemoglobin levels in six patients with significant mitigation of sickle phenotype at an average *in vivo* VCN of ~1 copies per diploid genome (cpdg).<sup>22</sup> In addition to *BCL11A*, *ZNF410* has also been shown to be a repressor of  $\gamma$ -globin expression.<sup>23, 24</sup> Combining, *BCL11A* and *ZNF410* shmiRs has been shown to increase HbF induction by an additional ~10% compared to knockdown of *BCL11A* alone, with enhanced anti-sickling results in SCD erythroid differentiated CD34+ cells.<sup>13</sup> Brusson

et al<sup>25</sup> reported a bifunctional lentiviral vector for SCD that expressed the same  $\beta^{\text{AS3}}$ -globin gene described here combined with an artificial microRNA to HbS. They observed a higher level of correction of parameters of SCD by this bifunctional vector compared to one expressing only the  $\beta^{\text{AS3}}$ -globin gene.

Here we have combined both approaches in a small, highly efficient vector for treating SCD. The bifunctional vector described here (UV1-DS) maintains high titers of production and high CD34+ cell transduction activity at reduced MOI, which may provide reduced cost of manufacturing for the LVV component of autologous hematopoietic stem cell gene therapy. We demonstrate that differentiation of hematopoietic stem/progenitor cells (HSPCs) transduced with the UV1-DS vector leads to reversal of the sickle cellular phenotype with cellular parameters equivalent to normal red blood cells.

## RESULTS

### Design and Assessment of UV1-DS Vector

Previous studies have shown success in ameliorating the sickle cell phenotype with both *BCL11A* shmiR and  $\beta^{\text{AS3}}$ -globin technologies. To determine whether combining these technologies could be accomplished with effective packaging and higher titers and improved gene transfer we cloned the *BCL11A* shmiR into the UV1 vector that expresses  $\beta^{\text{AS3}}$ -globin. To define the optimal location to incorporate the *BCL11A* shmiR into the UV1 vector, we cloned the *BCL11A* shmiR sequences into five different locations throughout the  $\beta^{\text{AS3}}$ -globin cassette. Multiple locations were selected due to the possibility of the shmiR disrupting  $\beta^{\text{AS3}}$ -globin RNA processing. The locations consisted of two sites in intron 1 (at the start of IVS1 and the end IVS1), two locations in intron 2 (intra $\Delta$  and the end IVS2), and one location in the 3'UTR (Figure 1.1A). Intron 2 had been previously modified by removing sequence that were detrimental to high titer vector production.<sup>26</sup> Removal of this region also reduced the length of the vector without significantly reducing  $\beta^{\text{AS3}}$ -globin expression. These locations were further screened bioinformatically to avoid mRNA splicing and branchpoint sequences.

The resulting vector plasmids were Sanger sequenced to confirm correct construction and packaged using a HEK293T *PKR* knockout cell line. A host *PKR* response is initiated when transfecting with opposite oriented expression cassettes leading to inhibited synthesis of viral proteins resulting in lower titers.<sup>27</sup> This producer line is designed to yield higher titer from the vectors with reverse orientation expression cassettes and has shown to increase titers of  $\beta^{\text{AS3}}$ -globin vectors by 2-5 fold.<sup>27,28</sup> We saw a minimal decrease in the unconcentrated viral titers with any of the five UV1-

shmiR vectors in comparison to the UV1 control (Figure 1.1B). In particular, the shmiR in the end IVS2 position retained a comparable titer to the parental UV1 vector.

To assess the functionality of the vectors, we transduced CD34+ cells at a MOI of 20 from two different healthy donors, each in triplicate, and performed erythroid differentiation. There was a mean VCN of 4.1 +/- 2.0 for UV1 and a range of mean VCN from 2.7 +/- 1.5 to 3.4 +/- 1.4 for all the UV1-shmiR vectors (Figure 1.1C).

High performance liquid chromatography (HPLC) analysis of hemoglobin species in the differentiated erythrocytes demonstrated that each of the UV1-shmiR vectors induced similar mean expression of  $\beta^{AS3}$ -globin per copy (3.9 +/- 1.2% to 4.2 +/- 0.8%) (Figure 1.1D), except for the vector with the *BCL11A* shmiR in the 3'UTR location. The UV1-shmiR vectors expressed 8.1 +/- 0.9% to 14.4 +/- 2.4% fetal globin per vector copy (Figure 1.1E). Percentages of total anti-sickling hemoglobins were calculated as Hb $\beta^{AS3}$  expression plus HbF expression. All UV1-shmiR combination vectors outperformed UV1 in the total anti-sickling hemoglobins produced.

From these data, we concluded that four out of the five locations performed well for inducing anti-sickling  $\beta^{AS3}$ -globin expression. The location at the End-IVS2 position (**UV1-SS**) was selected for further studies as it was the candidate vector that retained a high titer and high anti-sickling hemoglobin expression. At a VCN of 1, UV1-SS led to a mean expression of 17.1 +/- 2.4% anti-sickling hemoglobins (HbA $^{AS3}$  plus HbF) compared to 5.7 +/- 1.9% for UV1 (Figure 1.1F).<sup>16</sup>

UV1-DS, a double shmiR vector in the UV1 backbone incorporating the *ZNF410* shmiR with the *BCL11A* shmiR and  $\beta^{AS3}$ -globin

Incorporating the *ZNF410* shmiR with the *BCL11A* shmiR has been shown to increase fetal globin induction by an additional ~10%.<sup>13</sup> A vector was cloned to incorporate both the *BCL11A* and *ZNF410* shmiR at the END IVS2 location, creating a double shmiR vector in the UV1 backbone (**UV1-DS**) (Figure 1.2A). The performance of five candidate vectors were then compared, including UV1-DS, UV1-SS, UV1, double shmiR (DS)<sup>13</sup> (contains *ZNF410* and *BCL11A* shmiR), and single shmiR (SS)<sup>29</sup> (contains *BCL11A* shmiR). The shmiR sequences targeting *BCL11A* and *ZNF410* in the UV1 vectors are the same shmiR sequences in DS and SS. DS and SS will induce HbF expression but are not in the UV1 backbone and therefore do not encode the  $\beta$ -globin gene. Vectors were packaged in a HEK293T *PKR* knockout cell line<sup>8</sup> and unconcentrated and concentrated viral titers were determined through transduction of the HT-29 cell line. Unconcentrated titers across all vectors in the UV1 backbone were comparable with mean titers between 3.4 +/- 0.2e+06 TU/mL to 4.6 +/- 0.1e+06 TU/mL (Figure 1.2B). The vectors containing shmiRs, but not in the UV1 backbone, had titers approximately 20-fold lower, with mean titers between 1.8 +/- 0.0e+05 to 2.4 +/- 0.0e+05 TU/mL (Figure 1.2B).

Gene transfer efficiency into healthy donor CD34+ cells was analyzed using all five vectors. CD34+ cells were transduced with concentrated viral supernatant at four different MOIs. Transduced CD34+ cells were differentiated using myeloid cytokine stimulation (IL-3, IL-6, ckit ligand) for 14 days and VCN were measured using a droplet digital polymerase chain reaction (ddPCR) assay. Myeloid differentiation was used instead of erythroid differentiation as myeloid cell VCN has been demonstrated to more closely predict the VCN seen *in vivo* in bone marrow after xenotransplantation of

immune-deficient mice.<sup>12</sup> Notably, the VCN of DS and SS plateaued, with increasing vector concentration not resulting in increased gene transfer. Vectors with the UV1 backbone had higher gene transfer to CD34+ cells than the SS and DS vectors across a range of vector concentrations (Figure 1.2C).

#### *In vitro* assessment of the LVV using human SCD patient CD34+ cells

To evaluate the potential therapeutic impact of this series of lentiviral vectors in SCD patients, peripheral blood (PB) CD34+ cells from four different SCD donors were transduced with the 5 vectors at MOI of 50. These cells were then subsequently differentiated *in vitro* using erythroid cytokines for 18 days post transduction. VCNs were determined by qPCR, anti-sickling hemoglobin expression was determined by HPLC, and sickled cells were enumerated by microscopy after sodium metabisulfite (MBS) treatment. The VCN generated with this MOI were in an appropriate range for comparison in subsequent analyses, with VCN of all experimental arms between an average of 1.3 +/- 0.1 to 1.9 +/- 0.2 (Figure 1.3A). To assess the functionality of UV1-DS in retaining the properties of *BCL11A* shmiR and *ZNF410* shmiR, we examined their knockdown efficiency by evaluating mRNA expression levels. As shown in Figure 1.3B and 1.3C, UV1 had no impact on the expression of *BCL11A* and *ZNF410*. In contrast, the groups that included *BCL11A* shmiR (UV1-SS and SS) effectively suppressed *BCL11A* expression without affecting *ZNF410*. Notably, the groups incorporating both *BCL11A* and *ZNF410* shmiRs (UV1-DS and DS) exhibited a significant reduction in the expression of both *BCL11A* and *ZNF410*.

Each of the vectors induced expression of anti-sickling hemoglobins; the UV1-DS vector induced the highest levels of anti-sickling hemoglobin (HbA<sup>AS3</sup> and HbF) at an average

of 57.5 +/- 4.6%, while the UV1, SS, UV1-SS, and DS vector induced anti-sickling globin expression at an average of 26.3 +/- 0.1%, 23.5 +/- 9.0%, 46.9 +/- 0.6%, and 30.7 +/- 9.7% respectively (Figure 1.3E). When normalized to VCN, the UV1-DS vector induced the highest anti-sickling hemoglobin expression per VCN (30.7 +/- 2.5%), which was significantly higher compared to SS (18.1 +/- 3.0%) (Figure 1.3F).

Sodium MBS treatment of transduced enucleated erythroid cells was associated with decreased percentages of sickled erythrocytes with all vectors tested compared to non-transduced ("mock") controls. Cells treated with UV1-DS showed the fewest sickled cells at an average of 14.6 +/- 3.4%, while the average sickled cells in mock, UV1, SS, UV1-SS, DS were 58.8 +/- 5.0%, 32.7 +/- 0.4%, 37.8 +/- 4.5%, 19.8 +/- 2.9%, 31.8 +/- 6.0% respectively (Figure 1.3G). Taken together, these data demonstrate that incorporating a double shmiR vector in the UV1 backbone is feasible and further enhances anti-sickling hemoglobin concentrations, and largely mitigates the cellular phenotype of erythrocyte sickling.

#### *In Vivo* Analysis of Peripheral Blood from Berkeley SCD Mouse Model

To determine whether the UV1-DS vector can ameliorate characteristic SCD disease cellular phenotypes *in vivo*, we utilized a transplantation model with Berkeley SCD (BERK-SCD) mouse HSPCs as donor cells (Figure 1.4). We designed a UV1-DS(m) for murine BM cells, replacing the *ZNF410* shmiR with a *Zfp410* shmiR. *Zfp410* shmiR targeted the murine sequence and transduction led to the knock down of *Zfp410* and induction of *Hbb-y* mRNA expression in mouse erythroid leukemia (MEL) cells (Figure 1.5). We conducted a direct comparison of UV1-DS, UV1-SS, UV1, DS, and SS vectors at equivalent VCN after transduction of lineage-negative CD45.2+ BM cells from BERK-

SCD mice. Lineage-negative CD45.2+ BM cells from BERK-SCD mice were pre-stimulated for 36-40 hours and were transduced with vectors at different MOIs to achieve a VCN of 2 (based on prior assays). The transduced cells were injected into lethally irradiated (11 Gy) CD45.1+ BL/6 (B6.SJL-Ptprca Pepcb/BoyJ) recipient animals 24 hours post transduction. Two independent experiments were performed.

Engraftment was determined by flow cytometric enumeration of CD45.2+ donor cells; hemoglobin (HGB), hematocrit (HCT), reticulocyte (Retic) concentrations were measured on PB samples and the frequencies of CD71+Ter119+ erythroid precursors were measured by flow cytometry. The percentages of sickled RBCS were determined after sodium MBS treatment of blood samples *ex vivo*.

Engraftment across all experimental arms was above 87% and did not significantly differ between different vectors (Figure 1.6A). The HGB and HCT levels at 16 weeks were not significantly different when recipients of UV1-DS transduced cells were compared to the healthy donor arm (Figure 1.6B, 1.6C). The HGB of the UV1-DS recipient mice averaged 11.7 +/- 2.3 g/dl, while the HCT of the SS, UV1, and SCD arms averaged 9.8 +/- 1.3 g/dl, 8.5 +/- 0.5 g/dl, and 7.2 +/- 0.4 g/dl respectively. The HCT of the UV1-DS recipient mice averaged 46.0 +/- 7.2%, while the HCT of the SS, UV1, and SCD arms averaged 38.5 +/- 3.8%, 36.0 +/- 3.1%, and 30.0 +/- 1.4% respectively.

As a sensitive indicator of hemolysis, we measured reticulocyte and erythroid precursor cell populations in the PB. The reticulocyte percentages and the frequency of erythroid precursor cells of the UV1-DS experimental arm (6.5 +/- 2.4%, 5.6 +/- 1.9%) showed no statistically significant difference ( $p=0.14$ ,  $p=0.39$ ) with the healthy control (2.3 +/- 0.3%, 1.7 +/- 0.6%) (Figure 1.6D, 1.6F). The UV1-DS arm showed a significant decrease

( $p < 0.01$ ) in reticulocytes in comparison to the SS arm (6.5 $\pm$ 2.4% vs 13.1  $\pm$  2.9%) (Figure 1.6D).

*Ex vivo* quantification of sickled cells in blood harvested from mice at 16 weeks was performed. Red blood cells were exposed to 2% sodium MBS and incubated under hypoxic conditions at 37°C for 30 minutes and imaged. Mice transplanted with UV1-DS transduced cells showed an average of 10.9  $\pm$  3.3% sickled cells which was not statistically significantly different ( $p = 0.07$ ) than the healthy control arm with an average of 5.2  $\pm$  1.7% abnormally shaped cells (Figure 1.6E). These values were obtained with an average VCN of 1.5 c/dg. Overall, PB analysis suggested that the UV1-DS recipient animals demonstrated some significant improvement in hematologic parameters compared with the SS group, with no significant differences compared to animal transplanted with control cells from healthy donors.

#### Analysis of Bone Marrow after Gene Therapy in Berkeley SCD Mouse Model

Animals were sacrificed at 16 weeks for complete analyses including BM and spleen studies. The average BM VCNs were similar among all of the experimental arms with a median of 1.5 VCN/diploid genome and a range of 0.5 (Figure 1.7A).

To determine if UV1-DS demonstrated therapeutic levels of anti-sickling hemoglobin expression in BM cells, we measured  $\beta^{AS3}$ -globin and  $\gamma$ -globin RNA transcript levels by RT-qPCR in erythroid cells derived from BM purified by flow cytometry. We confirmed the SS and DS group demonstrated increased  $\gamma$ -globin while the UV1 group expressed  $\beta^{AS3}$ -globin and UV1-SS and UV1-DS induced both  $\gamma$ -globin and expressed  $\beta^{AS3}$ -globin (Figure 1.7B). The total anti-sickling globin mRNA content ( $\gamma$ -globin +  $\beta^{AS3}$ -globin) was

determined, and the expression was normalized to the VCN in BM for each mouse. The highest level of anti-sickling globin mRNA expression was seen the UV1-DS group with an average of 29.0 +/- 4.2% anti-sickling globin per 1 VCN/dg. This expression was significantly higher than the DS and SS group at 18.4 +/- 3.5% and 11.4 +/- 2.8% respectively ( $p < 0.0001$ ) (Figure 1.7C).

Spleen weights were measured as an indication of compensatory erythroid expansion. The UV1-DS treated group showed an average spleen mass of 0.16 +/- 0.05g and was not significantly different than the healthy control arm of 0.09 +/- 0.01g. UV1-DS treated spleen mass were significantly lower than the SS ( $p < 0.05$ ) group at 0.25 +/- 0.07g (Figure 1.7D). Taken together, these data demonstrate that UV1-DS transduced cells led to a robust rescue of all SCD RBC phenotypes examined.

## DISCUSSION

Gene therapies for SCD have made excellent advances in the past decade, with approaches using lentiviral vectors and CRSPR/Cas9 showing excellent clinical efficacy. A major limitation of gene therapy for SCD is the availability and access to a one-time potentially curative therapy for patients. With the cost of treatment being high (\$1-\$3 million per treatment) and the future reimbursement strategy for these expensive therapies not being clear, many individuals will face challenges in accessing approved therapies. One major factor is the high costs to produce lentiviral vectors that are produced at low titer and require relatively high MOI for effective transduction of CD34+ HSPC. The results presented here demonstrate the advantages in combining two technologies not only leading to increased efficacy, but increased titers and gene transfer which may lead to reduced production costs providing more accessible therapies.

Brendel et al<sup>21</sup> previously showed high induction of fetal globin expression per copy number with a *BCL11A* shmiR, while Morgan et al<sup>9</sup> showed that engineering smaller  $\beta^{\text{AS3}}$ -globin vectors can lead to higher titers. We wanted to investigate if combining shmiR and  $\beta^{\text{AS3}}$ -globin technologies could lead to a superior therapy which had the combined benefits of potent anti-sickling globin expression with significantly increased titers and gene transfer. We first showed proof of concept with the incorporation of the *BCL11A* shmiR into the UV1 backbone. ShmiRs integrated into the intronic regions of the  $\beta^{\text{AS3}}$ -globin cassette had minimal impact on  $\beta^{\text{AS3}}$ -globin protein expression, while incorporation into the 3'UTR region led to diminished expression of  $\beta^{\text{AS3}}$ -globin. We hypothesize that when the shmiR is processed out from the transcript in the 3'UTR, the

transcript is degraded, thereby reducing the amount of  $\beta^{\text{AS3}}$ -globin expression. Since the other shmiRs are located in intronic regions, the introns will be spliced out before translation and we hypothesize that this is before the shmiR will be processed out, therefore having minimal impact on the  $\beta^{\text{AS3}}$ -globin transcript. With the addition of the *BCL11A* shmiR, we witnessed fetal globin induction leading to an increase of total anti-sickling hemoglobins with the combination vectors. We did see that the incorporation of the shmiR lead to a small decrease in gene transfer when transduced at equal transduction units in comparison to UV1. We hypothesized that the incorporation of the shmiR which increases the length and complexity of the vector could be leading to an increase of incomplete viral genomic RNAs. This concept has been previously studied, and the data has shown that  $\beta$ -globin LVV viral genomic RNAs can be incomplete and released in vector particles leading to lower gene transfer in HSPCs.<sup>8</sup> We extracted viral RNA from UV1 and the UV1-shmiR unconcentrated viral supernatants and performed a ddPCR assay to assess the concentration of complete viral RNAs.<sup>8</sup> The data shows lower concentration of complete viral RNA in the UV1-shmiR vectors which could explain lower gene transfer (Figure 1.8).

We formulated a hypothesis suggesting that shmiRs present in UV1 would exhibit greater titer and gene transfer than in their original backbones. Upon evaluating the titer and gene transfer of the UV1-DS vector, we found both to be comparable to those of the UV1 control. This indicates that the additional shmiR sequence did not significantly affect titer and gene transfer. Moreover, the variation in backbone sequence of UV1, characterized by its smaller size and different complexity, yielded advantages in terms of higher titer and gene transfer, compared to SS and DS in a longer backbone.

Additionally, UV1-DS also showed a 2 to 8-fold increase in titer when packaged and titered head to head with alternative  $\beta^{\text{AS3}}$ -globin LVVs<sup>11,30</sup> (Figure 1.9B). The UV1 backbone has a substantial benefit because lower transduction concentrations can be used to achieve the same VCN, meaning lower viral volumes will be needed, leading to less potential toxicity for the cells and reduced cost per patient dose. For example, a five-thousand-fold scale up from  $1 \times 10^5$  to  $5 \times 10^8$  for a hypothetical 75 kg patient and a CD34+ cell dose of  $6 \times 10^6/\text{kg}$  would require the DS vector produced from four liters for one patient dose versus 0.2 liter for the UV1-DS vector; thus UV1-DS would yield 20-fold more patient doses per vector lot.

A crucial aspect in the development of this therapy involved evaluating whether the combination would result in improved efficacy. Previous research has demonstrated that achieving approximately 20% expression of anti-sickling hemoglobins leads to an improvement in the sickle cell disease phenotype.<sup>31</sup> Based on this, our hypothesis was that the UV1-DS vector would exhibit the highest expression of anti-sickling hemoglobins by combining  $\beta^{\text{AS3}}$ -globin expression and  $\gamma$ -globin induction, potentially allowing for a lower VCN requirement to ameliorate the disease. To evaluate the efficacy of the UV1-DS vector, we utilized human SCD CD34+ cells and conducted erythroid differentiation. We observed the highest induction of anti-sickling hemoglobins in UV1-DS treated cells (approximately 30% per VCN) resulting in a significant reduction in the sickling phenotype when compared with the UV1, SS, or DS vectors. Based on this experiment we would expect that human SCD CD34+ cells transduced with UV1-DS at VCN of 1 would achieve therapeutic levels. This finding supports the notion that the combination therapy holds promise for effectively addressing sickle cell disease at a

lower VCN, which in turn can lead to less safety concerns regarding insertional mutagenesis.

We then assessed amelioration of hematologic parameters of the sickle phenotype in the BERK SCD mouse model. We excluded mice with <97% donor engraftment at week 16 from the analysis to prevent confounding of phenotype correction by residual WT erythrocytes. We made the decision to perform *in vivo* experimentation at equal VCN to allow for intrinsic anti-sickling activity of each vector to be determined. However, using different amounts of each vector to achieve similar VCN obscures the potential advantages of smaller, higher titer vectors to transduce higher percentages of cells and achieves sufficient VCN to impede sickling. Of note, experimental arms using LVV in the UV1 backbone required less vector volume for transduction to achieve the same VCN as the shmiR only vectors in a longer backbone (SS and DS), which suggests not only less potential cell toxicity when translated clinically but also lowered costs of therapy. All hematologic parameters assessed in the PB of recipients treated with UV1-DS showed significant improvements when compared to mock treated and UV1 treated Lin<sup>-</sup> cells. Significant improvements in hematocrit and reticulocyte percentages were also seen when compared to the SS arm. When assessing BM, there was a significant increase in percentages of anti-sickling hemoglobins per VCN as well as a significant decrease in spleen weights of the UV1-DS treated group in comparison with the UV1 and SS arm. These data emphasize that the UV1-DS vector has potential to be advantageous in a clinical setting in comparison to  $\beta^{AS3}$ -globin and shmiR technologies separately.

We also performed *in vivo* experimentation in the Townes SCD mouse model (Figure 1.10-1.12) but saw minimal to no induction of HbF in mice treated with vectors

containing shmiRs. A recent article which has characterized the Townes mouse model by long-range sequencing showed the lack of distal gene-regulatory elements that may be necessary for HbF induction.<sup>32</sup> Several human assay for transposase-accessible chromatin sequencing (ATAC-seq) signals seen in the human  $\beta$ -globin locus were not included in the Townes mouse transgenes and are also not conserved in mice. These elements may be involved with hemoglobin switching, a mechanism that does not naturally occur in mice, leading to potential suboptimal expression of the *HBG1* transgene.<sup>32</sup> In addition, human ATAC-seq peaks for known gamma-globin regulators were also not seen in the Townes mouse transgene including: *HBBP1*<sup>33</sup> and *BGLT3*<sup>34,32</sup>. While *Bcl11a* and *Zfp410* may have been successfully downregulated in our Townes model, we hypothesize that lack of key regulatory elements may be the reason we witnessed minimal HbF induction. Our results suggest that the Townes mouse is not an optimal model when assessing HbF induction by *BCL11A* and *Zfp410* shmiRs.

The incorporation of shmiRs into the UV1 backbone has yielded a vector that not only induces high anti-sickling hemoglobin expression but also exhibits higher titer and gene transfer capabilities, compared to earlier constructs. It is worth noting that previous clinical trials involving  $\beta^{\text{AS3}}$ -globin and *BCL11A* shmiR technology have shown considerable success, but by utilizing UV1-DS vector, we can capitalize on the remarkable efficacy observed with shmiR technology, while simultaneously maintaining the high titers and gene transfer capabilities associated with the UV1 vector. Consequently, UV1-DS holds the potential for significant advantages in terms of efficacy, clinical-scale production, and cost reduction for autologous gene therapy.

There promising features make UV1-DS an exciting prospect for further exploration and potential implementation in future therapeutic approaches.

## MATERIALS AND METHODS

### *Cloning and Vector Production*

The UV1 vector, BCL11A shmiR, and ZNF410 shmiR, Zfp410 shmiR have been described previously.<sup>9,21,13,35</sup> To introduce the BCL11A shmiR, 5 sets of reverse-oriented primers with extended homology sequence were used to PCR amplify the UV1 plasmid backbone. The BCL11A shmiR oligos (Integrated DNA Technologies, San Diego, CA, USA) were combined and an oligo duplex was generated. With homology between the shmiR sequences and linearized plasmids, the shmiR sequences were joined using the NEBuilder HiFi DNA Assembly kit (New England Biolabs, Ipswich, MA, USA). All plasmids were sequence verified by Sanger sequencing (Laragen Inc, Culver City, CA, USA).

HEK293T *PKR* knockout cells<sup>8</sup> were plated on 10cm plates at a density of  $1 \times 10^7$  cells/mL and vectors were packaged through transient transfection, using 3<sup>rd</sup> generation lentiviral packaging plasmids.<sup>36</sup> Raw viral supernatant was collected 3 days post transfection, samples were used for titer determination, and the remainder of the viral supernatants were concentrated through ultracentrifugation. Titers were determined by performing transduction of HT-29 human colorectal carcinoma cell line at multiple dilutions of both raw and concentrated viral supernatant. 3 days post transduction, cells were harvested and titers were calculated through VCN determination by ddPCR assay using primers and probes for HIV-1 PSI (forward 5'-AAGTAGTGTGTGCCCGTCTG-3', reverse 5'-CCTCTGGTTTCCCTTTCGCT-3', 5'-56-FAM-AGCTCTCTC-ZEN-GACGCAGGACTCGGC-3IABkFQ-3') and the Human Syndecan 4 gene (SCD4) as a reference (forward 5'-CAGGGTCTGGGAGCCAAGT-3', reverse 5'-

GCACAGTGCTGGACATTGACA-3', 5'-5HEX-CCCACCGAA-ZEN-  
CCCAAGAACTAGAGGAGAAT-3IABkFQ-3').

*Peripheral Blood Healthy CD34+ Transduction and Erythroid Differentiation*

Peripheral blood CD34+ samples were obtained from healthy donors by plerixafor and granulocyte colony-stimulating factor (G-CSF) mobilization. Cells were thawed and plated at  $1 \times 10^6$  cells/mL on non-tissue culture treated 96 well plates pre-coated with retronectin (20ug/mL, Takara Shuzo, Otsu Japan). Cells were pre-stimulated for 24 hours in X-Vivo 15 medium (Lonza, Basel, Switzerland) supplemented with 1x glutamine, penicillin, and streptomycin (Gemini Bio-Products, Sacramento, CA, USA), human Flt-3 ligand (50 ng/mL), human stem cell factor (50 ng/mL), human thrombopoietin (50ng/mL), and human interleukin-3 (20 ng/mL) (cytokines: PeproTech, Rocky Hill, NJ, USA). CD34+ cells were transduced with concentrated viral supernatants at a transduction concentration of  $2 \times 10^7$  TU/mL (MOI: 20), without additional transduction enhancers. 24 hours post transduction, the cells were washed and plated under erythroid culture conditions. Days 2-7 post transduction the cells were cultured in erythroid differentiation base medium (EDM) consisting of Iscove modified Dulbecco's medium (Lonza, Basel, Switzerland), 1x glutamine, penicillin, and streptomycin (Gemini Bio-Products, Sacramento, CA, USA), Holo-human transferrin (330ug/mL) (Sigma-Aldrich, Burlington, MA), recombinant human insulin (10ug/mL) (Sigma-Aldrich, Burlington, MA, USA), heparin (2 IU/mL) (Sigma-Aldrich, Burlington, MA, USA), 5% human solvent detergent pooled plasma AB (Octapharma USA Inc., Paramus, NJ, USA) supplemented with hydrocortisone (1uM) (Sigma-Aldrich, Burlington, MA, USA), human interleukin-3 (5 ng/mL) (PeproTech, Rocky Hill, NJ, USA),

human stem cell factor (100 ng/mL) (PeproTech, Rocky Hill, NJ, USA), and erythropoietin (3 IU/mL) (Sigma-Aldrich, Burlington, MA, USA). Day 8-10 post transduction the cells were cultured in EDM supplemented with 3 IU/mL erythropoietin. Day 11-21 post transduction the cells were cultured in EDM without added cytokines. Cells were collected 14 days post transduction and ddPCR assays were used to analyze vector copy number (VCN) and mRNA expression for  $\gamma$ -globin and  $\beta^{\text{AS3}}$ -globin. At day 21, cells were collected for protein analysis using high-performance liquid chromatography (HPLC) to assess adult hemoglobin (HbA), fetal hemoglobin (HbF) and  $\beta^{\text{AS3}}$ -globin (Hb $\beta^{\text{AS3}}$ ).

#### *Myeloid Dose Response*

Peripheral blood CD34+ samples were obtained from healthy donors from plerixafor and G-CSF mobilization. Cells were thawed and plated at  $1 \times 10^6$  cells/mL on non-tissue culture treated 96 well plates pre-coated with retronectin (20ug/mL, Takara Shuzo, Otsu Japan). Cells were pre-stimulated for 24 hours in X-Vivo 15 medium (Lonza, Basel, Switzerland) supplemented with 1x glutamine, penicillin, and streptomycin (Gemini Bio-Products, Sacramento, CA, USA), human Flt-3 ligand (50 ng/mL), human stem cell factor (50 ng/mL), human thrombopoietin (50ng/mL), and human interleukin-3 (20 ng/mL) (cytokines: PeproTech, Rocky Hill, NJ). CD34+ cells were transduced with concentrated viral supernatants at transduction concentrations of  $2 \times 10^5$  TU/mL,  $6 \times 10^5$  TU/mL,  $2 \times 10^6$  TU/mL, and  $6 \times 10^6$  TU/mL with transduction enhancer Poloxamer 338 (1mg/mL) (BASF, Ludwigshafen, Germany). 24 hours post transduction, cells were washed and plated under myeloid culture conditions. Cells were cultured for 2 weeks post transduction in basal bone marrow media (BBMM) consisting of Iscove modified

Dulbecco's medium (Lonza, Basel, Switzerland) 1x glutamine, penicillin, and streptomycin (Gemini Bio-Products, Sacramento, CA, USA), 20% fetal bovine serum, and 0.52% bovine serum albumin (Sigma-Aldrich, Burlington, MA, USA) supplemented with human stem cell factor (25 ng/mL), human interleukin-3 (5 ng/mL), and human interleukin-6 (10 ng/mL).

#### *Transduction of human SCD patient CD34+ cells*

SCD patient CD34+ HSPCs were isolated from unmobilized PB following Boston Children's Hospital institutional review board (IRB) approval and informed patient consent. The SCD CD34+ HSPCs were enriched using the Miltenyi CD34 Microbead kit (Miltenyi Biotec, Auburn, CA, USA). CD34+ cells were prestimulated for 36–40 h at  $1 \times 10^6$  cells/mL in Stem Cell Growth Medium (CellGenix, Portsmouth, NH, USA) supplemented with stem cell factor (SCF), FMS-like tyrosinekinase 3 ligand (FLT3L), and thrombopoietin (TPO), all from Peprotech (Rocky Hill, NJ, USA). Cells were then enumerated and transduced with the vector at an MOI as indicated in presence of LentiBOOST enhancer (SIRION Biotech, Germany) for 24 h before downstream processing.

#### *In vitro erythroid differentiation of SCD CD34+ cells*

The *in vitro* erythroid differentiation protocol used is based on a three phase protocol adapted from Giarratana et al.<sup>37</sup> The cells were cultured in erythroid differentiation medium (EDM) consisting of Iscove modified Dulbecco's medium (Cellgro, Manassas, VA, USA) supplemented with 1% L-glutamine (Thermo Fisher, Waltham, MA, USA), and 1% penicillin-streptomycin (Thermo Fisher, Waltham, MA, USA), 330 mg/mL holo-

human transferrin (Sigma-Aldrich, Burlington, MA, USA), 10 mg/mL recombinant human insulin (Sigma-Aldrich, Burlington, MA, USA), 2 IU/mL heparin (Sigma-Aldrich, Burlington, MA, USA), 5% human solvent detergent pooled plasma AB (Rhode Island Blood Center, Providence, RI, USA), and 3 IU/mL erythropoietin (Amgen, Thousand Oaks, CA, USA). During the first phase of expansion (days 0–7), CD34+ cells were cultured in EDM in the presence of  $10^6$  mol/L hydrocortisone (Sigma-Aldrich, Burlington, MA, USA), 100 ng/mL SCF (Peprotech, Rocky Hill, NJ, USA), 5 ng/mL IL-3 (R&D Systems, Minneapolis, MN, USA), as EDM-1. In the second phase (days 7–11), the cells were resuspended in EDM supplemented with SCF, as EDM-2. For the third phase (days 11–18), the cells were cultured in EDM without additional supplements, as EDM-3.

#### *In vitro sickling assay*

At the completion of erythroid differentiation, enucleated RBCs were sorted, with the use of Hoechst 33342 (5 mg/mL; Invitrogen, Waltham, MA, USA), and subjected to an *in vitro* sickling assay. Sickling was induced by adding 500µL of freshly prepared 2% sodium MBS (Sigma-Aldrich, Burlington, MA, USA) solution prepared in PBS into enucleated cells resuspended with 500µL EDM-3 in a 24-well plate, followed by incubation at 37°C for 30 min. Live cell images were acquired using a Nikon Eclipse Ti inverted microscope (Nikon, Tokyo, Japan). More than 500 cells were counted for each sample; cells with irregular structure, protruding spikes, or sickle shape were counted as sickling cells.

#### *In vivo experiment in SCD mouse model*

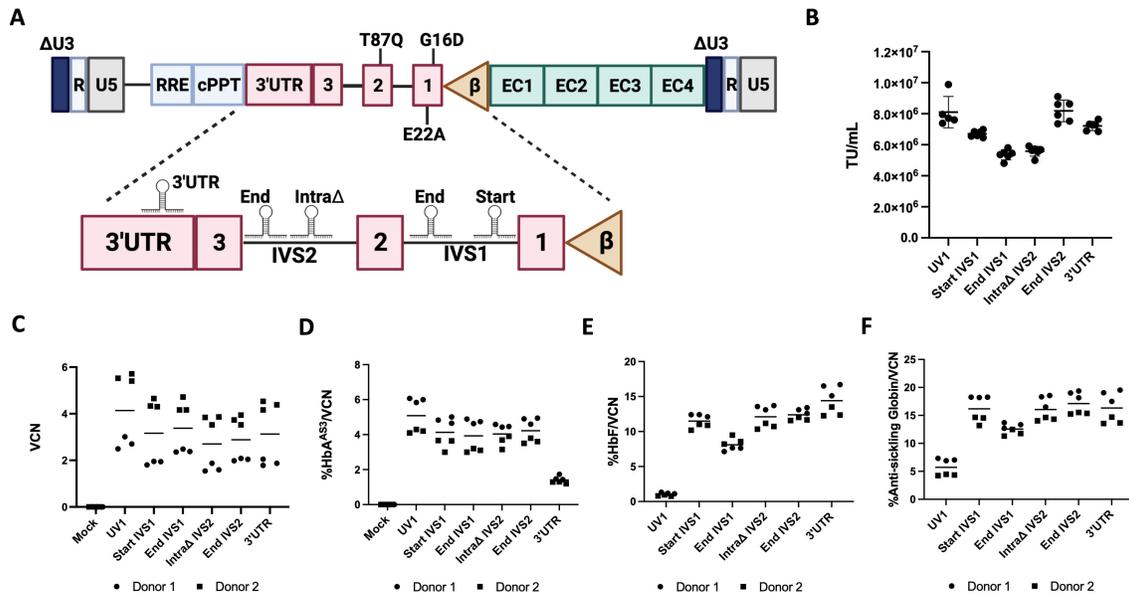
Lineage-negative (lin-) mouse BM cells were isolated by flushing femurs, tibias, and iliac crests of 6- to 8-week-old CD45.2 C57BL/6 or CD45.2 Berkeley SCD mice (BERK-SCD, JAX stock #003342) followed by lineage depletion using the Mouse Lineage Cell Depletion Kit (Miltenyi Biotec, Bergisch Gladbach, Germany). Lin- cells were pre-stimulated at  $1 \times 10^6$  cells/mL in Stem Cell Growth Medium (CellGenix) supplemented with mouse SCF (mSCF) (100 ng/mL), hTPO (100 ng/mL), mouse interleukin-3 (mIL-3) (20 ng/mL), and hFlt3-L (100 ng/mL), all from Peprotech (Rocky Hill, NJ, USA). Following a 36-40 h pre-stimulation, cells were transduced at a density of  $1 \times 10^6$  cells/mL in presence of LentiBOOST enhancer, and transduced cells (without sorting) were transplanted by retro-orbital injection into lethally irradiated (7 + 4 Gy, split dose) CD45.1 recipients (B6.SJL-Ptprca Pepcb/BoyJ, Jax Strain #002014) 24h after transduction. PB samples were collected at weeks 4, 8, 12, and 16 to measure engraftment by flow cytometry (CD45.2/CD45.1), determine RBC indices, and quantitate sickled cells. At week 16, mice were euthanized, and BM cells were used to measure engraftment by flow cytometry (CD45.2/CD45.1), VCN, and mRNA expression, spleens were collected to weigh.

#### *VCN assay for the BERK Mouse Studies*

Genomic DNA was extracted using the QIAGEN DNeasy protocol. VCN was assessed by qRT-PCR, performed with the use of TaqMan Fast Advanced Master Mix (Applied Biosystems, Waltham, MA, USA). VCN was calculated by using primers and probes HIV-1 PSI (forward 5'-CAGGACTCGGCTTGCTGAAG-3', reverse 5'-TCCCCGCTTAATACTGACG-3', probe FAM-50-CGCACGGCAAGAGGCGAGG-3') as a target and the human glycosyltransferase Like Domain Containing 1 gene (GTDC1)

as an internal reference standard (forward 5'-GAAGTTCAGGTTAATTAGCTGCTG-3', reverse 5'-TGGCACCTTAACATTTGGTTCTG-3', probe VIC-5'-ACGAACTTCTTGGAGTTGTTTGCT-3'). Standard curves were obtained by serial dilutions of a plasmid containing one copy of PSI and GTDC1 sequences. The number of PSI and GTDC1 copies in test samples was extrapolated from the standard curves.

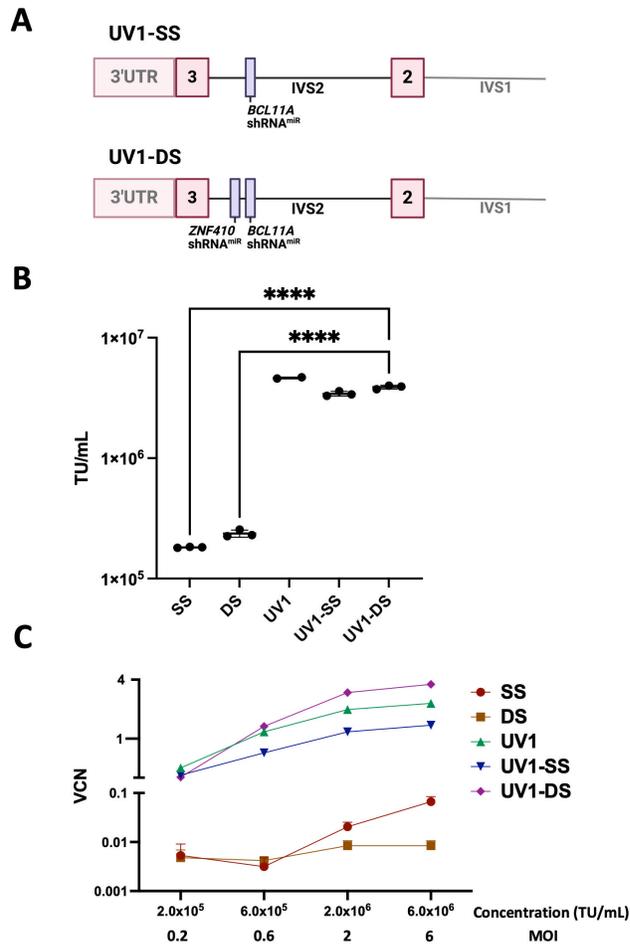
## FIGURES



**Figure 1.1**

(A) Schematic of UV1-shmiR lentiviral vector series and cloning design. The UV1 lentiviral vector (LVV) containing a  $\beta^{\text{AS3}}$ -globin cassette in reverse orientation expressed under the control of the  $\beta$ -globin promoter and Encode core (EC) enhancer regions derived and minimized from hypersensitive sites 1, 2, 3, 4 (HS1, HS2, HS3, HS4) of the  $\beta$ -globin locus control region.<sup>9</sup> (B) Unconcentrated viral titers of vectors in which the *BCL11A* shmiR<sup>35</sup> was introduced into 5 locations including: two sites in intron 1 (Start, End), two sites in IVS2 (Intra $\Delta$ , End), and one site in the 3'UTR. X axis denotes different vectors and Y axis shows titer as determined in materials and methods; (C-F) CD34+ HSPCs from 2 healthy donors were transduced in triplicate with vector constructs at  $2 \times 10^7$  TU/mL and differentiated *in vitro* under erythroid conditions. Cells were collected at day 14 for (C) VCN analysis

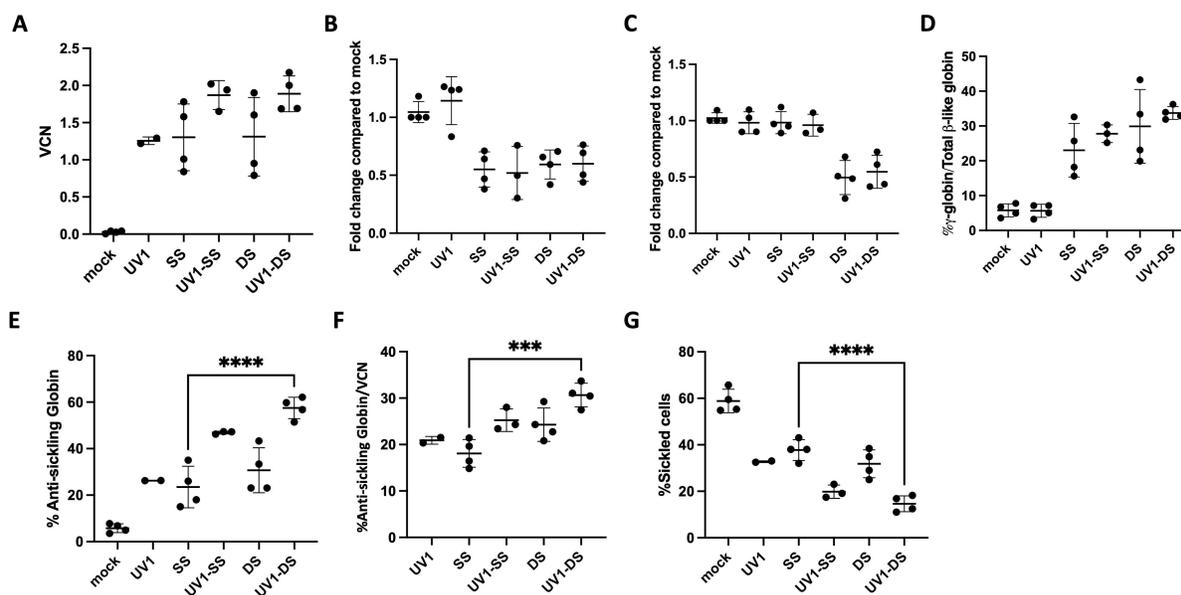
measured by ddPCR, and day 18 for (D) HbA<sup>AS3</sup> and (E) HbF by HPLC, (F) total anti-sickling globin (%HbA<sup>AS3</sup>-globin + %HbF) quantification measured by HPLC. Error bars represent means +/- standard deviation (SD). Image (A) created with BioRender.



**Figure 1.2**

(A) Map of **UV1-SS** (single shmiR) LRV containing the *BCL11A* shmiR at the end of IVS2 in the  $\beta^{\text{AS3}}$ -globin cassette. Map of **UV1-DS** (double shmiR) LRV containing both the *BCL11A* and *ZNF410* shmiR at the end of IVS2 in the  $\beta^{\text{AS3}}$ -globin cassette. Maps show detail of intron 1, site of shmiR insertion, and surrounding  $\beta$ -globin exons 2 and 3. (B) Vectors were packaged as described in *Materials and Methods* and titers were determined on HT-29 cells and quantified with ddPCR. Unconcentrated viral titer shown here. **SS** and **DS** represent vectors with single or double shmiR, expressed in a  $\beta$ -globin LCR-driven LRV previously described.<sup>13,29</sup> Each point on the plot represents vector packaged and titered from an individual 10cm plate. N=3. (C)

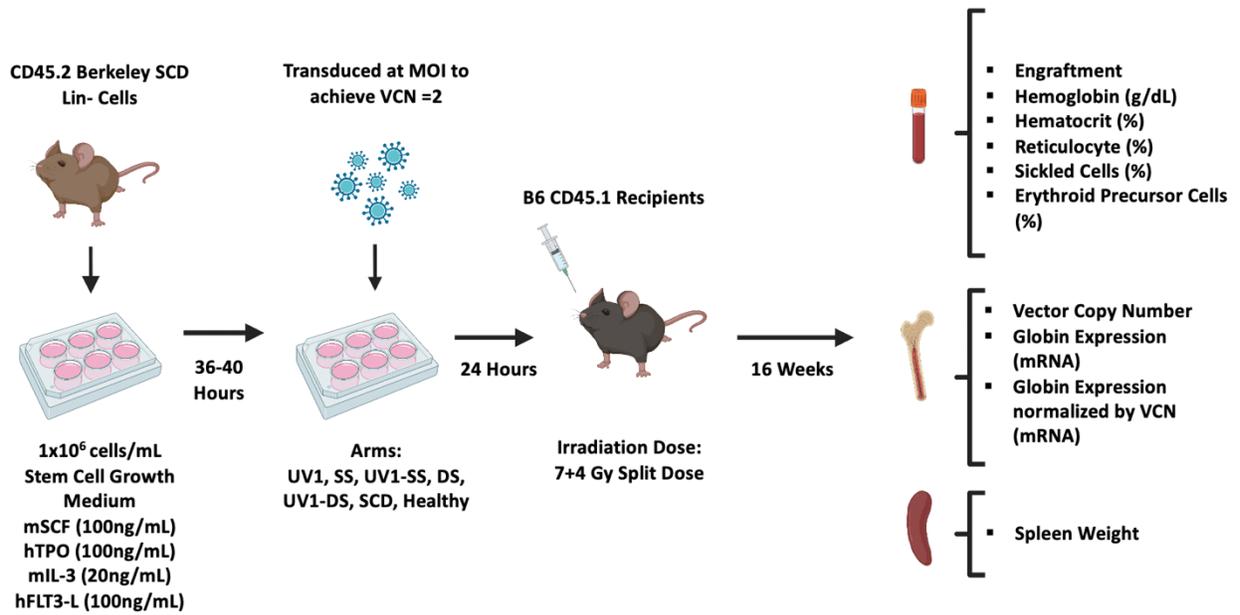
CD34+ cells from a healthy donor were transduced with constructs at  $2 \times 10^5$  TU/mL,  $6 \times 10^5$  TU/mL,  $2 \times 10^6$  TU/mL, and  $6 \times 10^6$  TU/mL (MOI: 0.2, 0.6, 2, and 6) and cultured for 14 days in myeloid differentiation conditions to assess levels of infectivity. Vector copy number (VCN) was measured by ddPCR. N=2, Error bars represent means +/- SD; Tukey's multiple comparisons test was performed on all arms with selected statistics shown. \*\*\*\*p < 0.0001. Image (A) created with BioRender.



**Figure 1.3**

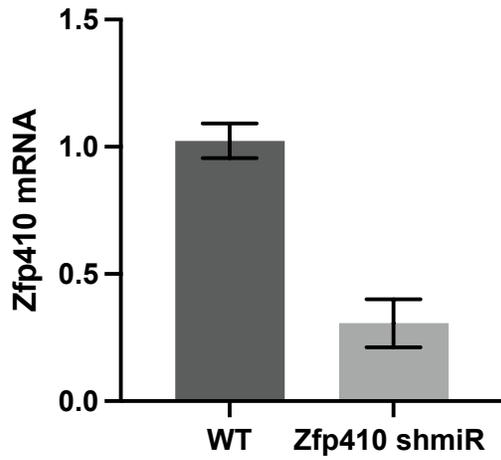
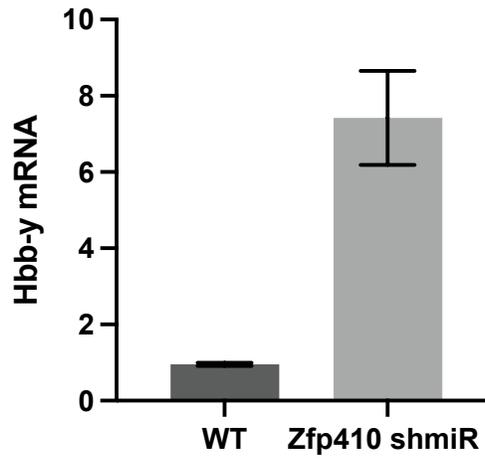
(A) Plerixafor-mobilized CD34+ HPSCs from patients with SCD were transduced with vectors at  $1 \times 10^7$  TU/mL and then differentiated *in vitro* for 18 days in erythroid culture conditions. VCN of transduced cells were determined by qRT-PCR. (B) Expression was measured by qRT-PCR with GAPDH as control on day 11 of differentiation for *BCL11A* and (C) *ZNF410*. (D) Induction of gamma-globin mRNA and (E) anti-sickling globin ( $HBB^{AS3} + HBB$ ) was determined on day 18 of differentiation by qRT-PCR. (F) Induction of anti-sickling globin ( $HBB^{AS3} + HBB$ ) normalized to VCN (G) Enucleated red blood cells were enriched by FACS and treated with sodium metabisulfite to induce sickling. Quantification of % sickled cells of enucleated erythroid cells differentiated from transduced mock vector (control) and various test vectors was assessed by phase contrast microscopy 30 min after sodium metabisulfite treatment. Error bars represent means  $\pm$  SD; each data point

represents data from cells of an individual SCD patient; Tukey's multiple comparisons test was performed on all arms with selected statistics shown. \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ .

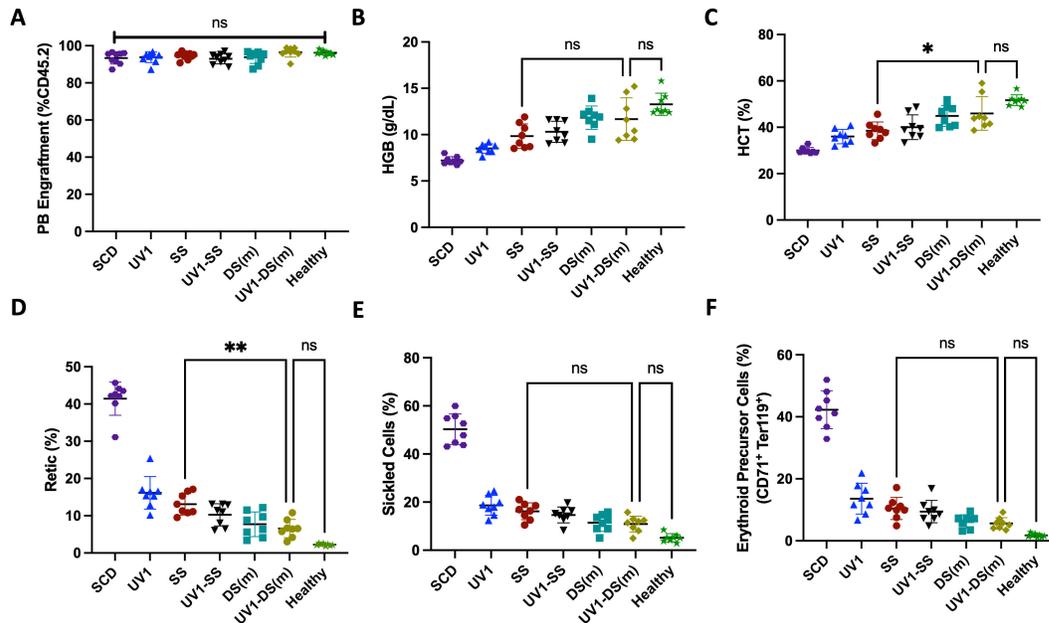


**Figure 1.4**

Schematic of experimental plan for the *in vivo* Berkeley SCD mouse model.

**A****B****Figure 1.5**

Efficient knockdown of Zfp410 by Zfp410 shmiR vector leads to high Hbb-y induction in erythroid differentiated MEL cells *in vitro*. Zfp410 and Hbb-y mRNA expression as measured by RT-qPCR with Gapdh as control. Data represent mean  $\pm$  SD.

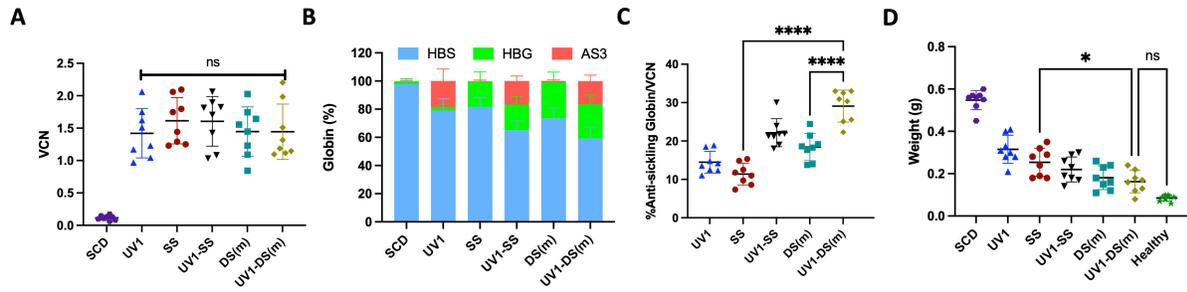


**Figure 1.6**

Lineage-negative (lin-) bone marrow cells from BERK mice (CD45.2) were transduced with each vector at MOIs adjusted to achieve similar VCN, or mock-transduced as controls, and transplanted into irradiated CD45.1+ BL/6 mouse recipients. Mice were bled at 16 weeks after transplant and peripheral blood (PB) was analyzed. (A) Engraftment was assessed in PB by flow cytometry (%CD45.2+ cells). (B) Hemoglobin (g/dL), (C) hematocrit and (D) reticulocyte counts (%) are shown. (E) PB was treated *ex vivo* with sodium metabisulfite for 30 min to induce sickling. Percentage of sickled RBCs from PB sample was quantified. (F) Percentages of CD71+ Ter119+ high erythroid precursor cell population in PB. Error bars represent mean ± SD. Symbols indicate mice transplanted with different shmiR vectors or non-transduced cells (SCD); each data point represents an individual mouse, N=8, ns, not significant; Tukey's multiple

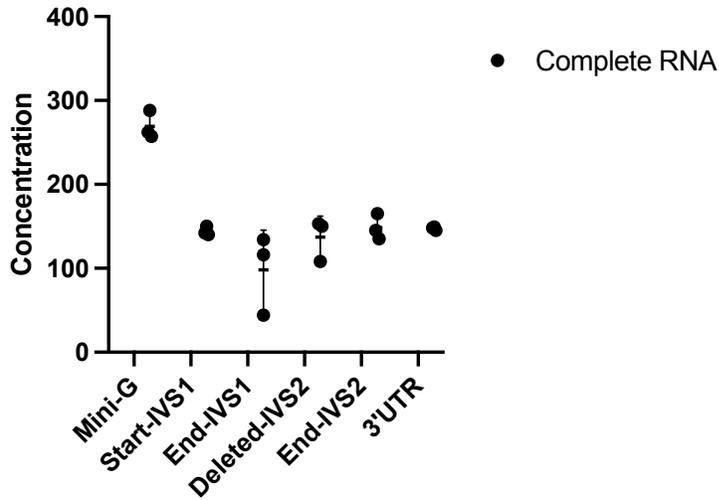
comparisons test was performed on all arms with selected statistics shown. \*p < 0.05;

\*\*p < 0.01



**Figure 1.7**

Mice were euthanized at 16 weeks after transplant and whole bone marrow (BM) and spleen was harvested and analyzed individually. (A) VCN in BM was determined by qPCR. (B) Percentages of globin mRNA expression of erythroid cells in BM determined by qRT-PCR. (C) Anti-sickling globin mRNA expression in erythroid cells in BM adjusted for VCN. (D) Spleen weights. Error bars represent mean  $\pm$  SD. Each data point represents an individual mouse. ns, not significant; Tukey's multiple comparisons test was performed on all arms with selected statistics shown. \* $p < 0.05$ , \*\*\*\* $p < 0.0001$ .

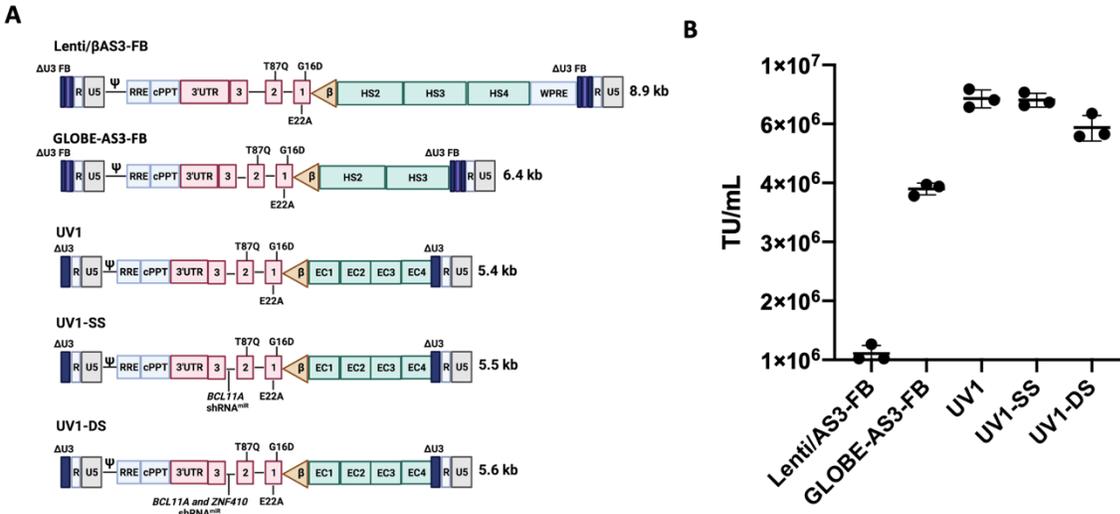


**Figure 1.8**

Vectors were packaged with an HEK293T *PKR* knock-out cell line. Unconcentrated viral supernatants were collected, and viral RNA was extracted and quantified with ddPCR.

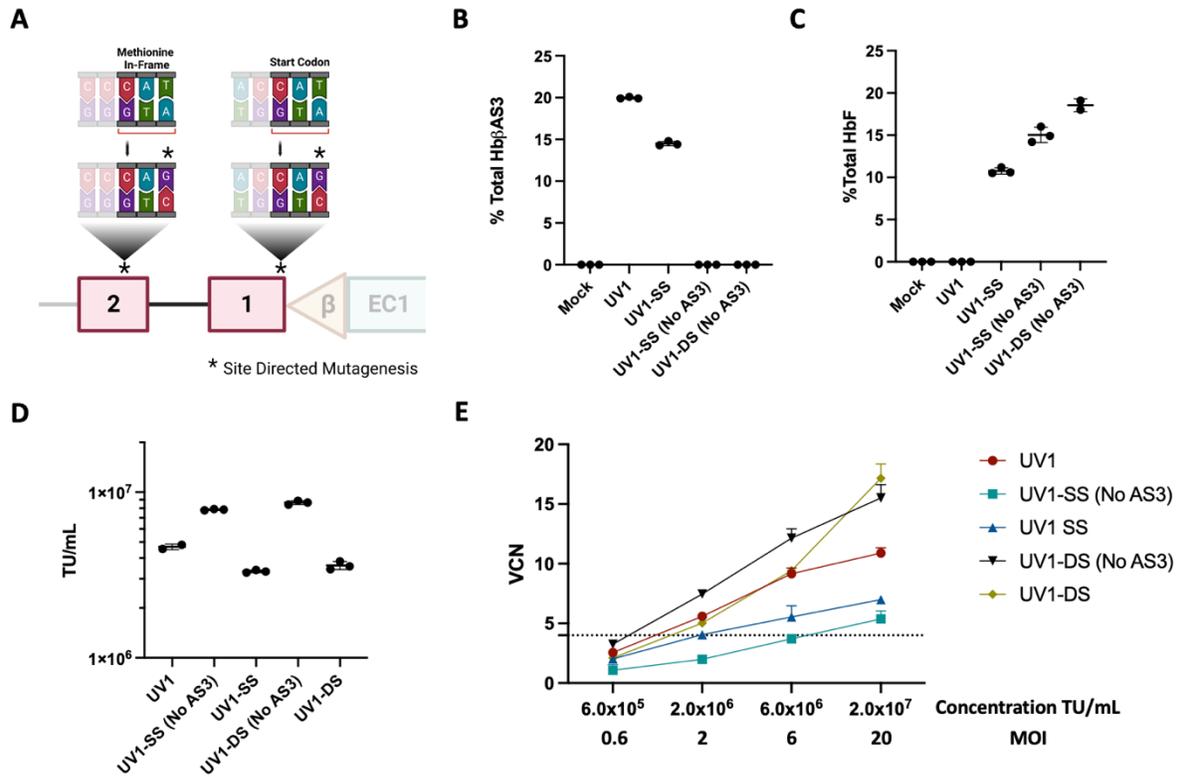
Each point on the plot represents a vector packaged from an individual 10cm plate.

Error bars represent mean  $\pm$  SD.



**Figure 1.9**

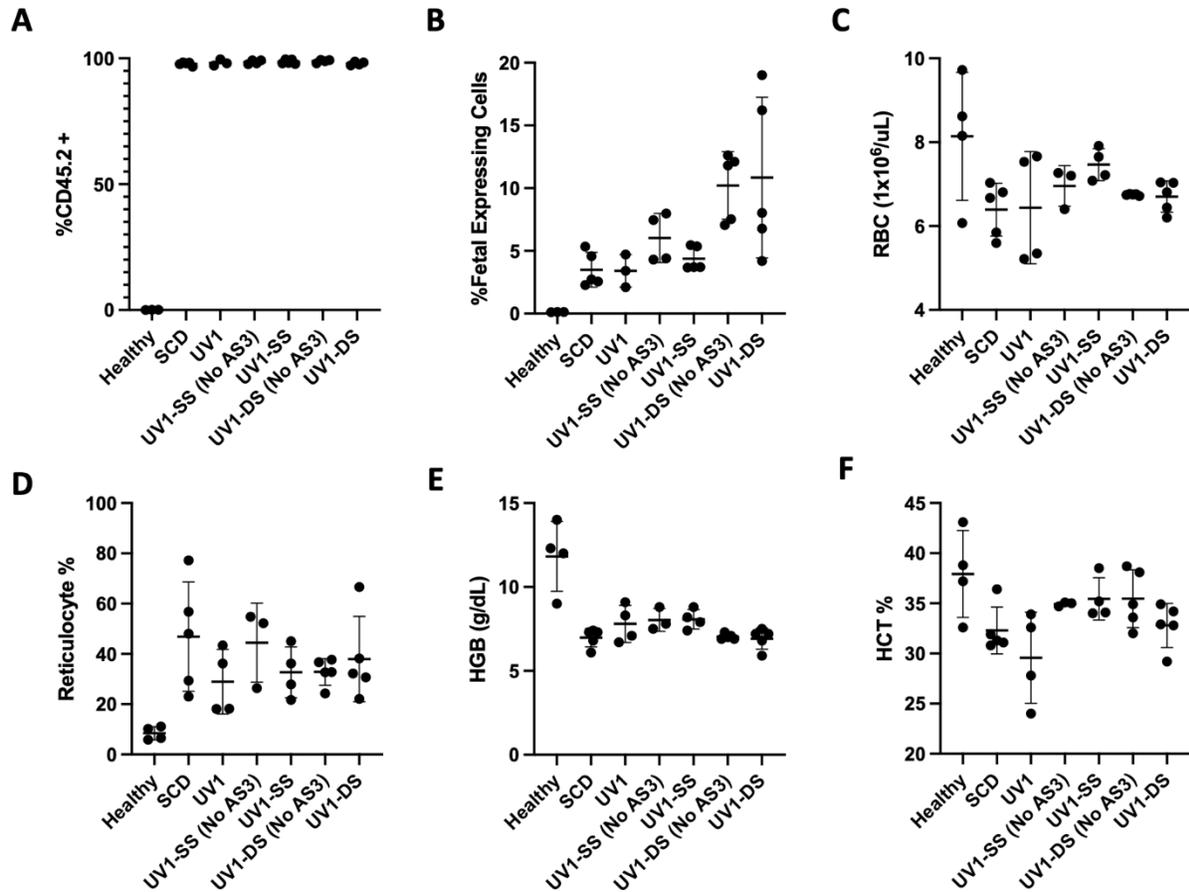
(A) Schematics of  $\beta^{AS3}$ -globin LVs: Lenti/ $\beta^{AS3}$ -FB<sup>2</sup> (8.9kb), GLOBE-AS3-FB<sup>4</sup> (6.4kb), UV1<sup>5</sup> (5.4kb), UV1-SS (5.5kb), UV1-DS (5.6kb). Vectors were packaged with an HEK293T *PKR* knock-out cell line and titers were determined by HT-29 cell line transduction, using unconcentrated viral supernatant, and quantified with ddPCR. Each point on the plot represents vector packaged and titered from an individual 10cm plate. Error bars represent mean  $\pm$  SD. Image created with BioRender.



**Figure 1.10**

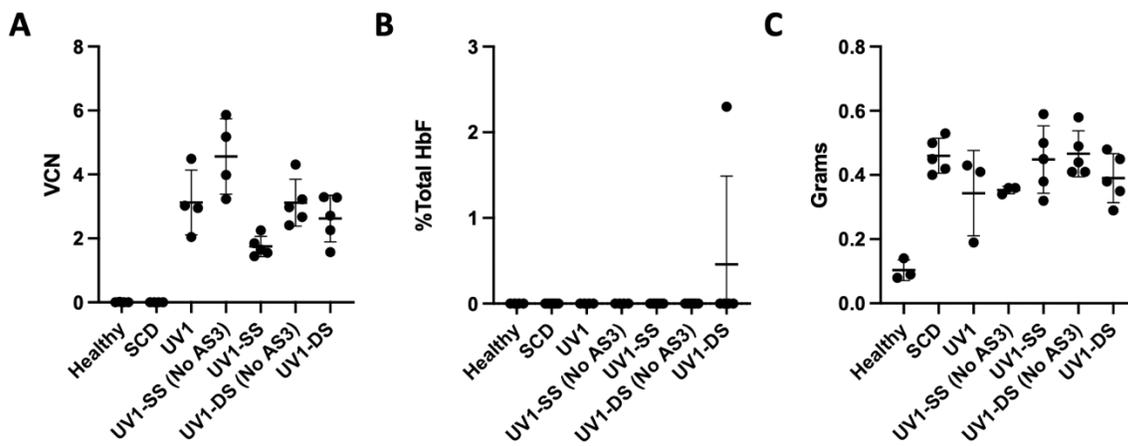
Fetal globin positive control vector design, titer, and gene transfer. (A) Site-directed mutagenesis strategy to change methionine codons (ATG) in the  $\beta^{AS3}$ -globin open reading frame (B) HUDEP-2 parental cells were transduced with vectors at  $3.0 \times 10^5$  TU/mL and then subjected to erythroid differentiation *in vitro* for 12 days and protein was assessed by HPLC. Induction of Hb $\beta^{AS3}$  (C) Induction of HbF (D) Vectors were packaged with an HEK293T *PKR* knock-out cell line and titers were determined by HT-29 cell line transduction, using raw viral supernatant, and quantified with ddPCR. Each point on the plot represents vector packaged and titered from an individual 10cm plate. (E) Lin<sup>-</sup> cells from Townes mice were transduced with constructs at  $6 \times 10^5$  TU/mL,  $2 \times 10^6$  TU/mL,  $6 \times 10^6$  TU/mL and  $2 \times 10^7$  TU/mL (MOI: 0.6, 2, 6, and 20) and cultured for 14 days under myeloid differentiation conditions to assess levels of infectivity. Vector copy

number (VCN) was measured by ddPCR. Error bars represent mean  $\pm$  SD. Image created with BioRender.



**Figure 1.11**

Peripheral blood sickle cell disease erythroid cell parameters *in vivo* in Townes SCD mouse model. Lineage negative (lin<sup>-</sup>) bone marrow cells from Townes mice (CD45.2) were transduced with each vector or mock-transduced as control and transplanted into irradiated B6 CD45.1+ (Pep Boy) mouse recipients. Mice were bled at 16 weeks after transplant and PB was analyzed. (A) Engraftment was assessed in PB by flow cytometry (%CD45.2+ cells). (B) Percentage of fetal globin expressing cells was assessed by intracellular staining and flow cytometry. (C) Red blood cell counts (1x10<sup>6</sup>/uL) (D) Reticulocyte counts (%). (E) Hemoglobin (g/dL). (F) Hematocrit. Error bars represent mean ± SD. Each data point represents an individual mouse.



**Figure 1.12**

Vector Copy Number, HbF Expression by HPLC, and Spleen Weight. Mice were euthanized at 16 weeks after transplant and whole bone marrow (BM) and spleen was harvested and analyzed individually. (A) VCN in BM was determined by ddPCR. (B) Percentages of HbF expression was determined by HPLC. (C) Spleen weights. Error bars represent mean  $\pm$  SD. Each data point represents an individual mouse.

## REFERENCES

1. CDC (2020). Complications and Treatments of Sickle Cell Disease | CDC. Centers for Disease Control and Prevention.  
<https://www.cdc.gov/ncbddd/sicklecell/treatments.html>.
2. Hassell, K.L. (2010). Population Estimates of Sickle Cell Disease in the U.S. *American Journal of Preventive Medicine* 38, S512–S521.  
10.1016/j.amepre.2009.12.022.
3. Charache, S., Terrin, M.L., Moore, R.D., Dover, G.J., Barton, F.B., Eckert, S.V., McMahon, R.P., and Bonds, D.R. (1995). Effect of hydroxyurea on the frequency of painful crises in sickle cell anemia. Investigators of the Multicenter Study of Hydroxyurea in Sickle Cell Anemia. *N Engl J Med* 332, 1317–1322.  
10.1056/NEJM199505183322001.
4. Niihara, Y., Miller, S.T., Kanter, J., Lanzkron, S., Smith, W.R., Hsu, L.L., Gordeuk, V.R., Viswanathan, K., Sarnaik, S., Osunkwo, I., et al. (2018). A Phase 3 Trial of L-Glutamine in Sickle Cell Disease. *New England Journal of Medicine* 379, 226–235.  
10.1056/NEJMoa1715971.
5. Ataga, K.I., Kutlar, A., Kanter, J., Liles, D., Cancado, R., Friedrisch, J., Guthrie, T.H., Knight-Madden, J., Alvarez, O.A., Gordeuk, V.R., et al. (2017). Crizanlizumab for the Prevention of Pain Crises in Sickle Cell Disease. *N Engl J Med* 376, 429–439.  
10.1056/NEJMoa1611770.
6. Vichinsky, E., Hoppe, C.C., Ataga, K.I., Ware, R.E., Nduba, V., El-Beshlawy, A., Hassab, H., Achebe, M.M., Alkindi, S., Brown, R.C., et al. (2019). A Phase 3

- Randomized Trial of Voxelotor in Sickle Cell Disease. *N Engl J Med* 381, 509–519.  
10.1056/NEJMoa1903212.
7. May, C., Rivella, S., Callegari, J., Heller, G., Gaensler, K.M., Luzzatto, L., and Sadelain, M. (2000). Therapeutic haemoglobin synthesis in beta-thalassaemic mice expressing lentivirus-encoded human beta-globin. *Nature* 406, 82–86.  
10.1038/35017565.
8. Han, J., Tam, K., Ma, F., Tam, C., Aleshe, B., Wang, X., Quintos, J.P., Morselli, M., Pellegrini, M., Hollis, R.P., et al. (2021).  $\beta$ -Globin Lentiviral Vectors Have Reduced Titers due to Incomplete Vector RNA Genomes and Lowered Virion Production. *Stem Cell Reports* 16, 198–211. 10.1016/j.stemcr.2020.10.007.
9. Morgan, R.A., Unti, M.J., Aleshe, B., Brown, D., Osborne, K.S., Koziol, C., Ayoub, P.G., Smith, O.B., O'Brien, R., Tam, C., et al. (2020). Improved Titer and Gene Transfer by Lentiviral Vectors Using Novel, Small  $\beta$ -Globin Locus Control Region Elements. *Molecular Therapy* 28, 328–340. 10.1016/j.ymthe.2019.09.020.
10. Seakins, M., Gibbs, W.N., Milner, P.F., and Bertles, J.F. (1973). Erythrocyte Hb-S Concentration AN IMPORTANT FACTOR IN THE LOW OXYGEN AFFINITY OF BLOOD IN SICKLE CELL ANEMIA. *J Clin Invest* 52, 422–432.
11. Levasseur, D.N., Ryan, T.M., Pawlik, K.M., and Townes, T.M. (2003). Correction of a mouse model of sickle cell disease: lentiviral/antisickling beta-globin gene transduction of unmobilized, purified hematopoietic stem cells. *Blood* 102, 4312–4319. 10.1182/blood-2003-04-1251.
12. Romero, Z., Urbinati, F., Geiger, S., Cooper, A.R., Wherley, J., Kaufman, M.L., Hollis, R.P., de Assin, R.R., Senadheera, S., Sahagian, A., et al. (2013).  $\beta$ -globin gene

transfer to human bone marrow for sickle cell disease. *J Clin Invest*, 67930.

10.1172/JCI67930.

13.Liu, B., Brendel, C., Vinjamur, D.S., Zhou, Y., Harris, C., McGuinness, M., Manis, J.P., Bauer, D.E., Xu, H., and Williams, D.A. (2022). Development of a double shmiR lentivirus effectively targeting both BCL11A and ZNF410 for enhanced induction of fetal hemoglobin to treat  $\beta$ -hemoglobinopathies. *Molecular Therapy*.

10.1016/j.ymthe.2022.05.002.

14.Platt, O.S., Thorington, B.D., Brambilla, D.J., Milner, P.F., Rosse, W.F., Vichinsky, E., and Kinney, T.R. (1991). Pain in Sickle Cell Disease. *New England Journal of Medicine* 325, 11–16. 10.1056/NEJM199107043250103.

15.Castro, O., Brambilla, D., Thorington, B., Reindorf, C., Scott, R., Gillette, P., Vera, J., and Levy, P. (1994). The acute chest syndrome in sickle cell disease: incidence and risk factors. The Cooperative Study of Sickle Cell Disease. *Blood* 84, 643–649.

10.1182/blood.V84.2.643.643.

16.Platt, O.S., Brambilla, D.J., Rosse, W.F., Milner, P.F., Castro, O., Steinberg, M.H., and Klug, P.P. (1994). Mortality In Sickle Cell Disease – Life Expectancy and Risk Factors for Early Death. *New England Journal of Medicine* 330, 1639–1644.

10.1056/NEJM199406093302303.

17.Murray, N., Serjeant, B.E., and Serjeant, G.R. (1988). Sickle cell-hereditary persistence of fetal haemoglobin and its differentiation from other sickle cell syndromes. *Br J Haematol* 69, 89–92. 10.1111/j.1365-2141.1988.tb07607.x.

18.Uda, M., Galanello, R., Sanna, S., Lettre, G., Sankaran, V.G., Chen, W., Usala, G., Busonero, F., Maschio, A., Albai, G., et al. (2008). Genome-wide association study

- shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. *Proc Natl Acad Sci U S A* 105, 1620–1625. 10.1073/pnas.0711566105.
19. Menzel, S., Garner, C., Gut, I., Matsuda, F., Yamaguchi, M., Heath, S., Foglio, M., Zelenika, D., Boland, A., Rooks, H., et al. (2007). A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. *Nat Genet* 39, 1197–1199. 10.1038/ng2108.
20. Xu, J., Peng, C., Sankaran, V.G., Shao, Z., Esrick, E.B., Chong, B.G., Ippolito, G.C., Fujiwara, Y., Ebert, B.L., Tucker, P.W., et al. (2011). Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. *Science* 334, 993–996. 10.1126/science.1211053.
21. Brendel, C., Negre, O., Rothe, M., Guda, S., Parsons, G., Harris, C., McGuinness, M., Abriss, D., Tsytsykova, A., Klatt, D., et al. (2020). Preclinical Evaluation of a Novel Lentiviral Vector Driving Lineage-Specific BCL11A Knockdown for Sickle Cell Gene Therapy. *Mol Ther Methods Clin Dev* 17, 589–600. 10.1016/j.omtm.2020.03.015.
22. Esrick, E.B., Lehmann, L.E., Biffi, A., Achebe, M., Brendel, C., Ciuculescu, M.F., Daley, H., MacKinnon, B., Morris, E., Federico, A., et al. (2021). Post-Transcriptional Genetic Silencing of BCL11A to Treat Sickle Cell Disease. *New England Journal of Medicine* 384, 205–215. 10.1056/NEJMoa2029392.
23. Vinjamur, D.S., Yao, Q., Cole, M.A., McGuckin, C., Ren, C., Zeng, J., Hossain, M., Luk, K., Wolfe, S.A., Pinello, L., et al. (2021). ZNF410 represses fetal globin by singular control of CHD4. *Nat Genet* 53, 719–728. 10.1038/s41588-021-00843-w.

24. Lan, X., Ren, R., Feng, R., Ly, L.C., Lan, Y., Zhang, Z., Aboreden, N., Qin, K., Horton, J.R., Grevet, J.D., et al. (2021). ZNF410 Uniquely Activates the NuRD Component CHD4 to Silence Fetal Hemoglobin Expression. *Molecular Cell* 81, 239-254.e8. 10.1016/j.molcel.2020.11.006.
25. Brusson, M., Chalumeau, A., Martinucci, P., Romano, O., Felix, T., Poletti, V., Scaramuzza, S., Ramadier, S., Masson, C., Ferrari, G., et al. (2023). Novel lentiviral vectors for gene therapy of sickle cell disease combining gene addition and gene silencing strategies. *Molecular Therapy - Nucleic Acids* 32, 229–246. 10.1016/j.omtn.2023.03.012.
26. Sadelain, M., Wang, C.H., Antoniou, M., Grosveld, F., and Mulligan, R.C. (1995). Generation of a high-titer retroviral vector capable of expressing high levels of the human beta-globin gene. *Proceedings of the National Academy of Sciences* 92, 6728–6732. 10.1073/pnas.92.15.6728.
27. Hu, P., Bi, Y., Ma, H., Suwanmanee, T., Zeithaml, B., Fry, N.J., Kohn, D.B., and Kafri, T. (2018). Superior lentiviral vectors designed for BSL-0 environment abolish vector mobilization. *Gene Ther* 25, 454–472. 10.1038/s41434-018-0039-2.
28. Han, J., Tam, K., Tam, C., Hollis, R.P., and Kohn, D.B. (2021). Improved lentiviral vector titers from a multi-gene knockout packaging line. *Molecular Therapy - Oncolytics* 23, 582–592. 10.1016/j.omto.2021.11.012.
29. Brendel, C., Guda, S., Renella, R., Bauer, D.E., Canver, M.C., Kim, Y.-J., Heeney, M.M., Klatt, D., Fogel, J., Milsom, M.D., et al. (2016). Lineage-specific BCL11A knockdown circumvents toxicities and reverses sickle phenotype. *J Clin Invest* 126, 3868–3878. 10.1172/JCI87885.

30. Poletti, V., Urbinati, F., Charrier, S., Corre, G., Hollis, R.P., Campo Fernandez, B., Martin, S., Rothe, M., Schambach, A., Kohn, D.B., et al. (2018). Pre-clinical Development of a Lentiviral Vector Expressing the Anti-sickling  $\beta$ AS3 Globin for Gene Therapy for Sickle Cell Disease. *Mol Ther Methods Clin Dev* 11, 167–179. 10.1016/j.omtm.2018.10.014.
31. Abraham, A., Hsieh, M., Eapen, M., Fitzhugh, C., Carreras, J., Keesler, D., Guilcher, G., Kamani, N., Walters, M.C., Boelens, J.J., et al. (2017). Relationship between Mixed Donor–Recipient Chimerism and Disease Recurrence after Hematopoietic Cell Transplantation for Sickle Cell Disease. *Biology of Blood and Marrow Transplantation* 23, 2178–2183. 10.1016/j.bbmt.2017.08.038.
32. Woodard, K.J., Doerfler, P.A., Mayberry, K.D., Sharma, A., Levine, R., Yen, J., Valentine, V., Palmer, L.E., Valentine, M., and Weiss, M.J. (2022). Limitations of mouse models for sickle cell disease conferred by their human globin transgene configurations. *Disease Models & Mechanisms* 15, dmm049463. 10.1242/dmm.049463.
33. Huang, P., Keller, C.A., Giardine, B., Grevet, J.D., Davies, J.O.J., Hughes, J.R., Kurita, R., Nakamura, Y., Hardison, R.C., and Blobel, G.A. (2017). Comparative analysis of three-dimensional chromosomal architecture identifies a novel fetal hemoglobin regulatory element. *Genes Dev.* 31, 1704–1713. 10.1101/gad.303461.117.
34. Ivaldi, M.S., Diaz, L.F., Chakalova, L., Lee, J., Krivega, I., and Dean, A. (2018). Fetal  $\gamma$ -globin genes are regulated by the BGLT3 long noncoding RNA locus. *Blood* 132, 1963–1973. 10.1182/blood-2018-07-862003.

35. Guda, S., Brendel, C., Renella, R., Du, P., Bauer, D.E., Canver, M.C., Grenier, J.K., Grimson, A.W., Kamran, S.C., Thornton, J., et al. (2015). miRNA-embedded shRNAs for Lineage-specific BCL11A Knockdown and Hemoglobin F Induction. *Mol Ther* 23, 1465–1474. 10.1038/mt.2015.113.
36. Dull, T., Zufferey, R., Kelly, M., Mandel, R. J., Nguyen, M., Trono, D., & Naldini, L. (1998). A Third-Generation Lentivirus Vector with a Conditional Packaging System. *J Virol* 72, 8463–8471. 10.1128/jvi.72.11.8463-8471.1998.
37. Giarratana, M.-C., Kobari, L., Lapillonne, H., Chalmers, D., Kiger, L., Cynober, T., Marden, M.C., Wajcman, H., and Douay, L. (2005). Ex vivo generation of fully mature human red blood cells from hematopoietic stem cells. *Nat Biotechnol* 23, 69–74. 10.1038/nbt1047.

## Chapter 2: Quantifying the Mutational Landscape of Retroviral and Lentiviral Vectors in Gene Therapy Patients

### ABSTRACT

Adenosine Deaminase Severe Combined Immunodeficiency (ADA-SCID) is a monogenic disorder caused by mutations in the *ADA* gene. Gene therapy utilizing  $\gamma$ -retroviral and lentiviral vector gene addition approaches have shown curative results. We sequenced the *ADA* transgene in transduced CD3<sup>+</sup> T-cells, and in peripheral blood cells from patients treated with autologous CD34<sup>+</sup> cells transduced with either a  $\gamma$ -retroviral or lentiviral *ADA* gene vector to assess transgene mutational profiles. In both CD3<sup>+</sup> T-cells and ADA-SCID patients' cells treated with the lentiviral vector, we observed significantly higher occurrences of guanine (G) to adenosine (A) base substitutions than with the  $\gamma$ -retroviral vector. We hypothesized that this G-to-A mutational signature was due to the APOBEC3 cytosine deaminase protein family. DNA substrate analysis showed a mean of  $86.12 \pm 4.89\%$  G-to-A mutations occurred at known APOBEC3 DNA substrate sites. APOBEC3 proteins may present a source of sequence errors in lentiviral vectors used for clinical gene therapy.

## INTRODUCTION

Adenosine Deaminase Severe Combined Immunodeficiency (ADA-SCID) is a monogenic disorder caused by mutations in the *ADA* gene of purine metabolism. ADA-SCID is a life-threatening condition because it results in severely impaired immune function. Specifically, ADA deficiency prevents the development of lymphocytes that are critical in protective adaptive immune responses. Enzyme replacement Therapy (ERT) with *ADA* is a long-term treatment that requires repeated weekly doses throughout the patient's life. Allogeneic Hematopoietic Stem Cell Transplant (HSCT) is a curative option but has the best results with an HLA-type matched sibling, to minimize the risk of Graft versus Host Disease (GvHD).

Gene therapy first became a course of action for the treatment of inherited monogenic diseases with the development of engineered retroviruses. Retroviral vectors have the ability to deliver an exogenous transgene into patients' cells by converting their RNA genome into double-stranded DNA with reverse transcriptase (RT) and integrating that DNA into the host's genome.<sup>1</sup> The first type of viral vector utilized a gamma-retrovirus, Murine Leukemia Virus (MLV), to successfully deliver the therapeutic transgene into the target cell. It has been nearly thirty-five years since the first patient received viral-vector gene therapy in 1990, for the inborn error of immunity Adenosine Deaminase Severe Combined Immune Deficiency (ADA-SCID).<sup>2</sup> The initial results proved that retroviruses can be developed as a gene transfer method to safely treat patients with monogenic diseases.

Additional studies of gene therapy followed for other inborn errors of immunity (X-linked SCID, X-linked Chronic Granulomatous Disease and Wiskott-Aldrich Syndrome), but

were soon challenged with profound concerns when some of the patients involved in this treatment developed leukemia-like leukoproliferative conditions.<sup>3</sup> The discovery of vector integrations near known proto-oncogenes in the leukemic clones halted the progress of viral vector based gene therapy.<sup>4,5</sup> New lentivirus-based vectors were developed based on the human immunodeficiency virus (HIV-1); to minimize the risk of insertional oncogenesis, the strong enhancer elements of the viral long terminal repeats were eliminated in self-inactivating (SIN) design. These lentiviral vectors exhibited a better safety profile and produced better clinical benefits due to their efficient gene transfer and stable expression in the modified cells demonstrated in long-term follow-up studies.<sup>6</sup>

The success of gene transfer is influenced by the viral vector's ability to generate a functional copy of the DNA through reverse transcription. However, RT is highly error-prone without a proof-reading function, as this contributes to the survival of the virus during infection by the generation of a swarm of quasi-species that can evade the immune response or anti-viral drugs.<sup>7,8</sup> DNA copies with imperfect sequences are expected to integrate into the patient's genome during gene therapy. The long-term consequences are often negligible when the RT generates synonymous mutations or missense mutations that result in a non-functional protein. However, certain genes can acquire gain-of-function and hyperactive mutations that can lead to malignant transformation.

Although other groups reported fidelity rates for HIV-1 and MLV RTs<sup>9</sup>, a comprehensive analysis of mutational profiles of incorporated transgenes involved in past and current viral vector gene therapy trials has not been assessed. More recently, it has been

suggested that Apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like (APOBEC) genes play key roles in mutagenesis during viral transduction.<sup>10</sup>

Duplex sequencing with targeted capture method was developed to detect low-frequency mutations.<sup>11</sup> This approach is especially helpful for viral vector transduction protocols where a low copy number of the therapeutic transgene is the ideal outcome (e.g. 1-5 copies per cell). We investigated the vector transgene from transduced cells and distinguished variants generated by the action of RT and APOBEC3. In this study, we performed duplexing sequencing and a targeted capture method to investigate the *ADA* transgenes from the patients' genomes, analyzing peripheral blood samples collected years after transplant with *ADA* gene vector modified CD34+ hematopoietic stem and progenitor cells (HSPC).

Gene therapy for adenosine deaminase-deficient severe combined immunodeficiency (ADA-SCID) has been one of the long-standing conditions studied in clinical trials of gene therapy to date, with one approach achieving approval by the European Medicines Agency in 2016.<sup>1,12</sup> From trials of gene therapy for ADA-SCID performed at the University of California Los Angeles between 2009-2016, we obtained peripheral blood leukocyte samples from five patients treated with a  $\gamma$ -retroviral vector (MND-ADA), and five patients treated with a lentiviral vector (EFS-ADA).

Using this strategy, we (1) reported low-frequency mutations in the *ADA* transgenes from patients in two different clinical trials; (2) compared the fidelity between gamma-retrovirus and lentivirus RTs, and (3) identified a relatively higher frequency of mutations generated by the action of the APOBEC3 cytidine deaminases.

## RESULTS

### Duplex Sequencing of CD3+ Cells Transduced with ADA-SCID $\gamma$ -Retrovirus and Lentivirus

Before sequencing the vector-derived *ADA* transgene and the endogenous *ADA* gene from ADA-SCID patients' peripheral blood mononuclear cells (PBMCs) and granulocytes, we optimized and assessed the sequencing pipeline to confirm the protocol was sensitive enough to detect the potentially rare mutations incorporated by RT. We acquired healthy donor PBMCs and isolated CD3+ T-cells (Figure 2.1A). CD3+ cells were transduced with either the  $\gamma$ -retrovirus (MND-ADA)<sup>2</sup> or the lentivirus (EFS-ADA)<sup>13</sup> at increasing concentrations to obtain samples with vector copy number (VCN) near to 1, to resemble the VCN of the patient PBMC samples. MND-ADA transduced CD3+ cells at a VCN of 0.7 and EFS-ADA transduced CD3+ cells at a VCN of 3.0 were selected for analyses.

A duplex sequencing library preparation protocol was selected due to the high sensitivity needed to detect rare variants and to distinguish true variants from errors incorporated through PCR amplification and sequencing, (Figure 2.1B). Duplex sequencing<sup>11</sup> utilizes dual in-line unique molecular identifiers (UMIs) that are ligated to sheared DNA which allows for the detection of errors incorporated during library preparation and sequencing. PCR amplification allows for the generation of two mirrored PCR products due to the dual UMIs. These products post sequencing, can be error corrected by grouping these reads into two mirrored read families. If a mutation is seen in all the reads in both read families, this mutation was deemed true. Mutations that were only seen in a few reads of one read family were removed as false and likely to have been

due to a mutation during PCR amplification or sequencing. Mutations seen in all the reads of one family but not the mirrored second read family were also called false and likely caused by a first-round PCR error.

An enrichment strategy was applied to the prepared library utilizing biotinylated hybridization probes. These probes were designed to bind to the  $\gamma$ -retroviral and lentiviral *ADA* cDNA transgene as well as exons 2, 7, 8, 9, 11 of the endogenous *ADA* gene. The enriched library was PCR amplified to produce the final library for sequencing.

CD3+ cell DNA was library prepped, enriched, and sequenced. FASTQ files were error-corrected and false mutations were removed. For data processing and error correction, an error-correction protocol that minimizes false positives but requires high raw coverage was selected. Data were aligned to the human genome assembly GRCh38 (hg38) as well as to the MND-ADA and EFS-ADA vector sequences. Before and after error correction and read filtering, we analyzed sequencing statistics and alignments of the *ADA* transgene and endogenous *ADA*, as well as the coverage across the *ADA* cDNA (Table 2.1).

A single hybridization capture enrichment method versus a double hybridization capture enrichment method was first compared to assess the percentage of on-target reads and to see the effects on variant calling. A double hybridization capture method has been shown to have a higher percentage of on-target reads when paired with duplex sequencing library preparation.<sup>14</sup> The data showed 30.44% and 41.58% on target alignments with a single hybridization capture and 43.26% and 98.70% on target alignments with a double hybridization capture in the MND-ADA and EFS-ADA samples

pre-error correction (Table 2.2). While higher on-target reads were seen in the samples with double hybridization capture, the MND-ADA sample had a 7-fold decrease in data in comparison to the EFS-ADA sample with a more drastic difference in enrichment between the two samples (Table 2.2). During the hybridization capture step, the libraries are pooled together to optimize reagents, and we concluded that if samples have different VCNs of the target region (*ADA* transgene), a double capture leads to greater pull down of the sample with higher VCN leading to unequal sequencing and enrichment of the pooled samples. After error correction, collapsing reads, and filtering, most alignments (greater than 90%) were aligned to the targeted regions of either the transgene or endogenous *ADA* regardless of utilizing a single or double capture with all samples having high coverage. (Table 2.2).

We then visualized the distribution of coverage across the transgene to make sure there was no drop-off of probes that could lead to low coverage in those regions. There was a uniform distribution across the EFS-ADA and MND-ADA transgene (Figure 2.2, 2.3).

#### Variant Calling of Transduced CD3+ Cells

DNA variant calling was performed using an allele frequency of 0.001 for the transgene *ADA* and the endogenous *ADA* in the samples with single capture hybridization probe enrichment. Variant calling data showed a higher frequency of mutations in the lentiviral transduced sample (Table 2.3). There was also a high frequency of G-to-A mutations in the lentiviral sample, which was not seen in the  $\gamma$ -retroviral sample. The  $\gamma$ -retroviral sample had very few mutations consisting of only pyrimidine to pyrimidine base substitutions. Variants in the samples with double capture hybridization probe enrichment were also assessed and showed a decrease in called variants in the  $\gamma$ -

retroviral sample and an increase in called variants in the lentiviral sample in comparison to the single capture samples (Table 2.4). While double capture may lead to greater sensitivity to detect rare sequence variants as shown in the lentiviral double capture sample, it can also potentially skew the variants due to unequal sequencing of samples, which can lead to losing rarer variants in the samples with lower VCN when pooled. Since the patient samples vary in VCN (Table 2.5) and will be pooled for the hybridization probe capture, we proceeded with a single capture to aim to prevent unequal sequencing and enrichment of samples with very low VCN.

The variants called in the endogenous *ADA* were then assessed to be used as a strategy to detect false mutations that were incorporated in the pipeline that were not caught by the error correction algorithm (i.e. DNA damage by thawing or sonication before UMI ligation). These variants in the exon and intronic regions detected were further analyzed in the database dbSNP<sup>15</sup> and all variants have been confirmed as either low-frequency somatic mutations or high-frequency SNPs seen in human populations (Table 2.6).

#### ADA-SCID Gene Therapy Treated Samples Variant Calling

Patient samples were obtained from clinical trials (#NCT00794508, #NCT01852071, #NCT02999984, #NCT01380990) and duplex library prepped, sequenced, and variant calling was performed (Figure 2.1C). Patient sample sequencing statistics are shown in Table 2.7 and 2.8. With this data set, mutation profiles comparing the  $\gamma$ -retroviral versus lentiviral treated groups as well as the mutational differences within the PBMCs and granulocytes were investigated.

Similarly to the CD3+ T-cell data, there was a significantly higher frequency of mutations in the PBMCs lentiviral transduced samples ( $p$ -value =  $1.2e-4$ ) with a significantly higher frequency of G to A mutations in PMBCs ( $p$ -value =  $9.9e-4$ ) and granulocytes ( $p$ -value = 0.02) (Table 2.9). These mutations are seen at low allele frequency. The mutation profile of the  $\gamma$ -retroviral patient samples did not show high frequencies of specific types of mutations, but there were slightly higher rates of purine-to-purine and pyrimidine-to-pyrimidine mutations than any other type of single base substitutions. There were also significantly more complex mutations in the  $\gamma$ -retroviral treated patients' granulocytes ( $p$ -value =  $5.3e-3$ ) in comparison to the lentiviral treated patients' granulocytes (Table 2.9).

The error rates were calculated for RT and APOBEC3 in the  $\gamma$ -retroviral and lentiviral patient PBMCs and granulocytes. The assumption was made that all G-to-A mutations were due to APOBEC3 and all other mutations would be due to RT. The error rate was calculated as the number of mutations called divided by the number of reads aligned to the *ADA* transgene. Due to the patient samples having varying VCNs, calculating the mutation rate per read normalizes for *ADA* coverage across patient samples with different *ADA* transgene sequencing depths. We calculated an error rate for all mutations (RT+APOBEC3) as well as an error rate for RT only and for APOBEC3 only (Figure 2.4). While error rates for (RT+APOBEC3) were not significantly different between  $\gamma$ -retroviral and lentiviral patients' granulocytes and PBMCs (Figure 2.4A, 2.4B), lentiviral RT had significantly lower mean error rates in both granulocytes and PBMCs ( $1.8e-4$ ,  $6.0e-5$ ) in comparison to  $\gamma$ -retroviral RT ( $3.8e-3$ ,  $8.0e-4$ ) (Figure 2.4C, 2.4D). APOBEC3 in the lentiviral patient samples had significantly higher mean error

rates in both granulocytes and PBMCs ( $4.0e-3$ ,  $2.0e-3$ ) in comparison to APOBEC3 in the 7retroviral patient samples ( $5.0e-4$ ,  $5.5e-4$ ) (Figure 2.4E, 2.4F).

The variants detected in the endogenous ADA gene in patient samples were then assessed (Table 2.10). Upon analysis, these were confirmed on the dbSNP database and some of the variants that display a high frequency in the human population were also found across multiple patient samples (Table 2.11). Furthermore, the sequencing analysis was able to detect the reported pathogenic mutations for ADA-SCID (Figure 2.5-2.12). On the other hand, our analysis also detected variants and deletions that were not reported on the dbSNP database (Table 2.12). Most of these are intron variants, which may suggest recent somatic mutations generated throughout the patient's lifetime or silent mutations from previous generations. Single nucleotide variants detected in the exons were confirmed to be synonymous protein mutations. However, our sequencing analysis detected large deletions that would be considered likely-pathogenic as they result in a truncated ADA protein. Some of the large deletions were identified in at least two patient samples. Since not all our findings were reported in databases, it is also possible that some of these mutations are results of sequencing errors. Overall, our results support the efficacy of our sequencing pipeline to recognize and confirm different variants in the patient's genome.

#### ADA-SCID Gene Therapy Treated Samples Common DNA Variants and motifs

We next were interested to see if any of the variants called were common across the different patient samples. Four groups were created (5 samples each consisting of the same viral treatment and cell type), and a high frequency of common variants were witnessed across the patient sample groups (Figure 2.13). Simulations were run to

estimate the probability of observing repeated mutations in the transgene. We calculated the average number of mutations across five patients in each condition and generated random permutations of the mutation profiles using the mean observed mutation rates for 100,000 rounds, assuming each position had an equal probability of mutating. Then, the chances of seeing any given mutations at given number of times was computed. Any recurrent mutations with a probability below 0.05 are determined to be highly unlikely and may indicate a possible hotspot for mutations. Mutations common to 3 and 4 samples were shown to be highly significant (i.e., unlikely by chance) in the MND-ADA granulocytes (p-value = 0.002, p-value = 1e-05) and PBMCs (p-value = 0.005, p-value= 4e-05). Mutations common to 4 samples were shown to be significant in the EFS-ADA granulocytes (p-value = 0.009). While our simulations did not calculate the likelihood of observing mutations common to 5 samples, these events can be inferred as similarly unlikely in MND-ADA granulocytes and PMBCs, as well as EFS-ADA granulocytes.

To assess our hypothesis that the G-to-A signature could be due to the APOBEC3 protein family, a substrate analysis on the common variants in the lentiviral patients was performed to see if there was an indication of known APOBEC3 substrate sites.<sup>16</sup> The substrate analysis showed a high frequency of mutations occurring at APOBEC3 substrate sites with a mean and standard deviation of  $86.12 \pm 4.89\%$  (Table 2.13). We also predicted which APOBEC3 family protein may have been responsible for the deamination based on the substrate signature with A3D and A3H sites with high occurrence in PBMCs (Figure 2.14).

## ADA-SCID Gene Therapy Samples Protein Analysis

We next analyzed how the DNA variants in each sample within the transgene would affect the coded ADA protein's amino acid sequence (Figure 2.15). Within viral treatment type, there was a similar frequency of synonymous, missense, nonsense, and complex/insertion/deletion when comparing PBMCs to granulocytes (p-values are nonsignificant). Comparing complex/insertion/deletion, we observed an increased number of mutations in granulocytes by  $\gamma$ -retrovirus (p-value =  $2e-2$ ). Comparing the lentiviral treated samples to the  $\gamma$ -retroviral treated samples, there was a decrease in missense in granulocytes (p-value =  $3.1e-2$ ) and an increase in complex/insertion/deletion (p-value= $1.4e-4$ ) in the  $\gamma$ -retroviral treated samples in comparison to the lentiviral samples. In PBMCs, we observed an increased mutation rate in nonsense (p-value =  $5.3e-3$ ) and a decrease in frameshift (p-value =  $3.5e-2$ ) in lentiviral-treated samples.

## DISCUSSION

While fidelity and mutational profiles of RTs (HIV-1 and M-MLV) have been previously studied, there is little research in RT fidelity in retroviral and lentiviral gene therapy. Gene therapies that utilize these vectors may be introducing unwanted genetic changes to the incorporated transgene due to errors from RT. While there are some genes where incorporation of a disruptive mutation will leave the protein non-functional like ADA, there are many other diseases and malignancies where the incorporation of the mutation can lead to oncogenesis. For example, an essential cytokine receptor for T cell development, can cause SCID when defective and is thus an important target for gene therapy but, activating mutations of *IL7R $\alpha$*  may be driver mutations in T-cell leukemias.<sup>17</sup> With this being the case, it is important to understand mutational profiles to make sure that retroviral and lentiviral approaches are safe for gene therapy for certain diseases and malignancies.

With unique access to ADA-SCID gene therapy patients' PMBCs and granulocytes, we performed duplex sequencing of the *ADA* cDNA transgene as well as the endogenous *ADA* gene to better understand the mutational landscapes in gene therapy utilizing retroviral and lentiviral delivery. Deep sequencing was performed which produced high raw coverage in the targeted regions, which allowed for the utilization of a stringent error correction bioinformatic protocol to remove any false mutations incorporated into the pipeline. After sequencing, data processing, and variant calling, the data illuminated a significantly higher mutation signature of G-to-A mutations in the lentiviral vector treated T-cells and patient samples. We hypothesized that this mutational signature may not be due to RT but to the APOBEC3 protein family. APOBEC3, which is a cytidine

deaminase that is part of the innate immune response that causes mutations in viral RNA or ssDNA during reverse transcription to prevent viral replication and infectivity.<sup>18,19</sup> By deaminating cytosines, which are then converted to uracil or thymine, APOBEC will lead to a DNA guanine to adenosine transitional mutations.

The fidelity of RT in both  $\gamma$ -retroviral and lentiviral-treated patient cells was next calculated. While it has been shown that MLV RT typically has a higher fidelity than HIV-1 RT<sup>20</sup>, our patient dataset showed that HIV-1 RT exhibited a higher fidelity. Some explanations for this could be the assumption made for the calculation that all G-to-A mutations were due to APOBEC3 when some may have been caused by RT. There also may have been mutations missed due to selective advantage in our PBMC population. For example, if a cell received a single *ADA* transgene (VCN =1) with a mutation deeming it non-functional, that cell would not survive and therefore those mutations would be missed in our analysis leading to a higher fidelity.

We also investigated if similar mutations were present across the different patient samples or if mutations occurred at random. While the  $\gamma$ -retroviral samples had far fewer mutations than the lentiviral samples, recurrent mutations across samples were observed at multiple sites, which is unlikely to be due just to chance but may represent “hot spots” for RT errors or APOBEC activity. A high frequency of common mutations in the lentiviral samples was also witnessed.

The effects of the alternate allele on the coded amino acid of the *ADA* protein were analyzed. We had hypothesized that we may see a higher proportion of synonymous protein mutations in the PMBCs than in the granulocytes in both the lentiviral and  $\gamma$ -retroviral samples because the lymphocytes (T, B, NK cells) among the PBMC fraction

need a functional copy of *ADA* to survive, while granulocytes do not, based on the observed strong selective advantage for *ADA* gene replete lymphocytes after gene therapy, but not for granulocytes<sup>2</sup>.

To our surprise we did not see a higher frequency of synonymous mutations in the PBMCs when compared to granulocytes within viral treated groups. One explanation for this is that the cells containing a missense, nonsense, or frameshift/complex mutation in the incorporated transgene, may also have an additional functional copy of the *ADA* transgene with no mutations incorporated. There only needs to be one functional copy of the *ADA* gene within the lymphoid cells for normal immune function as seen in heterozygous carriers who are asymptomatic.

A DNA substrate analysis of the G-to-A mutations in the lentiviral samples was performed to assess if there were any known APOBEC3 substrate sites at these mutation locations. APOBEC3 substrate sequences<sup>16</sup> were utilized and confirmed a high frequency of  $86.12 \pm 4.89\%$  across all lentiviral samples, showing that the majority of G to A mutations contained an APOBEC3 substrate sequence. Although further experimentation is necessary to validate this bioinformatic analysis, this data gives support to our hypothesis that the G-to-A mutational signature is due to APOBEC3 deamination.

The endogenous *ADA* gene was sequenced to define the baseline error rate in our pipeline for false mutations that were not corrected with the error correction protocol. We should only see the known pathogenic *ADA* mutations in the exons of each patients' samples that were their cause for ADA-SCID; all other exonic mutations could be used to create a baseline error rate. What we came to realize is that due to the hybridization

capture, we sequenced exon and parts of the intronic regions leading to the detection of low frequency alleles. We investigated these alternate alleles (majority in intronic regions) and saw that most of these alleles have been cited and documented in global populations and therefore, concluded that these could be true alleles carried by these individuals or low frequency somatic mutations. In addition, common variants that were identified in multiple patient samples further suggest that these are more true variant alleles than sequencing errors. We anticipate the new variants and large deletions that were not reported on any database can either be low frequency somatic mutations or sequencing errors. Through the detected current and potentially new variants that were identified in at least two patient samples, this established sequencing pipeline can distinguish changes in both the patient's genome and transgene.

We also questioned why we were not seeing APOBEC3 mediated restriction in our  $\gamma$ -retroviral samples. Both viruses have evolved methods to counteract APOBEC3 to better preserve their genome integrity. MLV has two mechanisms to inhibit APOBEC3 activity that could be the reasons as to why we see minimal G to A mutations in our data. MLV produces an alternatively spliced P50 protein from the *gag* RNA which prevents APOBEC3 from being packaged into the viral particles.<sup>21</sup> MLV also produces a glycosylated Gag (glycoGag) protein which prevents APOBEC3 interaction with the reverse transcription complex due to increasing stability of the viral core.<sup>21</sup> HIV-1 expresses the accessory protein VIF that counteracts the activity of APOBEC3, but standard lentiviral vector packaging systems have excluded the accessory proteins including VIF as they were not thought to be essential for vector production and it is important to minimize HIV-1 components in the vector production systems for biosafety.

Because APOBEC3-derived mutations were the predominant error observed in lentiviral vector proviruses, efforts to reduce APOBEC3 activity could produce more sequence consistency for clinical lentiviral vectors. Approaches to explore are including a VIF-expression plasmid during packaging or deleting the APOBEC3 gene complex from the packaging cells.

## MATERIALS AND METHODS

### *Vector Production*

The EFS-ADA and MND-ADA vector cloning design have been described previously.<sup>13</sup> HEK293T *PKR* knockout cells<sup>22</sup> were plated on 10cm plates at a density of  $1 \times 10^7$  cells/mL and vectors were packaged through transient transfection. EFS-ADA was packaged using third generation lentiviral packaging plasmids using the VSV-G protein envelope. Raw viral supernatant was collected 3 days post transfection and concentrated through ultracentrifugation. MND-ADA was packaged using MLV-based packaging plasmids<sup>23</sup> and VSV-G protein. Raw viral supernatant was collected 3 days post transfection and concentrated using Retro-X Concentrator (Takara Bio Inc., Shiga, Japan).

### *Healthy Donor T-Cell (CD3+) Transduction and Culturing*

Healthy donor PBMCs were acquired from the UCLA Virology Core. T-cells (CD3+) were selected from PBMCs using the CD3 Microbead kit (Miltenyi Biotec, Auburn, CA, USA). CD3+ cells were plated at  $1.5 \times 10^6$  cells/mL and stimulated with Immunocult CD3/CD28 activating cocktail (STEMCELL Technologies, Vancouver, Canada). 72 hours post-activation cells were plated on retronectin-coated plates at a concentration of  $1 \times 10^6$  cells/mL in X-Vivo 15 medium (Lonza, Basel, Switzerland) supplemented with 1x glutamine, penicillin, and streptomycin (Gemini Bio-Products, Sacramento, CA, USA), 1x Glutamax (Thermo Fisher Scientific, Irwindale, CA, USA), 5% Human AB Serum (Gemini Bio-Products, Sacramento, CA, USA), 100U/mL Human Interleukin-2 (Peprotech, Rocky Hill, NJ, USA) and transduced with ascending amounts of vector

(1ul, 3ul, 5ul, 10ul, 30ul, 50ul) and transduction enhancers Poloxamer 338 (1mg/mL) (BASF, Ludwigshafen, Germany) and Prostaglandin E2 (10uM) (Millipore Sigma, Burlington, MA, USA). Cells were cultured for two weeks and collected on day 14.

#### *Vector Copy Number Analysis*

VCN was determined from extracted genomic DNA for samples transduced with EFS-ADA by digital droplet polymerase chain reaction (ddPCR) assay using primers and probes for HIV-1 PSI (forward 5'-AAGTAGTGTGTGCCCGTCTG-3', reverse 5'-CCTCTGGTTTCCCTTTCGCT-3', 5'-56-FAM-AGCTCTCTC-ZEN-GACGCAGGACTCGGC-3IABkFQ-3'). VCN was determined for samples transduced with MND-ADA by ddPCR assay using primers and probes for Human ADA cDNA (forward 5'-GGTCCATCCTGTGCTGCAT-3', reverse 5'-CGGTCTGCTGCTGGTACTTCTT-3', 5'-56-FAM-CCAGCCCAA-ZEN-CTGGTCCCCCAAG-3IABkFQ-3'). The Human Syndecan 4 gene (SCD4) as a reference (forward 5'-CAGGGTCTGGGAGCCAAGT-3', reverse 5'-GCACAGTGCTGGACATTGACA-3', 5'-5HEX-CCCACCGAA-ZEN-CCCAAGAACTAGAGGAGAAT-3IABkFQ-3') for all samples.

#### *Library Preparation, Target Hybridization, and DNA Sequencing*

DNA libraries were prepared using the xGen cfDNA & FFPE DNA Library Prep v2 MC Kit (Integrated DNA Technologies, San Diego, CA, USA). For each sample, 250 ng of genomic DNA were processed. All DNA samples were sheared using the M220 Focused-ultrasonicated (Covaris) to create fragments with an average size of 250 base pairs. The end repair, adapter ligation and PCR amplification were performed according

to the manufacturer's manual. The adapters contain 8-bp unique molecular identifiers (UMIs) with optimized and fixed sequences (32). Because fixed sequences were used, any errors generated from sequencing or PCR can still be corrected during the analysis. Barcodes were introduced during PCR amplification (6 cycles) using xGen unique Dual Index (UDI) primer pairs (Integrated DNA Technologies, San Diego, CA, USA). For the enrichment of the library, a custom panel of probes was designed to target both codon-optimized and wild-type sequence of the ADA gene. Hybridization capture was performed with xGen Hybridization and Wash v2 Kit (Integrated DNA Technologies, San Diego, CA, USA), xGen Universal Blockers TS (Integrated DNA Technologies, San Diego, CA, USA), and xGen Library Amplification Primer Mix (Integrated DNA Technologies, San Diego, CA, USA), with an input of (1) ug of DNA library for each capture. Duplex Sequencing was performed with the Illumina NovaSeq X Plus (2x150).

### *Variant Calling*

The Integrated DNA technologies xGen cfDNA & FFPE DNA Library Prep Kit "Processing sequence data with unique molecular identifiers analysis guidelines" (xGen cfDNA & FFPE DNA Library Prep v2 MC Kit, #10010206, IDT, San Diego, CA, USA) was used for data processing and variant calling. Unmapped BAM files were constructed from demultiplexed FASTQ files using Picard<sup>24</sup> version 3.0 UMIs were extracted using fgbio version 2.2.1 with a "--read-structure=8M142T 8M142T". Reads were aligned using bwa<sup>25</sup> version 0.7.18 (r1243) to the reference genome hg38 and either MND-ADA or EFS-ADA reference sequences. We selected the collapsing combined read families error correction method. VarDictJava<sup>26</sup> version 1.8.3 was used

to call variants with an allele frequency of 0.001. All other commands and parameters were based on suggested parameters from the analysis guidelines.

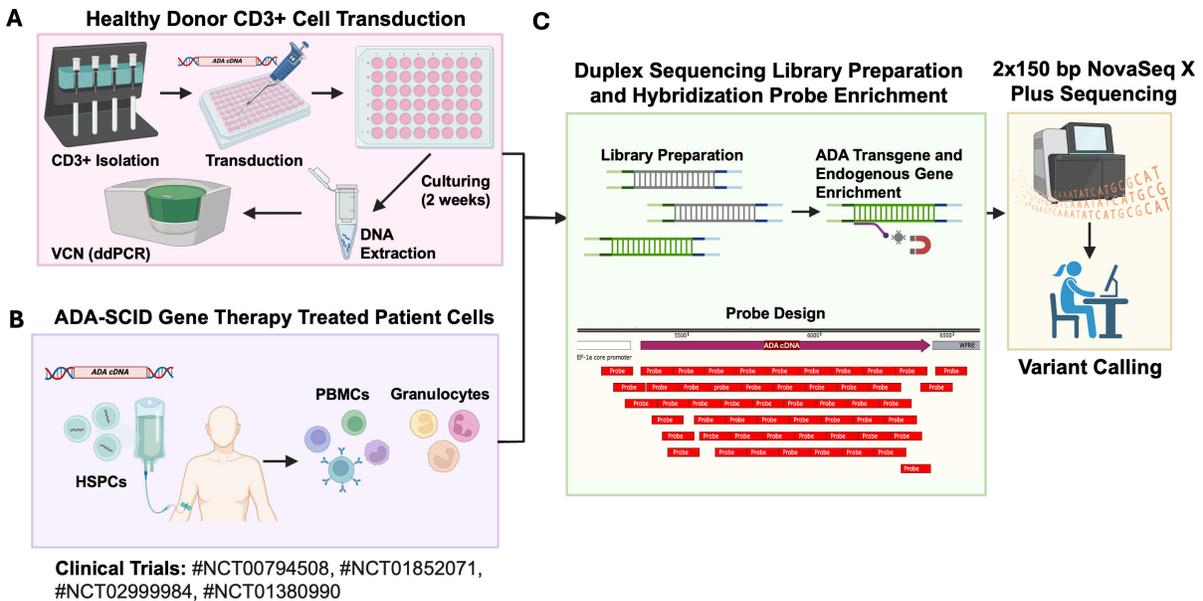
### *APOBEC3 Substrate Analysis*

APOBEC3 substrates were obtained from Maiti et al.<sup>16</sup> and G to A mutations were analyzed by comparing DNA sequence at the G-to-A site with APOBEC3 substrate sequences.

### *Statistics*

A t-test was applied to the patient variant callset to compare PMBCs and granulocytes, as well as  $\gamma$ -retroviral and lentiviral treated samples. To assess the likelihood of observing variants common to multiple samples, we ran simulations to estimate the probability of observing repeated mutations in the transgene. We calculated the average number of mutations across five patients in each condition and randomly generated mutations using the empirical mutation rates for 100,000 rounds, assuming each position had an equal probability of mutating. Then, we computed the chances of seeing any given mutations a given number of times. Any sites with a given number of mutations are determined to be highly unlikely if the simulated rates are under 0.05. Wilcoxon test was used in comparing the means of error rates between groups. This non-parametric test does not assume an underlying statistical distribution and a p-value below 0.05 shows that the error rate is significantly higher in one group than the other.

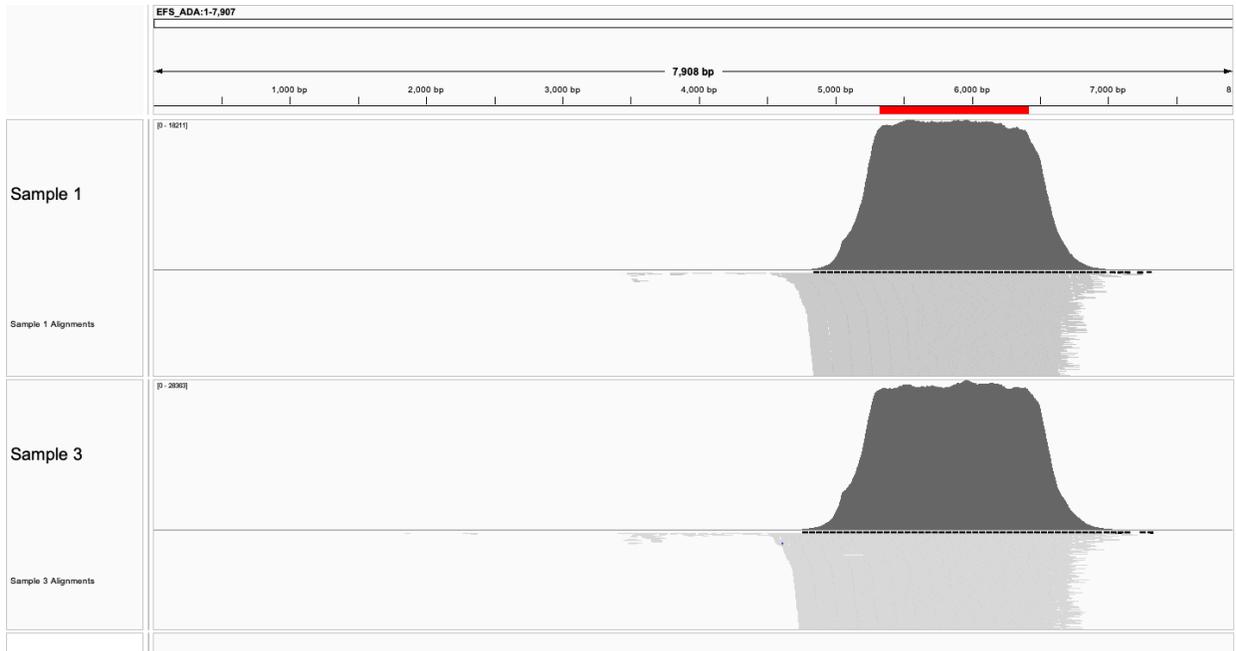
## FIGURES



**Figure 2.1**

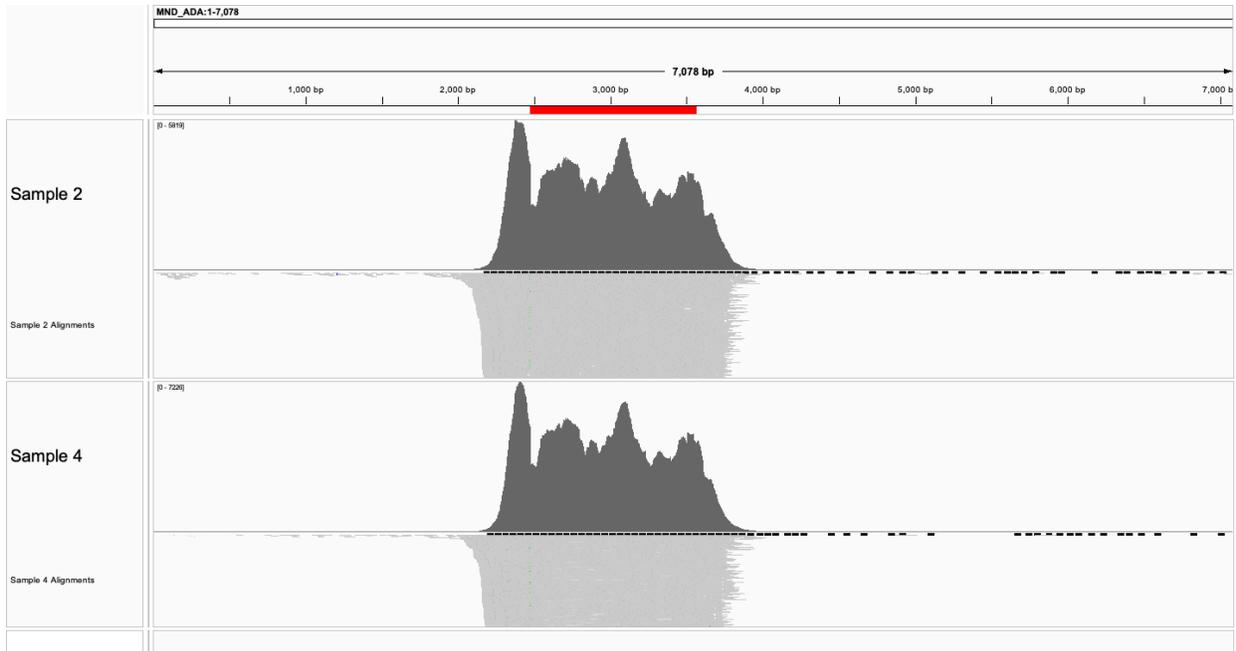
Schematic of project protocol and duplex sequencing pipeline. (A) CD3+ cells were isolated from healthy donor PBMCs and transduced at increasing viral volumes with either MND-ADA  $\gamma$ -retrovirus or EFS-ADA lentivirus. CD3+ cells were cultured for two weeks in T-cell culture conditions. VCN of transduced cells were determined by ddPCR. (B) ADA-SCID gene therapy patient PBMCs and granulocytes were acquired from clinical trials #NCT00794508, #NCT01852071, #NCT02999984, #NCT01380990. (C) Duplex sequencing library preparation protocol was performed on CD3+ and patient sample DNA. The ADA cDNA transgene and ADA endogenous gene were enriched through hybridization probe capture. Biotinylated probes were designed across the MND-ADA cDNA, EFS-ADA codon optimized cDNA, and across endogenous ADA exons. 2x150 paired end NovaSeq X Plus Sequencing was performed on the enriched

DNA library. FASTQ files were error corrected and aligned using Burrow-Wheeler Aligner (BWA) to the human genome (GRCh38) and MND-ADA and EFS-ADA vector sequences. Variant calling was performed with VarDictJava at an allele frequency of 0.001.



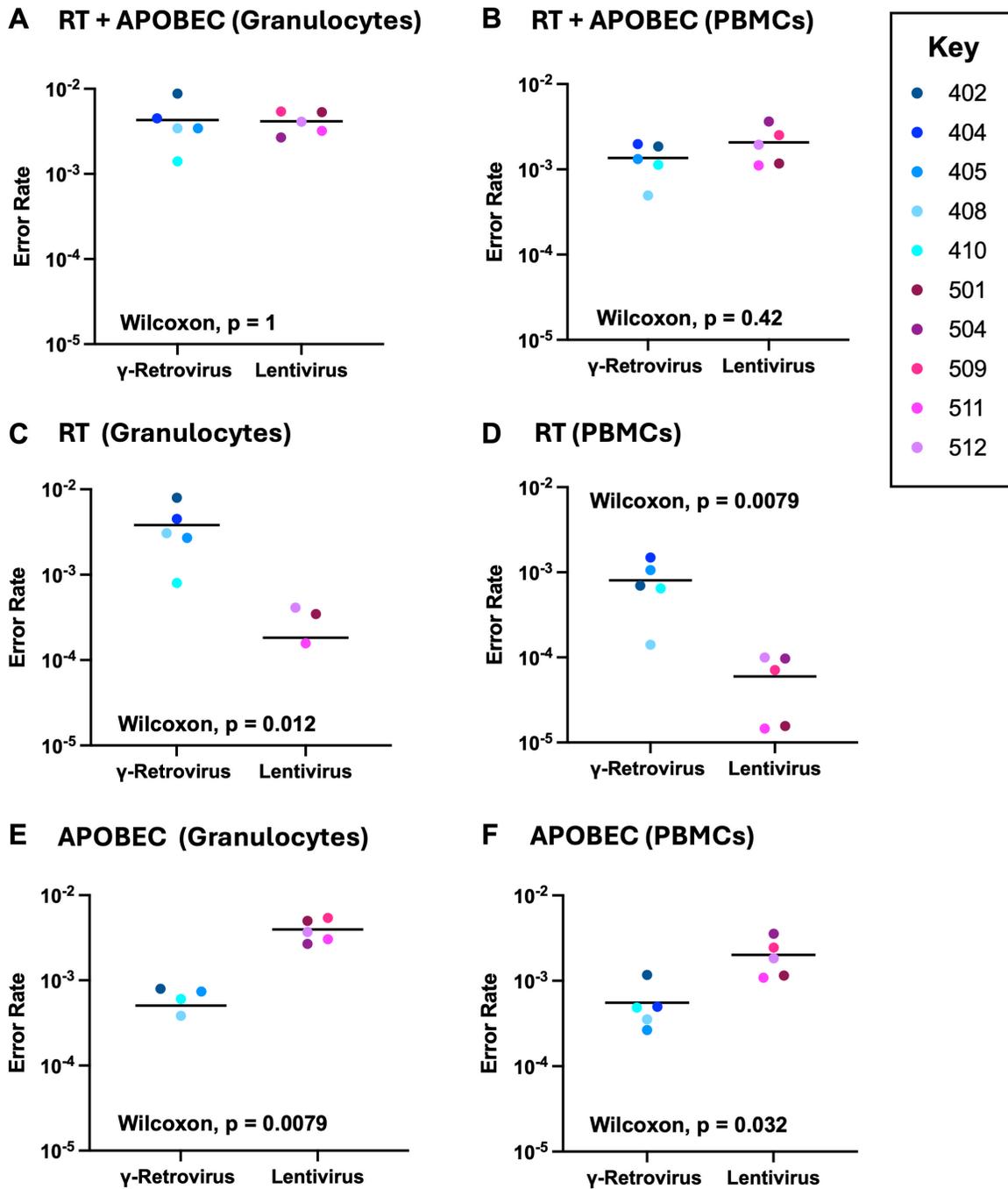
**Figure 2.2**

Integrative Genomics Viewer image of sample 1 and 3 alignments and coverage distribution of the EFS-ADA transgene. The red bar indicates the location of the ADA cDNA showing high coverage and uniform distribution across the cDNA.



**Figure 2.3**

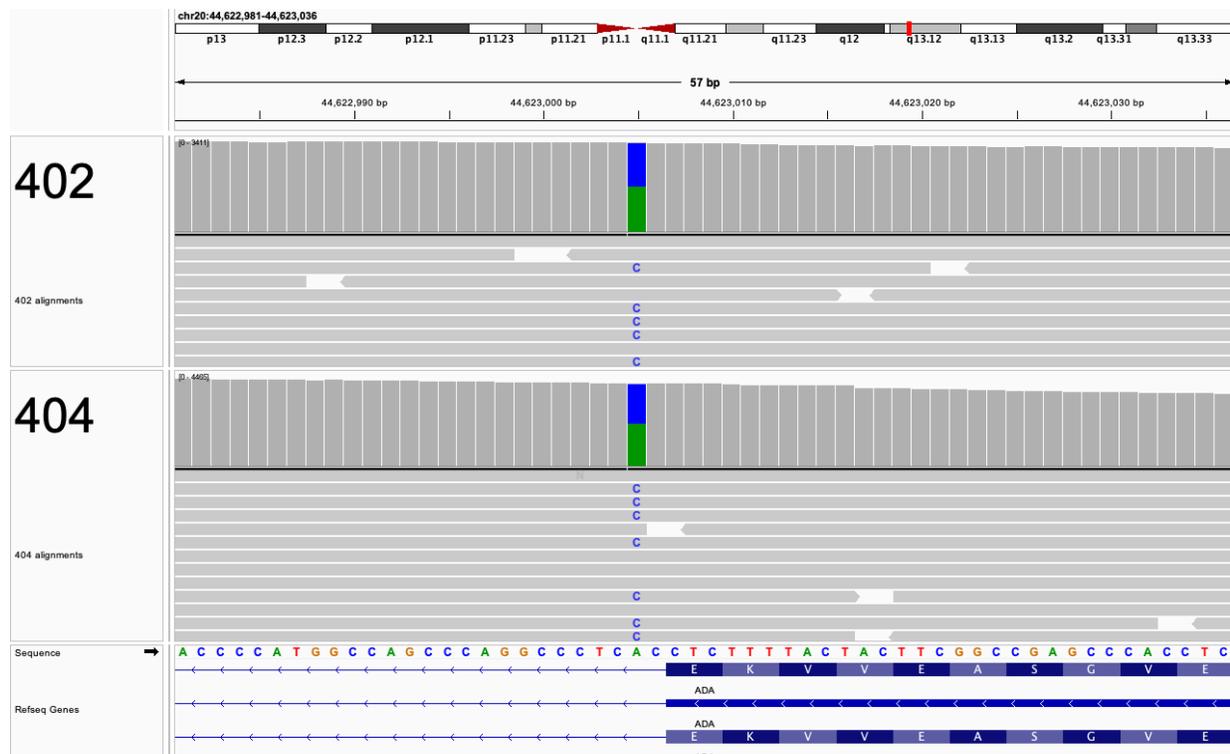
Integrative Genomics Viewer image of sample 2 and 4 alignments and coverage distribution of the MND-ADA transgene. The red bar indicates the location of the ADA cDNA showing high coverage across the cDNA.



**Figure 2.4**

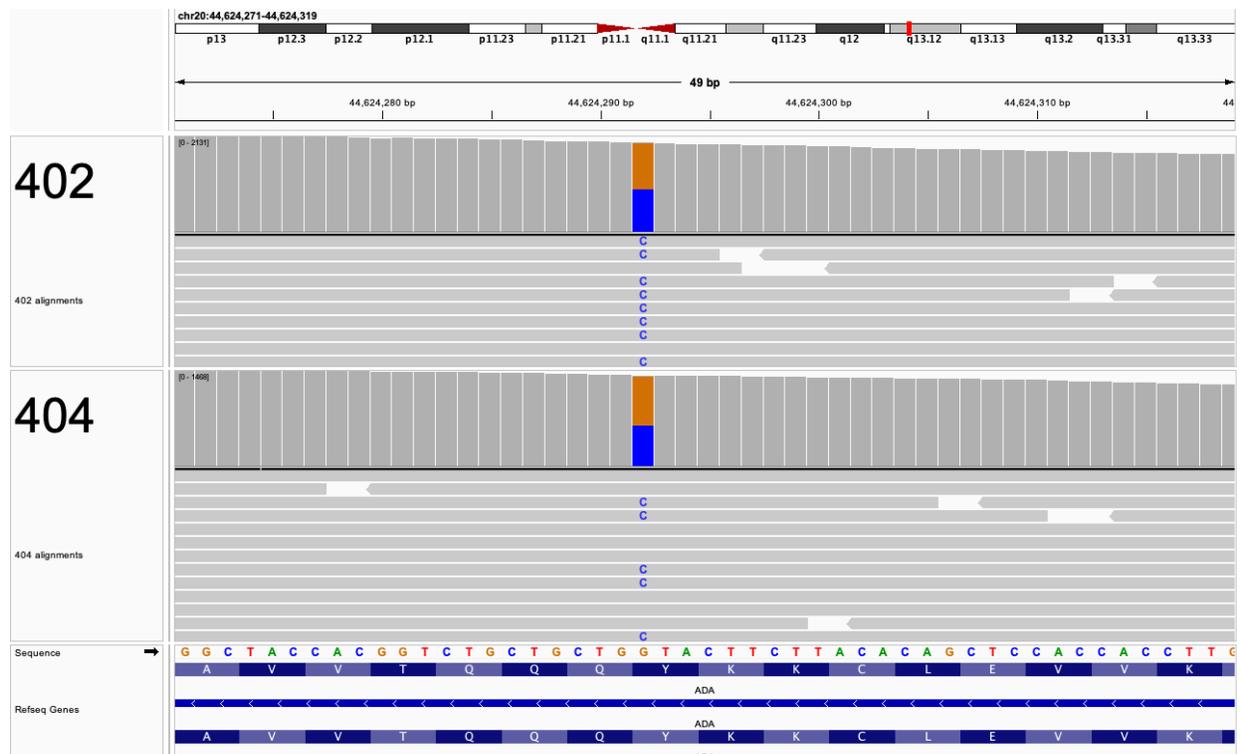
Error rates were calculated as the number of mutations called divided by the number of reads aligned to the ADA transgene post error correction, read collapsing, and filtering,

per patient sample. All G-to-A mutations were assumed to be caused by APOBEC3 and all other called mutations were deemed as RT mutations. (A) Error rate in granulocytes calculated with all mutations (RT + APOBEC3). (B) Error rate in PBMCs calculated with all mutations (RT + APOBEC3). (C) Error rate of RT in granulocytes. (D) Error rate of RT in PBMCs. (E) Error rate of APOBEC3 in granulocytes. (F) Error rate of APOBEC3 in PBMCs. Patient samples with no mutations in one of the categories (RT or APOBEC3) have no data point on the plot. A Wilcoxon rank sum test was applied to compare the mean error rate between  $\gamma$ -retroviral and lentiviral treated samples (n=5).



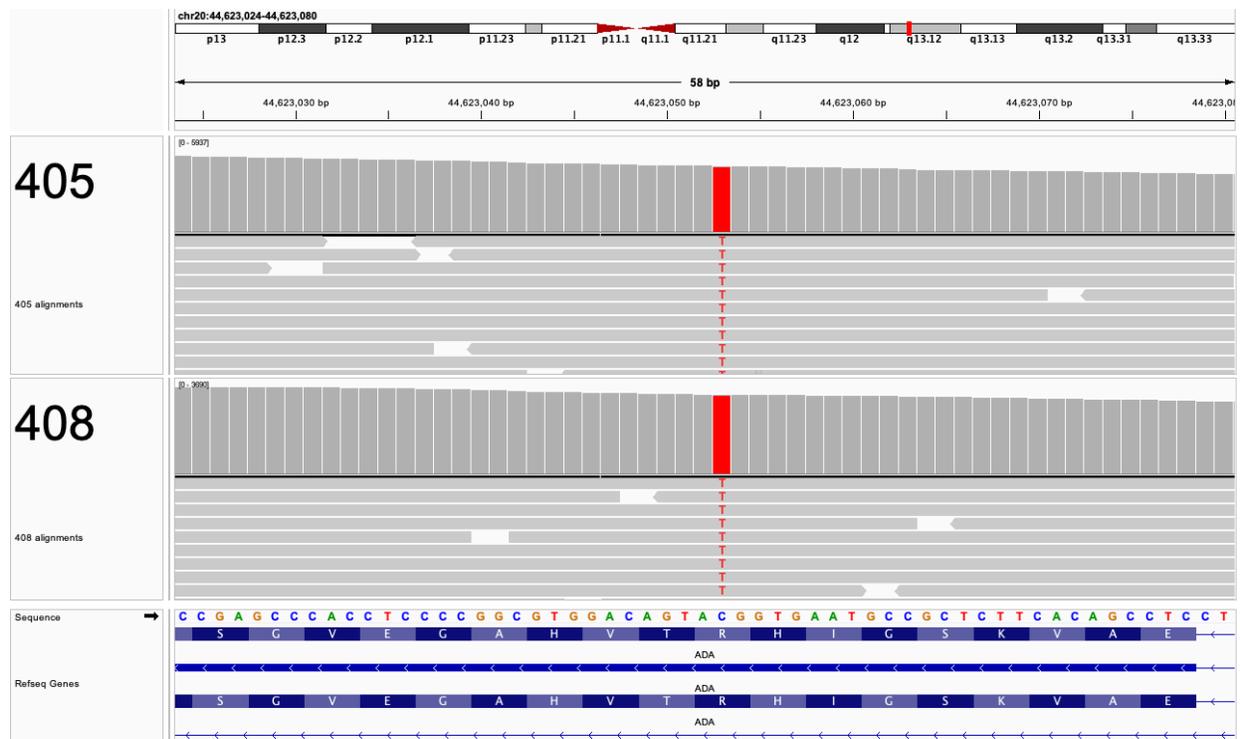
**Figure 2.5**

Integrative genomics viewer visualization of patients 402 and 404 ADA-SCID Mutations (T+2>G in IVS7).



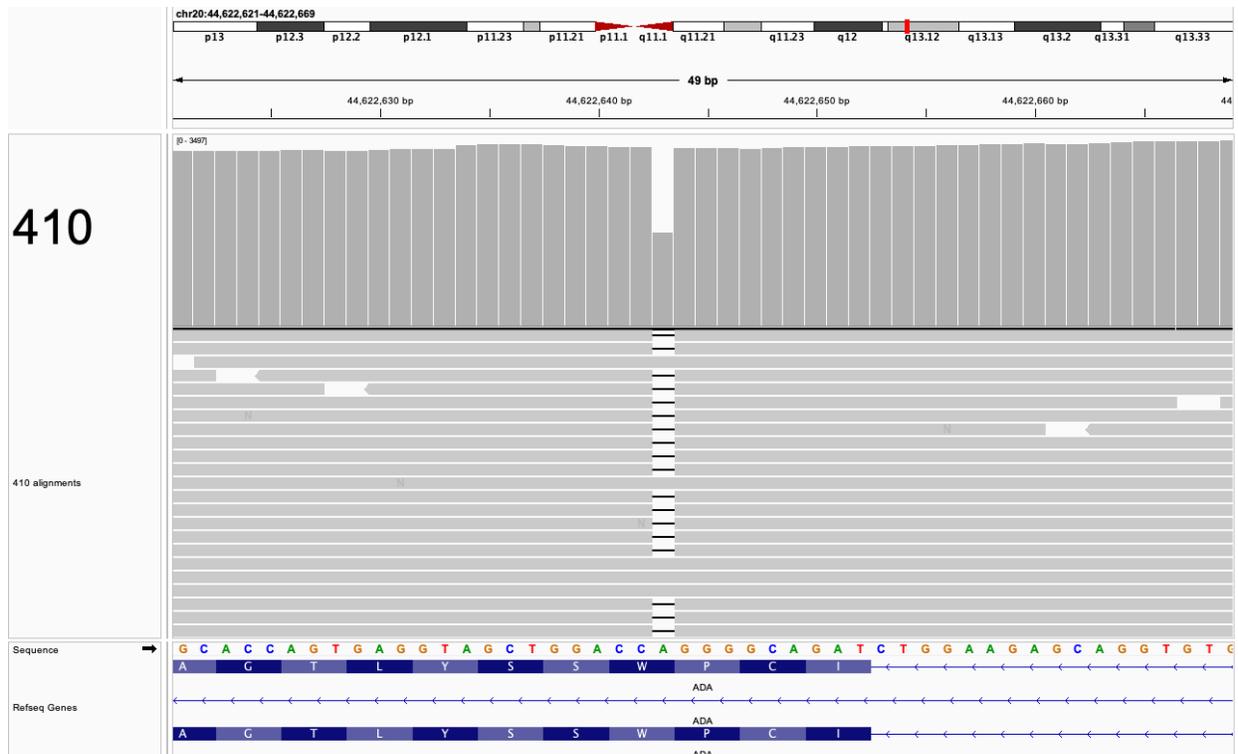
**Figure 2.6**

Integrative genomics viewer visualization of patients 402 and 404 ADA-SCID Mutations (Y172STOP).



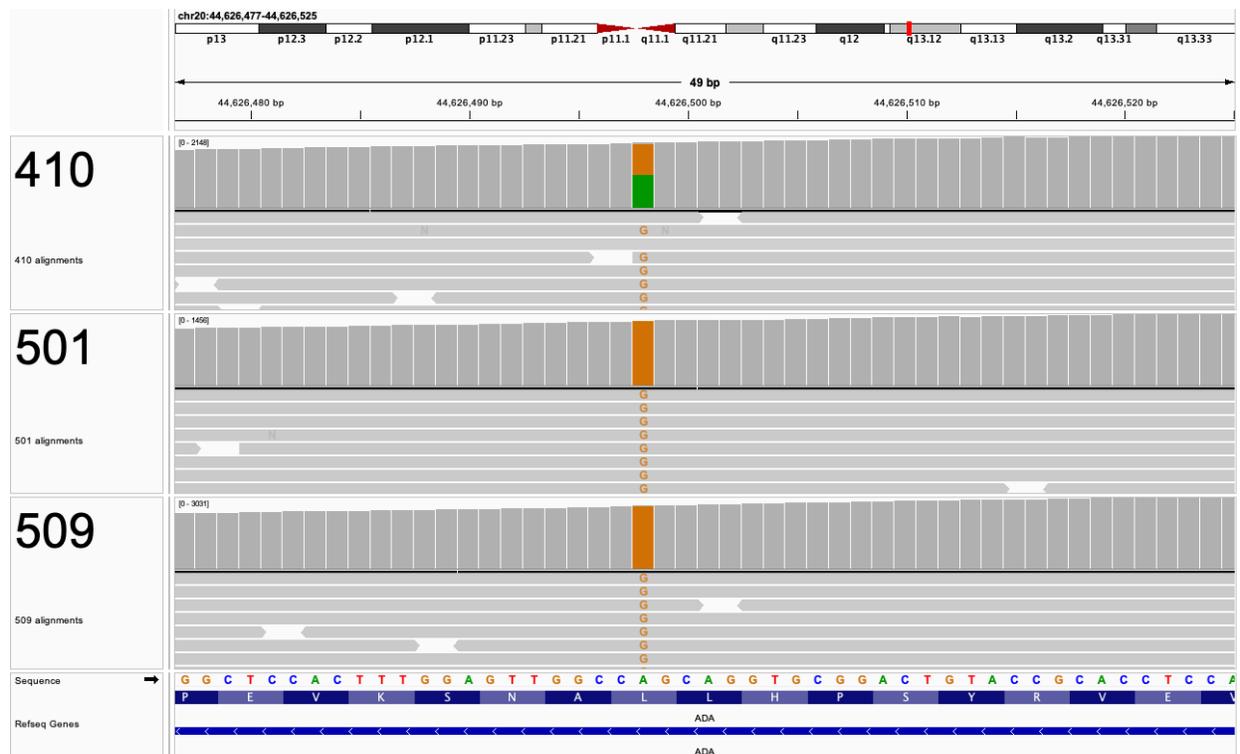
**Figure 2.7**

Integrative genomics viewer visualization of patients 405 and 408 ADA-SCID Mutations (R211H).



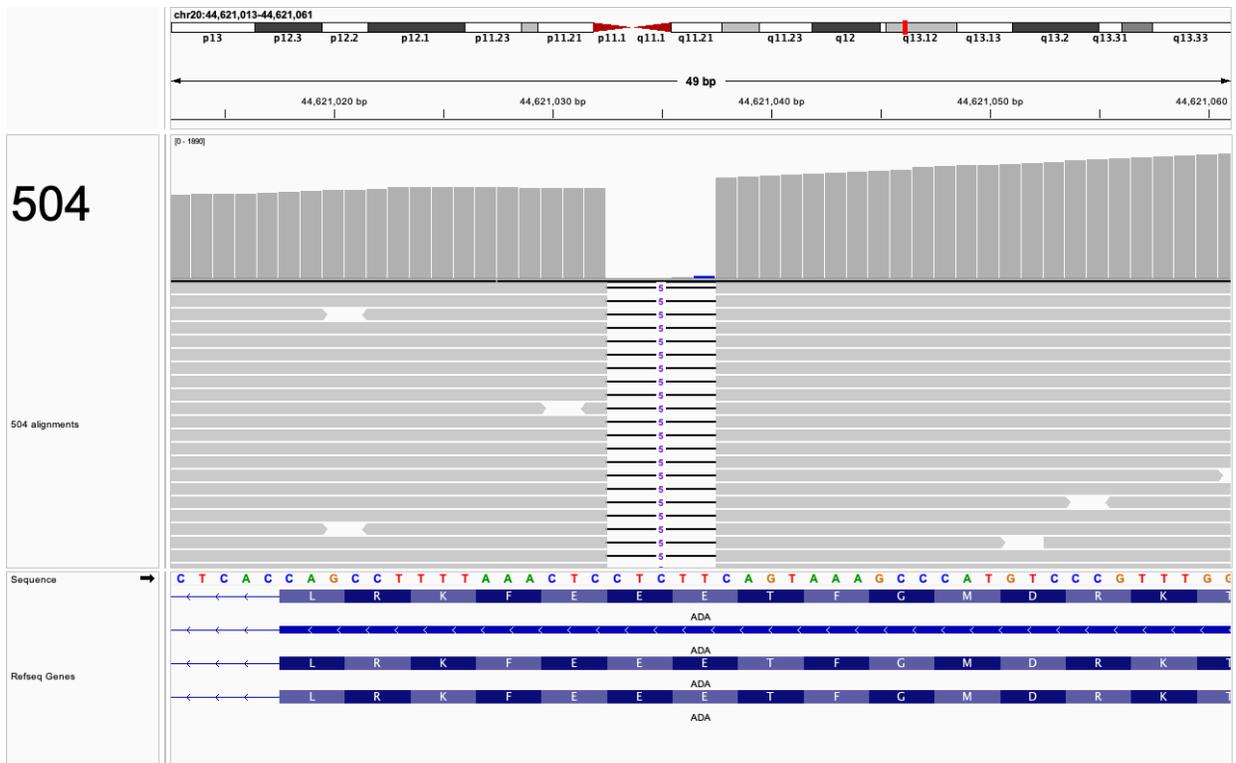
**Figure 2.8**

Integrative genomics viewer visualization of patients 405 and 408 ADA-SCID Mutation (c.790delT).



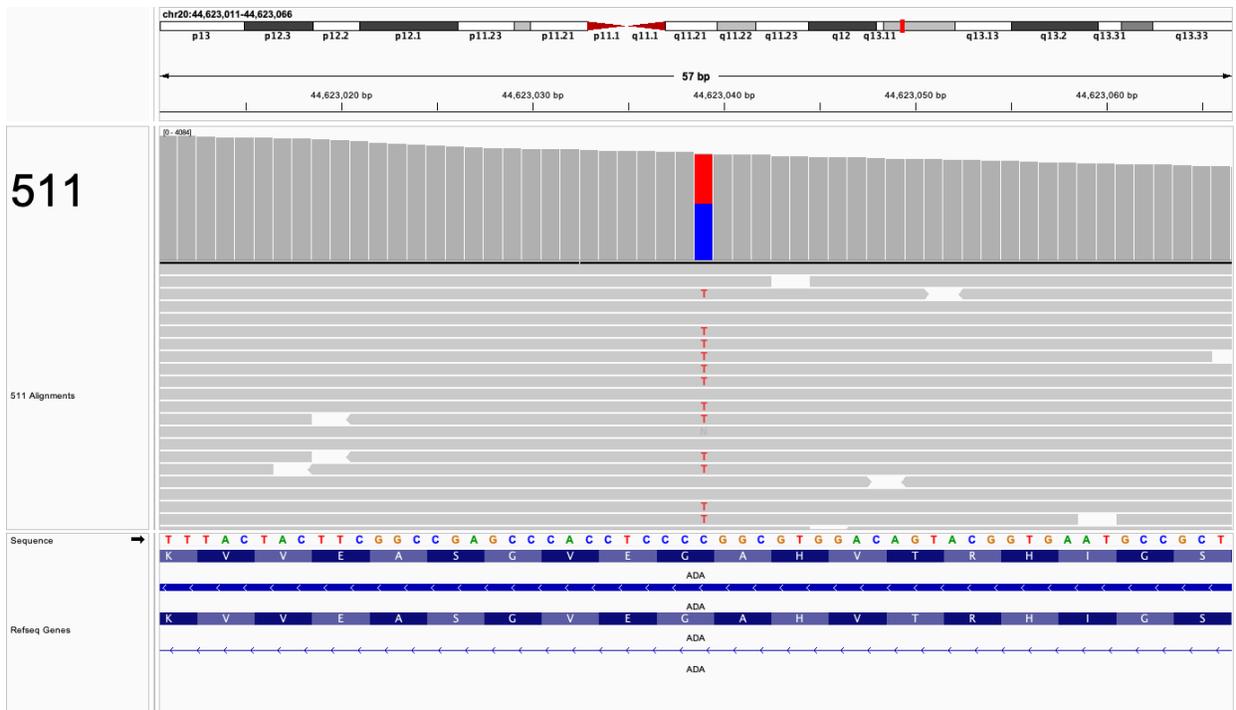
**Figure 2.9**

Integrative genomics viewer visualization of patients 410, 501, and 509 ADA-SCID Mutations (L107P)



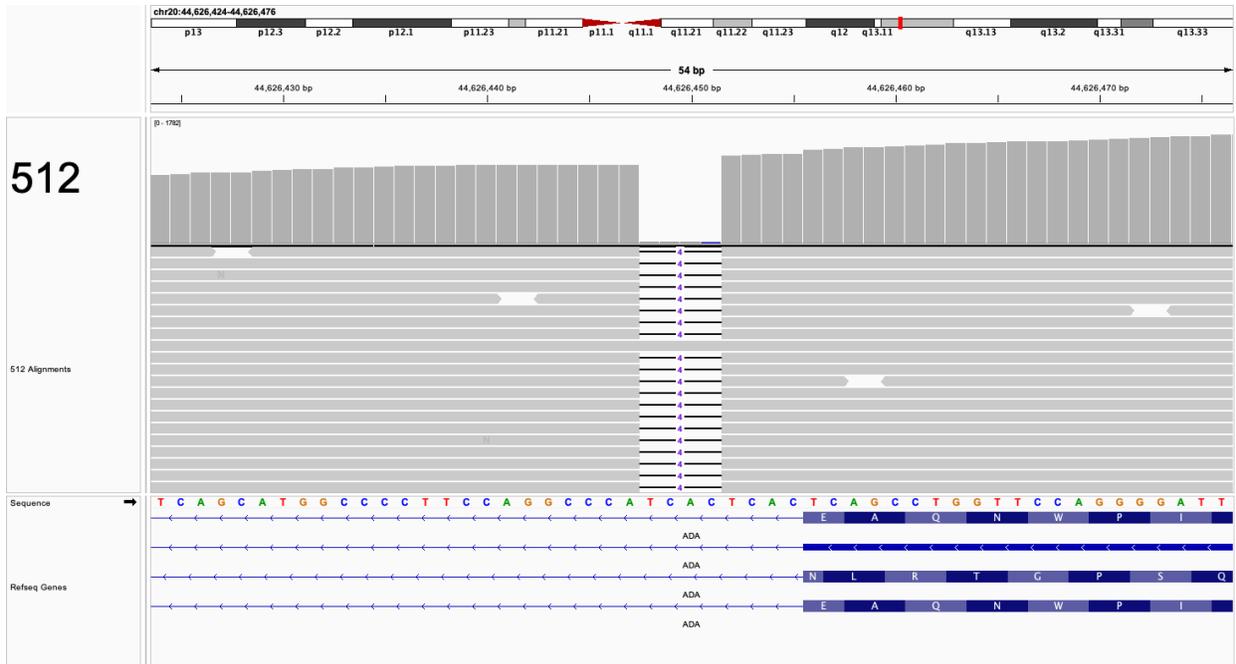
**Figure 2.10**

Integrative genomics viewer visualization of patient 504 mutations (delc.955-959).



**Figure 2.11**

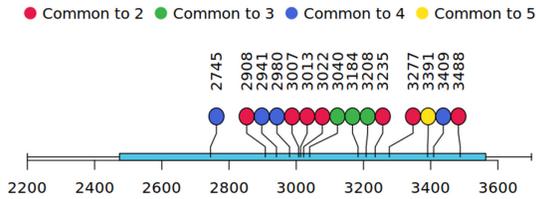
Integrative genomics viewer visualization of patient 411 ADA-SCID Mutation (G216R).



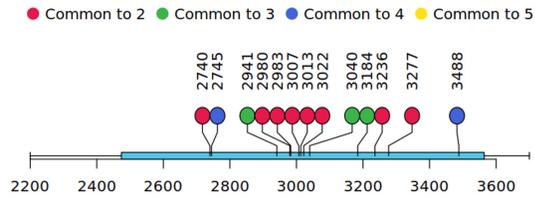
**Figure 2.12**

Integrative genomics viewer visualization of patient 512 ADA-SCID Mutations (c.360-362+1delTGAG)

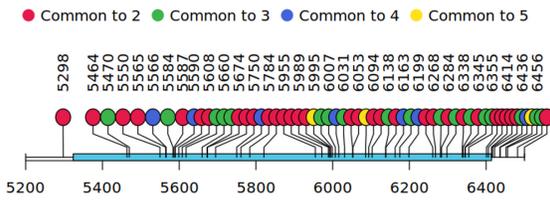
### A MND-ADA Granulocytes



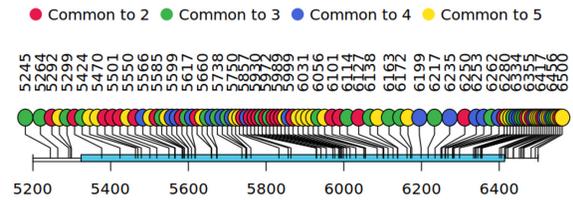
### B MND-ADA PBMCs



### C EFS-ADA Granulocytes

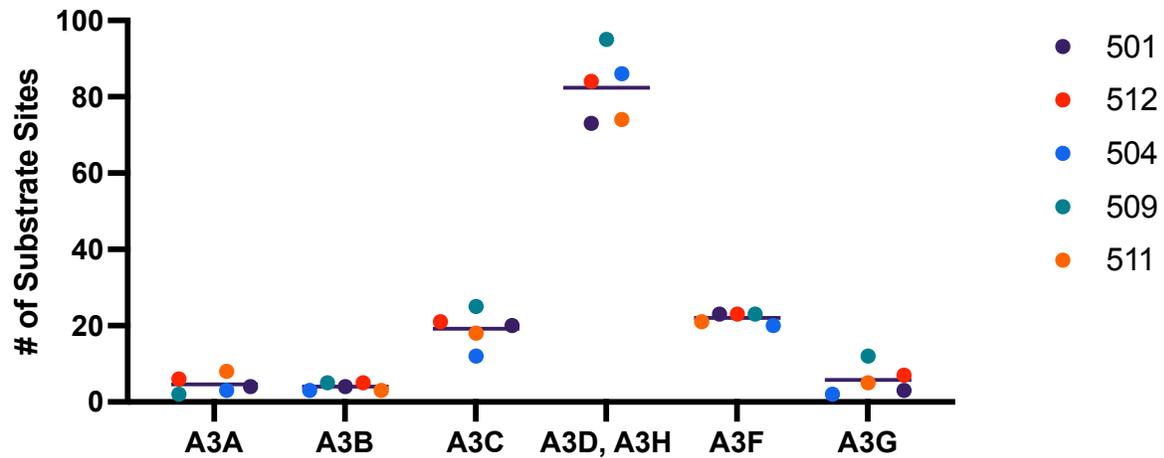


### D EFS-ADA PBMCs



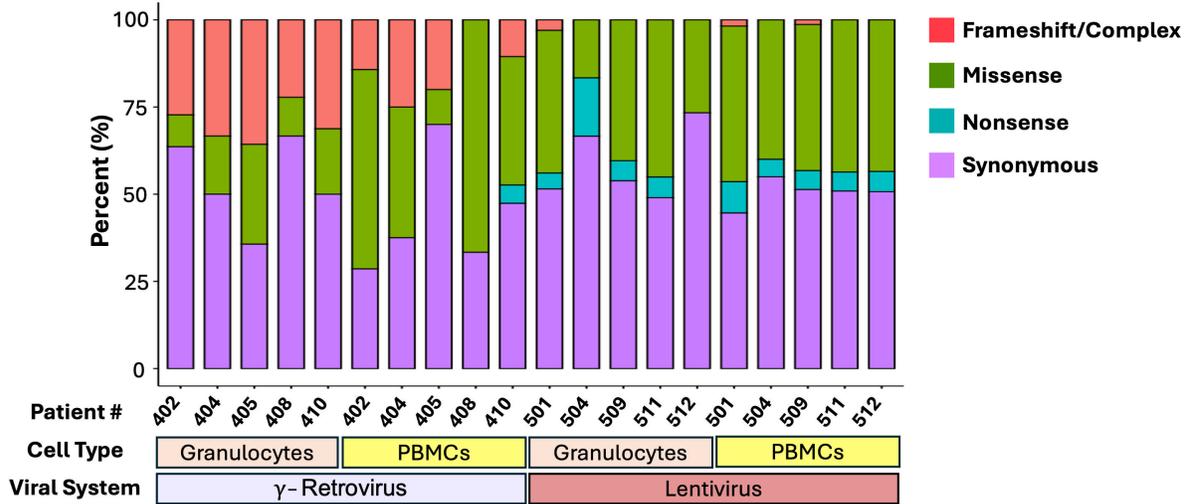
**Figure 2.13**

VardictJava was utilized to identify variants (allele frequency = 0.001) that were generated by gamma-retrovirus and lentivirus reverse transcriptase. Common mutations were observed among DNA samples from (A) Granulocytes and (B) PBMCs from patients who received gamma-retrovirus gene therapy. (C) Granulocytes and (D) PBMCs from patients who received lentivirus gene therapy.



**Figure 2.14**

APOBEC3 Family Protein Prediction in PBMCs. APOBEC3 substrate sites were analyzed to assess the DNA sequence at the mutation sites. A3D and A3H have the same substrate site and same number of calls so were grouped together.



**Figure 2.15**

Characterization of *ADA* mutations based on encoded amino acid outcome. Patient identification numbers in the 400s received  $\gamma$ -retrovirus gene therapy. Patient identification numbers in the 500s received lentivirus gene therapy. Only DNA mutations within the coordinates of *ADA* cDNA were analyzed. Percent was normalized by total mutations of each sample and called as either synonymous, nonsense, missense, or frameshift/complex.

TABLES

**Table 2.1:** Transduced CD3+ DNA Sequencing Statistics

Vector	Read Total	Total Primary Alignments	Total Alignments to Transgene	Total Alignments to Endogenous ADA	% On Target Alignments	ADA cDNA Mean Coverage
Sequencing Statistics Pre-Error Correction, Read Collapsing, and Filtering						
MND-ADA	863,334,033	862,560,476	88,601,959	173,968,486	30.44%	9.89e+05x
EFS-ADA	838,838,523	838,166,931	197,672,540	150,835,681	41.58%	9.92e+05x
Sequencing Statistics Post Error Correction, Read Collapsing, and Filtering						
MND-ADA	314,729	314,650	74,371	223,273	94.60%	3.35e+03x
EFS-ADA	439,440	439,439	231,324	195,670	97.17%	1.47e+04x

**Table 2.2:** Transduced CD3+ DNA Sequencing Statistics (Single and Double Hybridization Probe Capture Comparison)

Sample #	Vector	Capture	Data (Gb)	Read Total	Total Primary Alignments	Total Alignments to Transgene	Total Alignments to Endogenous ADA	% On Target Alignments	ADA cDNA Mean Coverage
Sequencing Statistics Pre Error-Correction, Read Collapsing, and Filtering									
2	MND-ADA (VCN: 0.7)	Single	42.99	863,334,033	862,560,476	88,601,959	173,968,486	30.44%	9.89e+05x
1	EFS-ADA (VCN: 3)	Single	41.77	838,838,523	838,166,931	197,672,540	150,835,681	41.58%	9.92e+05x
4	MND-ADA (VCN: 0.7)	Double	10.38	207,812,221	207,656,042	78,438,755	11,413,916	43.26%	9.78e+05x
3	EFS-ADA (VCN: 3)	Double	72.35	1,448,376,160	1,447,352,878	1,247,407,423	181,196,855	98.70%	9.92e+05x
Sequencing Statistics Post Error Correction, Read Collapsing, and Filtering									
2	MND-ADA (VCN: 0.7)	Single	0.22	314,729	314,650	74,371	223,273	94.60%	3.53e+03x
1	EFS-ADA (VCN: 3)	Single	0.32	439,440	439,439	231,324	195,670	97.17%	1.77e+04x
4	MND-ADA (VCN: 0.7)	Double	0.21	360,895	360,871	90,609	254,405	95.61%	4.42e+03x
3	EFS-ADA (VCN: 3)	Double	0.58	718,074	718,063	370,100	326,895	97.07%	2.72e+04x

**Table 2.3:** Transduced CD3+ Cell Transgene Variants with Single Capture Hybridization  
Probe Enrichment

	<b>γ-Retrovirus (MND-ADA)</b>	<b>Lentivirus (EFS-ADA)</b>
<b>Total Mutations</b>	2	53
<b>Purine to Purine</b>		
A to G	0	0
G to A	0	48
<b>Pyrimidine to Pyrimidine</b>		
C to T	1	0
T to C	1	1
<b>Purine to Pyrimidine</b>		
A to C	0	1
A to T	0	0
G to C	0	0
G to T	0	0
<b>Pyrimidine to Purine</b>		
C to A	0	0
C to G	0	0
T to A	0	0
T to G	0	0
<b>Deletion/Insertion</b>	0	1
<b>Complex</b>	0	2

**Table 2.4:** Transduced CD3+ Cell Transgene Variants (Single vs. Double Hybridization Capture Comparison)

	<b>γ-Retrovirus (MND-ADA)</b>		<b>Lentivirus (EFS-ADA)</b>	
<b>Sample #</b>	2	4	1	3
<b>Capture</b>	Single	Double	Single	Double
<b>Total Mutations</b>	2	1	53	72
<b>Purine to Purine</b>				
A to G	0	0	0	0
G to A	0	0	48	59
<b>Pyrimidine to Pyrimidine</b>				
C to T	1	1	0	1
T to C	1	0	1	2
<b>Purine to Pyrimidine</b>				
A to C	0	0	1	0
A to T	0	0	0	0
G to C	0	0	0	0
G to T	0	0	0	0
<b>Pyrimidine to Purine</b>				
C to A	0	0	0	0
C to G	0	0	0	1
T to A	0	0	0	0
T to G	0	0	0	0
<b>Deletion/Insertion</b>	0	0	1	6
<b>Complex</b>	0	0	2	3

**Table 2.5:** Gene Therapy Patient Demographics

<b>Patient ID</b>	<b>Years Post Gene Therapy</b>	<b>ADA-SCID Mutations</b>	<b>VCN (Granulocytes)</b>	<b>VCN (PBMCs)</b>
402 (γ-RV)	10	T+2>G in IVS7 / unknown- unidentified	0.0167	0.512
404 (γ-RV)	4.5	T+2>G in IVS7 / unknown- unidentified	0.0956	0.145
405 (γ-RV)	5	R211H / R211H	0.0406	0.474
408 (γ-RV)	7	R211H / R211H	0.183	1.29
410 (γ-RV)	6	L107P / c.790delT	0.204	0.91
501 (LVV)	6	L107P / L107P	0.466	2.02
504 (LVV)	8	delc.955-959 / delc.955-959	0.238	0.625
509 (LVV)	8	L107P / L107P	0.399	1.594
511 (LVV)	8	G216R / c.(Exon1)del	0.751	1.99
512 (LVV)	5	c.360-362+1delTGAG / c.360-362+1delTGAG	0.149	1.59

**Table 2.6:** Information on CD3+ T-Cell Variants Identified in the Endogenous *ADA* gene locus

Variant	Position	Region	Allele	Allele Frequency (Our Dataset)	Validated Frequency (Global Study)	Consequences
SNV	Chr20:44636867	Intron 1	A> <u>G</u> ,T	1	G=0.9719 (1000Genomes30x)	Benign
SNV	Chr20:44636286	Exon 2	C>A, <u>I</u>	1	T=0.99286 (ALFA) T=0.9816 (1000Genomes30x)	Synonymous, Benign
SNV	Chr20:44636062	Intron 2	T>A, <u>C</u> , G	0.5088	C=0.89038 (ALFA) C=0.9197 (1000Genomes30x)	Benign
SNV	Chr20:44636050	Intron 2	C> <u>G</u> ,T	0.511	G=0.6966 (1000Genomes30x)	Benign
SNV	Chr20:44628564	Intron 3	G> <u>I</u>	0.2857	T=0.000007 (gnomAD) T=0.000011 (TopMED)	Benign
SNV	Chr20:44626054	Intron 3	G> <u>I</u>	0.2023	T=0.07045 (ALFA) T=0.1138 (1000Genomes30x)	Benign
SNV	Chr20:44624853	Intron 5	C> <u>I</u>	0.1875	T=0.20628 (ALFA) T=0.2631 (1000Genomes30x)	Benign
SNV	Chr20:44624274	Exon 6	T> <u>C</u>	0.188	C=0.185627 (ALFA) C=0.2728 (1000Genomes30x)	Synonymous, Benign
SNV	Chr20:44623126	Intron 6	G> <u>A</u>	0.1603	A=0.12629 (ALFA) A=0.1668 (1000Genomes30x)	Benign
SNV	Chr20:44622793	Intron 8	C> <u>I</u>	0.1744	A=0.12629 (ALFA) A=0.1668 (1000Genomes30x)	Benign
SNV	Chr20:44622081	Intron 9	G> <u>A</u>	0.15	A=0.162783 (ALFA) A=0.2616 (1000Genomes30x)	Benign
SNV	Chr20:44620950	Intron 9	C> <u>I</u>	0.1711	T=0.063452 (ALFA) T=0.1106 (1000Genomes30x)	Benign

Mutations in the endogenous *ADA* gene that are identified in our analysis are recognized as common variants found in the Single Nucleotide Polymorphism (SNP) database. Validated frequency from two common genome sequencing datasets, Allele Frequency Aggregator (ALFA) dataset and 1000 Genomes Project Phase 3 (1000Genomes\_30x), are shown. Mutations that are also identified in our analysis are underlined. Consequences are classified in the Reference SNP Reports of dbSNP database. Benign variants are reported on ClinVar database.

**Table 2.7: Gene Therapy Patient Sequencing Statistics Pre-Error Correction, Read Collapsing and Filtering**

Sample #	Cell Type	Read Total	Total Primary Alignments	Total Alignments to Transgene	Total Alignments to Endogenous ADA	% On Target Alignments	ADA cDNA Mean Coverage
402	Granulocytes	733,080,304	731,926,939	822,451	28,242,229	3.97%	7.86+04x
402	PBMCs	842,459,801	841,678,430	7,389,807	73,381,043	9.60%	7.76+05x
404	Granulocytes	740,752,536	739,745,435	823,965	285,076,681	38.68%	8.30+04x
404	PBMCs	727,564,741	726,824,794	4,636,516	61,429,158	9.09%	4.89+05x
405	Granulocytes	801,851,736	800,880,869	2,079,981	31,738,329	4.2%	2.16+05x
405	PBMCs	883,139,674	879,659,663	9,243,470	81,099,137	10.3%	7.15+05x
408	Granulocytes	1,143,708,875	1,141,987,552	2,132,258	3,772,878	0.52%	2.20+05x
408	PBMCs	1,068,953,540	1,067,809,656	57,102,608	98,006,418	14.53%	9.92+05x
410	Granulocytes	635,950,786	634,658,951	2,555,917	26,240,126	4.5%	2.65+05x
410	PBMCs	969,919,482	965,932,333	21,636,634	82,880,816	10.82%	9.82+05x
501	Granulocytes	729,781,985	728,761,412	10,137,490	36,318,448	6.37%	9.62+05x
501	PBMCs	430,722,037	430,374,338	25,480,720	20,226,687	10.62%	9.92+05x
504	Granulocytes	736,640,959	735,283,398	4,646,323	39,505,055	6.00%	4.96+05x
504	PBMCs	597,588,757	596,972,362	12,018,633	28,964,866	6.87%	9.83+05x
509	Granulocytes	660,283,113	659,524,628	10,961,293	33,728,858	6.78%	9.77+05x
509	PBMCs	602,150,898	601,595,794	27,632,462	32,269,522	9.96%	9.92+05x
511	Granulocytes	609,124,973	608,352,580	12,982,860	29,753,015	7.02%	9.86+05x
511	PBMCs	439,207,503	438,873,190	26,806,684	21,046,498	10.90%	9.92+05x
512	Granulocytes	596,660,053	597,076,342	3,666,009	30,996,413	5.81%	3.93+05x
512	PBMCs	544,025,568	544,025,568	22,120,139	25,590,784	8.77%	9.92+05x

**Table 2.8:** Gene Therapy Patient Sequencing Statistics Post Error-Correction, Read Collapsing and Filtering

Sample #	Cell Type	Read Total	Total Primary Alignments	Total Alignments to Transgene	Total Alignments to Endogenous ADA	% On Target Alignments	ADA cDNA Mean Coverage
402	Granulocytes	56,314	56,268	1,254	53,256	96.88%	101.58x
402	PBMCs	69,446	69,444	4,290	62,004	95.46%	320.20x
404	Granulocytes	51620	51,607	1,330	48,853	97.24%	115.81x
404	PBMCs	55,406	55,394	4,031	50,012	97.56%	352.82x
405	Granulocytes	61,771	61,768	4,067	56,220	97.60%	366.18x
405	PBMCs	76,497	76,222	7,560	66,223	96.80%	624.68x
408	Granulocytes	74,244	74,231	2,614	67,655	94.66%	204.73x
408	PBMCs	101,772	101,772	28,359	69,674	96.32%	2.17+03x
410	Granulocytes	56,043	56,015	4,979	49,502	97.26%	424.67x
410	PBMCs	90,643	90,294	18,551	68,643	96.57%	1.55+03x
501	Granulocytes	57,150	57,143	14,433	41,921	98.92%	1.34+03x
501	PBMCs	110,374	110,374	64,128	45,408	99.24%	5.94+03x
504	Granulocytes	72,542	72,542	5,593	64,832	96.46%	461.53x
504	PBMCs	97,667	97,667	20,587	74,193	97.04%	1.66+03x
509	Granulocytes	69,749	69,749	12,578	55,272	97.28%	1.06+03x
509	PBMCs	121,305	121,305	42,111	76,526	97.8%	3.44+03x
511	Granulocytes	53,043	53,042	19,076	33,279	98.7%	1.75+03x
511	PBMCs	115,306	115,304	68,709	45,715	99.24%	6.35+03x
512	Granulocytes	39,125	39,123	4,857	33,648	98.42%	446.87x
512	PBMCs	104,009	103,994	50,068	52,815	98.93%	4.51+03x

**Table 2.9:** Transgene DNA Variants in  $\gamma$ -Retroviral and Lentiviral ADA-SCID Gene Therapy Patients

	$\gamma$ -Retrovirus (MND-ADA)										Lentivirus (EFS-ADA)									
	Granulocytes					PBMCs					Granulocytes					PBMCs				
Sample	402	404	405	408	410	402	404	405	408	410	501	504	509	511	512	501	504	509	511	512
<b>Total Mutations</b>	11	6	14	9	7	8	8	10	14	21	77	15	68	61	20	75	75	106	76	97
<b>Purine to Purine</b>																				
<b>A to G</b>	2	1	2	1	3	1	0	1	0	1	1	0	0	0	0	0	0	0	0	0
<b>G to A</b>	1	0	3	1	3	5	2	2	10	9	72	15	68	58	18	74	73	103	75	92
<b>Pyrimidine to Pyrimidine</b>																				
<b>C to T</b>	1	1	3	1	2	1	2	1	2	2	1	0	0	2	1	0	2	0	1	2
<b>T to C</b>	2	1	0	1	2	0	0	2	0	2	0	0	0	0	0	0	0	1	0	1
<b>Purine to Pyrimidine</b>																				
<b>A to C</b>	1	0	1	1	1	0	1	1	0	1	0	0	0	0	0	0	0	1	0	0
<b>A to T</b>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<b>G to C</b>	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<b>G to T</b>	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
<b>Pyrimidine to Purine</b>																				
<b>C to A</b>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<b>C to G</b>	1	1	0	1	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0
<b>T to A</b>	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
<b>T to G</b>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<b>Deletion/Insertion</b>	0	0	0	0	1	0	0	0	2	3	0	0	0	0	1	0	0	0	0	2
<b>Complex</b>	3	2	5	2	4	1	2	2	0	2	2	0	0	0	0	1	0	1	0	0

**Table 2.10:** Gene Therapy Patient Variants in the Endogenous *ADA* gene locus

	γ-Retrovirus (MND-ADA)										Lentivirus (EFS-ADA)									
	Granulocytes					PBMCs					Granulocytes					PBMCs				
Sample	402	404	405	408	410	402	404	405	408	410	501	504	509	511	512	501	504	509	511	512
<b>Total Mutations</b>	5	6	5	11	6	7	5	6	9	7	9	12	7	4	6	5	15	5	9	12
<b>Purine to Purine</b>																				
<b>A to G</b>	0	0	0	0	1	1	0	1	0	2	1	1	2	0	0	1	1	1	0	1
<b>G to A</b>	0	0	0	1	0	0	0	0	1	0	1	3	0	0	0	0	4	0	1	2
<b>Pyrimidine to Pyrimidine</b>																				
<b>C to T</b>	1	1	2	5	1	1	1	2	3	1	2	2	1	1	1	1	3	1	3	2
<b>T to C</b>	1	1	1	2	2	1	1	1	2	2	1	3	2	1	1	2	3	2	2	1
<b>Purine to Pyrimidine</b>																				
<b>A to C</b>	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
<b>A to T</b>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<b>G to C</b>	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
<b>G to T</b>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<b>Pyrimidine to Purine</b>																				
<b>C to A</b>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<b>C to G</b>	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1
<b>T to A</b>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<b>T to G</b>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<b>Deletion/Insertion</b>	0	0	1	2	1	1	0	1	2	1	2	3	1	0	3	1	4	0	3	5
<b>Complex</b>	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

**Table 2.11:** Identified Variants in the Endogenous ADA gene locus of patient samples; in database

Variant	Position	Region	Allele	Patient ID	Allele Frequency (Our Dataset)	Validated Frequency (Global Study)	Consequences
SNV	Chr20:44636867	Intron 1	A> <u>G</u> ,T	402, 405, 410, 509, 512	1, 1, 1, 1, 1	G=0.9719 (1000Genomes_30x)	Intron Variant
SNV	Chr20:44636580	Intron 1	T> <u>C</u>	410, 501, 509	0.5352, 0.4889, 1	C=0.00173 (ALFA) C=0.0003 (1000Genomes_30x)	Intron Variant
SNV	Chr20:44636286	Exon 2	C>A, <u>I</u>	ALL	1, 1, 1, 1, 1, 1, 1, 1, 1, 1	T=0.99286 (ALFA) T=0.9816 (1000Genomes_30x)	Synonymous, Benign
SNV	Chr20:44636062	Intron 2	T>A, <u>C</u> ,G	402, 404, 405, 408, 410, 501, 509, 511, 512	1, 1, 1, 1, 1, 1, 1, 1, 1	C=0.89038 (ALFA) C=0.9197 (1000Genomes_30x)	Intron Variant
SNV	Chr20:44636050	Intron 2	C> <u>G</u> ,T	402, 404, 405, 408, 410, 501, 509, 511, 512	1, 1, 1, 1, 1, 0.5322, 1, 1, 1	G=0.6583 (ALFA) G=0.6966 (1000Genomes_30x)	Intron Variant
SNV	Chr20:44629073	Exon 3	C> <u>I</u>	501, 511, 512	0.0096(G), 0.0033, 0.0262	T=0.0006 (1000Genomes_30x)	Synonymous, Benign
SNV	Chr20:44628765	Intron 3	C> <u>I</u>	504	0.9863	T=0.068580 (ALFA) T=0.0595 (1000Genomes_30x)	Intron Variant
SNV	Chr20:44626579	Exon 4	T> <u>C</u>	504	0.9995	C=0.063659 (ALFA) C=0.0517 (1000Genomes_30x)	K80R Missense Variant, Benign
<b>SNV</b>	<b>Chr20:44626498</b>	<b>Exon 4</b>	<b>A&gt;<u>G</u></b>	<b>410, 501, 509</b>	<b>0.4848, 1, 1</b>	<b>G=11 (ClinVar)</b>	<b>L107P Missense Pathogenic</b>
SNV	Chr20:44625657	Exon 5	C> <u>I</u>	408	0.4998	T=0.03738 (ALFA) T=0.0323 (1000Genomes_30x)	Synonymous
SNV	Chr20:44624853	Intron 5	C> <u>I</u>	408, 504	1, 1	T=0.20628 (ALFA) T=0.2631 (1000Genomes_30x)	Intron Variant
<b>SNV</b>	<b>Chr20:44624292</b>	<b>Exon 6</b>	<b>G&gt;<u>C</u></b>	<b>402, 404</b>	<b>0.4781, 0.4522</b>	<b>GTA &gt; CTA (Analysis)</b> <b>GTA&gt;TTA (ClinVar)</b>	<b>Y&gt;stop, Y&gt;stop, Pathogenic</b>
SNV	Chr20:44624274	Exon 6	T> <u>C</u>	408, 504	0.505, 1	C=0.185627 (ALFA) C=0.2728 (1000Genomes_30x)	Synonymous, Benign
SNV	Chr20:44624007	Intron 6	A> <u>G</u>	504	0.9944	G=0.07464 (ALFA) G=0.0623 (1000Genomes_30x)	Intron Variant
SNV	Chr20:44623126	Intron 6	G> <u>A</u>	504	0.9992	A=0.12629 (ALFA) A=0.1668 (1000Genomes_30x)	Intron Variant
<b>SNV</b>	<b>Chr20:44623053</b>	<b>Exon 7</b>	<b>C&gt;<u>I</u></b>	<b>405, 408</b>	<b>0.9995, 1</b>	<b>ACG&gt;ATG</b> <b>T=11 (ClinVar)</b>	<b>R211H Missense, Pathogenic</b>
<b>SNV</b>	<b>Chr20:44623039</b>	<b>Exon 7</b>	<b>C&gt;<u>I</u></b>	<b>511</b>	<b>0.4696</b>	<b>T=0.00011 (ALFA)</b> <b>T=13 (ClinVar)</b>	<b>G216R Missense, Pathogenic</b>
SNV	Chr20:44623005	Intron 7	A> <u>C</u>	402, 404	0.4858, 0.4822	C=0.000007 (gnomAD)	Splice Donor Variant
SNV	Chr20:44622793	Intron 8	C> <u>I</u>	408	0.49	A=0.12629 (ALFA) A=0.1668 (1000Genomes_30x)	Intron Variant
<b>DEL</b>	<b>Chr20:44622642</b>	<b>Exon 9</b>	<b>CA&gt;<u>C</u></b>	<b>410</b>	<b>0.4946</b>	<b>delA=3 (ClinVar)</b>	<b>Frameshift, Likely-Pathogenic</b>
SNV	Chr20:44622081	Intron 9	G> <u>A</u>	408, 504	0.4706, 1	A=0.162783 (ALFA)	Intron Variant

						A=0.2616 (1000Genomes_30x)	
<b>DEL</b>	<b>Chr20:44621032</b>	<b>Exon 10</b>	<b>CCTCTT &gt;C</b>	<b>504</b>	<b>0.9984</b>	<b>Deletion of E319-320</b>	<b>Frameshift</b>
SNV	Chr20:44620212	Intron 11	T>> <u>C</u>	504	1	C=0.06501 (ALFA) C=0.0754 (100Genomes_30x)	Intron Variant
SNV	Chr20:44620200	Intron 11	G>> <u>A</u>	504	1	A=0.07009 (ALFA) A=0.0732 (100Genomes_30x)	Intron Variant

Mutations in the endogenous *ADA* gene that are identified in our analysis are recognized as common variants found in the Single Nucleotide Polymorphism (SNP) database. Validated frequency from two common genome sequencing datasets, Allele Frequency Aggregator (ALFA) dataset and 1000 Genomes Project Phase 3 (1000Genomes\_30x), NHBLI's Trans-Omics for Precision Medicine Whole Genome Sequencing Project (TopMed), Genome Aggregation Database (gnomAD), are shown. For pathogenic mutations, the number of submissions on ClinVar are shown. Mutations that are also identified in our analysis are underlined. Consequences are classified in the Reference SNP Reports of dbSNP database. Benign and Pathogenic variants are reported on ClinVar database. Patient ID's are summarized for every identified variant. ADA-SCID mutations are in blue.

**Table 2.12:** Identified Variants in the Endogenous *ADA* gene locus of patient samples; but not in the current database

Variant	Position	Region	Allele	Patient ID	Allele Frequency (Our Dataset)	Potential consequences
DEL	Chr20:44636307	Intron 1	AAG>A	405, 504	0.002, 0.0019 (G)	Intron Variant
SNV	Chr20:44636307	Intron 1	G>A	408	1	Intron Variant
DEL	Chr20:44635902	Intron 2	G>GA	402, 405, 501, 504	0.0504, 0.0462, 0.381, 0.0435	Intron Variant
DEL	Chr20:44635902	Intron 2	GA>G	404, 405, 408, 501, 504	0.0366 (G), 0.0462 0.0495, 0.381, 0.0164	Intron Variant
INS	Chr20:44629417	Intron 2	A>AGT	408, 504, 509,	0.3351, 0.6364, 0.0103	Intron Variant
SNV	Chr20:44626545	Exon 4	G>A	501, 511, 512	0.0076, 0.0013, 0.0248	Synonymous, Benign
DEL	Chr20:44626447	Intron 4	ATCAC>A	512	0.9842	Intron Variant
DEL	Chr20:44625684	Exon 5	Large deletion	501, 511, 512	0.0119, 0.0055, 0.0455	Potentially Pathogenic
DEL	Chr20:44625291	Intron 5	AT>A	408	0.0208,	Intron Variant
DEL	Chr20:44622929	Exon 8	Large deletion	501, 512	0.0029 (G), 0.0065	Potentially Pathogenic
DEL	Chr20:44622652	Exon 9	Large deletion	501, 512	0.0028 (G), 0.0065	Potentially Pathogenic
SNV	Chr20:44621168	Intron 9	G>A	504, 512	0.0036, 0.0032	Intron Variant
DEL	Chr20:44620401	Exon 11	Large deletion	509, 512	0.0035, 0.0138	Potentially Pathogenic

**Table 2.13: APOBEC3 DNA Substrate Statistics in EFS-ADA Treated Patients**

Sample	Granulocytes					PBMCs				
	501	504	509	511	512	501	504	509	511	512
Total G to A Mutations	72	15	68	58	18	74	73	103	75	92
Mutations with APOBEC3 Substrate Signature	64	14	53	46	16	65	66	88	65	76
% APOBEC3 Signature	88.9%	93.3%	77.9%	79.3%	88.9%	87.8%	90.4%	85.4%	86.7%	82.6%

## REFERENCES

1. Dunbar, C. E. *et al.* Gene therapy comes of age. *Science* **359**, eaan4672 (2018).
2. Blaese, R. M. *et al.* T Lymphocyte-Directed Gene Therapy for ADA– SCID: Initial Trial Results After 4 Years. *Science* **270**, 475–480 (1995).
3. Maetzig, T., Galla, M., Baum, C. & Schambach, A. Gammaretroviral Vectors: Biology, Technology and Application. *Viruses* **3**, 677–713 (2011).
4. Kohn, D. B., Sadelain, M. & Glorioso, J. C. Occurrence of leukaemia following gene therapy of X-linked SCID. *Nat Rev Cancer* **3**, 477–488 (2003).
5. Hacein-Bey-Abina, S. *et al.* Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J Clin Invest* **118**, 3132–3142 (2008).
6. Morgan, R. A., Gray, D., Lomova, A. & Kohn, D. B. Hematopoietic Stem Cell Gene Therapy: Progress and Lessons Learned. *Cell Stem Cell* **21**, 574–590 (2017).
7. Hu, W.-S. & Hughes, S. H. HIV-1 Reverse Transcription. *Cold Spring Harb Perspect Med* **2**, a006882 (2012).
8. Ocorbin, I. P. & Filipenko, M. L. M-MuLV reverse transcriptase: Selected properties and improved mutants. *Computational and Structural Biotechnology Journal* **19**, 6315–6327 (2021).
9. Yasukawa, K. *et al.* Next-generation sequencing-based analysis of reverse transcriptase fidelity. *Biochemical and Biophysical Research Communications* **492**, 147–153 (2017).
10. Xu, W. K., Byun, H. & Dudley, J. P. The Role of APOBECs in Viral Replication. *Microorganisms* **8**, 1899 (2020).

11. Kennedy, S. R. *et al.* Detecting ultralow-frequency mutations by Duplex Sequencing. *Nat Protoc* **9**, 2586–2606 (2014).
12. Migliavacca, M. *et al.* Long-term and real-world safety and efficacy of retroviral gene therapy for adenosine deaminase deficiency. *Nat Med* **30**, 488–497 (2024).
13. Carbonaro, D. A. *et al.* Preclinical Demonstration of Lentiviral Vector-mediated Correction of Immunological and Metabolic Abnormalities in Models of Adenosine Deaminase Deficiency. *Molecular Therapy* **22**, 607–622 (2014).
14. Schmitt, M. W. *et al.* Sequencing small genomic targets with high efficiency and extreme accuracy. *Nat Methods* **12**, 423–425 (2015).
15. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308–311 (2001).
16. Maiti, A., Hou, S., Schiffer, C. A. & Matsuo, H. Interactions of APOBEC3s with DNA and RNA. *Curr Opin Struct Biol* **67**, 195–204 (2021).
17. Campos, L. W., Pissinato, L. G. & Yunes, J. A. Deleterious and Oncogenic Mutations in the IL7RA. *Cancers (Basel)* **11**, 1952 (2019).
18. Harris, R. S. *et al.* DNA Deamination Mediates Innate Immunity to Retroviral Infection. *Cell* **113**, 803–809 (2003).
19. Salter, J. D., Bennett, R. P. & Smith, H. C. The APOBEC Protein Family: United by Structure, Divergent in Function. *Trends Biochem Sci* **41**, 578–594 (2016).
20. Ji, J. & Loeb, L. A. Fidelity of HIV-1 reverse transcriptase copying RNA in vitro. *Biochemistry* **31**, 954–958 (1992).
21. Zhao, W., Akkawi, C., Mougel, M. & Ross, S. R. Murine Leukemia Virus P50 Protein Counteracts APOBEC3 by Blocking Its Packaging. *J Virol* **94**, e00032-20 (2020).

22. Han, J. *et al.*  $\beta$ -Globin Lentiviral Vectors Have Reduced Titers due to Incomplete Vector RNA Genomes and Lowered Virion Production. *Stem Cell Reports* **16**, 198–211 (2021).
23. Soneoka, Y. *et al.* A transient three-plasmid expression system for the production of high titer retroviral vectors. *Nucleic Acids Research* **23**, 628–633 (1995).
24. Picard Toolkit. *Broad Institute, Github repository* (2019).
25. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
26. Lai, Z. *et al.* VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res* **44**, e108 (2016).

## CONCLUSIONS AND FUTURE DIRECTIONS

Accessibility and safety will always be top priorities in the advancement of gene therapy. One of the major challenges in gene therapy is the limited accessibility to these treatments, largely due to the complex and expensive manufacturing processes. The high cost of gene therapies makes them inaccessible to many patients.<sup>15</sup> To address these barriers, several strategies are being explored. This work in developing the LV UV1-DS, aims to reduce costs of manufacturing by improving viral vector titers and enhancing gene transfer efficiency. Additionally, governments and healthcare providers are exploring new models for reimbursement and insurance coverage to make these therapies more affordable, particularly in cases where they provide long-term benefits that reduce overall healthcare costs. Expanding partnerships with global health organizations and biotechnology firms can also help facilitate the distribution of gene therapies to underserved regions.<sup>16</sup> Furthermore, efforts to train healthcare professionals and improve the infrastructure of healthcare systems in low- and middle-income countries can play a key role in making these therapies accessible to a broader population.

As insights into mutational profiles and the role of APOBEC3 in  $\gamma$ -retroviral and lentiviral gene therapy continue to emerge, future studies will need to focus on several key areas to enhance the safety of these approaches. Gene editing techniques aimed at deleting or silencing APOBEC3 in packaging cells may provide a more consistent, lower-risk gene therapy product by preventing unwanted mutations in the integrated transgene. Additionally, applying these techniques on a broader range of patient samples, including those with different genetic diseases will help establish a more comprehensive

understanding of the mutational landscapes associated with viral vector gene therapies. It will be crucial to investigate the long-term stability and functional consequences of mutations in treated cells, particularly for diseases where even small genetic changes could trigger oncogenesis, like in the case of *IL7R $\alpha$*  mutations. By integrating these patient-specific mutational insights with advances in vector engineering and packaging, future gene therapies may become safer and more precise, ultimately improving clinical outcomes across a broader range of genetic disorders and malignancies.

## BIBLIOGRAPHY

1. Morgan, R. A., Gray, D., Lomova, A. & Kohn, D. B. Hematopoietic Stem Cell Gene Therapy –Progress and Lessons Learned. *Cell Stem Cell* **21**, 574 (2017).
2. Viral Gene Delivery System - an overview | ScienceDirect Topics.  
<https://www.sciencedirect.com/topics/neuroscience/viral-gene-delivery-system>.
3. Escors, D. & Breckpot, K. Lentiviral vectors in gene therapy: their current status and future potential. *Arch. Immunol. Ther. Exp. (Warsz.)* **58**, 107–119 (2010).
4. CDC. Complications and Treatments of Sickle Cell Disease | CDC. *Centers for Disease Control and Prevention*  
<https://www.cdc.gov/ncbddd/sicklecell/treatments.html> (2020).
5. Robinson, T. M. & Fuchs, E. J. Allogeneic Stem Cell Transplantation for Sickle Cell Disease. *Curr. Opin. Hematol.* **23**, 524–529 (2016).
6. Leonard, A. & Tisdale, J. F. A new frontier: FDA approvals for gene therapy in sickle cell disease. *Mol. Ther.* **32**, 264–267 (2024).
7. Levasseur, D. N. *et al.* A recombinant human hemoglobin with anti-sickling properties greater than fetal hemoglobin. *J. Biol. Chem.* **279**, 27518–27524 (2004).
8. Liu, B. *et al.* Development of a double shmiR lentivirus effectively targeting both BCL11A and ZNF410 for enhanced induction of fetal hemoglobin to treat  $\beta$ -hemoglobinopathies. *Mol. Ther.* (2022) doi:10.1016/j.ymthe.2022.05.002.
9. Esrick, E. B. *et al.* Post-Transcriptional Genetic Silencing of BCL11A to Treat Sickle Cell Disease. *N. Engl. J. Med.* **384**, 205–215 (2021).
10. Blaese, R. M. *et al.* T Lymphocyte-Directed Gene Therapy for ADA– SCID: Initial Trial Results After 4 Years. *Science* **270**, 475–480 (1995).

11. Shaw, K. L. *et al.* Clinical efficacy of gene-modified stem cells in adenosine deaminase–deficient immunodeficiency. *J. Clin. Invest.* **127**, 1689–1699 (2017).
12. Reinhardt, B. *et al.* Long-term outcomes after gene therapy for adenosine deaminase severe combined immune deficiency. *Blood* **138**, 1304–1316 (2021).
13. Kohn, D. B. *et al.* Autologous Ex Vivo Lentiviral Gene Therapy for Adenosine Deaminase Deficiency. *N. Engl. J. Med.* **384**, 2002–2013 (2021).
14. Xu, W. K., Byun, H. & Dudley, J. P. The Role of APOBECs in Viral Replication. *Microorganisms* **8**, 1899 (2020).
15. Cornetta, K. *et al.* Gene therapy access: Global challenges, opportunities, and views from Brazil, South Africa, and India. *Mol. Ther.* **30**, 2122–2129 (2022).
16. Muigai, A. W. T. Expanding global access to genetic therapies. *Nat. Biotechnol.* **40**, 20–21 (2022).