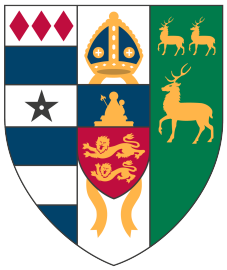
# Towards an Engineered Autologous Cellular Therapy for RAG2 Deficiency

## Advances in the Treatment of Severe Combined Immunodeficiency

**Cameron Lee Gardner**

*A thesis presented for the degree of*
*Doctor of Philosophy*






Sir William Dunn School of Pathology
National Institute for Allergy and Infectious Diseases
University of Oxford
National Institutes of Health
United Kingdom
United States of America
**Michaelmas 2020**

# Towards an Engineered Autologous Cellular Therapy for RAG2 Deficiency

The Future of Severe Combined Immunodeficiency Treatment

**Cameron Lee Gardner**

## Abstract

Severe combined immune deficiency (SCID) caused by RAG deficiency is a genetically-determined immune deficiency characterized by the virtual absence of adaptive immune cells. Unless treated with a hematopoietic stem cell transplantation (HSCT), RAG deficient patients succumb to severe infections early in life. HSCT, however, carries risks of graft-versus-host disease. Moreover, a high rate of graft failure and poor immune reconstitution have been reported after unconditioned HSCT. Expression of the RAG genes is tightly regulated, and preclinical attempts of gene therapy with heterologous promoters have led to controversial results. Using patient-derived induced pluripotent stem cells (iPSCs), patient hematopoietic stem and progenitor cells (HSPCs) and an *in vitro* artificial thymic organoid system as a model, here we demonstrate that gene editing rescues the progressive T-cell differentiation potential of RAG2 deficient cells to normal levels, with generation of a diversified T cell repertoire. These results suggest that gene editing may represent a novel autologous therapeutic option for correction of this immunodeficiency.

# Acknowledgements

I would like to thank the brilliant group of people that I was fortunate enough to work with during my training who have helped me develop as a scientist and who have contributed to the success of this thesis project.

To my supervisors Dr. Luigi Notarangelo and Profesor Oreste Acuto, thank you deeply for the opportunity you have given me to develop my passion and my skills in bioenegineering and cell/gene therapies. Not many students similarly get the chance to work on their dream project. Gigi, thank you for allowing me to take ownership of a project that I care deeply about. My PhD training has given me the skills I need to take into the future to continue to make an impact in gene and cell therapies. Oreste, your guidance has been fundamental to my development as a scientist, and your heart and humor have been an absolute gift over the last four years. I will remember fondly our frequent discussions over lunch and hope that they will continue well into the future. Thank you both for your mentorship and for changing my life as a scientist.

To my labmates, you have all helped me survive the at times perilous PhD journey. Specifically to Dr. Marita Bosticardo, with whom I worked to establish the organoid system in the lab, I thank you for helping me learn about immunological experimental work. You are a fantastic scientist and a fabulous mentor. Dr. Marita Bosticardo contributed to figure 3.7 and 3.8. To Kerry Dobbs, thank you for your endless work ethic and for all of the many things you juggle. Kerry Dobbs performed the reprogramming of iPSCs used in this thesis work from patient fibroblasts using a non-integrating Sendai virus. Justin Lack at the NIAID Collaborative Bioinformatics Resource Center performed the whole genome sequencing data processing.

To Julianne Cohen, thank you so much for your ever-present positive outlook and efficient work flow. You have enabled much of the work present in this thesis. To Dr. P.K. Reardon, I thank you for your enthusiastic outlook and for your help in processing some of the NGS data and circos plots (Figure 4.20 and Figure 4.21).

To my collaborators Professor Matthew Porteus and Dr. Mara Pavel-Dinu, I thank you for the opportunity to work together to advance a potential therapy for patients who needed treatment yesterday. Mara spearheaded many aspects of this project, including all of the *in vivo* work (not shown), production of the AAV used in editing studies, as well as some work in investigating reducing cytotoxicity in edited hematopoietic stem and progenitor cells which is included in this thesis (Figures 3.11-3.14). This project could not have been successful without your hard work and scientific merit.

To our collaborators at the University of California Los Angeles, I thank you for providing critical guidance in the establishment of the artificial thymic organoid (ATO) system, as well as for providing the murine stromal cell line used in all work demonstrated in this thesis. I would also like to thank Dr. Sally Cowley at the University of Oxford who helped provide some preliminary iPSC lines that were used to test this ATO system.

This work was funded by the National Institutes of Health and by the Wellcome Trust.

Finally, I would like to thank the NIH Oxford Cambridge Scholars Program, which has changed my life forever with the opportunity to carry out high grade research between two globally leading research institutions. The connections I have formed through this program will continue for the rest of my life, so I offer my gratitude for having had the opportunity to work alongside such a talented and compassionate group of young scientists.

The use of pronoun "we" will be found frequently throughout this thesis work instead of the pronoun "I". I would like to acknowledge each person who contributed intellectually - even in informal conversation - to the experiments that I performed.

With that said, I performed all experiments present within this thesis except where otherwise indicated above.

# Declaration

I, Cameron Lee Gardner, declare that this thesis and all of the research contents within represent my own original work unless otherwise indicated. The results contained within this thesis were generated at the National Institutes of Health as well as in the Sir William Dunn School of Pathology at the University of Oxford between Michaelmas 2016 and Trinity 2020 as a part of the NIH Oxford-Cambridge Scholars Biomedical Science DPhil Course under the supervision of Dr. Luigi Notarangelo and Prof. Oreste Acuto. No part of this thesis has been submitted for any other degree program or professional purpose apart from publication in a peer reviewed journal.

# List of Acronyms


**AAV** Adeno Associated Virus . . . . . 38

**ADA** Adenosine Deaminase . . . . . 20

**aGvHD** Acute Graft vs. Host Disease . . . . . 24

**AIRE** AutoImmune REgulator . . . . . 30

**APC** Antigen Presenting Cell . . . . . 30

**ATO** Artificial Thymic Organoid . . . . . 47

**BGH** Bovine Growth Hormone . . . . . 69

**CAI** Codon Adaptation Index . . . . . 70

**CBC** Complete Blood Count . . . . . 23

**CDR3** Complementary Determining Region 3 . . . . . 130

**CGD** Chronic Granulomatous Disease . . . . . 42

**cGvHD** chronic Graft vs. Host Disease . . . . . 24

**CMV** Cytomegalovirus . . . . . 22

**CRISPR** Clustered Regularly Interspaced Short Palindromic Repeat . . . . 39

**crRNA** CRISPR RNA . . . . . 40

**cTEC** Cortical Thymic Epithelial Cell . . . . . 29

**ddPCR** droplet digital PCR . . . . . 78

**DLL1** Delta-like 1 . . . . . 48

**DLL4** Delta-like 4 . . . . . 48

**DP** Double Positive . . . . . 29

**ERT** Enzyme Replacement Therapy . . . 23
**ESC** Embryonic Stem Cell . . . 44
**ETP** Early Thymic Progenitors . . . 28
**FACS** Fluorescence Activated Cell Sorting . . . 116
**gDNA** Genomic DNA . . . 91
**GFP** Green Fluorescent Protein . . . 69
**GVHD** Graft Versus Host Disease . . . 42
**hEMP** Human Embryonic Mesodermal Progenitor . . . 49
**HLA** Human Leukocyte Antigen . . . 24
**HDR** Homology Directed Repair . . . 40
**HSC** Hematopoietic Stem Cell . . . 47
**HSCT** Hematopoietic Stem Cell Transplantation . . . 23
**HSPC** Hematopoietic Stem and Progenitor Cell . . . 23
**HSV** Herpes Simplex Virus . . . 22
**HTS** High Throughput Sequencing . . . 130
**IEI** Inborn Errors of Immunity . . . 17
**IL2RG** Interleukin 2 Receptor Gamma . . . 20
**IL7** Interleukin 7 . . . 29
**INDEL** INsertion or DELetion . . . 33
**iPSC** Induced Pluripotent Stem Cell . . . 44
**ISP** Immature Single Positive . . . 80
**LTR** Long Terminal Repeat . . . 27
**MHC** Major Histocompatability Complex . . . 30
**MHCI** Major Histocompatability Complex I . . . 30
**MHCII** Major Histocompatability Complex II . . . 30

**MOI** Multiplicity of Infection . . . 68
**mTEC** Medullary Thymic Epithelial Cell . . . 30
**MUD** Matched Unrelated Donor . . . 25
**NGS** Next Generation Sequencing . . . 68
**NHEJ** Non Homologous End Joining . . . 33
**NK** Natural Killer . . . 80
**NSG** NOD-SCID-Gamma . . . 51
**pA** Polyadenylation Signal . . . 69
**PAM** Protospacer Adjacent Motif . . . 40
**PBMC** Peripheral Blood Mononuclear Cell . . . 60
**PHD** Plant Homeodomain . . . 32
**PID** Primary Immune Deficiency . . . 17
**PSC** Pluripotent Stem Cell . . . 44
**PTM** Post Translational Modification . . . 33
**qPCR** Quantitative PCR . . . 22
**RAG1** Recombination Activating Gene 1 . . . 20
**RAG2** Recombination Activating Gene 2 . . . 20
**RNP** Ribonucleoprotein . . . 75
**RSS** Recombination Signal Sequence . . . 21
**RSV** Respiratory Syncytial Virus . . . 22
**SCID** Severe Combined Immunodeficiency . . . 18
**SFFV** Spleen Focus Forming Virus . . . 69
**sgRNA** Short Guide RNA . . . 68
**SNP** Single Nucleotide Polymorphism . . . 58
**SP** Single Positive . . . 30

**ssDNA** Single Stranded DNA . . . . . 43

**TALEN** Transcriptional Activator-like Effector Nuclease . . . . . 38

**TCR** T Cell Receptor . . . . . 29

**tracrRNA** trans-activating CRISPR RNA . . . . . 39

**TRECs** T-cell receptor excision circles . . . . . 22

**tRNA** transfer RNA . . . . . 70

**TSP** Thymic Seeding Progenitors . . . . . 28

**vg** Vector Genomes . . . . . 85

**WAS** Wiskott-Aldrich Syndrome . . . . . 42

**WES** Whole Exome Sequencing . . . . . 23

**WGS** Whole Genome Sequencing . . . . . 23

**ZFN** Zinc Finger Nuclease . . . . . 38

# Contents


**1 Introduction** **17**

1.1 Primary Immune Deficiencies/Inborn Errors of Immunity . . . . . 17

1.1.1 Severe Combined Immunodeficiency . . . . . 19

1.1.2 Clinical Presentation and Diagnosis . . . . . 22

1.1.3 Issues with Standard of Care . . . . . 23

1.2 T cell Biology and Ontology . . . . . 28

1.3 RAG2 Structure and Function . . . . . 32

1.3.1 TCR Repertoire Generation . . . . . 34

1.4 Gene Editing . . . . . 37

1.4.1 Gene Therapy for *RAG2* Deficiency . . . . . 42

1.5 iPSC Modeling . . . . . 44

1.5.1 Overview of the Artificial Thymic Organoid System . . . . . 47

1.6 Aims and Objectives . . . . . 50

**2 Materials and Methods** **52**

2.1 Culturing and Engineering of Cell Lines . . . . . 52

2.1.1 sgRNAs and Plasmids . . . . . 52

2.1.2 iPSC Culture . . . . . 53

2.1.3 iPSC Editing . . . . . 54

2.1.4 Analysis of INDEL Frequency and Identity . . . . . 55

2.2 PCR, ddPCR, qPCR . . . . . 55

2.2.1 iPSC Reprogramming and Karyotyping . . . . . 57

2.3 Immunosequencing . . . 57
2.3.1 WGS of Edited and Unedited iPSCs . . . 58
2.3.2 HSPC Culture and Expansion . . . 59
2.3.3 HSPC Expansion and V-beta Profiling . . . 59
2.3.4 HSPC Editing . . . 60
2.3.5 Murine Stromal Cell line Culture . . . 61
2.4 Flow Cytometry . . . 62
2.5 Artificial Thymic Organoid System . . . 63
2.5.1 Derivation of Human Embryonic Mesodermal Progenitors . . . 63
2.5.2 Generation of Mature Naive T cells from iPSCs . . . 63
2.6 Statistical Analysis . . . 65

**3 Gene Editing and Differentiation in Hematopoietic Stem Cell Derived T Cells 66**

3.1 Characterizing Patient HSPCs . . . 66
3.1.1 Introduction . . . 66
3.1.2 Objectives and Hypotheses . . . 67
3.2 Genome Engineering in HSPCs . . . 68
3.2.1 Design of Viral Cargo . . . 69
3.2.2 Testing of sgRNAs . . . 71
3.2.3 Knock-in Efficiency Scales with Number of Cells Electroporated and MOI . . . 73
3.2.4 Knock in Efficiency Inversely Correlates with AAV Cargo Length 76
3.3 Artifical Thymic Organoids (ATOs) from HSPC Precursors . . . 78
3.3.1 ATOs Faithfully Recapitulate Physiological Development . . . 78
3.3.2 HSPC Derived T Cells Proliferate in Response to anti CD3 CD28 Antibodies and Display a Diverse V-beta Repertoire . . . 81
3.3.3 RAG Deficient Patient T Cell Developmental Blocks are Reproduced in the ATO System . . . 84
3.3.4 AAV Mediated Gene Editing Toxic in ATO System . . . 84

3.4 Amelioration of AAV-mediated Cytotoxicity in HSPC Editing . . . . 87
3.4.1 Further Expansion Time Allows for Nonlinear Increase in the Proliferation of HSPCs . . . . . 89
3.4.2 Electroporation Program Expands Post-Nucleofection Proliferation Capability of HSPCs . . . . . 90
3.4.3 Gentle Electroporation Programs Do not Confer Increased RNP Delivery . . . . . 91
3.4.4 Further HSPC Expansion Time Promotes Higher Levels of HDR 91
3.4.5 Optimized Conditions May Abrogate Cytotoxicity in ATO System . . . . . 94
3.5 Discussion . . . . . 97

**4 Gene Editing, Differentiation, and Characterization in Induced Pluripotent Stem Cell Derived T Cells 101**
4.1 Modeling RAG Deficiency and Rescue with Patient iPSCs . . . . . 101
4.1.1 Introduction . . . . . 101
4.1.2 Objectives and Hypotheses . . . . . 102
4.2 Characterizing Patient iPSCs . . . . . 103
4.2.1 Reprogramming RAG2 Deficient Patient Cells . . . . . 103
4.3 Genome Engineering in iPSCs . . . . . 105
4.3.1 Electroporation in iPSCs . . . . . 105
4.3.2 CRISPR/Cas9 RNP Activity in iPSCs . . . . . 106
4.3.3 CRISPR/Cas9 AAV Mediated Knock-in of the r.RAG2 Cassette in Patient iPSCs . . . . . 107
4.4 Analysis of Edited iPSC Line and Isogenic Precursor by WGS . . . . 111
4.5 Optimizing iPSC ATO System . . . . . 115
4.5.1 hEMP Induction from iPSCs . . . . . 116
4.5.2 Impaired T Cell Development in the ATO System from Low-Frequency hEMP Precursors . . . . . 117
4.5.3 Improving hEMP Differentiation . . . . . 118

4.6 r.RAG2 Rescues T Cell Development *In Vitro* . . . . . 120
4.6.1 r.RAG2 Edited Cells Demonstrate TCR Rearrangements Leading to Clearance of DP Block *In Vitro* . . . . . 120
4.6.2 Interrogation of iPSC Derived T Cell Receptors . . . . . 126
4.7 Discussion . . . . . 133

**5 Limitations** **137**

**6 Final Conclusions** **140**

**7 Appendices** **142**

# List of Figures


1.1 T Cell Development . . . . . 31

1.2 RAG2 Structure and Activity . . . . . 34

3.1 Design of Gene Editing Cassettes. . . . . 72

3.2 Testing and Location of sgRNAs . . . . . 74

3.3 Effect of Cellularity and MOI on Gene Targeting. . . . . 76

3.4 Further Effects of Increased Cellularity and MOI on Gene Targeting. 77

3.5 Effect of AAV Cargo Size on Genomic Integration. . . . . 79

3.6 Examination of Physiologically Relevant Developmental Markers. . . 81

3.7 HSPC Derived T cells Proliferate in Response to anti CD3/CD28 Antibodies and Display Diverse V$\beta$ Receptors . . . . . 83

3.8 Examination of RAG2 Deficiency Block. . . . . 85

3.9 AAV Causes Cytotoxicity in ATO System Panel A . . . . . 87

3.10 AAV Causes Cytotoxicity in ATO System Panel B . . . . . 88

3.11 HSPC Expansion Time Allows for Nonlinear Proliferation Increase . . 89

3.12 Electroporation Program Influences Cell Proliferation Capabilities of HSPCs . . . . . 90

3.13 Gentle Electroporation Program Reduces Cas9 On Target Activity . . 92

3.14 Single Cell Analysis of HDR Editing . . . . . 93

3.15 HDR Editing Deconvolved from Sanger Sequencing . . . . . 94

3.16 Optimized Healthy Donor ATOs Cell Types and Early T Cell Markers 96

3.17 Optimized Healthy Donor ATOs Mature T Cell Markers . . . . . 97

3.18 Optimized RAG2 Edited ATOs Cell Types and Early T Cell Markers 98

3.19 Optimized RAG2 Edited ATOs Mature T Cell Markers . . . . . 99

4.1 Confirmation of Pluripotency and G-Banding . . . . . 104
4.2 iPSC Nucleofection Optimization . . . . . 106
4.3 iPSC CRISPR/Cas INDEL Analysis . . . . . 108
4.4 iPSC Gene Editing and Sib Selection . . . . . 109
4.5 iPSC Genome Engineering Results . . . . . 110
4.6 PCR Verification of Cargo Insertion in Patient iPSCs . . . . . 112
4.7 Karyotype of Edited Patient iPSC Line . . . . . 113
4.8 WGS Off-target Analysis . . . . . 114
4.9 Overview of the ATO System for iPSC Input . . . . . 116
4.10 Preliminary hEMP Induction Results . . . . . 117
4.11 Preliminary Healthy Donor iPSC Derived T Cell Results . . . . . 119
4.12 Preliminary *RAG2* Deficient iPSC Derived T Cell Results . . . . . 120
4.13 Preliminary *RAG2* Edited iPSC Derived T Cell Results . . . . . 121
4.14 Optimized Induction of hEMP Intermediates . . . . . 122
4.15 Optimized Healthy Donor iPSC Derived T Cell Results . . . . . 123
4.16 Optimized *RAG2* Deficient iPSC Derived T Cell Results . . . . . 125
4.17 Optimized *RAG2* Edited iPSC Derived T Cell Results . . . . . 126
4.18 Comparison of Cell Line Differentiation Efficiency . . . . . 127
4.19 Single and Double Positive Sorting of iPSC Derived T Cells . . . . . 129
4.20 V-J Pairing in TCRs from iPSC Derived T Cells . . . . . 131
4.21 Comparison of V-J Pairing Between Experimental Patient iPSCs . . . 132
4.22 Complementary Determining Region 3 Data from Healthy Donor and Edited TCRs . . . . . 134

# Chapter 1

# Introduction

## 1.1 Primary Immune Deficiencies/Inborn Errors of Immunity

Primary Immune Deficiency (PID)s, also known as Inborn Errors of Immunity (IEI), comprise an array of genetic disorders which cause impairment in the development and/or function of the immune system (McCusker et al. 2018). Severity of IEIs can range from manageable to extreme, with more severe impairment requiring immediate treatment. The first IEI was formally recognized in the 1950s, yet retrospective analysis of patients dating back to the 1920s shows presumptive cases involving tell-tale IEI symptoms such as recurrent bacterial and viral infections (Ochs et al. 2012).

The discovery and categorization of systemic immune impairments hinged on the preceding discovery and characterization of the basic arms of the immune system. Going back as far as the 18th century, documents show experiments taking place on human subjects (Stiehm et al. 2005). These early experiments involved inoculating humans with biological material from animals infected with smallpox. Such

inoculation studies showed a marked decline in mortality from disease in inoculated humans, therefore suggesting that the toxins (later renamed antigens) present in infected tissues could transfer some kind of protection to a human host. At the time, the cause of disease was thought to be dependent on disturbances in the humors, consisting of phlegm, blood, and bile. An additional challenge to biological discovery was the lack of any treatment options to the infected population. Without clean water, antibiotics, and nourishing food, the deaths of many immunocompromised individuals were likely attributed to other causes. It wouldn't be until after the discovery of immune cells and proteins in the 1950s that the first IEI was identified as agammaglobulinemia at the Walter Reed Hospital (Ponader et al. 2014). By this time, medicine had advanced to the point where people were not dying as readily from bacterial infections, with the introduction and widespread use of drugs like Penicillin and sulfonamides. These medical advances allowed immunocompromised patients a greater chance to live long enough to be investigated by physicians, who were able to use the developing technologies at the time to study these patients more closely. Patient analysis at the time often consisted of fractionating blood samples for analysis. This approach was taken for the first patients diagnosed with Severe Combined Immunodeficiency (SCID). The first cases of SCID were recognized post-mortem in 1950, with identification of severe atrophy of lymphoid tissues associated with noted virtual lack of circulating lymphocytes ('lymphozythophthise') as reviewed in Kanegane et al. 2017. Shortly thereafter, it was appreciated that these patients also lacked immunoglobulins. Because the first cases were reported in Switzerland, and occurred both in males and in females, the disease was named

"Swiss-type agammaglobulinemia" to distinguish it from "Bruton's agammaglobulinemia", which had an X-linked inheritance. Also, unlike patients with Bruton's agammaglobulinemia, who typically manifested recurrent bacterial infections after 4-6 months of life (when maternally-derived IgG disappear), those with "Swiss-type agammaglobulinemia" presented with life-threatening infections sustained by various pathogens (bacteria, viruses, fungi) from the first weeks of life. As other patients continued to show similar characteristics affecting both humoral and cellular immunity, the disease was renamed SCID by the World Health Organization. It was also rapidly appreciated that SCID is genetically heterogeneous, and includes both X-linked and autosomal recessive variants.

In present day, we have an expanded toolkit which allows us to parse the genome of patients with suspected PIDs. We are able to now pair clinical presentation with next generation sequencing data to not only diagnose, but to discover new pathogenic mutations which shed further light on IEI function. Expanded use of animal and *in vitro* modeling has also led to a disruptive shift in our understanding of immunology and, when coupled with gene editing technologies, towards new treatment modalities which could improve, if not completely rescue impaired function.

### 1.1.1 Severe Combined Immunodeficiency

SCID is the most extreme form of IEI. SCID patients present with a block in the development of their adaptive immune system, leading to a lack of T cells. In some forms of SCID, B and/or NK cell development is also compromised (Fischer 2000). There are several genes critical to the development of the immune system, and so any mutation impairing the transcription of such lymphopoeisis related genes or the

function of their encoded proteins can cause impairment of downstream immune system development and function.

A frequent genotype reported for SCID is in the *Interleukin 2 Receptor Gamma (IL2RG)* gene, which causes X-linked SCID. The *IL2RG* gene encodes for the common gamma chain (gamma-c) protein which is shared by receptors for six different interleukins (IL-2, IL-4, IL-7, IL-9, IL-15, and IL-21), many of which are involved in the development, survival, proliferation and differentiation of T, B, and NK cells. Without functional interleukin signaling, the entire adaptive immune system fails to develop as it should (Puck et al. 1997). As a recessive X-linked disease this mutation most often affects patients with XY sex chromosomes.

Another common SCID causing mutation is in the gene in which a molecular defect was first reported, *Adenosine Deaminase (ADA)*. ADA is involved in breaking down purine compounds. An influx of purine (and in particular phosphorylated metabolites of adenosine and deoxy-adenosine) in the cell inhibits an enzyme called ribonucleotide reductase, which is responsible for generating the building blocks of DNA. Without effective DNA replication, the immune system cannot generate a diverse repertoire of thymocytes or respond to antigen challenges by expanding immune cells that can respond to the invading pathogen, thereby compromising the immune system (Flinn et al. 2018). The *ADA* gene is located on chromosome 20 and ADA deficiency is autosomal recessive.

The final SCID causing genes reported here are the focus of this thesis, *Recombination Activating Gene 1 (RAG1)* and *Recombination Activating Gene 2 (RAG2)*. The RAG1 and RAG2 proteins combine to form a heterotetramer with two copies of

each protein. The RAG heterotetramer binds to the Recombination Signal Sequence (RSS) that flank the Variable (V), Diversity (D) and Joining (J) gene segments at the T cell receptor (TCR) and immunoglobulin loci. The RAG complex initiates the process of V(D)J recombination, which is necessary to generate a large repertoire of T and B cells expressing distinct T and B Cell Receptors, respectively (Nishana et al. 2012). *RAG* mutations can result in several different clinical presentations depending on the identity of the mutation and as fully reviewed in Delmonte et al. 2018. Null mutations cause complete arrest of T and B cell development, and therefore are responsible for T- B- NK+ SCID. Patients with these mutations suffer from life threatening infections beginning shortly after birth and are in need of immediate treatment to reconstitute an adaptive immune system. In contrast, hypomorphic *RAG* mutations may allow for some residual development of T and B cells. A subgroup of these patients manifest Omenn syndrome, a condition characterized by generalized erythroderma, with infiltration of the skin and other organs by oligoclonal, activated T cells. Patients with RAG mutations causing Omenn syndrome lack circulating B cells, and have prominent features of immune dysregulation such as skin rashes, severe eosinophilia and elevated serum IgE. This condition leads to death early in life unless treated with an allogeneic hematopoietic stem cell transplantation. Finally, other patients with hypomorphic *RAG* mutations have a relatively milder phenotype, with delayed clinical onset, and clinical manifestations consisting of granulomatous lesions along with autoimmune problems and persistent infections.

### 1.1.2 Clinical Presentation and Diagnosis

For T-B- NK+ SCID patients, clinical presentation consists of early onset recurrent infection by viruses such as Cytomegalovirus (CMV), Herpes Simplex Virus (HSV), Respiratory Syncytial Virus (RSV), and influenza as well as fungal and bacterial infection including Candida, Aspergillus, and Pseudomonas (Niehues et al. 2010). Some RAG deficient patients can also present with eosinophilia along with indicators of infection including rash, diarrhea, and fever (Allenspach et al. 1993). Without V(D)J recombination happening in developing cells, lymphoid organs such as the spleen, lymph nodes, and tonsils are atrophic or underdeveloped. In particular, the thymus of RAG2 null patients shows a lack of architectural structure, with no Hassal's corpuscles or cortico-medullary distinction. Interestingly, allogeneic hematopoietic stem cell transplantation can rescue not only thymocyte development, but also maturation and differentiation of thymic epithelial cells in the atrophied thymic tissue (Til et al. 2014a), (Schuetz et al. 2014), (Haddad et al. 2018).

Diagnosis of SCID can be performed at birth or pre-birth with molecular diagnostics tools. In many states in the US and in other parts of the world, neonatal screening is employed for SCID detection, leading to faster diagnosis, more expedient treatment, and more favorable prognoses (Puck 2011), (Dezfouli et al. 2020), (King et al. 2018). The most common diagnostic tool for SCID detection is the analysis of T-cell receptor excision circles (TRECs). TRECs are a DNA by-product generated during V(D)J recombination in developing thymocytes. They are retained in naïve T cells that exit the thymus to populate the periphery, but are diluted at each T cell division. The TREC assay is performed by Quantitative PCR (qPCR), which

detects the amount of TRECs contained in a punch biopsy of a neonate dried blood spot (Borte et al. 2012). TREC levels correlate with thymopoietic activity, hence low or undetectable TREC levels are indicative of profound naïve T-cell lymphopenia, a hallmark of SCID. SCID diagnosis can be then confirmed by Complete Blood Count (CBC) and flow cytometry, with enumeration of T, B, and NK cells. As mentioned above, most patients with RAG deficiency manifest a T- B- NK+ phenotype. To tie the SCID diagnosis to a particular gene, molecular tools can be employed with use of a targeted gene panel of Whole Exome Sequencing (WES) or Whole Genome Sequencing (WGS) which can demonstrate single or bi-allelic mutations in *RAG2* or other SCID causing genes.

### 1.1.3 Issues with Standard of Care

Standard of care post SCID diagnosis typically takes one of three routes depending on the cause of the SCID: Hematopoietic Stem Cell Transplantation (HSCT), Enzyme Replacement Therapy (ERT), or gene therapy.

HSCT is the gold standard treatment for many types of SCID and has been in use for over 50 years (Heimall et al. 2017). HSCT consists of harvesting bone marrow or mobilized peripheral blood stem cells from a healthy donor and transplanting harvested cells into the immunocompromised patient. Because the bone marrow contains Hematopoietic Stem and Progenitor Cell (HSPC)s which are capable of differentiating into all blood cells, successful engraftment of HSPCs can solve many illnesses associated with either defects in blood cell development (autoimmunity and immunodeficiency) or blood cancers. Transplanted patient cells are allowed to differentiate in the body as they would during physiological development. Eventually,

the patient hematopoietic cells are replaced by those of the healthy donor. While HSCT has saved many lives, this treatment option carries many potential risks of complications.

One complication from HSCTs is related to donor/recipient HLA matching. The success of HSCT is dependent upon alignments in the Human Leukocyte Antigen (HLA) profiles of the donor and recipient (Yang et al. 2019). HLA typing is determined by a combination of alleles at hypervariable loci for HLA class I and class II antigens in the genome. HLA molecules are expressed on the surface of every nucleated cell and are critical for the restriction of antigen-specific adaptive immune response (Dendrou et al. 2018)s. The HLA system is what the body uses to discriminate self from non-self antigens. The best results of HSCT are obtained when there is full HLA match between the donor and the recipient. This is possible for matched sibling donors who receive the same HLA alleles from their parents. In contrast, HLA mismatching is associated with the risk of alloreactivity, which may translate into clinical manifestations of Acute Graft vs. Host Disease (aGvHD) and chronic Graft vs. Host Disease (cGvHD) (Gratwohl et al. 2017). The clinical manifestations of GvHD depend on the tissues in which the immune response is active and can range from diarrhea to liver damage to severe skin rash (in the case of aGvHD), and tissue fibrosis in the case of cGvHD (Blazar et al. 2020). GvHD can persist throughout the life of a patient and can be fatal. Donor-derived T cells play a critical role in the development of GvHD (Aversa et al. 2019). Therefore, in order to reduce the risk of GvHD in the case of HSCT from HLA-mismatched family donors, the donor's bone marrow is depleted of mature T cells. Alternatively,

donor-derived bone marrow or mobilized peripheral blood stem cells are positively selected. Matched Unrelated Donor (MUD)s are another source of hematopoietic stem cells wherein the donor and recipient are not genetically related. Although the donor and recipient are HLA-matched, nonetheless there is an increased risk of GvHD due to mismatching at minor HLA loci. Immunosuppressive drugs are used as a way to reduce the risk of GvHD after HSCT from a MUD (Gyurkocza et al. 2014).

Other complications from HSCT reflect use of pre-transplantation conditioning. This type of conditioning usually takes the form of chemotherapy, as host immune cells and precursors need to be eliminated in order for donor cells to engraft and proliferate - though there is debate on how much (if any) conditioning is needed for infants with SCID, and which chemotherapy agents are most safe and effective. Chemotherapy is associated with adverse conditions particularly in infants such as hepatic veno-occlusive disease (Malär et al. 2011), thyroid dysfunction (Slatter et al. 2004), and development of secondary malignancies later in life (Bomken et al. 2015).

Beyond matching, conditioning, and GvHD, the quality of immune reconstitution post-transplantation is another important factor that determines duration and quality of life after transplantation. For T-B- SCID patients, rescue of T cell development is usually a prerequisite for generation of B cells due to co-stimulation needed by T and B cell interactions (Haddad et al. 1998). Evidence of T cell reconstitution as noted by thymic activity, and the presence of circulating mature naive T cells is a good indication of success in engraftment, and an early presence of T cells post transplantation correlates with long term success (Sarzotti-Kelsoe et al.

2009).

Overall, there are many challenges with HSCT, but for most SCID patients this is the best treatment option. HSCT performed on infants aged 3.5 months or younger with a matched donor has a great success rate (>90% survival) and can successfully reboot the immune system leading to improved quality of life (Pai et al. 2014).

Enzyme replacement therapy (ERT) can be used in patients with SCID caused by ADA deficiency. ERT consists of weekly intramuscular injections of the recombinant ADA enzyme, conjugated with polyethylene-glycol (PEG), in order to increase solubility and stability of the enzyme. ERT has been shown to boost the immune function and reduce the frequency and severity of infections in patients with ADA deficiency (Weinberg et al. 1993). Unfortunately, this treatment is quite costly (projected to be 200,000 - 300,000 USD per patient annually (Chan et al. 2005)). In addition, there is evidence that immune function tends to decline long-term after initiation of ERT (Cagdas et al. 2018), (Chan et al. 2005). For these reasons, ERT is typically used as a place-holder, while preparing for HSCT or gene therapy.

Finally, gene therapies have been trialed as a treatment modality for SCID. The first IEI gene therapy took place over 20 years ago for ADA-SCID (Blaese et al. 1995). However, in that case, peripheral blood lymphocytes were used as targets, and the patient continued to receive ERT, which strongly limited the efficacy of gene therapy. A more robust attempt was performed a few years later in patients with X-linked SCID (Cavazzana-Calvo et al. 2000). Initially, gammaretroviral vectors were used to introduce the transgene of interest into the patient's hematopoietic stem

cells. Delivery of the common gamma chain transgene was driven by a retrovirus containing potent viral Long Terminal Repeat (LTR)s. The LTRs are viral DNA sequences which flank the transgene to be delivered. Sequences present in these retroviral LTRs contain enhancer sequences which can upregulate proto-oncogenes leading to tumorigenesis, specifically acute lymphoid leukemia (Hacein-Bey-Abina et al. 2003). Indeed, in the first series of 20 patients with X-linked SCID treated with this approach in Paris and London, six developed T-cell leukemia as a result of integration of the virus within or close to oncogenes.

This serious adverse event prompted the development of novel vectors. In particular, self-inactivating (SIN) retroviral vectors were developed, which remove enhancer functions of the viral LTR (Yu et al. 1986). Additionally, safer cellular promoters were introduced to drive expression of the gene of interest (Zhou et al. 2016). In the case of X-linked SCID, such a promoter was represented by the short form of elongation factor-1 alpha (Zhou et al. 2010). To further increase the safety profile, SIN-gammaretroviral vectors have been replaced by SIN-lentiviral vectors. While both gammaretroviruses and lentiviruses (another subset of retroviruses) are RNA viruses, retroviruses preferentially integrate at or near active enhancers, promoters, and CpG islands, which are all associated with actively transcribed regions of the genome (Wu et al. 2003), and thereby pose a higher risk for the development of insertional mutagenesis.

However, even with use of SIN-lentiviral vectors, there is a potential risk of leukemogenesis. This concern has promoted the development of alternative approaches to gene therapy, based on gene editing, as discussed later.

The cost of gene therapies should also be mentioned, which can range from 300,000 to 2.2 million USD.

## 1.2 T cell Biology and Ontology

The journey undertaken by T cells during development is complex. Indeed, the exact mechanisms responsible for some of the processes that occur during T-cell development are still ill-defined, yet many reviews exist covering known T cell development such as those in Germain 2002, Zuniga-Pflucker 2004, and Kumar et al. 2018. Hematopoietic progenitors are first found during human embryonic development in the yolk sac, a kind of bridge to the host's circulatory system during early stages of development and reviewed in Ross et al. 2020. This sac is then largely incorporated into the gut of the developing fetus, at which point these hematopoietic precursors migrate to the forming bone marrow as reviewed by Ciriza et al. 2013. The differentiation of these precursors then shifts either towards myeloid or toward lymphoid progenitor cells. The common lymphoid progenitor cells may further differentiate into T cells, or towards other lymphoid cells including B cells, which mature in the bone marrow. A series of signaling in the blood and bone marrow cause hematopoietic stem cells to migrate towards the thymus where they become Thymic Seeding Progenitors (TSP)s. TSPs enter the cortico-medullary junction of the thymus through a chemotaxi-slike process where they are deemed Early Thymic Progenitors (ETP)s. This point represents a bifurcation in differentiation, with ETPs either committing to T cell lineage specification or differentiating into natural killer cells or innate lymphoid cells. In humans, development of T cells is

first marked by the expression of CD7, CD5 and CD1a, followed by the expression of both CD4 and CD8 on so-called double-positive (DP) thymocytes Figure 1.1. Before this happens, developing thymocytes express neither CD4 nor CD8, and are therefore called double negative (DN) thymocytes. These cells are largely influenced by interacting with Notch ligands and signaling present in the thymus. Differentiating thymocytes then leave the corticomedullary junction and travel towards the subcapsular zone of the thymus. Traveling cells are exposed to stimulatory interactions with Cortical Thymic Epithelial Cell (cTEC)s and molecules like Interleukin 7 (IL7), which promote cell survival and rearrangement of the gamma/delta and beta T Cell Receptor (TCR) loci. At this point, RAG1 and RAG2 are expressed, allowing assembly of the pre-TCR complex. These pre-T cells begin to rapidly proliferate, while a subset of them commit to a $\gamma\delta$ T cell phenotype as opposed to an $\alpha\beta$ T cell phenotype Figure 1.1. Cells rearranging their TCRs then travel to the subcapsular zone of the thymus and undergo beta-selection, which requires developing cells to be able to signal through their functional pre-TCR complex consisting of a rearranged beta segment and an invariant pre-T alpha segment. Cells that progress through this phase of development are committed to T cell specification and begin to upregulate expression of CD4 and CD8 markers, thereby becoming human Double Positive (DP) cells. At this point, *RAG1* and *RAG2* gene expression is upregulated again, allowing rearrangements of the TCR alpha genomic segments to create a polyclonal repertoire of $\alpha\beta$ TCRs. DP cells test their newly formed TCR chains against peptides presented in the thymus by cTECs, dendritic cells, and fibroblasts in what is known as positive selection. Cells which have TCRs that bind

too strongly or too weakly to presented peptides on Major Histocompatability Complex (MHC) are eliminated, while those with intermediate binding strength proceed to the medulla. Because developing cells have TCRs that either bind preferentially to Major Histocompatability Complex I (MHCI) or Major Histocompatability Complex II (MHCII), that determines their final Single Positive (SP) cell fate. Cells which bind to MHCII become CD4 SP, and those which bind to MHCI become CD8 SP. After downregulating either CD4 or CD8 as appropriate, developing T cells now in the medulla face the final hurdle of negative selection. During negative selection, TCRs once again bind to peptides presented in the thymus by Medullary Thymic Epithelial Cell (mTEC)s expressing AutoImmune REgulator (AIRE). AIRE allows protein fragments from host-specific tissues to be presented on the surface of the cell to the developing T cells. If the TCR of a developing T cell binds too strongly to these peptides, such a cell has a high likelihood of attacking the host, causing autoimmune problems. Cells that bind too strongly to self peptides are eliminated by apoptosis (the so-called, negative selection process), so that the peripheral T cell repertoire is largely, albeit incompletely, purged of self-reactive T cell specificities. Cells that successfully proceed through negative selection then follow a chemical gradient out of the thymus and into circulation in the blood. Mature CD4+ cells include Tregs and Thelper cells, which recognize antigenic peptides via presentation by Antigen Presenting Cell (APC)s on MHCII molecules and provide support to the immune system by secreting molecules which enable the recruitment, modulation, and differentiation of immune cells. On the other hand, CD8+ T cells include cytotoxic cells which can directly lyse cells infected with a pathogen which can be

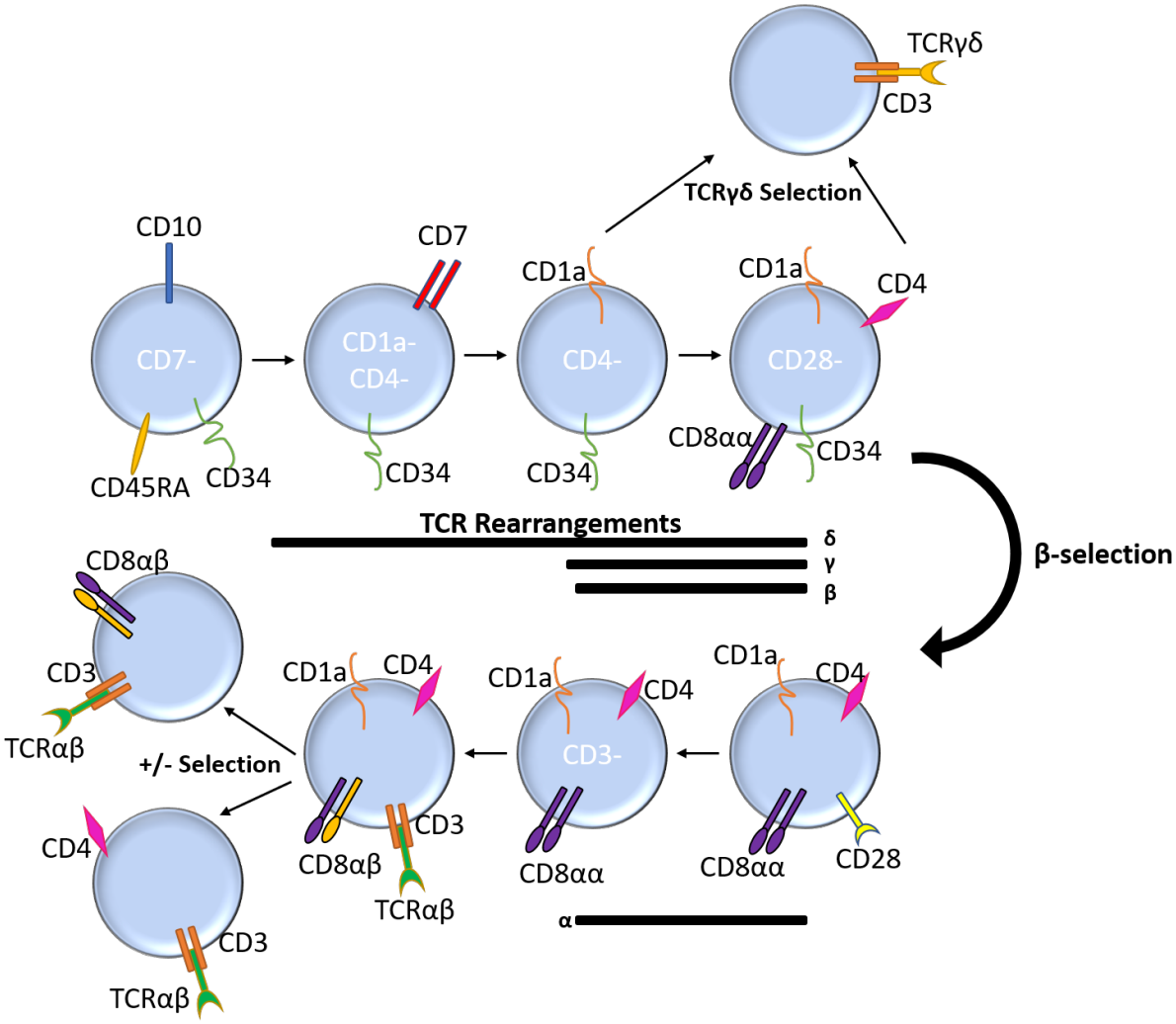



Figure 1.1: Illustration of the development, TCR rearrangement, and relevant surface markers of cells from the first point of entry to the thymus until release into the periphery as mature naive $\gamma\delta$ or $\alpha\beta$ T cells.

presented on MHCI molecules.

## 1.3 RAG2 Structure and Function

As discussed previously, RAG2 forms a heterotetrameric complex with its counterpart RAG1 to recombine all immunoglobulin and TCR loci during lymphocyte development. Without functional RAG1 or RAG2, this recombination cannot occur and developing lymphocytes arrest at this stage of development, causing SCID. The mechanisms of recombination for TCR loci will be described in the following chapter subsection. Here we will explore the domains of the RAG2 protein.

The human RAG2 protein is comprised of three domains and is 527 amino acids in length (Ward et al. 2018). The first domain from N to C terminus is the core domain and is built from the first 383 amino acids. The remaining amino acids contain the second and third domains which are within the non-core region of the protein and consist of the hinge domain and Plant Homeodomain (PHD) Figure 1.2 **A**.

The heterotetrameric RAG complex forms a 'Y' shape, with each copy of RAG2 forming one of the Y arms Figure 1.2 **B**. The RAG2 core domains create a six-bladed beta propeller which associates with two RAG1 domains, helping position the complex into a catalytically active state (M. S. Kim et al. 2015). Upon RSS recognition and DNA binding, the RAG complex further shifts in structure (Ru et al. 2015), with the RAG2 hinge domain allowing for extreme protein torsion to help facilitate V(D)J recombination (M. S. Kim et al. 2018).

Apart from aiding in conformational changes to the DNA-cleaving RAG1, the RAG2 PHD acts as a chromatin sensor to help ensure that V(D)J recombination is only active at an actively transcribed locus, which is demarcated by other histone

modifying proteins regulating cellular cycles and development (Ward et al. 2018). RAG2 forms an aromatic channel which binds to a trimethylated lysine at position four on histone 3 (H3K4me3), which is an epigenetic mark of active transcription and euchromatin, therefore helping regulate the activation of the recombinase to its appropriate expression time during lymphocyte maturation (Liu et al. 2007) (Shimazaki et al. 2009).

In addition to the enhancers that help regulate RAG expression during cellular differentiation, a further check on the activity of the RAG complex to preclude aberrant catalytic activity is the Post Translational Modification (PTM) of RAG2 through the cell cycle. RAG2 tyrosine residue 490 has been shown to be phosphorylated by cycline dependent kinase 2 during the G1 to S phase cell cycle cusp (Lee et al. 1999) with subsequent proteasomal degradation by polyubiquitination by the S-phase kinase-associated protein2/1-CUL1-F-box complex (Jiang et al. 2005). As such, the RAG2 protein can only be maintained in the cell during the G0 and G1 phases of the cell cycle. This limitation helps ensure that large genomic recombination events caused by the RAG complex are only active during the appropriate time of development and of cell cycle. Additionally, locking of the expression of RAG2 to the G0/G1 phase is interesting from a perspective of evolution, as the G0/G1 phase of cell division is where Non Homologous End Joining (NHEJ) predominates. NHEJ contributes to further diversity of TCR and BCR repertoires in the form of INsertion or DELetion (INDEL)s being added as a response to DNA damage during V(D)J recombination events (Mao et al. 2008).

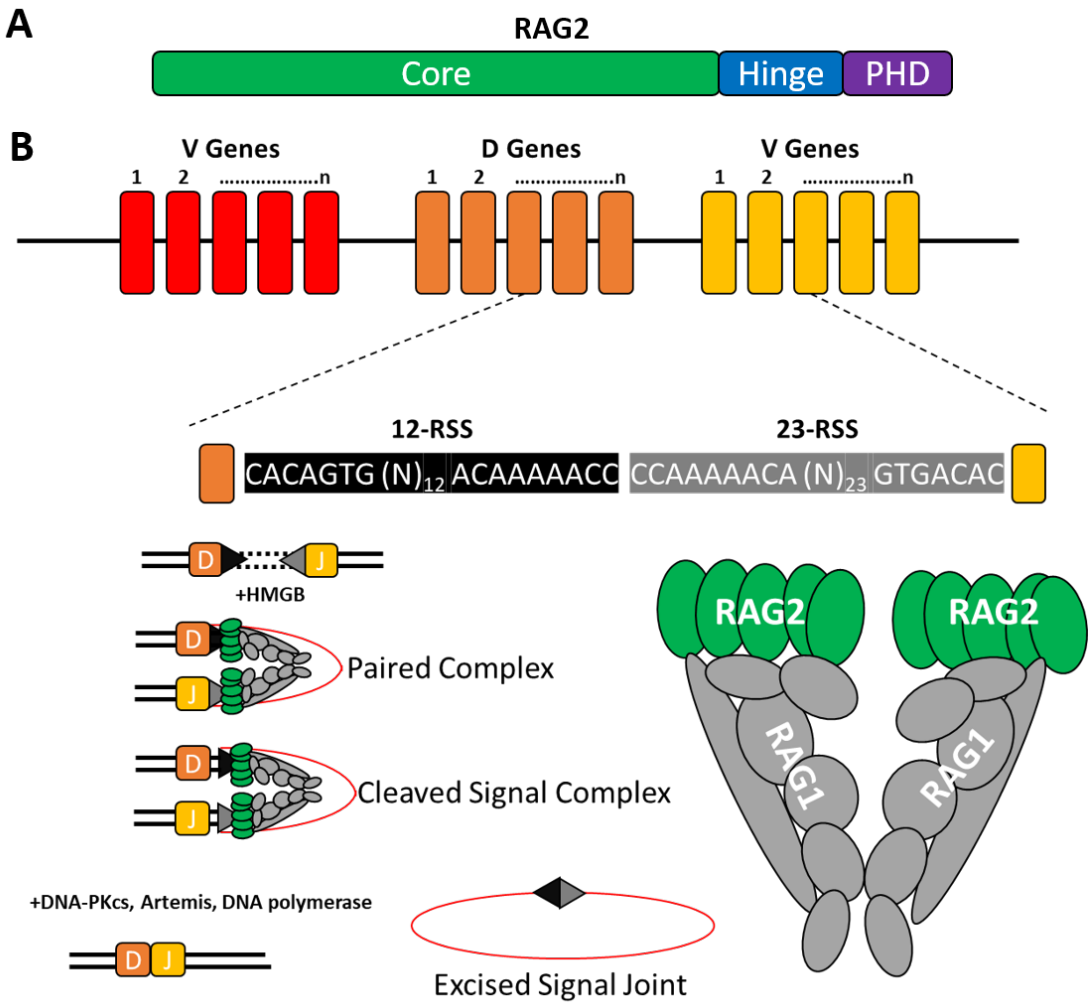



Figure 1.2: **Figure A** Map demarcating the three distinct components of the RAG2 gene. **Figure B** Illustration of V(D)J recombination and the RAG2 protein complexed with RAG1 to form the active heterotetrameric form.

### 1.3.1 TCR Repertoire Generation

In the human genome on chromosomes 7 and 14, variable (V), diversity (D), and joining (J) segments of DNA provide the cell with the fundamental blueprints to generate the TCR. A similar process is also established in B cells, though B cell

maturation is beyond the scope of this thesis. $\alpha\beta$ TCRs will be the focus of this thesis, as they make up roughly 95% of the TCRs present in the body.

When RAG1 and RAG2 first complex into their heterotetrameric active state during pre-T cell development, the complex scans these TCR genomic regions for V, D, and J segments which, in the first instance, correspond to the beta locus (Joachims et al. 2006). At the T cell receptor beta locus, the RAG complex recognizes a conserved nucleotide motif consisting of a heptamer of identity $CACAGTG$ and a nonamer of identity $ACAAAAACC$ separated by a spacer of either 12 or 23 nucleotides Figure 1.2 **B**. This motif is known as a Recombination Signal Sequence (RSS). These RSSs flank the V, (D) and J segments at all alpha, beta, gamma, and delta TCR loci, though of note the alpha and gamma chains do not contain diversity segments.

Within the beta locus, the RAG complex recombines the D and J segments first, with V(D)J recombination reviewed in Sadofsky 2001 and Mansilla-Soto et al. 2003. Post RSS detection, the complex binds randomly to the D and J regions and repositions them in physical space using the 12 or 23bp spacers to help orchestrate different segment pairings. RAG then nicks the DNA directly between the coding D/J segment and the RSS. This nick is then repositioned by the complex to attack the other strand of DNA, resulting in hairpin formation on the coding segment, and a blunt edge on the RSS segment. The blunt end piece of DNA containing all of the DNA in-between the D and J segment is then ligated and removed by a host of other proteins involved in this step of recombination including KU, DNA-Dependent Protein Kinase, XRCC4, ARTEMIS, Non-homologous End Joining 1 (CERNUNNOS/XLF)

protein, and DNA ligase 4. This excised signal joint called TRECs play no further part in TCR assembly though they represent useful diagnostic biomarkers in T-cell immunodeficiency and T-cell malignancies. The protein suite active in the excision of the lost fragment also facilitates joining of the D and J segments Figure 1.2 **B**. Once activated, ARTEMIS is responsible for opening the hairpins formed at the ends of these D and J segments. If ARTEMIS opens the hairpins from the center, it is thought that the ligation between the two segments happens in a blunt fashion. Otherwise, ARTEMIS opens the hairpin at an angle, exposing overhangs in the DNA within the hairpin, which allows for additional base pairing which can alter the amino acid sequence of the TCR being generated. Next, the open and nicked DNA strands are aligned and the DNA polymerase TdT adds additional bases at random to each strand, giving further diversity to the TCR repertoire. Finally, exonucleases act on the DJ segment to cleave any mispairing in the DNA generated by the recombination and polymerization process such that the two segments can be combined into one stretch of double stranded DNA with correct base pairing. This process is then repeated with the V segment being paired to the newly formed DJ segment within the beta locus.

The same sequence of events happens at the alpha and gamma loci, though the pairing begins and ends with the V and J genes, as alpha and gamma chain TCRs do not contain D genes. For $\alpha\beta$ T cells, the beta chain is rearranged first to create the pre-TCR complex with an invariant alpha chain. If this pre-TCR then survives beta selection, it will go on to proliferate extensively in the thymus in order to create as many opportunities as possible for the alpha chain of the TCR to also recombine

successfully to establish a functional TCR that passes both positive and negative selection. Of note, the alpha chain of the TCR does not recombine in an entirely stochastic manner. The study of patients with hypomorphic RAG mutations has shown that more downstream V and more upstream J genes (which are proximal to each other) are rearranged first, with distal V-J pairings observed more frequently in thymocytes that survive long enough and that have functional RAG proteins to allow multiple rounds of rearrangement (Villa et al. 2019), (Jouvin-Marche et al. 2009) and as reviewed in Jouvin-Marche et al. 2009. This process is how all TCRs are formed, and the primarily stochastic nature of the DNA repair process along with the V, (D), and J genes selected ensures that each immunocompetent person has a unique repertoire of TCRs which survive maturation and selection with the possibility of recognizing more than 300 billion targets (Jenkins et al. 2010).

## 1.4 Gene Editing

Accessible targeted gene editing is one of the most transformational technologies that has been developed in the past 20 years. After sequencing the human genome, scientists finally had a roadmap to begin to understand in more detail how genes are expressed, acted upon by *cis* and *trans* regulatory elements, and epigenetically modified. This information has played a critical role in elucidating many functions of the cell which were teased out by gene editing techniques using this genome roadmap as a primary tool. Additionally, gene editing has opened up an entirely new world of studying and ameliorating human diseases in both engineered cells and in animal models.

The first generation of engineered endonucleases was in Zinc Finger Nuclease (ZFN)s and Transcriptional Activator-like Effector Nuclease (TALEN)s. ZFNs are fusion proteins which consist of Zinc Finger domains, which can be engineered to bind specific DNA triplicates in an arbitrary combination, along with a FokI nuclease domain. The FokI nuclease is physically split onto two different zinc finger domains, which are engineered to bind to two separate but adjacent segments of DNA on opposite strands. When the ZFNs align on the target DNA sequence, the two halves of the FokI nuclease join to become catalytically active, cutting the target DNA at that site in a staggered manner (Urnov et al. 2010). ZFNs have many advantages including designed specificity, evolved sensitivity, derivation in a mammalian context which limits immunogenicity (no pre-existing adaptive immunity against ZFN proteins), and a short cargo length of approximately 1kb, allowing them to be packaged easily into clinically relevant Adeno Associated Virus (AAV) vectors. Downsides for ZFNs include engineering and manufacturing the zinc finger domains, which are not as accessible as other endonucleases for genome modification. Still, ZFNs have been further engineered and evolved to increase specificity, efficacy, and sensitivity (Paschon et al. 2019) (Miller et al. 2019)

TALENs are another designer endonuclease which work very much along the same lines as ZFNs. Instead of binding triplicate bases as the ZFN does, TALENS are comprised of DNA binding domains that recognize a single nucleotide (Romer et al. 2007). In this way, a string of DNA binding TALEN domains can be strung together with protein linkers to recognize any DNA sequence possible (Boch et al. 2009), (Moscou et al. 2009). Similarly to ZFNs, TALENs also fuse a split FokI

nuclease as the effector protein and take advantage of reverse strand binding of a second TALEN to become catalytically active and to further improve specificity. Stringing together each of these nucleotide binding domains rapidly creates a protein that is too large to be packaged into therapeutically viable vectors. The number of nucleotides that a TALEN protein recognizes can be reduced, reducing the size of the protein, but this reduction lowers specificity as fewer DNA binding domains means more potential off target sites. The typical functional and specific TALEN is about 3kb in size, which is still within the packaging limit of AAV, though without much room to spare to add any repair templates or reporter elements. TALENs have also been evolved and rationally redesigned to increase their specificity and targeted nuclease activity (Hubbard et al. 2015) (Sun et al. 2014).

Finally, the most recently deployed designer nuclease is Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-Cas9 technology. CRISPR is a bacterial immune system used to fend off invading bacteriophages attempting to infect the host cell, as reviewed in Barrangou et al. 2014. This defense system works by transcribing short guide-RNA sequences which complex with an endogenous nuclease (Cas9) to scan the genome for patterns of viral infection in the nucleotide space. Upon recognition, the RNA/nuclease complex cleaves the foreign DNA, which typically leads to cell death in prokaryotes. When the CRISPR/Cas system was discovered and redesigned for activity in cell types outside of its native context, the appeal of having an RNA-guided nuclease became very widespread due to the ease with which this tool could be used. As opposed to ZFNs and TALENs, modern CRISPR constructs require only a longer constant structural RNA called the trans-

activating CRISPR RNA (tracrRNA) and a variable region called the CRISPR RNA (crRNA) which can hybridize to any known sequence in the genome. After the first demonstrations of the CRISPR protein complex showing editing efficacy in human cells, the race began to use this new tool for discovery, engineering, and medicinal purposes. In mammalian cells, when a canonical CRISPR nuclease binds its target, the target strand is cleaved 3bp upstream of the derived organism specific conserved Protospacer Adjacent Motif (PAM), which is a short DNA sequence proximal to the target sequence. Post-cleavage, the DNA repair machinery of the cell is activated, leading to DNA repair by either NHEJ, which creates INDELs which can cause a frameshift mutation if in an open reading frame or splice junction, or by Homology Directed Repair (HDR), which can be taken advantage of to knock-in rationally designed DNA to the cut site. While the original Cas9 protein associated with CRISPR activity originated in Streptococcus Pyogenes (spCas9) (Mali et al. 2013), a variety of Cas homologues have been reported in other bacterial cells such as *Staphylococcus aureus* (Ran et al. 2015) and *Campylobacter jejuni* (E. Kim et al. 2017). These homologues have different characteristics including protein sizes, PAM sequence variations, and editing efficiencies. New homologues and orthologues are being reported and re-engineered every year, including Cas variants that can bind to and edit RNA (Freije et al. 2019), can induce nicks instead of doubles stranded breaks (Ran et al. 2013), and Cas fusion proteins that can up-regulate or down-regulate transcription of a target gene (Qi et al. 2013) (Maeder et al. 2013). Advantages for the CRISPR-Cas system of proteins are numerous not only in the effector functions described above, but also in diagnostic use, library screening, and

targeted genomic isolation for purposes such as high throughput sequencing. The ease of use is one of the primary draws; as opposed to ZFNs and TALENs which require designer protein sequences, CRISPR associated effector proteins in most cases require only the synthesis of a short RNA sequence which can be designed using a host of free webtools and ordered as either RNA or DNA from an array of accessible vendors. This allows for experiments to proceed much more quickly and for a much lower cost. One of the main cons of the CRISPR suite of proteins for clinical application is size restrictions for AAV incorporation, though shorter homologues are available such as that of *Campylobacter jejuni* which is about 2.5kb as opposed to the standard SpCas9 size of 4kb. Other clinical limitations include off-target activity of the guide RNAs which may cut at sites other than that which was intended, leading to potential oncogenic activity, as well as immunogenicity. As all known CRISPR proteins are derived from prokaryotes, the human body recognizes the repurposed therapeutic protein as foreign and can react in severe inflammatory signaling cascades. As such, many precautions need to be taken to use CRISPR proteins safely to edit human cells *in vivo*. Ongoing clinical trials involving CRISPR mediated genome modification are plentiful. China has been leading the way in clinical trial enrolment for CRISPR studies (clinicaltrials.gov). Prior *ex vivo* clinical CRISPR studies have shown efficacy in the treatment of sickle cell disease and $\beta$-thalassemia (Frangoul et al. 2020). The first *in vivo* application of CRISPR technology in the US is underway in a clinical trial run by Editas Medicine for treatment of Leber Congenital Amaurosis (clinicaltrials.gov). Preclinical data for this study has been promising (Maeder et al. 2019). Many more preclinical studies have demonstrated

exciting data promoting the use of CRISPR in the clinic for the treatment of blood based disorders including primary immune deficiencies, as reviewed in Surun et al. 2018. While not all therapies advance to the clinic, preclinical studies have shown the immense therapeutic potential contained within the CRISPR/Cas technology.

### 1.4.1 Gene Therapy for *RAG2* Deficiency

Gene therapy represents an exciting potential therapeutic option for patients suffering from *RAG* deficiency. Creating an *ex vivo* engineered autologous hematopoietic stem cell therapy for transplantation eliminates the need for finding matched donors, provides cells for editing immediately, and carries a lower risk of developing complications such as Graft Versus Host Disease (GVHD). To date, gene therapy has been performed to treat IEIs including ADA-SCID, X-linked SCID, Chronic Granulomatous Disease (CGD), and Wiskott-Aldrich Syndrome (WAS) with excellent efficacy and safety profiles (Thrasher et al. 2017). *RAG2* is an excellent candidate gene to add to this list. However, preclinical data have also highlighted potential limitations of this approach, as discussed below.

The results of prior studies aimed at elucidating the efficacy of replacing *Rag* genes in *Rag* deficient mice have been controversial. All prior work in mice has involved using retroviruses to integrate a functional copy of either *Rag1* or *Rag2* into the genome of murine HSPCs, rather than taking a targeted HDR mediated approach, whereby a double stranded DNA break causes a transgene to be inserted into the gnome at the site of the DNA break using homologous DNA sequences designed into the delivered DNA repair template. In one study, *Rag1* deficient

murine HSPCs were transduced with a lentivirus containing a functional copy of *Rag1*. The disparate heterologous promoters used in this work (PGK, SFFV, EFS, UCOE) drove variable restoration of lymphocyte development (Pike-Overzet et al. 2011). Another study similarly used a lentivirus to transduce *Rag1* deficient murine HSPCs with a functional copy of *Rag1* under promoters not used in the first study. The results of this transplantation showed inadequate reconstitution of T and B cells, along with autoimmunity and lymphocytic infiltrates in multiple organs (Til et al. 2014b), demonstrating some of the risks that can result from *Rag* gene transfer. When this lentiviral gene transfer strategy was applied to *Rag2*, better - though still sub-optimal - results were observed in *Rag2* deficient cells (Til et al. 2012). When a similar strategy was employed for correcting hypomorphic, rather than null *Rag2* mutations, reconstitution of T and B cell compartments remained relatively compromised (Capo et al. 2018).

Dysregulation of *RAG2* transcription may contribute both to this demonstrated sub-optimal lymphocyte rescue in *RAG* deficient cells as well as to the autoimmunity and lymphocytic infiltrates observed. As discussed above, *RAG2* is very strictly regulated at the transcriptional level as well as at the cell cycle and epigenetic levels. Genes delivered by a lentiviral vector must be accompanied by heterologous promoters to drive expression at the non-native site of integration, and as such cannot recapture physiological expression of target genes within their native locus.

AAV is a Single Stranded DNA (ssDNA) virus which does not integrate into the genome at significant levels, and therefore has a much lower risk of disrupting proto-oncogenes. AAV expresses its cargo episomally in the nucleus, and various serotypes

of AAV have been reported with specific tropism, broadening the uses of AAV to infect targeted tissues (Srivastava 2016). As a result, AAV has great promise for *in vivo* editing as well as *ex vivo* editing. When paired with CRISPR-Cas mediated double-stranded DNA breaks, the ssDNA cargo delivered by AAV can be used as a repair template to insert the transgene of interest into a precise location. Targeted insertion of a transgene into its endogenous locus allows for potential reestablishment of its native transcriptional landscape within that locus, including use of native promoters. As such, we are hopeful that a targeted promoterless *RAG2* gene transfer can retain wild-type-like transcriptional profiles during lymphocyte development, yielding efficient reversal of the SCID phenotype in human *RAG2* deficient cells without the challenges observed in prior studies.

## 1.5 iPSC Modeling

Coupled with gene editing advancements, the use of Induced Pluripotent Stem Cell (iPSC) and Embryonic Stem Cell (ESC) models has drastically advanced knowledge of the molecular underpinnings and treatment modalities available for a wide array of genetic diseases. The first human Pluripotent Stem Cell (PSC) line was reported in 1998 (Thomson et al. 1998). This cell line was able to differentiate into all three of the human germ layers preceding development of all known tissues. Since the demonstration of this technique to generate PSCs from fetal tissue, new technologies to generate PSCs have been developed. Such technologies include transplantation of the genome of a somatic cell into an enucleated oocyte (somatic cell nuclear transfer) (Tachibana et al. 2013) and from expressing a series of tran-

scription and reprogramming factors (Yamada et al. 2014) within a somatic cell type to transcriptionally reprogram the cell. As PSCs are primary cells which can renew indefinitely and differentiate into all known tissue types, they provide an excellent vehicle to study human pathogenesis.

With new PSC derived modeling technologies, human cell physiology can now be interrogated directly in many cases as opposed to relying solely on animal models and immortalized human cell lines, which differ physiologically from primary cells (Kaur et al. 2012). One fundamental strength of using PSCs to model disease is the ability to reprogram patient cells with known clinical phenotypes to pluripotency, and to then use those cells to derive and interrogate a cell-type of interest. Deriving PSCs from patients who present with an actual illness allows the study of the disease phenotype directly in a Petri dish. To further confirm the disease-causing effect of that specific gene variant, iPSC cells from healthy donors can be engineered with CRISPR-Cas to introduce the mutation of interest, and appropriate studies can then be performed to confirm that the same cellular phenotype as in the patient's iPSC line is observed, though this does not take into account potential epistatic events. Finally, CRISPR-Cas or other programmable nucleases can be used to correct such mutations in patient-derived iPSCs, which should result in phenotypic rescue *in vitro* in a perfectly isogenic model. Genetic diseases can produce various genotype phenotype dissimilarities which are difficult to quantitate in the sequence space. Deriving a model from actual patients helps preclude this phenotype genotype discrepancy from being a confounding issue.

Another very useful purpose for using iPSCs as a cell model is that they can

remain in culture indefinitely. This is similar to immortalized cell lines, but dissimilar in the fact that immortalized cells which are often cancer cell lines can contain genomic aberrations which cause polygenic variation and divergence from transcriptional similarity to primary tissues. PSCs are primary cells, but because they can divide indefinitely they provide an excellent chassis to perform all sorts of genome engineering that takes several iterations of modification to complete. Modifying somatic cells is often limited due to cellular senescence and cell death, so the fact that PSCs continue to grow means that complicated genetic manipulation can be performed over the course of many experiments. Unlimited cell division is also useful in regaining a sense of cellular homeostasis post-genome editing, when the cell is under stress from DNA damage and repair pathways are being activated. This allows the cell to incorporate rationally designed changes into the genome and reset before differentiation into the targeted tissue type. Indefinite renewal also allows for genetic rescue studies, wherein patient iPSCs are corrected and analyzed to see whether the designed and incorporated edits have established a healthy phenotype in the cell.

The advancement of the PSC modeling field has generated an array of cell media formulations which can be used to derive numerous target cell types. The list of derivable cell types from PSCs expands every year. Further refinement in disease modeling with PSCs has involved the hybridization of fields like molecular biology with engineering to produce more sophisticated models of complex heterogeneous tissues by combining microfluidics and PSCs along with other cell types to become what is known as an organoid, or alternatively an organ on chip device (Sances et

al. 2018). Organoids typically involve co-culturing or co-deriving various cell-types from iPSCs due to sequential or targeted exposure of cells to cytokines and growth factors which would be present in the surrounding environment in the body during physiological development of the targeted tissue type. These technologies have allowed the derivation of more specific terminally differentiated cells and of clusters of cell-types which colonize whole organs, thus providing another tool for drug discovery and ontological interrogation *in vitro*. Eventually, PSCs will move beyond uses in studying diseases, and will be used to derive and transplant therapeutically viable cells.

### 1.5.1 Overview of the Artificial Thymic Organoid System

The Artificial Thymic Organoid (ATO) system allows for the generation of mature naive T cells from either HSPCs or iPSCs by recapitulating the 3D thymic environment. This technology was first published by the lab of Gay Crooks from UCLA in 2018 using Hematopoietic Stem Cell (HSC) precursors to generate mature naive T cells (Seet et al. 2017). Derivation of T cells from hematopoietic cells was originally reported as a monolayer co-culturing system by Juan Carlos Zuniga-Pflucker's lab in 2002 (Schmitt et al. 2002), though problems with robustly generating mature single positive cells hindered broad applicability of the system. Taking this technology further, Zuniga-Pflucker's lab also demonstrated generation of T cells from human embryonic stem cells (Schmitt et al. 2004). One of the greatest takeaways from Zuniga-Pflucker's pioneering technology was the observation that Notch signalling seems to be essential for T cell lineage commitment from precursors (Dervovic et al. 2013) *in vitro*. Because there has been limited progress in propagating and immor-

talizing human thymic epithelial cells, a murine stromal cell line was engineered to express the human Delta-like 1 (DLL1) Notch ligand to help commit precursors to T cell etiology.

Expanding on this technology, the lab of Gay Crooks attempted to recreate the 3D environment of the thymus using an organoid co-culturing system. Similar to Zuniga-Pflucker's work, they also engineered a murine stromal cell line to express human Notch ligands DLL1 as well as Delta-like 4 (DLL4). The primary difference between these two technologies is that the ATO system makes use of the 3D environment to expose developing thymocytes to more plentiful signaling which may more accurately mimic thymic structures. Indeed, the ATO system more robustly produces mature naive T cells co-expressing TCR$\alpha\beta$ and CD3 along with single positive CD4 and CD8 cells, the former of which were not robustly demonstrated by prior technologies *in vitro*. Additionally, the ATO system encapsulates the developing T cells in a kind of stromal cell "skin" which sits upon a membrane and allows fluid exchange between the media placed in a well, and the membrane sitting on that well of media. Such a discrete system allows for a more delicate approach to development when compared to the monolayer system, which requires the cultivator to manually pipette developing cells during wash and media change steps potentially disrupting or destroying developing T cells.

The ATO system can derive T cells from HSCs and from iPSCs with slightly different protocols. The HSC input approach allows for the generation of mature naive T cells from a variety of HSC origins including umbilical cord blood, mobilized peripheral blood, and bone marrow samples, with umbilical cord blood derived HSCs

tending to perform the best. The HSC data generated in this thesis is from mobilized peripheral blood and bone marrow derived HSCs, which were sorted for CD34+ cells and CD3- cells to ensure a lack of artificial expansion of any contaminating mature T cell already present in the HSC vials. Purified HSCs are then cultured at a predefined ratio of HSC to engineered DLL4 expressing murine stromal cell and are exposed to the relevant cytokines and growth factors necessary to stimulate T cell development as detailed in the materials and methods section.

The protocol shifts slightly when incorporating iPSCs as the starting material, as there are additional cellular derivations required to differentiate PSCs into the HSCs used in the description above. In essence, a Human Embryonic Mesodermal Progenitor (hEMP) cell state is required for seeding into the ATO system. The mesoderm is one of the three germ layers derivable from PSCs. The mesoderm is the middle layer of tissue development in gastrulation and fetal development and the lateral plate mesoderm/hemangioblast cell type is what gives rise to all lymphoid cells including HSCs and subsequent T cells (Abramson et al. 1977). This cell state is indicated by the presence of CD56 and the absence of EpCAM, as will be discussed in detail in Chapter 4. These intermediate cells are then seeded with an increased ratio of hEMP cell to murine stromal cell to provide additional Notch signaling to developing cells to influence commitment to the T cell lineage. Further details can similarly be found in the materials and methods section.

## 1.6 Aims and Objectives

The objective for this research project is to investigate the possibility of creating an autologous engineered cell therapy for patients with RAG2 deficiency. An autologous approach would eliminate or abrogate many of the challenges resulting from allogeneic HSCT including issues finding matched donors and developing GVHD, the combination of which we hope will increase the likelihood of establishing robust immune reconstitution. Furthermore, maintaining endogenous regulation of gene expression would eliminate the risk of dysregulated RAG protein expression which could occur when using a heterologous promoter for lentivirus-based gene therapy. To this end, we have taken an *ex vivo* editing approach to knock in a promoterless *RAG2* cDNA construct at the endogenous locus using AAV mediated CRISPR/Cas9 genome engineering.

There are three legs to this project. The first is to examine and characterize the gene editing pipeline in patient iPSCs. iPSCs can renew indefinitely, and can therefore be used to most robustly characterize genome editing and select clones for more targeted analyses. To examine rescue of patient edited cells, the ATO system was used to generate mature naive T cells from patient RAG2 deficient iPSCs, patient RAG2 gene edited iPSCs, as well as an immunocompetent iPSC control line.

The second leg aims to discover whether the approach established in leg one can be adapted to HSCs. HSCs would be the target cell type for HSCT, and so demonstration of T cell developmental rescue in the ATO system with gene edited cells along with unedited isogenic precursors and immunocompetent control cells is

critical.

The final leg of this project has to do with showing reconstitution of the immune system in T-B- NOD-SCID-Gamma (NSG) mice transplanted with edited RAG2 deficient patient HSPCs. Transplantation and engraftment studies were carried out by our collaborators in Dr. Matthew Porteus' lab at Stanford, and so will not be covered in detail in this thesis work.

# Chapter 2

# Materials and Methods

## 2.1 Culturing and Engineering of Cell Lines

### 2.1.1 sgRNAs and Plasmids

All sgRNAs for gene targeting for this work were designed using the Benchling webtool and are listed in Appendix Table 7.2. Potential sgRNAs were screened using the CRISPR associated nuclease specific PAM sequence as a search reference within the target nucleotide space of interest. The Benchling webtool then applies an algorithm to score each 20bp sgRNA on its on target projected activity and its off-target projected activity. Highly scored sgRNAs also within a nucleotide reference frame of highest interest were then ordered from IDT with adaptor sequences to clone into a pX459 plasmid backbone (Addgene No. 62988) using a BbsI (New England Biolabs) restriction enzyme cut site. The pX459 plasmid backbone contains a U6 promoter to express the sgRNA cloned into it, as well as a high fidelity spCas9 variant under a CMV promoter. The pX459 plasmid also contains a Puromycin marker for plasmid selection in mammalian cells. High scoring sgRNAs assembled into this plasmid

were then lipofected into HEK293T cells to examine their cutting activity as measured by INDELs quantitated by sanger sequencing. Those sgRNAs which have the highest on target activity create the most frequent incidence of mismatched DNA repair products (INDELs) which are detectable by PCRing and sequencing around the target cut site. The same locus is PCRd from a sample which did not receive the sgRNA and Cas9 complex as a negative control. Free webtools such as ICE from Synthego (https://ice.synthego.com) or TIDE (https://tide.nki.nl/) can extract sequencing data from the experimental sample and negative control to infer sgRNA activity by the frequency of INDELs. Highest scoring sgRNAs were then purchased from Trilink Biotechnologies (San Diego, CA, USA) and were HPLC purified. All sgRNAs ordered from Trilink also contained three 2'-O-Methyl phosphorothioate modifications at the 5' and 3' ends. The 20bp CRISPR target sequence used for this work was of the sequence 5'-TCAGAAACTATGTCTCTGCA-3'. All other CRISPR target sequences for sgRNA design can be found in Table 7.2.

The only other plasmid used in this study was the pMaxGFP plasmid (VDC-1040) supplied by Lonza in their electroporation kits which used to optimize electroporation conditions.

### 2.1.2 iPSC Culture

iPSCs used in this study were cultured on Vitronectin (Thermo Fisher) coated tissue culture plates in Essential 8 Flex Medium (Thermo Fisher). Cells were split upon reaching 80% confluency using TryplE Express. Briefly, cells to be split were washed 2X with DPBS, then incubated with TryplE Express for 6 minutes at 37C. Trypsinized cells were then diluted with an additional 2X volume of DPBS be-

fore being centrifuged at 300xG for 5 minutes at room temperature. Pelleted cells were then resuspended in Essential 8 Flex Media containing 10uM ROCK Inhibitor (Tocris Bioscience, Bristol, UK). Cells were passaged at a 1:10 ratio and cell culture media was changed every two days. Cell viability and cellularity was assayed using a Countess II FL Automated Cell Counter (Thermo Fisher) along with a 0.4% Trypan Blue Stain (Thermo Fisher).

### 2.1.3 iPSC Editing

All single guide RNAs (sgRNAs) used in this study were purchased from Trilink Biotechnologies (San Diego, CA, USA) and were HPLC purified. All sgRNAs also contained three 2'-O-Methyl phosphorothioate modifications at the 5' and 3' ends. The 20bp CRISPR target sequence used for this work was of the sequence 5'-TCAGAAACTATGTCTCTGCA-3'. Recombinant High Fidelity Cas9 protein in supplied buffer was used in all editing experiments from Integrated DNA Technologies. All RNPs used for editing were created using a 1:3 Cas9:sgRNA ratio (3ug Cas9 protein to 1.3ug sgRNA) and were allowed to complex at room temperature for 30 minutes prior to electroporation.

$1x10^6$ iPSCs were resuspended in 20/uL OPTI-MEM (Thermo Fisher) and then combined with the RNP and inserted into a single well of a nucleovette strip (Lonza). Cells were electroporated using a Lonza 4D Nucleofector (program CA137) and all $1x10^6$ cells were then recovered in one well of a 6-well plate coated with Vitronectin (Thermo Fisher) in Essential 8 Flex Medium (Thermo Fisher) supplemented with 10uM ROCK Inhibitor (Tocris Bioscience, Bristol, UK). Plated cells were immediately transduced with rAAV6 at 50,000-250,000 vector genomes/cell as tittered by

qPCR from the manufacturer. AAV6 virus was produced, analyzed, and tittered by Dr. Mara Pavel-Dinu in the lab of Matthew Porteus at Stanford University.

### 2.1.4 Analysis of INDEL Frequency and Identity

INDEL frequencies and identities were examined using the ICE tool from Synthego. Genomic DNA from cells nucleofected with the target RNP was isolated 2 days post-electroporation. Harvested genomic DNA was then PCRd using primers forward: TIDE RAG2 F 5'-ATGTGGTTCTTTCAGCTGACG-3' and reverse: TIDE RAG2 R 5'-CGAAAAGTAACCTTTTTGTTGT-3' producing a 1kb amplicon. This PCR product was then cleaned up using a PCR Cleanup Kit (Qiagen) before being submitted for Sanger sequencing using the following primers (also used for PCR): forward: TIDE RAG2 F 5'-ATGTGGTTCTTTCAGCTGACG-3' and reverse: TIDE RAG2 R 5'-CGAAAAGTAACCTTTTTGTTGT-3' with an appropriate mock control genomic DNA sample used as the baseline for *in silico* deconvolution. All primers are reported in Appendix Table 7.1.

## 2.2 PCR, ddPCR, qPCR

Analysis of the percentage of cells with successful integration of the RAG2 cassette was carried out by a 5' junction ddPCR using the National Institutes of Health Core Genomics Facility. Genomic DNA extracted from all target cells was purified using a DNEasy Blood and Tissue Kit (Qiagen). 100ng of purified gDNA was then combined with WT FAM Probe Rag2sg1v2 dd ref FAM: 5' – CCCGAGGAACGTGACCATGGAGTGGC – 3' along with forward primer Rag2sg1v2 dd ref FAM F: 5' – GCACAGGAAGTTTAGCAGTG – 3' and reverse primer Rag2sg1v2

dd ref FAM R: 5' - GGGAATTCAAGACGCTCAGA – 3' and MUT HEX Probe, Rag2sg1ddIntegratedHEX: 5' – TACCATCAGAGACCATGTCACTCCAAATG – 3' along with forward primer Rag2sg1 dd Int HEX F: 5' – GCACCTTCGGCTAGTCTTTA – 3' and reverse primer Rag2sg1 dd Int HEX R: 5' – TCATGAGAGAGAAACCCGGT – 3' at a primer/probe ratio of 900nM/250nM and in a total reaction volume of 22uL including 11uL ddPCR Supermix for Probes (No dUTP) (BioRad). Droplet generation and PCR/reading was performed by the Genomics Core with cycling parameters of: 95C (10 minutes), [94C (30 seconds), 61.7C (30 seconds), 72C (2 minutes) – repeated steps in brackets 50 times], 98C (10 minutes), -4C (indefinite) on a QX200 Droplet Digital PCR System (Bio-Rad).

Additional junction PCRs were used to verify full insertion of the r.RAG2 cassette (see Appendix Table 7.1). 3' junction PCR primers were used to assay full target integration, amplifying a 1kb region spanning the 3' end of the cargo onto the adjacent genome beyond the 3' homology arm sequence. The primers used to verify this insertion were; forward: RAG2 int 3' junc F v1 5' - GCGTGATCTTCACCGACAAG - 3' reverse: RAG2 int 3' junc R v1 5' - GCGTGATCTTCACCGACAAG - 3'. We also assayed integration events on both alleles in clonal populations by PCRing over the insertion site. This design produced a 1kb band when the cargo was not inserted, and no band when the cargo was inserted due to the extension time of the PCR not being adequate to amplify the length of the integrated cargo. Primers for this check were of the sequence; forward: ver RAG2 KI F v1 5' - CTCTCCCAGAGCAACAAAGAC - 3' reverse: ver RAG2 KI R v1 5' - CAGGCAGTCAGCTACACTATTC - 3'. All PCR reactions used KAPA HiFi Polymerase (Kapa Biosciences)

using cycling conditions: 95C (3 minutes), [98C (20 seconds), 60C (30 seconds), 72C (1 minute) - repeated steps in brackets for 35 cycles], 72C (2 minutes), 4C (indefinite) on a T100 thermal cycler (Bio-Rad). All other oligos used in this study are included in Appendix A within Table 7.1.

qPCR was used to verify expression of pluripotency markers *OCT4*, *NANOG*, *ABCG2*, and *DNMT3B* in reprogrammed patient iPSCs and negative control patient fibroblasts. Pluripotency irrelevant housekeeping genes *ACTB* and *MYC* were used as controls. Sequences of qPCR primers can be found in Table 7.1.

### 2.2.1 iPSC Reprogramming and Karyotyping

Fibroblasts were cultured from a skin biopsy specimen obtained from the RAG2 deficient patient. Primary fibroblasts from patient and control foreskin fibroblasts (ATCC, Manassas, VA) were reprogrammed into iPSCs by Kerry Dobbs at the National Institutes of Health using a CytoTune -iSP 2.0 Sendai Reprogramming Kit (ThermoFisher) following kit instructions for feeder-dependent fibroblasts under IRB-approved protocol 16-I-N139. iPSC were analyzed by G-band karyotyping (Cell Line Genetics, Madison, WI) to ensure genomic integrity.

## 2.3 Immunosequencing

TCR sequencing was conducted by Adaptive Biotechnology (Seattle, WA, USA) using their immunoSEQ assay service and analyzer. <1ug of genomic DNA from CD3+ sorted ATO derived T cells was submitted to their facilities for sequencing. Sample data was generated using a two-step amplification bias-controlled multiplex PCR strategy to amplify each V and J gene. Amplicons were then amplified again

to adapt barcodes for Illumina next generation sequencing. Raw demultiplexed data was downloaded and analyzed using R to generate heatmaps. Briefly, recombination events between genes in the *TCRA* and *TCRB* regions were imported separately into the R programming environment. Once imported, the number of recombination events observed between each pair of genes was transformed into a percentage of total recombination events observed. Genomic loci for genes was downloaded from NCBI and genes were ordered by their chromosomal position. To aid in visualization the percentage of recombination events per gene pair was scaled by square root transformation. Software program Treemap version 2019.4.2 was used to generate treemap illustrations of CDR3 identities and frequencies within the sequenced population. The Shannon's Entropy Index score was tabulated using software program Past4.03.

### 2.3.1 WGS of Edited and Unedited iPSCs

Whole genome sequencing of the precursor and edited cell lines was performed by Justin Lack at the National Institutes of Health. Samples were sequenced to a median depth of 30X using the Illumina NovaSeq 6000 system. Reads were trimmed using Trimmomatic v0.39 and mapped using BWA-MEM v07.17 to the human hg38 reference genome with the inserted construct sequence added as an additional contig (Bolger et al. 2014). PCR duplicates were marked using Samblaster v0.1.2.5 (http://broadinstitute.github.io/picard/), and GATK v4.1.9.0 was used to perform base recalibration (Faust et al. 2014). To identify potential off-target mutations, Single Nucleotide Polymorphism (SNP)s and INDELs were called in a paired fashion using MuTect2 from GATK v4.1.9.0 and following the GATK Best Practices

(https://gatk.broadinstitute.org/hc/en-us/articles/360035531132–How-to-Call-somatic-mutations-using-GATK4-Mutect2) (Cibulskis et al. 2013). To reduce false-positive calls, variants were filtered with the following criteria: edited.sample.depth $> 15$, edited.sample.alt.count $> 5$, edited.sample.freq $> 0.1$, precursor.sample.depth $> 15$, precursor.sample.alt.count $= 0$.

### 2.3.2 HSPC Culture and Expansion

Donor HSPCs were thawed into a rich cytokine Stemspan SFEMII Media (Stem Cell Technologies) supplemented with SCF (100ng/mL Peprotech), TPO (100 ng/mL Peprotech), Flt3-Ligand (100ng/mL Peprotech), IL-6 (100ng/mL Peprotech), UM171 (35nM Stem Cell Technologies), 1X Primocin (100ug/mL Invivogen), and StemRegenin1 (SR1) (0.75uM Stem Cell Technologies). Cells were expanded at a density of $0.25x10^6$ cells/mL in 96 well v-bottom plates (Corning) at 37C and 5% CO2. Cells were counted after 48 hours and if they had expanded beyond $1x10^6$ cells/mL they were split further back to $0.25x10^6$ cells/mL. Expanded cells were used for gene editing and as an input to the ATO system. Cell viability and cellularity was assayed using a Countess II FL Automated Cell Counter (Thermo Fisher) along with a 0.4% Trypan Blue Stain (Thermo Fisher).

### 2.3.3 HSPC Expansion and V-beta Profiling

The V$\beta$ profiles of HSPC derived T cells was assayed by flow cytometry using the IOTest Beta Mark TCR Repertoire Kit (Beckman Coulter). This kit contains eight vials each containing three monoclonal antibodies against one of the 24 included V$\beta$s assayed which represent about 70% of possible V$\beta$s in human cells.

Each of the three antibodies in a vial is either conjugated to a PE fluorochrome, a FITC fluorochrome, or a dual labeled FITC/PE fluorochrome. T cells derived from healthy donor HSPCs in the ATO system were benchmarked against normal donor Peripheral Blood Mononuclear Cell (PBMC)s by staining 8 separate vials of cells per group which allows identification of all 24 V$\beta$s. Cells were also stained with anti-human CD3 PE-Cy7 clone OKT3 (Biolegend) and anti-human CD45 V500 clone HI30 (BD Biosciences) to ensure that V$\beta$s were only evaluated for CD3+CD45+ T lymphocytes.

Proliferation of cells was conducted using soluble anti-CD3 CD28 IL2 stimulation quantified by CellTrace Violet (Thermo Fisher) dilution. CellTrace Violet staining solution was added to harvested cells resuspended in 500uL PBS. Cells were incubated with CellTrace Violet for 20 minutes at 37C before CTS OpTmizer T cell expansion SFM (Thermo Fisher) was used to wash and resuspend stained cells. T cells were then stimulated with anti-CD3 CD28 antibodies along (Thermo Fisher) with 100U/mL IL-2 (R&D Systems) and analyzed via flow cytometry after five days.

### 2.3.4 HSPC Editing

The sgRNAs used for HSPC editing were of the same format as descibed for iPSC editing in subsection 2.1.3. Additional sgRNAs tested are included in Appendix A Table 7.2. Recombinant High Fidelity Cas9 protein was used in all editing experiments from Integrated DNA Technologies. All RNPs used for editing were created using a 1:3 Cas9:sgRNA ratio (3ug Cas9 protein to 1.3ug sgRNA) and were allowed to complex at room temperature in manufacturer supplied buffers for 30 minutes prior to electroporation.

$1x10^6$ HSPCs were resuspended in 20/uL OPTI-MEM (Thermo Fisher) and then combined with the RNP and inserted into a single well of a nucleovette strip (Lonza). Cells were electroporated using a Lonza 4D Nucleofector (program DS-112 and DZ-100) and all $1x10^6$ cells were then recovered in one well of a 6-well plate in Stemspan SFEMII Media (Stem Cell Technologies) supplemented with SCF (100ng/mL Peprotech), TPO (100 ng/mL Peprotech), Flt3-Ligand (100ng/mL Peprotech), IL-6 (100ng/mL Peprotech), UM171 (35nM Stem Cell Technologies), 1X Primocin (100ug/mL Invivogen), and StemRegenin1 (SR1) (0.75uM Stem Cell Technologies). Plated cells were immediately transduced with rAAV6 at 50,000-250,000 vector genomes/cell as tittered by qPCR from the manufacturer. Transduced cells were seeded into the ATO system and assayed for HDR events at various timepoints as discussed in chapter three of this thesis work. AAV6 virus was produced, analyzed, and tittered by Dr. Mara Pavel-Dinu in the lab of Matthew Porteus at Stanford University.

### 2.3.5 Murine Stromal Cell line Culture

The murine stromal cell line (MS5) edited to ectopically express human Notch ligand, delta-like 4 (hDLL4) was used as the hEMP co-culture cell type for the ATOs. MS5-hDLL4 cells were cultured in Dulbecco's Modified Essential Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) (Thermo Fisher Scientific, Waltham, MA, USA) and 100ug/mL Primocin (InvivoGen, San Diego, CA, USA). For ATO seeding, MS5-hDLL4 cells were used at a confluency of approximately 70% and were dissociated with TryplE Express. Cell viability and cellularity was assayed using a Countess II FL Automated Cell Counter (Thermo Fisher) along with a 0.4% Trypan

Blue Stain (Thermo Fisher).

## 2.4 Flow Cytometry

All flow cytometric analysis was performed on a BD LSR II Fortessa (BD Biosciences, San Jose, CA) and analyzed using FlowJo software version 10.5 (FlowJo, Ashland, OR, USA).. Cells derived from ATOs were stained with the following monoclonal antibodies: CD14 eFluor450 clone 61D3 (eBioscience), CD5 Pe-Cy7 clone UCHT2 (eBioscience), CD7 Alexa Fluor 700 clone eBio124-1D1 (eBioscience), TCRab APC clone IP26 (eBioscience), CD4 Alexa Fluor 700 clone OKT4 (eBioscience), CD19 PerCP-Cy5.5 clone HIB19 (Biolegend), CD56 FITC clone MEM-188 (Biolegend), CD1a APC clone HI149 (Biolegend), CD8a Pe/Dazzle clone RPA-T8 (Biolegend), TCRgd FITC clone B1 (Biolegend), CD34 PE clone 563 (BD Biosciences), CD45 V500 clone H130 (BD Biosciences), CD3 BV421 clone UCHT1 (BD Biosciences), CD8b PE clone 2ST8.5H7 (BD Biosciences). All cells were also stained with an intracellular Live/dead fixable UV dye (Invitrogen). This information is also present in Appendix table 7.3.

Derivation of the human embryonic mesodermal progenitor cell type was assessed with the antibodies CD56 FITC clone MEM-188 (Biolegend) and anti-human CD326 (EpCAM) PE/Cy7 clone CO17-1A (Biolegend).

## 2.5 Artificial Thymic Organoid System

### 2.5.1 Derivation of Human Embryonic Mesodermal Progenitors

Mesodermal progenitors were induced from iPSCs as described in Montel-Hagen et al. 2019 with certain optimizations. Briefly, iPSCs were cultured on Vitronectin coated played in Essential 8 complete medium. At the time of induction, cells are 60-70% confluent and are dissociated from the well using TryplE at 37C for 10 minutes. Dissociated cells are then washed and resuspended in X-Vivo 15 medium (Lonza) supplemented with rhActivin A (R&D Systems), rhBMP4 (R&D Systems), rhVEGF (R&D Systems), rhFGF (R&D Systems) at 10ng/mL and ROCK Inhibitor at 10uM (Tocris). After counting, $2.5x10^6$ cells are seeded into a single Vitronectin coated well in a 6-well dish in 2mL X-Vivo 15 media supplemented with the same factors. Medium was then changed each day with the same growth factors as above without ROCK inhibitor. On day 5, cells are incubated with 1X Accutase (Stem Cell, Vancouver, Canada) for 10 minutes at 37C and washed twice with PBS before staining with EPCAM and CD56 antibodies to prepare for fluorescence activated cell sorting. Stained cells were sorted on a FACSARIA instrument (BD Biosciences, San Jose, CA) for CD56+ EPCAM- cells.

### 2.5.2 Generation of Mature Naive T cells from iPSCs

ATOs were generated as described [19] with some minor alterations. Briefly, 10,000 EPCAM- CD56+ hEMP cells were combined with $0.5 * 10^6$ MS5-hDLL4 washed and counted cells per ATO resuspended directly in EGM2 (Lonza) media with

10uM TGF-BRI inhibitor SB-431542 (SB Blocker) (Tocris Bioscience) and ROCK Inhibitor. Enough cells were combined in a 1.5mL Eppendorf tube to accommodate 6 ATOs. Combined cells in 1.5mL Eppendorf tubes were centrifuged at 1400RPM for 5 minutes at 23C in a swing bucket centrifuge to pool cell aggregates to the bottom of the tube. The media was aspirated out and enough complete media was added to the cell slurry such that the total volume was 36uL (6uL per ATO with 6 ATOs in each tube). Two 6uL droplets of the resuspended cell slurry were then placed with a p20 pipette onto a pre-wet Millicell Transwell Insert (EMD Millipore, Billerica, MA) Membrane sitting in 1mL of complete EGM2 media in a single well of a six well plate. The media was then changed every other day. At day 7, the media was further supplemented with hematopoietic cytokines rhTPO (Peprotech), rhFLT3L (Peprotech) both at 5ng/mL, and rhSCF (Peprotech) at 50ng/mL. At day 14 the media is changed to "RB27" which consists of RPMI 1640 (Gibco, Waltham, MA, USA), 4% B27 (ThermoFisher Scientific, Grand Island, NY), 30uM L-Ascorbic Acid (Sigma-Aldrich, St. Louis, MO) resuspended in 1X PBS, 1% Glutamax (ThermoFisher Scientific, Grand Island, NY), 100ug/mL Primocin, rhIL7 and rhFLT3L at 5ng/mL as well as SCF at 10ng/mL. Media was changed twice weekly. From weeks 3 to 5 of RB27 culture, cells were harvested from ATOs by adding 1mL MACS buffer (PBS with 5% BSA and 0.5M EDTA) to each well, physically disrupting the ATOs using two p1000 tips, and then further pipetting to dissociate the ATO fragments from the membrane. Cells were then centrifuged and resuspended and counted in FACS buffer before being stained with the antibodies listed above on a BD LSR II Fortessa (BD Biosciences, San Jose, CA) and analyzed using FlowJo software

version 10.5 (FlowJo, Ashland, OR, USA).

## 2.6 Statistical Analysis

Details of all statistical analyses are including in the figure subtext. All statistical measures were calculated using Graphpad Prism 8 version 8.4.1 (San Diego, CA, USA).

# Chapter 3

# Gene Editing and Differentiation in Hematopoietic Stem Cell Derived T Cells

## 3.1 Characterizing Patient HSPCs

### 3.1.1 Introduction

Patients with inborn errors of immunity are born with an impaired immune system the severity of which is dependent upon the mutation driving pathogenesis. Severe Combined Immunodeficiency (SCID) patients require treatment in the first months of life to generate an adaptive immune system (Dorsey et al. 2017). As mentioned in the introduction chapter, the primary treatment modality for patients with severe combined immunodeficiency is an allogeneic Hematopoietic Stem Cell Transplantation (HSCT). While allogeneic transplantation options have been of great help to SCID patients, challenges remain including finding matched donors, acute and chronic graft vs. host disease, engraftment failure, and incomplete immune reconstitution. This work seeks to circumvent these issues by creating an edited autologous

transplantation treatment option.

### 3.1.2 Objectives and Hypotheses

The overarching objective for this thesis project is to provide pre-clinical data in support of a gene edited autologous hematopoietic stem cell transplantation as a safer and more efficacious treatment option when compared to other treatment methods including HSCT from matched unrelated donors and haploidentical familial donors. To that end, we hope to demonstrate that the gene editing approach that we've taken can robustly rescue not only the development of HSPCs and their block in T cell development, but that these cells also undergo a rigorous V(D)J recombination process to establish a polyclonal repertoire of TCRs.

As far as this thesis work, the HSPC *in vitro* work is the most closely correlated with a clinical application. The iPSC work can demonstrate functionality of the cargo and more thorough testing parameters in that it is a clonal population of edited cells, but iPSCs would not be transplanted into a RAG deficient patient due to the additional complexity introduced by *in vivo* differentiation of iPSCs and potential dangers introduced by unpredictable differentiation. The NSG mouse data is also of importance to demonstrate the engraftment and differentiation potential of these edited patient cells, though of course significant physiological dissimilarities exist between humans and mice, including differences in thymic development (Shultz et al. 2005).

The objective for this portion of the thesis project was to demonstrate high levels of integration of our redesigned and recoded r.RAG2 gene editing cargo with minimal cytotoxicity, and to then demonstrate that these edited patient HSPCs could rescue

T cell development when compared to their isogenic precursors. Much of the time spent towards this aim was used to troubleshoot editing toxicity which seemed to be amplified within the ATO system. Additional time was then spent to attempt to optimize the generation of mature naive T cells from this same system. There is some work that remains to be done on the HSPC portion of this project that will extend beyond this reported set of experimentation, namely repeating final experiments reported in this thesis to demonstrate whether or not our editing strategy rescues *RAG2* deficient patient HSPC T cell differentiation capabilities as well as confirming TCR rearrangement profiles by Next Generation Sequencing (NGS) and testing T cell function by proliferation and activation cell based assays.

## 3.2 Genome Engineering in HSPCs

The first task in correcting the **c.296C>A;c.1342C>A** compound heterozygous RAG2 deficient patient HSPCs gifted to us by the lab of Dr. Harry Malech by Dr. Suk See DeRavin was to design constructs for knocking in our codon optimized *RAG2* cargo. From there, we isolated variables involved in optimizing gene editing including: the Short Guide RNA (sgRNA) to Cas9 molar ratio, the number of cells electroporated, the activity of the sgRNA at the target site, the length of the cargo introduced by AAV, the electroporation program, and the Multiplicity of Infection (MOI) used to transduce targeted cells. We systematically tested these variables in order to optimize knock-in frequencies. This work was done under IRB protocol 04-09-113R.

### 3.2.1 Design of Viral Cargo

To look at the rescue of edited *RAG2* deficient patient cells, three different viruses were constructed using the latest ENSEMBL builds of the human *RAG2* consensus sequence.

Figure 3.1 details the variations in cargo design to comparatively examine potential optimization of transgene expression with efficiency of HDR integration. Of the three viruses, the shortest contains just the codon optimized recoded *RAG2* sequence (r.RAG2) from the latest ENSEMBL build of the gene's consensus sequence (ENSEMBL version ENSG00000175097), flanked by 400 basepairs of homology arms. Homology arms are sequences on either side of the cargo to be inserted into the genome which align to the 5' and 3' wild type DNA strands that are produced after the designed Cas9 sgRNA complex cuts at the target site. Proteins involved in the DNA damage response pathway flag these homologous regions as a template to repair the cut DNA, thereby inserting the homology arms and the cargo between them into the genome. This shortest virus was intended to have the highest insertion potential, as knock in efficiency scales down with cargo length (Paquet et al. 2016) (additional data informally published by Addgene). The next construct contained the same r.RAG2 element, but with an added Bovine Growth Hormone (BGH) Polyadenylation Signal (pA) which has been shown to stabilize viral transcripts (Wodrich et al. 2000). Finally, a reporter construct was designed which contained all elements of the r.RAG2 BGH pA cargo along with a Green Fluorescent Protein (GFP) reporter under the Spleen Focus Forming Virus (SFFV) promoter. This final construct is the largest of the designed cargos, and so was

expected to have the lowest knock-in efficiency, yet cells transduced with this virus produce a signal detectable by flow cytometry and so knock-in analysis is easier to perform with this fluorescent readout.

Another consideration in our design for this viral cargo was the effect that a knock-in may have on *RAG2* expression during development. *RAG2* is very tightly regulated at the transcriptional level, the epigenetic level, and the cell cycle level. As such, we intended to retain as much of the epigenetic and transcriptional regulatory landscape as possible by targeting the endogenous locus with a construct which does not contain any heterologous promoter to drive *RAG2* expression. We hoped that this design would take advantage of the native landscape to turn *RAG2* expression on and off in a way which is similar to that of immunocompetent unedited cells undergoing T and B cell maturation.

The final consideration in our viral cargo design was the codon optimization of the *RAG2* gene. Codon optimization or cDNA recoding is the process of tailoring the nucleotide sequence of a gene to cater to the preferences of the host organism as reviewed in (Plotkin et al. 2011). Recoding takes into account the transfer RNA (tRNA) profiles of the cell type of interest, which may have a bias in transferring an amino acid to a protein encoded by an often redundant codon for that particular amino acid. In the case of *RAG2* specifically, the catalytic activity of the RAG complex is indicative of transposases which do not have their origins in mammalian cells. As a result, the codon usage of the *RAG2* gene may not be primed for translation in human cells. The Codon Adaptation Index (CAI) score is a quantitation of codon bias usage in human cells and is reported for our designed

cargo in Figure 3.1 **A**. The CAI score is modestly higher for r.RAG2 than for that of the wild type sequence, which may give a competitive advantage to the translation of our recoded construct over its wild type counterpart in cells that are edited on only one allele. Perhaps more importantly than this boost in translational efficiency though is the de-risking of unintended homologous recombination products during gene editing. Without codon divergence, gene replacement strategies run the risk of producing partial gene recombinants given the high sequence similarity between the introduced repair template and the wild type sequence. Diverging the DNA sequence from that of the wild type gene reduces the risk of the DNA damage repair pathway partially recombining the template into the wild type gene, and instead promotes precise targeted integration using the homology arms flanking the cargo for insertion.

### 3.2.2 Testing of sgRNAs

The CRISPR system of mammalian genome engineering has been recently introduced and extensively characterized as an RNA programmable nuclease over the past 7 years. Four sgRNAs were designed to take advantage of the CRISPR system in our pursuit to knock in the r.RAG2 cassette into the endogenous locus. The sgRNA Cas9 complex will continuously cut at the location complementary to the sequence of the sgRNA, with the DNA repair machinery attempting to repair the DNA cut each time, until a mistake is made in the DNA repair which causes the sgRNA sequence to no longer be complementary to the target sequence. The relics of this failure to perfectly repair the cut site in a heterogeneous pool of genomic DNA is what is detectable by Sanger sequencing as INDELs, and the frequency

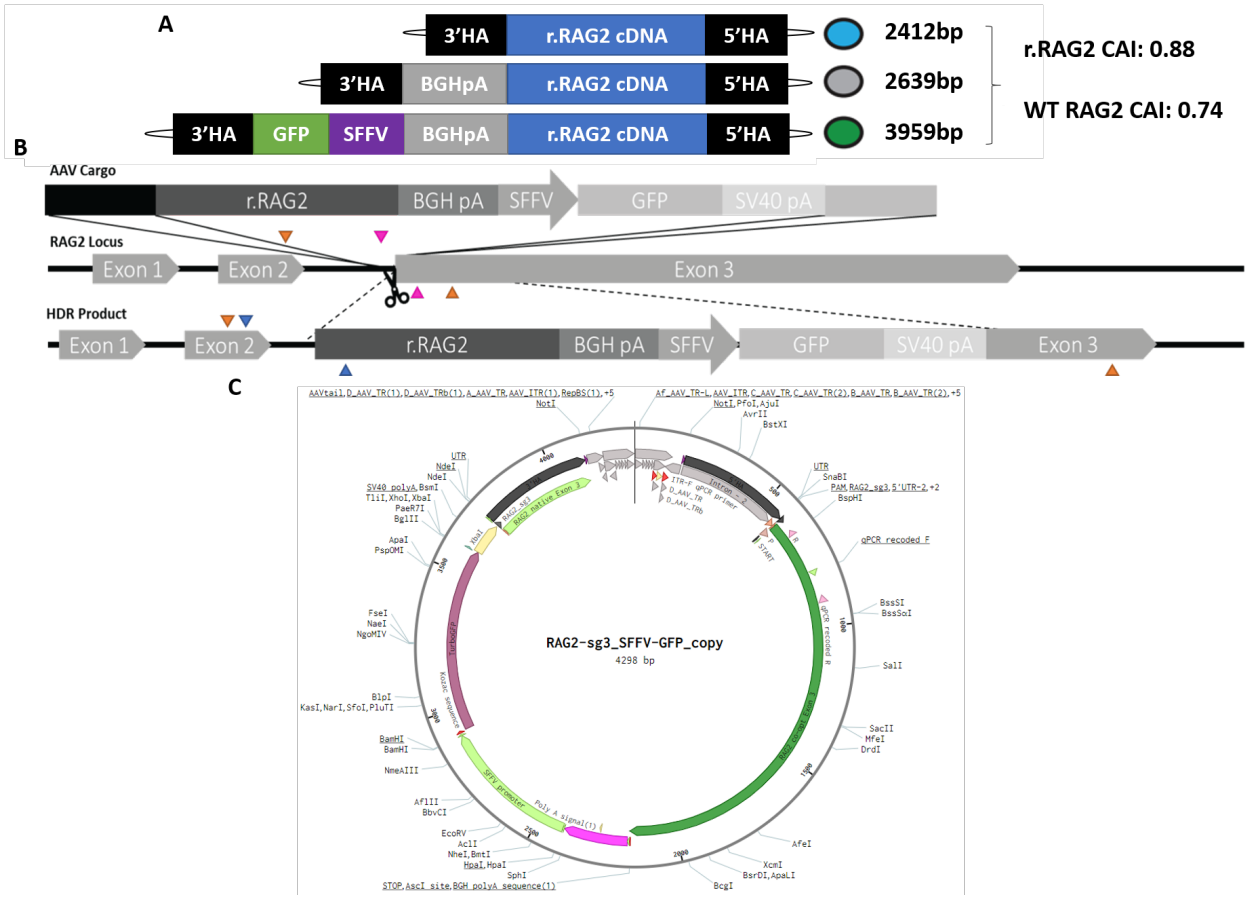



Figure 3.1: **Figure A**. Depicts the three variant virus payloads that we designed in order to interrogate the functionality of the cDNA. Shown to the right is the Codon Adaptation Index (CAI) score for the recoded gene against that of the wild type.**Figure B** Depiction of the r.RAG2 cargo being delivered by AAV6 into the target locus after the double-stranded DNA break induced by the RNP complex. Colored arrows indicate the location of primers used within this chapter. The orange triangles represent the location of primers used to check for gene targeting heterozygosity in Figure 3.14. The pink arrows represent the location of primers used to check INDEL activities as in Figure 3.2. The blue arrows represent the location of primers used to check gene targeting products by ddPCR as in Figure 3.5. **Figure C**. shows an example of the design within the plasmid backbone as visualized in Benchling.

of the INDELs correlates with the activity of that sgRNA. Canonically, the higher the activity of the sgRNA the better. To analyze this activity, four sgRNAs were synthesized and cloned into a px459 plasmid expressing high fidelity SpCas9. 1ug of each assembled plasmid was lipofected into 70% confluent HEK293T cells. After 48 hours, the genomic DNA of lipofected cells was purified and primers surrounding the *RAG2* start site were used to PCR a 1kb region containing the sgRNA target site. Sanger sequencing was then used to analyze the frequency of INDELs present in cells receiving each sgRNA compared with negative control cells that did not receive the gene editing machinery.

In the case of performing a knock-in at a target location, the efficiency of HDR drops dramatically the further away the insertion site is from the cut site (K. Wang et al. 2016) (Inui et al. 2014). sgRNA1 cuts 6 basepairs upstream of the start codon knock-in site, whereas sgRNA4 cuts 40bp downstream of the knock-in site Figure 3.2 **A**. For this reason, sgRNA1 was selected for further experimentation despite it having a lower on-target activity rate when compared to sgRNA4 Figure 3.2 **B**.

### 3.2.3 Knock-in Efficiency Scales with Number of Cells Electroporated and MOI

Taking our designs into the cell, we began to investigate the best methods for delivering the gene editing machinery into CD34 expressing HSPCs. As suggested in the existing literature at the time, we began our experiments in the range of $1x10^5$ using either 200,000 or 400,000 HSCs per gene editing condition (Gundry et al. 2016) (Brunetti et al. 2018).

What we found is that the efficiency of targeted integration correlated positively

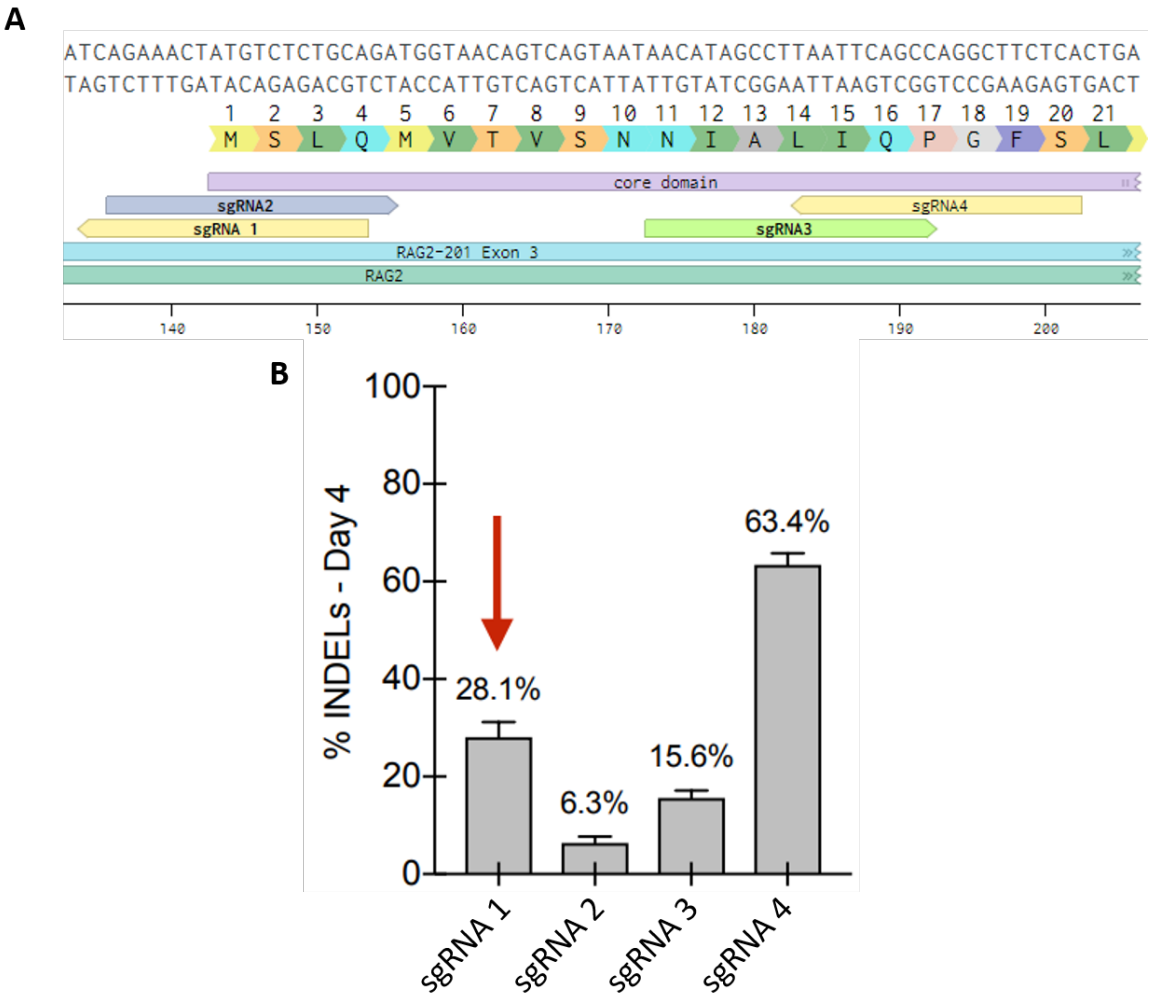



Figure 3.2: **Figure A**. Shows the genomic location of all four sgRNAs and their strand orientation. Sequence identity of these four sgRNAs can be found in Table 7.2. **Figure B**. On target sgRNA INDEL activity as measured by Sanger Sequencing. Primers used for sanger sequencing are entitled TIDE RAG2 F and TIDE RAG2 R in Table 7.1.

with the number of expanded HSPCs electroporated as demonstrated by flow cytometric GFP detection from the genomic integration of our r.RAG2 GFP reporter virus Figure 3.3. The 400,000 cell population modestly outperformed the 200,000

population in every measured condition. This higher cellularity-based editing efficiency boost during electroporation using the Lonza Nucleovette strips may result from a more consistent, higher density mass for the voltage applied across the cuvette's conducting plates to traverse. Also notable is another modest increase in knock-in efficiency between the Cas9/sgRNA molar ratio of 1:2.5 to 1:5. This modest increase suggests that a higher copy number of sgRNA is required during the Ribonucleoprotein (RNP) complexing step, and that the sgRNA may be rate limiting in active RNP formation. At this higher ratio, more gene editing constructs were available to begin the process of DNA damage response and activation of the HDR pathway, resulting in higher levels of cargo knock-in. As a result, the molar ratio of 1:5 was used for all future experiments. Finally, consistent with prior art we demonstrated that a higher MOI resulted in correspondingly higher frequencies of HDR mediated knock-in of the viral cargo. Higher knock-in rates are due to an increased frequency of the repair template being accessible to cells performing a DNA damage response as initiated by the sgRNA Cas9 RNP complex (Dai et al. 2019) (Gaj et al. 2017). To examine any autofluorescence detected by the flow cytometer, we used a control which received neither the RNP nor the AAV. To check for any episomal expression of GFP, we used an AAV only control which did not receive the RNP, abrogating genomic integration of the r.RAG2-GFP cassette into the genome. No autofluoresence or notable AAV episomal GFP expression were observed Figure 3.3.

Further examining cellularity, we pushed the number of cells to be electroporated from 400,000 to 1,000,000 per condition and noted another increase in HDR as read out by cells expressing the GFP tag downstream of the r.RAG2 cargo Figure 3.4.

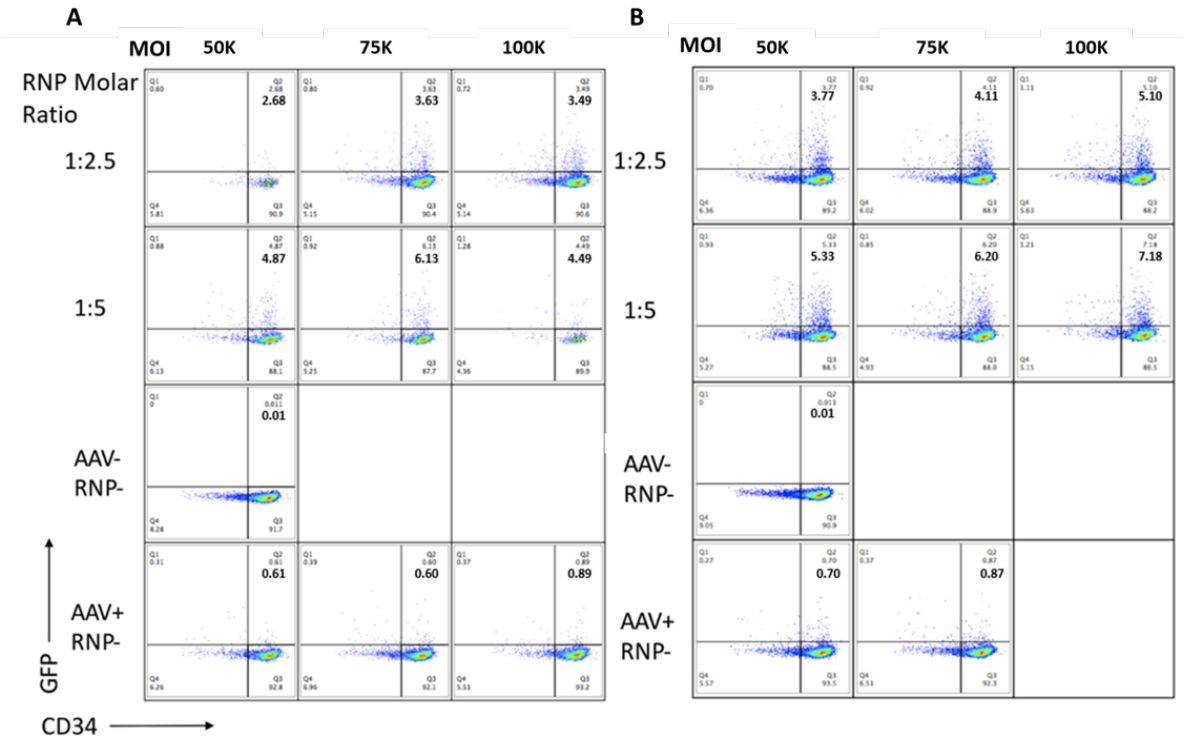



Figure 3.3: Flow cytometry data showing expression of the GFP reporter introduced into the genome of HSCs via three different MOIs for the coRAG2-GFP AAV cargo coupled with a 1:2.5 or 1:5 molar ratio of the Cas9 sgRNA ribonucleoprotein. **A** is representative of experimental conditions being applied to 200,000 electroporated HSCs, and **B** is representative of the same conditions being applied to a population of 400,000 HSCs. n=1

Similarly, an increase in MOI from 100,000 to 200,000 also showed a modest increase in cargo integration by HDR.

### 3.2.4 Knock in Efficiency Inversely Correlates with AAV Cargo Length

Prior experiments have demonstrated that shorter cassettes have a higher probability of being knocked into the genome via HDR (Li et al. 2014). Precise mechanisms

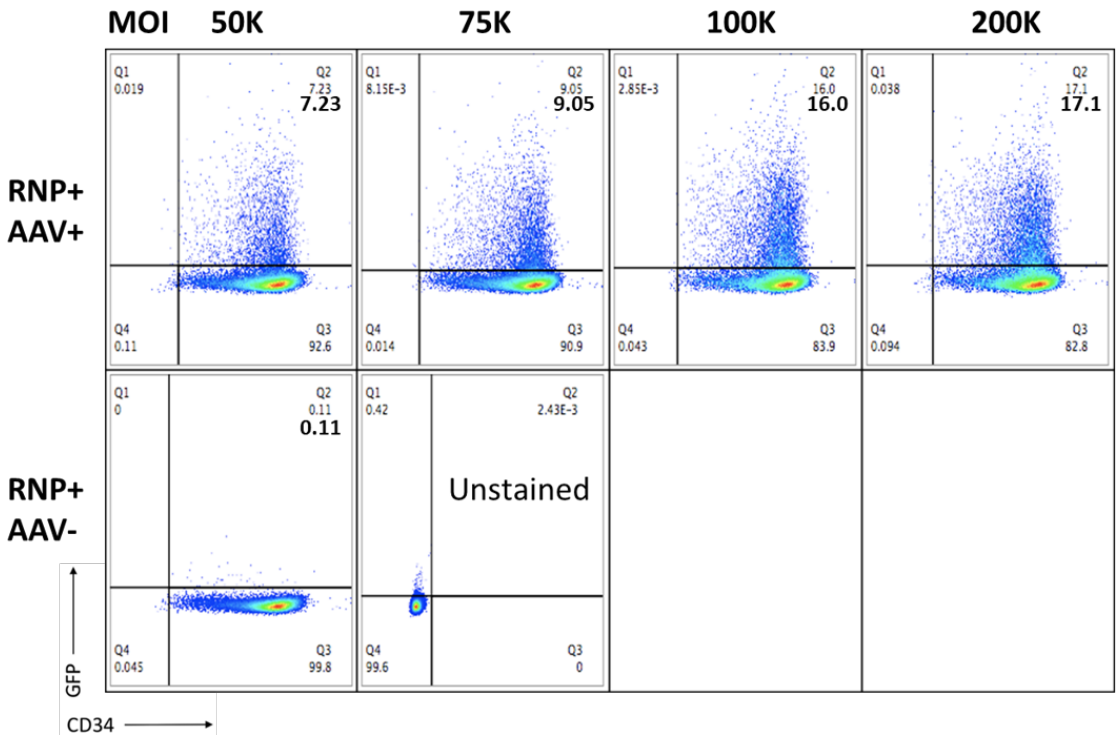



Figure 3.4: Flow cytometry data showing further increases in HDR efficiency contingent upon a higher cellularity of $1x10^6$ cells electroporated across a range of MOIs. n=1

for why this is the case are unknown, but may have to do with secondary DNA structures becoming more complex with larger cassettes, potentially obscuring the recognition of the homology arms on either end of the insertion sequence needed for recombination into the target cut site by effector proteins involved in the DNA damage repair pathway.

This phenomenon was recapitulated in our studies as well, with the shortest virus integrating most frequently into the genome and the largest virus integrating with the least efficiency Figure 3.5. Also of note is the concordance with which the fluorescence readout by flow cytometry matches the detection of HDR in the genomic

context by droplet digital PCR (ddPCR) Figure 3.5 **B**. ddPCR is a relatively new technique in which genomic DNA along with primers and probes against a target location and a reference gene are encapsulated in a droplet of oil in an immiscible fluid, creating roughly one copy of the genome per droplet analyzed by the machine. As the PCR reaction occurs, the probes bind to the increasing number of copies of the target sequence and produce a signal which is detectable by the instrument. By using primers and probes which detect a reference gene on the same chromosome as the gene editing target location of interest, gene modifications can be normalized and a percentage of gene edited alleles can be quantified. Using this technique, we noted the continued scale of HDR efficiency with MOI as shown in Figure 3.3.

## 3.3 Artifical Thymic Organoids (ATOs) from HSPC Precursors

### 3.3.1 ATOs Faithfully Recapitulate Physiological Development

To date, only two core protocols have been demonstrated to generate mature naive human T cells from HSPCs and/or iPSCs. The first technology to demonstrate this uses a monolayer co-culturing system wherein stem cells are cultured on top of the immortalized murine stromal cell line OP9 which has been engineered to express delta like ligand 1 or 4 (Schmitt et al. 2002). While results of this system are impressive and pioneering, the introduction of a 3D ATO culturing system was shown to significantly enhance the differentiation of functionally mature human T cells from stem cell precursors as is covered in more detail in Section 1.5.1 (Seet

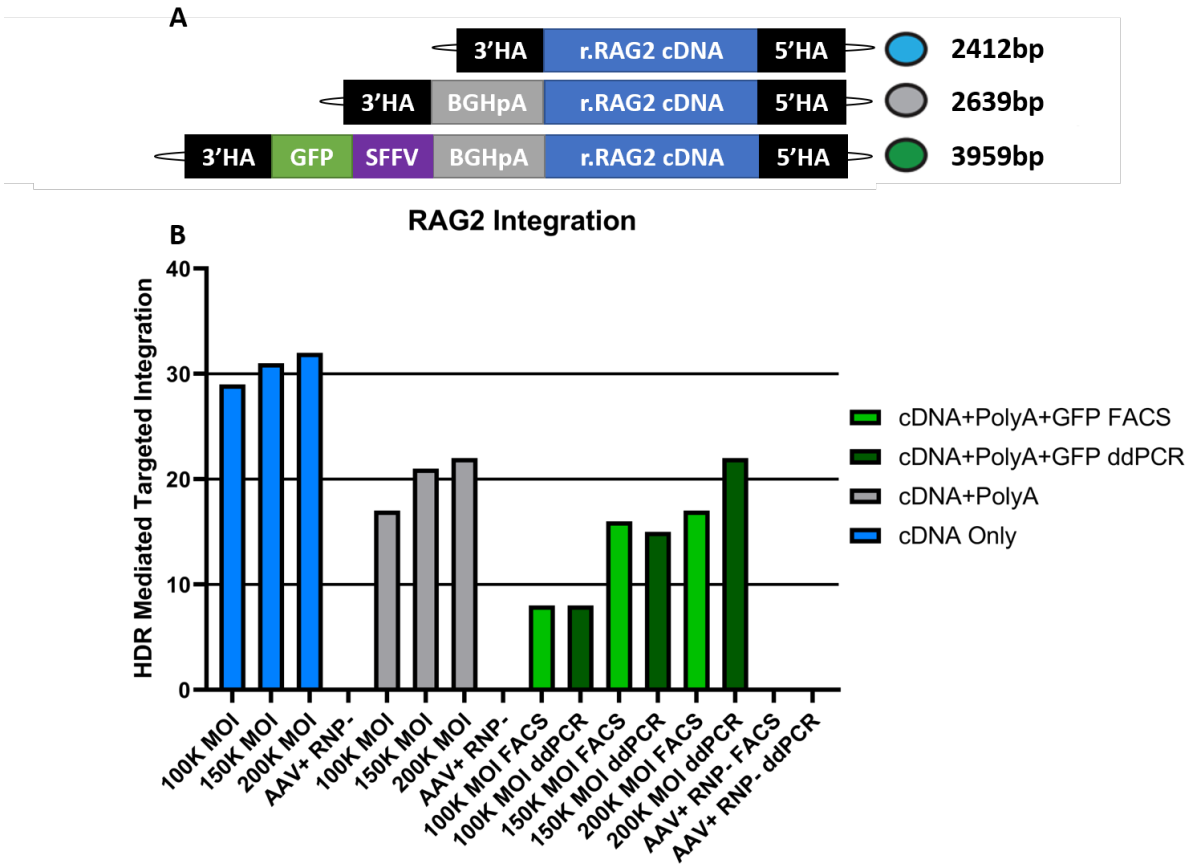



Figure 3.5: **Figure A** Structure of each virus being tested in this experiment. **Figure B** Detection of HDR by ddPCR for the three r.RAG2 viruses used for editing. Integration rate of the r.RAG2-GFP cargo demonstrates concordance of ddPCR results with fluorescence readout by flow cytometry. n=1

et al. 2017) (Montel-Hagen et al. 2019). The demonstration of this 3D organoid technology in deriving mature T cells from stem cells seemed an appropriate chassis for us to study the ontological effects that our gene editing approach for *RAG2* may have on developing lymphocytes *in vitro*.

Before moving forward with studying the effects that our gene editing had on developing T cells, we intended to confirm that the physiological markers of T cell

development which occur naturally in the thymus were present when looking at healthy donor HSPCs going through the ATO system. Healthy donor HSPCs derived from apheresed mobilized peripheral blood were generously provided by Dr. Suk See DeRavin. Pre-T cell markers such as CD1$\alpha$, CD5, and CD7 were detectable in robust levels after just weeks of culture in the ATO system Figure 3.6 **A**. While few cells showed expression of likely Natural Killer (NK) cell marker CD56, the majority of cells indicate very distinct T cell lineage commitment with more than 90% of cells expressing the indicated pre-T markers. Furthermore, low expression of CD34 indicates that almost all HSPC precursors had been depleted with none differentiating toward the B cell lineage as noted by lack of CD19 expression. Similarly, no monocytes were detected in this system as measured by CD14 expression.

Mature markers of T cell development such as expression or co-expression of CD4, CD8$\alpha$, TCR$\alpha\beta$, TCR$\gamma\delta$, and CD3 were evident beginning in week 3 Figure 3.6 **B**. From these results, we see few Immature Single Positive (ISP) cells expressing CD4 or CD8a without a TCR, with the majority of detectable cells co-expressing CD4 and CD8, becoming so-called Double Positive (DP) cells. A proportion of these DP cells mature to express TCR$\alpha\beta$ and co-receptor CD3, which is the final flow cytometric readout used to establish a mature naive T cell phenotype in this study. Interestingly, we also noted a physiologically relevant number of TCR$\gamma\delta$ cells, further establishing that this platform recapitulates a number of physiological processes which occur in the thymus during T cell development and as such represents a faithful chassis for us to study gene editing approaches to correct *RAG2* deficiency.

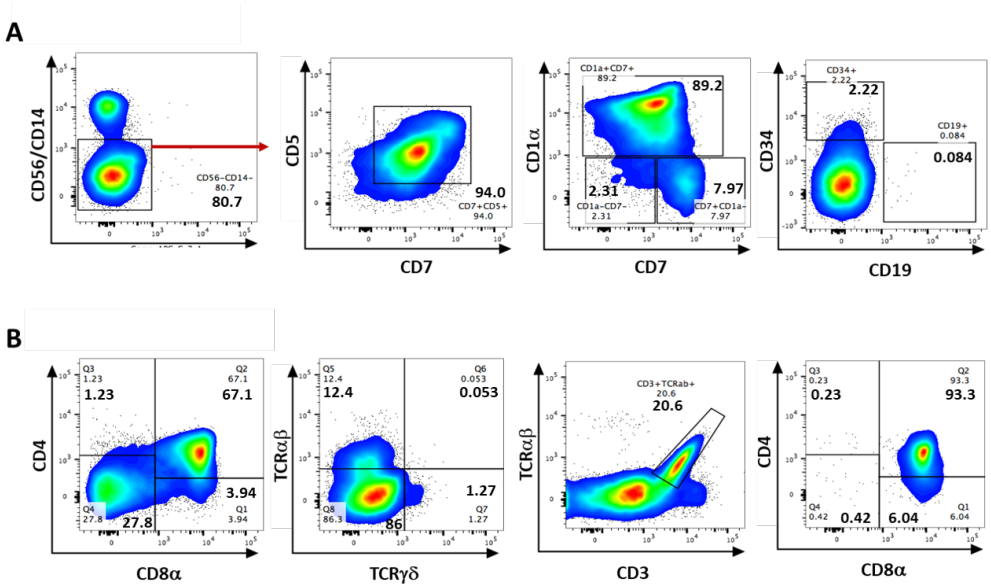



Figure 3.6: **Figure A** Examination of early T cell development markers as measured by flow cytometry. **Figure B** Analysis of mature T cell markers as measured by flow cytometry

### 3.3.2 HSPC Derived T Cells Proliferate in Response to anti CD3 CD28 Antibodies and Display a Diverse V-beta Repertoire

Beyond looking at surface markers, we wanted to verify and functionally characterize the activity of HSPC derived T cells generated within the ATO system.

Having successfully demonstrated expression of both TCR$\alpha\beta$ as well as a small proportion of TCR$\gamma\delta$ cells, we wanted to verify that the ATO environment was stimulating development of a polyclonal repertoire of TCRs. Several technologies exist to assay TCR diversity including NGS based approaches - which this thesis work will touch on later - as well as flow cytometric approaches which look at the expression of constant regions of V gene portions of the beta chain of the TCR (V$\beta$s).

Antibodies available for this flow cytometry based approach cover approximately 70% of possible V$\beta$ usage (Wei et al. 1994). In restricted TCR repertoires, typically there is a clonotype which dominates and expresses a single V$\beta$. A wide range of V$\beta$ detection by flow cytometry then intimates a polyclonal repertoire of TCRs. Using this approach we demonstrate that ATO derived T cells display similar V$\beta$ profiles to those of primary T cells Figure 3.7 **A** indicating that the ATO environment does a good job of physiologically replicating the maturation processes present in the thymus during TCR rearrangement.

Having now demonstrated robust V$\beta$ usage in ATO derived T cells, we wanted to verify functional activity by assaying the proliferative capabilities of these organoid derived T cells. To this end, we looked at antibody mediated expansion of T cells using anti-CD3 and anti-CD28 antibodies. In the human body, the TCR binds its cognate peptide MHC and activates its co-receptor complex CD3 leading to the activation of an intracellular signalling pathway which promotes cellular expansion and effector functions (Alarcon et al. 2003) (Meuer et al. 1984), (Haskins et al. 1984). Taking an *in vitro* approach to TCR stimulation of a polyclonal repertoire of T cells each with a unique TCR presents many challenges, as the antigen recognition domain of each TCR cannot engage with its cognate peptide. To subvert this problem, the TCR co-receptor molecule CD3 can be targeted with an antibody which can propagate intracellular T cell proliferation signals as well. Activated T cells quickly become anergic and undergo apoptosis with sole targeting of CD3 by antibodies (Schwartz 1990). Activation of the CD28 molecule circumvents this anergic state by propagating a complex and still mechanistically unresolved series of cellular changes

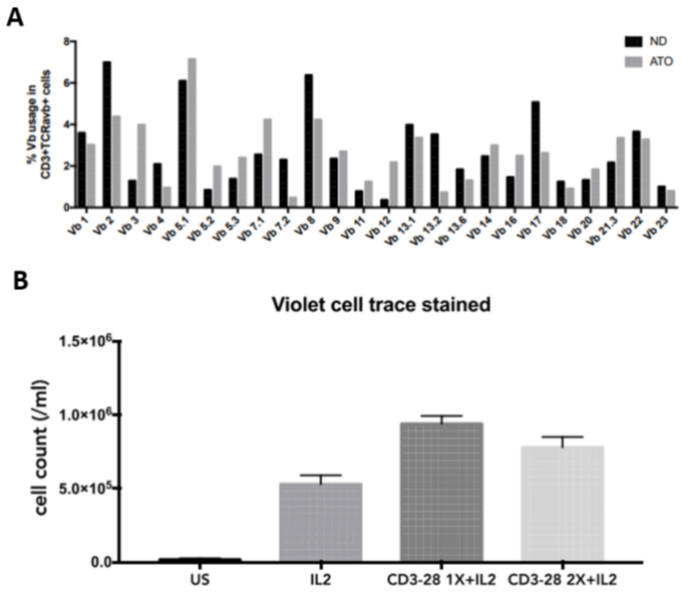



Figure 3.7: **Figure A** Flow cytometry data showing the frequency of V$\beta$ chains displayed on the surface of Normal Donor (ND) primary T cells or healthy donor HSPC derived T cells (ATO). **Figure B** Analysis of HSPC derived healthy donor T cells proliferation in response to stimulation with anti CD3/CD28 antibodies. Proliferation was measured by fluorescent readout using Celltrace Violet, with an unstained (US) control. IL-2 was used to sustain and propagate live cells, the effects of which were measured at two concentrations, 1X (100 U/mL) and 2X.

which upregulate the glycolytic properties of the cell and availability of IL-2, which allows for sustained function of activated and proliferating T cells (Jenkins et al. 1993), (Schwartz 1992), (Boise et al. 1995), (Frauwirth et al. 2002). The combination of antibodies against CD3 and CD28 then produces a more sustained proliferation of mature T cells *in vitro*. Applying this approach to the ATO derived T cells showed sustained co-receptor mediated proliferation which outperformed expansion due to inclusion of IL-2 alone, showing functional proliferation signalling and cellular properties in ATO derived cells Figure 3.6 **B**.

### 3.3.3 RAG Deficient Patient T Cell Developmental Blocks are Reproduced in the ATO System

Having established that the ATO system faithfully replicates common markers of T cell development, we wanted to confirm that this platform also recapitulates *RAG* null patient blocks in T cell maturation. Three separate patients with two copies of an inactivating mutation in either *RAG1* or *RAG2* were employed to verify this developmental block within the organoid environment. The *RAG2* deficient patient had a c.296C>A;c.1342C>A compound heterozygous genotype. *RAG1* deficient patient one and patient two had c.1228C>A;c.1520G>A, and c.413delT;c.2487-2488delinsTT compound heterozygous genotypes, respectively.

Consistent with patient presentation, all three *RAG* null patient cells undergoing differentiation in the ATO system failed to produce any significant amount of mature naive T cells Figure 3.8 **B** when compared to a healthy control. RAG inactivating mutations ablate any recombination in the V, D, and J, genes leading to the inability of patient cells to display a TCR, consistent with this flow cytometry data.

Having now demonstrated that the 3D thymic organoid system phenocopies patient developmental blocks in T cell maturation, we began the *RAG2* gene editing process to see if we could rescue development in *RAG* null patient cells.

### 3.3.4 AAV Mediated Gene Editing Toxic in ATO System

Transducing HSPCs with AAV has been well studied, with AAV6 predominating as the most efficient serotype for HDR mediated knock-in (Song et al. 2013) (Dudek et al. 2019). Prior art has also established that nuclease free targeted integration of AAV cargo is very limited (Rogers et al. 2019) and that high quality high titer

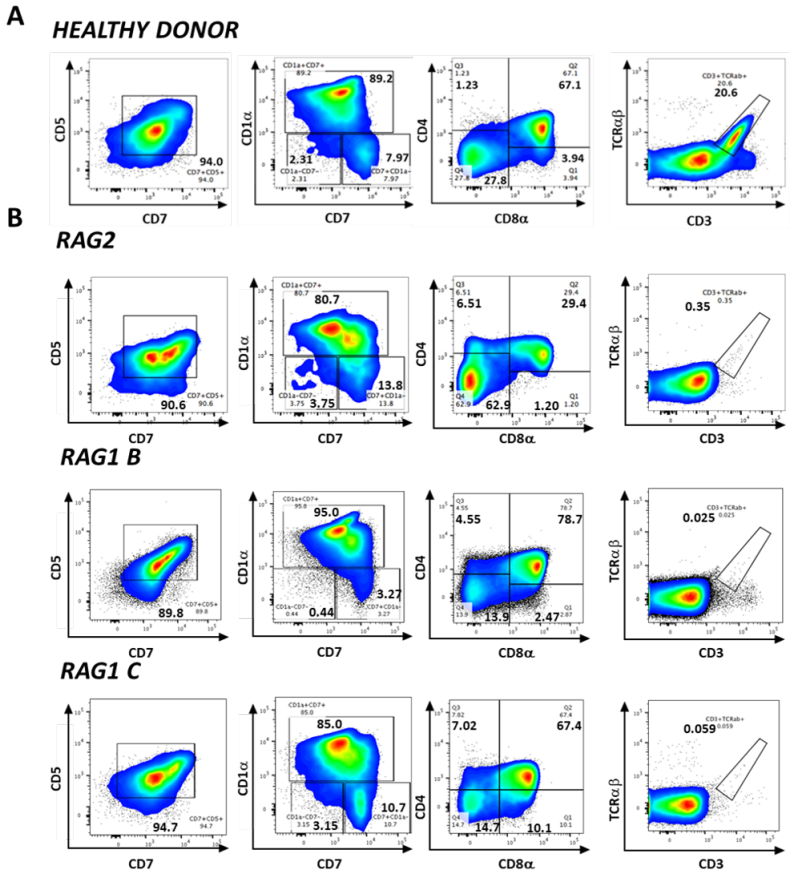



Figure 3.8: **Figure A** Flow cytometry data showing healthy and expected development of HSPC derived T cells from a healthy donor. **Figure B** Analysis of three patients with null mutations in *RAG2* (c.296C>A;c.1342C>A), *RAG1* (B: c.1228C>A;c.1520G>A, C: c.413delT;c.2487-2488delinsTT) showing a block at the double positive stage of development.

viruses stand the best chance at ameliorating cytotoxicity within the transduced cell (Chatterjee 2019) and as reviewed in (Chatterjee et al. 2020). With this in mind, all three viruses described in Figure 3.1 were produced at a titer of more than $1.88x10^9$ Vector Genomes (vg) per microliter as quantified by qPCR. The standardization of the exact number of reported vector genomes per cell has been contended due to

differences in quantification of the MOI. For example, titering by qPCR will give a different result when compared to titering from ddPCR. The range reported for vector genomes per cell in prior art spans from $1 - 10x10^4$ (J. Wang et al. 2016) to $1 - 3x10^6$ (De Ravin et al. 2017). We used an MOI of 200,000 which translates to a transduction in the range of $1 - 4x10^4$ vector genomes per cell, well within the range of previously reported success.

Irrespective of these previously corroborated ranges of successful HSPC transduction MOIs, we noticed stark cytotoxicity in the HSPCs that we attempted to gene correct with the r.RAG2-pA virus as we put them through the ATO system. Briefly, *RAG2* deficient patient HSPCs were thawed, expanded, and gene targeted with the r.RAG2-pA virus in an attempt to correct the *RAG2* deficient patient's inability to rearrange the TCRs of developing T cells in the ATO system. Unedited patient cells were used as a negative control, along with a positive control consisting of healthy donor cells which were not gene targeted. Healthy donor cells progressed through the 3D culture model quite well, with a majority of the cells analyzed being alive and expressing pre-T markers including CD5, CD7, and CD1$\alpha$ Figure 3.9, Figure 3.10. The *RAG2* deficient patient cells show a marked reduction in live cell count, with a viability of only 4.5% as noted by physical parameters Figure 3.9. Those few *RAG2* deficient cells that do survive in the ATO environment go on to develop and express CD5, CD7, and CD1$\alpha$, before becoming blocked at the CD4, CD8 double positive stage without further progression to co-expressing CD3 and a TCR Figure 3.9, Figure 3.10. However the *RAG2* deficient gene targeted cells show extreme cytotoxicity, with 0.1% of cells marked as viable Figure 3.9.

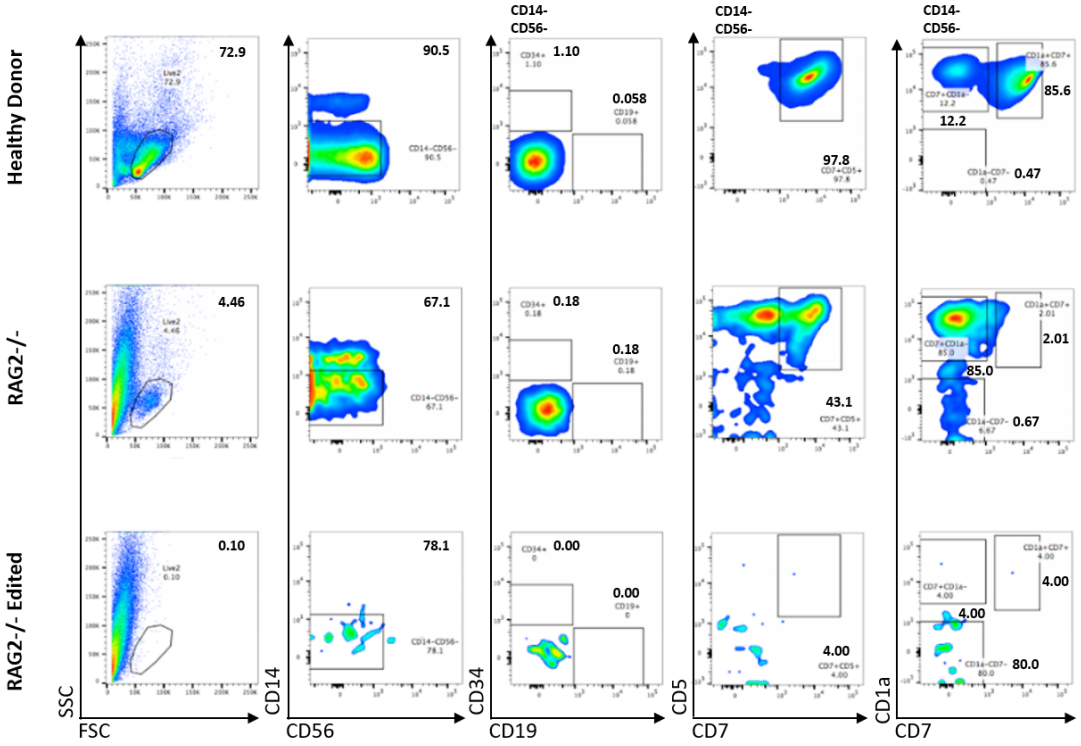



Figure 3.9: Flow cytometry data showing T cell development and differentiation of other cell types between the healthy donor, RAG2 deficient, and RAG2 edited patient HSPCs in the ATO system.

## 3.4 Amelioration of AAV-mediated Cytotoxicity in HSPC Editing

Our primary hypothesis for the cytotoxicity we noticed in HSPCs when compared to iPSCs - covered in Chapter 4 - was the amount of time available for HSPCs to recover from stressors in their editing pipeline, including the processes of thawing patient cells, expanding them *in vitro*, nucleofecting them with the RNP complexes, and then transducing them with a virus all in rapid succession. Whereas iPSCs can expand indefinitely, HSPCs need to be thawed quickly and expanded in order to

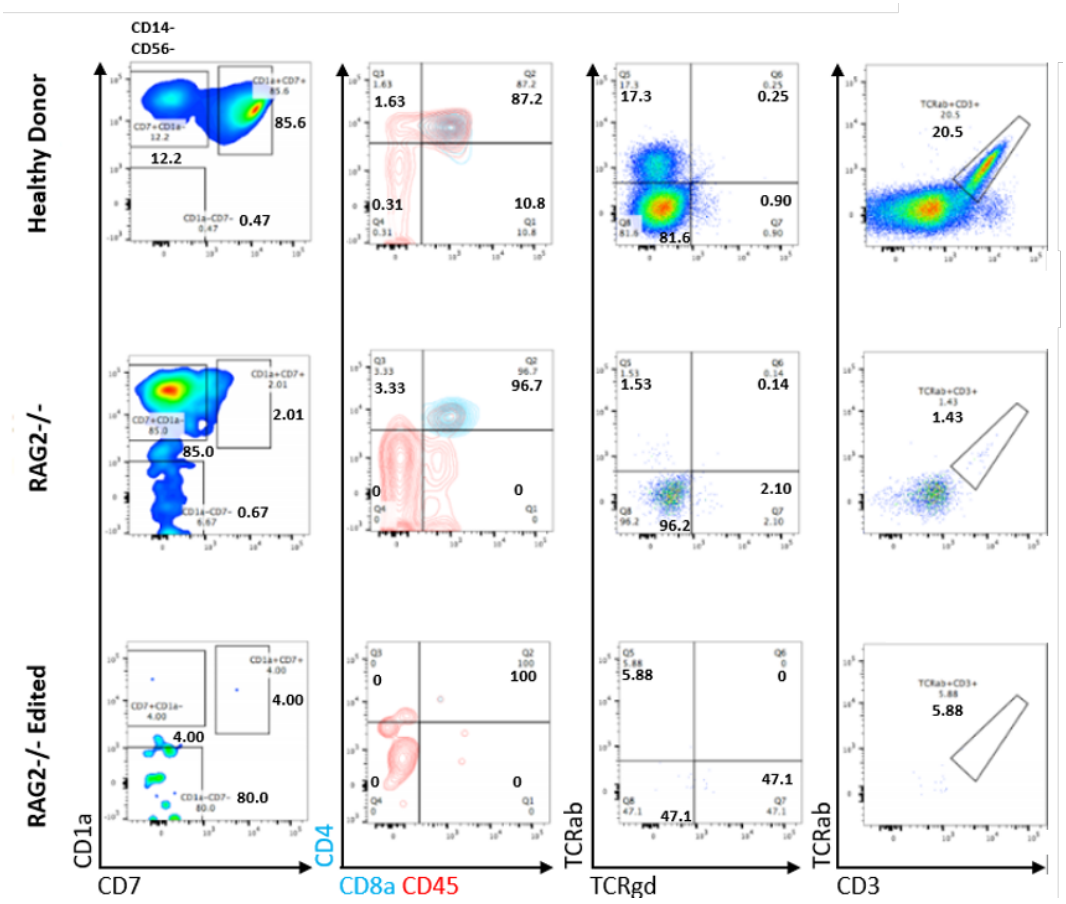



Figure 3.10: Flow cytometry data showing naive and mature T cell markers between the healthy donor, RAG2 deficient, and RAG2 edited patient HSPCs in the ATO system.

allow for cell division which predicates successful homologous recombination based gene editing activity (Orthwein et al. 2015). This quick succession of thawing, expanding, electroporating, transducing, and then seeding into an artificial 3D cell culture system we thought could be causing the cells to undergo apoptosis due to cellular stress induced by each step in the pipeline.

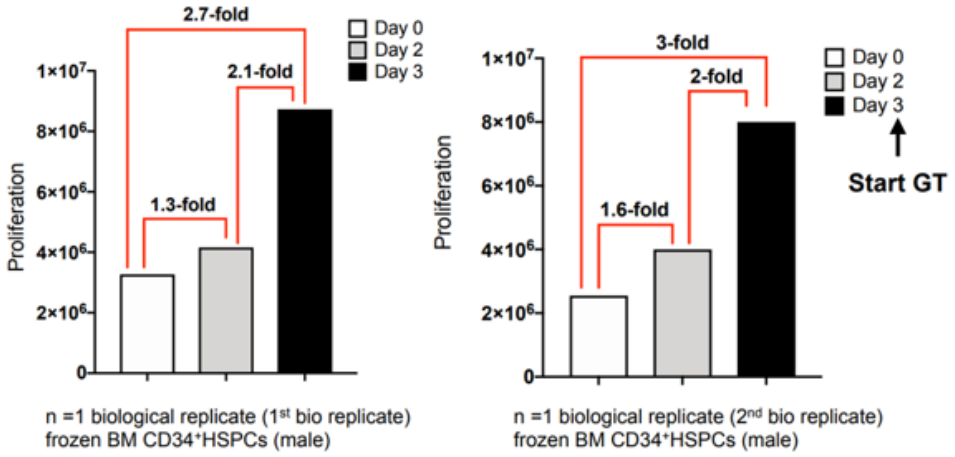



Figure 3.11: Increasing HSPC proliferation time from 48 hours to 72 hours more than doubles the proliferation rate as measured by hemocytometer. We used this indication as reason to start gene targeting (GT) on day 3.

### 3.4.1 Further Expansion Time Allows for Nonlinear Increase in the Proliferation of HSPCs

The first stress reducing approach we took for editing HSPCs was to extend the amount of time available for thawed cells to expand and reach some level of homeostasis. Typically, we would expand HSPCs for 2 days before beginning gene targeting experiments. To our surprise, the addition of just one day in the expansion pipeline lead to a non-linear jump in cellularity. We noted a 1.3-1.6 fold expansion of HSPCs after two days of culture, whereas we noticed a 2.7-3.0 fold expansion of HSPCs after three days of culture Figure 3.11. We suspected this more biologically active cell pool could be further amenable to gene targeting due to an increased amount of time to return to a more homeostatic state post thaw.

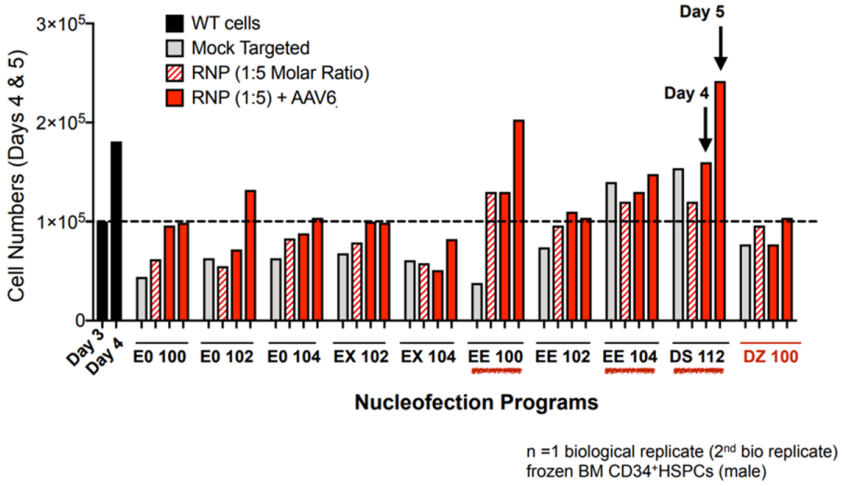



Figure 3.12: Ten electroporation programs were tested to examine their effect on cellular proliferation post transfection of a Cas9 sgRNA RNP. Mock targeted cells are those which were electroporated in empty buffer without RNP.

### 3.4.2 Electroporation Program Expands Post-Nucleofection Proliferation Capability of HSPCs

We next sought to reduce electroporation related cytotoxicity. Ten nucleofection programs were tested on the Lonza 4D system as recommended both by Lonza and from demonstrated work in prior art and were benchmarked against our prior program DZ100. Surprisingly, we found several other programs that seemed less harsh on the cell. These programs allowed for cells to continue proliferating post electroporation, indicating a minimized stressor effect on the cell. In particular, program DS112 showed almost mock levels of cellular proliferation post electroporation. This program was used alongside DZ100 in future experiments Figure 3.12.

### 3.4.3 Gentle Electroporation Programs Do not Confer Increased RNP Delivery

Having isolated a nucleofection program which seemed to leave HSPCs in a healthier state, we then looked at whether this program was as effective as our benchmark DZ100 at delivering the sgRNA:Cas9 RNP complex. Using our standard 1:5 molar ratio of sgRNA to recombinant high fidelity Cas9, we examined the INDELs generated by the RNPs delivered to HSPCs nucleofected with each program. We found that the electroporation programs we identified as being more gentle did not confer any increase in RNP delivery as identified by proxy as INDEL activity. The DS112 program showed a slight reduction in on-target activity when compared to DZ100, but we felt the impact of reducing RNP delivery efficiency was an acceptable tradeoff to having a healthier population of cells to input into the ATO system Figure 3.13.

### 3.4.4 Further HSPC Expansion Time Promotes Higher Levels of HDR

We next examined the effect of updated protocols for HSPC expansion and nucleofection on homologous recombination. Using single cell methylcellulose culturing techniques for colony forming assays, we were able to capture and expand single cells within a gelatinous matrix for Genomic DNA (gDNA) extraction and quantification by PCR of the r.RAG2 integration site. With amended protocols, we saw a jump in editing efficiency with 71% or 30 out of 42 clones having at least one copy of the r.RAG2 cassette inserted into the genome Figure 3.14 **A**. Primers across the junction of the insertion site alongside primers over the insertion site were used to

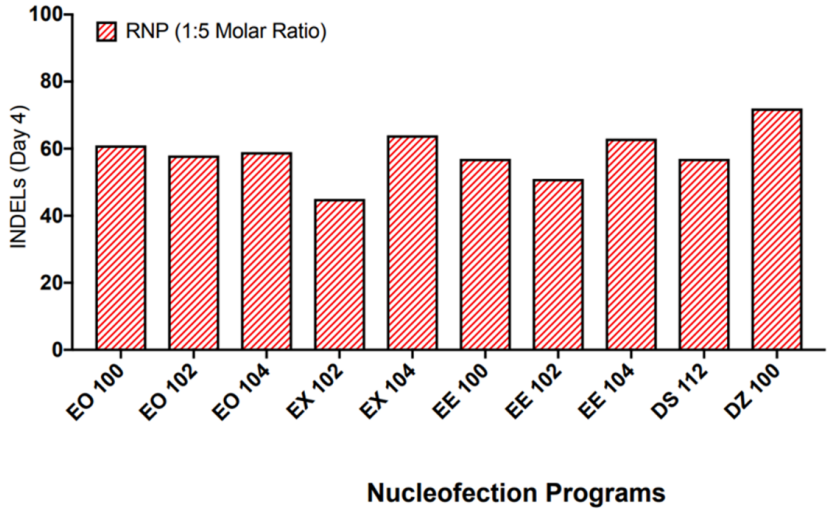



Figure 3.13: Ten electroporation programs were used to deliver the same amount of Cas9 RNP complex to HSPCs. Activity at the on-target site as noted by the presence and frequency of INDELs was analyzed by Sanger Sequencing with DZ100 performing the best.

determine either single or bi-allelic integration of the cassette Figure 3.14 **B**.

In addition to looking at the genotype of colonies from single cells, we devised a Sanger sequencing based approach to characterize the knock-in frequency from bulk gDNA. While not able to resolve hetero or homozygosity, this sequencing based approach is far cheaper and faster, and the outcome is in accordance with the data shown from methylcellulose colony formation assays. Briefly, the r.RAG2 cDNA cassette that we knocked into the genome was aligned to that of the wild type sequence at the same locus *in silico* Figure 3.15 **A**. While the sequence of the recoded cargo is diverged considerably from that of the wild type, there are stretches of DNA which remain identical and are long enough to design a short primer within. A reverse primer was designed within one of these sequence identical regions, with the forward primer being upstream on the chromosome in a region beyond the 400

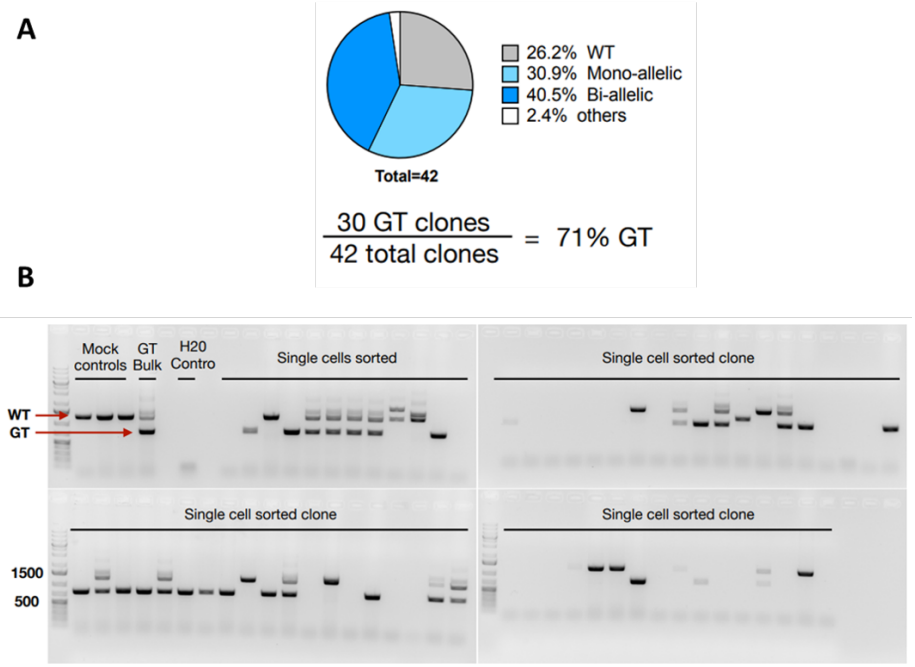



Figure 3.14: HSPCs cultured in methylcellulose supplemented media were interrogated by PCR for targeted integration of the r.RAG2 cassette. **Figure A** Quantification of all clones with either no knock in detected, one copy of the gene detected (1500bp and 500bp bands), or two copies of the gene detected (500bp band only). Some lanes produced unexpected bands and were labeled "Others". **Figure B** Gel image showing the PCR products of each clone analyzed for being either wild type (WT) or gene targeted (GT).

base pair homology arms, to preclude amplifying any episomal DNA delivered by the AAV. Using sequence deconvolution tools from the generated trace file, the HDR frequency of the r.RAG2 cassette was able to be resolved and a similar HDR frequency of 69% was reported Figure 3.14 **B**.




Figure 3.15: **Figure A** The wild type *RAG2* locus aligned to the theoretical knockin sequence of the r.RAG2 cassette. Grey areas indicate full alignment between both sequences. The forward and reverse primers are indicated by the small triangles. **Figure B** Sanger sequence deconvolution of alleles edited by HDR in samples receiving the RNP and AAV, only the AAV, and only the RNP.

### 3.4.5 Optimized Conditions May Abrogate Cytotoxicity in ATO System

Having further optimized all of the above conditions, we wanted to see if these ostensibly healthier edited cells could make it through the ATO system without the toxicity observed in prior attempts.

After applying the extended thawing and expanding conditions, both the healthy donor HSPCs as well as the *RAG2* null patient HSPCs were nucleofected with the

DS112 and DZ100 electroporation protocol before being transduced with the AAV virus containing the r.RAG2 cassette. Typically, three weeks into the differentiation protocol within the ATO system, a proportion of mature naive T cells are detectable expressing a TCR along with CD3. However in this case the three week timepoint did not seem to be sufficient to note robust expression of mature T cell surface markers. Indeed, looking at the healthy donor population of cells, we see an unusually high level of CD34 expressing cells which indicates the HSPC population has yet to be depleted, much less differentiated into T cells Figure 3.16. Looking at more terminal differentiation markers, we also see a limited population of immature single positive cells expressing CD4, further demonstrating that additional time would be needed to examine differentiation Figure 3.17. Overall the viability of the healthy donor cells was much lower than typically seen, which could be due to using an alternate lot of frozen healthy donor HSPCs which may also have affected the differentiation kinetics in this experiment. Alternatively, developing T cells may not have yet rearranged the beta chain of the TCR, which is the point at which thymocytes proliferate in the thymus.

Looking at the *RAG2* null and edited populations we see a more interesting result at the three week timepoint. Of note is a modest increase in the viability of cells edited with the newly established DS112 protocol when compared to the benchmark DZ100 program Figure 3.18. Additionally, the vast majority of these cells have progressed to co-express pre-T markers CD5, CD7, and CD1$\alpha$ despite there still being a residual amount of CD34 expressing HSPCs at this timepoint Figure 3.18. The mature T cell panel shows the cells beginning to shift from a CD4

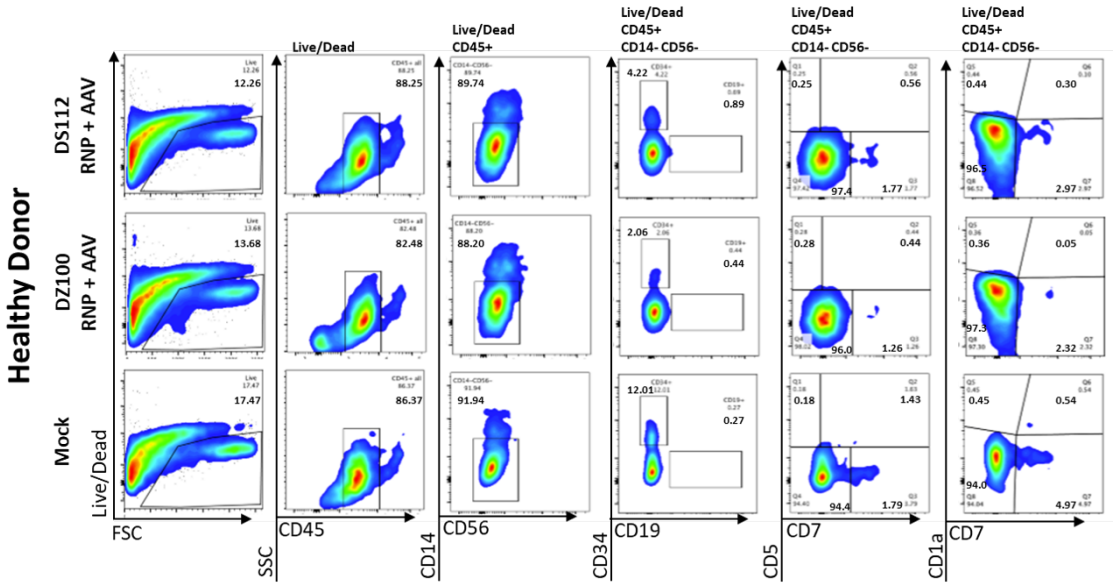



Figure 3.16: Flow cytometry data showing markers of early T cell development and differentiation of other cell types in the healthy immunocompetent donor and edited populations at 3 weeks.

immature single positive population to a CD4 CD8a double positive state, the next stage in T cell maturation Figure 3.19. A very small proportion of cells could be identified as co-expressing a TCR along with CD3, while this population was largely absent in the unedited controls Figure 3.19.

The results of this experiment were overall quite promising, particularly with regard to the marked decrease in cytotoxicity observed in the *RAG2* null edited patient cell population. Regrettably this experiment could not be carried out over a longer time period due to reagent delays at the onset of the ATO formation process. Initially this experiment was to be repeated, however lab closures resulting from the COVID-19 pandemic precluded this final experimentation.

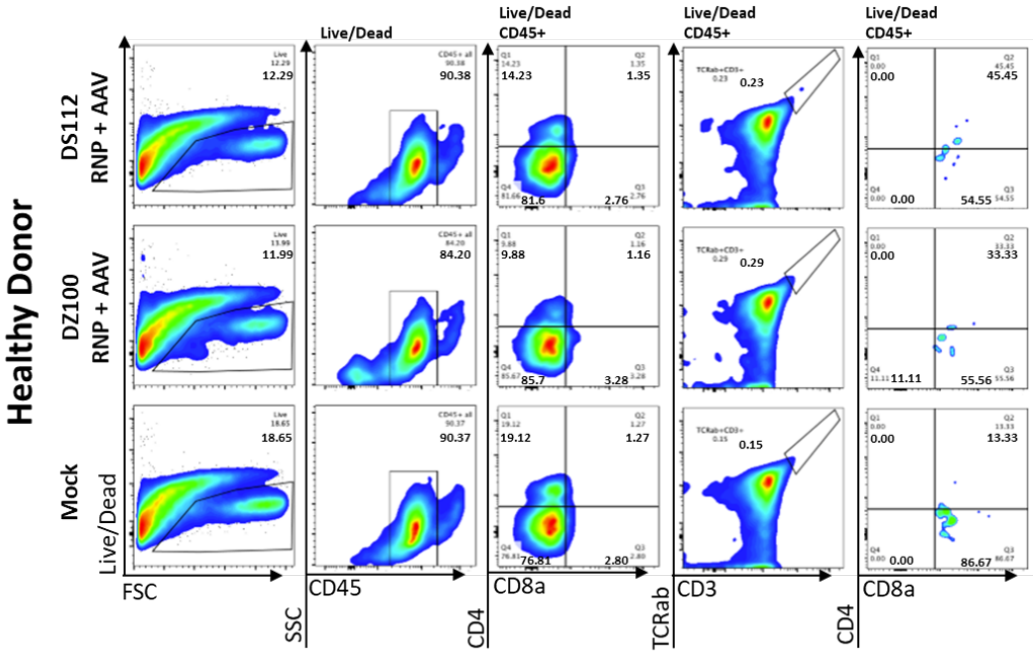



Figure 3.17: Flow cytometry data showing markers of mature T cell development in the healthy immunocompetent donor and edited populations at 3 weeks.

## 3.5 Discussion

The future of cellular and genetic therapies looks very promising with the addition of gene editing and next generation sequencing technologies. For CRISPR/Cas mediated therapies, one of the primary obstacles to overcome is the problem of immunogenicity. Most of the widely used CRISPR associated Cas proteins are derived from microbes. Microbial proteins face preexisting adaptive immune responses which makes Cas protein expression *in vivo* potentially problematic (Charlesworth et al. 2019). Non-canonical immune response environments such as the eye have an advantage for more immediate deployment of CRISPR/Cas technology, while most other tissue types face the issue of such immune response complications. A promis-

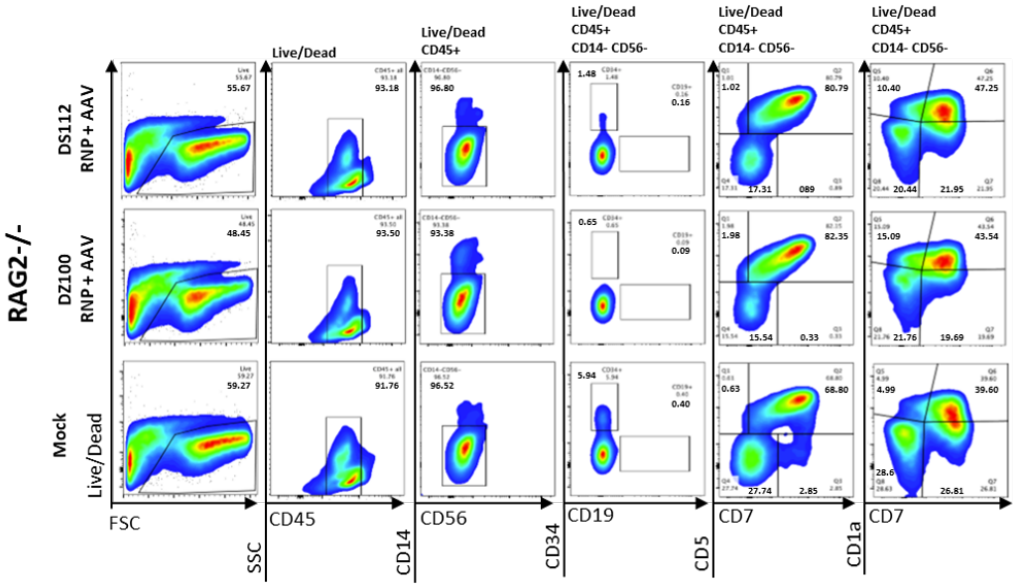



Figure 3.18: Flow cytometry data showing markers of early T cell development and differentiation of other cell types in the *RAG2* null and edited populations at 3 weeks.

ing alternative to *in vivo* editing is the strategy of *ex vivo* editing, wherein patient cells are extracted and edited outside of the body for subsequent reintroduction post modification. In this sense the body never has a chance to react to the Cas proteins, and the target cells still receive the gene editing machinery. Until immune recognition of Cas proteins can be attenuated or eliminated *in vivo*, the most promise for CRISPR/Cas technology lies in *ex vivo* approaches.

*Ex vivo* editing strategies may face fewer issues insofar as immune recognition, however there are still many challenges to be met in order to produce an efficacious edited cell therapy product. At the top of this list is the efficiency of gene editing. Many cell types are difficult to genetically manipulate due to the lack of established protocols or limited characterization of tissue specific tropism profiles for therapeuti-

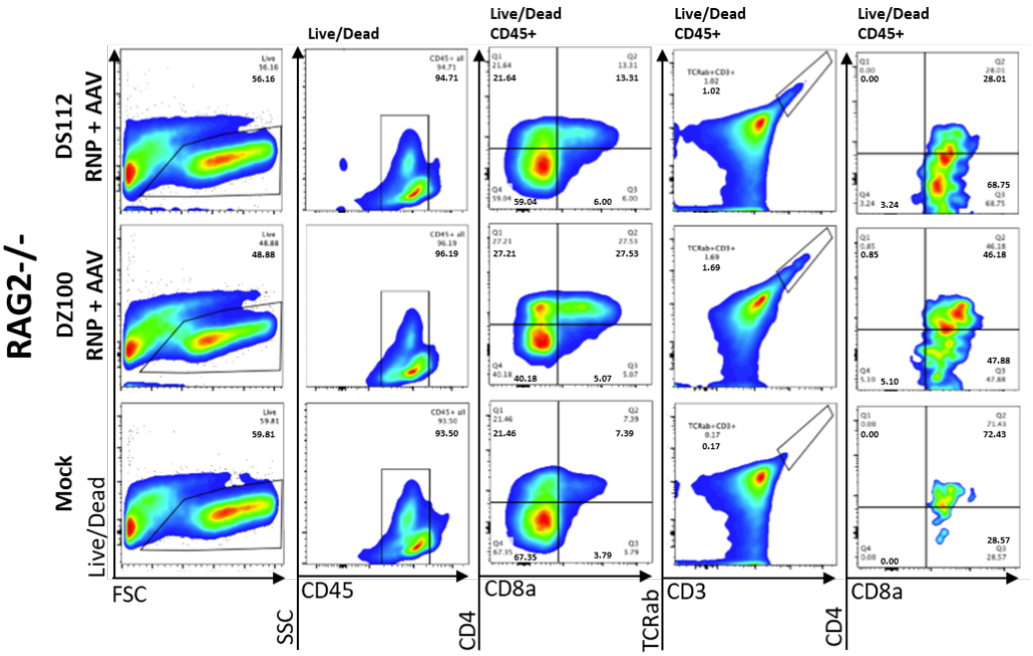



Figure 3.19: Flow cytometry data showing markers of mature T cell development in the *RAG2* null and edited populations at 3 weeks.

cally viable cargo delivering viruses. Fortunately, for blood diseases stemming from mutated Hematopoietic Stem and Progenitor Cell (HSPC)s, the tools for genetic manipulation are largely present.

*RAG2* deficiency falls into this blood disease category which can theoretically be addressed with gene editing approaches. Our work to produce and characterize edited patient cells still required a large effort to successfully reduce challenges presented by cytotoxicity in a 3D cell culture environment. Our data show that the processes we evaluated to be able to generate mature T cells from edited *RAG2* deficient patient cells were effective. The thaw and expansion time was extended to allow for HSPCs to establish a healthier homeostasis. The cellularity for nucleofection was

optimized to produce more HDR events. The electroporation program was altered to allow cells to continue proliferating after being disrupted. The transduction process was optimized to provide enough of the repair template. The summation of these processes in the ATO system intimate success, however it is indeed unfortunate to potentially have been one further week away from establishing this rescue more robustly by flow cytometry beyond the data presented in Figure 3.19. Nevertheless, we are hopeful that these pipeline improvements will yield rescued patient T cells which will demonstrate that an autologous edited cell therapy for *RAG2* deficiency is a prudent and needed improvement in transplantation treatment options.

# Chapter 4

# Gene Editing, Differentiation, and Characterization in Induced Pluripotent Stem Cell Derived T Cells

## 4.1 Modeling RAG Deficiency and Rescue with Patient iPSCs

### 4.1.1 Introduction

The introduction of reprogramming factors to derive Induced Pluripotent Stem Cell (iPSC)s from differentiated cell types created an entire field of disease modeling. iPSCs have been used to model a diverse swath of pathological drivers in fields from cardiac disease to Schizophrenia (C. Kim et al. 2013), (Brennand et al. 2011). The ability to study disease using iPSCs, however, impinges upon the availability and robustness of protocols to derive the target cell type. Looking specifically at adaptive immune deficiencies and related lymphocytes, there are only a few existing protocols to derive T and B cells. Protocols to derive T cells have been described

previously in this thesis work, but protocols to develop B cells have been even more limited in scope and in efficacy (French et al. 2015), (Vodyanik et al. 2005). A particular advantage of iPSC disease modeling is the ability to study pathology within patient cells as opposed to cells edited to contain a known pathogenic mutation. For example, a point mutation in *RAG2* could be engineered into a healthy donor derived iPSC line, and this mutation may recapitulate the loss of function observed in RAG deficient patients, however discrepancies between genotype and phenotype exist, and these epistatic factors can confound conclusions. For this reason and in this case, it is more compelling to use cells derived from a RAG deficient patient, as the clinical phenotype will have been well observed for the actual cells used in downstream experiments. This clinical data can also be used to benchmark the performance of patient iPSC derived immune cells when modeling RAG deficiency *in vitro*.

### 4.1.2 Objectives and Hypotheses

The objective for this portion of the thesis work was to demonstrate the efficacy of gene replacement strategy in rescuing the development of T cells derived from edited RAG deficient patient iPSCs, and to characterize those resultant cells. iPSCs can renew indefinitely, which allows for more robust engineering of components into the genome. Indefinite renewal also grants a longer window of opportunity to characterize edited cells, and allows us to pick edited clones to study without complications introduced from a heterogeneous pool of edited and unedited cells. We aimed to take advantage of these iPSC properties to robustly characterize development of the editing pipelines of engineered patient cells as well as to interrogate the function of

the recoded *RAG* gene in the cell through analyzing V-J pairings by NGS.

## 4.2 Characterizing Patient iPSCs

### 4.2.1 Reprogramming RAG2 Deficient Patient Cells

A male patient with a homozygous null c.831T>A mutation in *RAG2*, predicted to cause p.Y277X premature termination, presented with T-B- NK+ SCID. Fibroblasts from this patient were reprogrammed to pluripotency by Kerry Dobbs at the National Institutes of Health by infection with a non-integrating Sendai virus vector kit allowing transient expression of reprogramming factors OCT4, SOX2, c-MYC and KLF4. RAG2-deficient, patient-derived iPSCs were then expanded and subcloned. Use of a T-B- SCID patient presented the most stringent test case for rescue of T cell development, with obviation of any confounding factors produced by patients with some natural propensity to produce T cells such as those from leaky SCID patients or those with Omenn Syndrome (Shearer et al. 2014). Verification of pluripotency was observed by qPCR, using primers and an intercalating dye to quantitate levels of the mRNA from pluripotency genes being expressed natively in reprogrammed cells in which the Sendai reprogramming vector had been diluted out. When benchmarked against patient fibroblast mRNA, the reprogrammed cells showed significant differences in the expression of pluripotency markers as denoted by the CT value, or the number of cycles required for the signal produced by cDNA replication to be above a background threshold Figure 4.1 **A**. Verification of genomic integrity was also examined by G-banding. The karyotype was shown to be normal apart from one cell which demonstrated a non-clonal aberration as a result of a culturing

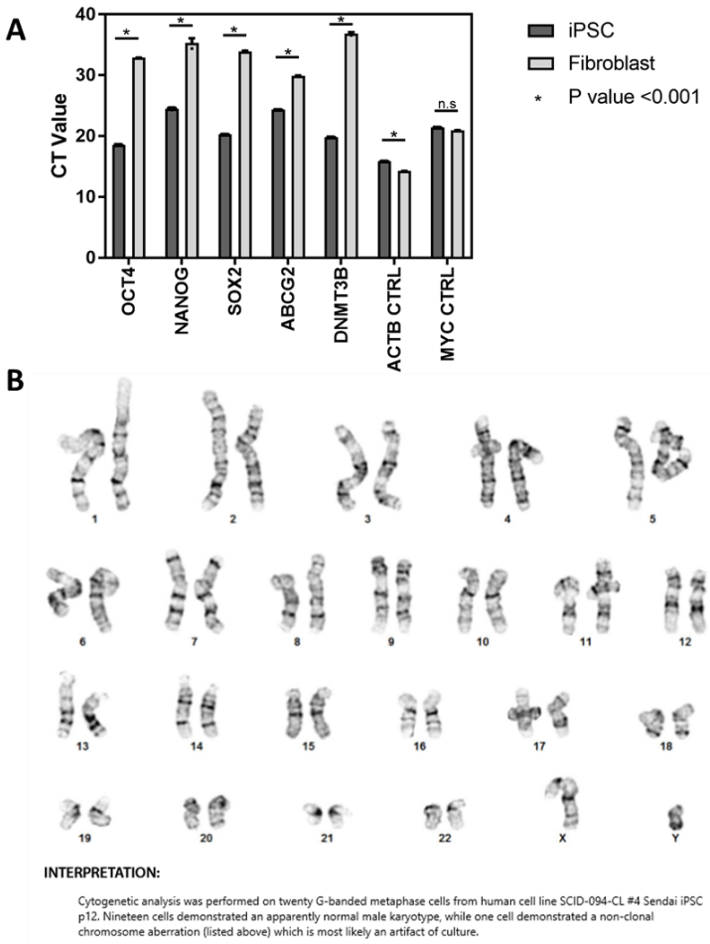



Figure 4.1: **Figure A** Representative qPCR data showing the CT values of pluripotency transcripts between reprogrammed iPSCs and control fibroblasts. **Figure B** Karyotyping of the *RAG2* deficient patient iPSCs, showing an apparently normal result apart from one cell which had a culturing artifact.

artifact Figure 4.1 **B**. This reprogrammed patient iPSC line was selected for use in downstream corrective investigation.

## 4.3 Genome Engineering in iPSCs

### 4.3.1 Electroporation in iPSCs

The viral cassettes introduced in the previous chapter containing the promoterless r.RAG2 cargo are the same used for experimentation here in patient iPSCs. In fact, many of the optimization lessons learned from our patient HSPCs work were carried over to iPSC editing. Electroporation conditions, however, needed to be altered to accommodate for a different cell type. Two primary nucleofection programs were selected based on existing literature and on the Lonza 4D Nucleofector suggestions (Byrne et al. 2014), (Steyer et al. 2018). Using a reporter GFP plasmid, programs CA-137 and CB-150 were interrogated in terms of resultant GFP expression as well as in viability of nucleofected cells. The electroporation buffer was also examined to determine if there was a more suitable buffer than the proprietary P3 buffer offered in limited quantities by Lonza. In parallel, we examined the number of cells electroporated and resulting effects on viability and GFP expression. We concluded that the nucleofection protocol CA-137 outperformed CB-150 in nearly every context in both healthy donor and *RAG2* deficient iPSCs Figure 4.1. For the patient iPSCs, the reduced serum media Opti-MEM served as a more efficient electroporation buffer. Opti-MEM is a significantly cheaper reagent than the Lonza provided P3 buffer, and so Opti-MEM was used for all electroporation experiments moving forward. We saw little difference in viability or GFP positivity as measured by flow cytometry due to differences in the either 250 or 500 thousand cells nucleofected. The results of this study suggested that we used Opti-MEM and nucleofection program CA-137, which we did for the remainder of gene editing experiments in iPSCs.

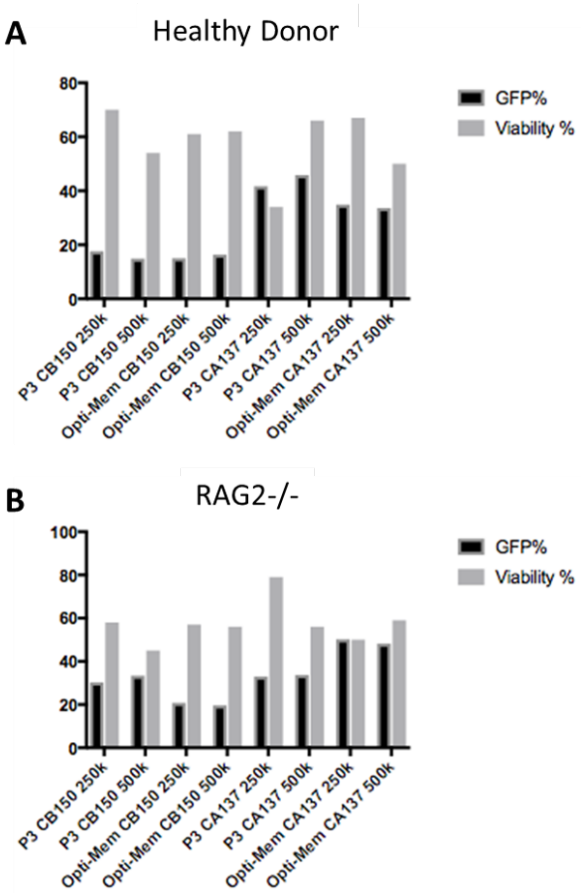



Figure 4.2: Electroporation data showing the GFP positive percentage of nucleofected cells as measured by flow cytometry, as well as viability measured by a hemocytometer for **A** a healthy donor iPSC line and **B** the RAG2 deficient patient iPSC line. n=1

### 4.3.2 CRISPR/Cas9 RNP Activity in iPSCs

With an appropriate nucleofection protocol in hand, we implemented the CRISPR/Cas optimization protocols derived from HSPCs into our iPSC editing pipeline. Using a 1:5 sgRNA to recombinant high fidelity Cas9 protein ratio, we examined delivery of the RNP complex into both healthy donor iPSCs as well as RAG2 deficient pa-

tient cells. We saw similar INDEL activities of 50.5% in healthy donor cells, and 50.0% in RAG2 deficient patient cells Figure 4.3. These INDEL results were similar not only to each other, but also to the editing activity detected in HSPCs. This was somewhat surprising given the inherent differences in epigenetic/methylation profiles of these three distinct cell populations. Prior work has shown that the accessibility of the complementary sgRNA sequence in the genome can be influenced by its exact positioning when wrapped around a nucleosome contained within either heterochromatin or euchromatin (Yarrington et al. 2018). Perhaps because expression of *RAG2* is turned off in both HSPCs and iPSCs, the epigenetic state and corresponding accessibility of this sgRNA sequence are similar.

### 4.3.3 CRISPR/Cas9 AAV Mediated Knock-in of the r.RAG2 Cassette in Patient iPSCs

With knowledge in hand of nucleofection optimization and RNP on-target activity, we were able to incorporate some of the optimization metrics demonstrated in patient HSPCs to our goal of knocking in the same r.RAG2-GFP cassette into patient iPSCs, namely the molar ratio used, the amount of time between nucleofection and transduction, and the number of cells electroporated. The same sgRNA was used in iPSCs as was used for HSPC experimentation to knock in the identical r.RAG2-GFP cargo into the endogenous *RAG2* locus Figure 4.5 **A**. Interestingly, we noticed a complete lack of any GFP signal post-transduction, even at the highest MOIs tested Figure 4.5 **B**. Of note is the cytotoxicity from the highest MOI of 200K shown, which caused a shift in granularity of the cells, pushing a subset of these transduced cells erroneously into the GFP+ gate. This lack of GFP indicated to

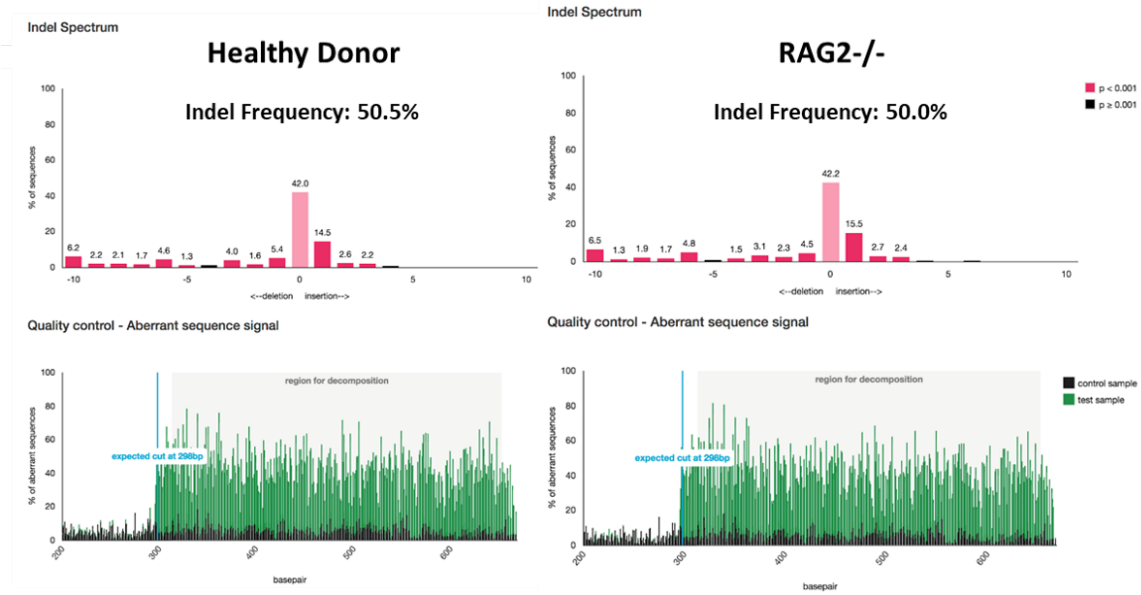



Figure 4.3: Sanger sequencing analysis of the sgRNA binding site in healthy donor and RAG deficient patient iPSCs. The top panels represent the indentity of INDELs, with the most frequent INDEL reported as a +1 basepair insertion. The lower panels show the difference in sequencing alignment confidence calls between a control (black) which did not receive the sgRNA Cas9 RNP and those that did (green), with expected divergence along the indicated sgRNA cut site. This data was generated using the TIDE webtool available at tide.nki.nl

us that the virus was not integrated into the genome. Because this AAV serotype is known to transduce iPSCs efficiently and because we saw robust HDR mediated editing in HSPCs, we looked into the genome of edited iPSCs instead via ddPCR to profile any integration events. What we noticed was relatively frequent integration of the cargo at rates of around 40% Figure 4.5 **C**. This indicated to us that the SFFV promoter in front of the GFP reporter was being epigenetically silenced in iPSCs, as has been shown in prior work (Pfaff et al. 2013), (Herbst et al. 2012). Through the process of sib-selection, pools of edited iPSCs containing integrated copies of the r.RAG2 cassette were enriched for by ddPCR screening. Sib-selection is the process

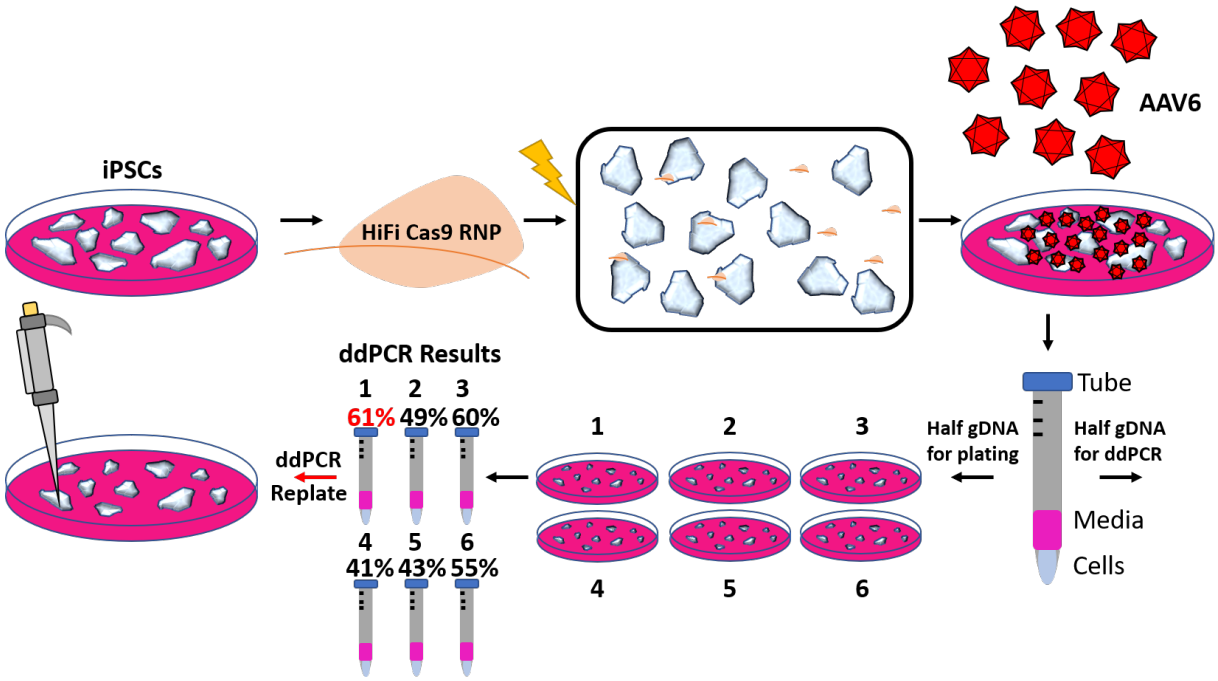



Figure 4.4: Overview of the genome engineering pipeline detailing iPSC culture, nucleofection with the SpCas9 sgRNA RNP complex, transduction with AAV6, and the process of sib-selection detailed in the bottom row which detects the percentage of HDR events by ddPCR and can be repeated to continue enrichment of edited cells before picking clones.

of examining pools of cells, in this case by ddPCR, to look for cell populations with the highest frequency of the desired characteristic, in this case rates of integration Figure 4.4. These cell populations are then passaged into additional wells, where the process repeats. After several passages, sib-selected cells were enriched to over 80% r.RAG2 containing cells Figure 4.5 **D**.

Enriched cells were then diluted and allowed to grow into well-spaced colonies. Resultant colonies were picked and screened by PCR for verification of the full insertion and of the hetero or homozygosity of the integrated r.RAG2 cassette in every clone. Primers were designed to assay integration of the 5' end of the insertion

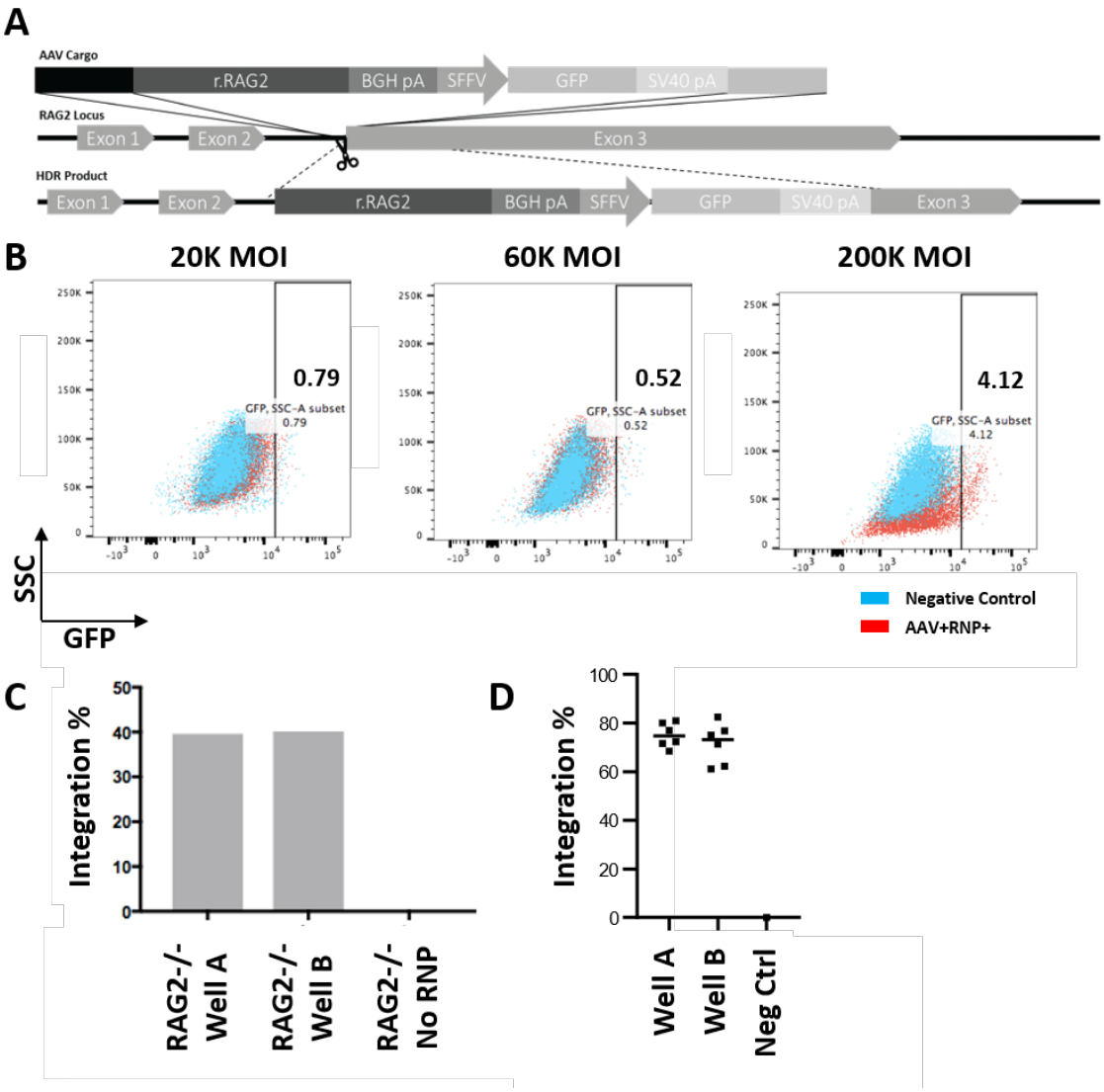



Figure 4.5: **A** Overview of the genome editing design detailing the intended knock-in of our r.RAG2 construct into the *RAG2* endogenous locus. **B** Flow cytometry data examining expression of GFP resulting from r.RAG2-GFP integration against a negative control. **C** Initial genomic integration rates of the r.RAG2 cargo by ddPCR. **D** Detection of genomic integration rates of the r.RAG2 cassette by ddPCR after sib-selection of edited cells from the precursor heterogeneous population shown in **C**.

junction, the 3' end of the insertion junction, and over the insertion site, to assay if clones had zero, one, or two copies of the r.RAG2 cargo integrated into the genome Figure 4.6 **A**. All three clones showed incorporation of the full r.RAG2 cassette by ddPCR of the 5' junction and PCR of the 3' junction site Figure 4.6 **B**. All three clones also showed bi-allelic integration, with no amplicons noted for the PCR over the insertion site apart from in the unedited control Figure 4.6 **C**. Clone two looked slightly cleaner and so was grown up to use in downstream experiments.

Clone two of the *RAG2* deficient edited patient iPSC line was karyotyped to ensure that no cytogenic abberrations were introduced through this large cassette integration process. All cells examined demonstrated a normal karyotype, indicating there were no large scale erroneous and detectable DNA damage events resulting from the gene editing process Figure 4.7.

## 4.4 Analysis of Edited iPSC Line and Isogenic Precursor by WGS

Most programmable nucleases can cause unwanted genomic effects stemming from activation of DNA repair pathways which can result in off-target double-stranded DNA breaks, chromosomal translocations, large deletions, and other aberrations as reviewed in (Jang et al. 2020). While detection of the targeted integration of a transgene is possible using simple primer designs, a more robust method to detect both transgene insertion and possible off-target effects is through WGS. WGS profiles the entire genome and reports variants from that of a reference genome. In this case, we sequenced both the *RAG2* deficient patient iPSC line as well as a clone of that line

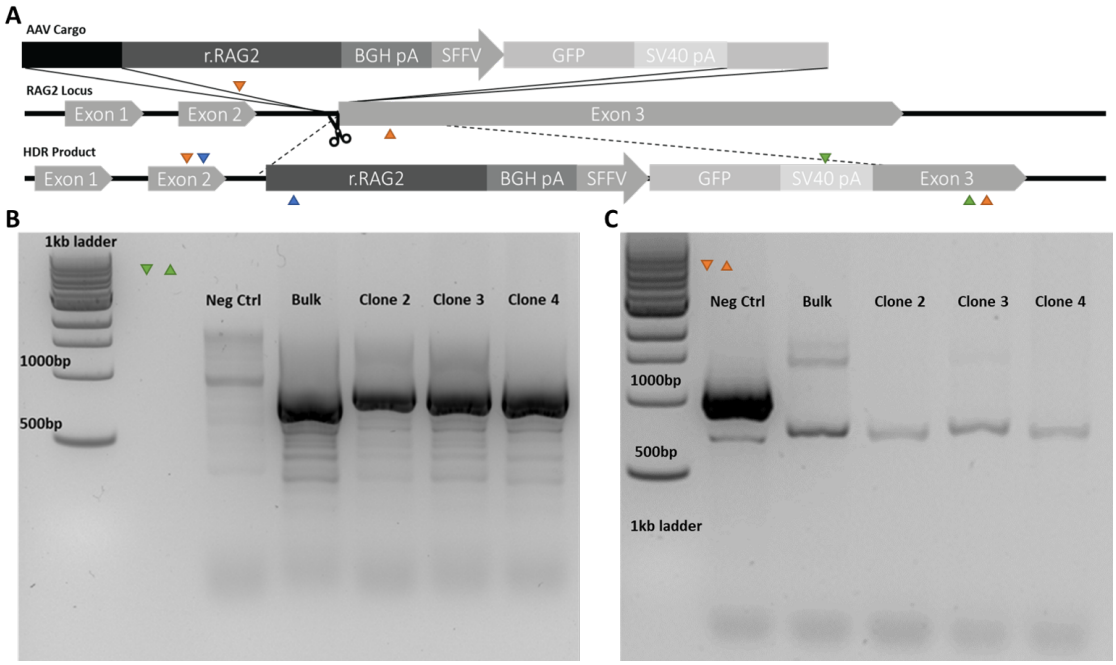



Figure 4.6: **A** Location of each primer pair for assaying cargo insertion into the genome. The blue triangles denote the primers used for all ddPCR results (Figure 4.5C, 4.5D). The green triangles denote the primer pairs used to assay the 3' insertion junction. The orange triangles denote the primer pair used to assay homozygosity by PCRing over the insertion site. **B** DNA electrophoresis gel image showing PCR amplification of clones containing or not containing the expected 883bp band corresponding to a 3' integration denoted by the green triangles. **C** DNA electrophoresis gel showing PCR amplification over the insertion site of the r.RAG2 cassette. Clones which have one or both alleles unmodified by the r.RAG2 cargo produce the expected 1062bp band. Locations of these primers are indicated by the orange triangles.

which had the r.RAG2-GFP cassette inserted on both alleles. In this way we were able to subtract out SNPs, or point mutations, present in the isogenic precursor from those in the edited cell line, which gives a better representation of any potential CRISPR associated mutations that may have been introduced during editing. Potential off-target sites were evaluated based on prediction algorithms which have

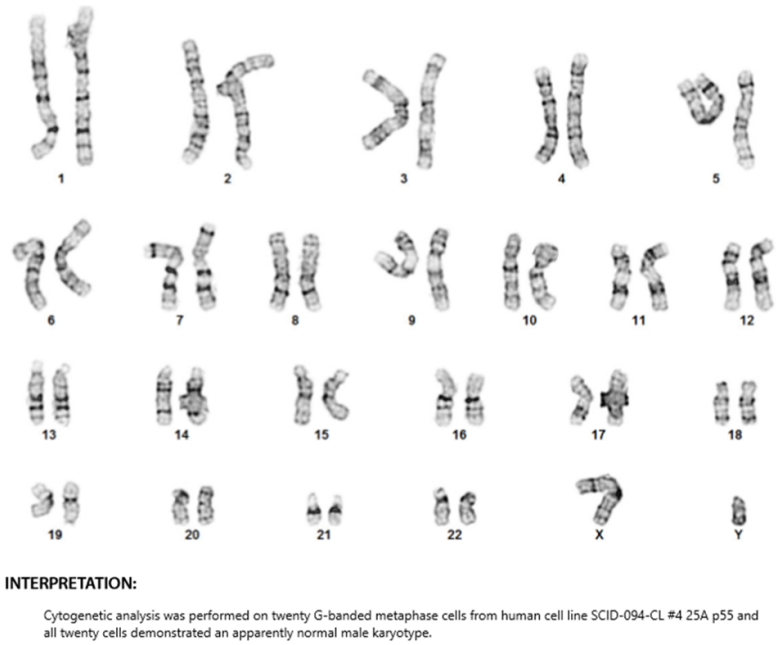



Figure 4.7: G-banding of the clonal *RAG2* deficient iPSC line edited with the r.RAG2 cassette on both alleles. No cytogenic aberrations were reported for any of the cells examined.

been previously described (Hsu et al. 2013), as well as by using the BLAST tool from NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to search the genome for any regions which share regions homologous to the sgRNA used for editing. Processing of the WGS data by Justin Lack at the National Institutes of Health revealed no INDELs associated with off-target events within 45kb of predicted regions of activity Figure 4.8.

Programmable nuclease associated large scale deletions or aberrant chromosomal events such as inversions, translocations, and truncations have been reported and can be difficult to detect with some WGS approaches (Cho et al. 2014), (D. Kim et al. 2015). Indeed, canonical WGS technologies, such as those used in this

| OT_chromosome | OT_position | mutation_chromosome | mutation_start | mutation_stop | distance | abs_distance |
|---|---|---|---|---|---|---|
| chr13 | 93562138 | chr13 | 93513373 | 93513374 | -48764 | 48764 |
| chr15 | 55631389 | chr15 | 55559800 | 55559801 | -71588 | 71588 |
| chr19 | 32798768 | chr19 | 32675809 | 32675810 | -122958 | 122958 |
| chr8 | 134177399 | chr8 | 134341593 | 134341594 | 164195 | 164195 |
| chr11 | 65503728 | chr11 | 65740110 | 65740111 | 236383 | 236383 |
| chr10 | 37058449 | chr10 | 37305186 | 37305187 | 246738 | 246738 |
| chr2 | 143993629 | chr2 | 144244174 | 144244175 | 250546 | 250546 |
| chr18 | 67938254 | chr18 | 67687278 | 67687279 | -250975 | 250975 |
| chr8 | 126433050 | chr8 | 126688975 | 126688976 | 255926 | 255926 |
| chr5 | 115098963 | chr5 | 114427659 | 114427660 | -671303 | 671303 |
| chr21 | 17476424 | chr21 | 18187117 | 18187118 | 710694 | 710694 |
| chr6 | 131394207 | chr6 | 130647377 | 130647378 | -746829 | 746829 |
| chr1 | 230258671 | chr1 | 231024958 | 231024959 | 766288 | 766288 |
| chr2 | 156357042 | chr2 | 155557328 | 155557329 | -799713 | 799713 |

Figure 4.8: Evaluation of any off-target (OT) mutations present in the edited *RAG2* deficient iPSC line edited with the r.RAG2 cassette on both alleles. Mutations shared by the isogenic precursor which did not receive the RNP were subtracted before OT analysis. The location and distance in basepairs to the nearest mutation to that of the predicted off-target site is shown from nearest to the predicted OT site to the farthest (abs_distance).

thesis work, make use of short reads produced by tagmentation or other genomic DNA shearing methodologies resulting in read lengths of 600bp or less (Amarasinghe et al. 2020). Sequences generated from these short genomic fragments are then assembled and mapped to a reference genome to start to build coverage of the submitted sample. Despite modern bioinformatics algorithms, when a more complicated CRISPR associated off-target event takes place such as a large deletion or other structural variant, these short reads can encounter difficulty being assembled and mapped to their genomic location (Treangen et al. 2011). Further, short read WGS has limitations in homopolymer regions of the genome (Mantere et al. 2019). A more recently deployed technology called long read sequencing can resolve the sequence identity of single molecules in excess of 10kb, dwarfing the 600bp read

length of traditional WGS based approaches (Amarasinghe et al. 2020). Long read sequencing is then better positioned to address more complex genomic abnormalities which may result from gene targeting, as such reads can span and report the distance from an abnormal event site onto a region of the genome which was otherwise unaffected. Long read sequencing comes with some disadvantages, such as lower accuracy, lower throughput, less coverage of the genome, fewer tools available for analysis, and higher costs, however progress in some of the nanopore based sequencing technologies which produce these longer reads continues to make this sequencing approach more appealing to many.

While long read sequencing was not performed for this study, paired-end variant calling using short read WGS along with karyotyping results from the edited and precursor iPSC lines used in this study did not detect chromosomal aberrations or off-target edits, suggesting genomic integrity of the iPSC lines used. However because specific structural variant analyses were not conducted for the short read WGS data generated from these cell lines, and limitations exist in short read processing for structural variants, it is possible that such variants exist and went unreported.

## 4.5 Optimizing iPSC ATO System

With our control, *RAG2* deficient, and *RAG2* deficient edited patient iPSC lines in hand, we had all of the components necessary to interrogate RAG function in the Artificial Thymic Organoid (ATO) system. While similar to the ATO protocol to generate T cells from HSPCs, the iPSC ATO protocol has a few additional steps. iPSCs are upstream of HSPC differentiation, and so the primary difference

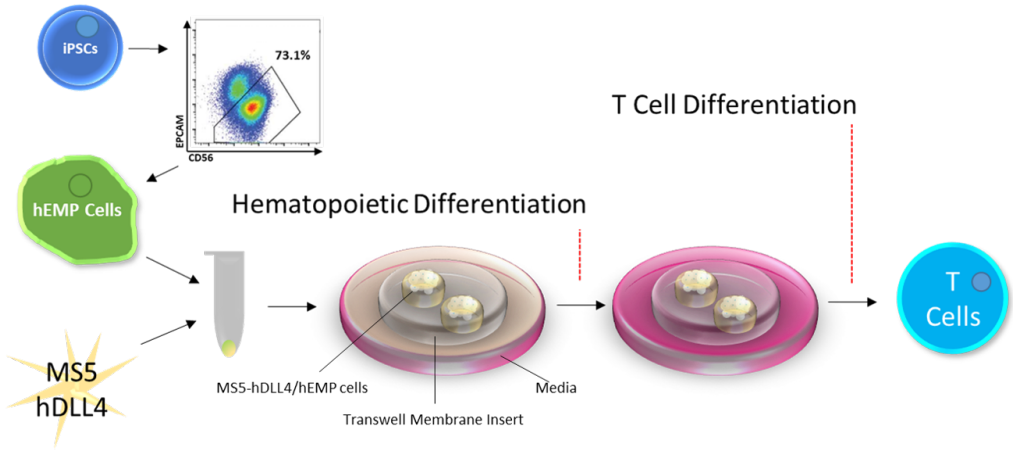



Figure 4.9: High level overview of the process of generating mature naive T cells from iPSCs using the 3D thymic organoid system.

between the two protocols is the need to derive an intermediate Human Embryonic Mesodermal Progenitor (hEMP) cell state from the iPSCs. hEMP cells are a precursor differentiation state to hematopoietic lineage cells. This hEMP intermediate cell state is quantifiable by the expression of membrane receptor CD56 and the absence of surface marker EPCAM and as such, hEMP cells can be sorted by Fluorescence Activated Cell Sorting (FACS). Sorted iPSC derived hEMP cells are then co-cultured with the same murine stromal cell line (MS5) expressing human Notch ligand Delta-like 4 (MS5-hDLL4) and seeded into the ATO system in media promoting first hematopoietic lineage commitment, then T cell maturation in a fashion similar to the HSPC input protocol from the prior chapter Figure 4.9.

### 4.5.1 hEMP Induction from iPSCs

During the optimization of the gene editing pipeline in *RAG2* deficient patient iPSCs, a preliminary experiment was conducted to investigate the efficiency with

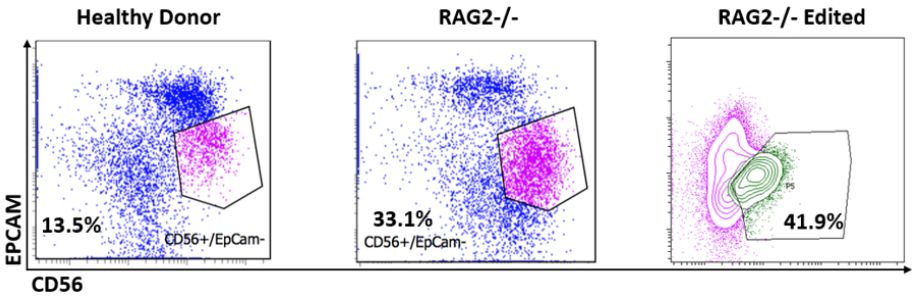



Figure 4.10: Preliminary hEMP induction experiment performed and analyzed by flow cytometry on a healthy donor iPSC line as well as the *RAG2* deficient patient iPSC line. Results show limited efficiency of differentiation.

which we could differentiate, or induce, this hEMP intermediate denoted as CD56+ EPCAM- by flow cytometry. Having not yet generated and isolated the patient cells containing the r.RAG2 insertion, we at first used only the isogenic precursor *RAG2* deficient cell line along with a healthy donor control. When the edited clones were isolated shortly thereafter, a similar experiment was repeated with that cell line. What we found is that we were able to induce the hEMP intermediate cell type, but only to a limited extent Figure 4.10. Prior art demonstrated over 50% hEMP differentiation efficiency before proceeding to differentiate T cells from hEMP precursors (Montel-Hagen et al. 2019).

### 4.5.2 Impaired T Cell Development in the ATO System from Low-Frequency hEMP Precursors

hEMP cells were seeded into the ATO system despite levels of hEMP differentiation being below that reported by the protocol originators. The differentiation from hEMP cells observed was severely impaired in all three cell lines including the healthy donor control, the *RAG2* deficient patient, and the *RAG2* deficient edited iPSC line.

What was interesting, however, was the stratification in developmental blocks within this 3D organoid system. The healthy donor iPSC line did not progress at all. The cellularity noted was surprisingly low, with only a few cells progressing to express the early leukocyte development marker CD45 Figure 4.11. The *RAG2* deficient patient line showed slightly better viability and cellularity, however CD45 expression was also similarly impaired, with a surprisingly elevated population of NK cells as noted by expression of CD56 Figure 4.12. This was an interesting observation because the *RAG2* deficient patient also showed elevated levels of NK cells. Still, this cell line did not progress to a pro-T cell phenotype with the demonstrated absence of CD1$\alpha$, CD5, and CD7. Finally, the *RAG2* deficient edited iPSC line did manage to progress through some stages of T cell development. Despite the low number of cells produced, the majority of live cells did express CD45 with the reversal of the higher presence of NK cells demonstrated in the isogenic precursor. Edited cells also showed expression of CD7 and CD5, indicating maturation to the pre-T stage of development. A few developing cells additionally showed expression of CD4 and CD8a, becoming double positive cells, though no cells showed clearance of this stage of development to display a TCR along with co-receptor CD3 Figure 4.13.

### 4.5.3 Improving hEMP Differentiation

Discussion with our collaborators who pioneered the 3D artificial thymic organoid system for iPSC input resulted in the establishment of a threshold for hEMP differentiation efficiency in order for cells to then progress into the ATO system. In their hands, an efficiency of 50% or less led to unfavorable differentiation outcomes. Each full iPSC ATO experiment takes more than 8 weeks to conduct, so we began trou-

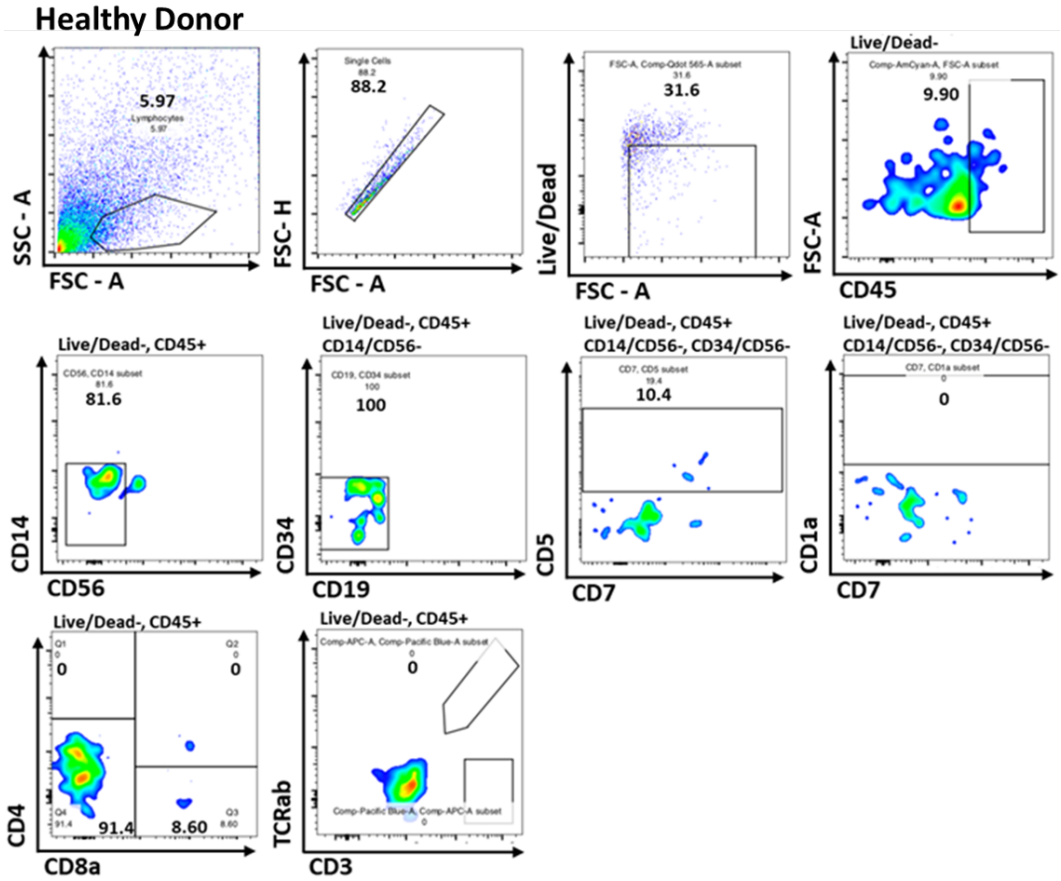



Figure 4.11: Representative flow cytometry data analyzing the T cell development of a healthy donor control iPSC line in the 3D thymic organoid system at week 3.

bleshooting with this shorter process of optimizing hEMP induction. The primary factor influencing the efficiency of differentiation was the length of induction. Other factors we analyzed were confluency of iPSCs prior to starting the induction process and the concentration and composition of cytokines used to stimulate differentiation. Lengthening the period of hEMP induction from 3 days to 5 days produced a significant enrichment for cells expressing CD56 and lacking EPCAM, clearing the 50% suggested hEMP threshold for all cell lines Figure 4.14.

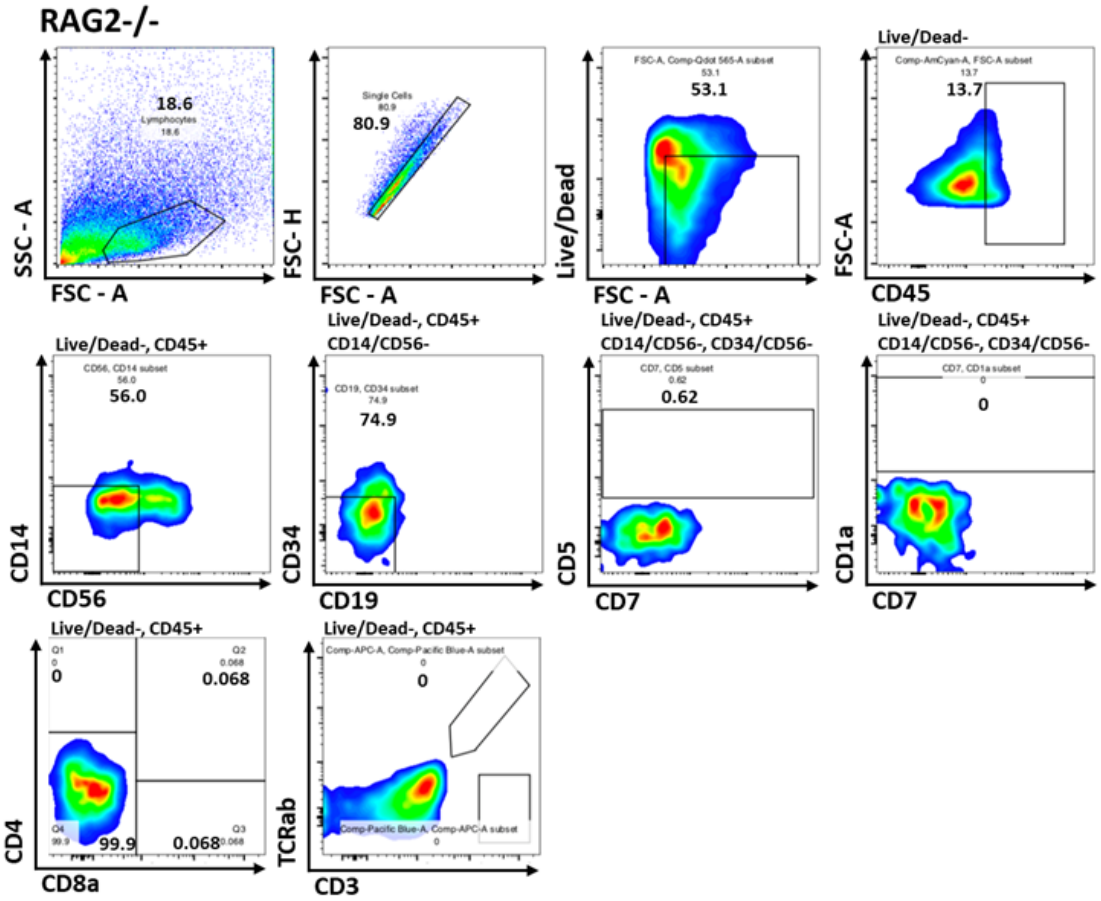



Figure 4.12: Representative flow cytometry data analyzing the T cell development of our *RAG2* deficient patient iPSC line in the 3D thymic organoid system at week 3.

## 4.6 r.RAG2 Rescues T Cell Development *In Vitro*

### 4.6.1 r.RAG2 Edited Cells Demonstrate TCR Rearrangements Leading to Clearance of DP Block *In Vitro*

After significant and lengthy troubleshooting, generated hEMP cells were seeded once again into the ATO system. The healthy donor, *RAG2* deficient, and *RAG2* deficient edited patient iPSC lines were examined after four weeks of T cell differ-

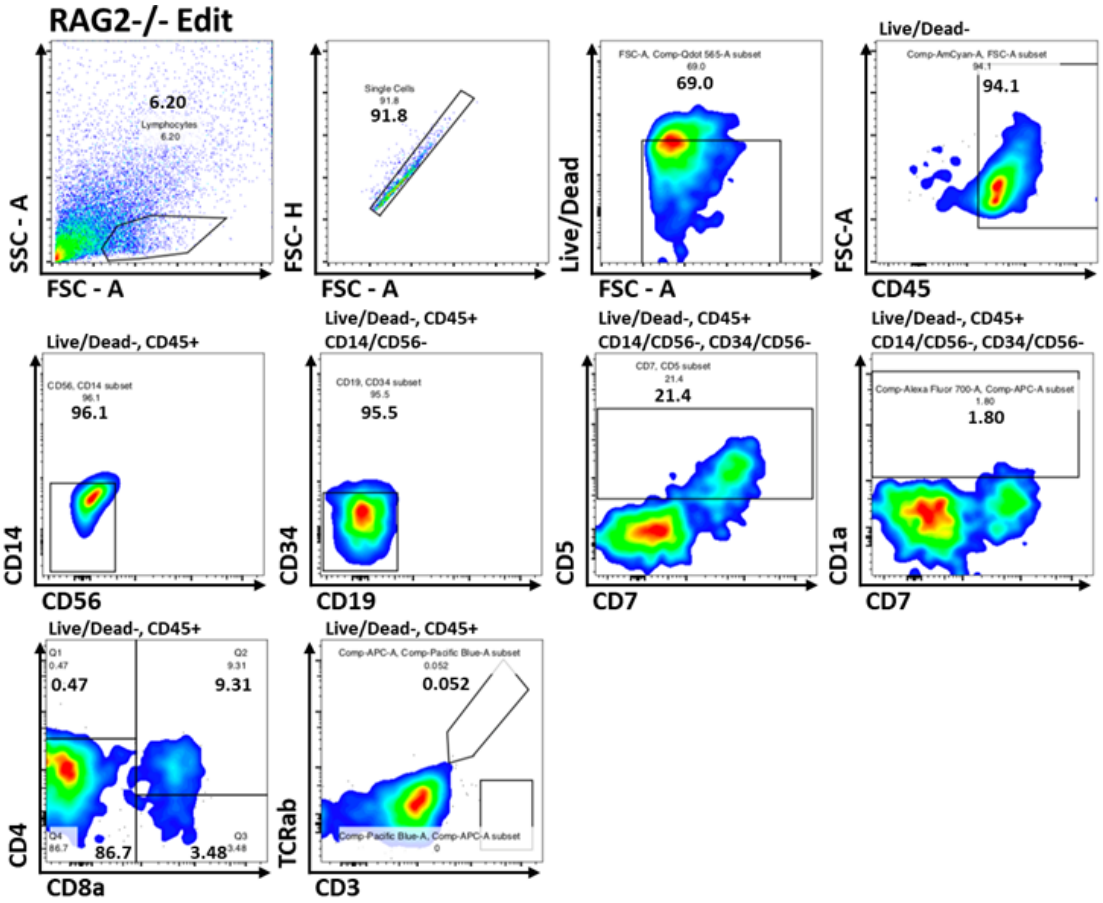



Figure 4.13: Representative flow cytometry data analyzing the T cell development of our *RAG2* deficient patient iPSC line edited with the r.RAG2 cassette in the 3D thymic organoid system at week 3.

entiation. The new set of results generated from optimizing the hEMP induction protocol showed an increase in viability and differentiation capabilities across all cell lines. The healthy donor line showed robust reconstitution when compared to the prior iteration, with more than 40% of the cells being viable. Lymphocytes differentiated in the ATO system showed almost complete commitment to the T cell lineage, with the vast majority of cells expressing CD45, CD5, CD7, and CD1$\alpha$. Fewer than

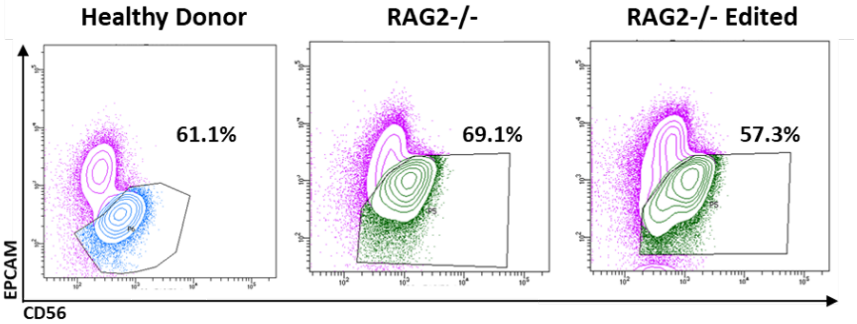



Figure 4.14: Flow cytometry data showing the results of an optimized hEMP induction period for the healthy donor, *RAG2* deficient, and *RAG2* deficient edited iPSC lines.

1% of the cells expressed non-T cell markers when examined by this antibody panel, including markers for monocytes, NK cells, HSCs, and B cells Figure 4.15. Indeed, the majority of cells expressed a TCR, with nearly 60% co-expressing CD3. While most of these cells had not progressed to becoming single positive cells, there was a slight shift from the CD4 CD8 double positive population towards a CD8 single positive mature naive T cell phenotype. Also noted were physiologically relevant number of TCR$\gamma\delta$ cells coming out of the system, suggesting a robust differentiation platform and significant improvements from earlier attempts.

The *RAG2* deficient patient iPSC line continued to struggle in the 3D organoid system. Despite physiological recapitulation of patient presentation, the number of cells generated in the ATO system was limited. This is likely due to developing cells failing to survive the rudimentary beta selection and positive selection occurring in

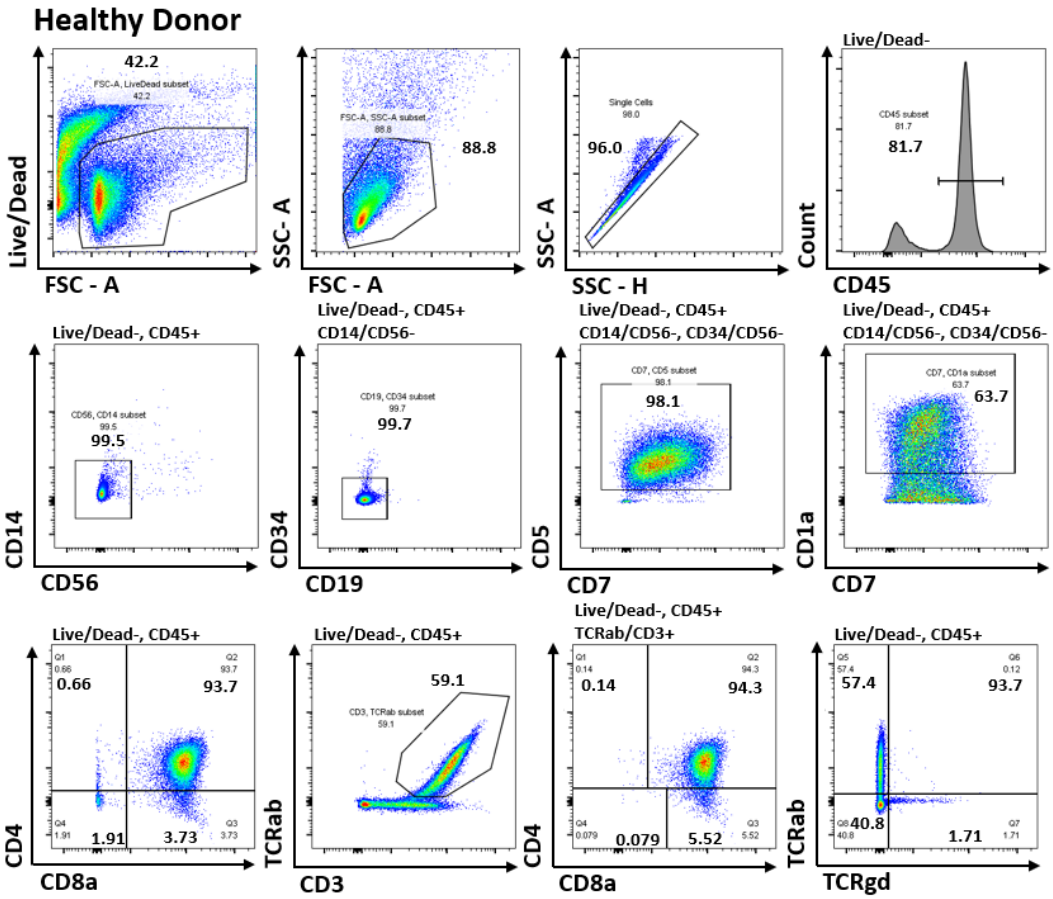



Figure 4.15: Representative flow cytometry data analyzing the T cell development of our healthy donor iPSC line in the 3D thymic organoid system at week 4.

the thymic organoid due to their lack of TCR rearrangement. When RAG rearranges the beta chain, the pre-TCR formed with an invariant alpha chain allows for developing lymphocytes to proliferate massively in the thymus, with the hope that these functionally rearranged beta chains will then be able to produce a functional alpha chain as well. Therefore without beta chain rearrangement it is not surprising that this *RAG2* deficient cell line does not generate many cells. Altogether, the few surviving developing cells do progress through to expressing pre-T markers CD5

and CD7, with limited CD1$\alpha$ expression, with many of those cells also expressing CD4 and CD8, either as ISPs or as DP cells, which is the terminal differentiation state possible for RAG deficient cells Figure 4.16 (Bosticardo et al. 2020). Perhaps unsurprisingly, developing hEMP derived cells seem to compensate for an inability to generate intended T cells as noted by a much higher percentage of CD45 negative cells, though the identity of these CD45 negative cells has not been defined. Additionally, there is a much higher skewing of monocyte development with cells expressing CD14, indicating there may be some sort of compensatory immune cell differentiation pathway.

Results from the differentiation of the *RAG2* deficient patient iPSC line edited with the r.RAG2 cassette showed a dominant reversal of the block in T cell differentiation observed in the isogenic precursor. In addition to the considerable viability and cellularity results, almost all observed cells progressed to T cell lineage commitment, with more than 90% of cells expressing CD45 and pre-T markers CD5 and CD7 Figure 4.17. The majority of these cells also co-expressed CD1$\alpha$, while very few cells expressed non-T cell development markers. Most importantly, edited cells robustly co-expressed TCR$\alpha\beta$ and co-activator CD3, with a clear population of DP cells as well as small demonstrable populations of CD4 single positive and CD8 single positive mature naive T cells. Fewer TCR$\gamma\delta$ cells were reported compared to the immunocompetent control, however this population was still observed to some extent. Rescued *RAG2* deficient cells actually displayed higher frequencies of TCR$\alpha\beta$+/CD3+ cells than our healthy donor control, and all prior experiments in the ATO system using HSPCs, suggesting very robust RAG function in developing

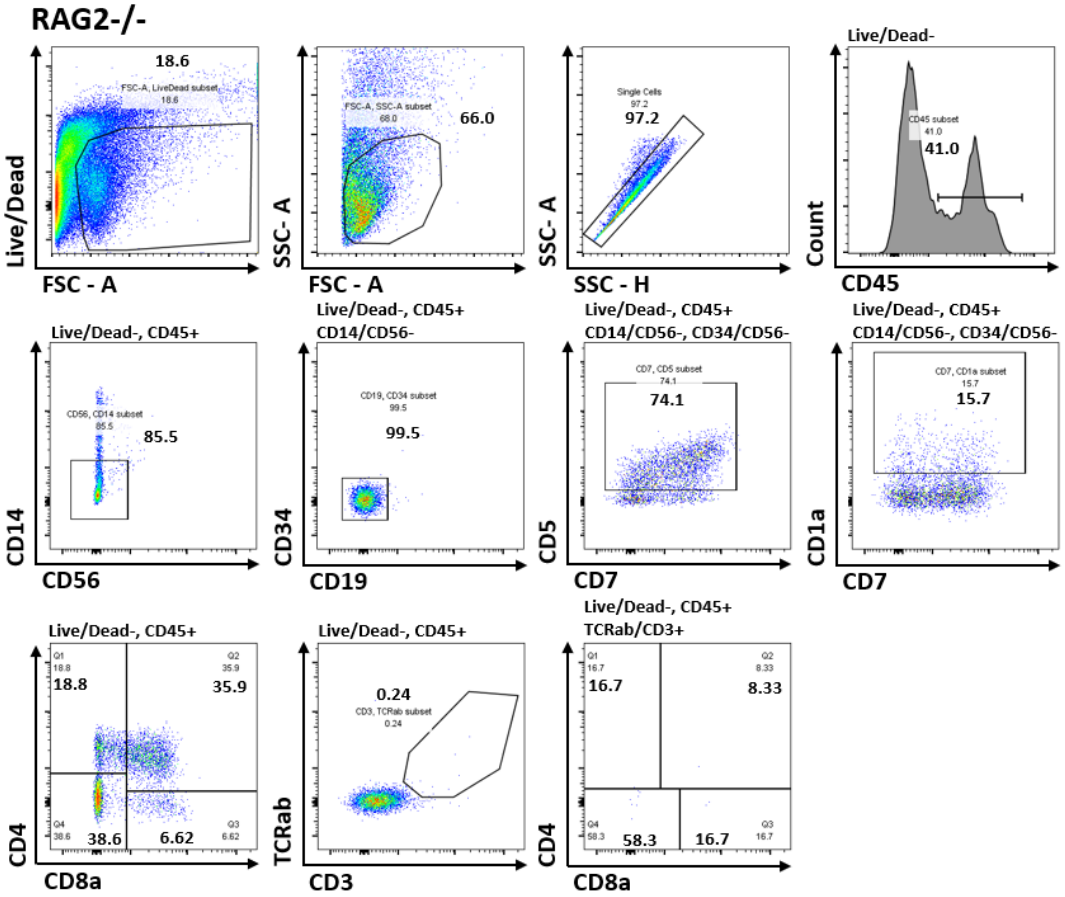



Figure 4.16: Representative flow cytometry data analyzing the T cell development of our *RAG2* deficient patient iPSC line in the 3D thymic organoid system at week 4.

edited cells.

A comparison of mature T cell markers displayed on all three cell lines shows more clearly the reversal of the T cell block in development between the *RAG2* deficient patient line and the cells rescued with the r.RAG2 cassette Figure 4.18 **A**. In addition, the comparative quantitation of cells per ATO demonstrates the lack of proliferation capabilities in the *RAG2* deficient cell line when compared to cells

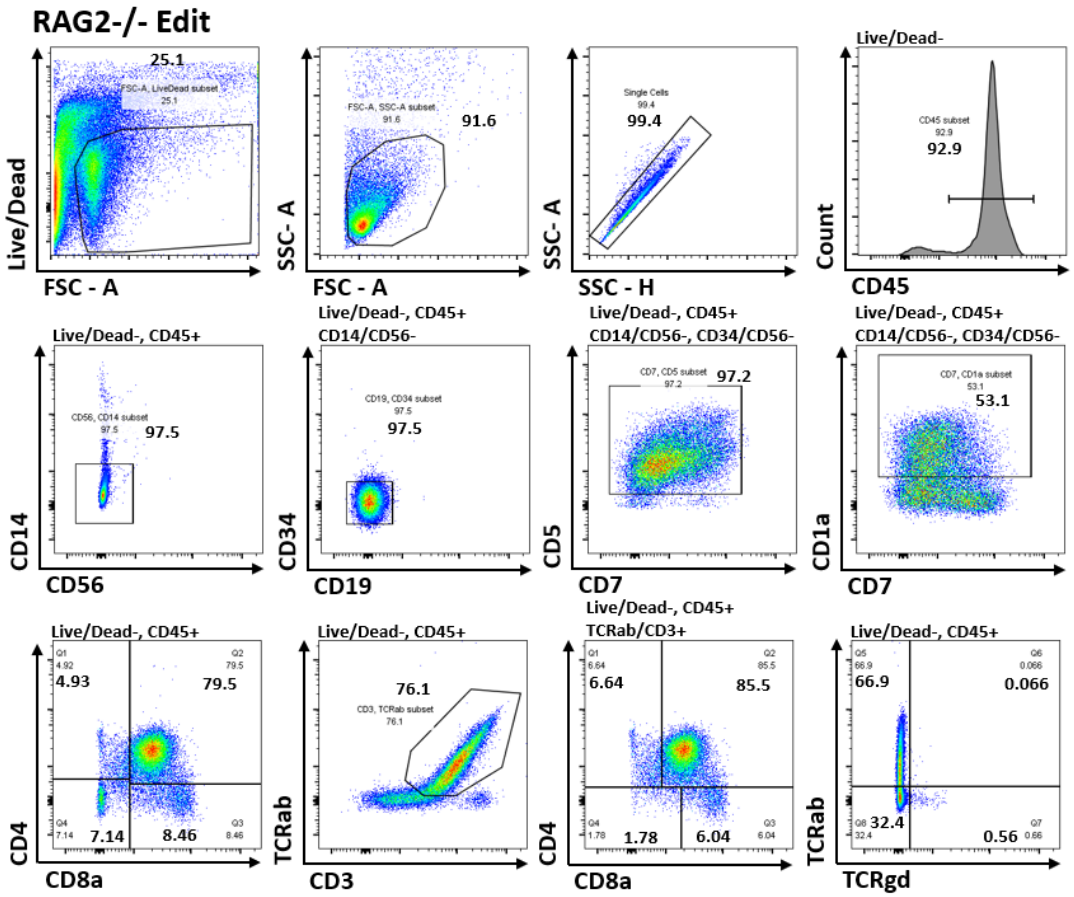



Figure 4.17: Representative flow cytometry data analyzing the T cell development of our *RAG2* deficient patient iPSC line edited with the r.RAG2 cassette in the 3D thymic organoid system at week 4.

able to clear beta selection Figure 4.18 **B**. Finally, we have demonstrated that this technology produces similar differentiation results in distinct biological experiments as read out by flow cytometry Figure 4.18 **C**.

### 4.6.2 Interrogation of iPSC Derived T Cell Receptors

Despite clear flow cytometric data showing expression of a TCR on the surface of edited *RAG2* deficient iPSC derived T cells, this flow panel does not interrogate the

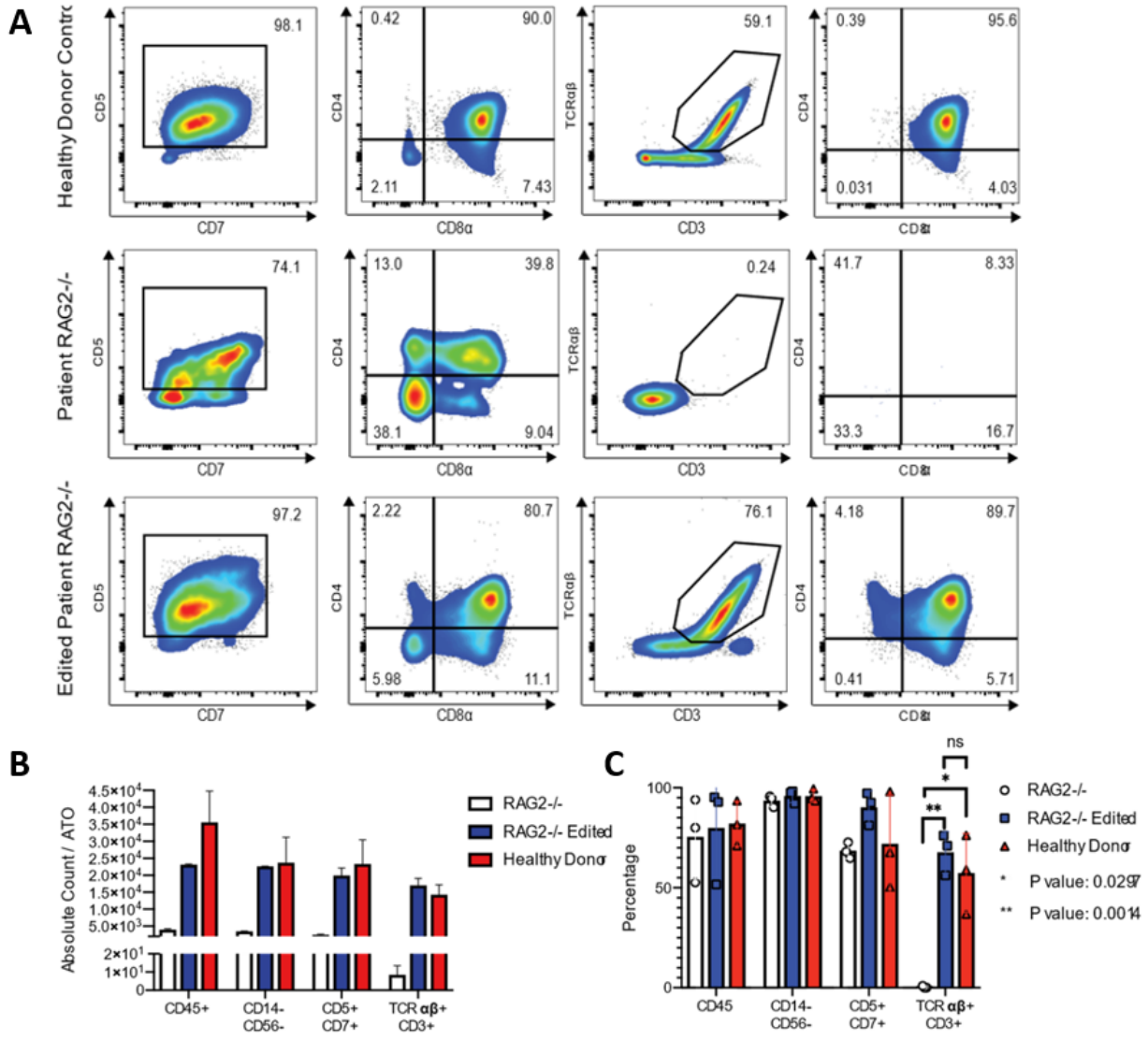



Figure 4.18: **Figure A** Comparative flow cytometry data showing markers of both early and mature T cell differentiation between the three examined cell lines. **Figure B** Quantification of the number of cells per ATO examined by hemocytometer. **Figure C** Flow cytometry gating results of derived cell populations over several distinct ATO experiments. Statistical significance was computed using multiple paired T tests with a Holm-Sidak correction.

identity of TCRs being displayed, or resolve any information on V(D)J pairings in establishing a TCR repertoire. To answer this question and explore whether or not a single or oligoclonal TCR repertoire was formed, we looked at the TCR V(D)J loci and RAG resultant rearrangements by Next Generation Sequencing (NGS).

We hoped to interrogate single positive and double positive T cells individually, to investigate if any sub-populations overlapped with TCR sequences commonly found in single positive populations in primary T cells. Despite excellent separation of these single and double positive T cells observed during cell sorting Figure 4.19, the quality of the genomic DNA from sorted populations was too low to produce reliable results through NGS. As a result, we gated on and analyzed the entire CD3+ mature T cell population in future experiments.

As reviewed more thoroughly in section 1.3.1, the nature of the TCR alpha chain rearrangement is not entirely stochastic. Within the chromosome housing the V and J genes of the alpha and delta TCR segments, cells containing a hypomorphic RAG complex or cells with otherwise impaired T cell development preferentially recombine *TRAD* and *TRAJ* genes that are the closest to each other in physical space Figure 4.20 **A**. These proximal pairings are detectable through sequencing of the TCR loci, looking for enrichment of downstream *TRAV* that pair with more upstream *TRAJ* genes. To demonstrate this enrichment of proximal V-J pairing in developmentally compromised cells, we attempted to derive T cells from iPSCs cultivated on a monolayer of murine stromal cells (OP9) expressing Delta-like 1 (DLL1) (OP9-DLL1) rather than in the 3D thymic organoid system which exposes developing cells to DLL4. This monolayer technology does not recapitulate the 3D environment of

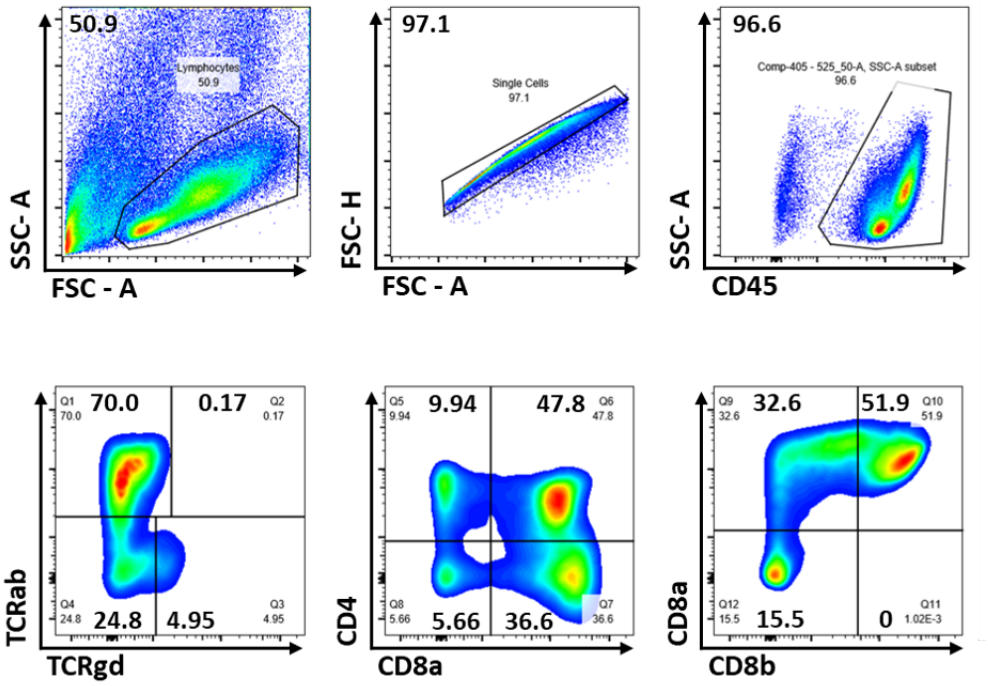



Figure 4.19: Flow cytometry data showing robust differentiation of iPSC derived T cells expressing a CD4 and CD8 single positive population along with a double positive population expressing primarily TCR$\alpha\beta$ and both a CD8$\alpha$ and CD8$\beta$ chain.

the thymus, and as such differentiating cells face impaired development. This impairment showed the attenuation of full V-J pairing when analyzed by NGS, with the predicted enrichment of proximal gene usage predominating Figure 4.20 **B**. We hoped to demonstrate a reversal of this preferential proximal pairing in both our healthy donor and r.RAG2 edited CD3+ T cells differentiating in the 3D thymic organoid environment. Consistent with our expectation, NGS data showed a very robust and quite similar V-J pairing profile for the gDNA harvested from both healthy donor and edited iPSC derived CD3+ T cells Figure 4.20 **C**, with no particular enrichment for any V-J pair which could indicate clonal expansion. Additionally, the

beta chain V-J pairs were similarly utilized, indicating that the r.RAG2 cassette was functionally allowing not only recombination of a TCR, but of a polyclonal repertoire of TCRs capable of producing T cells with a multiplicity of CDR3 binding domains.

We conducted a comparison of the V-J rearrangement profiles of the healthy donor iPSC derived T cells, the *RAG2* deficient edited iPSC derived T cells, and the healthy donor iPSC derived T cells cultivated on the monolayer system. We found similar V-J pairings in the *RAG2* deficient edited patient T cells as compared to the immunocompetent control T cells Figure 4.21 **ii.**, **iv.** indicating similarly rigorous recombination in the both lines. Unsurprisingly, the T cells generated from the OP9-DLL1 monolayer system did not contain many rearrangements not found in the healthy donor line or in the *RAG2* deficient edited line Figure 4.21 **vii.**, **viii.**, once again suggesting the 3D environment helps spur more physiologically relevant mechanisms of T cell development.

Beyond V-J pairing, we wanted to investigate whether the generated TCRs contained a multitude of unique Complementary Determining Region 3 (CDR3)s. Complementary determining regions are hypervariable segments of the alpha and beta chain of the TCR that confer specificity when scanning MHC molecules presenting antigen peptides Figure 4.22 **A** (Cole et al. 2014). There are three complementary determining regions. The first two complementary determining regions contain sequences derived primarily from V genes, while the third complementary determining region (CDR3) contains sequences from V, D (in the case of the beta chain) and J segments, making this region the most diverse within the TCR. Through High

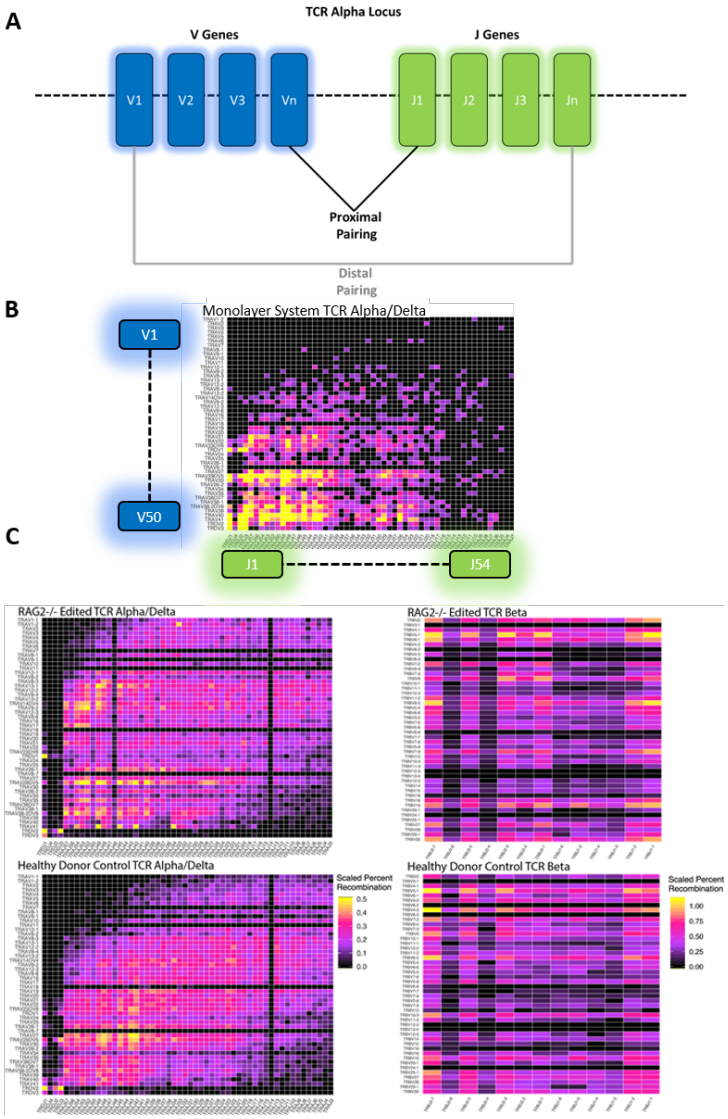



Figure 4.20: **Figure A** Illustration demonstrating how proximal and distal V-J pairings are recombined. **Figure B** NGS data showing V-J pairing in a monolayer T cell differentiation system using OP9 as a stromal cell line showing impaired rearrangements. **Figure C** Heatmaps showing the frequency of TCRs from CD3+ sorted cells with specific V and J gene pairings in both the $\alpha/\delta$ chain of the TCR as well as with the $\beta$ chain. V and J genes are laid out in their positioning in physical space along the chromosome.

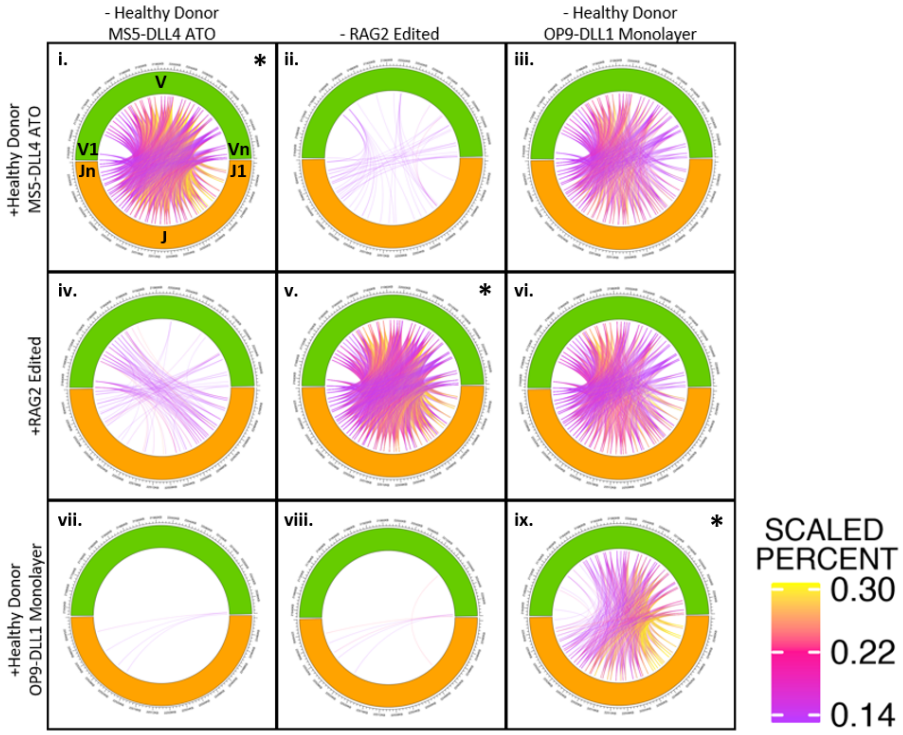



Figure 4.21: Circos plot comparison showing the pairing of V genes and J genes. The V and J genes have been arranged in order of their physical positioning in the genome from V1 to Vn as indicated in **i.**. Cells containing an '*' represent the full data set without any subtraction (cells **i., v., ix.**). The remaining cells show subtraction of one V-J pairing dataset with another as indicated by the labels along the X and Y axes.

Throughput Sequencing (HTS), we examined the CDR3 identity of CD3+ cells generated in the ATO system by both healthy donor and *RAG2* deficient edited cells. Perhaps unsurprisingly given the V-J pairing data, we saw a diverse repertoire of TCRs expressing unique CDR3s without a dominant clonotype Figure 4.22 **B**. This indicated to us that the ATO system was providing the requisite environment to allow for RAG recombination and activation of the r.RAG2 cassette in the *RAG2* deficient edited line. This diversity can also be translated to a statistical measure

using the Shannon's H Diversity Index (Shannon 1997), (Schneider-Hohendorf et al. 2016), (Cui et al. 2018). This index marks diversity as a function of entropy. Low entropy samples contain more repeats of identical sequences (clonal expansion) in the CDR3 pool, and as such have a lower index score. Conversely, higher entropy samples contain more unique CDR3 reads and are assigned a higher index score. We were happy to see a somewhat similar score without the predominance of any clone in cells edited with the r.RAG2 cassette, though we also recognize that the diversity is not as abundant in edited cells as in cells from a healthy immunocompetent donor Figure 4.22 **C**. Finally, looking at CDR3 length, we saw a fairly consistent distribution between the healthy donor and edited cells between the alpha/delta and beta TCR loci Figure 4.22 **D**. All TCRs generated from iPSCs in this system contain CDR3s that are shorter than those in primary T cells due to the silencing of the polymerase TdT during V(D)J recombination (Montel-Hagen et al. 2019). TdT has been shown to be absent in the fetal thymus during physiological development, and so the TCR identity of developing T cells out of the ATO system are more representative of fetal lymphocytes than adult mature T cells (Rother et al. 2016).

## 4.7 Discussion

The use of iPSC modeling has illuminated some of the occluded biological processes involved in disease progression. The increasing availability of new protocols to derive more and more complex cellular phenotypes allows us to ask questions of not only of cellular ontology, but of pathology and potential therapeutics. In the case of differentiating T cells, the stem cell biology field has now attempted to recapit-

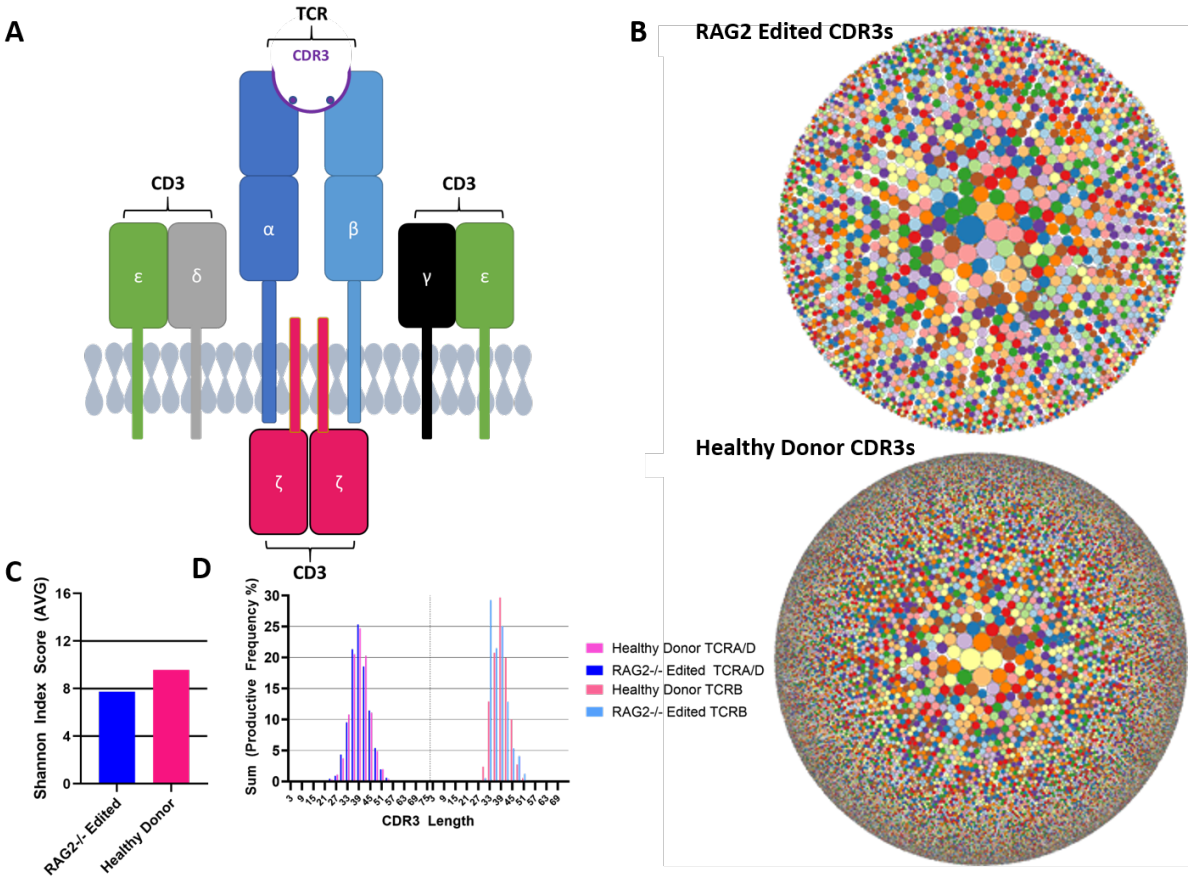



Figure 4.22: **A** Illustration of the alpha/beta TCR and co-activator CD3 domains. The segment highlighted in purple is meant to contain the complementary determining regions containing the CDR3. **B** Treemaps showing diversity of CDR3s in CD3+ iPSC derived T cells. Each dot represents a unique CDR3 sequence, with the size of each dot representing the frequency within the population. **C** Representation of the CDR3 data as a statistical measure of diversity given by the Shannon H Index Score. A score of 8 is often used to determine polyclonality in patient primary T cells. **D** Virtual spectratyping data showing the CDR3 lengths in both the TCR alpha/delta loci and the beta locus between the healthy donor cells and the *RAG2* deficient edited cells.

ulate more of the cellular environment that developing thymocytes are exposed to physiologically, leading away from more rudimentary monolayer modeling to the 3D

thymic organoid system used in this work. This 3D environment has shown a remarkable ability to reproduce many of the developmental properties of the thymic structure which precede T cell commitment and differentiation, which has allowed us to study our targeted therapy in patient cells.

Another fundamental property of iPSCs is their ability to proliferate indefinitely while avoiding artefacts often introduced from immortalizing somatic cell types. Indefinite renewal makes iPSCs a workhorse for gene editing. In HSPCs or other primary cell types, typically only a single round of gene editing is possible before cells become senescent and die. This short window of experimentation can limit both the amount of genome engineering possible as well as the amount of data able to be collected. iPSCs then represent an excellent chassis to introduce as many genetic alterations as desired, as well as to more thoroughly characterize a genotype phenotype relationship in a given cell line, which we were able to demonstrate here be isolating and characterizing a single edited clone.

One of the most powerful use of stem cells to study disease is the capability to generate iPSCs from patients with an established phenotype. Data generated with patient cells is more compelling than using a cell line or knocking in a known pathogenic mutation into a healthy donor cell because patient presentation gives robust physiological data on what differentiated cells should be able to recapitulate. Patient cells more robustly show therapeutic efficacy when compared to an engineered cell line as the therapeutic agent acts on the patient's known established diseased cell directly, accounting for epistatic effects and validating that the targeted therapy corrects the phenotype in that cell. If we were to simply knock out *RAG2*

from the healthy donor iPSC line, it would be unclear if rescue with the r.RAG2 cassette would actually produce functional T cells in a SCID patient, given potential confounding forces which may be present in patient cells arising from other epistatic variations.

We were able to take advantage of these iPSC strengths to demonstrate that a targeted gene therapy introduced to patient *RAG2* deficient cells can indeed rescue the developmental block in T cell differentiation. Data generated from iPSCs demonstrated rescue more robustly than our prior attempt in HSPCs, though we remain optimistic that the HSPC work will also show promising results in the near future. We hope that the results of our characterization of iPSC derived T cells and their TCRs will be consistent with those of HSPC derived T cells. We were able to show not only the robust generation of mature CD3+ TCR$\alpha\beta$+ T cells from *RAG2* deficient edited cells, but also that the RAG complex derived in part from our r.RAG2 cassette was able to recombine a multitude of V-J genes during V(D)J recombination both from proximal and distal genes, and that resulting T cells had a diverse array of cognate peptide binding targets when looking at the identities of CDR3s generated during the differentiation process. Overall, these data suggest that an endogenous *RAG2* gene targeting approach could restore differentiation and function of T lymphocytes in *RAG2* deficient patients, representing a new potential treatment modality for RAG deficiency.

# Chapter 5

# Limitations

The recent advances made in gene and cellular therapy, along with stem cell modeling, have opened up new doors to efficacious therapeutic treatment options. Despite this progress, many challenges remain to unlock the full potential of engineered cell and gene therapies. In the general sense, the most pressing concerns in gene therapy continue to be the targeted delivery of the gene editing machinery and components, immunogenicity from microbial derived nucleases, off-target editing effects from DNA damage, and efficiency of on-target editing. In stem cell biology, limitations continue to be a lack of robust protocols to derive a multitude of differentiated cell types, the 3D cell culture/organoid field still being in its infancy, complications from iPSC clones displaying different phenotypes from cell to cell, and full recapitulation of cellular mechanisms observed in primary cells within stem cell derived cells. Luckily, all of these challenges are being met and iteratively improved upon by labs both in industry and academia.

In addition to the challenges outlined above, this work specifically was limited by several other factors. The primary limiting consideration was the inconsistency of

the ATO system and the low number of cells produced in differentiation experiments which were successful. Despite considerable optimization and improvements in differentiation outcomes, the ATO system tended to either work well, or not at all, with all cells dying. This coupled to the high cost of using this platform, and the fact that these experiments take 6-8 weeks to get a readout led to a lot of time spent on planning downstream experiments which were stymied by cell death. Additional factors to consider for standardizing the differentiation efficiency of this protocol include: the lot number of cytokines and growth factors used, the expression of DLL4 on murine stromal cells, the culture time and confluency of iPSCs being differentiated into hEMP cells, the lot number of transwell membrane inserts - the composition of which was altered by the manufacturer in 2019 with noted mechanical differences, the concentration of cytokines and growth factors, the duration of differentiation of hEMP cells and developing lymphocytes, the number of ATOs seeded on each membrane, and the frequency of media changes. Other limiting factors included difficulty in isolating RNA from pre-T cells of sufficient quality to assay for expression of the recoded *RAG2* cassette, lack of further analysis of structural variants which could have resulted from the gene editing pipeline, and the lack of a reliable antibody against human RAG2 to confirm protein presence in developing cells.

Finally, challenges from the shutdown due to the global pandemic shortened the final experiments intended to be carried out in this thesis work. This work included the repetition of the editing and differentiation of patient derived HSPCs, and further characterization of those cells by PMA/Ionomycin stimulation and proliferation assays if successful. We also intended to look at the proliferation of iPSC derived T

cells by stimulation of the cells with soluble CD3 and CD28 antibodies and staining with CellTrace Violet as performed in experiments conducted with HSPCs. Ongoing work in the lab of Matthew Porteus includes all engraftment studies for gene targeted *RAG2* deficient patient cells into NSG mice. Engraftment studies will suggest whether or not this editing approach could be clinically viable in humans, and will provide data on reconstitution of immune cell subsets derived from cells containing the r.RAG2 cassette. Engraftment study data should be published in 2021.

Despite these limitations, we are excited to show that our *in vitro* studies gene correcting *RAG2* showed robust rescue of T cell developmental blocks and robust rearrangement of the V and J genes of the TCR, along with a diverse repertoire of CDR3s within patient iPSC derived mature naive T cells. While these experiments were curtailed by the pandemic, the two years of research spent on this project produced promising results and we hope that future work will more robustly characterize these *RAG2* deficient patient edited cells.

# Chapter 6

# Final Conclusions

The potential translational impacts of gene and cell therapies are immense. Progress in programmable nuclease implementation and disease modeling efforts have allowed not only molecular understanding of disease progression, but de-risking of targeted therapies. Unfortunately not all targets are commercially viable for development by large pharmaceutical companies. This is the case for most targets, including *RAG2*. We hope that elucidating the *in vitro* results from this pre-clinical work will help stimulate development in cellular treatments for Primary Immune Deficiency (PID)s.

To that end, we have built the foundation of a pipeline for potential employment in the clinic. We have designed clinically relevant cassettes delivered by AAV to be deployed in patient derived HSPCs. We have optimized their delivery and characterized their integration rates. We have reduced the cytotoxicity involved in the thawing, expansion, and editing process, and have demonstrated the potential rescue in T cell development, though further experiments will be needed to robustly demonstrate rescue. We have shown that the promoterless cargo we designed is expressed and is functional, reversing the block in T cell development observed in

isogenic precursors, and that the cargo provides robust V-J pairings in the alpha, delta, and beta loci leading to a diverse CDR3 repertoire in PSC derived T cells.

While further characterization efforts can strengthen the case for an *ex vivo* editing approach to correct *RAG2* deficiency, the data herein suggests this is a viable therapeutic approach that should be carefully considered for further development. We hope that all orphan diseases receive the attention they deserve and development of an appropriate and efficacious therapeutic. In summary, we hope that this thesis work moves the needle on developing a much needed autologous cell therapy for patients with *RAG2* deficiency, and that this target will soon join the still too-short list of other genes with effective cell and gene therapies available for suffering patients.

# Chapter 7

# Appendices

Table 7.1: Oligonucleotides used for PCR and ddPCR analysis.

| Name | Sequence |
|---|---|
| ver RAG2 KI F v1 | CTCTCCCAGAGCAACAAAGAC |
| ver RAG2 KI R v1 | CAGGCAGTCAGCTACACTATTC |
| ver RAG2 KI F v2 | CAAAGCTACACATCCCAGAGAG |
| ver RAG2 KI R v2 | GTATGATGTAGCACACCCAAATTC |
| ver RAG2 KI F v3 | CCAGAGCAACAAAGACAGATTTC |
| ver RAG2 KI R v3 | GATATTATTGGCAAGTGAATGTCCTC |
| ver P0.3 RAG2F v1 | TAATTCAGCCAGGCTTCTCAC |
| ver P0.3 RAG2R v1 | CCATGTTGCTTCCAAACCATATC |
| ver P0.3 RAG2F v2 | CTGTGGCTCTTTACTGACCTAAC |
| ver P0.3 RAG2F v3 | AAACATGTCCCAGCTCCTTG |
| ver P0.3 RAG2R v3 | GGTCTCCATCTCACGAATTTCTATC |
| ver P1.2.5 RAG2F v1 | GGATTCCACTCCCTTTGAAGAC |
| ver P1.2.5 RAG2R v1 | GCCAACTGCCGAATTGTTATG |
| ver P1.2.5 RAG2F v2 | GAGCAGACAACATTCACAAACAG |
| ver P1.2.5 RAG2R v2 | GCATAATCCAAAGGGTTGCTTG |
| ver P1.2.5 RAG2F v3 | CACTCCCTTTGAAGACTCTGAAG |
| ver P1.2.5 RAG2R v3 | AAGAACACTAATGTGGGATGTAGTAG |
| P0.3RAG2 5' G1 | CACCGTGTTTGCAAGAACAACAAAA |
| P0.3RAG2 3'rc G1 | AAACTGTTTGCAAGAACAACAAAAC |
| P0.3Rag2 5' G2 | CACCGTCGCTGCACAGAGAAAGACT |
| R0.3RAG2 3'rc G2 | AAACAGTCTTTCTCTGTGCAGCGAC |

Continued on next page

**Table 7.1 – continued from previous page**

| Name | Sequence |
|---|---|
| P0.3Rag2 5' G3 | CACCGATTAATGGAATGACCATATC |
| P0.3RAG2 3'rc G3 | AAACGATATGGTCATTCCATTAATC |
| P0.3Rag2 5' G4 | CACCGATATGGTCATTCCATTAATG |
| P0.3RAG2 3'rc G4 | AAACCATTAATGGAATGACCATATC |
| P0.3Rag2 5' G5 | CACCGTCGGCTGTACACCACATTAA |
| P0.3RAG2 3'rc G5 | AAACTTAATGTGGTGTACAGCCGAC |
| P1.2.5Rag2 5' G6 | CACCGCCAGATCCATGCACTGAGCA |
| RAG2 3'rc G6 | AAACTGCTCAGTGCATGGATCTGGC |
| P1.2.5Rag2 5' G7 | CACCGTGTTGCTTCCTGCTGACAGA |
| P1.2.5RAG2 3'rc G7 | AAACTCTGTCAGCAGGAAGCAACAC |
| P4 Cpf1 RAG2 5' G1 | AGATAAGCTGATAGCCACCAACAA |
| P4 Cpf1 RAG2 3'rc G1 | TTGTTGGTGGCTATCAGCTTAAAA |
| P4 Cpf1 RAG2 5' G2 | AGATCAAGCTGATAGCCACCAACA |
| P4 Cpf1 RAG2 3'rc G2 | TGTTGGTGGCTATCAGCTTGAAAA |
| P4 Cpf1 RAG2 5' G3 | AGATATTTTCAAGCTGATAGCCAC |
| P4 Cpf1 RAG2 3'rc G3 | GTGGCTATCAGCTTGAAAATAAAA |
| P4 Cpf1 RAG2 5' G4 | AGATGATTTTCAAGCTGATAGCCA |
| P4 Cpf1 RAG2 3'rc G4 | TGGCTATCAGCTTGAAAATCAAAA |
| P4 Cpf1 RAG2 5' G5 | AGATTGATTTTCAAGCTGATAGCC |
| P4 Cpf1 RAG2 3'rc G5 | GGCTATCAGCTTGAAAATCAAAAA |
| P4 Cpf1 RAG2 5' G6 | AGATTTGATTTTCAAGCTGATAGC |
| P4 Cpf1 RAG2 3'rc G6 | GCTATCAGCTTGAAAATCAAAAAA |
| P4 Cpf1 RAG2 5' G7 | AGATTTATTGTTGGTGGCTATCAG |
| P4 Cpf1 RAG2 3'rc G7 | CTGATAGCCACCAACAATAAAAAA |
| Rag2sg1ddIntegratedHEX | TACCATCAGAGACCATGTCACTCCAAATG |
| Rag2sg1ddReferenceFAM | TGCTCTCATGGGGATGGGCACTGGGT |
| Rag2sg1 dd Int HEX F | GCACCTTCGGCTAGTCTTTA |
| Rag2sg1 dd Int HEX R | TCATGAGAGAGAAACCCGGT |
| Rag2sg1 dd Ref FAM F | GACCCCAGACATTAAGCACA |
| Rag2sg1 dd Ref FAM R | TGACAGATGGATGAGTGTGC |
| TIDE RAG2 F | ATGTGGTTCTTTCAGCTGACG |
| TIDE RAG2 R | CGAAAAGTAACCTTTTTGTTGT |
| RAG2 qPCR recoded F | CCGCTTAGGTATCCTGCAAC |
| RAG2 qPCR recoded R | TCACGTATATCTTGTCGCTAACTTC |
| RAG2 qPCR wt F | CTTGGTAGGAGATGTTCCTGAAG |

Continued on next page

**Table 7.1 – continued from previous page**

| Name | Sequence |
|---|---|
| RAG2 qPCR wt R | GTGGGTAGAAGGCATGTATGAG |
| Rag2KO NALM6 | AACATAGCCTTAATTCAGCC |
| RAG2 int 3' junc F v1 | GCGTGATCTTCACCGACAAG |
| RAG2 int 3' junc R v1 | CATGTATGAGCGTCCTCCAAAG |
| RAG2 int 3' junc F v2 | CTTCACCGACAAGATCATCCG |
| RAG2 int 3' junc R v2 | CCTCCAAAGAGAACACCCATAC |
| Rag2 exon2 qpcr F | GCTCCTAAGCATCCAGACAG |
| Rag2 exon3 WT qpcr R | AGGCTATGTTATTACTGACTGTTACC |
| Rag2 exon3 REC qpcr R | CCGTCGAAGTTCATGAGAGAG |
| Rag2 rec mRNA verseq F | GACACCGTAACGACATCTCTG |
| Rag2 rec mRNA verseq R | TGACACATGGAAAGACAGTCC |
| Rag2 wt mRNA verseq F | CATCTTTAACAGGCAATGTCTCC |
| Rag2 wt mRNA verseq R | CAGGAACATCTCCTACCAAGTC |
| Rag2 3'int ver F | ATCGTGGAGTACCAGCAC |
| Rag2 3'int ver R | GGGAAGATCAACCCTTATTCTG |
| Rag2 ex1 mRNAseq v2 F | TTCTGCTTGCCACAGTCATAG |
| Rag2 wt mRNAseq v2 R | GGAACATCTCCTACCAAGTCTTTC |
| Rag2 rec mRNAseq v2 R | TCCCACGAGGTCCTTCTC |
| Rag2 KO sgRNA | CCCCACTGGAGTTTTCCATC |
| RAG2 3' int ver F v2 | CAACAAGATGAAGAGCACCAAAG |
| RAG2 3' int ver R v2 | GGCTTCAGTTTGACATGGTTATG |
| Rag2sg1v2 dd int HEX F | GCACCTTCGGCTAGTCTTTA |
| Rag2sg1v2 dd int HEX R | ATCAGAGAAAAGCCTGGCTG |
| Rag2sg1v2 dd ref FAM F | GCACAGGAAGTTTAGCAGTG |
| Rag2sg1v2 dd ref FAM R | GGGAATTCAAGACGCTCAGA |
| Rag2sg1v2 dd int HEX | GAGCCTGCAGATGGTGACCGTGTCCA |
| Rag2sg1v2 dd ref FAM | CCCGAGGAACGTGACCATGGAGTGGC |
| RAG2 wt/int common F | CTTCGGCTAGTCTTTATTC |
| RAG2 wt/int common R | TCAAAATCCACCAGGAA |
| RAG2 wt/int common Fv2 | GAGAGGCTCCAGCAATA |
| OCT4 qPCR F | CCTCACTTCACTGCACTGTA |
| OCT4 qPCR R | CAGGTTTTCTTTCCCTAGCT |
| NANOG qPCR F | TGAACCTCAGCTACAAACAG |
| NANOG qPCR R | TGGTGGTAGGAAGAGTAAAG |
| SOX2 qPCR F | CCCAGCAGACTTCACATGT |

Continued on next page

**Table 7.1 – continued from previous page**

| Name | Sequence |
|---|---|
| SOX2 qPCR R | CCTCCCATTTCCCTCGTTTT |
| ABCG2 qPCR F | TACCTGTATAGTGTACTTCAT |
| ABCG2 qPCR R | GGTCATGAGAAGTGTTGCTA |
| DNMT3B qPCR F | ATAAGTCGAAGGTGCGTCGT |
| DNMT3B qPCR R | GGCAACATCTGAAGCCATTT |
| ACTB qPCR F | TGAAGTGTGACGTGGACATC |
| ACTB qPCR R | GGACGAGCAATGATCTTGAT |
| MYC qPCR F | TGCCTCAAATTGGACTTTGG |
| MYC qPCR R | GATTGAAATTCTGTGTAACTGC |

Table 7.2: CRISPR Target Sequences Used for sgRNA Design

| **Name** | **Sequence** |
|---|---|
| RAG2 sgRNA1 | TGCAGAGACATAGTTTCTGA |
| RAG2 sgRNA2 | AGAAACTATGTCTCTGCAGA |
| RAG2 sgRNA3 | AACATAGCCTTAATTCAGCC |
| RAG2 sgRNA4 | TAATTCAGCCAGGCTTCTCA |

Table 7.3: Antibodies used for flow cytometric analysis.

| **Antibody** | **Clone** | **Fluorochrome** | **Vendor** |
|---|---|---|---|
| **Antibodies** | **Clone** | **Fluorochrome** | **Vendor** |
| CD14 | 61D3 | ef450 | Fisher Scientific |
| CD5 | UCHT2 | PECy7 | Fisher Scientific |
| CD7 | eBio123-1D1 | AF700 | Fisher Scientific |
| CD56 | MEM-188 | FITC | Biolegend |
| CD34 | 563 | PE | BD Biosciences |
| CD1a | HI149 | APC | Biolegend |
| CD45 | HI30 | V500 | BD Biosciences |
| CD19 | HIB19 | 5 | Biolegend |
| CD3 | UCHT1 | BV421 | BD Biosciences |
| CD4 | OKT4 | AF700 | Fisher Scientific |
| CD8a | RPA-T8 | UV | Biolegend |
| CD8b | 2ST8.5H7 | PE | BD Biosciences |
| TCRab | IP26 | APC | eBioscience |
| TCRgd | B1 | FITC | Biolegend |
| LIVE/DEAD Fixable Cell Stain Kit | NA | UV | Fisher Scientific |
| CD326 (EpCAM) | CO17-1A | PE/Cy7 | Biolegend |

# Bibliography

Abramson, S., R. G. Miller, and R. A. Phillips (1977). "The identification in adult bone marrow of pluripotent and restricted stem cells of the myeloid and lymphoid systems". In: *J Exp Med* 145.6, pp. 1567–79. ISSN: 0022-1007 (Print) 0022-1007 (Linking). DOI: 10.1084/jem.145.6.1567. URL: https://www.ncbi.nlm.nih.gov/pubmed/140917.

Alarcon, B. et al. (2003). "Initiation of TCR signaling: regulation within CD3 dimers". In: *Immunol Rev* 191, pp. 38–46. ISSN: 0105-2896 (Print) 0105-2896 (Linking). DOI: 10.1034/j.1600-065x.2003.00017.x. URL: https://www.ncbi.nlm.nih.gov/pubmed/12614350.

Allenspach, E., D. J. Rawlings, and A. M. Scharenberg (1993). "X-Linked Severe Combined Immunodeficiency". In: *GeneReviews((R))*. Ed. by M. P. Adam et al. Seattle (WA). URL: https://www.ncbi.nlm.nih.gov/pubmed/20301584.

Amarasinghe, S. L. et al. (2020). "Opportunities and challenges in long-read sequencing data analysis". In: *Genome Biol* 21.1, p. 30. ISSN: 1474-760X (Elec-

tronic) 1474-7596 (Linking). DOI: 10.1186/s13059-020-1935-5. URL: https://www.ncbi.nlm.nih.gov/pubmed/32033565.

Aversa, F. et al. (2019). "The Evolution of T Cell Depleted Haploidentical Transplantation". In: *Front Immunol* 10, p. 2769. ISSN: 1664-3224 (Electronic) 1664-3224 (Linking). DOI: 10.3389/fimmu.2019.02769. URL: https://www.ncbi.nlm.nih.gov/pubmed/31827475.

Barrangou, R. and L. A. Marraffini (2014). "CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity". In: *Mol Cell* 54.2, pp. 234–44. ISSN: 1097-4164 (Electronic) 1097-2765 (Linking). DOI: 10.1016/j.molcel.2014.03.011. URL: https://www.ncbi.nlm.nih.gov/pubmed/24766887.

Blaese, R. M. et al. (1995). "T lymphocyte-directed gene therapy for ADA- SCID: initial trial results after 4 years". In: *Science* 270.5235, pp. 475–80. ISSN: 0036-8075 (Print) 0036-8075 (Linking). DOI: 10.1126/science.270.5235.475. URL: https://www.ncbi.nlm.nih.gov/pubmed/7570001.

Blazar, B. R., G. R. Hill, and W. J. Murphy (2020). "Dissecting the biology of allogeneic HSCT to enhance the GvT effect whilst minimizing GvHD". In: *Nat Rev Clin Oncol* 17.8, pp. 475–492. ISSN: 1759-4782 (Electronic) 1759-4774 (Linking). DOI: 10.1038/s41571-020-0356-4. URL: https://www.ncbi.nlm.nih.gov/pubmed/32313224.

Boch, J. et al. (2009). "Breaking the code of DNA binding specificity of TAL-type III effectors". In: *Science* 326.5959, pp. 1509–12. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). DOI: 10.1126/science.1178811. URL: https://www.ncbi.nlm.nih.gov/pubmed/19933107.

Boise, L. H. et al. (1995). "CD28 costimulation can promote T cell survival by enhancing the expression of Bcl-XL". In: *Immunity* 3.1, pp. 87–98. ISSN: 1074-7613 (Print) 1074-7613 (Linking). DOI: 10.1016/1074-7613(95)90161-2. URL: https://www.ncbi.nlm.nih.gov/pubmed/7621080.

Bolger, A. M., M. Lohse, and B. Usadel (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data". In: *Bioinformatics* 30.15, pp. 2114–20. ISSN: 1367-4811 (Electronic) 1367-4803 (Linking). DOI: 10.1093/bioinformatics/btu170. URL: https://www.ncbi.nlm.nih.gov/pubmed/24695404.

Bomken, S. and R. Skinner (2015). "Secondary Malignant Neoplasms Following Haematopoietic Stem Cell Transplantation in Childhood". In: *Children (Basel)* 2.2, pp. 146–73. ISSN: 2227-9067 (Print) 2227-9067 (Linking). DOI: 10.3390/children2020146. URL: https://www.ncbi.nlm.nih.gov/pubmed/27417356.

Borte, S. et al. (2012). "Neonatal screening for severe primary immunodeficiency diseases using high-throughput triplex real-time PCR". In: *Blood* 119.11, pp. 2552–5. ISSN: 1528-0020 (Electronic) 0006-4971 (Linking). DOI: 10.1182/blood-2011-08-371021. URL: https://www.ncbi.nlm.nih.gov/pubmed/22130802.

Bosticardo, M. et al. (2020). "Artificial thymic organoids represent a reliable tool to study T-cell differentiation in patients with severe T-cell lymphopenia". In: *Blood Adv* 4.12, pp. 2611–2616. ISSN: 2473-9537 (Electronic) 2473-9529 (Linking). DOI: 10.1182/bloodadvances.2020001730. URL: https://www.ncbi.nlm.nih.gov/pubmed/32556283.

Brennand, K. J. et al. (2011). "Modelling schizophrenia using human induced pluripotent stem cells". In: *Nature* 473.7346, pp. 221–5. ISSN: 1476-4687 (Electronic) 0028-0836 (Linking). DOI: 10.1038/nature09915. URL: https://www.ncbi.nlm.nih.gov/pubmed/21490598.

Brunetti, L. et al. (2018). "Highly Efficient Gene Disruption of Murine and Human Hematopoietic Progenitor Cells by CRISPR/Cas9". In: *J Vis Exp* 134. ISSN: 1940-087X (Electronic) 1940-087X (Linking). DOI: 10.3791/57278. URL: https://www.ncbi.nlm.nih.gov/pubmed/29708546.

Byrne, S. M., P. Mali, and G. M. Church (2014). "Genome editing in human stem cells". In: *Methods Enzymol* 546, pp. 119–38. ISSN: 1557-7988 (Electronic) 0076-6879 (Linking). DOI: 10.1016/B978-0-12-801185-0.00006-4. URL: https://www.ncbi.nlm.nih.gov/pubmed/25398338.

Cagdas, D. et al. (2018). "ADA Deficiency: Evaluation of the Clinical and Laboratory Features and the Outcome". In: *J Clin Immunol* 38.4, pp. 484–493. ISSN: 1573-2592 (Electronic) 0271-9142 (Linking). DOI: 10.1007/s10875-018-0496-9. URL: https://www.ncbi.nlm.nih.gov/pubmed/29744787.

Capo, V. et al. (2018). "Efficacy of lentivirus-mediated gene therapy in an Omenn syndrome recombination-activating gene 2 mouse model is not hindered by inflammation and immune dysregulation". In: *J Allergy Clin Immunol* 142.3, 928–941 e8. ISSN: 1097-6825 (Electronic) 0091-6749 (Linking). DOI: 10.1016/j.jaci.2017.11.015. URL: https://www.ncbi.nlm.nih.gov/pubmed/29241731.

Cavazzana-Calvo, M. et al. (2000). "Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease". In: *Science* 288.5466, pp. 669–72. ISSN: 0036-8075 (Print) 0036-8075 (Linking). DOI: 10.1126/science.288.5466.669. URL: https://www.ncbi.nlm.nih.gov/pubmed/10784449.

Chan, B. et al. (2005). "Long-term efficacy of enzyme replacement therapy for adenosine deaminase (ADA)-deficient severe combined immunodeficiency (SCID)". In: *Clin Immunol* 117.2, pp. 133–43. ISSN: 1521-6616 (Print) 1521-6616 (Linking). DOI: 10.1016/j.clim.2005.07.006. URL: https://www.ncbi.nlm.nih.gov/pubmed/16112907.

Charlesworth, C. T. et al. (2019). "Identification of preexisting adaptive immunity to Cas9 proteins in humans". In: *Nat Med* 25.2, pp. 249–254. ISSN: 1546-170X (Electronic) 1078-8956 (Linking). DOI: 10.1038/s41591-018-0326-x. URL: https://www.ncbi.nlm.nih.gov/pubmed/30692695.

Chatterjee, S. (2019). "Efficient Nuclease-free HR by Clade F AAV Requires High MOIs with High Quality Vectors". In: *Mol Ther* 27.12, pp. 2058–2061. ISSN: 1525-

0024 (Electronic) 1525-0016 (Linking). DOI: 10.1016/j.ymthe.2019.11.002. URL: https://www.ncbi.nlm.nih.gov/pubmed/31735603.

Chatterjee, S., V. Sivanandam, and Jr. Wong Kk (2020). "Adeno-Associated Virus and Hematopoietic Stem Cells: The Potential of Adeno-Associated Virus Hematopoietic Stem Cells in Genetic Medicines". In: *Hum Gene Ther* 31.9-10, pp. 542–552. ISSN: 1557-7422 (Electronic) 1043-0342 (Linking). DOI: 10.1089/hum.2020.049. URL: https://www.ncbi.nlm.nih.gov/pubmed/32253938.

Cho, S. W. et al. (2014). "Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases". In: *Genome Res* 24.1, pp. 132–41. ISSN: 1549-5469 (Electronic) 1088-9051 (Linking). DOI: 10.1101/gr.162339.113. URL: https://www.ncbi.nlm.nih.gov/pubmed/24253446.

Cibulskis, K. et al. (2013). "Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples". In: *Nat Biotechnol* 31.3, pp. 213–9. ISSN: 1546-1696 (Electronic) 1087-0156 (Linking). DOI: 10.1038/nbt.2514. URL: https://www.ncbi.nlm.nih.gov/pubmed/23396013.

Ciriza, J. et al. (2013). "The migration of hematopoietic progenitors from the fetal liver to the fetal bone marrow: lessons learned and possible clinical applications". In: *Exp Hematol* 41.5, pp. 411–23. ISSN: 1873-2399 (Electronic) 0301-472X (Linking). DOI: 10.1016/j.exphem.2013.01.009. URL: https://www.ncbi.nlm.nih.gov/pubmed/23395775.

Cole, D. K. et al. (2014). "T-cell receptor (TCR)-peptide specificity overrides affinity-enhancing TCR-major histocompatibility complex interactions". In: *J Biol Chem* 289.2, pp. 628–38. ISSN: 1083-351X (Electronic) 0021-9258 (Linking). DOI: 10.1074/jbc.M113.522110. URL: https://www.ncbi.nlm.nih.gov/pubmed/24196962.

Cui, Jin-Huan et al. (2018). "TCR Repertoire as a Novel Indicator for Immune Monitoring and Prognosis Assessment of Patients With Cervical Cancer". In: *Frontiers in Immunology* 9.2729. ISSN: 1664-3224. DOI: 10.3389/fimmu.2018.02729. URL: https://www.frontiersin.org/article/10.3389/fimmu.2018.02729.

Dai, X. et al. (2019). "One-step generation of modular CAR-T cells with AAV-Cpf1". In: *Nat Methods* 16.3, pp. 247–254. ISSN: 1548-7105 (Electronic) 1548-7091 (Linking). DOI: 10.1038/s41592-019-0329-7. URL: https://www.ncbi.nlm.nih.gov/pubmed/30804551.

De Ravin, S. S. et al. (2017). "CRISPR-Cas9 gene repair of hematopoietic stem cells from patients with X-linked chronic granulomatous disease". In: *Sci Transl Med* 9.372. ISSN: 1946-6242 (Electronic) 1946-6234 (Linking). DOI: 10.1126/scitranslmed.aah3480. URL: https://www.ncbi.nlm.nih.gov/pubmed/28077679.

Delmonte, O. M., C. Schuetz, and L. D. Notarangelo (2018). "RAG Deficiency: Two Genes, Many Diseases". In: *J Clin Immunol* 38.6, pp. 646–655. ISSN: 1573-2592

(Electronic) 0271-9142 (Linking). DOI: 10.1007/s10875-018-0537-4. URL: https://www.ncbi.nlm.nih.gov/pubmed/30046960.

Dendrou, C. A. et al. (2018). "HLA variation and disease". In: *Nat Rev Immunol* 18.5, pp. 325–339. ISSN: 1474-1741 (Electronic) 1474-1733 (Linking). DOI: 10.1038/nri.2017.143. URL: https://www.ncbi.nlm.nih.gov/pubmed/29292391.

Dervovic, D. D. et al. (2013). "Cellular and molecular requirements for the selection of in vitro-generated CD8 T cells reveal a role for Notch". In: *J Immunol* 191.4, pp. 1704–15. ISSN: 1550-6606 (Electronic) 0022-1767 (Linking). DOI: 10.4049/jimmunol.1300417. URL: https://www.ncbi.nlm.nih.gov/pubmed/23851691.

Dezfouli, M. et al. (2020). "Newborn Screening for Presymptomatic Diagnosis of Complement and Phagocyte Deficiencies". In: *Front Immunol* 11, p. 455. ISSN: 1664-3224 (Electronic) 1664-3224 (Linking). DOI: 10.3389/fimmu.2020.00455. URL: https://www.ncbi.nlm.nih.gov/pubmed/32256498.

Dorsey, M. J. et al. (2017). "Treatment of infants identified as having severe combined immunodeficiency by means of newborn screening". In: *J Allergy Clin Immunol* 139.3, pp. 733–742. ISSN: 1097-6825 (Electronic) 0091-6749 (Linking). DOI: 10.1016/j.jaci.2017.01.005. URL: https://www.ncbi.nlm.nih.gov/pubmed/28270365.

Dudek, A. M. and M. H. Porteus (2019). "AAV6 Is Superior to Clade F AAVs in Stimulating Homologous Recombination-Based Genome Editing in Human HSPCs". In: *Mol Ther* 27.10, pp. 1701–1705. ISSN: 1525-0024 (Electronic) 1525-0016 (Linking). DOI: 10.1016/j.ymthe.2019.09.005. URL: https://www.ncbi.nlm.nih.gov/pubmed/31537456.

Faust, G. G. and I. M. Hall (2014). "SAMBLASTER: fast duplicate marking and structural variant read extraction". In: *Bioinformatics* 30.17, pp. 2503–5. ISSN: 1367-4811 (Electronic) 1367-4803 (Linking). DOI: 10.1093/bioinformatics/btu314. URL: https://www.ncbi.nlm.nih.gov/pubmed/24812344.

Fischer, A. (2000). "Severe combined immunodeficiencies (SCID)". In: *Clin Exp Immunol* 122.2, pp. 143–9. ISSN: 0009-9104 (Print) 0009-9104 (Linking). DOI: 10.1046/j.1365-2249.2000.01359.x. URL: https://www.ncbi.nlm.nih.gov/pubmed/11091267.

Flinn, A. M. and A. R. Gennery (2018). "Adenosine deaminase deficiency: a review". In: *Orphanet J Rare Dis* 13.1, p. 65. ISSN: 1750-1172 (Electronic) 1750-1172 (Linking). DOI: 10.1186/s13023-018-0807-5. URL: https://www.ncbi.nlm.nih.gov/pubmed/29690908.

Frangoul, H. et al. (2020). "CRISPR-Cas9 Gene Editing for Sickle Cell Disease and beta-Thalassemia". In: *N Engl J Med*. ISSN: 1533-4406 (Electronic) 0028-4793 (Linking). DOI: 10.1056/NEJMoa2031054. URL: https://www.ncbi.nlm.nih.gov/pubmed/33283989.

Frauwirth, K. A. et al. (2002). "The CD28 signaling pathway regulates glucose metabolism". In: *Immunity* 16.6, pp. 769–77. ISSN: 1074-7613 (Print) 1074-7613 (Linking). DOI: 10.1016/s1074-7613(02)00323-0. URL: https://www.ncbi.nlm.nih.gov/pubmed/12121659.

Freije, C. A. et al. (2019). "Programmable Inhibition and Detection of RNA Viruses Using Cas13". In: *Mol Cell* 76.5, 826–837 e11. ISSN: 1097-4164 (Electronic) 1097-2765 (Linking). DOI: 10.1016/j.molcel.2019.09.013. URL: https://www.ncbi.nlm.nih.gov/pubmed/31607545.

French, A. et al. (2015). "Human induced pluripotent stem cell-derived B lymphocytes express sIgM and can be generated via a hemogenic endothelium intermediate". In: *Stem Cells Dev* 24.9, pp. 1082–95. ISSN: 1557-8534 (Electronic) 1547-3287 (Linking). DOI: 10.1089/scd.2014.0318. URL: https://www.ncbi.nlm.nih.gov/pubmed/25519920.

Gaj, T. et al. (2017). "Targeted gene knock-in by homology-directed genome editing using Cas9 ribonucleoprotein and AAV donor delivery". In: *Nucleic Acids Res* 45.11, e98. ISSN: 1362-4962 (Electronic) 0305-1048 (Linking). DOI: 10.1093/nar/gkx154. URL: https://www.ncbi.nlm.nih.gov/pubmed/28334779.

Germain, R. N. (2002). "T-cell development and the CD4-CD8 lineage decision". In: *Nat Rev Immunol* 2.5, pp. 309–22. ISSN: 1474-1733 (Print) 1474-1733 (Linking). DOI: 10.1038/nri798. URL: https://www.ncbi.nlm.nih.gov/pubmed/12033737.

Gratwohl, A. et al. (2017). "Alloreactivity: the Janus-face of hematopoietic stem cell transplantation". In: *Leukemia* 31.8, pp. 1752–1759. ISSN: 1476-5551 (Electronic) 0887-6924 (Linking). DOI: 10.1038/leu.2017.79. URL: https://www.ncbi.nlm.nih.gov/pubmed/28270691.

Gundry, M. C. et al. (2016). "Highly Efficient Genome Editing of Murine and Human Hematopoietic Progenitor Cells by CRISPR/Cas9". In: *Cell Rep* 17.5, pp. 1453–1461. ISSN: 2211-1247 (Electronic). DOI: 10.1016/j.celrep.2016.09.092. URL: https://www.ncbi.nlm.nih.gov/pubmed/27783956.

Gyurkocza, B. and B. M. Sandmaier (2014). "Conditioning regimens for hematopoietic cell transplantation: one size does not fit all". In: *Blood* 124.3, pp. 344–53. ISSN: 1528-0020 (Electronic) 0006-4971 (Linking). DOI: 10.1182/blood-2014-02-514778. URL: https://www.ncbi.nlm.nih.gov/pubmed/24914142.

Hacein-Bey-Abina, S. et al. (2003). "LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1". In: *Science* 302.5644, pp. 415–9. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). DOI: 10.1126/science.1088547. URL: https://www.ncbi.nlm.nih.gov/pubmed/14564000.

Haddad, E. et al. (1998). "Long-term immune reconstitution and outcome after HLA-nonidentical T-cell-depleted bone marrow transplantation for severe combined immunodeficiency: a European retrospective study of 116 patients". In: *Blood* 91.10, pp. 3646–53. ISSN: 0006-4971 (Print) 0006-4971 (Linking). URL: https://www.ncbi.nlm.nih.gov/pubmed/9573000.

Haddad, E. et al. (2018). "SCID genotype and 6-month posttransplant CD4 count predict survival and immune recovery". In: *Blood* 132.17, pp. 1737–1749. ISSN: 1528-0020 (Electronic) 0006-4971 (Linking). DOI: 10.1182/blood-2018-03-840702. URL: https://www.ncbi.nlm.nih.gov/pubmed/30154114.

Haskins, K., J. Kappler, and P. Marrack (1984). "The major histocompatibility complex-restricted antigen receptor on T cells". In: *Annu Rev Immunol* 2, pp. 51–66. ISSN: 0732-0582 (Print) 0732-0582 (Linking). DOI: 10.1146/annurev.iy.02.040184.000411. URL: https://www.ncbi.nlm.nih.gov/pubmed/6085752.

Heimall, J. and M. J. Cowan (2017). "Long term outcomes of severe combined immunodeficiency: therapy implications". In: *Expert Rev Clin Immunol* 13.11, pp. 1029–1040. ISSN: 1744-8409 (Electronic) 1744-666X (Linking). DOI: 10.1080/1744666X.2017.1381558. URL: https://www.ncbi.nlm.nih.gov/pubmed/28918671.

Herbst, F. et al. (2012). "Extensive methylation of promoter sequences silences lentiviral transgene expression during stem cell differentiation in vivo". In: *Mol Ther* 20.5, pp. 1014–21. ISSN: 1525-0024 (Electronic) 1525-0016 (Linking). DOI: 10.1038/mt.2012.46. URL: https://www.ncbi.nlm.nih.gov/pubmed/22434137.

Hsu, P. D. et al. (2013). "DNA targeting specificity of RNA-guided Cas9 nucleases". In: *Nat Biotechnol* 31.9, pp. 827–32. ISSN: 1546-1696 (Electronic) 1087-

0156 (Linking). DOI: 10.1038/nbt.2647. URL: https://www.ncbi.nlm.nih.gov/pubmed/23873081.

Hubbard, B. P. et al. (2015). "Continuous directed evolution of DNA-binding proteins to improve TALEN specificity". In: *Nat Methods* 12.10, pp. 939–42. ISSN: 1548-7105 (Electronic) 1548-7091 (Linking). DOI: 10.1038/nmeth.3515. URL: https://www.ncbi.nlm.nih.gov/pubmed/26258293.

Inui, M. et al. (2014). "Rapid generation of mouse models with defined point mutations by the CRISPR/Cas9 system". In: *Sci Rep* 4, p. 5396. ISSN: 2045-2322 (Electronic) 2045-2322 (Linking). DOI: 10.1038/srep05396. URL: https://www.ncbi.nlm.nih.gov/pubmed/24953798.

Jang, H. K. et al. (2020). "Current trends in gene recovery mediated by the CRISPR-Cas system". In: *Exp Mol Med* 52.7, pp. 1016–1027. ISSN: 2092-6413 (Electronic) 1226-3613 (Linking). DOI: 10.1038/s12276-020-0466-1. URL: https://www.ncbi.nlm.nih.gov/pubmed/32651459.

Jenkins, M. K. and J. G. Johnson (1993). "Molecules involved in T-cell costimulation". In: *Curr Opin Immunol* 5.3, pp. 361–7. ISSN: 0952-7915 (Print) 0952-7915 (Linking). DOI: 10.1016/0952-7915(93)90054-v. URL: https://www.ncbi.nlm.nih.gov/pubmed/7688514.

Jenkins, M. K. et al. (2010). "On the composition of the preimmune repertoire of T cells specific for Peptide-major histocompatibility complex ligands". In: *Annu*

*Rev Immunol* 28, pp. 275–94. ISSN: 1545-3278 (Electronic) 0732-0582 (Linking). DOI: 10.1146/annurev-immunol-030409-101253. URL: https://www.ncbi.nlm.nih.gov/pubmed/20307209.

Jiang, H. et al. (2005). "Ubiquitylation of RAG-2 by Skp2-SCF links destruction of the V(D)J recombinase to the cell cycle". In: *Mol Cell* 18.6, pp. 699–709. ISSN: 1097-2765 (Print) 1097-2765 (Linking). DOI: 10.1016/j.molcel.2005.05.011. URL: https://www.ncbi.nlm.nih.gov/pubmed/15949444.

Joachims, M. L. et al. (2006). "Human alpha beta and gamma delta thymocyte development: TCR gene rearrangements, intracellular TCR beta expression, and gamma delta developmental potential–differences between men and mice". In: *J Immunol* 176.3, pp. 1543–52. ISSN: 0022-1767 (Print) 0022-1767 (Linking). DOI: 10.4049/jimmunol.176.3.1543. URL: https://www.ncbi.nlm.nih.gov/pubmed/16424183.

Jouvin-Marche, E., P. Fuschiotti, and P. N. Marche (2009). "Dynamic aspects of TCRalpha gene recombination: qualitative and quantitative assessments of the TCRalpha chain repertoire in man and mouse". In: *Adv Exp Med Biol* 650, pp. 82–92. ISSN: 0065-2598 (Print) 0065-2598 (Linking). DOI: 10.1007/978-1-4419-0296-2_7. URL: https://www.ncbi.nlm.nih.gov/pubmed/19731803.

Kanegane, H., K. Imai, and T. Morio (2017). "Severe combined immunodeficiency: From its discovery to the perspective". In: *Nihon Rinsho Meneki Gakkai Kaishi*

40.3, pp. 145–154. ISSN: 1349-7413 (Electronic) 0911-4300 (Linking). DOI: 10.2177/jsci.40.145. URL: https://www.ncbi.nlm.nih.gov/pubmed/28747600.

Kaur, G. and J. M. Dufour (2012). "Cell lines: Valuable tools or useless artifacts". In: *Spermatogenesis* 2.1, pp. 1–5. ISSN: 2156-5554 (Print) 2156-5554 (Linking). DOI: 10.4161/spmg.19885. URL: https://www.ncbi.nlm.nih.gov/pubmed/22553484.

Kim, C. et al. (2013). "Studying arrhythmogenic right ventricular dysplasia with patient-specific iPSCs". In: *Nature* 494.7435, pp. 105–10. ISSN: 1476-4687 (Electronic) 0028-0836 (Linking). DOI: 10.1038/nature11799. URL: https://www.ncbi.nlm.nih.gov/pubmed/23354045.

Kim, D. et al. (2015). "Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells". In: *Nat Methods* 12.3, 237–43, 1 p following 243. ISSN: 1548-7105 (Electronic) 1548-7091 (Linking). DOI: 10.1038/nmeth.3284. URL: https://www.ncbi.nlm.nih.gov/pubmed/25664545.

Kim, E. et al. (2017). "In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni". In: *Nat Commun* 8, p. 14500. ISSN: 2041-1723 (Electronic) 2041-1723 (Linking). DOI: 10.1038/ncomms14500. URL: https://www.ncbi.nlm.nih.gov/pubmed/28220790.

Kim, M. S. et al. (2015). "Crystal structure of the V(D)J recombinase RAG1-RAG2". In: *Nature* 518.7540, pp. 507–11. ISSN: 1476-4687 (Electronic) 0028-0836

(Linking). DOI: 10.1038/nature14174. URL: https://www.ncbi.nlm.nih.gov/pubmed/25707801.

Kim, M. S. et al. (2018). "Cracking the DNA Code for V(D)J Recombination". In: *Mol Cell* 70.2, 358–370 e4. ISSN: 1097-4164 (Electronic) 1097-2765 (Linking). DOI: 10.1016/j.molcel.2018.03.008. URL: https://www.ncbi.nlm.nih.gov/pubmed/29628308.

King, J. R. and L. Hammarstrom (2018). "Newborn Screening for Primary Immunodeficiency Diseases: History, Current and Future Practice". In: *J Clin Immunol* 38.1, pp. 56–66. ISSN: 1573-2592 (Electronic) 0271-9142 (Linking). DOI: 10.1007/s10875-017-0455-x. URL: https://www.ncbi.nlm.nih.gov/pubmed/29116556.

Kumar, B. V., T. J. Connors, and D. L. Farber (2018). "Human T Cell Development, Localization, and Function throughout Life". In: *Immunity* 48.2, pp. 202–213. ISSN: 1097-4180 (Electronic) 1074-7613 (Linking). DOI: 10.1016/j.immuni.2018.01.007. URL: https://www.ncbi.nlm.nih.gov/pubmed/29466753.

Lee, J. and S. Desiderio (1999). "Cyclin A/CDK2 regulates V(D)J recombination by coordinating RAG-2 accumulation and DNA repair". In: *Immunity* 11.6, pp. 771–81. ISSN: 1074-7613 (Print) 1074-7613 (Linking). DOI: 10.1016/s1074-7613(00)80151-x. URL: https://www.ncbi.nlm.nih.gov/pubmed/10626899.

Li, K. et al. (2014). "Optimization of genome engineering approaches with the CRISPR/Cas9 system". In: *PLoS One* 9.8, e105779. ISSN: 1932-6203 (Electronic) 1932-6203 (Linking). DOI: 10.1371/journal.pone.0105779. URL: https://www.ncbi.nlm.nih.gov/pubmed/25166277.

Liu, Y. et al. (2007). "A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement". In: *Immunity* 27.4, pp. 561–71. ISSN: 1074-7613 (Print) 1074-7613 (Linking). DOI: 10.1016/j.immuni.2007.09.005. URL: https://www.ncbi.nlm.nih.gov/pubmed/17936034.

Maeder, M. L. et al. (2013). "CRISPR RNA-guided activation of endogenous human genes". In: *Nat Methods* 10.10, pp. 977–9. ISSN: 1548-7105 (Electronic) 1548-7091 (Linking). DOI: 10.1038/nmeth.2598. URL: https://www.ncbi.nlm.nih.gov/pubmed/23892898.

Maeder, M. L. et al. (2019). "Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10". In: *Nat Med* 25.2, pp. 229–233. ISSN: 1546-170X (Electronic) 1078-8956 (Linking). DOI: 10.1038/s41591-018-0327-9. URL: https://www.ncbi.nlm.nih.gov/pubmed/30664785.

Malär, Reta et al. (2011). "Therapeutic drug monitoring is essential for intravenous busulfan therapy in pediatric hematopoietic stem cell recipients". In: *Pediatric Transplantation* 15.6, pp. 580–588. ISSN: 1397-3142.

Mali, P. et al. (2013). "RNA-guided human genome engineering via Cas9". In: *Science* 339.6121, pp. 823–6. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). DOI: 10.1126/science.1232033. URL: https://www.ncbi.nlm.nih.gov/pubmed/23287722.

Mansilla-Soto, J. and P. Cortes (2003). "VDJ recombination: Artemis and its in vivo role in hairpin opening". In: *J Exp Med* 197.5, pp. 543–7. ISSN: 0022-1007 (Print) 0022-1007 (Linking). DOI: 10.1084/jem.20022210. URL: https://www.ncbi.nlm.nih.gov/pubmed/12615895.

Mantere, T., S. Kersten, and A. Hoischen (2019). "Long-Read Sequencing Emerging in Medical Genetics". In: *Front Genet* 10, p. 426. ISSN: 1664-8021 (Print) 1664-8021 (Linking). DOI: 10.3389/fgene.2019.00426. URL: https://www.ncbi.nlm.nih.gov/pubmed/31134132.

Mao, Z. et al. (2008). "DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells". In: *Cell Cycle* 7.18, pp. 2902–6. ISSN: 1551-4005 (Electronic) 1551-4005 (Linking). DOI: 10.4161/cc.7.18.6679. URL: https://www.ncbi.nlm.nih.gov/pubmed/18769152.

McCusker, C., J. Upton, and R. Warrington (2018). "Primary immunodeficiency". In: *Allergy Asthma Clin Immunol* 14.Suppl 2, p. 61. ISSN: 1710-1484 (Print) 1710-1484 (Linking). DOI: 10.1186/s13223-018-0290-5. URL: https://www.ncbi.nlm.nih.gov/pubmed/30275850.

Meuer, S. C. et al. (1984). "The human T-cell receptor". In: *Annu Rev Immunol* 2, pp. 23–50. ISSN: 0732-0582 (Print) 0732-0582 (Linking). DOI: 10.1146/annurev.iy.02.040184.000323. URL: https://www.ncbi.nlm.nih.gov/pubmed/6443341.

Miller, J. C. et al. (2019). "Enhancing gene editing specificity by attenuating DNA cleavage kinetics". In: *Nat Biotechnol* 37.8, pp. 945–952. ISSN: 1546-1696 (Electronic) 1087-0156 (Linking). DOI: 10.1038/s41587-019-0186-z. URL: https://www.ncbi.nlm.nih.gov/pubmed/31359006.

Montel-Hagen, A. et al. (2019). "Organoid-Induced Differentiation of Conventional T Cells from Human Pluripotent Stem Cells". In: *Cell Stem Cell* 24.3, 376–389 e8. ISSN: 1875-9777 (Electronic) 1875-9777 (Linking). DOI: 10.1016/j.stem.2018.12.011. URL: https://www.ncbi.nlm.nih.gov/pubmed/30661959.

Moscou, M. J. and A. J. Bogdanove (2009). "A simple cipher governs DNA recognition by TAL effectors". In: *Science* 326.5959, p. 1501. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). DOI: 10.1126/science.1178817. URL: https://www.ncbi.nlm.nih.gov/pubmed/19933106.

Niehues, T., R. Perez-Becker, and C. Schuetz (2010). "More than just SCID–the phenotypic range of combined immunodeficiencies associated with mutations in the recombinase activating genes (RAG) 1 and 2". In: *Clin Immunol* 135.2, pp. 183–92. ISSN: 1521-7035 (Electronic) 1521-6616 (Linking). DOI: 10.1016/j.clim.2010.01.013. URL: https://www.ncbi.nlm.nih.gov/pubmed/20172764.

Nishana, M. and S. C. Raghavan (2012). "Role of recombination activating genes in the generation of antigen receptor diversity and beyond". In: *Immunology* 137.4, pp. 271–81. ISSN: 1365-2567 (Electronic) 0019-2805 (Linking). DOI: 10.1111/imm.12009. URL: https://www.ncbi.nlm.nih.gov/pubmed/23039142.

Ochs, H. D. and W. H. Hitzig (2012). "History of primary immunodeficiency diseases". In: *Curr Opin Allergy Clin Immunol* 12.6, pp. 577–87. ISSN: 1473-6322 (Electronic) 1473-6322 (Linking). DOI: 10.1097/ACI.0b013e32835923a6. URL: https://www.ncbi.nlm.nih.gov/pubmed/23095909.

Orthwein, A. et al. (2015). "A mechanism for the suppression of homologous recombination in G1 cells". In: *Nature* 528.7582, pp. 422–6. ISSN: 1476-4687 (Electronic) 0028-0836 (Linking). DOI: 10.1038/nature16142. URL: https://www.ncbi.nlm.nih.gov/pubmed/26649820.

Pai, S. Y. et al. (2014). "Transplantation outcomes for severe combined immunodeficiency, 2000-2009". In: *N Engl J Med* 371.5, pp. 434–46. ISSN: 1533-4406 (Electronic) 0028-4793 (Linking). DOI: 10.1056/NEJMoa1401177. URL: https://www.ncbi.nlm.nih.gov/pubmed/25075835.

Paquet, D. et al. (2016). "Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9". In: *Nature* 533.7601, pp. 125–9. ISSN: 1476-4687 (Electronic) 0028-0836 (Linking). DOI: 10.1038/nature17664. URL: https://www.ncbi.nlm.nih.gov/pubmed/27120160.

Paschon, D. E. et al. (2019). "Diversifying the structure of zinc finger nucleases for high-precision genome editing". In: *Nat Commun* 10.1, p. 1133. ISSN: 2041-1723 (Electronic) 2041-1723 (Linking). DOI: 10.1038/s41467-019-08867-x. URL: https://www.ncbi.nlm.nih.gov/pubmed/30850604.

Pfaff, N. et al. (2013). "A ubiquitous chromatin opening element prevents transgene silencing in pluripotent stem cells and their differentiated progeny". In: *Stem Cells* 31.3, pp. 488–99. ISSN: 1549-4918 (Electronic) 1066-5099 (Linking). DOI: 10.1002/stem.1316. URL: https://www.ncbi.nlm.nih.gov/pubmed/23307570.

Pike-Overzet, K. et al. (2011). "Correction of murine Rag1 deficiency by self-inactivating lentiviral vector-mediated gene transfer". In: *Leukemia* 25.9, pp. 1471–83. ISSN: 1476-5551 (Electronic) 0887-6924 (Linking). DOI: 10.1038/leu.2011.106. URL: https://www.ncbi.nlm.nih.gov/pubmed/21617701.

Plotkin, J. B. and G. Kudla (2011). "Synonymous but not the same: the causes and consequences of codon bias". In: *Nat Rev Genet* 12.1, pp. 32–42. ISSN: 1471-0064 (Electronic) 1471-0056 (Linking). DOI: 10.1038/nrg2899. URL: https://www.ncbi.nlm.nih.gov/pubmed/21102527.

Ponader, S. and J. A. Burger (2014). "Bruton's tyrosine kinase: from X-linked agammaglobulinemia toward targeted therapy for B-cell malignancies". In: *J Clin Oncol* 32.17, pp. 1830–9. ISSN: 1527-7755 (Electronic) 0732-183X (Linking). DOI:

10.1200/JCO.2013.53.1046. URL: https://www.ncbi.nlm.nih.gov/pubmed/24778403.

Puck, J. M. (2011). "Neonatal screening for severe combined immunodeficiency". In: *Curr Opin Pediatr* 23.6, pp. 667–73. ISSN: 1531-698X (Electronic) 1040-8703 (Linking). DOI: 10.1097/MOP.0b013e32834cb9b0. URL: https://www.ncbi.nlm.nih.gov/pubmed/22001765.

Puck, J. M. et al. (1997). "Mutation analysis of IL2RG in human X-linked severe combined immunodeficiency". In: *Blood* 89.6, pp. 1968–77. ISSN: 0006-4971 (Print) 0006-4971 (Linking). URL: https://www.ncbi.nlm.nih.gov/pubmed/9058718.

Qi, L. S. et al. (2013). "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression". In: *Cell* 152.5, pp. 1173–83. ISSN: 1097-4172 (Electronic) 0092-8674 (Linking). DOI: 10.1016/j.cell.2013.02.022. URL: https://www.ncbi.nlm.nih.gov/pubmed/23452860.

Ran, F. A. et al. (2013). "Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity". In: *Cell* 154.6, pp. 1380–9. ISSN: 1097-4172 (Electronic) 0092-8674 (Linking). DOI: 10.1016/j.cell.2013.08.021. URL: https://www.ncbi.nlm.nih.gov/pubmed/23992846.

Ran, F. A. et al. (2015). "In vivo genome editing using Staphylococcus aureus Cas9". In: *Nature* 520.7546, pp. 186–91. ISSN: 1476-4687 (Electronic) 0028-0836 (Link-

ing). DOI: 10.1038/nature14299. URL: https://www.ncbi.nlm.nih.gov/pubmed/25830891.

Rogers, G. L. et al. (2019). "Homologous Recombination-Based Genome Editing by Clade F AAVs Is Inefficient in the Absence of a Targeted DNA Break". In: *Mol Ther* 27.10, pp. 1726–1736. ISSN: 1525-0024 (Electronic) 1525-0016 (Linking). DOI: 10.1016/j.ymthe.2019.08.019. URL: https://www.ncbi.nlm.nih.gov/pubmed/31540849.

Romer, P. et al. (2007). "Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene". In: *Science* 318.5850, pp. 645–8. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). DOI: 10.1126/science.1144958. URL: https://www.ncbi.nlm.nih.gov/pubmed/17962564.

Ross, C. and T. E. Boroviak (2020). "Origin and function of the yolk sac in primate embryogenesis". In: *Nat Commun* 11.1, p. 3760. ISSN: 2041-1723 (Electronic) 2041-1723 (Linking). DOI: 10.1038/s41467-020-17575-w. URL: https://www.ncbi.nlm.nih.gov/pubmed/32724077.

Rother, M. B. et al. (2016). "Decreased IL7Ralpha and TdT expression underlie the skewed immunoglobulin repertoire of human B-cell precursors from fetal origin". In: *Sci Rep* 6, p. 33924. ISSN: 2045-2322 (Electronic) 2045-2322 (Linking). DOI: 10.1038/srep33924. URL: https://www.ncbi.nlm.nih.gov/pubmed/27658954.

Ru, H. et al. (2015). "Molecular Mechanism of V(D)J Recombination from Synaptic RAG1-RAG2 Complex Structures". In: *Cell* 163.5, pp. 1138–1152. ISSN: 1097-4172 (Electronic) 0092-8674 (Linking). DOI: 10.1016/j.cell.2015.10.055. URL: https://www.ncbi.nlm.nih.gov/pubmed/26548953.

Sadofsky, M. J. (2001). "The RAG proteins in V(D)J recombination: more than just a nuclease". In: *Nucleic Acids Res* 29.7, pp. 1399–409. ISSN: 1362-4962 (Electronic) 0305-1048 (Linking). DOI: 10.1093/nar/29.7.1399. URL: https://www.ncbi.nlm.nih.gov/pubmed/11266539.

Sances, S. et al. (2018). "Human iPSC-Derived Endothelial Cells and Microengineered Organ-Chip Enhance Neuronal Development". In: *Stem Cell Reports* 10.4, pp. 1222–1236. ISSN: 2213-6711 (Electronic) 2213-6711 (Linking). DOI: 10.1016/j.stemcr.2018.02.012. URL: https://www.ncbi.nlm.nih.gov/pubmed/29576540.

Sarzotti-Kelsoe, Marcella et al. (2009). "Thymic output, T-cell diversity, and T-cell function in long-term human SCID chimeras". eng. In: *Blood* 114.7, p. 1445. ISSN: 0006-4971.

Schmitt, T. M. and J. C. Zuniga-Pflucker (2002). "Induction of T cell development from hematopoietic progenitor cells by delta-like-1 in vitro". In: *Immunity* 17.6, pp. 749–56. ISSN: 1074-7613 (Print) 1074-7613 (Linking). DOI: 10.1016/s1074-7613(02)00474-0. URL: https://www.ncbi.nlm.nih.gov/pubmed/12479821.

Schmitt, T. M. et al. (2004). "Induction of T cell development and establishment of T cell competence from embryonic stem cells differentiated in vitro". In: *Nat Immunol* 5.4, pp. 410–7. ISSN: 1529-2908 (Print) 1529-2908 (Linking). DOI: 10.1038/ni1055. URL: https://www.ncbi.nlm.nih.gov/pubmed/15034575.

Schneider-Hohendorf, T. et al. (2016). "CD8(+) T-cell pathogenicity in Rasmussen encephalitis elucidated by large-scale T-cell receptor sequencing". In: *Nat Commun* 7, p. 11153. ISSN: 2041-1723 (Electronic) 2041-1723 (Linking). DOI: 10.1038/ncomms11153. URL: https://www.ncbi.nlm.nih.gov/pubmed/27040081.

Schuetz, C. et al. (2014). "SCID patients with ARTEMIS vs RAG deficiencies following HCT: increased risk of late toxicity in ARTEMIS-deficient SCID". In: *Blood* 123.2, pp. 281–9. ISSN: 1528-0020 (Electronic) 0006-4971 (Linking). DOI: 10.1182/blood-2013-01-476432. URL: https://www.ncbi.nlm.nih.gov/pubmed/24144642.

Schwartz, R. H. (1990). "A cell culture model for T lymphocyte clonal anergy". In: *Science* 248.4961, pp. 1349–56. ISSN: 0036-8075 (Print) 0036-8075 (Linking). DOI: 10.1126/science.2113314. URL: https://www.ncbi.nlm.nih.gov/pubmed/2113314.

— (1992). "Costimulation of T lymphocytes: the role of CD28, CTLA-4, and B7/BB1 in interleukin-2 production and immunotherapy". In: *Cell* 71.7, pp. 1065–8. ISSN: 0092-8674 (Print) 0092-8674 (Linking). DOI: 10.1016/s0092-8674(05)80055-8. URL: https://www.ncbi.nlm.nih.gov/pubmed/1335362.

Seet, C. S. et al. (2017). "Generation of mature T cells from human hematopoietic stem and progenitor cells in artificial thymic organoids". In: *Nat Methods* 14.5, pp. 521–530. ISSN: 1548-7105 (Electronic) 1548-7091 (Linking). DOI: 10.1038/nmeth.4237. URL: https://www.ncbi.nlm.nih.gov/pubmed/28369043.

Shannon, C. E. (1997). "The mathematical theory of communication. 1963". In: *MD Comput* 14.4, pp. 306–17. ISSN: 0724-6811 (Print) 0724-6811 (Linking). URL: https://www.ncbi.nlm.nih.gov/pubmed/9230594.

Shearer, W. T. et al. (2014). "Establishing diagnostic criteria for severe combined immunodeficiency disease (SCID), leaky SCID, and Omenn syndrome: the Primary Immune Deficiency Treatment Consortium experience". In: *J Allergy Clin Immunol* 133.4, pp. 1092–8. ISSN: 1097-6825 (Electronic) 0091-6749 (Linking). DOI: 10.1016/j.jaci.2013.09.044. URL: https://www.ncbi.nlm.nih.gov/pubmed/24290292.

Shimazaki, N., A. G. Tsai, and M. R. Lieber (2009). "H3K4me3 stimulates the V(D)J RAG complex for both nicking and hairpinning in trans in addition to tethering in cis: implications for translocations". In: *Mol Cell* 34.5, pp. 535–44. ISSN: 1097-4164 (Electronic) 1097-2765 (Linking). DOI: 10.1016/j.molcel.2009.05.011. URL: https://www.ncbi.nlm.nih.gov/pubmed/19524534.

Shultz, L. D. et al. (2005). "Human lymphoid and myeloid cell development in NOD/LtSz-scid IL2R gamma null mice engrafted with mobilized human hemopoietic stem cells". In: *J Immunol* 174.10, pp. 6477–89. ISSN: 0022-1767 (Print)

0022-1767 (Linking). DOI: 10.4049/jimmunol.174.10.6477. URL: https://www.ncbi.nlm.nih.gov/pubmed/15879151.

Slatter, M A et al. (2004). "Thyroid dysfunction after bone marrow transplantation for primary immunodeficiency without the use of total body irradiation in conditioning". In: *Bone Marrow Transplantation* 33.9, p. 949. ISSN: 0268-3369.

Song, L. et al. (2013). "Optimizing the transduction efficiency of capsid-modified AAV6 serotype vectors in primary human hematopoietic stem cells in vitro and in a xenograft mouse model in vivo". In: *Cytotherapy* 15.8, pp. 986–98. ISSN: 1477-2566 (Electronic) 1465-3249 (Linking). DOI: 10.1016/j.jcyt.2013.04.003. URL: https://www.ncbi.nlm.nih.gov/pubmed/23830234.

Srivastava, A. (2016). "In vivo tissue-tropism of adeno-associated viral vectors". In: *Curr Opin Virol* 21, pp. 75–80. ISSN: 1879-6265 (Electronic) 1879-6257 (Linking). DOI: 10.1016/j.coviro.2016.08.003. URL: https://www.ncbi.nlm.nih.gov/pubmed/27596608.

Steyer, B. et al. (2018). "Scarless Genome Editing of Human Pluripotent Stem Cells via Transient Puromycin Selection". In: *Stem Cell Reports* 10.2, pp. 642–654. ISSN: 2213-6711 (Electronic) 2213-6711 (Linking). DOI: 10.1016/j.stemcr.2017.12.004. URL: https://www.ncbi.nlm.nih.gov/pubmed/29307579.

Stiehm, E. R. and Jr. Johnston R. B. (2005). "A history of pediatric immunology". In: *Pediatr Res* 57.3, pp. 458–67. ISSN: 0031-3998 (Print) 0031-3998 (Linking).

DOI: 10.1203/01.PDR.0000151692.05422.4C. URL: https://www.ncbi.nlm.nih.gov/pubmed/15611343.

Sun, N. et al. (2014). "SunnyTALEN: a second-generation TALEN system for human genome editing". In: *Biotechnol Bioeng* 111.4, pp. 683–91. ISSN: 1097-0290 (Electronic) 0006-3592 (Linking). DOI: 10.1002/bit.25154. URL: https://www.ncbi.nlm.nih.gov/pubmed/24255022.

Surun, D., H. von Melchner, and F. Schnutgen (2018). "CRISPR/Cas9 genome engineering in hematopoietic cells". In: *Drug Discov Today Technol* 28, pp. 33–39. ISSN: 1740-6749 (Electronic) 1740-6749 (Linking). DOI: 10.1016/j.ddtec.2018.08.001. URL: https://www.ncbi.nlm.nih.gov/pubmed/30205879.

Tachibana, M. et al. (2013). "Human embryonic stem cells derived by somatic cell nuclear transfer". In: *Cell* 153.6, pp. 1228–38. ISSN: 1097-4172 (Electronic) 0092-8674 (Linking). DOI: 10.1016/j.cell.2013.05.006. URL: https://www.ncbi.nlm.nih.gov/pubmed/23683578.

Thomson, J. A. et al. (1998). "Embryonic stem cell lines derived from human blastocysts". In: *Science* 282.5391, pp. 1145–7. ISSN: 0036-8075 (Print) 0036-8075 (Linking). DOI: 10.1126/science.282.5391.1145. URL: https://www.ncbi.nlm.nih.gov/pubmed/9804556.

Thrasher, A. J. and D. A. Williams (2017). "Evolving Gene Therapy in Primary Immunodeficiency". In: *Mol Ther* 25.5, pp. 1132–1141. ISSN: 1525-0024 (Elec-

tronic) 1525-0016 (Linking). DOI: 10.1016/j.ymthe.2017.03.018. URL: https://www.ncbi.nlm.nih.gov/pubmed/28366768.

Til, N. P. van et al. (2012). "Correction of murine Rag2 severe combined immunodeficiency by lentiviral gene therapy using a codon-optimized RAG2 therapeutic transgene". In: *Mol Ther* 20.10, pp. 1968–80. ISSN: 1525-0024 (Electronic) 1525-0016 (Linking). DOI: 10.1038/mt.2012.110. URL: https://www.ncbi.nlm.nih.gov/pubmed/22692499.

Til, N. P. van et al. (2014a). "Recombination-activating gene 1 (Rag1)-deficient mice with severe combined immunodeficiency treated with lentiviral gene therapy demonstrate autoimmune Omenn-like syndrome". In: *J Allergy Clin Immunol* 133.4, pp. 1116–23. ISSN: 1097-6825 (Electronic) 0091-6749 (Linking). DOI: 10.1016/j.jaci.2013.10.009. URL: https://www.ncbi.nlm.nih.gov/pubmed/24332219.

— (2014b). "Recombination-activating gene 1 (Rag1)-deficient mice with severe combined immunodeficiency treated with lentiviral gene therapy demonstrate autoimmune Omenn-like syndrome". In: *J Allergy Clin Immunol* 133.4, pp. 1116–23. ISSN: 1097-6825 (Electronic) 0091-6749 (Linking). DOI: 10.1016/j.jaci.2013.10.009. URL: https://www.ncbi.nlm.nih.gov/pubmed/24332219.

Treangen, T. J. and S. L. Salzberg (2011). "Repetitive DNA and next-generation sequencing: computational challenges and solutions". In: *Nat Rev Genet* 13.1,

pp. 36–46. ISSN: 1471-0064 (Electronic) 1471-0056 (Linking). DOI: 10.1038/nrg3117. URL: https://www.ncbi.nlm.nih.gov/pubmed/22124482.

Urnov, F. D. et al. (2010). "Genome editing with engineered zinc finger nucleases". In: *Nat Rev Genet* 11.9, pp. 636–46. ISSN: 1471-0064 (Electronic) 1471-0056 (Linking). DOI: 10.1038/nrg2842. URL: https://www.ncbi.nlm.nih.gov/pubmed/20717154.

Villa, A. and L. D. Notarangelo (2019). "RAG gene defects at the verge of immunodeficiency and immune dysregulation". In: *Immunol Rev* 287.1, pp. 73–90. ISSN: 1600-065X (Electronic) 0105-2896 (Linking). DOI: 10.1111/imr.12713. URL: https://www.ncbi.nlm.nih.gov/pubmed/30565244.

Vodyanik, M. A. et al. (2005). "Human embryonic stem cell-derived CD34+ cells: efficient production in the coculture with OP9 stromal cells and analysis of lymphohematopoietic potential". In: *Blood* 105.2, pp. 617–26. ISSN: 0006-4971 (Print) 0006-4971 (Linking). DOI: 10.1182/blood-2004-04-1649. URL: https://www.ncbi.nlm.nih.gov/pubmed/15374881.

Wang, J. et al. (2016). "Highly efficient homology-driven genome editing in human T cells by combining zinc-finger nuclease mRNA and AAV6 donor delivery". In: *Nucleic Acids Res* 44.3, e30. ISSN: 1362-4962 (Electronic) 0305-1048 (Linking). DOI: 10.1093/nar/gkv1121. URL: https://www.ncbi.nlm.nih.gov/pubmed/26527725.

Wang, K. et al. (2016). "Efficient Generation of Orthologous Point Mutations in Pigs via CRISPR-assisted ssODN-mediated Homology-directed Repair". In: *Mol Ther Nucleic Acids* 5.11, e396. ISSN: 2162-2531 (Print) 2162-2531 (Linking). DOI: 10.1038/mtna.2016.101. URL: https://www.ncbi.nlm.nih.gov/pubmed/27898095.

Ward, A. et al. (2018). "The RAG-2 Inhibitory Domain Gates Accessibility of the V(D)J Recombinase to Chromatin". In: *Mol Cell Biol* 38.15. ISSN: 1098-5549 (Electronic) 0270-7306 (Linking). DOI: 10.1128/MCB.00159-18. URL: https://www.ncbi.nlm.nih.gov/pubmed/29760281.

Wei, S. et al. (1994). "The extent of the human germline T-cell receptor V beta gene segment repertoire". In: *Immunogenetics* 40.1, pp. 27–36. ISSN: 0093-7711 (Print) 0093-7711 (Linking). DOI: 10.1007/BF00163961. URL: https://www.ncbi.nlm.nih.gov/pubmed/8206523.

Weinberg, K. et al. (1993). "T lymphocyte ontogeny in adenosine deaminase-deficient severe combined immune deficiency after treatment with polyethylene glycol-modified adenosine deaminase". In: *J Clin Invest* 92.2, pp. 596–602. ISSN: 0021-9738 (Print) 0021-9738 (Linking). DOI: 10.1172/JCI116626. URL: https://www.ncbi.nlm.nih.gov/pubmed/8349799.

Wodrich, H., A. Schambach, and H. G. Krausslich (2000). "Multiple copies of the Mason-Pfizer monkey virus constitutive RNA transport element lead to enhanced HIV-1 Gag expression in a context-dependent manner". In: *Nucleic Acids Res*

28.4, pp. 901–10. ISSN: 1362-4962 (Electronic) 0305-1048 (Linking). DOI: 10.1093/nar/28.4.901. URL: https://www.ncbi.nlm.nih.gov/pubmed/10648781.

Wu, X. et al. (2003). "Transcription start regions in the human genome are favored targets for MLV integration". In: *Science* 300.5626, pp. 1749–51. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). DOI: 10.1126/science.1083413. URL: https://www.ncbi.nlm.nih.gov/pubmed/12805549.

Yamada, M. et al. (2014). "Human oocytes reprogram adult somatic nuclei of a type 1 diabetic to diploid pluripotent stem cells". In: *Nature* 510.7506, pp. 533–6. ISSN: 1476-4687 (Electronic) 0028-0836 (Linking). DOI: 10.1038/nature13287. URL: https://www.ncbi.nlm.nih.gov/pubmed/24776804.

Yang, B. et al. (2019). "Haploidentical versus matched donor stem cell transplantation for patients with hematological malignancies: a systemic review and meta-analysis". In: *Bone Marrow Transplant* 54.1, pp. 99–122. ISSN: 1476-5365 (Electronic) 0268-3369 (Linking). DOI: 10.1038/s41409-018-0239-9. URL: https://www.ncbi.nlm.nih.gov/pubmed/29988061.

Yarrington, R. M. et al. (2018). "Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo". In: *Proc Natl Acad Sci U S A* 115.38, pp. 9351–9358. ISSN: 1091-6490 (Electronic) 0027-8424 (Linking). DOI: 10.1073/pnas.1810062115. URL: https://www.ncbi.nlm.nih.gov/pubmed/30201707.

Yu, S. F. et al. (1986). "Self-inactivating retroviral vectors designed for transfer of whole genes into mammalian cells". In: *Proc Natl Acad Sci U S A* 83.10, pp. 3194–8. ISSN: 0027-8424 (Print) 0027-8424 (Linking). DOI: 10.1073/pnas.83.10.3194. URL: https://www.ncbi.nlm.nih.gov/pubmed/3458176.

Zhou, S. et al. (2010). "A self-inactivating lentiviral vector for SCID-X1 gene therapy that does not activate LMO2 expression in human T cells". In: *Blood* 116.6, pp. 900–8. ISSN: 1528-0020 (Electronic) 0006-4971 (Linking). DOI: 10.1182/blood-2009-10-250209. URL: https://www.ncbi.nlm.nih.gov/pubmed/20457870.

Zhou, S. et al. (2016). "Evaluating the Safety of Retroviral Vectors Based on Insertional Oncogene Activation and Blocked Differentiation in Cultured Thymocytes". In: *Mol Ther* 24.6, pp. 1090–1099. ISSN: 1525-0024 (Electronic) 1525-0016 (Linking). DOI: 10.1038/mt.2016.55. URL: https://www.ncbi.nlm.nih.gov/pubmed/26957223.

Zuniga-Pflucker, J. C. (2004). "T-cell development made simple". In: *Nat Rev Immunol* 4.1, pp. 67–72. ISSN: 1474-1733 (Print) 1474-1733 (Linking). DOI: 10.1038/nri1257. URL: https://www.ncbi.nlm.nih.gov/pubmed/14704769.