

LENTIVIRAL GENE THERAPY PROVIDES ROBUST PROTECTION AGAINST  
PANCYTOPENIA IN AN INDUCIBLE CODANIN-1 KNOCKOUT MOUSE MODEL

Perry Alexandr Demsko

A DISSERTATION

in

Cell and Molecular Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2025

Supervisor of Dissertation

Stefano Rivella, PhD, Kwame Ohene Frempong Chair on Sickle-cell Anemia

Graduate Group Chairperson

Daniel Kessler, PhD, Associate Professor of Cell & Developmental Biology

Dissertation Committee

Peter Kurre, PhD, Buck Family Chair in Pediatric Hematology

Saar Gill, MD, PhD, Associate Professor of Medicine

Wei Tong, PhD, Professor of Pediatrics

Hamideh Parhiz, PhD, Assistant Professor of Systems Pharmacology and Translational  
Therapeutics

LENTIVIRAL GENE THERAPY PROVIDES ROBUST PROTECTION AGAINST  
PANCYTOPENIA IN AN INDUCIBLE CODANIN-1 KNOCKOUT MOUSE MODEL

COPYRIGHT

2025

Perry Alexandr Demsko

This work is licensed under the Creative Commons Attribution – Non-Commercial – ShareAlike  
4.0 License

To view a copy of this license, please visit  
<http://creativecommons.org/licenses/by-nc-sa/4.0/us/>

*To Audrey,*

*You're my moon, I'm your moon*

## ACKNOWLEDGEMENTS

There are too many of you to list; friends, mentors, and family who have gotten me to this juncture. My parents and siblings gave me the environment to be intellectually curious. The Baltimore Underground Science Space let me lift a pipette for the first time. Dr. Mariaelena Pierobon took a chance on an unproven undergrad. Mirek hammered in the finer aspects of molecular cloning and viral production. Mouse nugget had the genotype I needed. The whole GTV student body kept me grounded, provided mentorship and levity and allowed me to return in kind. Jake and Alex provided friendship, candor, and hope for the future. Jonathan provided essential mentorship and went above and beyond.

Lastly, I'd like to thank Stefano Rivella for having a contagious passion for improving the lives of patients, and for providing me an environment where I could learn from and work with some of the smartest and most capable scientists I have ever known, himself included. Ama, Lucas, and Naoto have had outsized roles in making me the scientist I am today.

To all of you mentioned here, and to all those countless others who are not: thank you.

I think I can, I think I can, I think I can.

## ABSTRACT

### LENTIVIRAL GENE THERAPY PROVIDES ROBUST PROTECTION AGAINST PANCYTOPENIA IN AN INDUCIBLE CODANIN-1 KNOCKOUT MOUSE MODEL

Perry Demsko

Stefano Rivella

The Congenital Dyserythropoietic Anemias (CDAs) are a diverse group of rare heritable blood disorders characterized by anomalies in erythroid maturation and red blood cell production. Mechanistically varied, CDAs have lacked suitable animal models which has inhibited development of tractable therapies. CDA1a is a macrocytic CDA caused by biallelic hypofunctional mutations in Codanin-1, an enigmatic protein whose precise function has not been conclusively established. CDA1a causes morbidity and mortality through moderate to severe anemia and concurrent iron overload. We sought to rectify the lack of animal model by producing a suitable inducible Codanin-1 knockout mouse line with two disparate mechanisms of cre-based deletion, allowing for assessment of treatments for CDA1a. Using our model, we found that deletion of Codanin-1 is incompatible with life with death invariably resulting upon induction. Seeking to take advantage of this opportunity we utilized lentiviral gene therapy to restore Codanin-1 functionality, thereby establishing the ability to use gene therapy to correct CDA1a. Using our model and lentiviral system, we sought to explore human Codanin-1 mutations for better understanding of the disease pathology of CDA1a. We found that human Codanin-1 mutations do not cause disease in mice, allowing us to hypothesize about fundamental differences between human and mouse erythropoiesis. Finally, we deployed machine learning to interrogate *in silico* the functionality of Codanin-1 and uncovered a putative disease mechanism for CDA1a.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	iv
ABSTRACT .....	v
LIST OF TABLES .....	viii
LIST OF FIGURES .....	ix
<b>CHAPTER 1 – INTRODUCTION</b> .....	<b>1</b>
<b>Hematopoiesis</b> .....	<b>1</b>
<b>Genetic Diseases of the Hematopoietic Lineage</b> .....	<b>4</b>
<b>Gene Therapy</b> .....	<b>5</b>
<b>Lentivirus Biology and Lifecycle</b> .....	<b>6</b>
<b>The Lentivirus as a Gene Therapy Vector</b> .....	<b>9</b>
<b>Flow Cytometry</b> .....	<b>10</b>
<b>Immunophenotyping of the Hematopoietic Lineage</b> .....	<b>12</b>
<b>Complete Blood Counts</b> .....	<b>13</b>
<b>The Goals of this Dissertation</b> .....	<b>13</b>
<b>CHAPTER 2 – LENTIVIRAL GENE THERAPY PROVIDES ROBUST PROTECTION AGAINST PANCYTOPENIA IN AN INDUCIBLE CDANIN-1 KNOCKOUT MODEL</b> .....	<b>21</b>
<b>Abstract</b> .....	<b>21</b>
<b>Introduction</b> .....	<b>22</b>
<b>Results</b> .....	<b>23</b>
<i>Inducible <math>Cdan1^{\Delta Mx}</math> knockout mice develop pancytopenia.</i> .....	<b>23</b>
<i>Lentiviral gene therapy with human CDAN1 rescues <math>Cdan1^{\Delta LNP}</math> knockout mice from non-engraftment and sustains long-term hematopoiesis.</i> .....	<b>25</b>
<i>Expression of human CDAN1 in <math>Cdan1^{\Delta LNP}</math> knockout cells is associated with mild abnormalities in erythropoiesis.</i> .....	<b>27</b>
<b>Discussion</b> .....	<b>28</b>
<b>Figures</b> .....	<b>38</b>
<b>Supplemental Figures</b> .....	<b>47</b>
<b>CHAPTER 3 – MUTATED CDAN1 VECTORS AND MACHINE LEARNING REVEAL ASPECTS OF CDAN1 FUNCTIONALITY</b> .....	<b>52</b>
<b>Introduction</b> .....	<b>52</b>
<b>Results</b> .....	<b>52</b>
<i>CDAN1 vectors with multiple mutated residues do not rescue <math>Cdan1^{\Delta LNP}</math> mice.</i> .....	<b>53</b>
<i>Machine learning provides a putative mechanism for CDA1a pathology.</i> .....	<b>54</b>
<b>Discussion</b> .....	<b>58</b>

<b>Methods.....</b>	<b>59</b>
<b>Figures.....</b>	<b>60</b>
<b>Tables.....</b>	<b>68</b>
<b>CHAPTER 4: CONCLUSIONS AND FUTURE DIRECTIONS.....</b>	<b>70</b>
<b>The Mouse as a Model Organism .....</b>	<b>70</b>
<b>Deficiencies of the Mouse Model Organism .....</b>	<b>71</b>
<b>Mouse Modeling of CDA1a .....</b>	<b>73</b>
<b>The Curative Power of Gene Therapy .....</b>	<b>74</b>
<b>Machine Learning as a Tool for Biological Exploration.....</b>	<b>75</b>
<b>Codanin-1 as a Central Axis of Cell Fate.....</b>	<b>76</b>
<b>Future Lines of Inquiry .....</b>	<b>76</b>
<i>AlphaFold Validation Via Co-immunoprecipitation.....</i>	<i>77</i>
<i>Assessment of ASF1 phosphorylation in CDA1a.....</i>	<i>78</i>
<i>Xenotransplantation for Faithful Recapitulation of CDA1a.....</i>	<i>78</i>
<b>Final Thoughts .....</b>	<b>79</b>
<b>BIBLIOGRAPHY.....</b>	<b>80</b>

## LIST OF TABLES

Table 3.1. PTM and iPTM scores for the AlphaFold generated ASF1-CDAN1-CDIN1 complex.....	72
Table 3.2. Residues phosphorylated on ASF1 isoforms.....	73

## LIST OF FIGURES

Figure 1.1. Simplified diagram of hematopoiesis.....	16
Figure 1.2. Simplified diagram of erythropoiesis.....	18
Figure 1.3. Graphical overview of factors involved in regulation of erythropoiesis.....	19
Figure 1.4. Graphical representation of steps involved in lentiviral reverse transcription converting the ssDNA genome into dsDNA for integration into the infected cell's genome.....	20
Figure 2.1. The impact of Cdan1 deletion on survival and steady state hematopoiesis.....	39
Figure 2.2. The impact of Cdan1 deletion on endogenous erythropoiesis.....	40
Figure 2.3. The impact of Cdan1 deletion on hematopoiesis and bone marrow integrity.....	41
Figure 2.4. Rescue of engraftment of Cdan1 <sup>ΔLNP</sup> marrow via lentiviral gene therapy.....	43
Figure 2.5. Long-term assessment of hematopoietic parameters in lentiviral corrected Cdan1 <sup>ΔLNP</sup> HSCs.....	45
Figure 2.6. Long-term assessment of lentiviral corrected Cdan1 <sup>ΔLNP</sup> marrow for maintenance of hematopoiesis.....	46
Figure 2.7. Histopathological assessment of lentiviral corrected Cdan1 <sup>ΔLNP</sup> bone marrow.....	48
Figure 2.S1.....	50
Figure 2.S2. Assessment of Secondary transplant chimerism, Cdan1 deletion, and VCN.....	52
Figure 2.S3. Assessment of liver iron build-up, liver/spleen to body weight ratios, and peripheral blood morphology.....	53
Figure 3.1. Complete blood count values of Cdan1 <sup>ΔLNP</sup> marrow treated with mutant lentiviral gene therapy vectors.....	63
Figure 3.2. AlphaFold generated model of CDAN1 with CDA1a causing residues indicated.....	64
Figure 3.3. AlphaFold generated model of the ASF1-CDAN1-CDIN1 structure.....	65
Figure 3.4. AlphaFold generated model of the ASF1-CDAN1-CDIN1 structure, obverse side...66	
Figure 3.5. Detail of ASF1 C-terminus within CDIN1.....	67
Figure 3.6. Specific residues implicated in ASF1-CDIN1 binding.....	68
Figure 3.7. pLDDT scores for each residue in the ASF1-CDAN1-CDIN1 complex.....	70

## CHAPTER 1 – INTRODUCTION

### Hematopoiesis

The mammalian blood system is an intricate milieu of numerous cell types with essential functions varying from oxygen transport, tissue maintenance, and both innate and adaptive immunity. A majority of blood cell lineages are short-lived, necessitating constant replenishment from progenitor cell lineages, which themselves are replenished by hematopoietic stem and progenitor cells (HSPCs).<sup>1</sup> At the apex of this hierarchy resides the long term hematopoietic stem cell (LT-HSC), capable of repopulating the entire hematopoietic compartment via a capacity for self-renewal and asymmetrical multilineage differentiation.<sup>2</sup> Indeed, it has been demonstrated that a single transplanted LT-HSC can regenerate the entire blood system in mice when provided with sufficient supporting cells and transplantation has been the experimental gold standard for confirmation of LT-HSC identity and function with utility well beyond cell-surface markers.<sup>3-6</sup> After the LT-HSC, short-term HSCs (ST-HSC) are a more active and less quiescent population, handling much of the requirements of hematopoiesis in the immediate term. Multipotent progenitors (MPPs) are a heterogeneous population produced via asymmetrical division by HSCs, characterized by specific lineage biases that influence downstream cell type production. Two subsets, MPP2 and MPP3, are myeloid biased, with MPP4s exhibiting a preference for lymphoid biased cell production.<sup>7</sup> There is some evidence for lineage bias derived from specific HSC clones as well.<sup>8</sup>

Perhaps the most incredible subset of hematopoiesis in terms of raw cell count is erythropoiesis, the generation of red blood cells (RBCs). In the adult human steady-state

erythropoietic output is legion, generating approximately 2.5 million RBCs every second.<sup>9</sup> This process continues indefinitely for decades of adult life resulting in quadrillions of cells over the human lifespan. Erythropoiesis in adults under physiologically normal conditions is performed in the bone marrow within erythroblastic islands, clusters of erythroid precursors around a central nurse macrophage.<sup>10-13</sup> This clustering maintains essential cell-cell contacts that enforce cellular fate and proliferation and down-regulate apoptosis of the developing erythroblasts as the erythroblasts mature into reticulocytes and ultimately RBCs.<sup>14-16</sup> Beginning with erythroid-committed progenitors that arise from HSPCs, erythropoiesis can be broadly bifurcated into three general stages, early phase, late/terminal phase, and lastly reticulocyte maturation, inclusive of many steps within each.<sup>17</sup>

Within the early phase of erythropoiesis burst-forming unit-erythroid (BFU-E) cells, so named because of their explosive growth curve that leaves characteristically large colonies, establish themselves then mature into colony forming unit erythroid (CFU-E) cells, which begin the association with the nurse macrophage.<sup>18</sup> These two populations, while utilized for convenience of classification, are highly heterogenous and much overlap exists between BFU-Es and CFU-Es.<sup>19</sup> During the early stage extrinsic factors regulate the proliferation and differentiation programming of BFU-Es and CFU-Es. Early erythroid progenitor proliferation has been shown to be advanced by IL-3, with most of the effect seen at the BFU-E stage.<sup>20</sup> Stem cell factor (SCF) is also an essential early phase cytokine that acts on both BFU-Es and CFU-Es, despite its receptor, CD117, having extensive functionality in total hematopoiesis and beyond.<sup>21-</sup><sup>24</sup> Erythropoietin (EPO) is perhaps the hormone most associated with erythropoiesis. Generated in the kidney and upregulated by hypoxic stress, EPO upregulates RBC generation and angiogenesis.<sup>25,26</sup> Indeed, recombinant forms of EPO are used both as a treatment for anemia

and abused as a sports doping agent.<sup>27-29</sup> The Erythropoietin receptor (EPO-R) is found at low levels on BFU-Es, with much higher levels of expression subsequently found on CFU-Es.<sup>30</sup> At the CFU-e stage erythroid precursors are dependent on EPO for survival, providing the organism with a high degree of control over RBC production in response to environmental and physiological conditions.<sup>31,32</sup>

Late or terminal phase erythropoiesis features cells with much more distinct, identifiable morphologies, leading to more convenient classifications.<sup>33</sup> It should be noted however, as with much of hematopoiesis, there remains a continuum of differentiation resulting in a blurring of cell types when interrogated beyond physical attributes.<sup>34</sup> At this stage the erythroblastic island finalizes its formation, with a central nurse macrophage interacting with over two dozen erythroblasts. Late phase begins with the proerythroblast, then proceeds through basophilic, polychromatic, and orthochromatic before finally producing the reticulocyte, an immature stage immediately before a true RBC.<sup>35</sup> Each stage progressively accumulates hemoglobin, decreases cell volume, and prepares for enucleation by reducing the size of the nucleus and undergoing chromatin condensation.<sup>36</sup> By the end of the orthochromatic stage the nucleus has been compacted by an order of magnitude and is subsequently ejected as a so-called pyrenocyte, which is phagocytosed by the nurse macrophage.<sup>37</sup> Once an enucleated reticulocyte is established, it enters into peripheral blood circulation and undergoes external and internal membrane remodeling plus a loss of mitochondria and organelles, at last becoming an RBC.<sup>38-42</sup>

As with the early phase, many intrinsic and extrinsic factors contribute to the process of late-stage erythropoiesis. EPO-R levels begin to decrease as the late phase proceeds, leading to what is termed EPO-independent erythropoiesis, where the erythroblasts are not dependent on

the presence of EPO to survive and proliferate.<sup>43</sup> During late-phase erythropoiesis, the central nurse macrophage maintains essential cell-cell contacts and further assists by consuming nuclei as they are ejected from developing erythroblasts. The transferrin receptor-1 (CD71) is highly expressed at this stage, importing iron from circulation for use in hemoglobin production and is furthermore a useful marker for experimental analysis of erythropoiesis.<sup>44,45</sup> At this stage, erythroblasts signal the liver via erythroferrone to increase the available iron in circulation.<sup>46</sup> As iron is an essential co-factor of hemoglobin, the regulation and cross-talk between erythropoiesis and iron homeostasis is not surprising. In fact, erythropoiesis is the largest single metabolic iron sink in humans and jawed vertebrates, but the intricacies of iron metabolism are beyond the scope of this dissertation.

### **Genetic Diseases of the Hematopoietic Lineage**

Given the importance of hematopoiesis to sustaining life, it is not surprising to find numerous disorders arising from mutations in cell types emerging from this compartment. Severe combined immunodeficiency (SCID) manifests with a failure of maturation, function or maintenance of multiple leukocyte lineages, including T-cells, B-cells, NK-cells and others.<sup>47,48</sup> SCID can arise from numerous molecular mechanisms, including loss of the common gamma chain in x-linked SCID, and defects in purine metabolism as seen in Adenosine deaminase deficiency SCID.<sup>49,50</sup> Sickle cell disease (SCD) is perhaps the hematopoietic disorder most well known in the public consciousness. Caused by a mutation in beta-globin that replaces a hydrophilic glutamic acid residue with a hydrophobic valine, leading to aggregation of the hemoglobin heterotetramer in RBCs and subsequent, characteristic RBC sickling and rigidity.<sup>51</sup> Presenting with high morbidity, SCD creates an untold economic and personal burden.<sup>52</sup> Patients

can suffer from strokes along with debilitating and painful vaso-occlusive crises, depriving them of many life opportunities.<sup>53</sup> The thalassemias are caused by the absence of, or significant reduction in, production of hemoglobin. Alpha thalassemia is caused by deficiencies in the alpha globin component of hemoglobin, whereas beta thalassemia corresponds to the absence or reduction of beta-globin.<sup>54,55</sup> Both of these conditions are the cause of much economic cost and morbidity. Many patients are dependent on blood transfusions and can develop iron overload due to dysregulation of iron metabolism caused by chronic anemia and transfusions.<sup>56</sup> Additional diseases have further, disparate molecular mechanisms. Dyskeratosis congenita is caused by mutations in components related to telomere maintenance, including TERC, an lncRNA and essential cofactor of Telomerase reverse transcriptase.<sup>57,58</sup> The failure of telomere maintenance eventually results in bone marrow failure, essentially from early aging of the marrow. Fanconi anemia is a family bone marrow failure disorders, arising from errors of DNA repair within the FA/BRCA pathway, with at least 21 proteins implicated.<sup>59</sup> Over time, DNA damage accumulates, leading to bone marrow failure. Both Dyskeratosis congenita and Fanconi anemia have drastically increased instances of cancer as an additional complication.<sup>60,61</sup> The disease focus of this dissertation will be on the Congenital Dyserythropoietic Anemias (CDAs), with an emphasis on CDA1a, so these will be discussed in depth in chapter 2.

## **Gene Therapy**

Gene therapy is a uniquely powerful and curative treatment paradigm wherein a deficient, disease-causing gene is replaced with a functional, therapeutic copy.<sup>62</sup> The technologies used to introduce therapeutic genetic material into patient's cells are referred to as vectors and are often based on viruses found in nature.<sup>63,64</sup> Utilizing recombinant DNA technology, molecular biology

has the ability to adapt viruses to deliver designer cargos, such as therapeutic genes. Viruses have been finely crafted by evolution and natural selection to deliver their genetic material into cells, having billions of years to solve the puzzle of cellular entry and heterologous gene expression. Many viral vectors have been deployed in gene therapy. The most common vector deployed in approved gene therapies is the Adeno-associated virus (AAV), having a good safety profile and modest immunogenic potential.<sup>65,66</sup> Unfortunately, its packaging capacity is limited at 4.7 kb and it delivers a non-integrating cassette, a limitation that will be discussed in a later section.<sup>67</sup> Adenoviral vectors and gammaretroviral vectors have been deployed in a clinical context, but have fallen out of favor due to poor safety.<sup>68,69</sup> After AAV, the most commonly utilized viral vector is the lentiviral vector, which will form the crux of this dissertation and will be expanded upon later. Additional viral vectors have found utility in a basic research context, for example *baculovirus*, capable of handling large, polycistronic transfer cassettes.<sup>70</sup> Non-viral vectors such as Lipid Nanoparticles (LNPs) are on the horizon, accelerated by their use for vaccination in response to the COVID-19 pandemic.<sup>71-73</sup> Gene therapy as a treatment paradigm is poised to completely resolve the underlying mechanisms of disease, rather than treating emerging symptoms.<sup>74-78</sup>

## **Lentivirus Biology and Lifecycle**

Lentiviruses belong to the genus *Lentivirus*, within the *Retroviridae* family. The *Retroviridae* are eponymic with their lifecycle requirement of reverse transcription, wherein the ssRNA viral genome is converted to dsDNA via a conserved reverse transcriptase and integrated into the host genome to form the provirus.<sup>79</sup> This strategy benefits the virus in multiple ways, firstly by removing the viral genome from innate immune nucleic acid sensors, and by allowing

the virus to propagate into daughter cells within mitotic cell types, ensuring higher numbers of viral progeny.<sup>80</sup> To wit: Human immunodeficiency virus (HIV), the causative agent of acquired immunodeficiency syndrome (AIDS), can avoid immune surveillance for years while simultaneously targeting CD4<sup>+</sup> T-cells, further inhibiting immunity and resulting in a failure of cellular immunity.<sup>81</sup> This immunocompromised state results in death from opportunistic infections.

The lentiviral genome is simple and compact, generally ranging from 7-12 kbp, depending on species. For the sake of simplicity, HIV-1 will be used in this discussion as it is the most well characterized lentivirus. Flanked by long-terminal repeats (LTRs), the quintessential lentivirus encodes three structural protein coding genes, *gag*, *pol*, and *env*, and numerous additional accessory genes that aid in virion assembly, maturation, and immune evasion.<sup>82</sup> A small recognition motif, *psi*, is present on the viral genome and targets it for packaging within assembling particles. Encoding the main elements constituting the viral structure is *gag*, expressing a sophisticated polyprotein that is proteolytically processed into the p17 matrix protein, the p24 capsid protein, the NC nucleocapsid zinc-finger protein, and finally p6. A second polyprotein, *pol*, encodes the machinery required to perform reverse transcription of the viral genome, integration into the host genome, and maturation of the virion via proteolytic cleavage both of *gag* and *pol* itself. Intriguingly, *pol* utilizes a ribosomal frameshift to generate alternative transcripts in a 20:1 stoichiometric ratio to facilitate faithful virion assembly. *Env* encodes the gp141 fusogen protein, responsible for cellular attachment and entry.

To establish infection of a host cell, the virus first attaches itself to a cell surface receptor, in the case of HIV-1, this receptor is CD4, along with CCR5 or CXCR4 as a coreceptor

depending on strain.<sup>83</sup> Once the viral fusogen binds the receptor and co-receptor, a tremendous structural change occurs in the fusogen, providing the energy needed to merge the viral envelope with the cell membrane, depositing the viral core within and completing cellular entry.<sup>84</sup> Now exposed to the cytoplasm, the viral core begins the process of uncoating while simultaneously translocating to the nucleus and beginning the process of reverse transcription, forming the reverse transcription complex.<sup>85</sup> To convert the ssRNA viral genome into dsDNA ready to be integrated, the virus utilizes a host tRNA as a primer to initiate DNA polymerase activity of the reverse transcriptase (RT) with the ssRNA as a template, starting at the tRNA binding site and proceeding towards the 5' end. As the RT proceeds, the ssRNA template is degraded by the RT's RNase H activity, leaving a negative-sense ssDNA intermediate. Once the RT reaches the 5' end of the ssRNA template, it translocates to the 3' end, using the nascent ssDNA strand generated from the 5' end to bind, taking advantage of the complementary nature of the 5' and 3' LTRs. The generation of the full minus strand then proceeds as previously performed, with RNase H activity removing the ssRNA template as the RT proceeds. Two purine-rich tracts on the ssRNA template preserve these sections of RNA, which are used as primers for the final round of reverse transcription, forming the full dsDNA proviral genome ready for integration as part of the pre-integration complex (PIC).<sup>86</sup> This entire process occurs within the viral capsid, obscuring the reverse transcription process from innate immune pattern recognition while the capsid traffics to the nucleus, exploiting host active-transport machinery. Once passed through the nuclear pore complex, the PIC undergoes partial disassembly and integrates the viral payload into the host genome, forming the provirus. To complete the viral lifecycle, the 5' LTR acts as a promoter, simultaneously producing transcripts of the ssRNA genome and expressing the proteins necessary for the production of new virions. Newly assembled virions bud from the cell

membrane, enveloping themselves in the lipid bilayer as they egress, further reducing immune interrogation and completing the viral life cycle.<sup>87</sup>

### **The Lentivirus as a Gene Therapy Vector**

The unique features of the lentivirus make it an ideal gene therapy vector for diseases of the hematopoietic compartment. Firstly, the ability to integrate genetic cargo is necessary to sustain fidelity of treatment in hematopoiesis as non-integrative therapeutic cargos, as found in AAVs and similar, would be diluted out over the continuous rounds of cell division that occur in hematopoiesis, eventually rendering the therapy ineffective. Secondly, lentiviruses can effectively transduce non-dividing cells, such as quiescent HSCs, unlike many retroviruses which require dissolution of the nuclear envelope as encountered in cells within the cell cycle.<sup>88</sup> Lastly, and perhaps most importantly, lentiviral integration patterns have been established to be within the bodies of genes, away from promoter or enhancer elements, significantly reducing the risk of insertion-based oncogenesis, a known issue with early gammaretroviral vectors.<sup>89</sup>

Conversion of the lentivirus into a gene therapy vector is straightforward. Current third generation lentiviral systems utilize four plasmids co-transfected into producer cell lines.<sup>90</sup> The essential viral elements are split up into multiple plasmids, with *gag* and *pol* on one plasmid; and *rev*, an essential packaging accessory, on a second. Two additional plasmids complete the set, with *env* (or more often, a heterologous envelope fusogen, such as the G-protein of Vesicular stomatitis virus, VSV-G) on its own, then finally the transfer cassette on the fourth. The transfer cassette contains the therapeutic gene of interest to be transduced under control of its own promoter and flanked by so-called self-inactivating LTRs that upon integration as a provirus lack promoter activity, improving the safety profile. Due to the lentivirus' genome size, the transfer

cassette can be modestly large, in excess of 9kb, nearly twice the size of AAV. In fact, transfer cassettes as large as 12 kb have been experimentally deployed, but vector production efficiency is low enough to leave this as an experimental curiosity.<sup>91</sup>

A limitation of lentiviral vectors is their reduced efficacy when deployed *in vivo*, in contrast to AAV which is extensively deployed clinically in the human body. Most lentiviral vectors are pseudotyped with VSV-G, which binds the LDL receptor (LDL-R), which is ubiquitously expressed, preventing tissue targeting. Furthermore, due to an evolutionary arms race between mammals and *Rhabdoviridae*, VSV-G is inactivated by the complement system.<sup>92</sup> While extensive work has been performed on alternative pseudotypes, including by this author, much refinement is still required in this space. Clinical application of lentiviral vectors in the present day utilize *ex vivo* transduction protocols. Briefly, patient cells are removed from the body, transduced in culture, then returned after extensive quality assurance checks.<sup>93</sup> In the human hematopoietic context, the cell type of interest is the CD34<sup>+</sup> HSC, which is obtained either from bone marrow aspirate, or by leukapheresis of cytokine mobilized peripheral blood.<sup>94</sup>

## **Flow Cytometry**

Flow cytometry is an essential cellular and molecular biology technique for the assessment of biological samples.<sup>95</sup> Individual cells are decorated with fluorescent markers, either chemical dyes, modified amino acids, expressed fluorescent proteins, or most commonly, fluorophore-tagged antibodies specific for proteins of interest. The labeled cells are then suspended in a buffered solution and passed via a fluidic system past multiple sets of lasers producing discrete wavelengths of light. The laser light sources interact with each cell and the fluorescent markers, exciting the fluorescent tags to reemit light in a specific manner based on

the laser light frequency and the excitation and emission properties of the tag. Additionally, light scatter is measured both straight on, revealing front-scatter or the relative size of the cell, and at 90 degrees revealing side scatter or the internal complexity/granularity of the cell. A partial limitation of flow cytometry is the possibility of crossover of emission spectra between two or more tags, leading to noisy signals with uncertain sources. For example, APC and PE, two common fluorophore tags both demonstrate partial emission overlap at between 600 and 650 nm. To remedy this phenomenon, flow cytometry utilizes a technique called compensation that, in conjunction of different laser wavelengths and bandpass filters, allows for signal processing techniques to remove the errant crossover signal, revealing each marker independently and faithfully. With an up-to-date flow cytometer, it is possible to perform experiments containing in excess of two dozen individual markers, allowing for intensive and robust exploration of biological systems. Certain tags can never be utilized together, even with compensation. For example, 7-AAD, a marker for differentiating between live or dead cells, is indistinguishable from mCherry, a red fluorescent protein. Coupled with charged electrodes, one can even physically sort cells into different collection receptacles, allowing for isolation of discrete cell types for downstream experimentation.

Many fields of study utilize flow cytometry, including immunology, cancer biology, and infectious disease research. In chapter 2, this dissertation makes extensive use of immunophenotyping, the interrogation of heterogeneous cell populations via flow cytometry utilizing fluorescent antibodies specific for discrete cell markers. For a general example, T-cells can be assessed by detecting which cells express CD45 and CD3 simultaneously, then further divided into helper or cytotoxic sub-populations depending on whether they also express CD4 or CD8. Indeed, a full T<sub>reg</sub> panel contains approximately 15 different markers.

## **Immunophenotyping of the Hematopoietic Lineage**

An application of immunophenotyping germane to this dissertation is its ability to assess fundamental hematopoietic processes, including the presence of hematopoietic stem cells and the performance of erythropoiesis. Identification of murine HSCs is relatively straight forward. The schema is as follows: after identification of the appropriate single cell population, live cells not expressing any differentiated hematopoietic lineage marker (eg: CD3 $\epsilon$ , Ly-6G/Ly-6C, CD11b, B220, Ter119) are placed on a gate with Sca-1 and CD117 on individual axes. Cells double positive for these markers are drilled down to as a lower gate, and within this gate cells not expressing CD48, but positive for CD150 are canonical murine LT-HSCs. Erythropoiesis is observed utilizing a similar, but distinct gating logic. After selecting for viable single cells, the erythroid compartment is identified based on expression of CD71, the transferrin receptor, and Ter119, a marker enriched in late-stage erythroblasts. Through this gate, cells are shown with CD44 vs. front scatter, or size. As an erythroblast matures, levels of CD44 drop, along with overall size. Across this gate five different puncta of cell populations can be assessed as individual erythroblast populations, referred to as P1-P5. P1+P2 are proerythroblasts and basophilic erythroblasts, P3 are polychromatic erythroblasts; P4 are orthochromatic erythroblasts and reticulocytes, and lastly P5 are mature RBCs. The presence and relative proportions of these populations can give insight as to the state of erythropoiesis in the organism and reveal potential deviations from steady state. Furthermore, the presence of increased amounts of P1-P4 populations in the spleen can indicate extramedullary hematopoiesis as a response to anemic stress.

## **Complete Blood Counts**

The collective assessment of multiple peripheral blood parameters is referred to as a complete blood count (CBCs) and provides insights into the state of hematopoiesis in the body and the overall health of the organism. Conveniently, unlike immunopathology of the spleen or marrow, CBCs can be measured repeatedly over the life of a mouse model organism. Collected generally via insertion of a heparinized glass capillary into the retroorbital sinus or via puncture of the skin via a sterile lance, CBCs can indicate the number of RBCs in circulation, the quantity of hemoglobin in the blood, the presence of white blood cells, etc. In general, it is good practice to keep the site of blood collection consistent across an experiment to achieve quality results.

## **The Goals of this Dissertation**

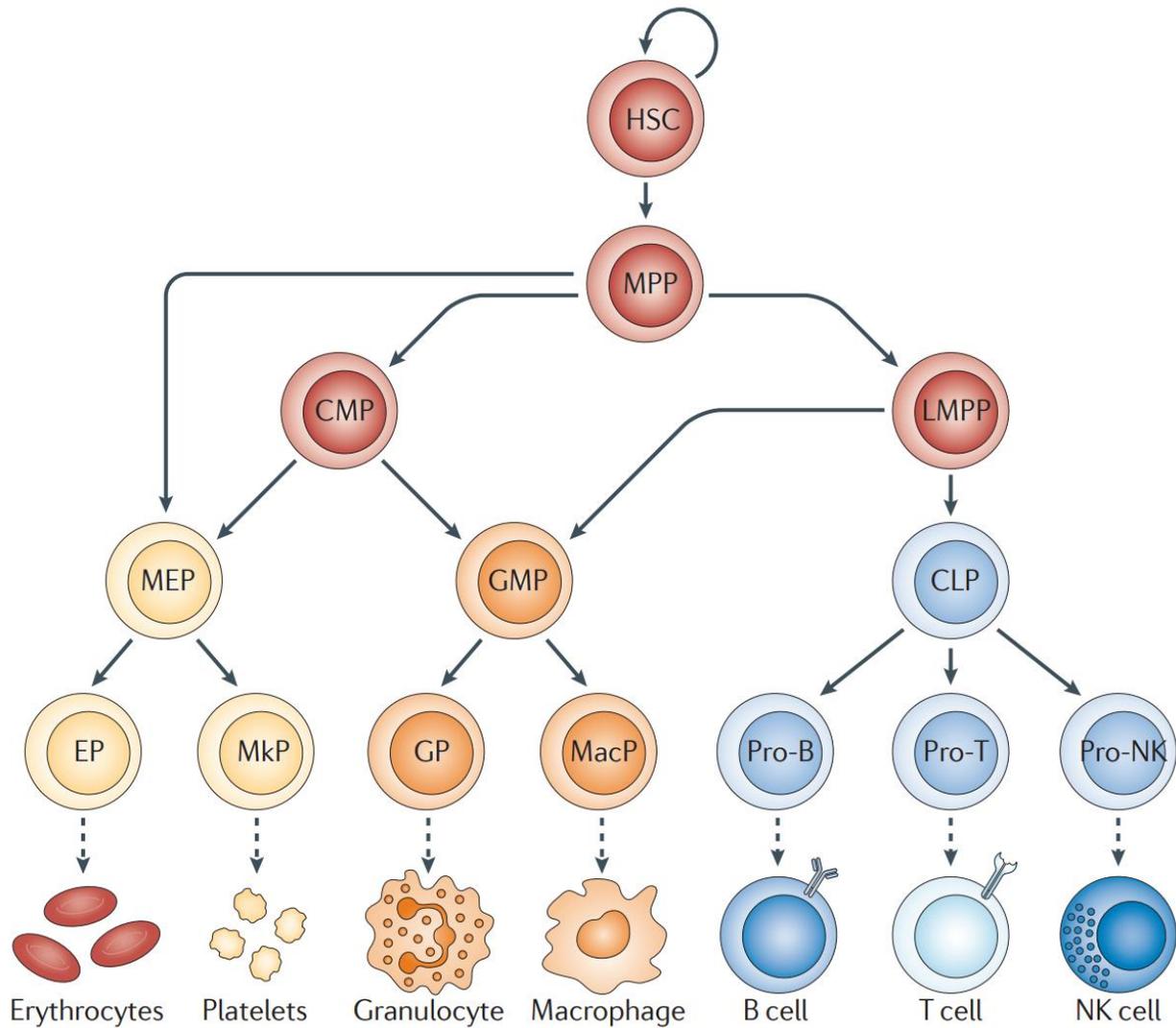
The exploration and treatment of rare diseases is often confounded via the absence of viable animal models, either due to difficulties deploying the disease mechanism in animals or lack of attention from the scientific community. Clearly, developing treatments for rare human conditions can be hamstrung by a lack of experimental models to validate them in a pre-clinical setting. Generally speaking, knock-out and knock-in models have been the experimental standard for decades, but given the enormous differences between human and animal physiology the fact that certain diseases cannot be readily modeled via animal models is not surprising.

The goals setting this dissertation in motion can be divided into two primary experimental aims. In the first portion of chapter two we explore the attempted generation of a mouse model of CDA1a via inducible knock-out of *Cdan1* utilizing a pair of disparate Cre recombination methodologies. We characterize the non-viability of inducible *Cdan1* knock-out

in adult mice. In the second portion of chapter 2 we demonstrate rescue of the *Cdan1* knock-out mouse models via lentiviral gene therapy delivering the human *Cdan1* orthologue. In chapter 3 we discuss the implications of the work moving forward and indicate a potential CDA1a disease mechanism informed in part by artificial intelligence. We further discuss the implications of *Cdan1* as a central regulator of cell fate and division.

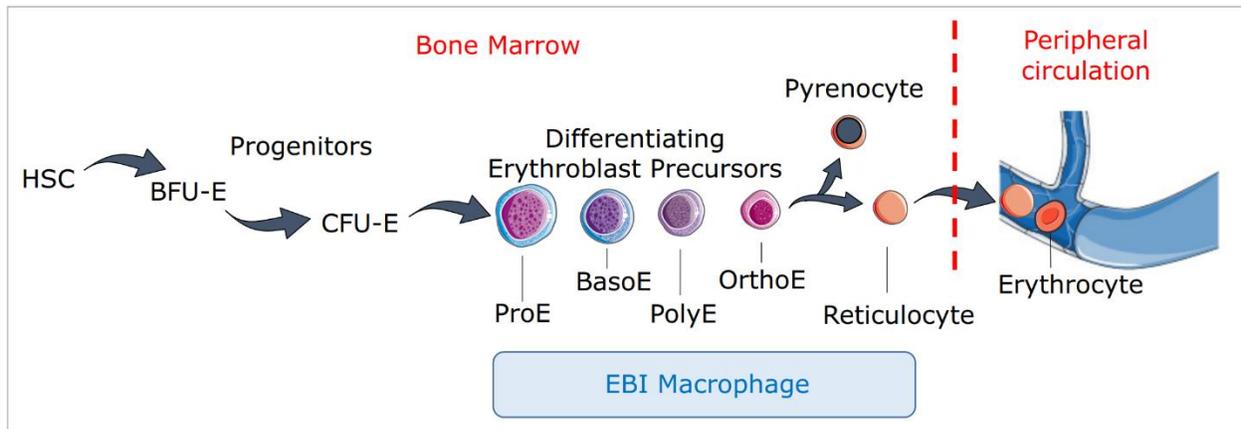
In totality, this dissertation demonstrates the power of gene therapy, along with the insufficiency of knock-out animal models in certain contexts, how animal models can be poor facsimiles of human physiology, and how artificial intelligence is poised to revolutionize fundamental biological research.

## Figures

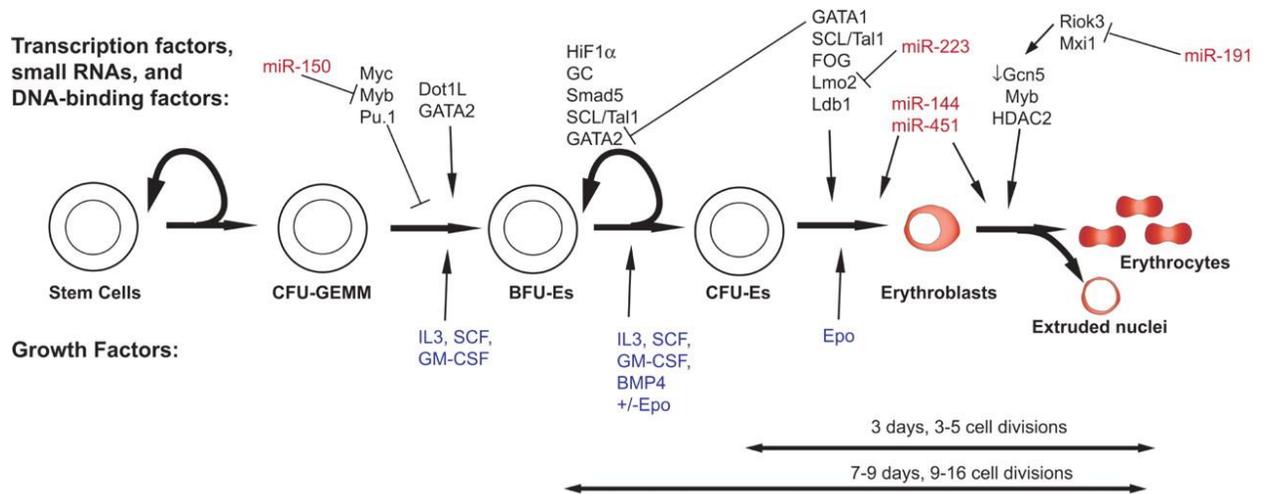


**Figure 1.1. Simplified diagram of hematopoiesis.** The hematopoietic stem cell (HSC) produces daughter cells that differentiate in a progressive manner. As these cells progress, they become progressively restricted in their ability to adopt different lineages, becoming committed to a single lineage type. Further differentiation steps beyond committed progenitors are simplified as dashed lines. Recursive arrow at HSC stage indicates the self-maintenance function of these essential stem cells. HSC: hematopoietic stem cell, MPP: multipotent

progenitor, CMP: common myeloid progenitor, LMPP: lymphoid primed multipotent progenitor, MEP: megakaryocyte-erythrocyte progenitor, GMP: granulocyte-macrophage progenitor, CLP: common lymphoid progenitor, EP: erythrocyte progenitor, MkP: megakaryocyte progenitor, GP: granulocyte progenitor, MacP: macrophage progenitor, Pro-B: B cell committed progenitor, Pro-T: T cell committed progenitor, Pro-NK: natural-killer committed progenitor. Adapted from Cedar and Bergman, 2011<sup>96</sup>

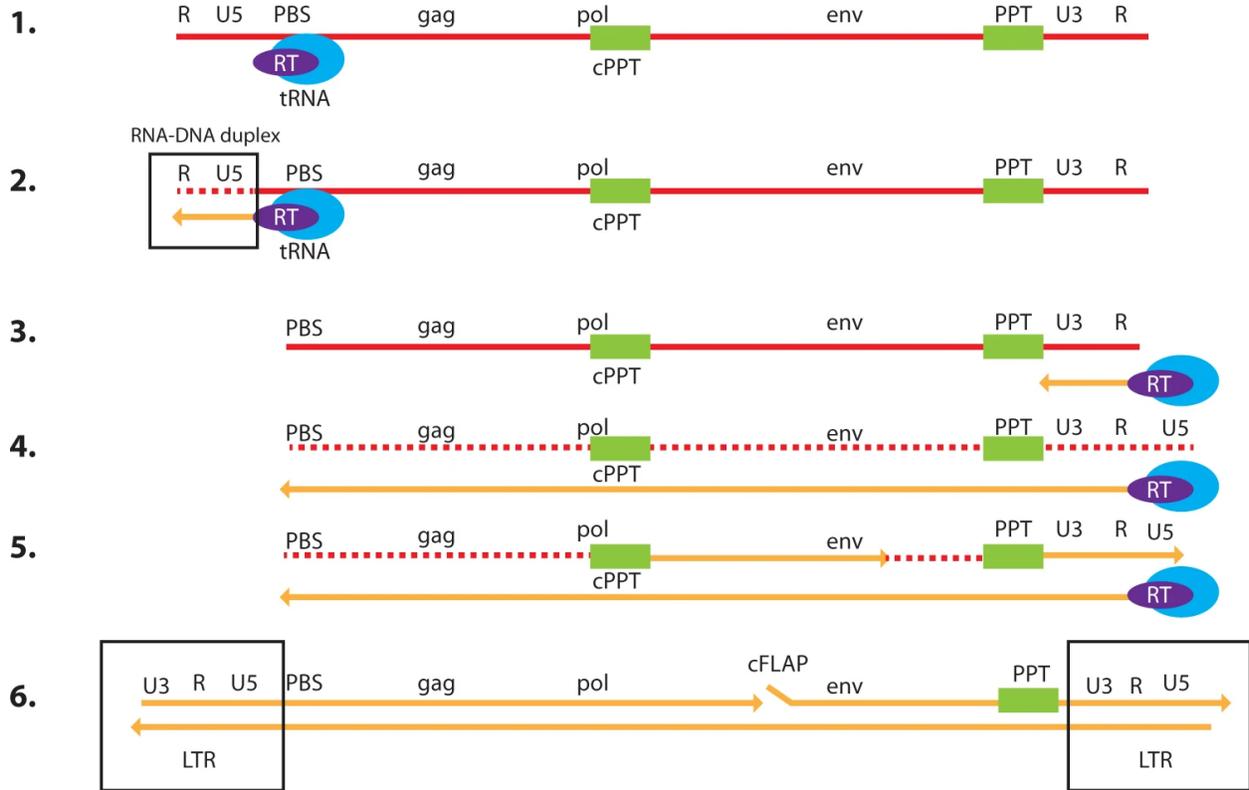


**Figure 1.2. Simplified diagram of erythropoiesis.** Erythropoiesis begins with the hematopoietic stem cell (HSC) then progressively moves through multiple steps of sequential differentiation and expansion starting at BFU-Es, and ending with the mature erythrocyte in circulation. Erythroblasts associate with the erythroblastic island macrophage (EBI) as they mature. Ejection of the nucleus as a pyrenocyte leads to the maturing reticulocyte that undergoes extensive membrane and organelle remodeling to become a mature erythrocyte. BFU-E: burst-forming unit-erythroid, CFU-E: colony-forming unit-erythroid, ProE: proerythroblast, BasoE: basophilic erythroblast, PolyE: polychromatic erythroblast, OrthoE: orthochromatic erythroblast. Adapted from Ginzburg, et al. 2023.<sup>17</sup>



**Figure 1.3. Graphical overview of factors involved in regulation of erythropoiesis.** Multiple signaling molecules, miRNAs, and transcription factors function at each step to induce sequential differentiation programming. Cytokines are indicated in blue, miRNAs are indicated in red, and transcription factors are indicated in black. Horizontal arrows indicate approximate timeline for a given step. Adapted from Hattangadi, et al. 2011.<sup>97</sup>

## Reverse transcription



**Figure 1.4. Graphical representation of steps involved in lentiviral reverse transcription converting the ssDNA genome into dsDNA for integration into the infected cell's genome.**

DNA is indicated in yellow, RNA is indicated in red. **1.** The tRNA primer is based paired to the primer-binding site (PBS). **2.** Initiation of reverse transcription by the reverse transcriptase (RT) indicated as a purple ellipse, begins at the tRNA primer, replicating the 5' end of the genome complete with the U5 and R sequences. This produces an RNA/DNA duplex. RNase-H activity removes the RNA template as the RT progresses. **3.** Transferring the minus-strand to the 3' end, DNA synthesis continues. **4.** Reverse transcription process continues to the remainder of the genome, with purine-rich tracts (PPT) of RNA left intact. **5.** The PPTs act as primers for plus strand DNA synthesis. **6.** Both plus and minus strands of DNA are elongated and completed,

leaving a central flap (cFLAP) which is essential in integrating the genome subsequently.

Adapted from Milone and O'Doherty, 2018.<sup>98</sup>

## CHAPTER 2 – LENTIVIRAL GENE THERAPY PROVIDES ROBUST PROTECTION AGAINST PANCYTOPENIA IN AN INDUCIBLE CODANIN-1 KNOCKOUT MODEL

Chapter 2 is adapted from a manuscript with the same title in preparation by Perry Demsko, Carlo Castruccio Castracani, Lucas Tricoli, Amaliris Guerra, Laura Breda, Pankaj Sharma, Ariel Rivera, Megan Fedorky, Naoto Tanaka, Tyler Papp, Yink Tam, Barbara Mui, Hamideh Parhiz, and Stefano Rivella. I contributed to this work by formulating experiments, performing experiments, soliciting feedback, analyzing data, composing the initial draft, and performing project management.

### Abstract

Congenital Dyserythropoietic Anemia 1a (CDA1a) is a macrocytic anemia caused by biallelic mutations in Codanin-1 (*CDANI*), a putative regulator of chromatin assembly. The present treatment paradigm for CDA1a is limited to blood transfusions, with allogeneic hematopoietic stem cell (HSC) transplant being the only curative option. Because of embryonic lethality in *Cdan1* knockout mice, the lack of a CDA1a model has limited the development of novel therapies. Here, we report an inducible *Cdan1<sup>f/f</sup>* mouse model for the study of CDA1a and the development of gene therapies. Deletion of *Cdan1* in adult animals was achieved via a dsRNA analog, poly(I:C), administered to *Cdan1<sup>f/f</sup>* Mx-Cre mice (*Cdan1<sup>ΔMx</sup>*). Upon induction with poly(I:C), *Cdan1<sup>ΔMx</sup>* animals developed an aplastic bone-marrow phenotype by 7-10 days post-treatment. As an alternative approach, we delivered Cre-encoding mRNA via targeted lipid-nanoparticles (LNPs) to *Cdan1<sup>f/f</sup>* HSCs as part of an *ex vivo* transplant (*Cdan1<sup>ΔLNP</sup>*). *Cdan1<sup>ΔLNP</sup>* HSCs failed to engraft. However, pre-treatment of *Cdan1<sup>ΔLNP</sup>* HSC with a lentiviral vector expressing human *CDANI* under the control of a constitutive promoter led to the complete rescue of *Cdan1<sup>ΔLNP</sup>* animals.

These observations indicate an essential role of *Cdan1* in HSC survival and provide evidence of a potential therapy for patients with CDA1a.

## **Introduction**

The Congenital Dyserythropoietic Anemias (CDAs) are a group of disorders characterized by ineffective erythropoiesis characterized by anomalous erythroid differentiation and replication pathways.<sup>99,100</sup> CDAs represent a large group of mechanistically diverse, hypoplastic anemias divided into several categories: (1) CDA types I-III, (2) transcription factor-related CDAs, and (3) CDA variants.<sup>101</sup> The lack of suitable animal models for CDAs has limited the development of novel treatments, including gene therapies.

Mutations in CDA1a account for approximately 12% of all CDA cases caused by biallelic hypofunctional mutations in *CDAN1*.<sup>102</sup> CDAN1 is thought to negatively regulate chromatin deposition during S-phase, but full elucidation of its functions is not understood.<sup>103</sup> Patients with CDA1a present with moderate to severe macrocytic anemia, insufficient reticulocyte response, and assorted skeletal and limb deformities.<sup>104</sup> Their bone marrow exhibits erythroid hyperplasia, replete with up to 10% erythroblasts with characteristic binucleation or internuclear chromatin and cytoplasmic bridging between cells.<sup>105</sup> Treatment options for patients with CDA1a are very limited, often restricted to blood transfusions and management of iron overload caused by ineffective erythropoiesis and chronic transfusions.<sup>106</sup> Although an interferon (IFN)- $\alpha$ -based treatment regime is available, side effects limit its long-term use as a therapy.<sup>107</sup> Indeed, besides allogeneic bone marrow transplant, there is no cure.

Previous work has shown that global deletion of *Cdan1* is lethal in early embryogenesis by E7 in mice, whereas erythroid-specific deletion via the EpoR-Cre system induces lethality from

anemia at E14.5.<sup>108,109</sup> Critically, EpoR-Cre animals exhibit CDA1a-characteristic spongy chromatin with binucleated erythroblasts.<sup>109</sup> However, this model does not provide information about the potential role of *Cdan1* in other hematopoietic lineages. Given the embryonic lethality of congenital *Cdan1* deletion, the limited understanding of its role in developmental and adult hematopoiesis, and the absence of therapies for patients, there is a critical need for preclinical models of CDA1a. We hypothesized that the inducible, conditional deletion of *Cdan1* in adult animals would allow us to generate viable mice with CDA1a-like disease, thereby facilitating the development of novel therapies.

Here, we report two related *Cdan1*<sup>fl/fl</sup> transgenic mouse models that allow the conditional disruption of *Cdan1* in adult animals. In the first model (*Cdan1*<sup>ΔMx</sup>), we show that the deletion of *Cdan1* in transgenic Mx1-Cre *Cdan1*<sup>fl/fl</sup> mice can be induced by poly(I:C), a dsRNA analogue that activates Cre downstream of the *Mx1* promoter upon induction of interferon (IFN) signaling.<sup>110</sup> As a second model (*Cdan1*<sup>ΔLNP</sup>), we utilize LNP-based delivery of Cre mRNA to induce the deletion into HSC of *Cdan1* *ex vivo*, followed by transplantation into adult lethally irradiated mice. Similar systems have been previously deployed by our group to test other congenital blood disorders, including lethal α-thalassemia and X-linked sideroblastic anemia.<sup>111,112</sup> In this work, we identify the essential requirement for CDAN1 in HSC survival and the ability of a lentiviral vector to rescue *Cdan1* deficiency, providing a path toward gene therapy in patients with CDA1a.<sup>90</sup>

## Results

***Inducible Cdan1*<sup>ΔMx</sup> knockout mice develop pancytopenia.** Given the embryonic lethal nature of prior models of *Cdan1* deficiency, including germline and *in utero* deletion of *Cdan1*<sup>108,109</sup>, we generated a *Cdan1*<sup>fl/fl</sup> transgenic line with LoxP sites flanking exons 7-10 of the *Cdan1* gene

(**Supplemental Figure 2.1A**). *Cdan1*<sup>fl/fl</sup> animals were then crossed with transgenic Mx1-Cre mice to generate our first inducible deletion model (*Cdan1*<sup>ΔMx</sup>). Induction of Mx1-Cre expression was achieved upon administration of the dsRNA analogue poly(I:C), which activates Toll-like receptor 3 and induces type I IFN signaling.<sup>113</sup> This, in turn, stimulates the *Mx1* promoter to drive Cre-recombinase expression, leading to the deletion of the floxed *Cdan1* allele. Without exception, Poly(I:C)-treated *Cdan1*<sup>ΔMx</sup> mice required euthanasia between days 10 and 21 post-treatment (**Figure 2.1B**) due to severe anemia manifested by diminished red blood cell (RBC) and low hematocrit (HTC, **Figure 2.1C-G**). This was consistent with failure to generate multilineage hematopoietic cells, including white blood cells (WBC), neutrophils, and reticulocytes 7-9 days post-induction (**Figure 2.1E-G**). Neutrophils have a short half-life of less than 24 hours in mice, therefore neutrophil deficiency in such a short time span as demonstrated in **Figure 2.1E** can be predictive of hematopoietic failure.<sup>114</sup>

We characterized erythropoiesis in mice, 7-9 days post poly(I:C) administration using previously established markers<sup>115</sup>. We found diminished erythropoiesis in the bone marrow, characterized by the lack of CD71<sup>+</sup>/Ter119<sup>+</sup> nucleated progenitor erythroid cells (**Figure 2.2A**; top right quadrant) and persistence of CD71<sup>-</sup>/Ter119<sup>+</sup> mature, mostly enucleated, erythroid cells or RBC (lower quadrant), which are less affected by poly(I:C). Lack of erythroid progenitors was further characterized using an approach that dissects the different stages of erythroid maturation (I-V) by CD44 and size within the TER119<sup>+</sup> population in **Figure 2.2B**. In the BM white cell compartment, we also detected a severe depletion of long-term HSC (Lin<sup>-</sup>cKit<sup>+</sup>Sca1<sup>+</sup>CD150<sup>+</sup>CD48<sup>-</sup>), in the *Cdan1*<sup>ΔMx</sup> mice treated with Poly(I:C) (**Figure 2.3A-B**). The decreased number of granulocytes, identified as Gr1<sup>+</sup> within the CD45<sup>+</sup> population, further

confirmed the pancytopenia at the FACS analysis (**Figure 2.3C**). Moreover, the histopathological assessment revealed hypocellular marrow and gross aplasia (**Figure 2.3D**).

***Lentiviral gene therapy with human CDAN1 rescues  $Cdan1^{ALNP}$  knockout mice from non-engraftment and sustains long-term hematopoiesis.*** In  $Cdan1^{AMx}$  animals, the deletion of  $Cdan1$  may not be limited to the hematopoietic compartment since IFN responsiveness is well known to induce  $Mx1$  in various cell types.<sup>116-118</sup> Therefore, to generate a hematopoietic-specific model of  $Cdan1$  depletion, we deployed lipid nanoparticles (LNPs) chemically ligated to the surface of the mRNA-LNP (LNP<sup>CD117Cre</sup>) to target the c-Kit receptor in HSC<sup>119</sup> and deliver Cre-encoding mRNA *in vitro*. Our group has previously described this system to induce efficient and preferential targeting of HSCs *ex vivo* and *in vivo*.<sup>120</sup> We hypothesize that deleting  $Cdan1$  *ex vivo* followed by engraftment of  $Cdan1$ -KO HSC would provide further insights on the role of  $Cdan1$  on hematopoiesis. To that end, we isolated  $Cdan1^{fl/fl}$   $lin^-$  bone marrow cells and transfected them with LNP<sup>CD117Cre</sup> in cell culture, followed by the transplantation of these cells into congenic, lethally irradiated recipients, which we refer to as  $Cdan1^{ALNP}$  knockout mice (**Figure 2.4A**).

The lethality phenotype of the  $Cdan1^{ALNP}$  knockout animals mirrored that of lethally irradiated, non-transplanted placebo control mice, suggesting failure of engraftment by  $Cdan1$ -deleted  $lin^-$  cells (**Figure 2.4B**), and in line with the phenotype in the  $Cdan1^{AMx}$  knockout mice. We reasoned that this model could serve as a platform to develop a lentiviral  $CDAN1$  gene therapy. We generated a lentiviral vector encoding human  $CDAN1$  under the control of the constitutively active phosphoglycerate kinase (PGK) promoter, supplemented with an ankyrin insulator (AP) as well as the foamy virus insulator and posttranscriptional regulatory elements (FWP), hereafter referred to as AP-CDAN1-FWP (**Supplemental Figure 2.1B**).<sup>121-125</sup> Remarkably,  $Cdan1^{ALNP}$  animals transplanted with  $lin^-$  cells pre-treated with AP-CDAN1-FWP at a 25 MOI exhibited full

bone marrow reconstitution, in line with that of lethally irradiated *Cdan1<sup>fl/fl</sup>* transplanted controls. AP-CDAN1-FWP treated *Cdan1<sup>ALNP</sup>* mice maintained normal CBC values through the 16-week post-transplant follow-up period, consistent with normal steady-state erythropoiesis (**Figures 2.4C**) and concomitant high donor chimerism (**Supplemental Figure 2.1C**).

We confirmed the deletion of endogenous *Cdan1* alleles, as well as vector copy number (VCN), as a surrogate of transgenic *Cdan1* allelic levels, in gDNA isolated from peripheral nucleated blood cells, using droplet digital PCR (ddPCR) (**Figure 2.4D-F, Supplemental Figure 2.1A**).<sup>126</sup> The probes used to assess the endogenous levels of *Cdan1* were specific for an intronic sequence not present in the lentiviral vector, thereby ensuring faithful accounting of both endogenous *Cdan1* and transgenic levels. Endogenous donor *Cdan1* levels approached the lowest detection limit (**Figures 2.4D-E**), while AP-CDAN1-FWP VCN stabilized at approximately two copies/genome at six months post-transplant (**Figure 2.4F**). Recombination was further validated using end-point PCR (**Supplemental Figure 2.1D**).

To further investigate the lack of viability of *Cdan1<sup>ALNP</sup>* HSC, a colony formation assay was performed.<sup>127</sup> We observed severe impairment in the ability of *Cdan1<sup>ALNP</sup>* *lin*<sup>-</sup> cells to generate colonies and proliferate. By contrast, Cre-deleted *Cdan1<sup>ALNP</sup>* cells treated with AP-CDAN1-FWP expanded and produced suitable colonies, albeit at lower levels than non-deleted *Cdan1<sup>fl/fl</sup>* control cells (**Figure 2.4G**), likely because not all the KO colonies carried an integrated vector. These results are consistent with the hypothesis that deletion of *Cdan1* in HSCs limits their capacity to survive and proliferate. Furthermore, these results confirm that expression of human *CDAN1* can at least partially rescue homozygous deficiency of murine *Cdan1* in HSCs.

***Human CDAN1 lentiviral gene therapy facilitates reconstitution of hematopoiesis by *Cdan1<sup>ALNP</sup>* knockout HSCs.*** Lentiviral gene therapy-treated *Cdan1<sup>ALNP</sup>* mice were followed for six months

post-transplant and then euthanized to assess hematopoietic reconstitution. CBC parameters taken immediately prior to euthanasia in animals treated with gene therapy were similar to those of non-deleted, transplanted *Cdan1<sup>ΔLNP</sup>* healthy control mice (**Figure 2.5**). To determine the self-renewing capacity of transduced *Cdan1<sup>ΔLNP</sup>* LT-HSCs, lethally irradiated recipients were transplanted with marrow from primary chimeras, setting up secondary transplants.<sup>128</sup> Secondary chimeras successfully engrafted and showed comparable CBC values at six months (**Figure 2.5**). Indeed, donor chimerism in these mice averages  $95.0\% \pm 0.45$ , comparable with secondary mice infused with non-deleted, *Cdan1<sup>ΔLNP</sup>* primary marrow ( $95.3\% \pm 0.79$ ) (**Supplemental Figure 2.2A**). Deletion levels of endogenous *Cdan1* levels were conserved and remained close to the lowest detection limit (**Supplemental Figure 2.2B**). In secondary transplanted animals, the lentiviral cassette (AP-CDAN1-FWP) was detected at approximately one copy per genome, compared to two copies per genome in primary transplantation (**Supplemental Figure 2.2C-D**).

***Expression of human CDAN1 in *Cdan1ΔLNP* knockout cells is associated with mild abnormalities in erythropoiesis.*** Gene therapy treatment of *Cdan1<sup>ΔLNP</sup>* knockout hematopoietic cells led to the reconstitution of erythropoiesis in mice (**Figure 2.6A-B**). However, analysis of the bone marrow showed minor differences, with a slight increase in the erythrocyte population (V) compared to non-deleted, transplanted control animals (**Figure 2.6A-B**). Considering the differences in the marrow-based erythropoiesis, we analyzed erythropoiesis in the spleen, which showed concomitant higher proportion of erythrocytes (V) and a mildly lower proportion of orthochromatic cells and reticulocytes (IV). Altogether, these observations suggest an increased ability of erythroid progenitor cells to survive and differentiate into enucleated RBC without affecting the overall production of terminally differentiated RBC, as shown by peripheral blood values. LSK and LT-HSC populations of gene therapy-treated mice were found to be

interchangeable with non-deleted, transplanted controls, demonstrating that *CDANI* gene complementation can rescue the hematopoietic stem cell compartment and hematopoiesis in *Cdan1*-deficient mice (**Figures 2.7A-B and E**).<sup>129</sup> White blood cell populations, including granulocytes and macrophages, were also successfully reconstituted (**Figures 2.7C-D and E**), with macrophages showing minimal differences comparing healthy control and treated mice.

To further assess cellular morphology and iron buildup, we performed histopathology on relevant organs, including bone marrow and liver. No iron buildup was observed in the liver (**Supplemental Figure 2.3A**). However, this was somewhat anticipated since iron overload in patients with CDA1a is often secondary to transfusions, especially in fetal-onset CDA1a.<sup>130,131</sup> Bone marrow morphological and hematological characteristics in the treated and healthy cohorts showed no appreciable differences (**Figure 2.8**). While liver weights in gene-complemented animals were comparable to those of non-deleted transplanted control animals (**Supplemental Figure 2.3B**), we observed slight splenic hypoplasia in gene-complemented animals, corroborating the mild increase in enucleated RBC observed in the bone marrow and spleen (**Supplemental Figure 2.3C**). Peripheral blood smears were indistinguishable between deleted and control transplanted animals (**Supplemental Figure 2.3D**).<sup>108</sup> Thus, *Cdan1* deletion in the hematopoietic compartment is associated with lethality in mice, and this can be rescued upon human *CDANI* gene lentiviral delivery.

## Discussion

We found that inducible deletion of *Cdan1* in adult mice is incompatible with life, either due to total loss of endogenous HSC function in our poly(I:C) treated *Cdan1*<sup>AMx</sup> mice, or because

of bone marrow engraftment failure in *Cdan1*<sup>ALNP</sup> animals. Our work expands upon the understanding of *Cdan1* as an essential gene in hematopoiesis. Previous attempts to generate models of the human disease CDA1a via global or *in utero* genetic deletion of *Cdan1* resulted in embryonic lethality.<sup>108,109</sup> Unlike CDA1a in humans, which is associated with anemia and hepatosplenomegaly<sup>104,105</sup>, we found that the inducible deletion of *Cdan1* in adult mice caused particularly severe disease manifested by pancytopenia and lethality. Thus, although our mouse models of *Cdan1* deficiency do not faithfully model CDA1a in humans and have a phenotype not observed in humans due to presumed *in utero* lethality<sup>101</sup>, the system still permitted the development and testing of a lentiviral-based gene therapy that may become useful in treating patients with CDA1a. Remarkably, expression of human *CDANI* in hematopoietic cells led to functional rescue of murine *Cdan1* deficiency, demonstrating that the human gene is functional in mice.

We observed death associated with hematopoietic failure in inducible Mx1-Cre *Cdan1*<sup>fl/fl</sup> poly(I:C) treated animals. Since Mx1-Cre expression can be induced in all IFN-responsive cell types, and *Cdan1* is ubiquitously expressed, deletion of *Cdan1* in the *Cdan1*<sup>ΔMx</sup> model may have occurred both within and outside the hematopoietic compartment.<sup>132</sup> As *Cdan1* is ubiquitously expressed,<sup>132</sup> it stands to reason that other cell types would have undergone deletion of *Cdan1*, and that this may have impacted non-hematopoietic proliferating cell types since *Cdan1* is thought to be expressed primarily during the S phase.<sup>133</sup> Therefore, we cannot exclude the possibility that the lethality phenotype in the *Cdan1*<sup>ΔMx</sup> model is also related to the deletion of *Cdan1* outside the hematopoietic compartment. Furthermore, the inflammation triggered by poly(I:C) might have had additional effects in mice compromised by widespread deletion of *Cdan1*. For example, the gut

epithelium simultaneously undergoes rapid cellular turnover while being tightly regulated by IFN signaling.<sup>134,135</sup>

Furthermore, a genome-wide association study (GWAS) showed that human *CDANI* has a role in maintaining cardiac function, possibly by regulating cellular turnover,<sup>136</sup> but we did not evaluate the heart conditions in this mouse model. In the future, cardiac-, epithelial- and other tissue-specific deletions of *Cdan1* could be explored to define functions of *Cdan1* outside the hematopoietic compartment. Nevertheless, our *Cdan1*<sup>ΔLNP</sup> model confirmed that hematopoietic expression of *Cdan1* is essential for hematopoietic colony formation in vitro and for homing and multilineage reconstitution of the bone marrow and peripheral blood in lethally irradiated animals.

Given the important role of *Cdan1* in chromatin regulation,<sup>137</sup> the erythroid lineage appears to be uniquely sensitive to loss of *Cdan1*, likely due to the intricate sequence of chromatin remodeling that occurs in rapidly dividing cells that are simultaneously undergoing proliferation and differentiation, all the while compacting their nucleus to prepare it for expulsion.<sup>138-141</sup> The *Cdan1*<sup>ΔLNP</sup> model extends the study of *Cdan1* beyond erythropoiesis by more broadly demonstrating the essential nature of *Cdan1* in HSC. The *Cdan1*<sup>ΔLNP</sup> model achieves this by segregating *Cdan1* deletion to hematopoietic cells exclusively, without the confounding variable of IFN activation in other cells. Thus, the *Cdan1*<sup>ΔLNP</sup> model ensures that the effects observed are limited to the hematopoietic compartment and unrelated to the disruption of other cells in the bone marrow microenvironment.<sup>142,143</sup> Given the recursive rounds of asymmetrical differentiation and division that are undertaken to fully recover hematopoiesis, it stands to reason that the removal of an essential chromatin regulator such as *Cdan1* would affect other cell types beyond erythrocytes. However, before this study, that had not been shown.<sup>144,145</sup>

One limitation of our study is the expression of the human CDAN1 protein in mouse hematopoietic cells in the *ex vivo* gene therapy model. This may explain the slight variation in erythrocytes and macrophage (**Figure 6**), whose increased number may be triggered by the slight increase in enucleated erythroid cells. Future studies using *CDANI* human HSC cells *in vitro* and *in vivo* will address whether these differences are species-specific or triggered by the vector design.

Another limitation of our study is that *Cdan1* deletion does not fully mimic the clinical disease phenotypes observed in patients with CDA1a. Given the lack of biallelic-null *CDANI* patients in the medical literature and the essential nature of *Cdan1* based on CRISPR-based knockout library screens in cancer cells, this is perhaps unsurprising.<sup>146</sup> Indeed, we found that the inducible deletion of *Cdan1* remarkably affected HSC, in contrast to what has been described in patients with CDA1a, who exhibit more limited disease phenotypes due to hypomorphic, disease-causing *CDANI* variants rather than complete loss of *CDANI*. Future studies may model hypomorphic *Cdan1* variants in animals to mimic the human disease more accurately.

Nevertheless, the potential implications of our animal model for developing therapies are far-reaching. We demonstrated the ability of human *CDANI*, delivered via a lentiviral vector, to rescue engraftment of *Cdan1* knockout bone marrow, as well as the ability of human *CDANI* to sustain hematopoiesis through secondary transplantation, directly informing the ability of this gene therapy to maintain marrow health and HSC stemness in the context of murine *Cdan1* deficiency. It is conceivable that while the human *CDANI*-expressing animals appear to undergo steady-state erythropoiesis, there may be more subtle anomalies, possibly due to the expression of the human gene in mouse hematopoietic cells.<sup>147,148</sup> This will require further investigation. Additionally, we found that only a single copy of the human *CDANI* gene is required to prevent the disease, consistent with a lack of dominant-negative *CDANI* mutations in humans.<sup>101</sup> Taken together, these

findings directly inform the potential of a gene complementation-based strategy to treat CDA1a using a lentiviral vector.

Our study illustrates differences between murine and human erythropoiesis hinging on the *Cdan1* axis. The AP-CDAN1-FWP vector, while rescuing mice from lethality, still led to minor deviations from steady-state erythropoiesis. The CDAN1 protein is a known interactor of ASF1a and ASF1b, canonical histone chaperones.<sup>103</sup> In our model system where human CDAN1 replaces its orthologue in mice, mild dysregulation of erythropoiesis might be due to CDAN1-mediated effects on murine ASF1a and ASF1b<sup>149</sup>.

Continuation of this study would entail further exploration of the molecular mechanisms of *Cdan1* function in adult animals during lineage-specific hematopoiesis, as well as outside the hematopoietic compartment, using the models here described. Chromosome conformation capture<sup>150</sup> experiments may also help clarify functions of *Cdan1* during hematopoiesis, as chromatin topology is essential for regulating cell fate, especially since *Cdan1* appears strategically located on the chromatin regulatory axis.<sup>151</sup> Finally, we believe that our findings, including human CDAN1-complementation therapy, could potentially benefit patients suffering from CDA1a.

**Acknowledgments:** This project has been supported by The National Institute of Diabetes and Digestive and Kidney Diseases Institute of the National Institutes of Health (R01 DK133475, R01 DK095112), Grant Penn-RNA “LNP technology to treat and cure hematological disorders”, Grant IRM-UPenn “Use of in vivo mRNA delivery for the cure of severe combined immunodeficiencies (SCID), The Sickle Cell and Red Cell Disorders Curative Therapy Center (CuRED) & Molecular Therapies for Inborn Errors of Metabolism-Frontier Program awarded to S.R. Dr. Peter Kurre

(Children's Hospital of Philadelphia) and Dr. Osheiza Abdulmalik (Children's Hospital of Philadelphia) provided the use of essential equipment. The Children's Hospital of Philadelphia Translational Core Lab and the University of Pennsylvania School of Veterinary Medicine Pathology core provided key technical or core facility support. Pathology scoring was performed by Dr. Julie Feldstein (Histowiz). Figures in the paper were generated with GraphPad Prism and Biorender.com.

## **Methods**

***Creation and validation of  $Cdan1^{fl/fl}$  Mice.***  $Cdan1^{fl/fl}$  Mice of a C57BL/6J CD45.2 background were produced by Cyagen (Cyagen, Santa Clara, CA, USA) via CRISPR/Cas9 facilitated insertion of LoxP sites flanking exons 7 and 10 of the endogenous mouse  $Cdan1$  gene. The presence of the LoxP sites was validated via PCR (Forward: GACAACCTAGTTGATTCTGGCTACA, Reverse: GGGTTTGTCTTCTAATTTGGCAGAT) per Cyagen instructions.

***Induction of  $Cdan1$  Deletion in  $Cdan1^{AMx}$  Animals via poly(I:C).*** Animals were injected intraperitoneally with 400  $\mu$ g poly(I:C) (GE Healthcare Life Sciences, Chicago, IL, USA) once every other day for a total of three injections. First injection counts as day 1.

***Complete blood count analysis.*** CBCs were processed via a Sysmex XT-2000iV (Sysmex, Kobe, Hyogo, Japan) at the Children's Hospital of Philadelphia Translational Core Lab.

***Ex vivo Bone Marrow Treatment and Transplantation.*** BM cells were extracted from the long bones, pelvis, and sternum of 8-12 week old  $Cdan1^{fl/fl}$  animals via mortar and pestle within a sterile hood and resuspended in PBS (Gibco, Waltham, MA, USA) supplemented with 4% heat-inactivated fetal bovine serum (Hyclone, Logan, UT, USA). Red blood cells were lysed utilizing ACK Lysis buffer (Gibco, Waltham, MA, USA), filtered through a 40  $\mu$ m strainer (Corning, Glendale, AZ, USA), then counted and viability assessed by AOPI staining in conjunction with a

Cellometer Auto 2000 (Nexcelom Bioscience, Lawrence, MA, USA). Lin<sup>-</sup> cells were enriched via immunomagnetic separation using a mouse lineage depletion kit (Miltenyi Biotec, Auburn, CA, USA) and seeded at a 1.5x10<sup>6</sup>/mL concentration in DMEM (Cellgro, Manassas, VA) supplemented with 50 ng/mL mSCF, 6 ng/mL mL-3, and 10 ng/mL mL-6 (PeproTech, Rocky Hill, NJ, USA) and 1% penicillin/streptomycin. For lentiviral transduction Lin<sup>-</sup> cells were incubated overnight at a multiplicity of infection of 25 with the preceding media supplemented with 2 µl/mL Lentiblast Premium (OZ Biosciences, San Diego, CA, USA). 18-hours post transduction LNP<sup>CD117</sup>Cre was added to the media at 1 µg/1x10<sup>6</sup> cells depending on experimental condition. 6-hours post LNP addition, Lin<sup>-</sup> cells were washed with PBS then injected retro-orbitally into lethally-irradiated 8-12 week old Pepc<sup>b</sup>/BoyJ CD45.1 recipient mice (The Jackson Laboratory, Bar Harbor, ME, USA) at 250,000 cells/animal. Recipient animals were irradiated with 2 doses of 550 cGy separated by 4 hours with an X-Rad320 Irradiator (Precision, Madison, CT, USA). For secondary transplants, BM cells were isolated as previously, but not Lin<sup>-</sup> enriched and injected at 1.0x10<sup>7</sup> cells/animal immediately upon isolation.

***Vector Production and Titration.*** The AP-CDAN1-FWP viral vector was generated utilizing cotransfection of 3<sup>rd</sup>-generation lentiviral vector plasmids consisting of the AP-CDAN1-FWP transfer vector, envelope (VSV-G), core lentiviral machinery (pMDLg/pRRE), and genome packaging machinery (pRSV-REV) into HEK-293T cells. 5x10<sup>6</sup> 293T cells are seeded into 10cm culture dishes 24 hours prior to transduction in Dulbecco's Modified Eagle Medium (DMEM, Cellgro, Manassas, VA, USA) supplemented with 10% heat-inactivated fetal bovine serum, and 1% Penicillin/Streptomycin at 37°C and 5% CO<sub>2</sub>. Following 1,000x concentration by ultracentrifugation, serial dilutions of viral media (4, 0.4, 0.04 µl) are deployed to infect 5x10<sup>4</sup> NIH 3T3 murine fibroblasts (ATCC, Manassas, VA, USA) in 1ml of transfection media (IMDM,

supplemented with 10% heat inactivated fetal bovine serum, 1% Penicillin/Streptomycin, and 1% polybrene). After an overnight incubation the transfection media is changed to maintenance media and four days after infection gDNA extracted (Qiagen DNeasy blood and tissue kit, Valencia, CA, USA). Multiplicity of infection was determined via a ddPCR assay specific to the lentiviral Psi sequence and the following formula: Number of cells x dilution factor x VCN.

***Droplet Digital PCR (ddPCR) for analysis of VCN and Cdan1 Deletion.*** Droplet Digital PCR was performed with a BioRad QX200 Droplet Digital PCR System (BioRad, Hercules, CA, USA). gDNA was extracted from peripheral blood using a Qiagen DNeasy Kit (Qiagen, Valencia, CA, USA). For VCN a primer/probe set specific for lentiviral Psi sequence was utilized (Forward: ACCTGAAGCGAAAGGGAAAC, Reverse: CGCACCCATCTCTCCTTCT, Probe: FAM-AGCTCTCTCGACGCAGGACTCGGC). For Cdan1 levels a primer/probe set specific for the sequence flanked by the LoxP sites was deployed (Forward: TGCATCCCTTGTCTCCTTTCTACTG, Reverse: AGCTGTTTTTGGAAAAGGCGGACAA, Probe: FAM-ATGGGTCAACTTTCAGCAGCCTGCAGC). Both assays used PCBP2 as an endogenous control (Forward: CCAGTCTGCTTGGCATGAAA, Reverse: GGTACCCTTAGCAGCAGACA, Probe: HEX-CCCATCCCTCTCCTGGCTCTAA).

***PCR for analysis of Cdan1 Deletion.*** A primer set flanking the LoxP sites in the *Cdan1* gene was used to validate deletion of the intervening sequenced upon recombination (Forward: GGTTTGGTCGTAAGCAAGGATAC, Reverse: GG AACTCAAATGGGAGGACCAG).

***RNA Synthesis and preparation of LNP<sup>CD117</sup>Cre.*** Codon-optimized Cre recombinase sequence was cloned into an IVT-mRNA production plasmid utilizing a T7 promoter, 5' and 3' UTR, Kozak sequence, and 101 poly(a) tail elements. IVT produced mRNA was produced using a linearized

IVT template and the MEGAScript T7 kit (Thermo Fisher Scientific, Waltham, MA, USA) and decorated with the m<sup>1</sup>Ψ-5'-triphosphate (TriLink, N-1081, San Diego, CA) nucleoside. The IVT mRNAs were capped co-transcriptionally using the trinucleotide cap1 analog (Clean Cap, TriLink, N-7413, San Diego, CA, USA). Cellulose-purified RNAs were encapsulated in self-assembling LNPs. The LNPs used in this study are proprietary to Acuitas Therapeutics (Vancouver, BC, Canada) and described in patent US10,221,127. To target the LNP-mRNA carrier, the LNP was modified using a post-insertion technique to add maleimide functional groups, and reacted with sulfhydryl group modified rat anti-mouse CD117 antibody (Clone 2B8, BioLegend, San Diego, CA, USA). These two components were conjugated together to form the LNP<sup>CD117</sup>Cre vehicle, as described previously.

**Flow cytometry.** Flow cytometry analysis was performed on a BD FACSCanto (BD Biosciences, Franklin Lakes, NJ, USA). Erythroid populations made use of PE Anti-Human/Mouse CD71 (12-0711-82) and PE/Cy7 Anti-Mouse Ter119 (25-5921-82) (ThermoFisher, Waltham, MA, USA), and APC Anti-Human/Mouse CD44 (103012) (Biolegend, San Diego, CA, USA). Chimerism was assessed via PE Anti-Mouse CD45.2 (12-0454-82) (ThermoFisher, Waltham, MA, USA) and APC Anti-Mouse CD45.1 (110714) (Biolegend, San Diego, CA, USA). HSC population made use of CD150 PeCy5, Lin+ PE/Cy7, CD48 APC, Sca1 PE, CD117 APC/Cy7. Lymphoid population made use of TerB FITC, CD3 PE, B220 APC. Macrophage and granulocyte populations made use of CD11b FITC, F480 PE, Gr1 PE/Cy7, CD45 APC. Live/dead staining was accomplished with 7-AAD, where applicable (BD Pharmagen).

**Colony Formation Assay.** 1,000 Lin<sup>-</sup> cells were plated in Methocult M3434 (STEMCELL Technologies, Vancouver, BC, Canada) media supplemented with 1% penicillin/streptomycin within a SmartDish meniscus-free 6-well plate (STEMCELL Technologies, Vancouver, BC,

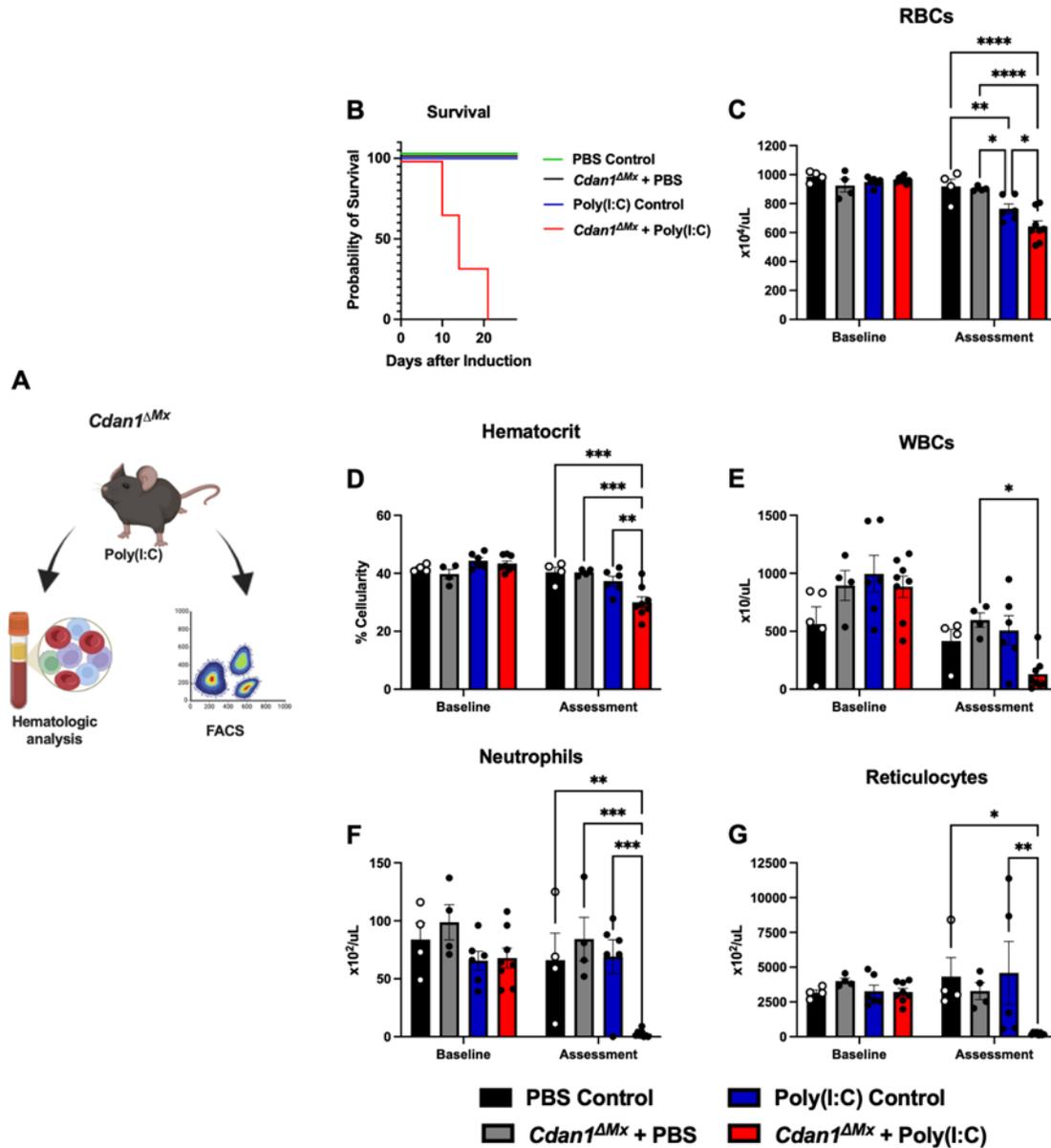
Canada), then imaged and analyzed at 1 week post plating on a STEMvision automated colony counter (STEMCELL Technologies, Vancouver, BC, Canada).

***RAL Diff-Quik stain of peripheral blood.*** 2  $\mu$ L of peripheral blood was smeared on a standard glass microscope slide, then air dried. Slides were then fixed, then stained in Diff-Quik solution I, II, and II for five seconds each, washed with distilled water, then air dried according to manufacturer instructions utilizing the RAL Diff-Quik kit (Siemens, Munich, Germany).

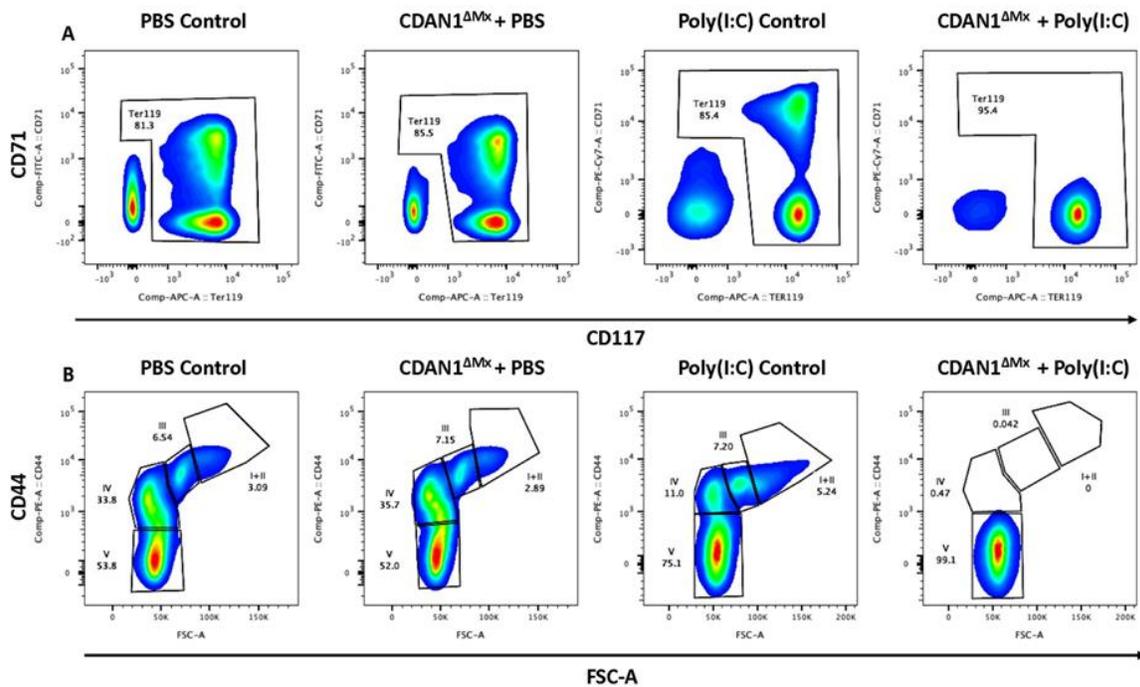
***Institutional Animal Care and Use Committee (IACUC) Regulation.*** The Children's Hospital of Philadelphia IACUC approved all experiments performed on mice in this study (Protocol #1173).

***Data and code availability.*** Data that support the findings of this study are available from the corresponding author upon request.

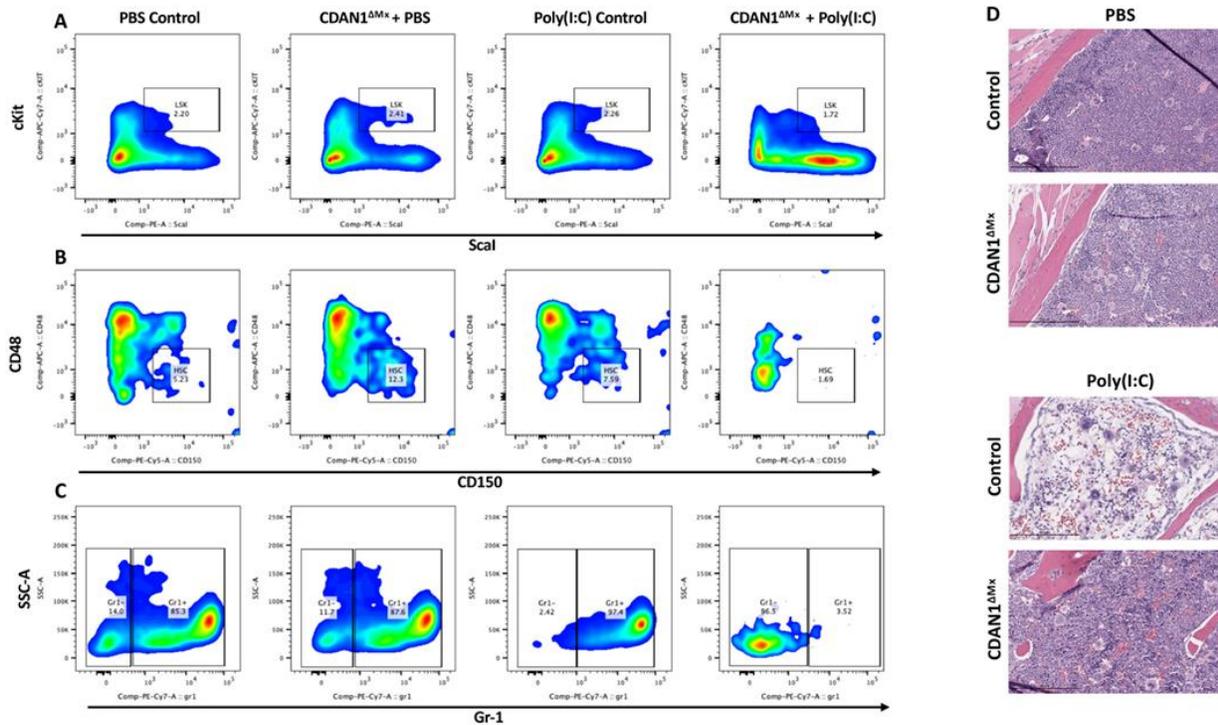
## Figures



**Figure 2.1.** The impact of *Cdan1* deletion on survival and steady state hematopoiesis. (A) Experimental schematic of *Cdan1*<sup>ΔMx</sup> mice injected with poly(I:C) or PBS. (B) Survival curve for wild-type and *Cdan1*<sup>ΔMx</sup> animals treated with PBS and poly (I:C). N=4 per condition. (C-G) CBC analysis of *Cdan1*<sup>ΔMx</sup> animals taken at 7 days before poly(I:C) injection (baseline) and 7-9 days post injection (assessment). N=4-7 per condition. Data were expressed as mean ± SEM and analyzed via Tukey's multiple comparison test after one-way ANOVA, \*p< 0.05, \*\*p<0.01, \*\*\*p<0.001, \*\*\*\*p<0.0001.

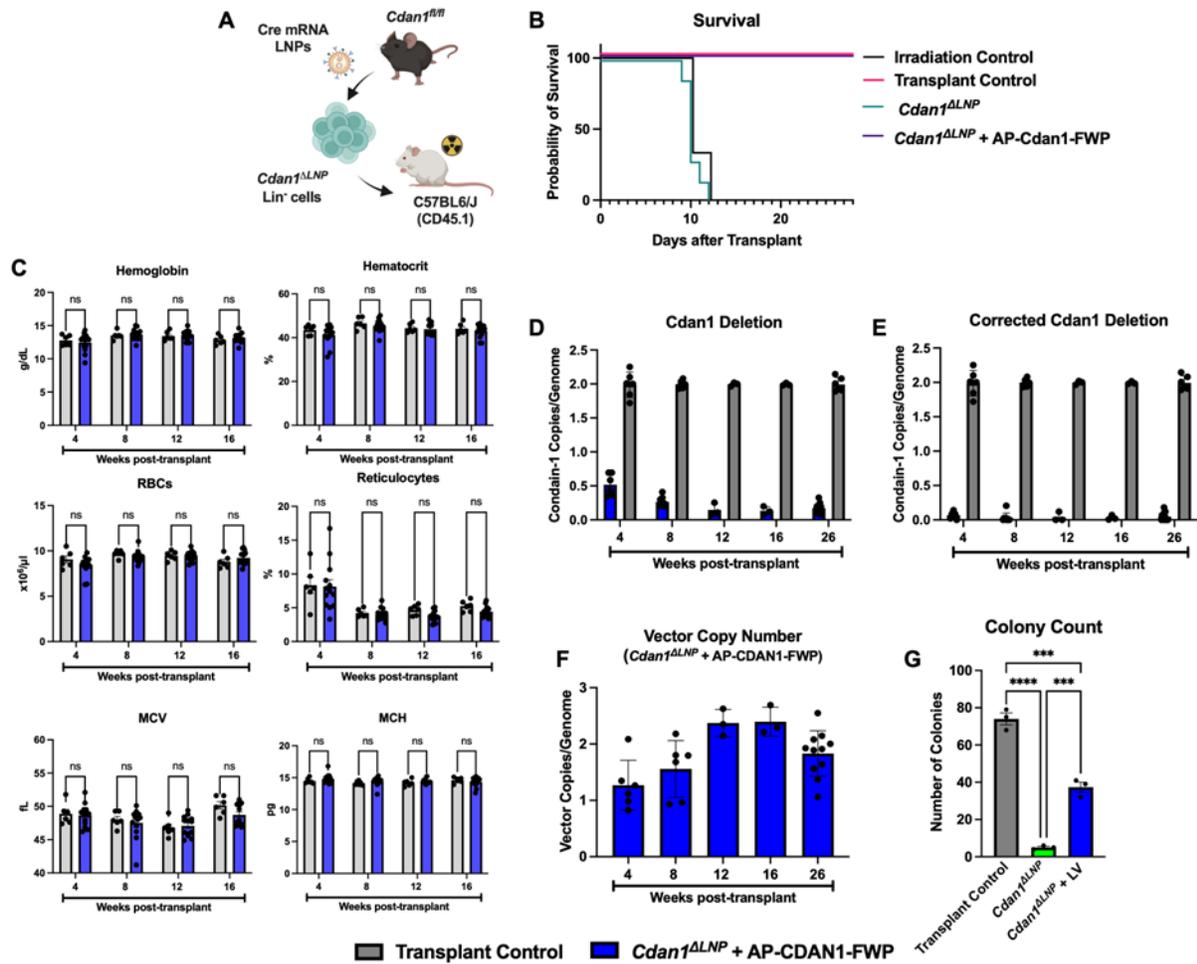


**Figure 2.2. The impact of *Cdan1* deletion on endogenous erythropoiesis.** (A) Representative flow cytometry analysis of *Cdan1*<sup>ΔMx</sup> of bone marrow erythroid cells at 7-9 days post poly (I:C). Data are shown as Ter119/CD71. (B) Representative FACS analysis of the erythroid populations at 7-9 days post poly(I:C). Data are shown as FSC-A/CD44 sub-gated on the Ter119<sup>+</sup> population. I+II = Proerythroblasts/basophilic erythroblasts; III = Polychromatic erythroblasts; IV: Orthochromatic erythroblasts/reticulocytes; V: Erythrocytes.



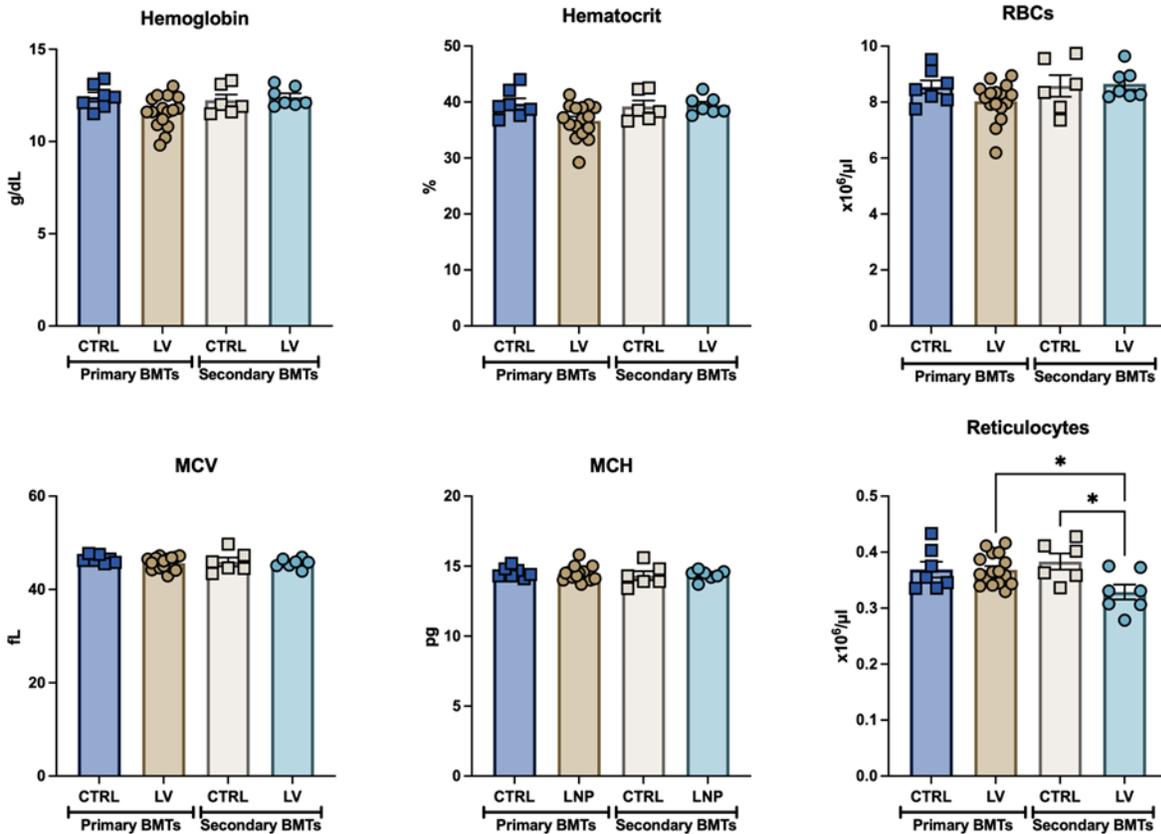
**Figure 2.3. The impact of *Cdan1* deletion on hematopoiesis and bone marrow integrity. (A)**

Representative FACS analysis *Cdan1*<sup>ΔMx</sup> of bone marrow LSK cells 7-9 days post poly(I:C). Data are shown as Sca1/cKit. **(B)** Representative FACS analysis of *Cdan1*<sup>ΔMx</sup> bone marrow LT-HSCs 7-9 days post poly(I:C). Data are shown as CD150/CD48 sub-gated on the LSK+ population **(C)** Representative FACS analysis of *Cdan1*<sup>ΔMx</sup> bone marrow granulocytes 7-9 days post poly(I:C). Data are shown as Gr-1/SSC-A. **(D)** Representative H&E stains of *Cdan1*<sup>ΔMx</sup> bone marrow sections, 10X.

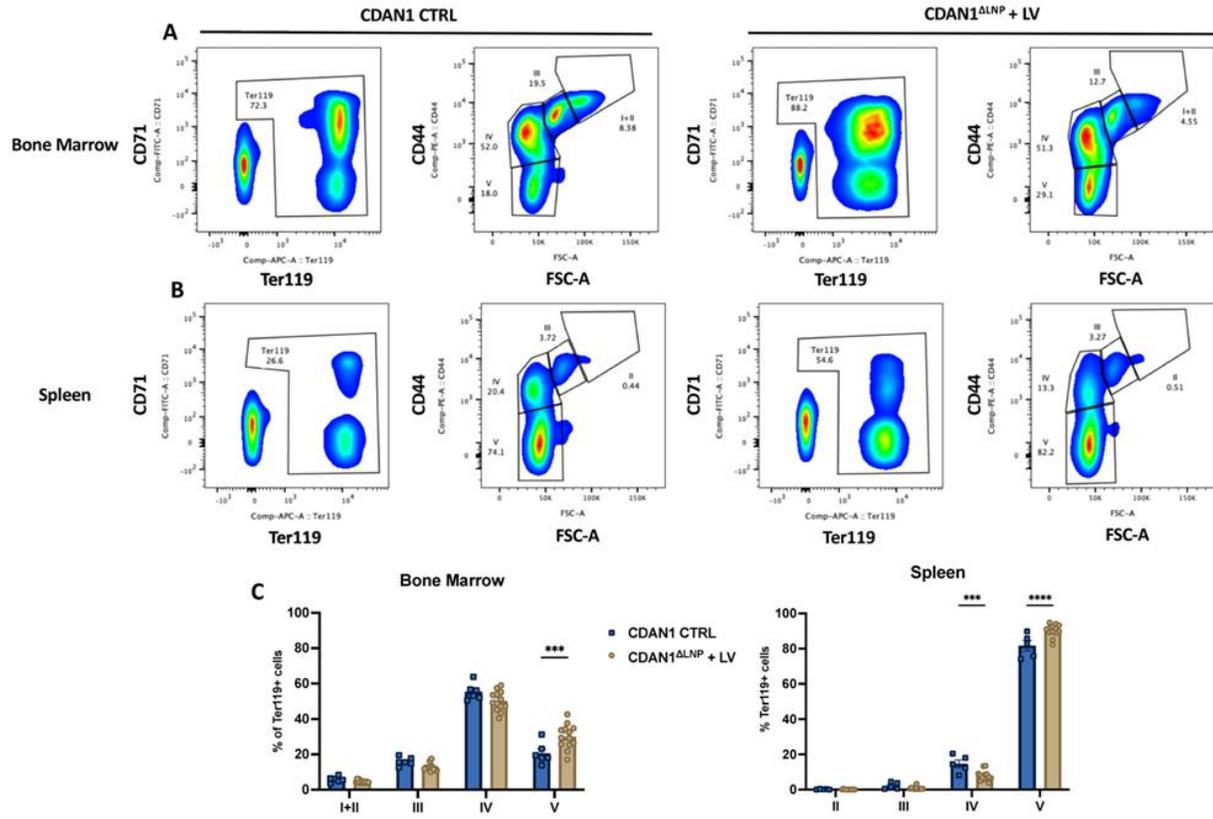


**Figure 2.4. Rescue of engraftment of *Cdan1*<sup>ΔLNP</sup> marrow via lentiviral gene therapy.** (A) Schematic of *Cdan1*<sup>ΔLNP</sup> experimental workflow. *Cdan1*<sup>fl/fl</sup> lin<sup>-</sup> cells are isolated and treated *ex vivo* with *LNP*<sup>CD117</sup>*Cre*, then transplanted into lethally irradiated recipients. (B) Survival curve for lethally irradiated *Cdan1*<sup>ΔLNP</sup> recipient animals. N=4-15. (C) CBC values at 16 weeks post-transplant. N=9-15 per condition. Data were expressed as mean ± SEM and analyzed via an unpaired two-tailed t-test. (D) Quantification of endogenous *Cdan1* levels via ddPCR from peripheral blood of *Cdan1*<sup>ΔLNP</sup> animals up to 26 weeks post-transplant. (E) *Cdan1* deletion values (D) normalized by donor chimerism levels to subtract recipient *Cdan1* copies. (F) ddPCR quantification of AP-*CDAN1*-FWP vector copy number in *Cdan1*<sup>ΔLNP</sup> animals post-transplant. (G) Comparison of total colony count in controls (*Cdan1*<sup>fl/fl</sup>), *Cdan1*<sup>ΔLNP</sup> lin<sup>-</sup>, and treated *Cdan1*<sup>ΔLNP</sup> lin<sup>-</sup> cells 1-week post-

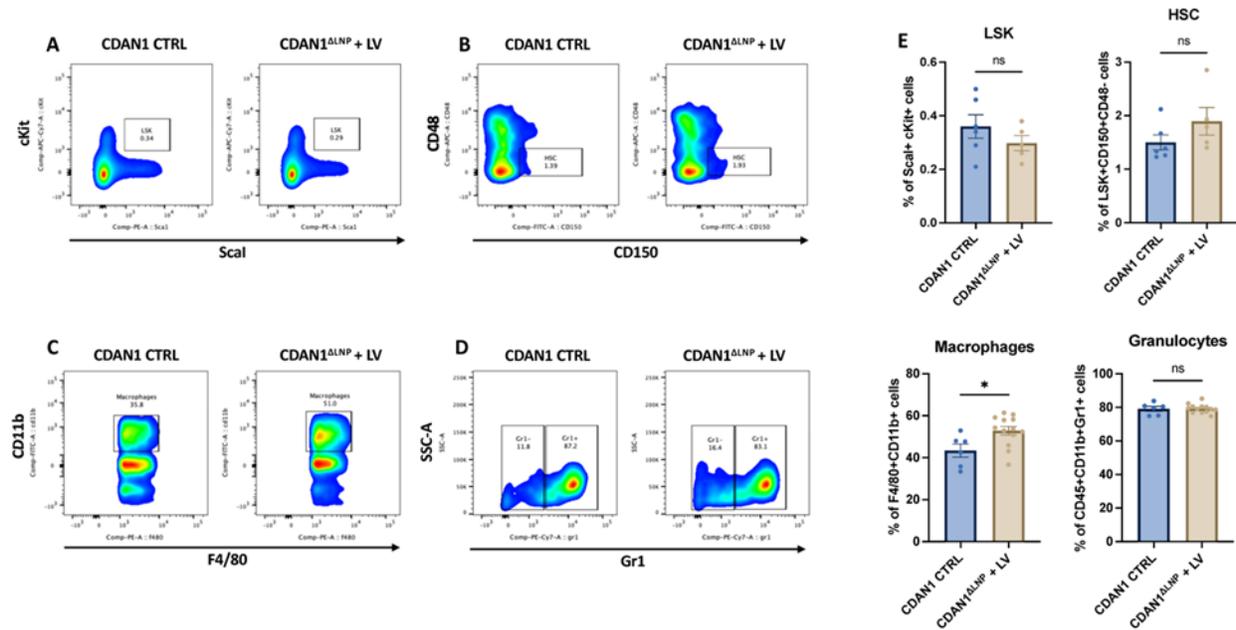
plating (N= 3). Data were expressed as mean  $\pm$  SEM and analyzed via Tukey's multiple comparison test after one-way ANOVA, \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ .



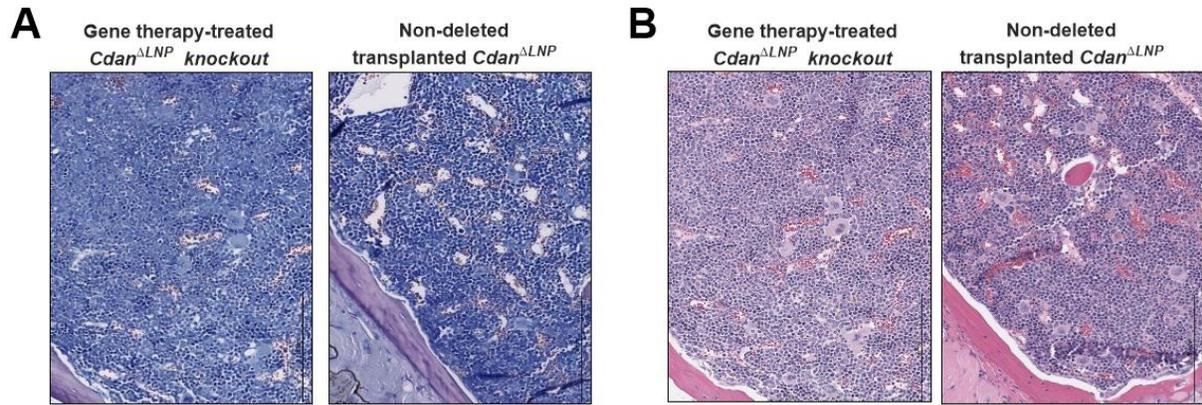
**Figure 2.5. Long-term assessment of erythroid parameters in peripheral blood of primary and secondary *Cdan1<sup>ΔLNP</sup>* treated and healthy control chimeras.** (A-F) Lethally irradiated primary and secondary *Cdan1<sup>ΔLNP</sup>* recipient mice were assessed 6 months post-transplant for complete blood counts. Results obtained from primary chimeras were obtained from 9-15 mice across three replicates, whereas results in secondary chimeras were obtained from 6-7 mice across one cohort for control and treated mice. Data were expressed as mean ± SEM and analyzed via Tukey's multiple comparison test after one-way ANOVA, \*p < 0.05.



**Figure 2.6. Long-term assessment of lentiviral corrected *Cdan1*<sup>ALNP</sup> marrow for maintenance of erythropoiesis.** (A) Representative FACS analysis of *Cdan1*<sup>ALNP</sup> bone marrow erythropoiesis 6 months post-transplant, after treatment with AP-CDAN1-FWP and in mice transplanted with control BM (*Cdan1*<sup>fl/fl</sup>). Erythroid cells are gated according to CD71/Ter119 distribution, and erythroid maturation is dissected using FSC-A/CD44 distribution sub-gated on the erythroid Ter119<sup>+</sup> population. (B) Representative FACS analysis of *Cdan1*<sup>ALNP</sup> spleen erythropoiesis a 6-months post-transplant following the gating strategy shown in A. (C) Quantification of erythroid populations in bone marrow and spleen, from A and B, respectively. FSC-A/CD44 distribution in A-B: I+II = Proerythroblasts/basophilic erythroblasts; II: Basophilic erythroblasts; III = Polychromatic erythroblasts; IV: Orthochromatic erythroblasts/reticulocytes; V: Erythrocytes. N=6-15 between three replicates. Data were expressed as mean ± SEM and analyzed via an unpaired two-tailed t-test, \*p< 0.05, \*\*p<0.01.



**Figure 2.7. Long-term assessment of lentiviral corrected *Cdan1*<sup>ALNP</sup> marrow.** (A-B) Representative FACS analysis of *Cdan1*<sup>ALNP</sup> bone marrow LSK cells and LT-HSC 6 months post-transplant. Data are shown as (A) Sca1/cKit (LSK cells) and (B) CD150<sup>+</sup>/CD48<sup>-</sup> (LT-HSC). (C) Representative FACS analysis of macrophages in *Cdan1*<sup>ALNP</sup> bone marrow. Data are shown as F480/Cd11b. (D) Representative FACS analysis of granulocytes in *Cdan1*<sup>ALNP</sup> bone marrow. Data are shown as Gr-1/SSC-A. (E) Quantification of FACS analysis is shown in (A-D). N=6-15. Data were expressed as mean  $\pm$  SEM and analyzed via an unpaired two-tailed t-test, \* $p < 0.05$ .



**Figure 2.8. Histopathological assessment of lentiviral corrected *Cdan<sup>ΔLNP</sup>* bone marrow.**

(A) Representative Wright-Giemsa stains of sectioned bone marrow six months post-transplant.

(B) Representative hematoxylin & eosin stains of sectioned bone marrow six months post-transplant.

Supplemental Figures

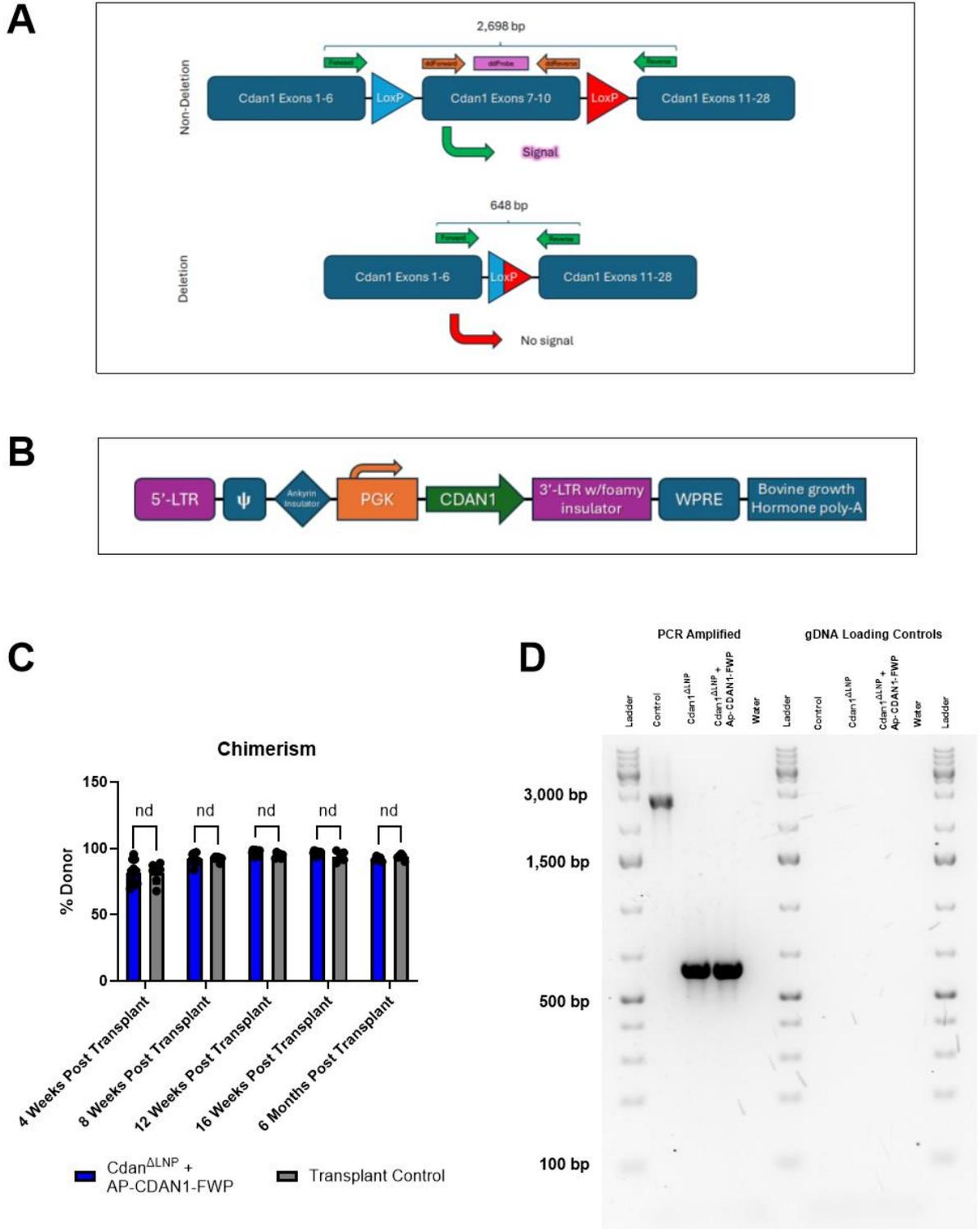
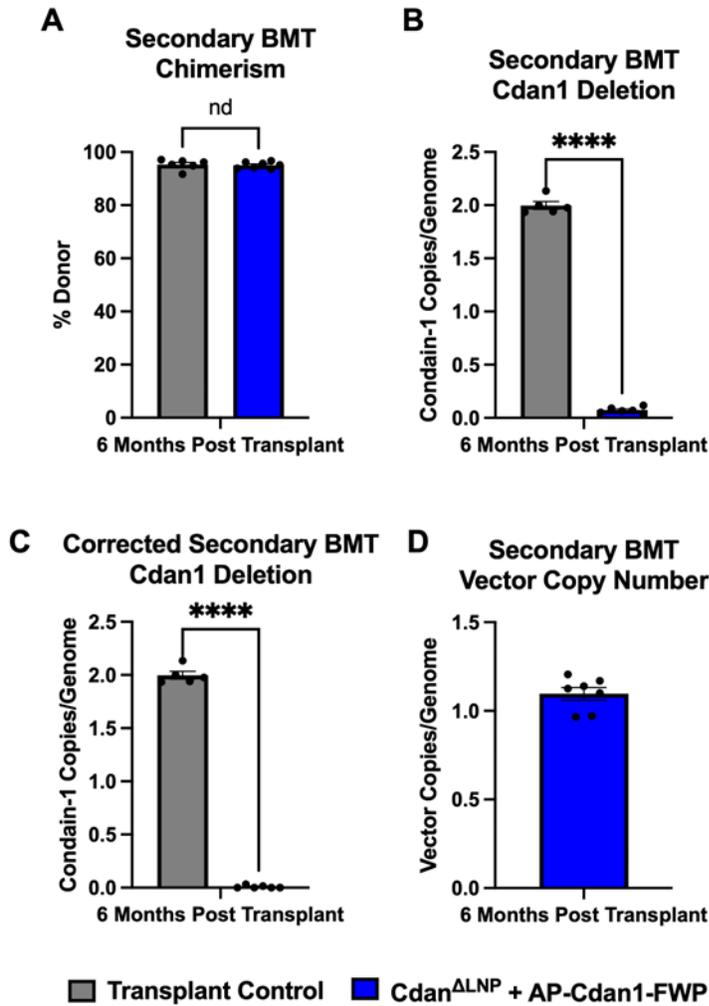
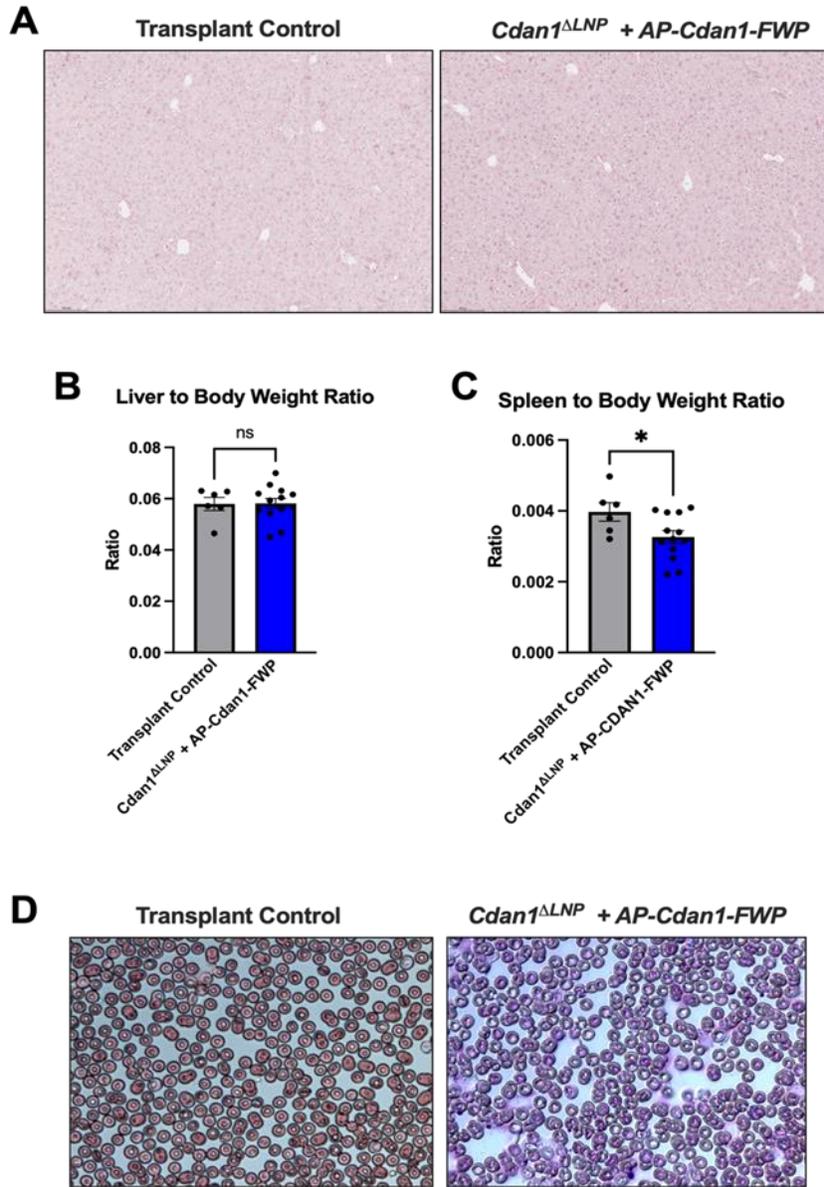


Figure S1

**Figure 2.S1.** (A) Schematic representation of the location of LoxP sites in the *Cdan1* gene and relative locations of ddPCR primer/probe set (ddForward, ddProbe, ddReverse) flanked by LoxP sites, along with expected output depending on gene deletion: signal upon non-deletion or no signal upon deletion. Traditional PCR primer sets are located outside of the LoxP sites (Forward, Reverse), and the expected amplicon size depending on gene deletion: 2,698 bp upon non-deletion and 648 bp upon deletion. (B) Schematic representation of the AP-CDAN1-FWP lentiviral vector cassette. The *CDAN1* open reading frame is under the control of the PGK-constitutive promoter and insulated on the 5' end by the ankyrin insulator and on the 3' by the foamy insulator located within the 3'-LTR, which in the integrated proviral sequence will replace the 5'-LTR. WPRE: Woodchuck hepatitis virus post-transcriptional regulatory element. (C) Peripheral blood chimerism levels of *Cdan1*<sup>ALNP</sup> animals, post-transplant. N=9-15. Data were expressed as mean ± SEM and analyzed via an unpaired two-tailed t-test, \*p< 0.05. (D) PCR-based assessment of *Cdan1* deletion, verifying deletion of LoxP flanked segment in *Cdan1*<sup>ALNP</sup> lin<sup>-</sup> cells 24 hours post-LNP<sup>CD117</sup>Cre, showing the expected amplified bands as indicated in S1A. The water-amplified and the gDNA non-amplified DNA were uploaded to check for potential contaminations.



**Figure 2.S2. Assessment of Secondary transplant chimerism, *Cdan1* deletion, and VCN.** (A) Analysis of *Cdan1*<sup>ALNP</sup> secondary transplant peripheral blood chimerism at six months. N=6-7. Data were expressed as average  $\pm$  SEM and analyzed via an unpaired two-tailed t-test, \* $p < 0.05$ . (B) Analysis of endogenous *Cdan1* levels in *Cdan1*<sup>ALNP</sup> secondary transplant animals from the peripheral blood at six months. Corrected values utilize chimerism levels to correct for recipient *Cdan1* levels. N=6-7. Data were expressed as average  $\pm$  SEM and analyzed via an unpaired two-tailed t-test, \*\*\*\* $p < 0.0001$ . (C) Assessment of *Cdan1*<sup>ALNP</sup> secondary transplant peripheral blood AP-CDAN1-FWP VCN at six months.



**Figure 2.S3. Assessment of liver iron build-up, liver/spleen to body weight ratios, and peripheral blood morphology.** (A) Representative Perl's Prussian blue stained liver sections of primary *Cdan1*<sup>ΔLNP</sup> mice at six months post-transplant. 10X magnification. (B) Liver to body-weight ratio of primary *Cdan1*<sup>ΔLNP</sup> mice at six months post-transplant. N=6-13. Data were expressed as mean  $\pm$  SEM and analyzed via unpaired two-tailed t-test, \* $p < 0.05$ . (C) Spleen to body-weight ratio of primary *Cdan1*<sup>ΔLNP</sup> mice at six months post-transplant. N=6-13. Data were expressed as mean  $\pm$  SEM and analyzed via an unpaired two-tailed t-test, \* $p < 0.05$ .

(D) Representative peripheral blood smears of control and treated mice at six months post-transplant. RAL Diff-Quik staining, 40X magnification.

## CHAPTER 3 – MUTATED CDAN1 VECTORS AND MACHINE LEARNING REVEAL ASPECTS OF CDAN1 FUNCTIONALITY

### Introduction

The results in the previous chapter, while evocative and demonstrating a gene therapy based rescue of *Cdan1* knock out, clearly do not recapitulate CDA1a as a disease. Spontaneous bone marrow failure and resulting aplasia is not a feature of CDA1a, nor such rapid and total lethality. We reasoned that if we have the ability to deliver a curative, rescue *Cdan1* cassette we could trivially deliver known disease causing mutations via the same system, thereby reconstituting the disease. The most common causative *Cdan1* mutation causing disease is R1042W, the so-called Bedouin mutation. Naturally, this variant was our first choice, but given uncertainties regarding differences between mouse and human physiology we elected to test another mutation, P672L, in parallel with R1042W. These variants were selected due to their relatively high proportions in CDA1a patient populations and because these residues are conserved between humans and mice.<sup>137</sup> These experiments deployed the same *ex vivo* LNP based *Cdan1*<sup>ΔLNP</sup> methodology utilized in chapter 2, simply replacing the wild type CDAN1 open reading frame with either CDAN1 P672L or CDAN1 R1042W. These vectors are referred to as AP-CDAN1-P672L-FWP and AP-CDAN1-R1042W-FWP, respectively and hereafter collectively referred to as mutant vectors.

### Results

*Cdan1*<sup>ΔLNP</sup> mice transduced with the mutant vectors engraft in-line with AP-CDAN1-FWP transduced mice and transplant controls, indicating a reconstitution of hematopoiesis. However, the CBC results for these animals are comparable with wild-type controls, lacking

classical indications of CDA1a. (Figure 3.1 A-F) Mutant vector treated animals have wild-type hemoglobin levels, RBC counts, hematocrit and most perhaps the most telling, wild-type MCV values, which should be inflated within a CDA1a condition. Furthermore, at 4-weeks post-transplant, the mutant vector treated animals exhibit elevated reticulocyte counts, showing an adaptation to anemia downstream from the transplant procedure. In CDA1a, patients do not produce a sufficient reticulocyte response as a result of infection or trauma due to ineffective erythropoiesis. The presence of an elevated reticulocyte count at 4-weeks, in line with controls, demonstrates a full recapitulation of erythropoiesis that can respond to dynamic changes in physiological requirements. Potentially, this could be as a result of the *CDAN1* gene being downstream of a constitutive promoter producing supraphysiological levels of the CDAN1 protein, rescuing the deficient mutation via saturation. This is unlikely to be the case as CDAN1 is known to form a 1:1 stoichiometric association with CDIN1, the protein implicated in the pathophysiology of CDA1b. Given this ratio, excess CDAN1 levels would be superfluous as they would lack an endogenous CDIN1 binding partner past a certain level of expression. This is in contrast to diseases wherein an enzyme is deficient in performing a chemical reaction. In such a case, it is conceivable that sufficiently high levels of deficient enzyme, provided that the variant at hand does not totally lack catalytic activity, could rescue the disease state.

***CDAN1 vectors with multiple mutated residues do not rescue *Cdan1*<sup>ALNP</sup> mice.*** Intrigued by this phenomenon, we generated additional mutant vectors that combine CDA1a-causing mutations into pairs. We rationalized that while there are no known human CDAN1 variants with more than one altered amino acid residue, at least in homozygosity and presumably due to embryonic lethality, mice might be able to tolerate a pair of alterations due to the lack of disease state seen in the mutant vector line of inquiry shown previously. The double mutants selected

were P672L, F714W and P672L, F868I; selected due to these residues being retained between humans and mice, as with the single mutant vectors. All of the double mutant transplanted mice died in line with irradiation controls, indicating a failure to engraft mirroring *Cdan1* cKO animals. This result offers further evidence for the sensitivity of CDAN1 function to sequence alterations.

***Machine learning provides a putative mechanism for CDA1a pathology.*** Given the available evidence that CDA1a disease pathology is downstream of faulty protein-protein interactions, we assessed the locations of CDA1a mutations on the 3D structure of CDAN1. Historically, this would have been a major undertaking, requiring either x-ray crystallography and/or Cryo-electron microscopy (cryo-EM). To bypass this limitation, we turned to a machine learning (ML) algorithm, AlphaFold. AlphaFold is a ML model trained on thousands of experimentally determined and extensively curated protein structures that have been obtained over decades of research by countless groups.<sup>152</sup> In brief, the algorithm was exposed to the primary amino acid sequences of numerous proteins which were then associated with a known 3D structure. The algorithm develops a mathematical representation of these relationships, then compares it to an isolated test set of primary and 3D structures, recursively improving itself over many training cycles, or epochs. In this manner the machine develops a heuristic understanding of protein folding, without requiring intensive molecular dynamic calculations which would be intractable for all but the most powerful supercomputing clusters. Based on an AlphaFold simulation of CDAN1, we found that the mutations tended to cluster around a prominent groove formation near the c-terminus of the molecule. (**Figure 3.2**) The prominence of this formation is highly evocative, appearing to be an ideal location for interaction with other macromolecules. CDAN1 has high structural homology to CNOT1, a scaffolding protein essential in the functionality of

the CCR4-Not complex, which post-transcriptionally regulates gene expression via mRNA deadenylase activity.<sup>153</sup> Based on this, disruptions in protein complexing is the most parsimonious explanation for the CDA1a disease state.

Extensive work by other groups has elucidated the binding partners of CDAN1. As touched upon earlier, CDAN1 is known to bind histone chaperones ASF1a and ASF1b via conserved basic residues in its N-terminus, so-called B-domains. Through this interaction with ASF1, CDAN1 complexes with s-phase histones H3.1 and H4 along with importin-4 as indicated by co-immunoprecipitation and mass spectroscopy. Mutational studies utilizing a mutated variant of ASF1a, V94R, that does not bind Histones indicates that the interaction of CDAN1 with these additional binding partners is mediated through ASF1a, as H3, H4 and importin-4 do not pull down in complex with CDAN1, but ASF1a V94R retains association.<sup>103</sup> Furthermore, CDIN1, previously referred to in the literature as C15orf41, is known to associate with the c-terminus of CDAN1 based upon studies performed with truncated versions of CDAN1.<sup>109</sup>

Fortuitously, more recent versions of AlphaFold have been developed that have been trained on protein complexes, allowing for the *in silico* exploration of protein interactions, informing directions of research for wet lab experimentation and validation of the ML findings.<sup>152</sup> We leveraged this ability to interrogate the ASF1-CDAN1-CDIN1 complex. To assess the CDAN1 complex we utilized AlphaFold 3 multimer, deploying the human orthologues of ASF1, CDAN1, and CDIN1 into the system to interact and complex *in silico*. Both ASF1a and ASF1b were complexed in parallel. AlphaFold faithfully recapitulated the canonical location of the ASF1 binding location to the N-terminal B-domains of CDAN1, lending credence to the accuracy of the AlphaFold prediction. Additionally, AlphaFold associated CDIN1 with the

C-terminus of CDAN1, another faithful reconstitution of experimental wet lab evidence. **(Figures 3.3, 3.4)** Evocatively, AlphaFold indicated that CDIN1 inserts itself into the aforementioned groove formation that exhibits CDA1a mutational clustering. Given that CDIN1 is known to be stabilized by interaction with CDAN1, this interaction appears to be performing the function of a chaperone protein, ensuring faithful folding of CDIN1 by CDAN1 and prolonging the lifespan of CDIN1. Indeed, heterologous expression of CDIN1 in the absence of CDAN1 results in no protein production.

Most intriguing, however, was a putative non-canonical interaction between the c-terminus of ASF1 and the open structure of CDIN1 as held into shape by CDAN1, analogous to a hand (CDAN1) wielding a baseball glove (CDIN1) to catch a ball (c-terminus of ASF1). **(Figure 3.5)** As discussed previously, the c-terminus of ASF1 is phosphorylated by tousel-like kinases as a form of regulation.<sup>154</sup> It is conceivable that the interaction of ASF1 and CDIN1, if real, performs a negative-regulatory function by steric inhibition of phosphorylation on the c-terminus of ASF1. Furthermore, the c-terminus of ASF1 is known to associate with histones H3 and H4, members of the same importin-IV complex addressed previously. Loss of the c-terminus of ASF1 is shown to reduce the association between ASF1 and H3/H4 by a factor of 200.<sup>155</sup> A closer inspection of the ASF1-CDIN1 interface reveals multiple contacts, primarily driven by hydrophobic and van der Waals interactions at less than 15 angstroms, providing a plausible interaction mechanism **(Figure 3.6)**.

Fortuitously, AlphaFold simultaneously produces metrics designed to convey the level of confidence of the protein structure and complexes it produces, allowing for a measure of quality assurance. Predicted Alignment Error (PAE) estimates how faithfully the model predicts the 3D

positioning and internal rotation of a given structural object, often referred to in ML terms as a token. PAE estimates the error between location and orientation of two tokens within the output. This score is useful for deriving confidence levels within the minutiae of the protein structure. Lower values indicate lower predicted error, and thusly a higher confidence in the prediction. This alignment is derived from the protein backbone in 3D space, with reference conformers used for co-factors and post-translational modifications. Predicted template modeling (pTM) assesses the accuracy of the totality of the structure, essentially asking if the manifold-smoothed surface occupies the 3D space it is intended to occupy. A further development of pTM is ipTM, or interface predicted template modeling which assesses the location of individual monomers within a complex and their 3D association with other complex members. Both pTM and ipTM are intended as total structure confidence metrics. Lastly, predicted local distance difference test (pLDDT) is a confidence metric measuring distance to other polymers on a per-atom basis as a further development of an IDDT-C $\alpha$  metric based on alpha carbons used historically.<sup>156</sup> The pLDDT metric is useful when a given protein has both structured and unstructured regions, allowing the researcher to curate known disorganized protein regions that might cause the model to score outputs with these proteins as lower.

Scoring of the CDAN1 complex was assessed, with moderate pTM values and ipTM values for the structure and interfaces, respectively (**Table 3.1**). The values found are marginal, as a good rule of thumb for ML protein scoring is that a pTM value in excess of 0.5 is more than likely representative of the true structure. For ipTM values in excess of 0.8 are considered highly confident, whereas values of less than 0.6 are often considered a failed prediction. Values between 0.6 and 0.8 form an interpretive gray area. These scores are not absolute and require human interpretation as the dynamic nature of protein structure and interaction can be lost in

current ML processes. The structure and interface predictions all score greater than 0.6 but by a marginal extent. (TABLE) Per-atom pLDDT values across the complex suggest that much of the uncertainty in scoring is derived from unstructured regions of both CDAN1 and ASF1, reflecting experimental data as ASF1 is known to have an unstable c-terminus.<sup>155</sup> An additional feature of AlphaFold 3 is its ability to assess post-translational modifications, including phosphorylation, glycosylation, etc. Taking advantage of this feature, we re-ran the simulation with phosphorylated ASF1a and ASF1b c-terminal residues known to be acted upon by tyrosine-like kinases. (Table 3.2) This resulted in a minor drop in pTM and ipTM, suggesting a shift in interaction, but the change was very small, so this result is inconclusive. However, pLDDT data demonstrates an increase in confidence in per-atom alignment between all three proteins when ASF1 is phosphorylated (Figure 3.7). This effect is most pronounced in the unstructured c-terminus of ASF1.

## Discussion

A major limitation of this data is the fact that as sophisticated as modern ML algorithms are, they can still make mistakes and ‘hallucinate.’<sup>157</sup> Hallucinations in ML can run the gambit from the machine effectively acting on a ‘hunch’ to gross errors that do not make any logical sense to an educated human observer. This phenomenon is derived from the fact that a ML algorithm does not understand or logically reason with the information it has learned, rather it has found a ‘path of least resistance’ through the high level dimensional space it navigates to derive a solution.<sup>158</sup> It is conceivable that this interaction, if real, was missed in previous CDAN1 complex studies due to artifacts often encountered in co-ip experiments.<sup>159</sup> In fact, a direct interaction between ASF1 and CDIN1 is incidentally noted in protein-protein interaction

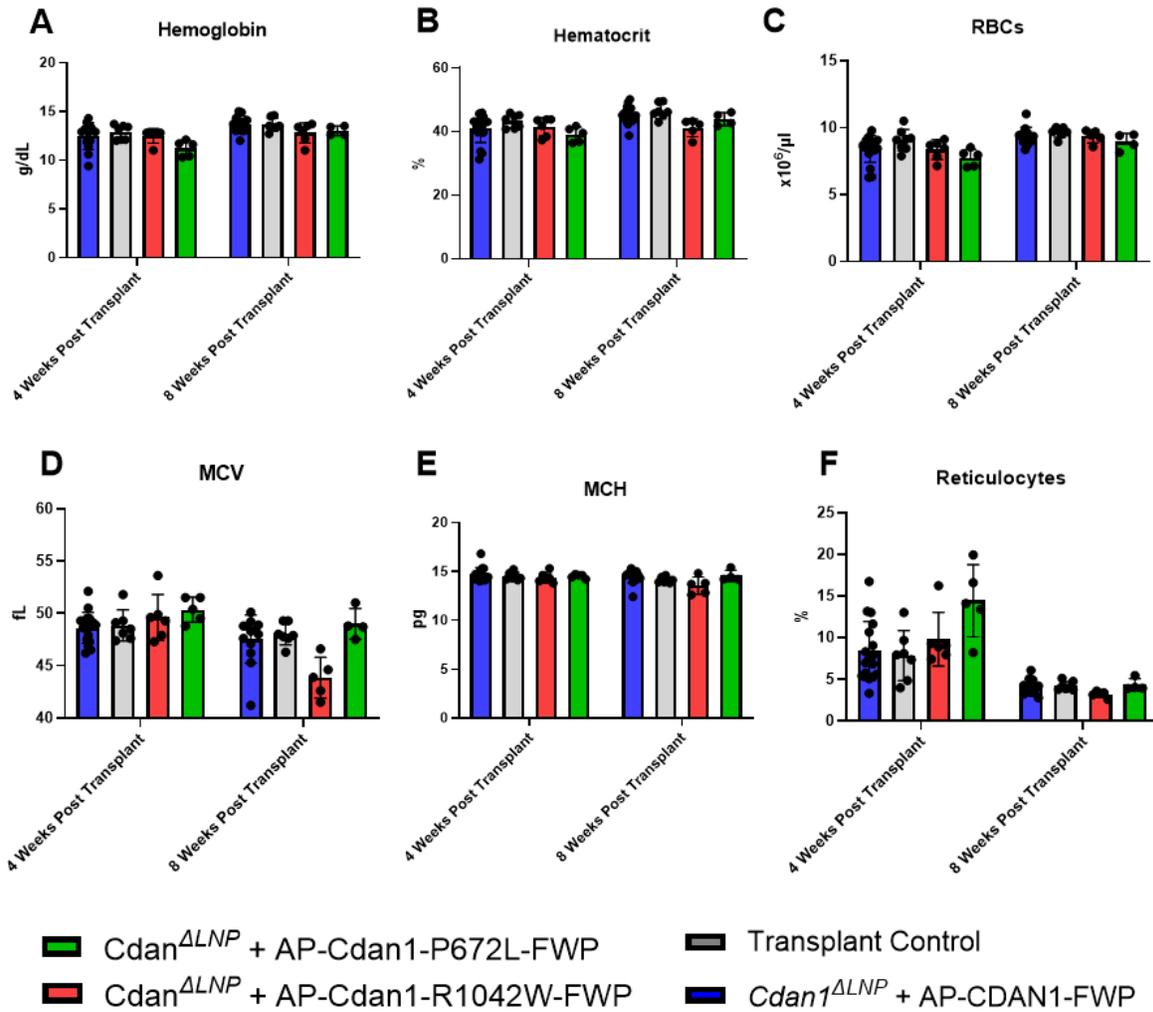
studies not focused on the CDAN1 complex.<sup>132</sup> There are numerous limitations in this line of inquiry, namely a present lack of direct experimental evidence of this interaction in experiments designed specifically to look for the direct ASF1-CDIN1 association. An additional issue is the fact that AlphaFold and many other ML-based protein structure algorithms are trained on cryo-EM and x-ray crystallography derived structures, which do not reflect the dynamic nature of active protein biomolecules. This algorithmic preference for mature proteins and complexes could, in fact, hide dynamic protein sub-unit interactions or understate their activity.

## **Methods**

***Machine Learning-based Determination of CDAN1-ASF1-CDIN1 complex.*** Primary amino acid sequences for human ASF1A, ASF1B, CDAN1, and CDIN1 were obtained from the UniProt database (Entry numbers: ASF1A: Q9Y294, ASF1B: Q9NVP2, CDAN1: Q8IWY9, CDIN1: Q9Y2V0). Complexes were fed into AlphaFold 3 via the AlphaFold server (courtesy of Google DeepMind, <http://alphafoldserver.com>) and generated utilizing a random seed.

***Analysis of CDAN1-ASF1-CDIN1 complex.*** AlphaFold 3 generated structure files were assessed via ChimeraX, an open source protein structure program produced by the University of California San Francisco Resource for Biocomputing, Visualization, and Informatics and can be downloaded via <https://www.cgl.ucsf.edu/chimerax/>.

## Figures



**Figure 8**

**Figure 3.1. Complete blood count values of  $Cdan1^{\Delta LNP}$  marrow treated with mutant lentiviral gene therapy vectors.** (A-F) Selected  $Cdan1^{\Delta LNP}$  CBC values assessed through 8-weeks of post-transplant follow-up. Results are derived from 5-15 animals per condition across 2-3 replicates and were analyzed via Tukey's multiple comparison test after one-way ANOVA, \* $p < 0.05$ , only results indicating significance shown.

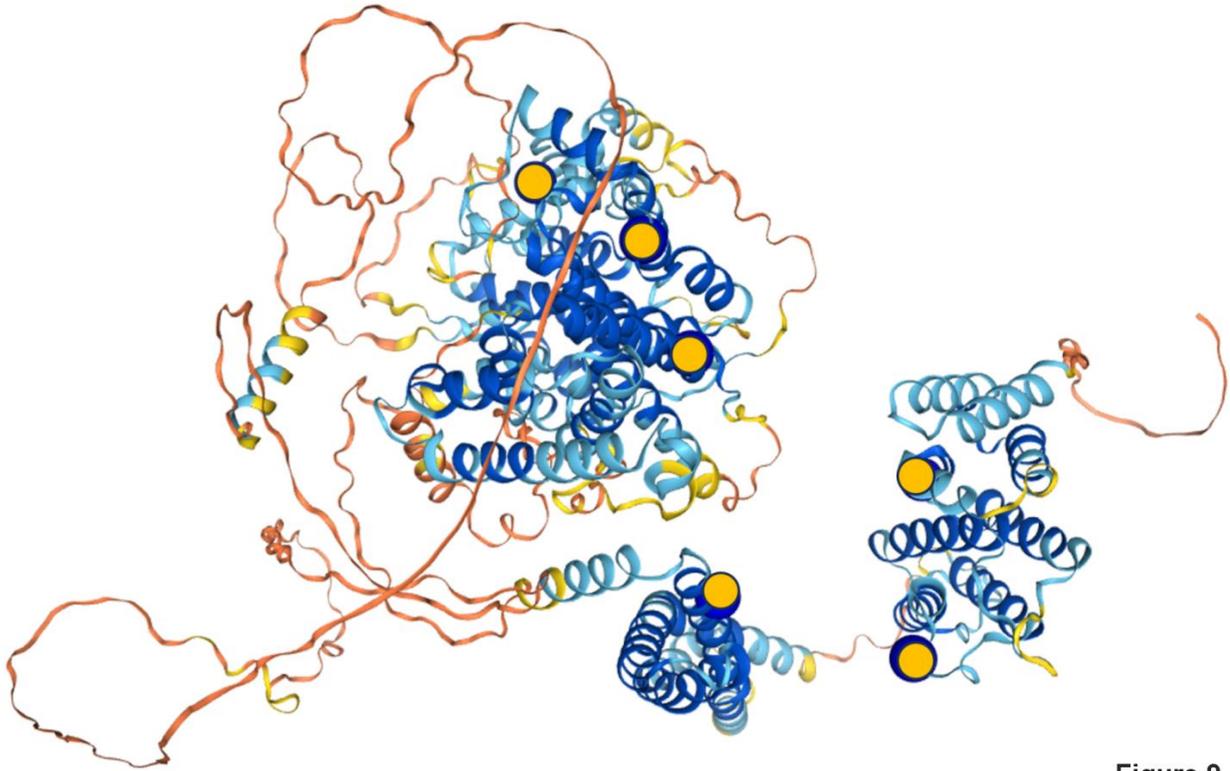
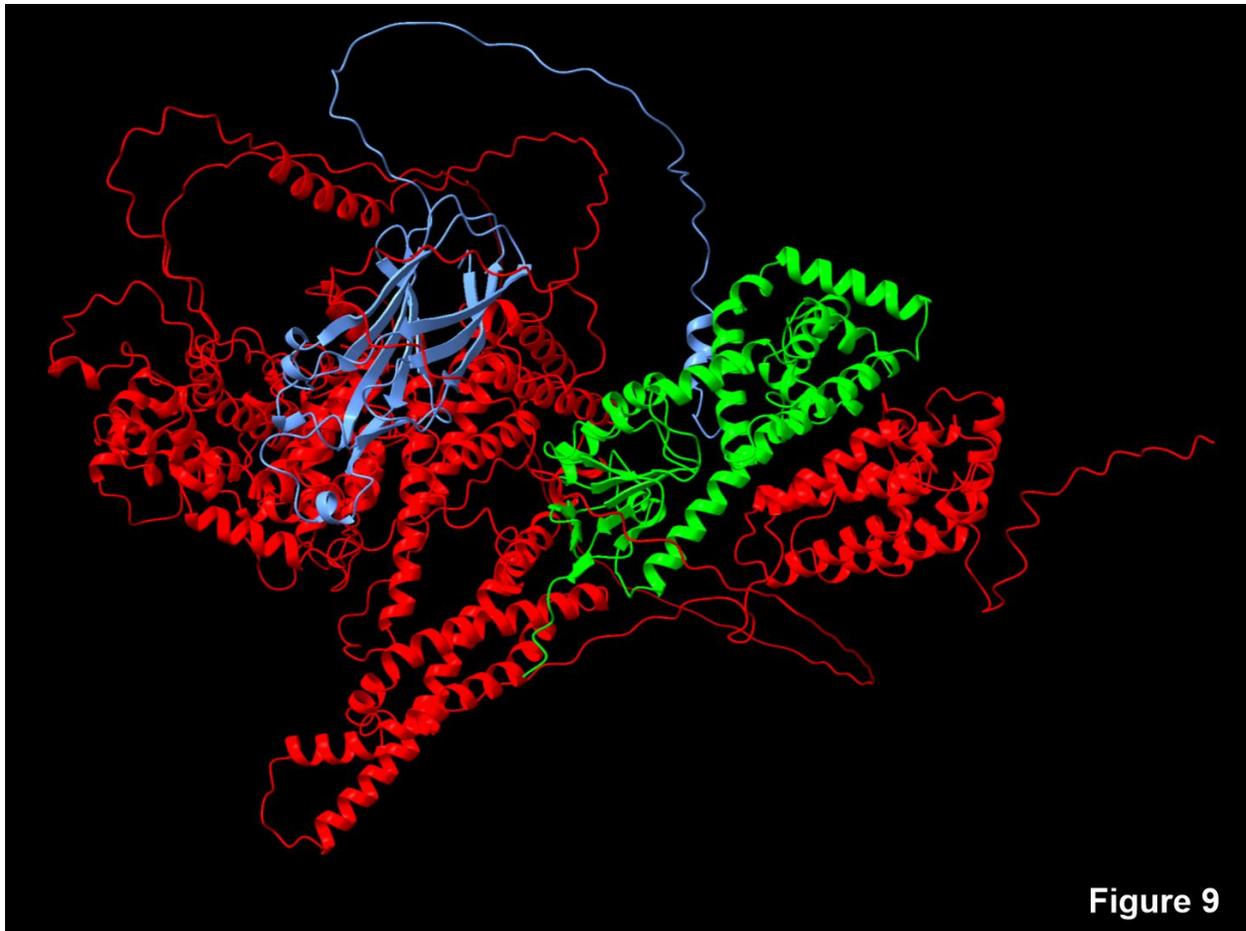


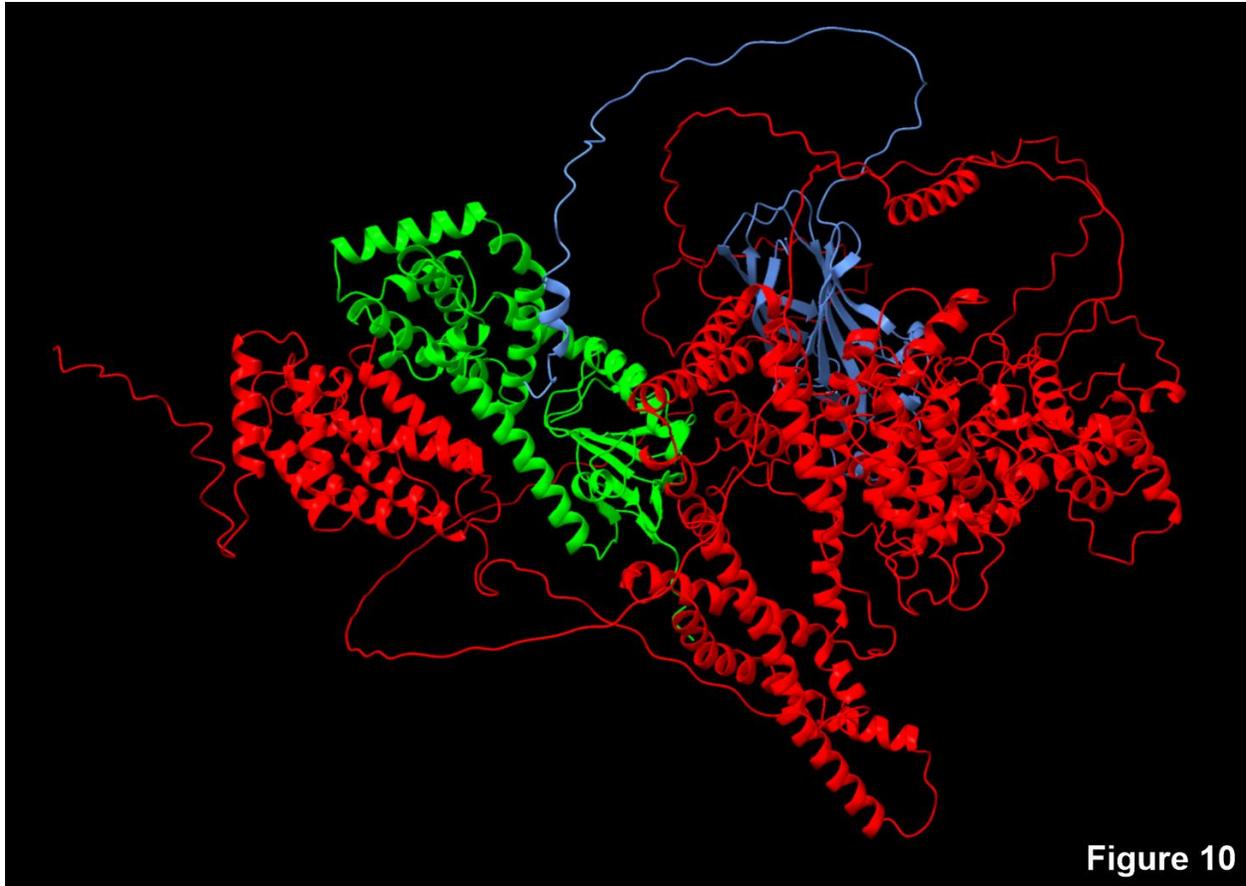
Figure 9

**Figure 3.2. AlphaFold generated model of CDAN1 with CDA1a causing residues indicated.**

Ribbon model of CDAN1, confidence levels in structure indicated via coloration, with bluer being increased confidence, redder less confidence. Disordered regions inherently have lower confidence levels. Residues associated with CDA1a disease state are marked with yellow circles, and appear to cluster around a large groove-like structure implying a functional association.



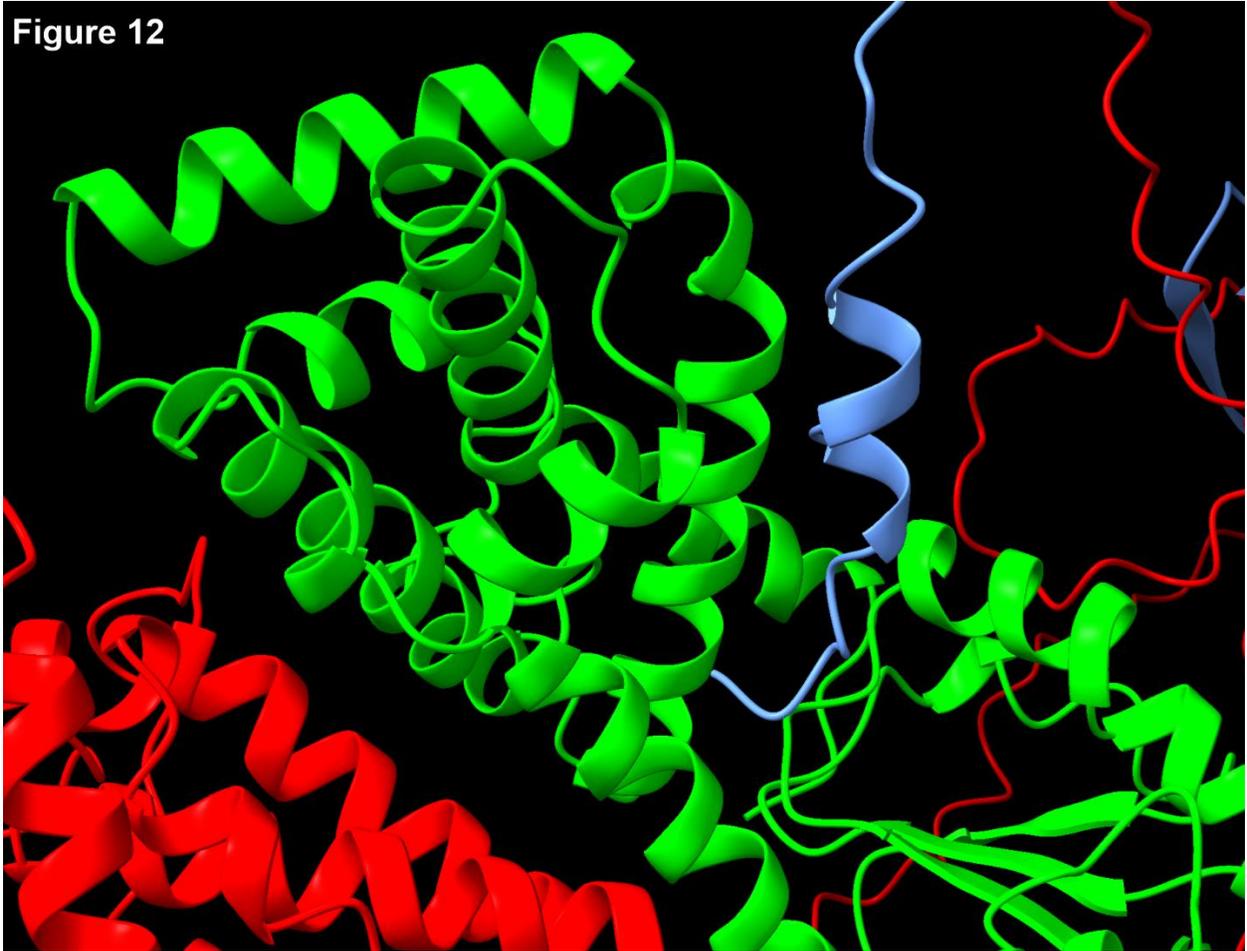
**Figure 3.3. AlphaFold generated model of the ASF1-CDAN1-CDIN1 structure.** ASF1 (blue) associates with the n-terminus of CDAN1 (red) and CDIN1 (green). CDIN1 binds to the groove noted in CDAN1 and the c-terminal region of ASF1 associates with CDIN1 in a conformation facilitated by CDAN1.



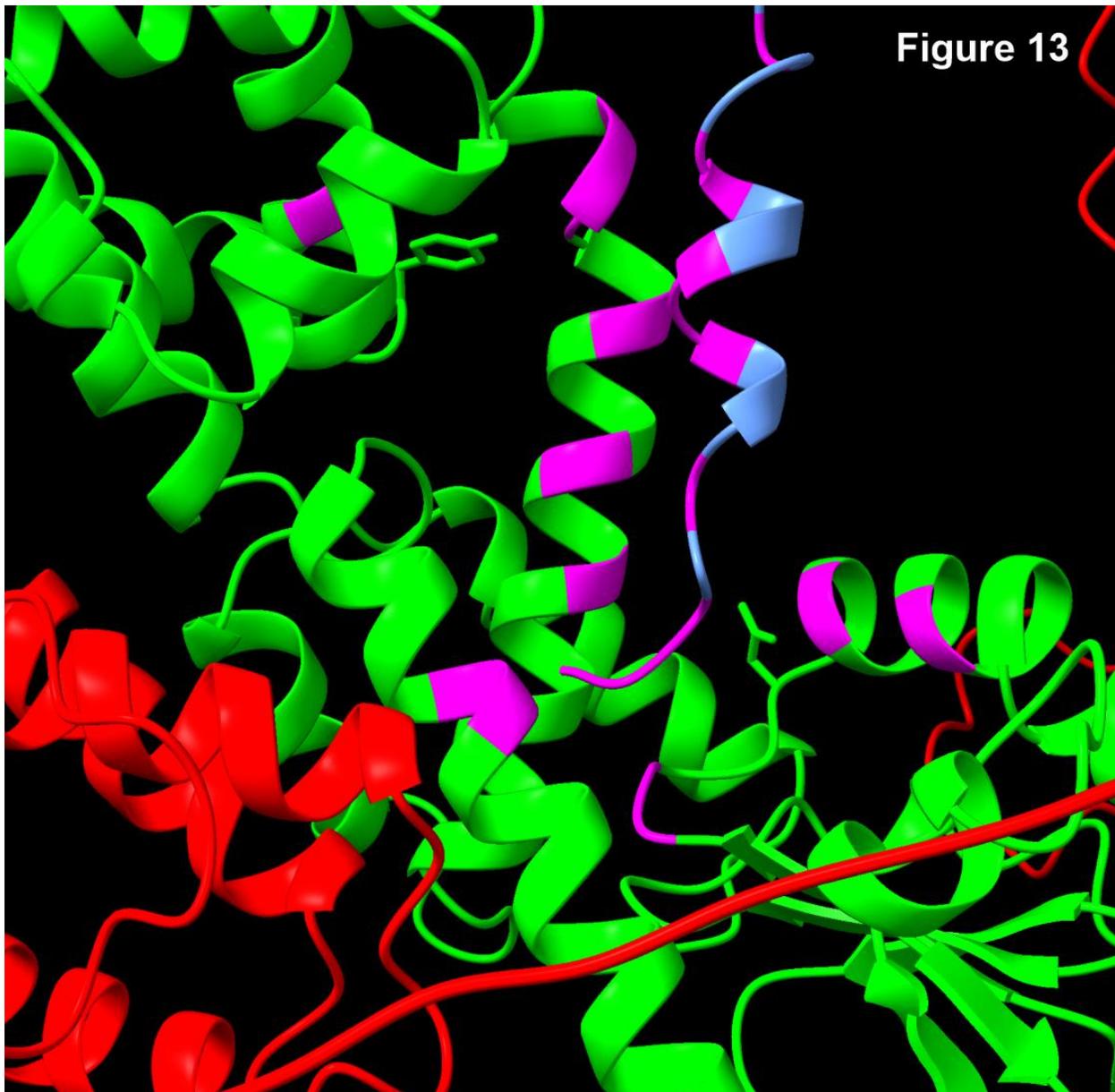
**Figure 10**

**Figure 3.4.** AlphaFold generated model of the ASF1-CDAN1-CDIN1 structure, obverse side. ASF1 (blue) associates with the n-terminus of CDAN1 (red) and CDIN1 (green). CDIN1 binds to the groove noted in CDAN1 and the c-terminal region of ASF1 associates with CDIN1 in a conformation facilitated by CDAN1.

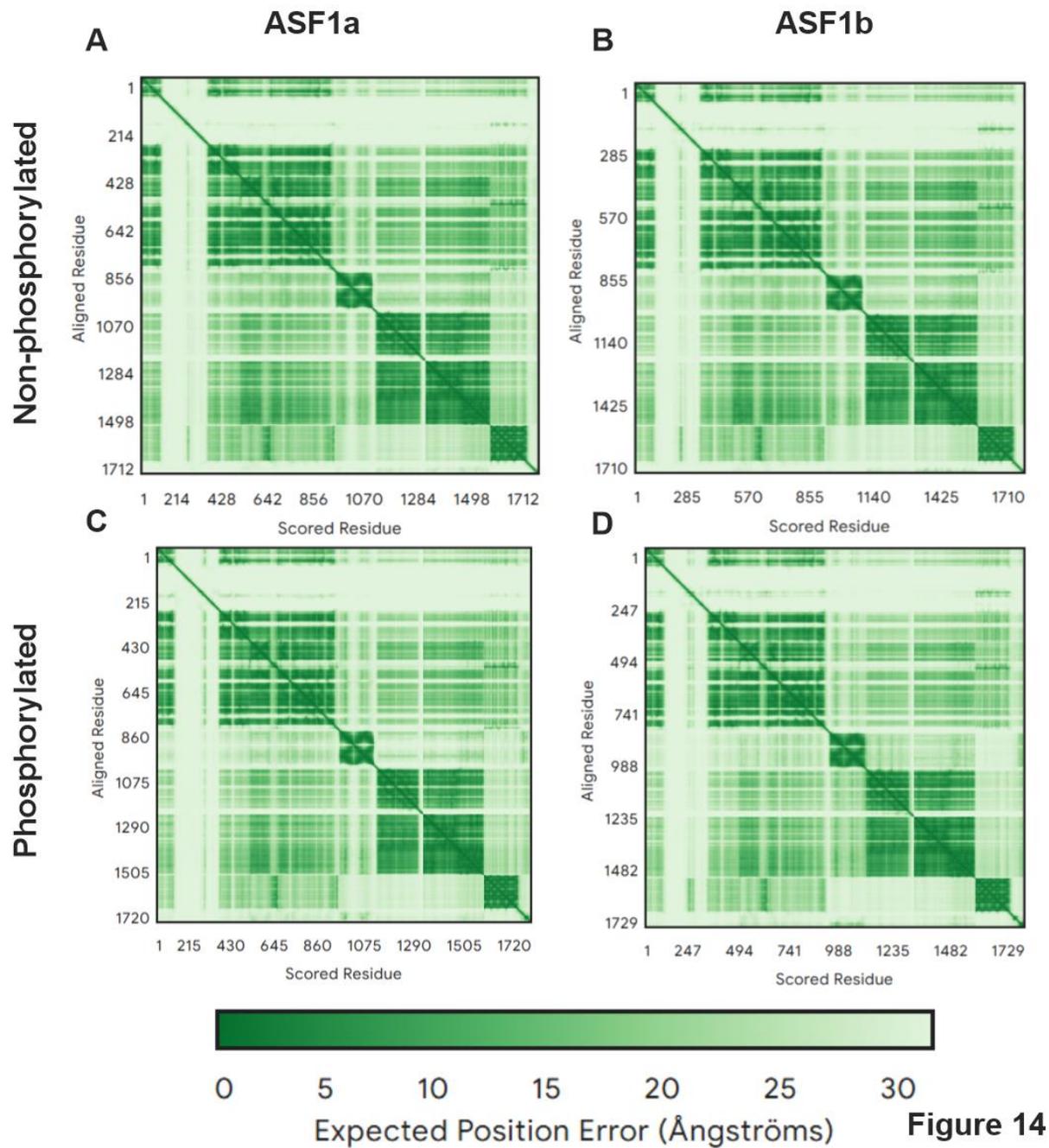
Figure 12



**Figure 3.5. Detail of ASF1 C-terminus within CDIN1.** Close association of ASF1 and CDIN1 facilitated by CDAN1.



**Figure 3.6. Specific residues implicated in ASF1-CDIN1 binding.** ChimeraX curated residues implicated in ASF1-CDIN1 binding are marked in violet. All interactions are at less than 15 angstroms.



**Figure 14**

**Figure 3.7. pLDDT scores for each residue in the ASF1-CDAN1-CDIN1 complex.** The ASF1-CDAN1-CDIN1 complex was generated with either ASF1a or ASF1b. Additionally, phosphorylated residues of both ASF1a and ASF1b were also assessed. Darker green indicates

higher confidence/lower anticipated spatial deviation from the model. Phosphorylation of both ASF1a and ASF1b increases positional confidence on select regions of the complex.

## Tables

	<b>ASF1a</b>	<b>pASF1a</b>	<b>ASF1b</b>	<b>ppASF1b</b>
<b>PTM</b>	0.61	0.58	0.60	0.57
<b>iPTM</b>	0.65	0.63	0.63	0.61

**Table 3.1. PTM and iPTM scores for the AlphaFold generated ASF1-CDAN1-CDIN1 complex.** ASF1 isoform participating in the complex indicated in top row, along with phosphorylation status. iPTM, or interface metric remains relatively constant between phosphorylated and non-phosphorylated status, whereas PTM decreases with phosphorylation, suggesting increased flexibility. All models exhibit moderate confidence scores, indicating a highly dynamic protein complex.

Isoform	Residue
ASF1a	192S: Phosphoserine
ASF1b	179T: Phosphothreonine 198S: Phosphoserine

**Table 3.2. Residues phosphorylated on ASF1 isoforms.** Residues phosphorylated in ASF1-CDAN1-CDIN1 AlphaFold modeling per ASF1 isoform participating in complex. Residues selected are most commonly post-translationally modified by tousel-like kinases.

## CHAPTER 4: CONCLUSIONS AND FUTURE DIRECTIONS

This dissertation simultaneously demonstrates the curative power of gene therapy while touching upon the deficiencies sometimes encountered in animal models. Furthermore, this dissertation indicates the emergence of artificial intelligence as a powerful tool to inform future rounds of experimental inquiry.

### **The Mouse as a Model Organism**

It is perhaps trivial to state that mice (*Mus musculus*), while extensively utilized simulacra of the human condition, are not humans. Mice and humans are separated by over 90 million years of evolution, a gulf of time great enough to surpass the creation of new continents and endure the occurrence of multiple mass extinction events including the demise of the non-avian dinosaurs. Indeed, the most parsimonious explanation of the utilization of mice is one of convivence and necessity. Mice are small, docile, and have a rapid reproductive lifespan that produces large litters of pups. Beginning in the early 20<sup>th</sup> century, scientists had revived from obscurity Gregor Mendel's laws of inheritance and sought to assess if Mendel's laws applied to animals and ultimately mammals.<sup>160</sup> Inspired by the inheritance patterns of Mendel's pea flowers and buoyed by the advent of fancy mouse breeding as a hobby in the United States, mice were assessed for inheritance in coat patterning, which indeed followed Mendel's laws of inheritance.<sup>161,162</sup> A major coup occurred when Abbie Lathrop, a fancy mouse breeder, intrigued by recurring skin lesions on one of her mouse lines, partnered with Leo Loeb of the University of Pennsylvania to investigate the phenomenon.<sup>163</sup> Confirming that the lesions were cancerous, Lathrop and Loeb began a collaboration that established that different lines of mice had differing rates of cancer.<sup>164</sup> This work, along with the work of Ernest Tyzzer of Harvard, the first

researcher to use the term ‘somatic mutation’ in 1916, established the genetic basis for cancer for the first time.<sup>165</sup>

To standardize their mouse strains, scientists sought to develop inbred mouse strains exhibiting little to no genetic variation. This level of back-breeding had been demonstrated before in plants, but never before in animals. It was here that Clarence Cook Little succeeded in producing the first inbred mouse strain, DBA (abbreviation of Dilute, Brown, Agouti) allowing for the transplantation of tumors between mice and eventually resulting in the discovery of the major histocompatibility complex.<sup>166</sup> Eventually Little would produce the ubiquitous C57BL/6, or Black-6, line of mouse commonly used in biomedical research across the globe. Advances in mouse experimentation would undergo another profound leap in 1988 with the advent of the knockout mouse, allowing for the direct demonstration of gene functionality via absence in a complex organism.<sup>167</sup> This technique would go on to win the 2007 Nobel Prize in Physiology or Medicine. Further technological developments would utilize recombinases to allow for tissue specific and temporal control of gene knockout in adult mice or during specific stages of embryonic development.<sup>168</sup> Lastly, gene knock-in techniques have been developed that allow for expression of human genes within mice, allowing for more faithful reproduction of human physiology.<sup>169</sup> Today there are thousands of bespoke knock-out and knock-in mouse lines.

### **Deficiencies of the Mouse Model Organism**

It is clear that mice, not only wild-type but knock-in and knock-out forms, have been important in the understanding of human disease and basic animal biology. However, in many cases mice fail to recapitulate the human condition. Metachromatic leukodystrophy (MLD) is a lysosomal storage disease with an autosomal recessive pattern of inheritance caused by

enzymatic deficiency in arylsulfatase A, causing a toxic accumulation of cerebroside-3-sulfate.<sup>170</sup> In patients, it manifests with progressive neurological degradation downstream of progressive demyelination of their central nervous system, often resulting in death by early childhood. Mice knocked out for *As2*, the murine arylsulfatase A orthologue, accumulate much of the same metabolic signs of human patients, with characteristic accumulation of sulfatides in the central nervous system and visceral organs. However, these animals do not exhibit the widespread demyelination seen in human patients and do not exhibit pronounced lethality.<sup>171</sup> *As2* KO mice do produce neuromotor defects, but they manifest at advanced age and are not nearly as severe as in humans. This has resulted in difficulties assessing potential treatments for MLD, including gene therapies, as behavioral assays for interventions are often noisy with a small statistical effect, pathological and biochemical assays must be used as proxies for efficacy. Cystic fibrosis (CF) is caused by loss-of-function mutations in the *CFTR* gene, encoding a gated anion channel.<sup>172</sup> These mutations result in anomalous viscosity in mucus membrane secretions, primarily affecting the lungs, pancreas, and intestines.<sup>173</sup> Patients experience progressive respiratory failure in conjunction with chronic lung infections, which are the primary causes for CF morbidity. CF mouse models fail to recapitulate the multi-organ pathology or lung infections encountered in the human population.<sup>174</sup> Simulations of opportunistic infection via infective challenge result in a high degree of variability and validity, as the utilized human specific opportunistic pathogens do not necessarily behave the same manner in a non-human host.<sup>175</sup> Alzheimer's disease (AD) research makes extensive use of mouse models of the disease. AD is a complex neurodegenerative disorder with an uncertain etiology.<sup>176</sup> Part of the difficulties in determining with a mechanistic certainty the nature of AD is the inability of mice to faithfully recapitulate the disease, even with supraphysiological overexpression of mutations implicated in

familial AD.<sup>177</sup> Furthermore, 90% of all drug candidates, extensively and exhaustively validated in-part by utilizing murine pre-clinical testing, fail in the clinical trial stage directly partially demonstrating the insufficiency of mouse-based animal modeling.<sup>178,179</sup> There are several efforts designed to replace the mouse as the standard biomedical model. One such proposal seeks to utilize the mouse lemur (*Microcebus murinus*) as a bridge between mice and humans.<sup>180</sup> Like the mouse it is small and has a rapid lifespan, but most importantly it is a primate. Difficulties adopting such a model include inertia and familiarity, as science has over a century of experience working with mice.

### **Mouse Modeling of CDA1a**

In our *Cdan1<sup>ALNP</sup>* model we indicate that human CDAN1 mutations do not cause a disease state in mice, at least upon cursory examination. Indeed, it is conceivable that mice, with their much lower blood volumes and higher RBC turnover rate, have already solved the problem of rapid cell proliferation coupled with genome organization and compaction as encountered in erythropoiesis.<sup>147</sup> Differences between human and mouse hematopoiesis should not be surprising given that one of the cell surface markers commonly utilized to enrich human HSCs, CD34, is not found on murine LT-HSCs at all. Also, with our *Cdan1<sup>AMx</sup>* model, we demonstrate difficulties in utilizing a knock-out protocol with a gene that can have functionality beyond the axis of pathology encountered in a specific disease manifestation. To wit: *Cdan1* KO resulted in gross aplasia rather than an erythroid-specific pathology in the *Cdan1<sup>AMx</sup>* model and a failure of engraftment within *Cdan1<sup>ALNP</sup>* mice. Our models provide an example how utilizing a gene knock-out does not necessarily reconstitute a disease state, controlling for species-level physiological differences. Perhaps obviously, gene knock-out does not emulate the presence of a

hypofunctional allele, as no function is not equivalent with partially diminished function as seen here with *Cdan1*.

To be clear, mouse models, despite any issues they might have in a given experimental or modeling context as discussed in the previous section, are still immensely useful tools in biological exploration. The researcher must keep the compromises endemic in the use of mouse models in mind and plan accordingly. In fact, there was significant uncertainty within our work of the ability of the human *CDAN1* orthologue to rescue the *Cdan1*<sup>ALNP</sup> mice prior to performing the first round of pilot transplant and transduction experiments. The noted differences between humans and mice encountered in our work have the potential to unveil additional aspects of both human and murine physiology.

### **The Curative Power of Gene Therapy**

We demonstrated the ability of gene therapy to rescue *Cdan1* knock-out animals from lethality. *Cdan1*<sup>ALNP</sup> mice treated with AP-CDAN1-FWP produce robust CBC values and full erythropoiesis. Furthermore, the therapy is effective through secondary transplantation, revealing reconstitution of full stemness within the LT-HSC population and full hematopoiesis. Some deviations upon histological inspection are noted but are almost certainly downstream from partial compatibility of the human CDAN1 orthologue and an artifact of the murine model system. This phenomenon requires more exploration.

In the absence of any recorded dominant-negative effects of CDAN1 mutations, our work validates the ability of a gene-addition based strategy for treatment of CDA1a and expands the envelop for this powerful treatment paradigm.

## Machine Learning as a Tool for Biological Exploration

Our work demonstrates the profound utility of ML in exploring the biological universe. As with most current ML applications, tools such as AlphaFold function the best as an adjunct to wet lab work, rather than a true replacement. While AlphaFold performs the best at predicting static structures resembling cryo-EM or x-ray crystallography, there are many emerging ML utilities that are capable of predicting dynamic, conformational changes in protein molecules, either in response to changes in environment, such as pH, or as a function of docking with another protein molecule.<sup>181-183</sup> There are even programs such as protein MPNN and ESMFold that can generate novel proteins with unique functions entirely *in silico*, including enzymes and binders useful for targeting epitopes of interest.<sup>184,185</sup>

A word of caution is that ML outputs, while exciting, enticing and plausible, will always require extensive wet lab experimentation to validate their findings. No matter how extensive the computational fluid dynamics simulations, or how well tuned the finite element analysis reconstructions are, a rocket must still be flown in multiple test campaigns before being brought into service. Indeed, most of the current ML utility is derived from informing future paths of research rather than acceptance of their outputs as ground truth. Machines are sublimely capable of finding the easiest path towards an output, regardless of the underlying reality. At this point, they do not understand the systems that they are trained on, rather they are optimizing a loss function to maximize a return. In some ways ML can be thought of as a genie, providing a wish-maker with what they were asking for in a cold and literal sense, focused through the Rorschach test of the user's biases and goals. However, this field is advancing so rapidly, by time this dissertation is deposited it will almost certainly be out of date.

## **Codanin-1 as a Central Axis of Cell Fate**

Based on our findings, we can speculate on the function of Cdan1 as a primary regulator of cell viability and cell fate, primarily through its interactions with ASF1 and CDIN1 as informed by AlphaFold in conjunction with the literature. Interestingly, while ASF1a is almost completely conserved between humans and mice, ASF1b differs more significantly between the species at the C-terminus, where regulation of ASF1b occurs via phosphorylation.<sup>154</sup> Furthermore, the ASF1a and ASF1b isoforms perform different functions, with ASF1b being more highly expressed in rapidly developing cell lineages.<sup>186</sup> Conceivably, the CDAN1 protein may act as a regulator of ASF1b phosphorylation by offering steric bulk and/or alternative binding to the ASF1b C-terminus. In our heterologous model system where human CDAN1 replaced its orthologue in mice, mild dysregulation of erythropoiesis might be due to CDAN1-mediated effects on ASF1b. In fact, there is additional evidence towards this line of thinking via a murine knock-out model of ASF1b. ASF1b-KO mice exhibit a slight anemia and increase in MCV, and are performing stress erythropoiesis in the spleen in an approximation of human CDA1a.<sup>149</sup> Furthermore, these animals continue to express fetal hemoglobin into adulthood, showing how ASF1, and by association, how CDAN1 is involved in the regulation of erythropoietic programming and cell fate.

## **Future Lines of Inquiry**

This dissertation shows that modeling of CDA1a, and the molecular function of CDAN1, is more complex than initially speculated and requires more experimentation to build of the preceding results.

***AlphaFold Validation Via Co-immunoprecipitation.*** Clearly, the AlphaFold results necessitate experimental validation. The first and simplest step would be a co-immunoprecipitation experiment designed to look for the ASF1-CDIN1 interaction. As the specificity of the protein-protein interactions directly informs much of the stability of the complex, especially the CDAN1-CDIN1 interaction being necessary for faithful CDIN1 folding and functionality, the co-IP experiment must include all three members of the complex, and cannot rely on yeast expression of only ASF1 and CDIN1, for example. Furthermore, ASF1 is known to contain many disordered regions, making the expression of excessively truncated versions of ASF1, for example expressing exclusively the ASF1 c-terminus equipped with a tag, difficult.

A path forward around this obstacle would be to simultaneously express the n-terminal region of ASF1 that is a known CDAN1 interactor and a heterologous protein, such as GFP, fused with the c-terminus of ASF1 (GFP-ASF1-c). In this manner the putative ASF1 c-terminal ASF1-CDIN1 interaction can be observed by pulling down with a GFP-specific antibody. If GFP-ASF1-c pulls down CDIN1 in complex, the AlphaFold results are verified. Varying salt concentrations can add a level of certainty by reducing transient and non-specific interactions. It is conceivable that this interaction is highly dynamic, and this could frustrate co-IP based attempts to resolve it. In such a case, alternative techniques such as FRET or surface plasmon resonance could establish interaction kinetics. Utilizing aspartic-acid or alanine substituted residues to generate phosphomimetic or phosho-inhibited versions of ASF1 in these contexts could also prove enlightening.

***Assessment of ASF1 phosphorylation in CDA1a.*** Even if the putative ASF1-CDAN1 interaction is not validated, it is conceivable that there are alterations in ASF1 phosphorylation in the presence of mutations in CDAN1. Intracellular localization of ASF1 is known to be perturbed in conjunction with the well-established CDAN1 R1042W variant, with excessive accumulations of ASF1 in the nucleus as compared to controls.<sup>103</sup> Given that ASF1 activity is regulated via phosphorylation, it is conceivable that altered interactions with CDAN1 can perturb the activity of the tousel-like kinases that phosphorylate ASF1, leading to anomalous phosphorylation levels and aberrant activity. A simple experiment utilizing a human erythroid cell line, such as K562, could be engineered to express CDAN1 R1042W or similar, then western blot performed on separated cytoplasmic and nuclear cell fractions. If this hypothesis is correct, aberrantly high levels of phosphorylated ASF1 should be seen, primarily in the nucleus.

***Xenotransplantation for Faithful Recapitulation of CDA1a.*** Xenotransplantation of immune-compromised mice with human marrow offers a path forward to reduce species-level differences in erythropoiesis and faithfully model CDA1a in an animal model. In brief, immune-compromised mice, which lack an adaptive immune system and thusly cannot mount a response to human cells as ‘non-self’ can be transplanted with human CD34<sup>+</sup> stem cells and repopulate the periphery with human blood lineages.<sup>187</sup> There are specialized transgenic models that express human cytokines and other human proteins, further enhancing the human hematopoietic facsimile. Indeed, MISTRG mice with a  $\text{Rag}^{-/-}$ ,  $\text{IL2R}\gamma^{-/-}$  background expressing from their native murine loci human M-CSF, IL3/GM-CSF, SIRP  $\alpha$ , and thrombopoietin have been used to successfully model myelodysplastic syndromes (MDS).<sup>188</sup> There is some debate regarding the accuracy of murine xenotransplant hematopoiesis, however.

Difficulties in utilizing this approach are primarily related to procuring human CD34<sup>+</sup> stem cells expressing CDA1a mutations. Patients with CDA1a are very rare, with an estimated prevalence of less than 0.75 per million live births.<sup>189</sup> Furthermore, the procurement of HSCs from these individuals in their chronic disease state is likely to be ethically problematic, except in rare cases when the patient is already scheduled for an allogenic bone marrow transplant to alleviate an extreme case of CDA1a. Genome editing provides a path forward, wherein CD34<sup>+</sup> cells can be obtained from healthy volunteer donors, then edited to contain CDA1a causing mutations.<sup>190</sup> A possible complication of this technique is the fact that no genome editing technique has an efficacy of 100%, meaning that CD34<sup>+</sup> cells transplanted would be a mixed population of healthy and CDA1a-diseased cells, potentially confounding results.

### **Final Thoughts**

In the end, it is our hope that the reader can derive some insights into the applications of viral vectors and machine learning in the biological sciences and some insights into CDA1a pathology and the power of gene therapy as a treatment paradigm.

## BIBLIOGRAPHY

1. Orkin, S.H., and Zon, L.I. (2008). Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. *Cell* 132, 631-644. 10.1016/j.cell.2008.01.025.
2. Schoedel, K.B., Morcos, M.N.F., Zerjatke, T., Roeder, I., Grinenko, T., Voehringer, D., Göthert, J.R., Waskow, C., Roers, A., and Gerbaulet, A. (2016). The bulk of the hematopoietic stem cell population is dispensable for murine steady-state and stress hematopoiesis. *Blood* 128, 2285-2296. 10.1182/blood-2016-03-706010.
3. Nishi, K., Sakamaki, T., Sadaoka, K., Fujii, M., Takaori-Kondo, A., Chen, J.Y., and Miyanishi, M. (2022). Identification of the minimum requirements for successful haematopoietic stem cell transplantation. *Br J Haematol* 196, 711-723. 10.1111/bjh.17867.
4. Osawa, M., Hanada, K., Hamada, H., and Nakauchi, H. (1996). Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science* 273, 242-245. 10.1126/science.273.5272.242.
5. Ema, H., Morita, Y., Yamazaki, S., Matsubara, A., Seita, J., Tadokoro, Y., Kondo, H., Takano, H., and Nakauchi, H. (2006). Adult mouse hematopoietic stem cells: purification and single-cell assays. *Nat Protoc* 1, 2979-2987. 10.1038/nprot.2006.447.
6. Eaves, C.J. (2015). Hematopoietic stem cells: concepts, definitions, and the new reality. *Blood* 125, 2605-2613. 10.1182/blood-2014-12-570200.
7. Pietras, E.M., Reynaud, D., Kang, Y.A., Carlin, D., Calero-Nieto, F.J., Leavitt, A.D., Stuart, J.M., Göttgens, B., and Passegué, E. (2015). Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions. *Cell Stem Cell* 17, 35-46. 10.1016/j.stem.2015.05.003.
8. Lu, R., Czechowicz, A., Seita, J., Jiang, D., and Weissman, I.L. (2019). Clonal-level lineage commitment pathways of hematopoietic stem cells in vivo. *Proceedings of the National Academy of Sciences* 116, 1447-1456. doi:10.1073/pnas.1801480116.
9. Higgins, J.M. (2015). Red blood cell population dynamics. *Clin Lab Med* 35, 43-57. 10.1016/j.cll.2014.10.002.
10. Bessis, M. (1958). [Erythroblastic island, functional unity of bone marrow]. *Rev Hematol* 13, 8-11.
11. Chasis, J.A., and Mohandas, N. (2008). Erythroblastic islands: niches for erythropoiesis. *Blood* 112, 470-478. 10.1182/blood-2008-03-077883.
12. Bessis, M.C., and Breton-Gorius, J. (1962). Iron metabolism in the bone marrow as seen by electron microscopy: a critical review. *Blood* 19, 635-663.
13. Yokoyama, T., Kitagawa, H., Takeuchi, T., Tsukahara, S., and Kannan, Y. (2002). No apoptotic cell death of erythroid cells of erythroblastic islands in bone marrow of healthy rats. *J Vet Med Sci* 64, 913-919. 10.1292/jvms.64.913.
14. Hanspal, M., and Hanspal, J.S. (1994). The association of erythroblasts with macrophages promotes erythroid proliferation and maturation: a 30-kD heparin-binding protein is involved in this contact. *Blood* 84, 3494-3504.
15. Hanspal, M., Smockova, Y., and Uong, Q. (1998). Molecular identification and functional characterization of a novel protein that mediates the attachment of erythroblasts to macrophages. *Blood* 92, 2940-2950.

16. Soni, S., Bala, S., Gwynn, B., Sahr, K.E., Peters, L.L., and Hanspal, M. (2006). Absence of erythroblast macrophage protein (Emp) leads to failure of erythroblast nuclear extrusion. *J Biol Chem* *281*, 20181-20189. 10.1074/jbc.M603226200.
17. Ginzburg, Y., An, X., Rivella, S., and Goldfarb, A. (2023). Normal and dysregulated crosstalk between iron metabolism and erythropoiesis. *Elife* *12*. 10.7554/eLife.90189.
18. Iscove, N.N., and Sieber, F. (1975). Erythroid progenitors in mouse bone marrow detected by macroscopic colony formation in culture. *Exp Hematol* *3*, 32-43.
19. Li, J., Hale, J., Bhagia, P., Xue, F., Chen, L., Jaffray, J., Yan, H., Lane, J., Gallagher, P.G., Mohandas, N., et al. (2014). Isolation and transcriptome analyses of human erythroid progenitors: BFU-E and CFU-E. *Blood* *124*, 3636-3645. 10.1182/blood-2014-07-588806.
20. Goodman, J.W., Hall, E.A., Miller, K.L., and Shinpock, S.G. (1985). Interleukin 3 promotes erythroid burst formation in "serum-free" cultures without detectable erythropoietin. *Proc Natl Acad Sci U S A* *82*, 3291-3295. 10.1073/pnas.82.10.3291.
21. Sui, X., Krantz, S.B., and Zhao, Z.J. (2000). Stem cell factor and erythropoietin inhibit apoptosis of human erythroid progenitor cells through different signalling pathways. *Br J Haematol* *110*, 63-70. 10.1046/j.1365-2141.2000.02145.x.
22. Broudy, V.C. (1997). Stem Cell Factor and Hematopoiesis. *Blood* *90*, 1345-1364. 10.1182/blood.V90.4.1345.
23. Rossi, P., Sette, C., Dolci, S., and Geremia, R. (2000). Role of c-kit in mammalian spermatogenesis. *J Endocrinol Invest* *23*, 609-615. 10.1007/bf03343784.
24. Wehrle-Haller, B. (2003). The role of Kit-ligand in melanocyte development and epidermal homeostasis. *Pigment Cell Res* *16*, 287-296. 10.1034/j.1600-0749.2003.00055.x.
25. Kobayashi, H., Liu, J., Urrutia, A.A., Burmakin, M., Ishii, K., Rajan, M., Davidoff, O., Saifudeen, Z., and Haase, V.H. (2017). Hypoxia-inducible factor prolyl-4-hydroxylation in FOXD1 lineage cells is essential for normal kidney development. *Kidney International* *92*, 1370-1383. 10.1016/j.kint.2017.06.015.
26. Anusornvongchai, T., Nangaku, M., Jao, T.-M., Wu, C.-H., Ishimoto, Y., Maekawa, H., Tanaka, T., Shimizu, A., Yamamoto, M., Suzuki, N., et al. (2018). Palmitate deranges erythropoietin production via transcription factor ATF4 activation of unfolded protein response. *Kidney International* *94*, 536-550. 10.1016/j.kint.2018.03.011.
27. Trinh, K.V., Diep, D., Chen, K.J.Q., Huang, L., and Gulenko, O. (2020). Effect of erythropoietin on athletic performance: a systematic review and meta-analysis. *BMJ Open Sport Exerc Med* *6*, e000716. 10.1136/bmjsem-2019-000716.
28. Atkinson, T.S., and Kahn, M.J. (2020). Blood doping: Then and now. A narrative review of the history, science and efficacy of blood doping in elite sport. *Blood Reviews* *39*, 100632. <https://doi.org/10.1016/j.blre.2019.100632>.
29. Hayat, A., Haria, D., and Salifu, M.O. (2008). Erythropoietin stimulating agents in the management of anemia of chronic kidney disease. *Patient Prefer Adherence* *2*, 195-200. 10.2147/ppa.s2356.
30. Broudy, V., Lin, N., Brice, M., Nakamoto, B., and Papayannopoulou, T. (1991). Erythropoietin receptor characteristics on primary human erythroid cells. *Blood* *77*, 2583-2590. 10.1182/blood.V77.12.2583.2583.
31. Udupa, K.B., and Reissmann, K.R. (1979). In vivo erythropoietin requirements of regenerating erythroid progenitors (BFU-e, CFU-e) in bone marrow of mice. *Blood* *53*, 1164-1171.

32. Mungamuri, S.K., and Ghaffari, S. (2008). Role of mTOR Signaling in Erythropoiesis. *Blood* *112*, 3870. <https://doi.org/10.1182/blood.V112.11.3870.3870>.
33. Peter, V., Guntram, B., Igor, T., Iris, Z.U., Ulrich, G., Reinhard, S., Karl, S., Wolfgang, F., Peter, B., Michael, P., et al. (2018). Normal and pathological erythropoiesis in adults: from gene regulation to targeted treatment concepts. *Haematologica* *103*, 1593-1603. 10.3324/haematol.2018.192518.
34. Gillespie, M.A., Palii, C.G., Sanchez-Taltavull, D., Shannon, P., Longabaugh, W.J.R., Downes, D.J., Sivaraman, K., Espinoza, H.M., Hughes, J.R., Price, N.D., et al. (2020). Absolute Quantification of Transcription Factors Reveals Principles of Gene Regulation in Erythropoiesis. *Molecular Cell* *78*, 960-974.e911. <https://doi.org/10.1016/j.molcel.2020.03.031>.
35. Valent, P., Büsche, G., Theurl, I., Uras, I.Z., Germing, U., Stauder, R., Sotlar, K., Füreder, W., Bettelheim, P., Pfeilstöcker, M., et al. (2018). Normal and pathological erythropoiesis in adults: from gene regulation to targeted treatment concepts. *Haematologica* *103*, 1593-1603. 10.3324/haematol.2018.192518.
36. Hu, J., Liu, J., Xue, F., Halverson, G., Reid, M., Guo, A., Chen, L., Raza, A., Galili, N., Jaffray, J., et al. (2013). Isolation and functional characterization of human erythroblasts at distinct stages: implications for understanding of normal and disordered erythropoiesis in vivo. *Blood* *121*, 3246-3253. 10.1182/blood-2013-01-476390.
37. Toda, S., Segawa, K., and Nagata, S. (2014). MerTK-mediated engulfment of pyrenocytes by central macrophages in erythroblastic islands. *Blood* *123*, 3963-3971. 10.1182/blood-2014-01-547976.
38. Chasis, J.A., Prenant, M., Leung, A., and Mohandas, N. (1989). Membrane assembly and remodeling during reticulocyte maturation. *Blood* *74*, 1112-1120.
39. Liu, J., Guo, X., Mohandas, N., Chasis, J.A., and An, X. (2010). Membrane remodeling during reticulocyte maturation. *Blood* *115*, 2021-2027. 10.1182/blood-2009-08-241182.
40. Waugh, R.E., McKenney, J.B., Bauserman, R.G., Brooks, D.M., Valeri, C.R., and Snyder, L.M. (1997). Surface area and volume changes during maturation of reticulocytes in the circulation of the baboon. *J Lab Clin Med* *129*, 527-535. 10.1016/s0022-2143(97)90007-x.
41. Kundu, M., Lindsten, T., Yang, C.Y., Wu, J., Zhao, F., Zhang, J., Selak, M.A., Ney, P.A., and Thompson, C.B. (2008). Ulk1 plays a critical role in the autophagic clearance of mitochondria and ribosomes during reticulocyte maturation. *Blood* *112*, 1493-1502. 10.1182/blood-2008-02-137398.
42. Gronowicz, G., Swift, H., and Steck, T.L. (1984). Maturation of the reticulocyte in vitro. *J Cell Sci* *71*, 177-197. 10.1242/jcs.71.1.177.
43. Ugo, V., Marzac, C., Teyssandier, I., Larbret, F., Lécluse, Y., Debili, N., Vainchenker, W., and Casadevall, N. (2004). Multiple signaling pathways are involved in erythropoietin-independent differentiation of erythroid progenitors in polycythemia vera. *Exp Hematol* *32*, 179-187. 10.1016/j.exphem.2003.11.003.
44. Gautier, E.F., Ducamp, S., Leduc, M., Salnot, V., Guillonnet, F., Dussiot, M., Hale, J., Giarratana, M.C., Raimbault, A., Douay, L., et al. (2016). Comprehensive Proteomic Analysis of Human Erythropoiesis. *Cell Rep* *16*, 1470-1484. 10.1016/j.celrep.2016.06.085.
45. Richard, C., and Verdier, F. (2020). Transferrin Receptors in Erythropoiesis. *Int J Mol Sci* *21*. 10.3390/ijms21249713.

46. Kautz, L., Jung, G., Valore, E.V., Rivella, S., Nemeth, E., and Ganz, T. (2014). Identification of erythroferrone as an erythroid regulator of iron metabolism. *Nat Genet* *46*, 678-684. 10.1038/ng.2996.
47. Aloj, G., Giardino, G., Valentino, L., Maio, F., Gallo, V., Esposito, T., Naddei, R., Cirillo, E., and Pignata, C. (2012). Severe combined immunodeficiencies: new and old scenarios. *Int Rev Immunol* *31*, 43-65. 10.3109/08830185.2011.644607.
48. Fischer, A., Notarangelo, L.D., Neven, B., Cavazzana, M., and Puck, J.M. (2015). Severe combined immunodeficiencies and related disorders. *Nature Reviews Disease Primers* *1*, 15061. 10.1038/nrdp.2015.61.
49. Giblett, E.R., Anderson, J.E., Cohen, F., Pollara, B., and Meuwissen, H.J. (1972). Adenosine-deaminase deficiency in two patients with severely impaired cellular immunity. *Lancet* *2*, 1067-1069. 10.1016/s0140-6736(72)92345-8.
50. Puck, J.M. (1996). IL2RGbase: a database of gamma c-chain defects causing human X-SCID. *Immunol Today* *17*, 507-511. 10.1016/0167-5699(96)30062-5.
51. Stamatoyannopoulos, G. (1972). The molecular basis of hemoglobin disease. *Annu Rev Genet* *6*, 47-70. 10.1146/annurev.ge.06.120172.000403.
52. Ballas, S.K., Kesen, M.R., Goldberg, M.F., Luty, G.A., Dampier, C., Osunkwo, I., Wang, W.C., Hoppe, C., Hagar, W., Darbari, D.S., and Malik, P. (2012). Beyond the definitions of the phenotypic complications of sickle cell disease: an update on management. *ScientificWorldJournal* *2012*, 949535. 10.1100/2012/949535.
53. Jang, T., Poplawska, M., Cimpeanu, E., Mo, G., Dutta, D., and Lim, S.H. (2021). Vaso-occlusive crisis in sickle cell disease: a vicious cycle of secondary events. *J Transl Med* *19*, 397. 10.1186/s12967-021-03074-z.
54. Musallam, K.M., Cappellini, M.D., Coates, T.D., Kuo, K.H.M., Al-Samkari, H., Sheth, S., Viprakasit, V., and Taher, A.T. (2024). Alpha-thalassemia: A practical overview. *Blood Reviews* *64*, 101165. <https://doi.org/10.1016/j.blre.2023.101165>.
55. Origa, R. (2017).  $\beta$ -Thalassemia. *Genetics in Medicine* *19*, 609-619. 10.1038/gim.2016.173.
56. (2014). In *Guidelines for the Management of Transfusion Dependent Thalassaemia (TDT)*, M.D. Cappellini, A. Cohen, J. Porter, A. Taher, and V. Viprakasit, eds. (Thalassaemia International Federation © 2014 Thalassaemia International Federation.).
57. Alter, B.P., Baerlocher, G.M., Savage, S.A., Chanock, S.J., Weksler, B.B., Willner, J.P., Peters, J.A., Giri, N., and Lansdorp, P.M. (2007). Very short telomere length by flow fluorescence in situ hybridization identifies patients with dyskeratosis congenita. *Blood* *110*, 1439-1447. 10.1182/blood-2007-02-075598.
58. Vulliamy, T., Marrone, A., Szydlo, R., Walne, A., Mason, P.J., and Dokal, I. (2004). Disease anticipation is associated with progressive telomere shortening in families with dyskeratosis congenita due to mutations in TERC. *Nat Genet* *36*, 447-449. 10.1038/ng1346.
59. Hoover, A., Turcotte, L.M., Phelan, R., Barbus, C., Rayannavar, A., Miller, B.S., Reardon, E.E., Theis-Mahon, N., and MacMillan, M.L. (2024). Longitudinal clinical manifestations of Fanconi anemia: A systematized review. *Blood Reviews* *68*, 101225. <https://doi.org/10.1016/j.blre.2024.101225>.
60. Rosenberg, P.S., Greene, M.H., and Alter, B.P. (2003). Cancer incidence in persons with Fanconi anemia. *Blood* *101*, 822-826. 10.1182/blood-2002-05-1498.

61. Alter, B.P., Giri, N., Savage, S.A., and Rosenberg, P.S. (2009). Cancer in dyskeratosis congenita. *Blood* *113*, 6549-6557. 10.1182/blood-2008-12-192880.
62. Sayed, N., Allawadhi, P., Khurana, A., Singh, V., Navik, U., Pasumarthi, S.K., Khurana, I., Banothu, A.K., Weiskirchen, R., and Bharani, K.K. (2022). Gene therapy: Comprehensive overview and therapeutic applications. *Life Sci* *294*, 120375. 10.1016/j.lfs.2022.120375.
63. Cline, M.J. (1985). Perspectives for gene therapy: Inserting new genetic information into mammalian cells by physical techniques and viral vectors. *Pharmacology & Therapeutics* *29*, 69-92. [https://doi.org/10.1016/0163-7258\(85\)90017-8](https://doi.org/10.1016/0163-7258(85)90017-8).
64. Bulcha, J.T., Wang, Y., Ma, H., Tai, P.W.L., and Gao, G. (2021). Viral vector platforms within the gene therapy landscape. *Signal Transduction and Targeted Therapy* *6*, 53. 10.1038/s41392-021-00487-6.
65. Wang, J.-H., Gessler, D.J., Zhan, W., Gallagher, T.L., and Gao, G. (2024). Adeno-associated virus as a delivery vector for gene therapy of human diseases. *Signal Transduction and Targeted Therapy* *9*, 78. 10.1038/s41392-024-01780-w.
66. Nathwani, A.C., Tuddenham, E.G., Rangarajan, S., Rosales, C., McIntosh, J., Linch, D.C., Chowdary, P., Riddell, A., Pie, A.J., Harrington, C., et al. (2011). Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. *N Engl J Med* *365*, 2357-2365. 10.1056/NEJMoal108046.
67. Wu, Z., Yang, H., and Colosi, P. (2010). Effect of Genome Size on AAV Vector Packaging. *Molecular Therapy* *18*, 80-86. 10.1038/mt.2009.255.
68. Wilson, J.M. (2009). Lessons learned from the gene therapy trial for ornithine transcarbamylase deficiency. *Molecular Genetics and Metabolism* *96*, 151-157. <https://doi.org/10.1016/j.ymgme.2008.12.016>.
69. Hacein-Bey-Abina, S., Garrigue, A., Wang, G.P., Soulier, J., Lim, A., Morillon, E., Clappier, E., Caccavelli, L., Delabesse, E., Beldjord, K., et al. (2008). Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J Clin Invest* *118*, 3132-3142. 10.1172/jci35700.
70. Freisen, D., and Miller, L. (2001). *Insect viruses Fields' Virology 4th edn ed DM Knipe et al.* Philadelphia, PA: Lippincott Williams and Wilkins.
71. Taberero, J., Shapiro, G.I., LoRusso, P.M., Cervantes, A., Schwartz, G.K., Weiss, G.J., Paz-Ares, L., Cho, D.C., Infante, J.R., Alsina, M., et al. (2013). First-in-humans trial of an RNA interference therapeutic targeting VEGF and KSP in cancer patients with liver involvement. *Cancer Discov* *3*, 406-417. 10.1158/2159-8290.Cd-12-0429.
72. Hajj, K.A., and Whitehead, K.A. (2017). Tools for translation: non-viral materials for therapeutic mRNA delivery. *Nature Reviews Materials* *2*, 17056. 10.1038/natrevmats.2017.56.
73. Baden, L.R., El Sahly, H.M., Essink, B., Kotloff, K., Frey, S., Novak, R., Diemert, D., Spector, S.A., Rouphael, N., Creech, C.B., et al. (2021). Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *N Engl J Med* *384*, 403-416. 10.1056/NEJMoal2035389.
74. Maguire, A.M., Bennett, J., Aleman, E.M., Leroy, B.P., and Aleman, T.S. (2021). Clinical Perspective: Treating RPE65-Associated Retinal Dystrophy. *Mol Ther* *29*, 442-463. 10.1016/j.ymthe.2020.11.029.
75. Al-Zaidy, S., Pickard, A.S., Kotha, K., Alfano, L.N., Lowes, L., Paul, G., Church, K., Lehman, K., Sproule, D.M., Dabbous, O., et al. (2019). Health outcomes in spinal muscular

- atrophy type 1 following AVXS-101 gene replacement therapy. *Pediatr Pulmonol* 54, 179-185. 10.1002/ppul.24203.
76. Coppens, M., Pipe, S.W., Miesbach, W., Astermark, J., Recht, M., van der Valk, P., Ewenstein, B., Pinachyan, K., Galante, N., Le Quellec, S., et al. (2024). Etranacogene dezaparvovec gene therapy for haemophilia B (HOPE-B): 24-month post-hoc efficacy and safety data from a single-arm, multicentre, phase 3 trial. *The Lancet Haematology* 11, e265-e275. 10.1016/S2352-3026(24)00006-1.
  77. Lancaster, V., Richardson, M., Beaudoin, F.L., Synnott, P.G., Rind, D.M., Herce-Hagiwara, B., Campbell, J.D., and Pearson, S.D. (2022). The effectiveness and value of betibeglogene autotemcel for the management of transfusion-dependent beta-thalassemia. *J Manag Care Spec Pharm* 28, 1316-1320. 10.18553/jmcp.2022.28.11.1316.
  78. Magrin, E., Semeraro, M., Hebert, N., Joseph, L., Magnani, A., Chalumeau, A., Gabrion, A., Roudaut, C., Marouene, J., Lefrere, F., et al. (2022). Long-term outcomes of lentiviral gene therapy for the  $\beta$ -hemoglobinopathies: the HGB-205 trial. *Nat Med* 28, 81-88. 10.1038/s41591-021-01650-w.
  79. Zheng, J., Wei, Y., and Han, G.Z. (2022). The diversity and evolution of retroviruses: Perspectives from viral "fossils". *Virol Sin* 37, 11-18. 10.1016/j.virs.2022.01.019.
  80. Roe, T., Reynolds, T.C., Yu, G., and Brown, P.O. (1993). Integration of murine leukemia virus DNA depends on mitosis. *Embo j* 12, 2099-2108. 10.1002/j.1460-2075.1993.tb05858.x.
  81. Lohse, N., Hansen, A.B., Pedersen, G., Kronborg, G., Gerstoft, J., Sørensen, H.T., Vaeth, M., and Obel, N. (2007). Survival of persons with and without HIV infection in Denmark, 1995-2005. *Ann Intern Med* 146, 87-95. 10.7326/0003-4819-146-2-200701160-00003.
  82. Durand, S., and Cimarelli, A. (2011). The inside out of lentiviral vectors. *Viruses* 3, 132-159. 10.3390/v3020132.
  83. Moore, J.P., Trkola, A., and Dragic, T. (1997). Co-receptors for HIV-1 entry. *Current Opinion in Immunology* 9, 551-562. [https://doi.org/10.1016/S0952-7915\(97\)80110-0](https://doi.org/10.1016/S0952-7915(97)80110-0).
  84. Kondo, N., Miyauchi, K., Meng, F., Iwamoto, A., and Matsuda, Z. (2010). Conformational Changes of the HIV-1 Envelope Protein during Membrane Fusion Are Inhibited by the Replacement of Its Membrane-spanning Domain\*. *Journal of Biological Chemistry* 285, 14681-14688. <https://doi.org/10.1074/jbc.M109.067090>.
  85. Bukrinsky, M.I., Sharova, N., McDonald, T.L., Pushkarskaya, T., Tarpley, W.G., and Stevenson, M. (1993). Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection. *Proceedings of the National Academy of Sciences* 90, 6125-6129. 10.1073/pnas.90.13.6125.
  86. Miller, M.D., Wang, B., and Bushman, F.D. (1995). Human immunodeficiency virus type 1 preintegration complexes containing discontinuous plus strands are competent to integrate in vitro. *J Virol* 69, 3938-3944. 10.1128/jvi.69.6.3938-3944.1995.
  87. Sundquist, W.I., and Kräusslich, H.G. (2012). HIV-1 assembly, budding, and maturation. *Cold Spring Harb Perspect Med* 2, a006924. 10.1101/cshperspect.a006924.
  88. Lewis, P., Hensel, M., and Emerman, M. (1992). Human immunodeficiency virus infection of cells arrested in the cell cycle. *Embo j* 11, 3053-3058. 10.1002/j.1460-2075.1992.tb05376.x.

89. Shao, L., Shi, R., Zhao, Y., Liu, H., Lu, A., Ma, J., Cai, Y., Fuksenko, T., Pelayo, A., Shah, N.N., et al. (2022). Genome-wide profiling of retroviral DNA integration and its effect on clinical pre-infusion CAR T-cell products. *Journal of Translational Medicine* 20, 514. 10.1186/s12967-022-03729-5.
90. Dull, T., Zufferey, R., Kelly, M., Mandel, R.J., Nguyen, M., Trono, D., and Naldini, L. (1998). A third-generation lentivirus vector with a conditional packaging system. *J Virol* 72, 8463-8471. 10.1128/jvi.72.11.8463-8471.1998.
91. Sweeney, N.P., and Vink, C.A. (2021). The impact of lentiviral vector genome size and producer cell genomic to gag-pol mRNA ratios on packaging efficiency and titre. *Molecular Therapy Methods & Clinical Development* 21, 574-584. 10.1016/j.omtm.2021.04.007.
92. Schaubert-Plewa, C., Simmons, A., Tuerk, M.J., Pacheco, C.D., and Veres, G. (2005). Complement regulatory proteins are incorporated into lentiviral vectors and protect particles against complement inactivation. *Gene Ther* 12, 238-245. 10.1038/sj.gt.3302399.
93. Tucci, F., Scaramuzza, S., Aiuti, A., and Mortellaro, A. (2021). Update on Clinical *Ex Vivo* Hematopoietic Stem Cell Gene Therapy for Inherited Monogenic Diseases. *Molecular Therapy* 29, 489-504. 10.1016/j.ymthe.2020.11.020.
94. Dreger, P., Haferlach, T., Eckstein, V., Jacobs, S., Suttorp, M., Löuffler, H., And, W.M.-R., and Schmitz, N. (1994). G-CSF-mobilized peripheral blood progenitor cells for allogeneic transplantation: safety, kinetics of mobilization, and composition of the graft. *British Journal of Haematology* 87, 609-613. <https://doi.org/10.1111/j.1365-2141.1994.tb08321.x>.
95. McKinnon, K.M. (2018). Flow Cytometry: An Overview. *Curr Protoc Immunol* 120, 5.1.1-5.1.11. 10.1002/cpim.40.
96. Cedar, H., and Bergman, Y. (2011). Epigenetics of haematopoietic cell development. *Nature Reviews Immunology* 11, 478-488. 10.1038/nri2991.
97. Hattangadi, S.M., Wong, P., Zhang, L., Flygare, J., and Lodish, H.F. (2011). From stem cell to red cell: regulation of erythropoiesis at multiple levels by multiple proteins, RNAs, and chromatin modifications. *Blood* 118, 6258-6268. <https://doi.org/10.1182/blood-2011-07-356006>.
98. Milone, M.C., and O'Doherty, U. (2018). Clinical use of lentiviral vectors. *Leukemia* 32, 1529-1541. 10.1038/s41375-018-0106-0.
99. Iolascon, A., Andolfo, I., and Russo, R. (2020). Congenital dyserythropoietic anemias. *Blood* 136, 1274-1283. 10.1182/blood.2019000948.
100. King, R., Gallagher, P.J., and Khoriaty, R. (2022). The congenital dyserythropoietic anemias: genetics and pathophysiology. *Curr Opin Hematol* 29, 126-136. 10.1097/moh.0000000000000697.
101. Iolascon, A., Heimpel, H., Wahlin, A., and Tamary, H. (2013). Congenital dyserythropoietic anemias: molecular insights and diagnostic approach. *Blood* 122, 2162-2166. 10.1182/blood-2013-05-468223.
102. Dgany, O., Avidan, N., Delaunay, J., Krasnov, T., Shalmon, L., Shalev, H., Eidelitz-Markus, T., Kapelushnik, J., Cattan, D., Pariente, A., et al. (2002). Congenital dyserythropoietic anemia type I is caused by mutations in codanin-1. *Am J Hum Genet* 71, 1467-1474. 10.1086/344781.

103. Ask, K., Jasencakova, Z., Menard, P., Feng, Y., Almouzni, G., and Groth, A. (2012). Codanin-1, mutated in the anaemic disease CDAI, regulates Asf1 function in S-phase histone supply. *Embo j* 31, 2013-2023. 10.1038/emboj.2012.55.
104. Tamary, H., Dgany, O., Proust, A., Krasnov, T., Avidan, N., Eidelitz-Markus, T., Tchernia, G., Geneviève, D., Cormier-Daire, V., Bader-Meunier, B., et al. (2005). Clinical and molecular variability in congenital dyserythropoietic anaemia type I. *Br J Haematol* 130, 628-634. 10.1111/j.1365-2141.2005.05642.x.
105. Heimpel, H., Kellermann, K., Neuschwander, N., Högel, J., and Schwarz, K. (2010). The morphological diagnosis of congenital dyserythropoietic anemia: results of a quantitative analysis of peripheral blood and bone marrow cells. *Haematologica* 95, 1034-1036. 10.3324/haematol.2009.014563.
106. Roy, N.B.A., and Babbs, C. (2019). The pathogenesis, diagnosis and management of congenital dyserythropoietic anaemia type I. *Br J Haematol* 185, 436-449. 10.1111/bjh.15817.
107. Marwaha, R.K., Bansal, D., Trehan, A., and Garewal, G. (2005). Interferon therapy in congenital dyserythropoietic anemia type I/II. *Pediatr Hematol Oncol* 22, 133-138. 10.1080/08880010590907221.
108. Renella, R., Roberts, N.A., Brown, J.M., De Gobbi, M., Bird, L.E., Hassanali, T., Sharpe, J.A., Sloane-Stanley, J., Ferguson, D.J., Cordell, J., et al. (2011). Codanin-1 mutations in congenital dyserythropoietic anemia type 1 affect HP1 {alpha} localization in erythroblasts. *Blood* 117, 6928-6938. 10.1182/blood-2010-09-308478.
109. Noy-Lotan, S., Dgany, O., Marcoux, N., Atkins, A., Kupfer, G.M., Bosques, L., Gottschalk, C., Steinberg-Shemer, O., Motro, B., and Tamary, H. (2021). Cdan1 Is Essential for Primitive Erythropoiesis. *Front Physiol* 12, 685242. 10.3389/fphys.2021.685242.
110. Kühn, R., Schwenk, F., Aguet, M., and Rajewsky, K. (1995). Inducible gene targeting in mice. *Science* 269, 1427-1429. 10.1126/science.7660125.
111. Chappell, M.E., Breda, L., Tricoli, L., Guerra, A., Jarocha, D., Castruccio Castracani, C., Papp, T.E., Tanaka, N., Hamilton, N., Triebwasser, M.P., et al. (2024). Use of HSC-targeted LNP to generate a mouse model of lethal  $\alpha$ -thalassemia and treatment via lentiviral gene therapy. *Blood* 144, 1633-1645. 10.1182/blood.2023023349.
112. Castruccio Castracani, C., Breda, L., Papp, T.E., Guerra, A., Radaelli, E., Assenmacher, C.A., Finesso, G., Mui, B.L., Tam, Y.K., Fontana, S., et al. (2024). An erythroid-specific lentiviral vector improves anemia and iron metabolism in a new model of XLSA. *Blood*. 10.1182/blood.2024025846.
113. Schröder, M., and Bowie, A.G. (2005). TLR3 in antiviral immunity: key player or bystander? *Trends Immunol* 26, 462-468. 10.1016/j.it.2005.07.002.
114. Pillay, J., den Braber, I., Vrisekoop, N., Kwast, L.M., de Boer, R.J., Borghans, J.A., Tesselaar, K., and Koenderman, L. (2010). In vivo labeling with 2H2O reveals a human neutrophil lifespan of 5.4 days. *Blood* 116, 625-627. 10.1182/blood-2010-01-259028.
115. Liu, J., Zhang, J., Ginzburg, Y., Li, H., Xue, F., De Franceschi, L., Chasis, J.A., Mohandas, N., and An, X. (2013). Quantitative analysis of murine terminal erythroid differentiation in vivo: novel method to study normal and disordered erythropoiesis. *Blood* 121, e43-49. 10.1182/blood-2012-09-456079.

116. Velasco-Hernandez, T., Säwén, P., Bryder, D., and Cammenga, J. (2016). Potential Pitfalls of the Mx1-Cre System: Implications for Experimental Modeling of Normal and Malignant Hematopoiesis. *Stem Cell Reports* 7, 11-18. 10.1016/j.stemcr.2016.06.002.
117. Joseph, C., Quach, J.M., Walkley, C.R., Lane, S.W., Lo Celso, C., and Purton, L.E. (2013). Deciphering hematopoietic stem cells in their niches: a critical appraisal of genetic models, lineage tracing, and imaging strategies. *Cell Stem Cell* 13, 520-533. 10.1016/j.stem.2013.10.010.
118. Ulyanova, T., Priestley, G.V., Nakamoto, B., Jiang, Y., and Papayannopoulou, T. (2007). VCAM-1 ablation in nonhematopoietic cells in MxCre<sup>+</sup> VCAM-1<sup>f/f</sup> mice is variable and dictates their phenotype. *Exp Hematol* 35, 565-571. 10.1016/j.exphem.2007.01.031.
119. Czechowicz, A., Palchaudhuri, R., Scheck, A., Hu, Y., Hoggatt, J., Saez, B., Pang, W.W., Mansour, M.K., Tate, T.A., Chan, Y.Y., et al. (2019). Selective hematopoietic stem cell ablation using CD117-antibody-drug-conjugates enables safe and effective transplantation with immunity preservation. *Nat Commun* 10, 617. 10.1038/s41467-018-08201-x.
120. Breda, L., Papp, T.E., Triebwasser, M.P., Yadegari, A., Fedorky, M.T., Tanaka, N., Abdulmalik, O., Pavani, G., Wang, Y., Grupp, S.A., et al. (2023). In vivo hematopoietic stem cell modification by mRNA delivery. *Science* 381, 436-443. 10.1126/science.ade6967.
121. Romero, Z., Campo-Fernandez, B., Wherley, J., Kaufman, M.L., Urbinati, F., Cooper, A.R., Hoban, M.D., Baldwin, K.M., Lumaquin, D., Wang, X., et al. (2015). The human ankyrin 1 promoter insulator sustains gene expression in a  $\beta$ -globin lentiviral vector in hematopoietic stem cells. *Mol Ther Methods Clin Dev* 2, 15012. 10.1038/mtm.2015.12.
122. Burgess, D.J. (2015). Gene therapy: insulating from genotoxicity. *Nat Rev Genet* 16, 130-131. 10.1038/nrg3904.
123. Browning, D.L., and Trobridge, G.D. (2016). Insulators to Improve the Safety of Retroviral Vectors for HIV Gene Therapy. *Biomedicines* 4. 10.3390/biomedicines4010004.
124. Goodman, M.A., Arumugam, P., Pillis, D.M., Loberg, A., Nasimuzzaman, M., Lynn, D., van der Loo, J.C.M., Dexheimer, P.J., Keddache, M., Bauer, T.R., Jr., et al. (2018). Foamy Virus Vector Carries a Strong Insulator in Its Long Terminal Repeat Which Reduces Its Genotoxic Potential. *J Virol* 92. 10.1128/jvi.01639-17.
125. Zufferey, R., Donello, J.E., Trono, D., and Hope, T.J. (1999). Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors. *J Virol* 73, 2886-2892. 10.1128/jvi.73.4.2886-2892.1999.
126. Lin, H.T., Okumura, T., Yatsuda, Y., Ito, S., Nakauchi, H., and Otsu, M. (2016). Application of Droplet Digital PCR for Estimating Vector Copy Number States in Stem Cell Gene Therapy. *Hum Gene Ther Methods* 27, 197-208. 10.1089/hgtb.2016.059.
127. Sarma, N.J., Takeda, A., and Yaseen, N.R. (2010). Colony forming cell (CFC) assay for human hematopoietic cells. *J Vis Exp*. 10.3791/2195.
128. Krause, D.S., Theise, N.D., Collector, M.I., Henegariu, O., Hwang, S., Gardner, R., Neutzel, S., and Sharkis, S.J. (2001). Multi-organ, multi-lineage engraftment by a single bone marrow-derived stem cell. *Cell* 105, 369-377. 10.1016/s0092-8674(01)00328-2.
129. Challen, G.A., Pietras, E.M., Wallscheid, N.C., and Signer, R.A.J. (2021). Simplified murine multipotent progenitor isolation scheme: Establishing a consensus approach for multipotent progenitor identification. *Exp Hematol* 104, 55-63. 10.1016/j.exphem.2021.09.007.

130. Mezmarich, J.A., Draper, L., Christensen, R.D., Yaish, H.M., Luem, N.D., Pysker, T.J., Jung, G., Nemeth, E., Ganz, T., and Ward, D.M. (2018). Fetal presentation of congenital dyserythropoietic anemia type 1 with novel compound heterozygous CDAN1 mutations. *Blood Cells Mol Dis* 71, 63-66. 10.1016/j.bcmd.2018.03.002.
131. Shalev, H., Kapelushnik, J., Moser, A., Dgany, O., Krasnov, T., and Tamary, H. (2004). A comprehensive study of the neonatal manifestations of congenital dyserythropoietic anemia type I. *J Pediatr Hematol Oncol* 26, 746-748. 10.1097/00043426-200411000-00011.
132. Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. *Science* 347, 1260419. 10.1126/science.1260419.
133. Noy-Lotan, S., Dgany, O., Lahmi, R., Marcoux, N., Krasnov, T., Yissachar, N., Ginsberg, D., Motro, B., Resnitzky, P., Yaniv, I., et al. (2009). Codanin-1, the protein encoded by the gene mutated in congenital dyserythropoietic anemia type I (CDAN1), is cell cycle-regulated. *Haematologica* 94, 629-637. 10.3324/haematol.2008.003327.
134. Banjac, I., Maimets, M., and Jensen, K.B. (2023). Maintenance of high-turnover tissues during and beyond homeostasis. *Cell Stem Cell* 30, 348-361. 10.1016/j.stem.2023.03.008.
135. Kotredes, K.P., Thomas, B., and Gamero, A.M. (2017). The Protective Role of Type I Interferons in the Gastrointestinal Tract. *Front Immunol* 8, 410. 10.3389/fimmu.2017.00410.
136. Small, A.M., Peloso, G.M., Linefsky, J., Aragam, J., Galloway, A., Tanukonda, V., Wang, L.C., Yu, Z., Sunitha Selvaraj, M., Farber-Eger, E.H., et al. (2023). Multiancestry Genome-Wide Association Study of Aortic Stenosis Identifies Multiple Novel Loci in the Million Veteran Program. *Circulation* 147, 942-955. 10.1161/circulationaha.122.061451.
137. Olijnik, A.A., Roy, N.B.A., Scott, C., Marsh, J.A., Brown, J., Lauschke, K., Ask, K., Roberts, N., Downes, D.J., Brolih, S., et al. (2021). Genetic and functional insights into CDA-I prevalence and pathogenesis. *J Med Genet* 58, 185-195. 10.1136/jmedgenet-2020-106880.
138. Andrieu-Soler, C., and Soler, E. (2022). Erythroid Cell Research: 3D Chromatin, Transcription Factors and Beyond. *Int J Mol Sci* 23. 10.3390/ijms23116149.
139. Ludwig, L.S., Lareau, C.A., Bao, E.L., Nandakumar, S.K., Muus, C., Ulirsch, J.C., Chowdhary, K., Buenrostro, J.D., Mohandas, N., An, X., et al. (2019). Transcriptional States and Chromatin Accessibility Underlying Human Erythropoiesis. *Cell Rep* 27, 3228-3240.e3227. 10.1016/j.celrep.2019.05.046.
140. Kreto, D.A., Folkes, L., Mora-Martin, A., Walawalkar, I.A., Imrat, Syedah, N., Vanuytsel, K., Moxon, S., Murphy, G.J., and Cifuentes, D. (2024). The miR-144/Hmgn2 regulatory axis orchestrates chromatin organization during erythropoiesis. *Nat Commun* 15, 3821. 10.1038/s41467-024-47982-2.
141. Li, D., Zhao, X.Y., Zhou, S., Hu, Q., Wu, F., and Lee, H.Y. (2023). Multidimensional profiling reveals GATA1-modulated stage-specific chromatin states and functional associations during human erythropoiesis. *Nucleic Acids Res* 51, 6634-6653. 10.1093/nar/gkad468.
142. Galán-Díez, M., Cuesta-Domínguez, Á., and Kousteni, S. (2018). The Bone Marrow Microenvironment in Health and Myeloid Malignancy. *Cold Spring Harb Perspect Med* 8. 10.1101/cshperspect.a031328.

143. Fröbel, J., Landspersky, T., Percin, G., Schreck, C., Rahmig, S., Ori, A., Nowak, D., Essers, M., Waskow, C., and Oostendorp, R.A.J. (2021). The Hematopoietic Bone Marrow Niche Ecosystem. *Front Cell Dev Biol* 9, 705410. 10.3389/fcell.2021.705410.
144. Beckmann, J., Scheitza, S., Wernet, P., Fischer, J.C., and Giebel, B. (2007). Asymmetric cell division within the human hematopoietic stem and progenitor cell compartment: identification of asymmetrically segregating proteins. *Blood* 109, 5494-5501. 10.1182/blood-2006-11-055921.
145. Brummendorf, T.H., Dragowska, W., Zijlmans, J., Thornbury, G., and Lansdorp, P.M. (1998). Asymmetric cell divisions sustain long-term hematopoiesis from single-sorted human fetal liver cells. *J Exp Med* 188, 1117-1124. 10.1084/jem.188.6.1117.
146. Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343, 80-84. 10.1126/science.1246981.
147. Dholakia, U., Bandyopadhyay, S., Hod, E.A., and Prestia, K.A. (2015). Determination of RBC Survival in C57BL/6 and C57BL/6-Tg(UBC-GFP) Mice. *Comp Med* 65, 196-201.
148. Vilarrasa-Blasi, R., Soler-Vila, P., Verdaguer-Dot, N., Russiñol, N., Di Stefano, M., Chapaprieta, V., Clot, G., Farabella, I., Cuscó, P., Kulis, M., et al. (2021). Dynamics of genome architecture and chromatin function during human B cell differentiation and neoplastic transformation. *Nat Commun* 12, 651. 10.1038/s41467-020-20849-y.
149. Papadopoulos, P., Kafasi, A., De Cuyper, I.M., Barroca, V., Lewandowski, D., Kadri, Z., Veldhuis, M., Berghuis, J., Gillemans, N., Benavente Cuesta, C.M., et al. (2020). Mild dyserythropoiesis and  $\beta$ -like globin gene expression imbalance due to the loss of histone chaperone ASF1B. *Hum Genomics* 14, 39. 10.1186/s40246-020-00283-3.
150. Akgol Oksuz, B., Yang, L., Abraham, S., Venev, S.V., Krietenstein, N., Parsi, K.M., Ozadam, H., Oomen, M.E., Nand, A., Mao, H., et al. (2021). Systematic evaluation of chromosome conformation capture assays. *Nat Methods* 18, 1046-1055. 10.1038/s41592-021-01248-7.
151. Winick-Ng, W., Kukalev, A., Harabula, I., Zea-Redondo, L., Szabó, D., Meijer, M., Serebreni, L., Zhang, Y., Bianco, S., Chiariello, A.M., et al. (2021). Cell-type specialization is encoded by specific chromatin topologies. *Nature* 599, 684-691. 10.1038/s41586-021-04081-2.
152. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630, 493-500. 10.1038/s41586-024-07487-w.
153. Murphy, Z.C., Getman, M.R., Myers, J.A., Burgos Villar, K.N., Leshen, E., Kurita, R., Nakamura, Y., and Steiner, L.A. (2020). Codanin-1 mutations engineered in human erythroid cells demonstrate role of CDAN1 in terminal erythroid maturation. *Exp Hematol* 91, 32-38.e36. 10.1016/j.exphem.2020.09.201.
154. Simon, B., Lou, H.J., Huet-Calderwood, C., Shi, G., Boggon, T.J., Turk, B.E., and Calderwood, D.A. (2022). Tousled-like kinase 2 targets ASF1 histone chaperones through client mimicry. *Nat Commun* 13, 749. 10.1038/s41467-022-28427-0.
155. Dennehey, B.K., Noone, S., Liu, W.H., Smith, L., Churchill, M.E., and Tyler, J.K. (2013). The C terminus of the histone chaperone Asf1 cross-links to histone H3 in yeast and promotes interaction with histones H3 and H4. *Mol Cell Biol* 33, 605-621. 10.1128/mcb.01053-12.

156. Mariani, V., Biasini, M., Barbato, A., and Schwede, T. (2013). IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* 29, 2722-2728. 10.1093/bioinformatics/btt473.
157. Farquhar, S., Kossen, J., Kuhn, L., and Gal, Y. (2024). Detecting hallucinations in large language models using semantic entropy. *Nature* 630, 625-630. 10.1038/s41586-024-07421-0.
158. Jiang, H. (2023). A latent space theory for emergent abilities in large language models. arXiv preprint arXiv:2304.09960.
159. Zheng, X., Ho, C.Q.W., Zheng, X., Lee, K.L., Gradin, K., Pereira, T.S., Berggren, P.O., and Ali, Y. (2018). Co-immunoprecipitation Assay Using Endogenous Nuclear Proteins from Cells Cultured Under Hypoxic Conditions. *J Vis Exp*. 10.3791/57836.
160. Hickman, M., and Cairns, J. (2003). The Centenary of the One-Gene One-Enzyme Hypothesis. *Genetics* 163, 839-841. 10.1093/genetics/163.3.839.
161. Cuénot, L. (1902). La loi de Mendel et l'hérédité de la pigmentation chez les souris. *Archives de Zoologie Expérimentale et Générale* 3 Series.
162. Cuénot, L. (1908). L'hérédité de la pigmentation chez les souris. *Zeitschrift für induktive Abstammungs-und Vererbungslehre* 1, 129-129.
163. Steensma, D.P., Kyle, R.A., and Shampo, M.A. (2010). Abbie Lathrop, the "mouse woman of Granby": rodent fancier and accidental genetics pioneer. *Mayo Clin Proc* 85, e83. 10.4065/mcp.2010.0647.
164. Lathrop, A.E.C., and Loeb, L. (1919). Further Investigations on the Origin of Tumors in Mice: IV. The Tumor Incidence in Later Generations of Strains with Observed Tumor Rate. *The Journal of Cancer Research* 4, 137-179. 10.1158/jcr.1919.137.
165. Wunderlich, V. (2006). Early references to the mutational origin of cancer. *International Journal of Epidemiology* 36, 246-247. 10.1093/ije/dyl272.
166. Auchincloss, H., and Winn, H.J. (2004). Clarence Cook Little (1888–1971): The Genetic Basis of Transplant Immunology. *American Journal of Transplantation* 4, 155-159. <https://doi.org/10.1046/j.1600-6143.2003.00324.x>.
167. Mansour, S.L., Thomas, K.R., and Capecchi, M.R. (1988). Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* 336, 348-352. 10.1038/336348a0.
168. Orban, P.C., Chui, D., and Marth, J.D. (1992). Tissue- and site-specific DNA recombination in transgenic mice. *Proc Natl Acad Sci U S A* 89, 6861-6865. 10.1073/pnas.89.15.6861.
169. Westphal, C.H., and Leder, P. (1997). Transposon-generated 'knock-out' and 'knock-in' gene-targeting constructs for use in mice. *Curr Biol* 7, 530-533. 10.1016/s0960-9822(06)00224-7.
170. Shaimardanova, A.A., Chulpanova, D.S., Solovyeva, V.V., Mullagulova, A.I., Kitaeva, K.V., Allegrucci, C., and Rizvanov, A.A. (2020). Metachromatic Leukodystrophy: Diagnosis, Modeling, and Treatment Approaches. *Front Med (Lausanne)* 7, 576221. 10.3389/fmed.2020.576221.
171. Hess, B., Saftig, P., Hartmann, D., Coenen, R., Lüllmann-Rauch, R., Goebel, H.H., Evers, M., von Figura, K., D'Hooge, R., Nagels, G., et al. (1996). Phenotype of arylsulfatase A-deficient mice: relationship to human metachromatic leukodystrophy. *Proc Natl Acad Sci U S A* 93, 14821-14826. 10.1073/pnas.93.25.14821.

172. Ong, T., and Ramsey, B.W. (2023). Cystic Fibrosis: A Review. *Jama* 329, 1859-1871. 10.1001/jama.2023.8120.
173. Gibson-Corley, K.N., Meyerholz, D.K., and Engelhardt, J.F. (2016). Pancreatic pathophysiology in cystic fibrosis. *J Pathol* 238, 311-320. 10.1002/path.4634.
174. Dickinson, P., Dorin, J.R., and Porteous, D.J. (1995). Modelling cystic fibrosis in the mouse. *Mol Med Today* 1, 140-148. 10.1016/s1357-4310(95)80092-1.
175. Vaillancourt, M., Aguilar, D., Fernandes, S.E., and Jorth, P.A. (2024). A chronic *Pseudomonas aeruginosa* mouse lung infection modeling the pathophysiology and inflammation of human cystic fibrosis. *bioRxiv*. 10.1101/2024.10.07.617039.
176. Zhang, J., Zhang, Y., Wang, J., Xia, Y., Zhang, J., and Chen, L. (2024). Recent advances in Alzheimer's disease: mechanisms, clinical trials and new drug development strategies. *Signal Transduction and Targeted Therapy* 9, 211. 10.1038/s41392-024-01911-3.
177. Goedert, M. (2015). Alzheimer's and Parkinson's diseases: The prion concept in relation to assembled A $\beta$ , tau, and  $\alpha$ -synuclein. *Science* 349, 1255555. doi:10.1126/science.1255555.
178. Harrison, R.K. (2016). Phase II and phase III failures: 2013-2015. *Nat Rev Drug Discov* 15, 817-818. 10.1038/nrd.2016.184.
179. Dowden, H., and Munro, J. (2019). Trends in clinical success rates and therapeutic focus. *Nat Rev Drug Discov* 18, 495-496. 10.1038/d41573-019-00074-z.
180. Ezran, C., Karanewsky, C.J., Pendleton, J.L., Sholtz, A., Krasnow, M.R., Willick, J., Razafindrakoto, A., Zohdy, S., Albertelli, M.A., and Krasnow, M.A. (2017). The Mouse Lemur, a Genetic Model Organism for Primate Biology, Behavior, and Health. *Genetics* 206, 651-664. 10.1534/genetics.116.199448.
181. Kang, L., Wu, B., Zhou, B., Tan, P., Kang, Y., Yan, Y., Zong, Y., Li, S., Liu, Z., and Hong, L. (2025). AI-enabled Alkaline-resistant Evolution of Protein to Apply in Mass Production. *eLife Sciences Publications, Ltd.*
182. Lu, W., Zhang, J., Huang, W., Zhang, Z., Jia, X., Wang, Z., Shi, L., Li, C., Wolynes, P.G., and Zheng, S. (2024). DynamicBind: predicting ligand-specific protein-ligand complex structure with a deep equivariant generative model. *Nature Communications* 15, 1071. 10.1038/s41467-024-45461-2.
183. Harmalkar, A., Lyskov, S., and Gray, J.J. (2025). Reliable protein-protein docking with AlphaFold, Rosetta, and replica-exchange. *eLife Sciences Publications, Ltd.*
184. Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R.J., Milles, L.F., Wicky, B.I.M., Courbet, A., de Haas, R.J., Bethel, N., et al. (2022). Robust deep learning-based protein sequence design using ProteinMPNN. *Science* 378, 49-56. 10.1126/science.add2187.
185. Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* 379, 1123-1130. 10.1126/science.ade2574.
186. Corpet, A., De Koning, L., Toedling, J., Savignoni, A., Berger, F., Lemaitre, C., O'Sullivan, R.J., Karlseder, J., Barillot, E., Asselain, B., et al. (2011). Asf1b, the necessary Asf1 isoform for proliferation, is predictive of outcome in breast cancer. *Embo j* 30, 480-493. 10.1038/emboj.2010.335.
187. Perez, L.E., Rinder, H.M., Wang, C., Tracey, J.B., Maun, N., and Krause, D.S. (2001). Xenotransplantation of immunodeficient mice with mobilized human blood CD34+ cells

- provides an in vivo model for human megakaryocytopoiesis and platelet production. *Blood* 97, 1635-1643. 10.1182/blood.V97.6.1635.
188. Song, Y., Rongvaux, A., Taylor, A., Jiang, T., Tebaldi, T., Balasubramanian, K., Bagale, A., Terzi, Y.K., Gbyli, R., Wang, X., et al. (2019). A highly efficient and faithful MDS patient-derived xenotransplantation model for pre-clinical studies. *Nature Communications* 10, 366. 10.1038/s41467-018-08166-x.
  189. Heimpel, H., Matuschek, A., Ahmed, M., Bader-Meunier, B., Colita, A., Delaunay, J., Garcon, L., Gilsanz, F., Goede, J., Högel, J., et al. (2010). Frequency of congenital dyserythropoietic anemias in Europe. *Eur J Haematol* 85, 20-25.
  190. Ferrari, S., Vavassori, V., Canarutto, D., Jacob, A., Castiello, M.C., Javed, A.O., and Genovese, P. (2021). Gene Editing of Hematopoietic Stem Cells: Hopes and Hurdles Toward Clinical Translation. *Frontiers in Genome Editing* 3. 10.3389/fgeed.2021.618378.