© Copyright 2018

Aaron Seo

Genomic Insights into Inherited Bone Marrow Failure

Aaron Seo

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Mary-Claire King, Chair

Marshall Horwitz

Sioban Keel

Program Authorized to Offer Degree:

Genome Sciences

University of Washington

Abstract

Genomic Insights into Inherited Bone Marrow Failure

Aaron Seo

Chair of the Supervisory Committee:

Professor Mary-Claire King

Department of Medicine (Medical Genetics) and Genome Sciences

Hematopoiesis is a tightly regulated hierarchical differentiation process where hematopoietic stem cells differentiate into lineage-specific progenitors and in time, mature blood cells. Hematopoiesis takes place in the bone marrow. Bone marrow failure occurs when the marrow has either impaired hematopoietic cellularity or problems in differentiation into at least one lineage. Inherited forms of bone marrow failure can present in either syndromic fashion or restricted to complications from defective hematopoiesis. Although many genes have been identified that, when mutated, cause inherited bone marrow failure, the genetic basis of a subset of patients remains unresolved. We hypothesized that these patients harbor previously uncharacterized mutations in genes that would help us better understand why specific mutations are deleterious and help elucidate the importance of these genes in cell function and hematopoiesis.

We utilized massively-parallel sequencing to sequence the genomes of patients and informative family members. We then identified candidate mutations likely to explain the cause of bone marrow failure. In this dissertation, we highlight key insights made in four genes: *FAM111B*, *THPO*, *BRCA1*, and *GALE*. We expanded the phenotype associated with *FAM111B* to include marrow failure and exocrine pancreatic dysfunction. We identified homozygous loss-of-function mutations in *THPO* that cause a hematopoietic extrinsic marrow failure but is treatable with a THPO receptor agonist. We provide a mechanism for survival of homozygous nonsense mutations in *BRCA1*. Finally, we report a homozygous missense mutation in *GALE*, an important enzyme for glycosylation, as the cause of inherited thrombocytopenia in a large Bedouin family. Our findings highlight the importance of studying the genetic basis of rare hematopoietic diseases to inform clinical management, provide accurate genetic counseling, and better understand molecular pathways and interactions important for hematopoiesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First, I would like to thank Mary-Claire King, who has shown me what it truly means to have a passion for science and research. She taught me to be scientifically rigorous and to think critically for myself. I will miss our scientific chats as well as the more fun and casual weekend chats. I would like to also thank Akiko Shimamura, who exemplifies the essence of being a MD/PhD physician scientist. She has shown me how you can be both a compassionate and caring physician as well as think deeply about the basic biology of hematopoietic diseases. Together, Mary-Claire and Akiko have been the ideal blend of mentors for me. They have been my biggest cheerleaders and have encouraged me to never give up. They have taught me that being a good physician scientist is more than publishing high impact papers or bringing in a lot of grant money. They have taught me to put the patients first, to be a good person, and to strive to be my best self.

I would like to give special thanks to Tom Walsh, Suleyman Gulsuner, Sarah Pierce, and Silvia Casadei in the King Lab for their mentorship and technical expertise in not just science but also important topics such as good hiking/backpacking trails and bread-baking. I would like to thank other past and present members of the King Lab and Swisher Lab, including Ming Lee, James Moore, Anne Thornton, Mary Eng, Sunday Stray, Amanda Watts, Hilal Gulsuner, Barb Norquist, Shwetha Jayakumar, Ryan Carlson, Sarah Baxter, Maribel Harrell, Piri Welcsh, Sarah Bernards, and Marc Radke, for their friendship, scientific help, encouragements, and making the lab a wonderful place to be.

I would like to thank members of the Shimamura Lab, including Mike Zhang, Take Furuyama, Nick Burwick, Marilyn Sanchez-Bonilla, Melisa Ruiz-Gutierrez, Phoenix Ho, and Ang Scott, for being amazing colleagues in my early formative PhD years, where I was both excited and

nervous about the progress and future of my research projects. Thank you for your cheerful attitudes and encouraging spirits.

I have had the privilege of working with a number of amazing collaborators. I would like to thank Sergei Doulatov for teaching me all I know about induced pluripotent stem cells and hematopoietic assays and the rest of his lab, including Jasper Hsu and Massiel Stolla. They made the tissue culture room a lot more exciting to be in. Thank you to all my other collaborators: Hannah Tamary, Judy Fridovitch-Keil, and Hazel Holden.

Thank you to all the patients and their families for their invaluable contribution to the research process. Thank you for allowing us to dive deep into your genomes and help uncover new insights into biology and disease.

I would like to thank the rest of my thesis committee, including Marshall Horwitz, Sioban Keel, and Phil Bradley, for their sage advice and helping guide my dissertation project to completion.

Thank you to my friend, Danny Tabor, who helped me come up with my thesis title.

Thank you to my partner, Jill Mizokawa, for not letting me eat doughnuts and pizza 24/7 and giving me big hugs to get me through the week.

Finally, I would like to thank my family, my parents Daniel and Grace and my brother David. My parents came to the United States in pursuit of a brighter future. They taught me what it means to put God and family first and to work hard. The day David was born was the happiest day of my life. Thank you, David, for bringing so much joy and laughter into our lives.

Chapter 1. INTRODUCTION

1.1 INHERITED BONE MARROW FAILURE

The bone marrow is the site of hematopoiesis. Hematopoietic differentiation is a tightly regulated and hierarchical process in which hematopoietic stem cells give rise to lineage specific progenitors and ultimately to mature blood cells. Inherited bone marrow failure syndromes (IBMFS) are caused by problems in this process of differentiation and survival, leading to failure of the bone marrow to produce blood cells. Mutations responsible for IBMFS generally lie in genes that are either key transcriptional regulators of hematopoiesis, such as *GATA2* and *ETV6*,[1,2] or in genes that affect global cellular processes of survival and maintenance, such as the Fanconi Anemia/BRCA DNA repair pathway, *SBDS*, and ribosomal subunit genes.[3] Although many causes of inherited bone marrow failure have been identified, the genetic basis of IBMFS for 20-25% of patients remains unknown.[4]

IBMFS are a heterogeneous group of genetic disorders characterized by impaired hematopoiesis and a hypoproliferative marrow. Many of these syndromes are also associated with congenital malformations and clonal evolution.[3] Examples of IBMFS include Fanconi anemia, dyskeratosis congenita, and Shwachman-Diamond syndrome. Early and accurate diagnosis of IBMFS is critical to inform medical management, because IBMFS do not respond to standard medical therapies utilized for acquired conditions. The significant proportion of unsolved cases also reflects incomplete mechanistic understanding of the molecular pathways responsible for disease.

1.2 TOOLS FOR DIAGNOSING THE GENETIC BASIS OF BONE MARROW FAILURE

Traditionally, clinical genetic diagnosis of inherited bone marrow failure has done with PCR amplification of genomic loci of interest followed by Sanger sequencing of these amplicons. Although this method is highly accurate, it is a single locus or single gene query, making the method low-throughput. The onus is on the clinician to choose the right genes to evaluate first based on clinical findings. However, not every patient with a mutation in a known gene that causes bone marrow failure presents with the same phenotype. Furthermore, the general allelic heterogeneity and overlap of clinical features amongst various genes known to cause bone marrow failure make Sanger sequencing a non-ideal method for timely accurate genetic diagnosis of bone marrow failure.

The advent of massively-parallel sequencing technologies, coupled with targeted capture, has significantly increased sequencing throughput, allowing clinicians and researchers to sequence many genes simultaneously at a much lower cost than with Sanger sequencing.[5–8] The two main approaches to massively-parallel sequencing include targeted gene capture panels and whole exome capture. Targeted gene capture panels consist of oligo hybridization probes against the genes that make up a panel. Gene panels can be themed to a specific disease phenotype or a set of diseases that affect the same organ system. For example, MarrowSeq™ is a gene panel consisting of all genes whose mutations are known to cause inherited bone marrow failure and myelodysplastic syndromes.[6] MarrowSeq™ is actively used in clinical genetic testing to diagnose the genetic basis of marrow failure and myelodysplasia in patients. Whole exome capture is another approach to probing genes for massively-parallel sequencing.[5] Whole exome capture targets the coding sequence of all genes. When coupled with good clinical phenotyping in a family with a likely inherited Mendelian disorder, whole exome sequencing of informative family members is a powerful approach to identify mutations

causative for disease. When compared to traditional Sanger sequencing-based diagnostics approaches, both targeted gene capture panel and whole exome capture sequencing represent unbiased approaches to identify likely disease-causing mutations.

1.3 THESIS OUTLINE

The goal of this dissertation project was to identify and characterize genes responsible for marrow failure and related cancer predisposition. I hypothesized that patients with inherited marrow failure of previously unidentified genetic cause harbor mutations in as-yet-uncharacterized genes that are important for hematopoiesis, clonal evolution, and human development. I also hypothesized that previously uncharacterized mutations in already known bone marrow failure or hematopoietic disorder genes may reveal new and important insights into gene function and biology. The impact of understanding inherited BMF/MDS lies both in direct benefit to the patients and in the realization that genes responsible for rare inherited disease yield novel insights into global molecular mechanisms for both acquired and inherited disease. For example, identification of genes disrupting telomere maintenance in dyskeratosis congenita led to the demonstration of abnormal telomere maintenance in some patients with apparently acquired aplastic anemia or myelodysplastic syndrome.

In Chapter 2, I describe a new association between a previously uncharacterized mutation in the gene *FAM111B* and exocrine pancreatic dysfunction and bone marrow failure discovered with whole exome sequencing in a large pedigree whose affected individuals also present with facial rash, sparse hair, hypohidrosis, and swelling of the extremities. In Chapter 3, I characterize two new homozygous loss-of-function mutations in *THPO* (thrombopoietin) identified with MarrowSeq^{TM} in 5 individuals across 3 families that cause bone marrow failure unresponsive to transplant. I show that romiplostim, a thrombopoietin receptor agonist, treatment in these

patients restores trilineage hematopoiesis. In Chapter 4, I describe a mechanism for survival of Fanconi Anemia patients with homozygous nonsense mutations in *BRCA1* identified with MarrowSeq^{TM}. In Chapter 5, I report a homozygous missense mutation in *GALE* identified with whole exome sequencing in a large Bedouin family whose affected individuals present with severe thrombocytopenia and intracranial bleeding. I show that the mutation causes impairment of enzyme function and leads to protein instability. Appendices A, B, and C contain supplementary materials for Chapters 2, 3, and 5, respectively.

Chapter 2. *FAM111B* MUTATION IS ASSOCIATED WITH INHERITED EXOCRINE PANCREATIC DYSFUNCTION

Note: This chapter is based on the following published paper:

Seo A, Walsh T, Lee M, Ho P, Hsu E, Sidbury R, King MC, and Akiko Shimamura. FAM111B Mutation is Associated with Inherited Exocrine Pancreatic Dysfunction. Pancreas. 2015 Oct 22. PMCID: PMC4841754

2.1 ABSTRACT

Objectives: Few genetic causes of exocrine pancreatic dysfunction have been described to date. We identified a family with multiple affected members manifesting exocrine pancreatic dysfunction. Additional associated features included facial rash, sparse hair, hypohidrosis, and swelling of the extremities. The transmission pattern of these clinical features was consistent with an autosomal dominant mode of inheritance. The two proband siblings also had transient elevated liver transaminases with hepatic steatosis early in life. This study identifies the genetic cause of exocrine pancreatic dysfunction in this family.

Methods: Whole exome sequencing was performed to identify the genetic cause of exocrine pancreatic dysfunction.

Results: A heterozygous germline in-frame deletion in the gene *FAM111B* (*c.1261_1263delAAG,* p.Lys421del) co-segregated with the phenotype: the variant was present in all affected relatives genotyped and absent in all unaffected relatives genotyped. The variant is also absent from public control sequence databases.

Conclusions: Our findings implicate *FAM111B* in autosomal dominantly inheritable exocrine pancreatic dysfunction.

2.2 INTRODUCTION

Exocrine pancreatic dysfunction is characterized by impaired production or release of digestive enzymes. The most common cause of inherited exocrine pancreatic dysfunction is cystic fibrosis, followed by Shwachman-Diamond Syndrome.[9] These are autosomal recessive disorders caused by mutations in the *CFTR* and *SBDS* genes, respectively.[10–13] Other causes of heritable exocrine pancreatic dysfunction are rare, and many remain genetically unidentified.

We evaluated a 3-year-old girl with exocrine pancreatic dysfunction referred to hematology clinic for suspected Shwachman-Diamond syndrome. Her sister shared similar symptoms. Genetic testing failed to reveal mutations in the *SBDS* gene and immunoblotting showed normal SBDS protein levels. Exome sequencing analysis was performed to identify the genetic basis of disease in this family.

2.3 METHODS

Patient Samples.

Informed consent was obtained from all study subjects in accordance with a protocol approved by the institutional review boards (IRBs) of Fred Hutchinson Cancer Research Center and Seattle Children's Hospital. Bone marrow stromal cells and lymphoblasts were prepared and cultured as previously described.[14]

gDNA Processing.

Genomic DNA was purified from patient blood samples using the AllPrep DNA/RNA Mini Kit (Qiagen, Venlo, The Netherlands). Saliva samples were collected with the Oragene Dx collection kit (DNA Genotek, Kanata, Canada). Genomic DNA was purified from samples using the prepIT-L2P reagent and protocol (DNA Genotek, Kanata, Canada).

Targeted Gene Capture Panel Sequencing and Analysis.

Targeted gene capture panel sequencing and analysis was performed as previously described.[6]

Exome Sequencing and Analysis.

Genomic DNA libraries were prepared using the KAPA Library Preparation Kit (KAPA Biosystems, Wilmington, MA) and then hybridized to the SeqCap EZ HGSC VCRome Design Library (NimbleGen Roche, Basel, Switzerland). Captured sequences were sequenced on an

Illumina HiSeq2500 using 2x101bp paired-end reads (Illumina, San Diego, CA). Average depth of coverage was 120.3x, with 93.8% of targeted reads having >10x coverage. The reads were filtered for quality and then aligned to the exome using BWA against human genome assembly hg19. Single nucleotide and small indel (1-23 bp) variants were called using GATK and Pindel. Copy number variant analysis was performed as previously described.[15] Variant annotation and classification was performed using custom Perl scripts. Variants were classified as either intergenic, intronic, 5' UTR, 3' UTR, splice site, or coding. Coding sequence variants were also classified as either nonsense, missense, frameshift, in-frame deletion/insertion, silent, or splice site alteration. The likelihood of damaging effect of missense variants was predicted using PolyPhen2 and SIFT as well as search of the published literature. Sequence conservation scores such as GERP were also used to inform the likelihood of the damaging effect of a variant.

Common variants were excluded by comparison to an in-house database of >1000 exomes, dbSNP v142, the 1000Genomes project, NHLBI Exome Variant Server (EVS), and the Exome Aggregation Consortium (ExAC) database. Variants found with frequency >1% in dbSNP, >5 in EVS, and >20 in ExAC were excluded. Missense variants with PolyPhen2 scores ≤0.90 and/or SIFT P >0.05 were excluded as candidate variants.

Sanger Sequencing.

Direct sequencing was performed using BigDye v3.1 Terminator (Life Technologies/Applied Biosystems, Carlsbad, CA) on PCR template generated using primers flanking the variant sites. Sequence reads were collected via capillary electrophoresis with the 3130xl Genetic Analyzer (Life Technologies/Applied Biosystems, Carlsbad, CA)

Immunoblotting.

Protein lysates were prepared from patient-derived lymphoblasts as previously described.[14] Immunoblotting was performed using antibodies against FAM111B (Sigma-Aldrich, St. Louis, MO) (1:250), SBDS (1:10,000)[14], or γ-tubulin (Abcam) (1:10,000) and quantitated followed by the LI-COR two-color fluorescence detection system (LI-COR, Lincon, NE).

2.4 RESULTS

Clinical Features

Clinical features of affected members in the study family (Figure 2.1) are summarized in Table 2.1 and Table 2.2. The proband, patient V-2, was referred to Hematology Clinic at 5 years of age for diagnostic evaluation of possible Shwachman-Diamond Syndrome (SDS) after assessment by pediatric gastroenterologists at several institutions since her first year of life for liver enzyme elevation and steatorrhea. Upon presentation at our center at age 3 years, AST was 178 IU/L (normal: 5-41), ALT 251 IU/L (normal: 5-40), GGT 407 IU/L (normal: 5-55), alkaline phosphatase 506 IU/L (normal: 95-380), triglycerides 182 mg/dL (normal: 60-135), albumin 4.9g/dL (normal: 3.8-5.4). Liver ultrasound and magnetic resonance cholangiopancreatography were unremarkable. Workup was negative for infantile sclerosing cholangitis and inborn errors of metabolism, including cholesterol-bile acid biosynthetic pathway defects or peroxisomal beta-oxidation of bile acid intermediates. Screening tests for alpha-1 antitrypsin deficiency, celiac disease, and cystic fibrosis (including sweat chloride testing) were negative. A percutaneous liver biopsy was performed at 2 years of age, and the histopathology revealed diffuse microvesicular steatosis without cholestasis or inflammation. At 3 years of age she developed poor weight gain, recurrent abdominal pain and frequent malodorous loose bowel movements typically 4-5 times per day. She was diagnosed with exocrine pancreatic dysfunction when fecal pancreatic elastase testing revealed < 50 µg elastase / g stool (normal: 200 – 500 µg elastase / g stool) on two separate occasions. Pancreatic enzyme replacement therapy resulted in improvement of her gastrointestinal symptoms. Serum trypsinogen (3.1

ng/mL; lower limit normal (LLN) 10 ng/mL) and pancreatic isoamylase (6 units/L; LLN 13 units/L) were decreased, consistent with ongoing pancreatic dysfunction.[16] The findings of microvesicular steatosis on liver biopsy in conjunction with the pancreatic insufficiency prompted referral to hematology clinic for possible Shwachman-Diamond syndrome.[17] During initial hematology evaluation at age 5 years, the complete blood count (CBC) was normal. Given the concerns about possible Shwachman-Diamond syndrome[18], a bone marrow aspirate and biopsy was performed, which was notable for hypocellularity for age (biopsy cellularity 50%), no dysplasia and no cytogenetic clonal abnormalities noted by metaphase analysis or by fluorescent *in situ* hybridization. No sideroblasts or vacuoles were noted to suggest Pearson syndrome. Telomere length analysis was normal.

Past medical history was significant for bilateral facial rash on cheeks, sparse hair particularly of the eyebrows and eyelashes, hypohydrosis with heat intolerance, and intermittent swelling of hands and feet. Additional history included vesiculoureteral reflux leading to recurrent urinary tract infections which resolved after bilateral ureteral implantation. She also had amblyopia, and one episode of cellulitis on her arm without apparent antecedent cause. Mild stiffness of her heel cords were noted on physical exam. Workup for a rheumatologic disorder was negative.

The patient's older sister, patient V-1, shared similar symptoms summarized in Table 2.1. She developed frequent emesis in early infancy and was diagnosed with gastroesophageal reflux treated with Prevacid. She had a dermatologic evaluation for severe eczema and photodistributed poikiloderma characterized by mid to lateral facial reticulate erythema, telangiectasia, mottled pigmentation (alternatively described as hyper and hypopigmentation) on her lower neck and trunk, and mild epidermal atrophy involving the nasal tip but otherwise sparing the central face. Xerosis and follicular papules were noted on her lower trunk. Poilkiloderma and mild non-pitting edema were noted on the dorsum of both hands. No

psoriasiform change noted. No erysipelas-like changes noted. No nail changes or sclerosis of the digits noted. Percutaneous liver biopsies at 1 and 3 years of age revealed macrovesicular and microvesicular steatosis, with minimal local portal and lobular inflammation. At 7 years of age, pancreatic isoamylase was noted to be low (5 units/L; LLN 13 units/L) despite the absence of overt malabsorption symptoms. Patient V-1 also had a history of hypohidrosis, sparse eyebrows and eyelashes, and slow-growing scalp hair. Dermatologic history is also notable for intermittent spontaneous occurrence of hemorrhagic bullae on her hands and feet. Skin biopsies were non-diagnostic. Her past medical history was also notable for amblyopia and hyperopia.

Family history was notable for numerous paternal relatives (Table 2.1, Table 2.2, and Figure 2.1) similarly affected with poikiloderma, truncal pigmentary changes, lymphedema, hypohydrosis/heat intolerance, and exocrine pancreatic dysfunction. In addition, older affected family members had sclerosis of the digits, hand numbness or pain upon cold exposure, and lung disease.

Gene Discovery

Genetic testing for cystic fibrosis and Shwachman-Diamond Syndrome was negative: there were no mutations in the *CFTR* or *SBDS* genes. Furthermore, SBDS protein levels were comparable to that of healthy controls, arguing against promoter or regulatory mutations affecting *SBDS* expression levels (Supplementary Figure 1). Targeted gene capture panel followed by next-generation sequencing of genes causing bone marrow failure syndromes and myelodysplastic syndrome[6] was performed in individuals IV-4, IV-5, V-1, and V-2 (Figure 2.1A) and did not reveal any candidate variants. Exome sequencing of family members IV-4, IV-5, V-1, and V-2 identified protein-altering variants in 5 genes, *FAM111B*, *PRF1, SMC4, TRIO,* and *WFDC12,* affecting evolutionarily conserved amino acids predicted to be deleterious and

segregating with exocrine pancreatic dysfunction in an autosomal dominant pattern. After Sanger sequencing of these five mutations in additional family members, III-2, III-3, III-5, III-6, IV-1, IV-4, IV-5, IV-6, V-1, and V-2, co-segregatation with the phenotype was noted for only one variant: a heterozygous germline *FAM111B* variant, c.1261_1263delAAG (NM_198947.3), encoding p.Lys421del (NP_945185.1), at chr11:58,892,831-58,892,833 (hg19). *FAM111B* missense mutations were recently described to cause a similar clinical phenotype, including poikiloderma, eczema, lymphedema of extremities, sclerosis of digits, hypohydrosis/heat intolerance, and restrictive pulmonary disease (Figure 2.2A).[19] Additional phenotypic features included muscle weakness and myopathies. The *FAM11B* p.Lys421del variant has not been reported in public databases dbSNP139, the Exome Variant Server, the 1000 Genomes Project, the ExAC database of 61,486 exomes, and is not present in our in-house exome database. The lysine residue at position 421 is highly conserved in primate species (Figure 2.2B). Direct Sanger sequencing was used to validate the mutation and to further genotype the extended family (Figure 2.1B, Supplementary Figure 2). Seven affected and three unaffected individuals in our kindred were genotyped and demonstrated co-segregation of p.Lys421del with the phenotype. Immunoblotting showed that FAM111B protein levels were similar across affected and unaffected family members (Figure 2.2C,D).

2.5 Discussion

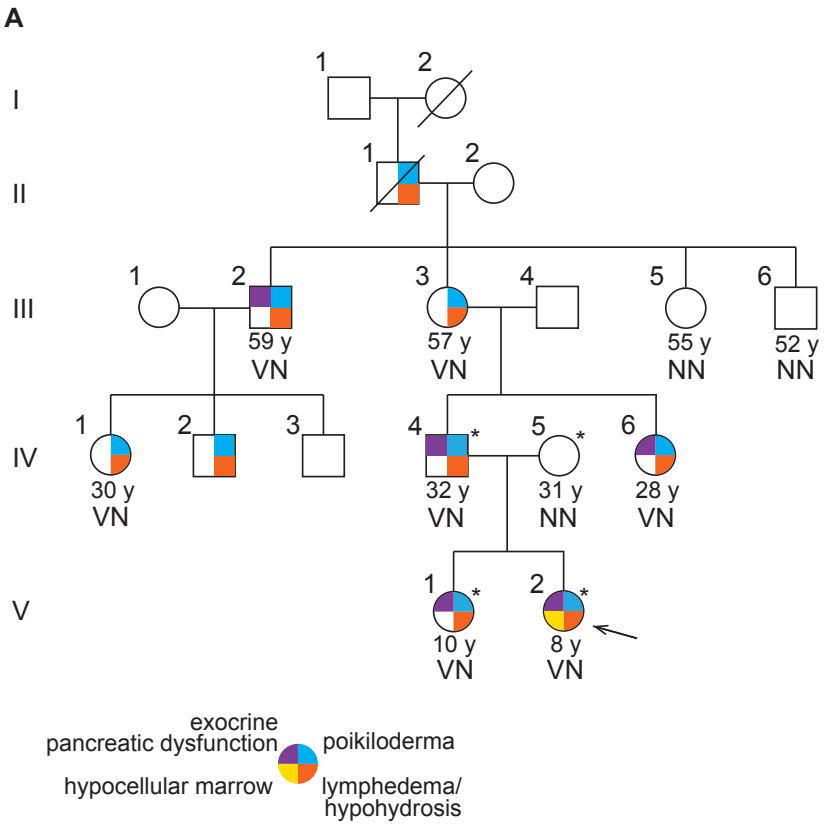
We identified a familial *FAM111B* mutation co-segregating with exocrine pancreatic dysfunction, liver transaminase elevation early in life, and skin rash in an autosomal dominant fashion. The proband also had a hypocellular marrow but no peripheral blood cytopenias. Together, this clinical phenotype shares overlapping features with Shwachman-Diamond Syndrome which typically presents with exocrine pancreatic dysfunction due to pancreatic lipomatosis, marrow failure, as well as eczema and liver transaminase elevation early in life. The phenotypic similarities between our family and those previously reported with heterozygous variants in

FAM111B[19] further supports the notion that mutations in *FAM111B* are responsible for producing the shared phenotype. Exocrine pancreatic dysfunction and bone marrow cellularity had not been previously evaluated in published reports of patients with FAM111B mutations. Of note, pancreatic lipomatosis was noted on autopsy in one previously reported affected individual harboring a FAM111B mutation.[19] Previously described *FAM111B* heterozygous missense mutations were all located within the linker region of the trypsin-like cysteine/serine peptidase domain of FAM111B, whereas the heterozygous deletion of the conserved lysine residue at amino acid position 421 in this study lies upstream of this conserved domain, which might contribute to the phenotypic variation (Figure 2.2A).

FAM111B is member B of the two-gene "Family With Sequence Similarity 111" gene family. The function of FAM111B is currently unknown. FAM111B contains a conserved trypsin-like cysteine/serine peptidase domain, which is split into two parts joined by a linker region 41 amino acids in length. Mutations in FAM111B produce a complex phenotype affecting many organ systems, including the exocrine pancreas, skin, liver, skeletal muscle, and lung.[19] Given the marrow hypocellularity noted in the proband, additional characterization of the hematologic features of this syndrome is of interest. The liver dysfunction in SDS typically presents with elevated liver enzymes early in life that then resolves.[10] An increased risk of hepatic complications with stem cell transplant has been reported for patients with SDS.[20,21] Since little is known about *FAM111B*, future investigations into the molecular and cellular function of this gene are of high interest, especially in the context of the phenotypic similarities with SDS.

Patients with *FAM111B* mutations may benefit from monitoring for potential long-term medical complications such as pulmonary fibrosis.[19] Mutations in *FAM111B* should be considered in the differential diagnosis of patients presenting with heritable exocrine pancreatic dysfunction.

2.6 FIGURES



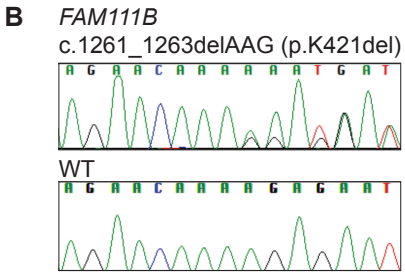


Figure 2.1. Pedigree and Sanger sequences.

(A) Family Pedigree. Asterisks indicate individuals evaluated with whole exome sequencing. The arrow indicates the proband. The *FAM111B* variant allele (c.1261_1263delAAG; p.K421del) is indicated with a V, and wildtype alleles are indicated with a N. Clinical test results for exocrine pancreatic dysfunction were available for individuals III-2, IV-4, IV-6, V-1, and V-2. Marrow hypocellularity was evaluated in individual V-2. NA= not available (B) Representative variant and wildtype Sanger sequencing traces.

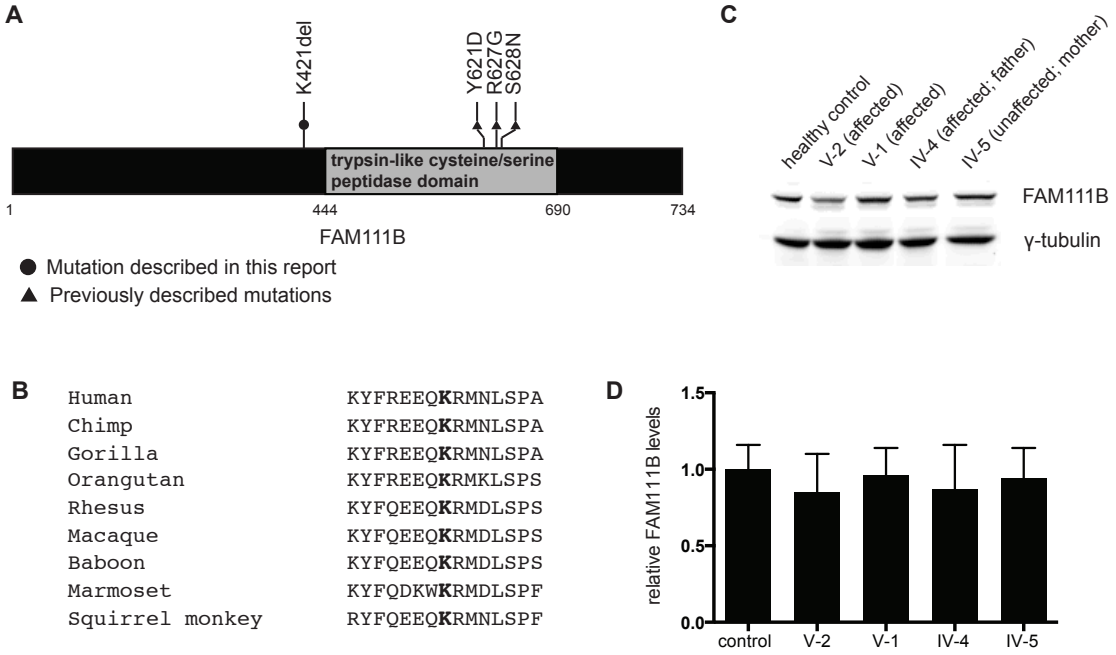


Figure 2.2. *FAM111B* mutations, domains, conservation, and immunoblot.

(A) *FAM111B* Mutations and Domains. The conserved trypsin-like cysteine/serine peptidase domain (light gray) is a two-part domain joined by a loop linker region (dark gray). Lines mark the positions of reported variants in this protein. All previously described variants are located in the loop linker region.[19] (B) Amino acid conservation across vertebrate species. Bold font indicates the lysine residue at position 421 of the human protein sequence. Seven amino acid residues flanking position 421 are shown. (UCSC Genome Browser). (C) Immunoblot against FAM111B. Protein lysates of EBV-immortalized lymphoblastoid cell lines from the indicated study subjects were evaluated by immunoblot for FAM111B together with γ-tubulin as an internal loading control. (D) Densitometry quantification of FAM111B immunoblot levels relative to healthy control. The results from 4 independent experiments are shown.

2.7 TABLES

Table 2.1. Clinical Characteristics of Affected Individuals

	I-2	II-1	III-2	III-3	IV-1	IV-2	IV-4	IV-6	V-1	V-2
Age at last contact	56 yr (deceased)	67 yr (deceased)	58 yr	57 yr	30 yr	27 yr	28 yr	31 yr	10 yr	8 yr
Congenital poikiloderma/red cheeks	NA	+	+	+	+	+	+	+	+	+
Truncal hypo/hyperpigmentation	NA	-	+	+	+	+	+	+	+	+
Hemorrhagic bullae hands/ feet	NA	-	-	-	-	-	-	-	+	-
Eczema	+	+	-	-	-	-	+	+	+	-
Lymphedema/swelling of extremities	NA	+	+	+	+	+	+	+	+	+
Raynaud's or numbness/pain with cold exposure	NA	+	+	+	-	-	+	-	-	-
Sclerosis of digits/inability to straighten fingers/scleroderma	+	+	+	+	-	-	+	+	-	-
Hypohidrosis/heat intolerance (red in face with heat)	NA	+	+	+	+	+	+	+	+	+
Hypotrichosis (scalp, eyebrows, eyelashes)	NA	+	+	+	+	-	+	+	+	+
Lung disease	NA	+	NA	-	-	-	-	-	-	-

Key: - indicates not present, + indicates present; abbreviations used: NA = Not Available

Table 2.2. Gastrointestinal and Hematologic Features

	III-2	IV-4	IV-6	V-1	V-2
Abnormal hepatic biopsy	ND	ND	ND	+	+
Transaminitis	-	+	+	+	+
Low pancreatic isoamylase	+	+	+	+	+
Steatorrhea/diarrhea	+	+	-	-	+
Hypocellular marrow	ND	ND	ND	ND	+

Key: - indicates not present, + indicates present; abbreviations used: ND = Not Determined

Chapter 3. BONE MARROW FAILURE UNRESPONSIVE TO BONE MARROW TRANSPLANT IS CAUSED BY MUTATIONS IN *THPO*

Note: This chapter is based on the following published paper:

Seo A*, Ben-Harosh M*, Sirin M*, Stein J, Dgany O, Kaplelushnik J, Hoenig M, Pannicke U, Lorenz M, Schwarz K, Stoccklausner C, Walsh T, Gulsuner S, Lee MK, Sendamarai A, Sanchez-Bonilla M, King MC, Cario H, Kulozik AE, Debatin KM, Schulz A**, Tamary H**, Shimamura A**. Bone Marrow Failure Unresponsive to Bone Marrow Transplant is Caused by Mutations in *thrombopoietin*. Blood. 2017 Aug 17; 130(7):875-880. PMCID: PMC5561901

*Contributed equally to this work

3.1 ABSTRACT

We report 5 individuals in 3 unrelated families with severe thrombocytopenia progressing to trilineage bone marrow failure (BMF). Four of the children received hematopoietic stem cell transplants and all showed poor graft function with persistent severe cytopenias even after repeated transplants with different donors. Exome and targeted sequencing identified mutations in the gene encoding thrombopoietin (*THPO)*: THPO R99W, homozygous in affected children in two families and THPO R157X, homozygous in the affected child in the third family. Both mutations result in a lack of THPO in the patients' serum. For the two surviving patients, improvement in trilineage hematopoiesis was achieved following treatment with a thrombopoietin receptor agonist. These studies demonstrate that biallelic loss of function mutations in *THPO* cause BMF which is unresponsive to transplant due to a hematopoietic cell-extrinsic mechanism. These studies provide further support for the critical role of the MPL-THPO pathway in hematopoiesis and highlight the importance of accurate genetic diagnosis to inform treatment decisions for BMF.

3.2 INTRODUCTION

Bone marrow failure (BMF) is a potentially life-threatening inherited or acquired disorder characterized by low blood counts and an empty marrow. Mutations in key intracellular pathways are responsible for the development of inherited bone marrow failure syndromes (BMF syndromes)[3,22] Replacement of the impaired hematopoietic system with a hematopoietic stem cell transplant constitutes the standard curative treatment approach for bone marrow failure disorders.

Here, we report 5 children from 3 families presenting with early onset thrombocytopenia evolving into BMF. Four of the children received hematopoietic stem cell transplants and all showed poor graft function even after repeated transplants with different donors. Genetic investigation was pursued and identified a hematopoietic cell-extrinsic factor deficiency causing BMF.

3.3 METHODS

Patient Samples: This study was conducted in accordance with a protocol approved by the Institutional Review Board of the University of Washington, Seattle Children's Hospital, Boston Children's Hospital, Rabin Medical Center (Israel) and Ulm University (Germany) and in accordance with the Declaration of Helsinki. Informed consent was obtained from all study subjects.

Genetic analysis: Targeted gene capture panel sequencing and analysis were performed as previously described.[6] Whole-exome capture was performed with the Agilent SureSelect V5 enrichment capture kit. The enriched library was then sequenced on an Illumina HiSeq 4000 (100 bases paired end). Filtering for single-nucleotide variants (SNVs) and small indels excluded sequences with a read depth < 5 and a quality score < 20. We included only

homozygous variants ($freq_{var} \geq 0.85$) on exons +/- 2bp, with non-existent dbSNP annotation and no entry or a frequency of < 0.0001 in the ExAC database of 60,706 exomes. Only SNVs with a CADD score >10 were considered.

Cell Culture & Transfections: HEK-293T cells (ATCC; CRL-1573) and Baf3-Mpl cells (kind gift of Dr. Kenneth Kaushansky, Stony Brook University) were cultured as described (Supplemental Methods). Transfections were performed using TransIT-LT1 (Mirus Bio, Madison, WI).

Plasmids: The THPO WT and R99W (295C>T) cDNAs were purchased from Genscript (Genscript, Piscataway, NJ) and cloned into the pcDNA3.1/myc-His expression vector (Genscript, Piscataway, NJ).

THPO ELISA quantification: THPO levels from patient serum and secreted THPO from cell culture supernatant, respectively, were quantitated using the Human Thrombopoietin Quantikine ELISA Kit (R&D Systems, Minneapolis, MN) in triplicate (Coefficient of Variation <10%).

qPCR: Quantitative PCR for *THPO* gene expression was performed using iTaq universal SYBR green (Bio-Rad, Hercules, CA). Primer sequences are available in the Supplemental Methods.

Proliferation/Survival Assay: BaF3-Mpl cells were plated, in triplicate, in complete media +/- 20 ng/mL murine IL-3 (mIL-3) or 10 ng/mL TPO. Cells were counted in triplicate for each condition for 3 independent experiments using trypan blue (Life Technologies, Carlsbad, CA). Cell viability was measured by Alamar Blue (Bio-Rad, Hercules, CA).

Cell Culture & Transfections: HEK-293T cells (ATCC; CRL-1573) were cultured with DMEM supplemented with 10% FBS, 2 mM glutamine, and penicllin/streptomycin. BaF3-Mpl cells (kind gift of Dr. Kenneth Kaushansky, Stony Brook University) were cultured in RPMI-1640 (Life Technologies, Carlsbad, CA) supplemented with 10% FBS, 2 mM glutamine, penicillin/streptomycin, and 20 ng/mL murine IL-3 (Peprotech, Rocky Hill, NJ).

Structural Modeling: Crystal structures were superimposed using PDBeFold SSM[23] and modeled using PyMOL.[24]

3.4 RESULTS

Clinical Features

Five children from three families (Figure 3.1A, Supplementary Table 2) presented with thrombocytopenia which progressed to severe pancytopenia with marrow hypocellularity. No dysmorphic features were noted on exam. Evaluation for known causes of inherited or acquired thrombocytopenia or bone marrow failure failed to reveal a diagnosis (see Supplement for details of clinical presentation). Four of the children received hematopoietic stem cell transplants and all showed poor graft function even after repeated transplants with different donors.

Family A: A Bedouin boy with consanguineous parents presented at the age of 9 months with a platelet count of $40x10^9$/L (Family A III-1; Figure 3.1A, Supplementary Table 2). The remainder of his blood counts and his physical examination were unremarkable. Macrocytosis [mean corpuscular volume (MCV) of 100fL] was first noticed at 3 years of age followed one year later by pancytopenia with a white blood cell count (WBC) $2.45x10^9$/L, absolute neutrophil count (ANC) $0.5x10^9$/L, hemoglobin (Hb)- 9.9 g/dL, MCV 98 fL, and platelet (PLT) count $9x10^9$/L. The BM appeared hypocellular (5-10% cellularity) with an absence of megakaryocytes. He soon became transfusion-dependent and underwent bone marrow transplantation from an HLA-A allele mismatched unrelated female donor. Due to failure of engraftment with persistence of the Y chromosome in 100% of bone marrow cells, a second transplant using G-CSF-mobilized peripheral blood derived stem cells from the same donor was performed. He remained profoundly pancytopenic and a bone marrow aspirate on day +70 revealed empty spicules but 100% XX cells by FISH. Immune suppression was discontinued without improvement in the blood counts. Four months after the second transplant, severe gastrointestinal and pulmonary hemorrhage led to his demise.

The younger brother (Family A III-6; Figure 3.1A, Supplementary Table 2) of patient III-1 in Family A was thrombocytopenic at birth (PLT $21x10^9$/L). Around 3 years of age, macrocytosis with pancytopenia developed (Hb 6.5g/dL, WBC $4.5x10^9$/L, ANC $0.51x10^9$/L, PLT $3x10^9$/L Empiric therapy with romiplostim (5 µg/kgBW/week) was followed by improved trilineage hematopoiesis with increased PLT, WBC, ANC and Hb levels (Figure 3.2). THPO levels, drawn after he was no longer requiring platelet transfusions, were undetectable.

Family B: A 2.5 year-old boy (Family B II-1; Figure 3.1A, Supplementary Table 2) from a second family was referred to our clinic on April 1996 for evaluation of cutaneous bleeding. He was born after normal pregnancy and delivery to Arabic parents who were first-degree cousins. No other abnormalities were noted on physical examination. His blood count revealed a platelet count of 18 $x10^9$/L, Hb of 10.4g/dL and MCV of 92.3fL which progressed to pancytopenia and an aplastic marrow. As there was no HLA-matched related donor, two courses of anti-thymocyte globulin and cyclosporine A were administered without improvement. Therefore, at age 4.5 years he underwent an unrelated HLA-matched (low resolution HLA-A and HLA-B, high resolution HLA-DRB1) cord blood transplant. The patient remained pancytopenic with no hematopoietic recovery for 60 days. There was no evidence of donor cell engraftment by Restriction Fragment Length Polymorphism (RFLP) analysis of the patient's blood. A subsequent transplant was performed with the patient's haploidentical father as the donor, but the patient remained pancytopenic. The patient died with multi-organ system failure three months after the first stem cell graft.

During the search for an HLA-matched family donor, the patient’s 2 year-old brother (Family B II-2) was found to be thrombocytopenic (PLT $22x10^9$/L) with mild neutropenia (ANC $1.4x10^9$/L) which progressed to pancytopenia with a hypoplastic marrow. He had no physical anomalies on exam. At the age of 5.5 years, the patient was transplanted with haploidentical paternal

peripheral blood derived stem cells. Two months later, in the absence of any sign of hematopoietic engraftment, he received a second transplant with maternal haploidentical CD34+ enriched stem cells, but a disseminated infection with *pseudallescheria boydii* led to multi-organ failure and death after three weeks.

Family C: In a third family, the patient (Family C II-5; Figure 3.1A, Supplementary Table 2) was the fifth child of a consanguineous family of Saudi-Arabian origin. Following spontaneous vaginal delivery, the girl presented with isolated thrombocytopenia ($16x10^9$/L) unresponsive to IVIG or steroids. Maternal alloimmune thrombocytopenia was ruled out. She required platelet transfusions 1-2 times per week but had no severe bleeding complications. At the age of 3 months she developed moderate neutropenia with ANC $0.5\text{-}0.9x10^9$/L but no recurrent or severe infections. At the age of 10 months, a bone marrow biopsy revealed a hypocellular marrow without clonal cytogenetic aberrations. Bone marrow transplantation was performed at the age of 19 months with a graft from an HLA-identical brother who had normal blood counts. He remained pancytopenic and transfusion-dependent for platelets and red cells. The bone marrow biopsy performed on day+88 post-BMT was hypocellular. Persistent complete donor chimerism was documented in her bone marrow and peripheral blood. Serum thrombopoietin levels were undetectable, drawn 6 days after the last platelet transplant when the platelet count was $21x10^9$/L. Serum thrombopoietin levels of the donor were normal. Treatment with the thrombopoietin receptor (MPL) agonist romiplostim was initiated with a subsequent rise in platelet counts to $>100x10^9$/L, reticulocytes rose above 20%, and ANC rose above $1x10^9$/L (Figure 3.2). Bone marrow biopsy now showed regeneration trilineage hematopoiesis with only mild morphologic signs of dysplasia. The last follow-up was 13 months after BMT with romiplostim being administered every 2 weeks with a normal full blood count.

Gene Discovery

Targeted capture followed by next-generation sequencing of candidate genes (MarrowSeq™)[6] was performed on Family A (Supplementary Table 1; Figure 3.1A). Sequencing revealed one homozygous germline variant in *THPO* co-segregating with the phenotype in an autosomal recessive pattern. The *THPO* variant 295C>T (cDNA sequence, NM_000460.3[25]), resulted in THPO R99W (NP_000451.1), at chr3:184,091,304 (hg19). THPO R99W was not present in public databases including dbSNP141, the Exome Variant Server, the 1000 Genomes Project, the ExAC database of 60,706 exomes, and has not been seen in our in-house databases (1,419 individuals). *In silico* analyses predict THPO R99W to be damaging (PolyPhen-2: 1.00, SIFT: 0.001). Arg99 lies in the EPO-like receptor-binding domain (Figure 3.1C) and is highly conserved in vertebrate species (Figure 3.1D). Sanger sequencing of the entire *THPO* gene in Family B identified the same homozygous THPO R99W variant in affected individuals. Whole exome sequencing of the proband, II-5, in Family C identified a homozygous nonsense mutation *THPO* 469C>T (NM_000460.3; chr3:184,090,894 (hg19), R157X (NP_000451.1), THPO R157X co-segregated with the marrow failure phenotype and was also not present in the public databases listed above.

Functional Characterization

The *THPO* gene encodes a 353 amino acid protein including the 21-amino acid signal peptide. To test whether THPO R99W affects THPO function, normal and R99W THPO were obtained in the supernatant of HEK-293T cells transfected with the corresponding expression vectors. Both normal and R99W THPO supported MPL-dependent cell survival and proliferation (Figure S2). Coupled with the absence of THPO in serum of patients homozygous for THPO R99W (Figure 3.1B), these data suggest that THPO R99W severely reduces THPO levels without affecting THPO function. R157X introduces a stop codon in the middle of the protein, resulting in the absence of the C-terminal glycan domain. Previously published studies have shown that THPO

C-terminal deletion of the glycan domain markedly impaired THPO production through reduced THPO secretion.[26–28]

3.5 Discussion

We identified five individuals in three families with early onset thrombocytopenia evolving into hypoplastic BMF unresponsive to transplant and caused by homozygous *THPO* mutations co-segregating with clinical findings in an autosomal recessive fashion. Bone marrow transplant of three patients resulted in engraftment failure or persistent BMF despite donor cell engraftment, consistent with the cause of the marrow failure being extrinsic to hematopoietic cells. Although serum THPO levels are typically elevated with BMF, THPO R99W and R157X resulted in low or absent endogenous THPO levels. THPO is produced primarily in the liver, kidney, bone marrow stroma, spleen, and nervous system,[29–32] so transplantation of hematopoietic cells failed to cure the BMF. The combination of absent endogenous THPO and discovery of these THPO mutations led us to treat the two surviving patients with romiplostim, a thrombopoietin receptor agonist, resulting trilineage improvement in blood counts. The long-term effects of romiplostim therapy remain to be determined for this patient population. Future studies are needed to address whether romiplostim treatment is associated with the development of clonal hematopoiesis in patients with THPO deficiency where the goal is to replace a missing factor rather than to deliver a supra-physiologic stimulus. Since the liver is a major source of THPO production, liver transplant might offer a potential alternative therapy for this disorder.

A previous study reported homozygosity for THPO R38C (NP_000451.1) in a child with hypoplastic bone marrow failure and in a sibling with clinically asymptomatic cytopenias.[33] THPO R38C did not affect endogenous THPO levels but impaired the ability of THPO to support the growth of a THPO-dependent cell line. It was hypothesized that THPO R38C might affect

the receptor-binding domain of THPO. Another report described three unrelated patients with thrombocytopenia and *de novo* 3q26.33-3q27.2 microdeletions that encompassed *THPO*.[34] None of these prior reports comment on therapies for these patients.

The THPO-MPL pathway is important early in hematopoietic development and is critical for hematopoietic stem cell function and trilineage hematopoiesis. The interaction of thrombopoietin with its receptor, MPL, is essential for production and function of megakaryocytes, as well as for the maintenance of hematopoietic stem cells.[35] Heterozygous gain-of-function mutations in *THPO* result in a myeloproliferative syndrome characterized by thrombocythemia,[36–39] while autosomal recessive loss-of-function mutations in MPL, which is expressed on the surface of hematopoietic stem cells, result in congenital amegakaryocytic thrombocytopenia (CAMT).[40–42] Patients with CAMT typically present with thrombocytopenia at birth and later develop trilineage bone marrow failure.[43] Knockout of either *mpl* or *thpo* in murine models results in reduced trilineage hematopoiesis.[44–46]

Additional studies of the clinical features and outcomes of patients with *THPO* mutations are needed. Despite similarities in clinical presentation, BMF due to mutations in *THPO* versus *MPL* differ dramatically in their treatment.[40] While hematopoietic stem cell transplantation is curative for CAMT, patients with THPO mutations do not benefit from HSCT even with full donor engraftment. This difference is analogous to the differences in the hematopoietic stem cell-intrinsic vs cell-extrinsic defects associated with mutations in another hematopoietic cytokine/receptor pair, SCF and c-kit.[47–50] THPO levels provide an important clue to the underlying diagnosis. This report implicates the importance of accurate diagnosis for optimal management of patients with inherited bone marrow failure and specifically points to the importance of investigation of *THPO* mutations to guide treatment of inherited BMF syndromes.

3.6 FIGURES

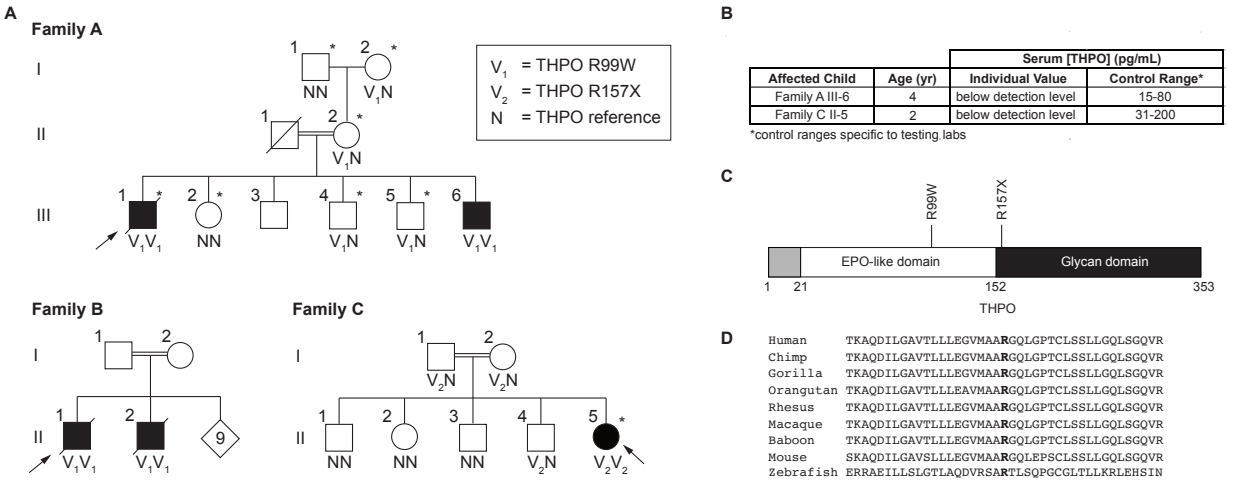


Figure 3.1. Pedigrees, serum THPO, domains, and conservation.

Family pedigrees. (A) The arrow indicates the proband. Family A and Family B: Asterisks indicate individuals evaluated with targeted gene capture panel/high-throughput sequencing. The variant allele (295C>T; R99W) is indicated with V_1, and wild type alleles are indicated with a N. Family C: Asterisk indicates individual evaluated with whole exome sequencing. The variant allele (469C>T; R157X) is indicated with V_2, and reference alleles are indicated with a N. (B) Serum THPO levels in affected individuals. Serum THPO concentrations (pg/mL) are given for the affected child along with the testing lab's control ranges. (C) THPO protein domains. The signal peptide (amino acid residues 1-21) are shaded in grey. The positions of the THPO mutations are indicated on the structure. (D) THPO amino acid conservation across vertebrate species. Bold font indicates the arginine residue at position 99 of the human THPO sequence. 20 amino acid residues flanking position 99 are shown (NCBI RefSeq).

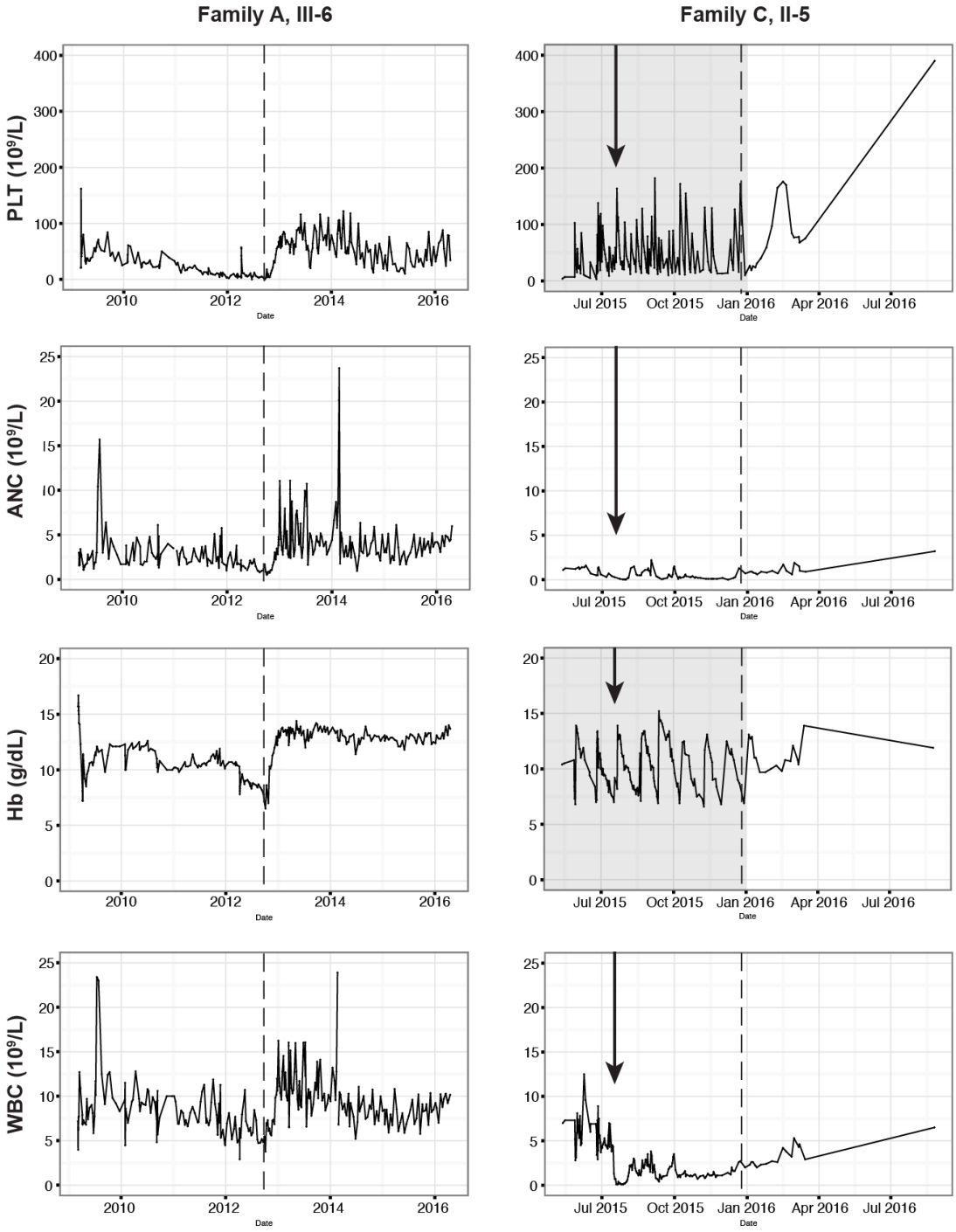


Figure 3.2. Clinical response to romiplostim.

Trilineage response to romiplostim therapy. Platelet (PLT), Absolute neutrophil counts (ANC), Hemoglobin (Hb), and white blood cell (WBC) counts are shown for affected individuals Family A, III6 and Family C, II-5. Arrow indicates date of bone marrow transplantation. Grey shading indicates timeframe of platelet and erythrocyte transfusion dependence. Dotted line indicates initiation of romiplostim therapy. Note difference in time scale (x-axis) between left and right panels.

Chapter 4. A MECHANISM FOR SURVIVAL OF HOMOZYGOUS *BRCA1* NONSENSE MUTATIONS

Note: This chapter is based on the following published paper:

Seo A, Steinberg O, Unal S, Casadei S, Walsh T, Gumruk F, Shimamura A, Akarsu NA, Tamary H, King MC. A Mechanism for Survival of Homozygous *BRCA1* Nonsense Mutations. Proc. Natl. Acad. Sci. USA. Published ahead of print April 30, 2018. https://doi.org/10.1073/pnas.1801796115

4.1 ABSTRACT

Complete loss of the tumor suppressor gene *BRCA1* is lethal during early embryonic development. Patients who are compound heterozygous for *BRCA1* truncating mutations and hypomorphic missense alleles may survive, albeit with very high risk of early onset breast or ovarian cancer and features of Fanconi anemia. However, a mechanism enabling survival of patients homozygous for *BRCA1* truncating mutations has not been described. We studied two unrelated families in which four children presented with multiple congenital anomalies and severe chromosomal fragility. One child developed T-cell acute lymphocytic leukemia (ALL) and a second child developed neuroblastoma. Each of the four children was homozygous for a nonsense mutation in *BRCA1* exon 11. Homozygosity for the nonsense mutations was viable thanks to the presence of a naturally occurring alternative splice donor in *BRCA1* exon 11 that lies 5-prime of the mutations. The mutations did not affect the alternative splice donor, but transcription from this alternative splice site produced a message with deletion of 3309 base pairs, and a short but in-frame BRCA1 isoform. These patients extend the range of *BRCA1*-related phenotypes and illustrate how naturally occurring alternative splicing can enable survival, albeit with severe consequences, of otherwise lethal genotypes of an essential gene.

4.2 SIGNIFICANCE

Many breast and ovarian cancer patients carry inherited cancer-predisposing mutations in *BRCA1*. However, virtually no patients have two inherited mutations in *BRCA1*, because one copy of *BRCA1* is essential for embryonic development. We discovered that patients with two nonsense mutations from a specific region of *BRCA1* may survive, as the result of naturally occurring alternative splicing that yields a short but partially functional BRCA1 protein. These patients are extremely rare, and are characterized by severe chromosomal fragility, congenital anomalies, and predisposition to childhood cancers.

4.3 INTRODUCTION

The BRCA/Fanconi DNA repair pathway is responsible for repair of DNA double strand breaks via homologous recombination[51]. Among the components of the pathway are *BRCA1, BRCA2*, and *PALB2*, each of which harbors mutations predisposing to breast, ovarian, and prostate cancer[52–55]. Homozygosity or compound heterozygosity for mutations in several of the genes in the pathway cause Fanconi anemia, presenting with congenital anomalies, bone marrow failure, and predisposition to both solid tumors and leukemia[51]. However, despite the existence of relatively common cancer-predisposing *BRCA1* alleles in several populations, homozygotes for these founder alleles have not been reported. Indeed, patients homozygous for BRCA1 truncating alleles have not been expected, because evidence from mouse models has shown that at least one copy of *BRCA1* is necessary for embryonic development[56–58]. Consistent with this observation, patients with *BRCA1*-related Fanconi anemia and very young onset breast or ovarian cancer have been shown to be compound heterozygous for a *BRCA1* truncating mutation and a *BRCA1* missense allele retaining partial DNA repair activity[59,60].

In the course of genomic analysis of consanguineous families with severe childhood illnesses, we evaluated two families with congenital abnormalities and childhood cancers. In each family, the affected children were homozygous for an inherited *BRCA1* nonsense mutation. Because we had believed inherited homozygous pathogenic mutations in *BRCA1* to be incompatible with life, we explored the molecular mechanism enabling survival of these children. We discovered that these mutant alleles exploited an alternate, naturally occurring donor splice site in *BRCA1* exon 11, yielding an in-frame short BRCA1 protein. This short BRCA1 isoform is known to retain partial capacity for DNA repair[61], thus enabling embryonic survival but not fully normal development.

4.4 METHODS

Patients. The study was approved by the institutional review boards of the University of Washington, Seattle Children's Hospital, Boston Children's Hospital, Rabin Medical Center, and Hacettepe University. Informed consent was obtained from all study subjects or their parents.

Genetic analysis. Chromosome breakage was analyzed from peripheral blood, as previously described.[62] Genomic analysis for family A was based on targeted capture and multiplexed sequencing using the MarrowSeq panel of 241 genes[6] and for family B by whole exome sequencing. Bioinformatics analysis and variant interpretation were carried out as previously described.[7]

Transcript analysis. *BRCA1* transcripts were evaluated by RT-PCR. RNA was extracted using Trizol (Thermo Fisher Scientific) then treated with DNase I (Qiagen). cDNA was transcribed using SuperScript II (Thermo Fisher Scientific). *BRCA1* full-length, *del11q*, and *del11* transcripts were amplified with the following primers[63]: BRCA1ex9–10for2: 5'-CAACTTATTGCAGTGTGGGAGA-3'; BRCA1ex11FLrev: 5'-GGAGTCCGCCTATCATTACATG-3'; and BRCA1ex12rev: 5'-CCAGATGCTGCTTCACCCT-3'. PCR products were analyzed using D1000 ScreenTape and 2200 TapeStation (Agilent) and Sanger sequenced.

Tissue culture. Cell line UWB1.289[64] was grown in 50% MEGM, 50% RPMI-1640 supplemented with 3% FBS and penicillin/streptomycin (Thermo Fisher Scientific). UWB1.289+BRCA1, the same cell line transfected with wild type BRCA1, was grown in the same media with G418 supplementation (200 μg/ml). Breast cancer cell line MCF7 was grown in MEM supplemented with 10% FBS and penicillin/streptomycin. Human fibroblasts were grown in DMEM supplemented with 10% FBS, 2 mM glutamine, and penicillin/streptomycin.

Patient fibroblasts were grown in Chang D Media (Irvine Scientific) supplemented with penicillin/streptomycin. For inhibition of nonsense-mediate decay, skin fibroblasts were cultured with 500 ug/mL puromycin (Sigma) for 4 hours prior to cell harvesting.

Western blots. Nuclear lysates from all cell lines were extracted using the NE-PER cytoplasmic/nuclear kit according to manufacturer protocols (Thermo Fisher Scientific). 40μg nuclear lysates were resolved on a 4-12% Bis-Tris gel (Thermo Fisher Scientific) under denaturing conditions. Wet transfer was performed at 100V for 2 hours at 4°C. Membranes were blocked in LI-COR Odyssey Blocking Buffer. Primary antibody incubation was performed in the cold room overnight. Antibody to detect BRCA1 was MS110 (EMD Millipore), diluted 1:500. Other antibodies were against β-actin, 1:10,000 (Sigma) and αMs-700nm (LI-COR). Membranes were developed using the LI-COR Odyssey Imager (LI-COR).

4.5 RESULTS

Patients and clinical analysis. Clinical features of the four children in the two families in the study are summarized in Table 4.1. Family A, consanguineous and Arab ancestry, has two affected daughters (Figure 4.1A). Family B, consanguineous and Turkish ancestry, has an affected daughter and an affected son (Figure 4.1B). All four children had failure to thrive, manifesting from birth or postnatally. All had microcephaly, intellectual disability, microphthalmia, and skin pigmentation lesions (hyper- and hypo-pigmented). Limb defects (small hands or feet, hypoplastic thumb) were present in three. Dysmorphic features were noted in both affected children in family A, and brain gliosis was observed in both affected children in family B. Other manifestations (e.g. congenital heart defects, optic nerve hypoplasia) were present in only one patient. Chromosomal breakage was substantially increased in all cases (Table 4.1). Proportions of cells with radial chromosomes (e.g. Figure 1A) were very similar to those of patients with other forms of Fanconi anemia; proportions of cells with other forms of chromosome breaks were somewhat lower than among Fanconi anemia patients generally[65]. Notably, except during periods of active malignant disease (see below), none of the children had anemia or bone marrow failure.

In family A, proband A.II.1 presented at age 5y with anemia (hemoglobin 5.1 g/dL), hyperleukocytosis ($156x10^9$/L) and generalized lymphadenopathy, and was diagnosed with T-cell acute lymphocytic leukemia (ALL) with chromosomal aberrations t(1,16) and +1q. She began treatment according to the ALL IC-BFM 2002 protocol[66] including intrathecal methotrexate and prednisone. On day 8 of treatment, her blast count was 8000 and she received one dose of vincristine (0.75mg/m2). Her vincristine dose was half the usual, given results of her chromosome breakage tests and hence a diagnosis of possible Fanconi anemia. (Her genotype was not then known.) Her leukemia did not respond well to chemotherapy, and

two weeks later she developed hepatitis, fulminant varicella infection, enterococcal sepsis, and succumbed.

Patient A.II.4, the younger sister of A.II.1, also has elevated chromosomal sensitivity to DEB and MMC. She remains cancer free at age 8y. Family A also has two healthy children. Another pregnancy was terminated after *in utero* diagnosis of increased chromosomal breakage. The patients' mother and father are cancer free at ages 36y and 43y, respectively.

In family B, the proband (B.II.4) was diagnosed at age 2y with stage IV neuroblastoma and completed the treatment protocol with full chemotherapy doses. At age 5y she was diagnosed with possible Fanconi anemia based on her clinical features and chromosome breakage testing. She received no transfusions except during neuroblastoma treatment; serum ferritin was 911 ng/mL, and abdominal MRI revealed absence of iron deposition.

Patient B.II.2, the older brother of B.II.4, had chromosome breakage testing at the same time as his sister. He remains cancer free and with no signs of bone marrow failure at age 15y 6m. The genotypes of the children were not known when possible Fanconi anemia was diagnosed. Family B also has two healthy children. The patients' mother and father are cancer free at ages 37y and 44y, respectively.

Mutation discovery. Genomic analysis was carried out for family A using the MarrowSeq panel of 241 genes[6] and for family B by whole exome sequencing. For family A, the only damaging mutation consistent with recessive inheritance of the children's phenotype was BRCA1 p.W372X (c.1115G>A, NM_007294) at chr17:41,246,433 (hg19). The mutation was homozygous in both affected siblings, heterozygous in their parents, and absent from the healthy sibling (Figure 4.1A). For family B, the only damaging mutation consistent with recessive

inheritance was BRCA1 p.L431X (c.1292T>G, NM_007294) at chr17:41,246,256 (hg19). The mutation was homozygous in both affected siblings, heterozygous in their parents and one healthy sibling, and absent from the other healthy sibling (Figure 4.1B).

Homozygosity for complete loss of *BRCA1* function is embryonic lethal[56–58] , so we first considered the possibility of somatic reversion of the nonsense allele to wild type. Somatic reversion can be evaluated by screening for mosaic sequence at the mutant site in affected individuals. From patient A.II.4, both sequencing at 250x coverage of genomic DNA from peripheral blood, and Sanger sequencing of genomic DNA from her skin fibroblasts, yielded only the mutant allele (Figure 4.1A). Similarly, from patient B.II.4, Sanger sequencing of genomic DNA from peripheral blood yielded only the mutant allele, and sequence from genomic DNA of his father B.I.2 did not reveal any loss of the mutant allele (Figure 4.1B).

To assess whether variants in *BRCA1* distal exon 11 might be less severe with respect to cancer risk than mutations elsewhere in *BRCA1*, we evaluated whether cancer-predisposing mutations in BRCA1 were less likely to lie in this region. Distal exon 11 includes 3309 of the 5589 base pairs of the coding sequence of *BRCA1* (59.2%). Of all known *BRCA1* mutations (other than large genomic deletions) indicated as "pathogenic," 1318 of 2238 mutations (58.9%) lie in distal exon 11.[67] We conclude that cancer-predisposing mutations are as likely to occur in this region as elsewhere in *BRCA1*.

Transcription and translation of the mutant alleles. Transcripts and proteins from the patients were evaluated in order to determine the molecular mechanism by which homozygosity for these nonsense mutations could be viable (Figure 4.2A). Wild-type *BRCA1* generates multiple minor transcripts, including several lacking part or all of exon 11.[68] One such naturally

occurring transcript exploits an alternate donor splice at *BRCA1 c.787* at chr17:41,246,761 (Figure 4.2B). Transcription from this alternate donor site leads to loss of nucleotides c.788-4096 of exon 11, and translation yields a short BRCA1 isoform of 759 amino acids, lacking residues 263-1365 of the normal 1863 residue protein. The short isoform has normal 5-prime and 3-prime UTR sequences. We call this transcript and isoform *BRCA1 del11q*. Another naturally occurring minor transcript, *BRCA1 del11*, completely lacks *BRCA1* exon 11 and yields an isoform of 760 amino acids with normal 5-prime and 3-prime UTR sequences.[61] BRCA1 p.W372X and BRCA1 p.L431X lie 3-prime of the alternate donor splice site at *BRCA1 c.787* and consequently are deleted from the *BRCA1 del11q* mutant transcript, as well as from the *BRCA1 del11* mutant transcript. The BRCA1 del11q isoform from the homozygous children would therefore not be subject to premature stops, but would be only 40% of the length of normal BRCA1.

Profiles of *BRCA1* transcripts of A.II.4 compared to controls reflected these effects (Figure 4.2C). In RNA from control fibroblasts, 66% of *BRCA1* transcripts were full-length and 34% were *del11q*. In contrast, in RNA from fibroblasts of A.II.4, 23% of *BRCA1* transcripts were full-length and 77% were *del11q* ($P<0.001$). *BRCA1 del11* was not detected in either sample. After inhibition of nonsense-mediated decay by puromycin, expression of full length and *del11q* transcripts was similar in A.II.4 fibroblasts, as in controls. This result suggests that from the mutant allele, enrichment of *del11q* transcripts is the result primarily of nonsense-mediated decay of full-length transcripts, rather than of increased alternative splicing.

Analysis of proteins from fibroblasts of A.II.4 compared to controls revealed differences in BRCA1 isoforms (Figure 4.2D). Cells of A.II.4 had no detectable full-length BRCA1 protein, but instead had BRCA1 protein corresponding to the size of the del11q isoform.

4.6 DISCUSSION

Many thousands of families have been identified with inherited cancer-predisposing mutations in *BRCA1*, yet virtually no patients with two *BRCA1* mutations are known. Previous reports described two patients, each compound heterozygous for a truncating and a missense mutation in *BRCA1*, with features of Fanconi anemia (FANCS) and very young onset breast or ovarian cancer.[59,60] These reports were at first surprising, because it had been known for more than 20 years that at least one copy of *BRCA1* is necessary for murine embryonic development.[56–58] Characterization of the *BRCA1* mutations of the human patients resolved the paradox, revealing that *BRCA1* compound heterozygosity was compatible with life because the missense allele of each patient retained enough function to enable embryonic survival.[59,60]

Analyses of the children in families A and B suggest that homozygosity for their nonsense mutations is compatible with life for a different reason: the existence of a naturally occurring, minor isoform of BRCA1, generated by an alternate splice donor site in *BRCA1* exon 11. The short isoform created by the alternate splice is only 40% the length of normal BRCA1 protein, but retains the normal reading frame. The nonsense mutations of families A and B do not create or alter the alternate splice donor, but lie 3-prime of it in exon 11, so that the nonsense mutations are spliced out of the short isoform. Notably, splicing at the alternate donor site is not elevated in cells of patients compared to controls, but loss of full-length mutant *BRCA1* transcripts (but not full-length normal *BRCA1* transcripts) by nonsense-mediated decay yields significant enrichment of alternately spliced transcripts in patients' cells (Figure 4.2C). Additional support for this explanation is provided by the very recent report of another child with Fanconi-like features who is homozygous for nonsense mutation BRCA1 p.C903X (c.2709T>A).[69] This mutation also lies in exon 11 at a site 3-prime of the alternate splice donor, so likely also yields the *BRCA1 del11q* isoform. Despite the existence of relatively common *BRCA1* truncating

alleles in several populations, patients homozygous for *BRCA1* founder alleles have not been encountered. The absence of such patients is likely due to the positions in *BRCA1* of these founder alleles: none lie in distal exon 11.

Response to chemotherapy varied considerably among the four cancer patients who were homozygous or compound heterozygous for *BRCA1* mutations. One of the two previously reported patients had multiple congenital anomalies, was diagnosed with ovarian cancer at age 28, and developed severe neutropenia, anemia, and thrombocytopenia during chemotherapy. Her BRCA1 missense mutation, V1736A, had reduced, but not completely lost, capacity to repair double strand DNA breaks.[59] The other previously reported patient had multiple congenital anomalies, elevated formation of radial chromosomes on chromosome breakage assay, and was diagnosed with breast cancer at age 25. She was compound heterozygous for a *BRCA1* frameshift and for R1699W. Fibroblasts from this patient retained partial capacity for DNA repair.[60] This patient tolerated chemotherapy and radiation without unusual toxicity. Of the patients in the present study, A.II.1 did not respond well to steroids. Her fatal complications (varicella and sepsis) may have been the result of both her active leukemia and her treatment. In contrast, patient B.II.2 tolerated chemotherapy for her stage IV neuroblastoma very well, which is unusual for a Fanconi anemia patient.

The BRCA1 del11q isoform has been shown to have lost approximately 50% capacity to repair DNA damage.[61] This deficiency is reflected in sensitivity in chromosome breakage assays, in congenital anomalies, and in cancer predisposition of the homozygous children. Other functions of BRCA1, including interactions with PALB2, BRCA2, RAD51, and CTIP, remain intact in this isoform.[61] Binding of these protein partners to the C-terminal BRCT domain is critical to BRCA1 function, and the BRCT domain is intact in BRCA1 del11q.

All patients with two mutations in *BRCA1*, both those previously described and our patients, had severe but not complete loss of DNA repair capacity.[59,60] Prior to their cancer diagnoses, all had normal complete blood counts and no signs of bone marrow failure (Table 4.1). This combination of normal blood counts and predisposition to early-age leukemia and solid tumors is strikingly reminiscent of the clinical features of Fanconi anemia subtype D1 (FANCD1), caused by biallelic mutations of *BRCA2.*[70,71] In mouse, complete deletion of *Brca1* specifically in the bone marrow leads to hematopoietic defects resembling Fanconi anemia.[72] The children of families A and B presented with all features of Fanconi anemia except bone marrow failure. Age of onset of bone marrow failure in other types of Fanconi anemia is highly variable, with mean 7.6y.[3] It is possible that these children may not have had bone marrow failure simply because they were too young, although the previously described patients with compound hterozygous *BRCA1* mutations had no bone marrow failure at ages 28 and 23y.[59,60]

The discovery and characterization of two families with children homozygous for nonsense mutations in *BRCA1* offers insights into biological and clinical features of the gene. First, homozygosity for some *BRCA1* nonsense mutations can be compatible with life. Second, these patients have clinical features that include both Fanconi syndrome anomalies and predisposition to childhood cancers. Third, increased expression of an alternate in-frame BRCA1 isoform can partially compensate for loss of the full-length protein.

Finally, we suggest that evaluating other consanguineous families with Fanconi anemia, childhood cancers, or young onset breast or ovarian cancers may reveal other homozygous *BRCA1* mutations that are compatible with life. Homozygosity for *BRCA1* in patients with chromosomal fragility or childhood cancers may be under-reported because many Fanconi

anemia gene panels do not yet include *BRCA1*. Identifying the genotypes and phenotypes of such patients would contribute to understanding the full spectrum of clinical phenotypes associated with *BRCA1* mutations. This nuanced understanding can lead to better clinical management, inform genetic counseling, and help to further reveal the role of *BRCA1* in cell biology and human development.

4.7 FIGURES

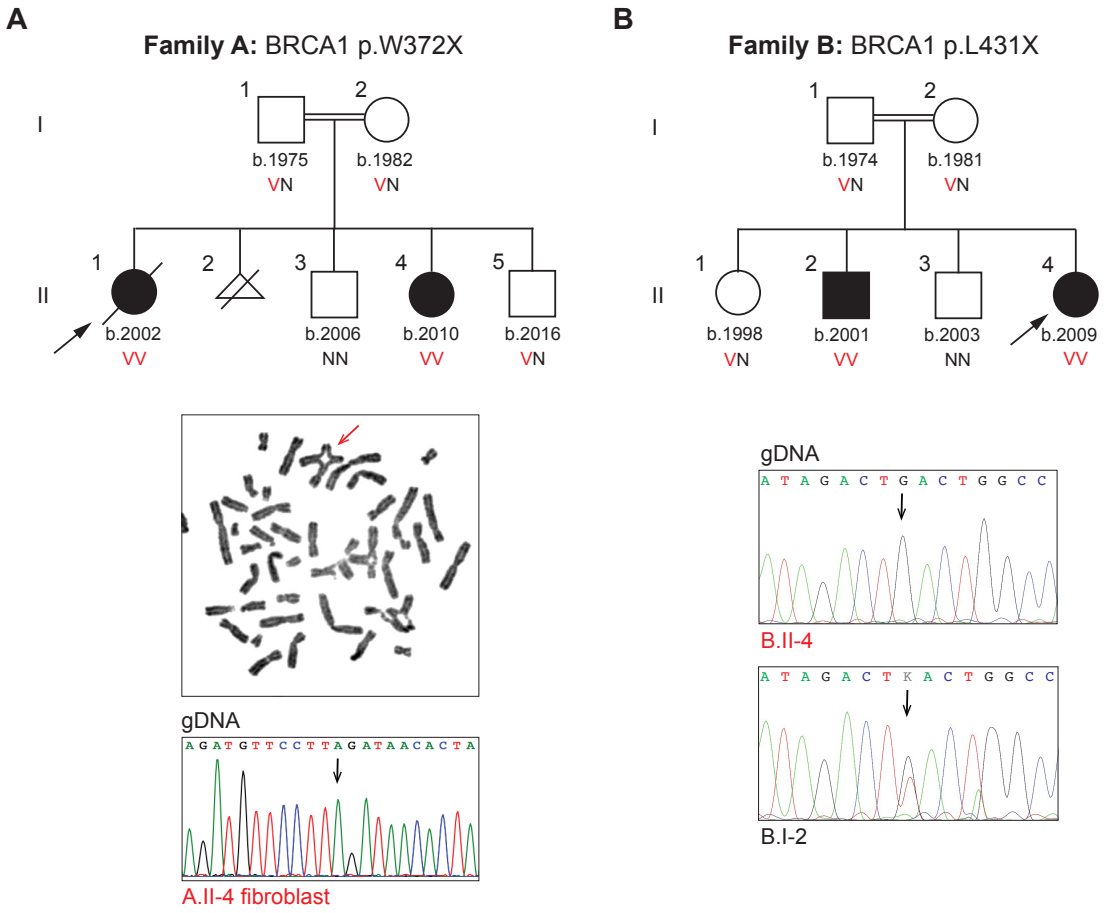


Figure 4.1. Genetic and clinical features of the families.

(*A*) *Top*: Co-inheritance of BRCA1 p.W372X (c.1115G>A, NM_007294) with chromosomal breakage and congenital anomalies. *Middle*: Example of a radial chromosome from patient A.II.4. *Bottom:* Sequence at BRCA1 c.1115G>A from skin fibroblasts of A.II.4, indicating no reversion of the mutant to wild-type sequence. (*B*) *Top:* Co-inheritance of BRCA1 p.L431X (c.1151T>G, NM_007294) chromosomal breakage and congenital anomalies. *Bottom:* Sequence at BRCA1 p.L431X (c.1151T>G) from B.II.4 compared to his father B.I.2, indicating no reversion of the mutant to wild-type sequence. V indicates variant allele, N indicates reference allele, arrows in sequence traces indicate variant sites.

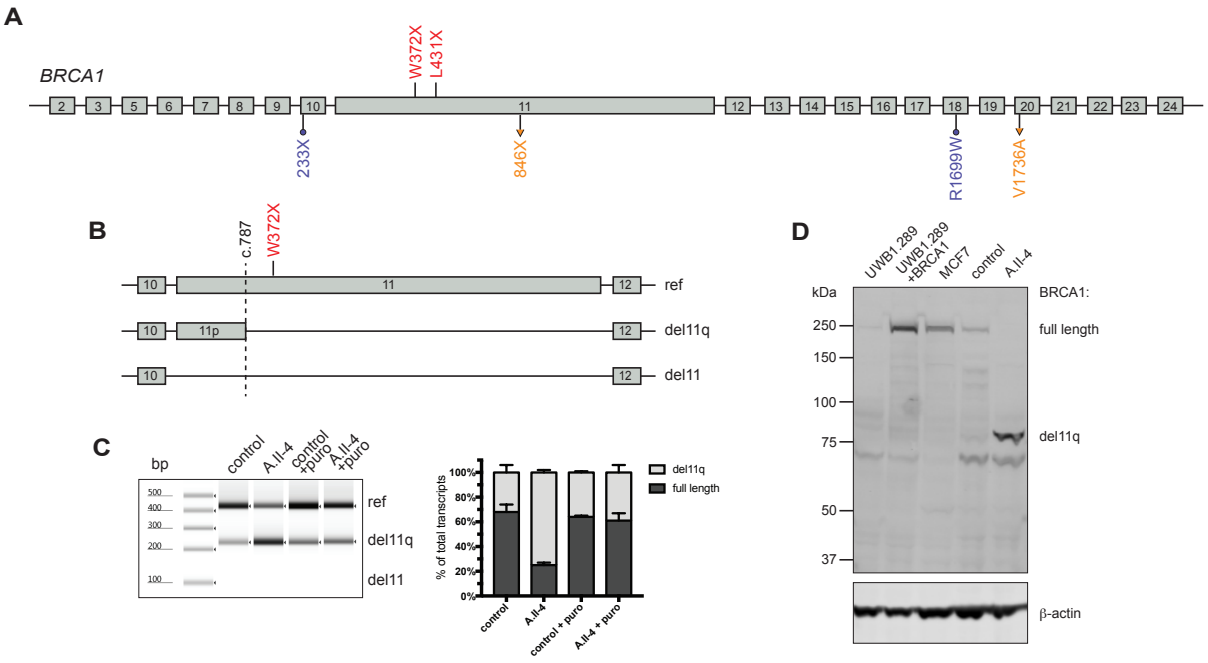


Figure 4.2. Transcript and protein analysis of *BRCA1* mutant alleles.

(A) Homozygous mutations in families A and B, in red, and previously reported compound heterozygous mutations V1736A / 846X (c.2457delC) in orange and R1999W / 233X (c.594_597del) in purple. (B) Schematic overview of splicing at *BRCA1* exons 10-12. The dashed line at *BRCA1 c.787* indicates a naturally occurring alternate splice site that yields an in-frame transcript, *BRCA1 del11q*, lacking 3309bp of exon 11. Another in-frame transcript, *BRCA1 del11,* lacking all 3429bp of exon 11 has also been reported. (*C*) *Left: BRCA1* transcripts, measured by RT-PCR of RNA isolated from fibroblasts from a control and from patient A.II.4, then repeated from fibroblasts treated with puromycin to suppress nonsense-mediated decay. *Right*: relative amounts of *BRCA1* transcripts from three replicate experiments. Results suggest that both full length and *del11q* transcripts are made in all cells, with full-length transcripts from the mutant allele subject to nonsense-mediated decay. Transcripts completely lacking exon 11 were not detected. (*D*) Western blot of nuclear fractions of protein lysate from fibroblasts of A.II.4 compared to controls. UWB1.289 is a *BRCA1*-null ovarian cancer cell line[i]. UWB1.289+BRCA1 is UWB1.289 complemented with full-length *BRCA1* transcript. MCF7 is a breast cancer cell line with normal *BRCA1* sequence. β-actin is an internal loading control.

4.8 TABLES

Table 4.1. Clinical Characteristics of patients with two *BRCA1* mutations

	A.II.1	A.II.4	B.II.2	B.II.4	Ref. 9	Ref. 10
BRCA1 genotype	p.W372X homozygote	p.W372X homozygote	p.L431X homozygote	p.L431X homozygote	p.846X/p.V1736A	p.233X/p.R1699W
Sex and karyotype	female; 46;XX	female; 46;XX	male; 46;XY	female; 46;XX	female; 46;XX	female; 46;XX
Age last follow-up	5y (deceased)	6y	15.5y	7y	28y (deceased)	25y (deceased)
Growth at birth postnatal	SGA: 33 wks, 1.47 kg 5y: 12 kg, 92 cm	AGA: 38 wks, 2.79 kg FTT	SGA: term, 1.6 kg height <3%ile	AGA: 2.9 kg height <3%ile	nd short stature	SGA: 1.99 kg 25y: 40 kg, 135 cm
Intellectual disability	developmental delay	mild developmental delay	yes (IQ 50-69)	yes (IQ 62)	developmental delay	yes
Microcephaly	yes (39 cm at 5y)	yes	yes	yes	yes	yes (microsomia)
Microphthalmia	yes	yes	yes	yes	nd	yes (microsomia)
Skin pigment lesions hypopigmented hyperpigmented café-au-lait spots	 yes yes yes	 yes yes yes	 yes generalized yes	 yes no yes	 nd nd nd	 yes yes no
Limb defects	small edematous palms, soles	small palms	hypoplastic thumb (Blauth 2)	none	none	proximally inserted thumbs, second digit captodactyly, 2-3 toe syndactyly
Dysmorphic features	upslanting palpebral fissures micrognathia	triangular facies, upslanting eyes, epicanthus, cupped ears, short neck micrognathia	none	none	low anterior hairline, prominent nasal bridge, small alae nasi macrognathia	upslating palpebral fissures, sparse hair, hypertelorism, epicanthal folds, ptosis, strabismus, blepharophimosis, broad nasal bridge/nasal tip
Other	optic nerve hypoplasia, bilateral hip dysplasia	R first rib hypoplasia; L 11 ribs	R undescended testis, adrenal insufficiency	heart defect (ASD and VSD), growth hormone deficiency		duodenal stenosis, hyperextensible knees, history of hip dislocation, dental malocclusion
Malignancy (age dx)	T-cell ALL (5y)	none	none	neuroblastoma (2y)	ovarian ca (28y)	breast ca (25y)
Chromosome breakage (case vs control) spontaneous induced, DEB induced, MMC radial chromosomes	 13.3% vs 6.6% 70% vs 7% nd increased	 nd 32% vs 8% 92% vs 34% increased	 6% vs 0 1.44 vs 1.0 nd nd	 2% vs 0 2.14 vs. 1.0 nd nd	 nd nd nd nd	 0.025 vs. 0.025 aberrations/cell 2.2 vs. 0.05 aberrations/cell nd 30% vs 0%
CBC/HbF	normal	normal	normal/normal	normal/normal	normal	normal
Bone marrow	t(1,16)+1q	nd	15y: cellular, no dysplasia	2y: cellular, adipocyte infiltration, no dysplasia 7y: normal	nd	nd
Brain MRI	nd	nd	gliosis: periventricular and subcortical	gliosis, prominent cortical sulci	nd	nd
Cancer family history	intestinal and urological ca, 2nd and 3rd degree relatives		uterine, esophageal, and lung ca in 2nd and 3rd degree relatives		breast, ovarian, and intestinal ca in 1st-4th degree relatives	

AGA, appropriate for gestational age; ASD, atrial septal defect; CBC, complete blood count; DEB, diepoxybutane; dx, diagnosis; FTT, failure to thrive; HbF, fetal hemoglobin; IQ, intelligence quotient; MMC, mitomycin C; MRI, magnetic resonance imaging; nd, not described; %ile, percentile; SGA, small for gestational age.

Chapter 5. MUTATION OF UDP-GALACTOSE-4-EPIMERASE (*GALE*) ASSOCIATED WITH INHERITED THROMBOCYTOPENIA

Note: This chapter is based on the following material:

Seo A, Ben-Harosh M, Walsh T, Lee MK, Gulsuner S, Shimamura A, Tamary H, King MC. Mutation of UDP-Galactose-4-Epimerase (*GALE*) Associated with Inherited Thrombocytopenia. *Manuscript in preparation.*

5.1 ABSTRACT

We report six individuals with severe thrombocytopenia in two generations in a large and consanguineous Bedouin family. Affected individuals have intracranial bleeding, and histology shows dysplastic megakaryocytes. Three individuals underwent successful bone marrow transplant. Whole exome sequencing revealed a mutation in the *GALE* gene encoding the enzyme UDP-galactose-4-epimerase: GALE p.R51W, homozygous in all affected individuals and either heterozygous or not present in unaffected relatives. GALE p.R51W was the only coding sequence variant segregating with thrombocytopenia in the family. GALE plays an important role in galactose metabolism and glycosylation resulting from two enzymatic functions: the interconversion of UDP-galactose with UDP-glucose and the interconversion of UDP-N-acetylgalactosamine with UDP-N-acetylglucosamine. GALE p.R51W has impaired enzymatic activity towards both interconversions. GALE p.R51W has a significantly lower melting point than the normal protein. Proper glycosylation is important for normal hematopoiesis, especially megakaryocyte and platelet development, with mutations in genes either involved in glycosylation or encoding glycoproteins shown to cause thrombocytopenia. One previous clinical report describes a patient presenting with thrombocytopenia and dysplastic megakaryocytes as well as GALE enzyme deficiency. Our genetic and functional studies suggest that GALE p.R51W is functionally defective, potentially disrupting glycosylation and thereby impairing normal megakaryocyte and platelet development. These studies further emphasize the important role of studying glycosylation in the context of hematopoiesis.

5.2 INTRODUCTION

Inherited thrombocytopenia is a rare hematological condition with a wide range of clinical presentation. Patients can present with or without congenital anomalies, such as absent radii, and can present with small, normal, or large platelets. Inherited thrombocytopenia can be sub-categorized either by clinical phenotype (absence/present of megakaryocytes; platelet size) or by genetic cause. Genetic studies over the past 15 years have revealed multiple genes responsible for inherited thrombocytopenia.[3,6,73] The genes responsible can generally be categorized into different aspects of thrombopoiesis (platelet biogenesis): megakaryopoiesis (*ANKRD26, ETV6, FLI1, FYB, GATA1, GFI1B, HOXA11, MPL, NBEAL2, RBM8A, RUNX1, THPO*), platelet product (*ACTN1, CYCS, GP1BA, GP1BB, GP9, ITGA2B, ITGB3, MKL1, MYH9, PRKACG, TUBB1, WAS*), and platelet clearance (*ABCG6, ABCG8, ADAMTS13, GNE, SLFN14, STIM1, vWF*).[73] Conversely, congenital disorders of glycosylation can also present with thrombocytopenia.[74–77]

Here, we report 6 affected individuals across two generation in a large Bedouin family. The affected individuals presented at early ages with severe thrombocytopenia, mild anemia, and febrile neutropenia. Some of the older individuals developed intracranial bleeding, and histology revealed dysplastic megakaryocytes. Three of the older individuals underwent successful bone marrow transplant. We pursued the genetic cause of disease and identified a homozygous mutation in a gene important for galactose metabolism and glycosylation, but not previously associated with thrombocytopenia.

5.3 METHODS

Patients: The study was conducted in accordance with a protocol approved by the Institutional Review Board of the University of Washington, Seattle Children's Hospital, Boston Children's Hospital, and Rabin Medical Center (Israel) in accordance with the Declaration of Helsinki. Informed consent was obtained from all study subjects.

Genetic analysis: Genomic DNA was extracted from peripheral blood of patients. Whole exome capture was performed using VCRome v2.1 (NimbleGen) and then sequenced on a HiSeq 250 (Illumina) with mean targeted depth of coverage of 120X. More than 96.5% of targeted basepairs had >8 high-quality reads. Sequencing reads were aligned to hg19 with BWA. Variant calling was performed using the Genome Analysis Toolkit.[78] Variant annotation and filtering were performed as previously described.[2,79,80] Candidate variants were validated with Sanger sequencing. Dense SNP genotyping was performed with patients' genomic DNA using the HumanOmniExpress-24 v1.1 BeadChip (Illumina). Identity-by-descent (IBD) analysis and homozygosity mapping were performed using PLINK.[81]

Biochemistry: The human WT GALE cDNA sequence was PCR-amplified with flanking NdeI and XhoI restriction sites and cloned into the pET-31 expression vector (Novagen; kind gift of Dr. Hazel Holden, University of Wisconsin-Madison). Mutant alleles p.R51W, p.V94M, and p.D103G were created with the Q5 site-directed mutagenesis kit (New England Biolabs). Expression vectors were transformed into Rosetta2(DE3) *E. coli* (Millipore Sigma). Single clones were grown in LB media at 37°C until OD600 = 0.6. Bacterial cultures were then induced with 500 μM IPTG and then grown at 18°C overnight. Bacterial cultures were then pelleted and lysed with xTractor Cell Lysis Buffer (Clontech). Recombinant hGALE proteins were then purified with column chromatography using nickel agarose columns (Clontech) and the following buffers:

Equilibration Buffer = 25 mM Tris, pH = 8, 250 mM NaCl, 10% glycerol (TBSG), 20 mM imidazole; Wash Buffer = TBSG, 40 mM imidazole; Elution Buffer = TBSG, 300 mM imidazole. Elution fractions were dialyzed with TBSG to remove imidazole.

Enzyme assays were adapted from previous procedures.[82–84] Unless otherwise specified, 10 ng enzyme was combined with either 0.8 mM UDP-gal or UDP-galNAc and 0.5 mM NAD in 40 mM glycine (pH 8.7). Reactions were incubated for 30 min. at 37°C and quenched at a 1:1 ratio with acetonitrile (ACN). Quenched reactions were centrifuged at 3,800 x rpm for 15 min. using the SH-3000 rotor (Sorvall), and supernatant was further diluted with ACN to achieve a 80:20 ACN:H_2O ratio. To analyze the amounts of sugar substrate and product, 5 μL post-quenched sample was analyzed on the CORTECS UPLC HILIC Column, 90 Å, 1.6um, 2.1mm x 150mm (Waters) using the Acquity UPLC/Xevo TQ MS/MS system (Waters). Detailed liquid chromatography and mass spectrometry procedures are provided in Supplementary Materials.

For differential scanning calorimetry (DSC), protein samples were freshly dialyzed overnight prior to experiment. Proteins were diluted to 10 μM prior to loading onto the Microcal VP-Capillary DSC system (Malvern). 12 scans were performed per run, with starting temperature 25°C, final temperature 65°C, at a scan rate of 60°C/hour. Calorimetry scans were analyzed using the Microcal VP-DSC software (Malvern). Scans were normalized to concentration and intrascan baseline prior to determining the melting points (Tm).

Hematopoietic experiments: For lentivirus preparation, shRNA for *GALE* (TRCN0000049461) was purchased from Sigma in pLKO.1 lentivral vector. pLKO-GFP-luciferase was a kind gift from Dr. Sergei Doulatov (University of Washington). Lentiviral particles were produced by transfecting HEK-293T cells with the lentiviral plasmids and third-generation packaging

plasmids. Virus was harvested 48 hours after transfection and concentrated with ultracentrifugation. Viruses were titered with serial dilution on HEK-293T cells.
For quantitative PCR, RNA was extracted using Trizol reagent (Thermo). Reverse transcription was performed using Superscript II (Thermo). Quantitative PCR was carried out in triplicate with TaqMan Gene Expression Assay reagents (Applied Biosystems). TaqMan probes for GALE (Hs00166181_m1) and control GAPDH were used. Assays were performed on the ABI 7900HT Real-Time PCR System.

Human cord blood CD34+ cells were purchased from AllCells. Thawed cells were expanded for 4 hours at 37°C in tissue culture incubator with X-VIVO10 media (Lonza) supplemented with 1% BSA (Thermo), 2 mM L-glutamine (Thermo), penicillin/streptomycin (Thermo), SCF(100 ng/mL), FLT3 (100 ng/mL), TPO (50 ng/mL), and IL-6 (20 ng/mL). All cytokines were purchased from Peprotech. Cells were seeded on RetroNectin-coated (Takara; 10 $\mu g/cm^2$) non-TC-treated 96-well plates at a density of 10-100,000 cells per well. The multiplicity of infection (MOI) was 50 for all viruses. Viruses were concentrated onto cells via centrifugation at 2,500 x rpm for 30 min. Viruses were removed with washing 24 hours after infection. GFP+ cells were sorted on Day 3 with the BD FACS Aria cytometer.

For methycellulose colony formation assays, 10,000 cells were added to 4 mL MethoCult (H3434; StemCell Technologies). 1 mL of mixture was plated on to 35 mm dishes in triplicate and maintained in a humidified chamber for 14-16 days. Colonies were counted manually using a light microscope.

For collagen-based megakaryocyte colony formation assays, 50,000 cells were added to 1.2 mL MegaCult-C media containing cytokines (04901; StemCell Technologies). Cells in media were

mixed with collagen and plated on to dual chamber slides and maintained in a humidified chamber for 10 days. Slides were dehydrated, fixed, and stained according to manufacturer's instructions (StemCell Technologies).

For liquid culture expansion of cells in megakaryocyte growth conditions, CD34+ cells were placed into the following media. Megakaryocyte (MK) Stage 1: SFEM supplemented with penicillin/streptomycin (Thermo), LDL (20 μg/mL; StemCell Technologies), TPO (50 ng/mL), SCF (50 ng/mL), IL-6 (10 ng/mL), and IL-9 (15 ng/mL). MK Stage 2: same as MK Stage 1 except [SCF] is 1 ng/mL. All cytokines were purchased from Peprotech.

5.4 RESULTS

Clinical Features

Six individuals in two generations of a large consanguineous Bedouin family from Israel presented with thrombocytopenia, and in some individuals, mild anemia and febrile neutropenia (Figure 5.1). The proband was a 13-month old girl who presented at a few months of age with thrombocytopenia (PLT 44 x 10^9/L) and neutropenia (ANC 1.00 x 10^9/L). The rest of the CBC was normal. Peripheral blood histology showed big and pale platelets. Bone marrow aspiration revealed normal cellularity but markedly dysplastic megakaryocytes with hypolobated nuclei hypogranularity (Figure 5.1B). Physical exam was normal. The proband had two affected aunts and three affected cousins once removed. Relatives IV-1, IV-6, IV-16, and IV-22 have thrombocytopenia, dysmegakaryopoiesis, mild dyserythropoiesis, and febrile neutropenia. All affected adult relatives presented as children with increased bleeding tendency including intracranial hemorrhages in three of the adults. Three of the relatives underwent successful bone marrow transplant. The detailed clinical features of this family were originally described in 2003.[85] At the time, the proband had not been born. Also at that time, four genes associated with congenital thrombocytopenia (*NFE2, FLI1, FOG1, and GFI1B1*) were Sanger sequenced, but all sequences were normal.[86–90]

Gene Discovery

To identify the gene and mutations responsible for thrombocytopenia in this family, whole exome sequencing was performed on the affected proband (V-1), unaffected parents (IV-7, IV-8), and one affected aunt (IV-1). All available family members were then genotyped for all variants co-segregating with disease under a recessive model. The only damaging mutation consistent with recessive inheritance of the affected individuals' phenotype was GALE p.R51W (c.C151T, NM_001127621) at chr1:24,125,191 (hg19) (Figure 5.1A). The mutation was

homozygous in all affected family members, heterozygous in parents, and either heterozygous or absent from healthy siblings. GALE p.R51W has been seen three times in the ExAC database of 60,706 exomes as heterozygotes, but is absent from dbSNP150, the Exome Variant Server, the 1000 Genomes Project, and our in-house databases (exomes of 1,384 individuals, many from the Middle East).

Due to the high level of consanguinity in the family, we carried homozygosity mapping on all 6 affected individuals and 14 unaffected individuals (Figure S1-2). Homozygosity mapping revealed only homozygous region genome-wide longer than 200kb that was shared by all 6 affected individuals and not homozygous in any unaffected relatives. The critical region is chr1:23,907,702-27,666,378 (hg19) and contains our candidate variant GALE p.R51W. Whole genome sequencing was also performed to evaluate noncoding variants in the critical region. Please see Supplementary Materials (Appendix C) for additional details.

All *in silico* analyses predict GALE p.R51W to be damaging (PolyPhen-2: 1.00, SIFT: 0.000, PROVEAN: -7.78). Arg51 is completely conserved through yeast (*S. cerevisiae*) (Figure 5.2A). Structurally, R51W does not interfere with the binding pockets of the substrate or the NAD cofactor, but the residue is in close proximity to the C-terminal end of GALE (Figure 5.2B).

Functional Characterization

<u>p.R51W has impaired enzyme activity for both UDP-gal/UDP-glc and UDP-galNAc/UDP-glcNAc interconversions</u>

GALE encodes UDP-galactose-4-epimerase, an enzyme important for galactose metabolism and the production of glycosylation substrates. GALE has two reversible enzymatic functions: the conversion between UDP-galactose (UDP-gal) and UDP-glucose (UDP-glc) and the

conversion between UDP-N-acetylgalactosamine (UDP-galNAc) and UDP-N-acetylglucosamine (UDP-glcNAc). A severe defect in the ability to convert UDP-gal to UDP-glc can lead to classic features of galactosemia, including vomiting, hypotonia, seizures, jaundice, galactosuria, and hepatomegaly (Figure 5.2C).[84,91] However, less severe defects in enzyme activity either do not lead to a clinical phenotype or lead to a different phenotype. For example, GALE enzyme activity defect was detected in a young patient with thrombocytopenia, leukopenia, and dysplastic megakaryocytes without clinical signs of galactosemia.[92]

Differences in clinical presentation may be due to a range of enzyme activity impairments with respect to the two enzyme functions. Whereas inability to convert UDP-gal to UDP-glc can lead to the buildup of toxic galactose metabolites leading to the classic clinical features of galactosemia, UDP-galNAc and UDP-glcNAc are important building blocks of glycosylation; buildup of UDP-galNAc and UDP-glcNAc is not expected to lead to clinical signs of galactosemia but may lead to other defects.

To understand the heterogeneity of genotype-clinical phenotype correlations within GALE, we compared GALE p.R51W to two previously reported homozygous GALE mutations: p.V94M and p.D103G.[84,93] GALE p.V94M is the only GALE allele known to cause the severe classic galactosemia phenotype including vomiting, hypotonia, jaundice, galactosuria, and hepatomegaly. This severe galactosemia phenotype is attributed to the nearly absent enzyme activity for UDP-gal→UDP-glc. In contrast, homozygosity for GALE p.D103G does not lead to a clinical phenotype, presumably because the enzyme activity is far less severely impaired.

To test the consequence of the various GALE homozygous mutations on the initial rates of the enzyme functions, assays were performed using recombinant human GALE proteins purified from *E. coli*. and substrate and product sugars analyzed with UPLC-MS/MS (Figure 5.3). GALE

p.R51W is impaired for both UDP-gal→UDP-glc (39.5% activity relative to WT) (Figure 5.3A) and UDP-galNAc→UDP-glcNAc (38.7% activity relative to WT) (Figure 5.3B). GALE p.V94M and GALE p.D103G activities were consistent with levels seen previously. GALE p.V94M was the most severely affected mutant, with 0.9% activity relative to WT for UDP-gal→UDP-glc and 4.9% activity relative to WT for UDP-galNAc→UDP-glcNAc. GALE p.D103G, UDP-gal→UDP-glc activity (51% relative to WT) was slightly, but not significantly, higher than this activity of GALE p.R51W. GALE p.D103G UDP-galNAc→UDP-glcNAc activity (48% relative to WT) was significantly higher than this activity of GALE p.R51W (P=0.01).

The absence of a galactosemia phenotype in our patients makes sense in light of the enzyme studies because activity for UDP-gal→UDP-glc is comparable between GALE p.R51W and GALE p.D103G, the allele identified in a patient with no clinical phenotype. The absence of a galactosemia phenotype is also consistent with the observation that both GALE p.R51W and GALE p.D103G rescue yeast growth upon galactose challenge in the context of *gal10* knockout, the analogous yeast *GALE* gene (Figure S3).[82,84] In yeast, Gal10p has only one function, the interconversion of UDP-gal and UDP-glc.

Protein stability of GALE p.R51W

GALE Arg51 is not located in the binding pockets of the substrate, but the residue is in close proximity to the C-terminal end of the protein (Figure 5.2B). We wondered if substitution of Trp for Arg at this site destabilizes the GALE protein. To test this hypothesis, we performed differential scanning calorimetry (DSC) to measure melting points (Tm) of various purified alleles of GALE proteins (Figure 5.4). These experiments showed that wild-type GALE has a Tm of 48.9°C ± 0.126°C and GALE p.R51W has a Tm of 40.5°C ± 0.16°C (Figure 5.4A,B). The Tm of GALE p.R51W was significantly lower than any other allele tested (Figure 5.4C). The 8.45°C

difference in Tm's of WT vs. p.R51W strongly suggests that p.R51W is unstable relative to WT and likely also unstable relative to GALE p.V94M, and p.D103G.

GALE shRNA knockdown may impair megakaryocyte proliferation *in vitro*

To evaluate the functional consequence of GALE p.R51W on hematopoiesis, we performed *in vitro GALE* gene shRNA knockdown experiments in cord blood-derived CD34+ hematopoietic progenitor cells. *GALE* knockdown did not show significant difference with control for erythroid, myeloid, and megakaryocyte lineage total colony formation potential (Figure 5.5,Figure 5.6A). However, the proportion of large megakaryocyte colonies (≥50 cells) was significantly increased in *GALE* knockdown, suggesting there may be more immature colonies (Figure 5.6B). Total proliferation of hematopoietic cells in megakaryocyte growth conditions was significantly decreased with *GALE* knockdown (Figure 5.7).

5.5 DISCUSSION

We report the identification of the homozygous missense mutation GALE p.R51W as the possible cause of disease in a family with inherited thrombocytopenia. Affected individuals have dysplastic megakaryocytes, severe intracranial bleeding, and some have mild anemia and febrile neutropenia. GALE p.R51W is the only coding sequence change that perfectly co-segregates with the phenotype in the family and that lies within the only region of homozygosity shared by all affected individuals. Arg51 is a highly conserved amino acid, and GALE p.R51W is predicted to be highly damaging by *in silico* tools. Homozygosity for GALE p.R51W has not been observed in any database.

GALE encodes UDP-galactose-4-epimerase, which catalyzes two different reversible reactions of galactose metabolism: interconversions of UDP-gal with UDP-glc and UDP-galNAc with UDP-glcNAc.[94,95] Products of the GALE reaction are important for production of energy from galactose metabolism, for clearance of potentially toxic galactose metabolites, and for glycosylation. We showed that p.R51W impairs enzyme function of both reactions and that p.R51W destabilizes GALE.

Missense mutations in *GALE* cause epimerase deficiency galactosemia (EDG), with a wide range of clinical severity.[84,93,96–102] The condition may be peripheral, affecting only blood cells and generally considered benign, or generalized, with acute clinical features including hypotonia, poor feeding, vomiting, jaundice, hepatomegaly/liver dysfunction, aminoaciduria, and cataracts.[100,103] Acute signs of galactosemia are not present in our family. To date, only one homozygous allele in GALE, p.V94M, has been reported to cause the acute generalized form of EDG. Most individuals with EDG present with the peripheral form, with either mild or no clinical phenotype. Most of these individuals have compound heterozygous mutations, and less severe

enzyme deficiencies than GALE p.V94M homozygotes. This observation suggests that enzyme function deficiency must reach a low threshold before clinical symptoms manifest.

A fundamental question is why patients homozygous for GALE p.V94M or p.D103G do not exhibit hematopoietic defects, whereas patients homozygous for p.R51W present with a severe thrombocytopenia and bleeding phenotype. We hypothesized that p.R51W has a distinctive combination of enzyme dysfunction and protein instability that affects hematopoietic cells. To test this hypothesis, we performed *in vitro* enzyme assays and differential scanning calorimetry on GALE WT, p.R51W, p.V94M, and p.D103G. For the interconversion of UDP-gal to UDP-glc, GALE p.V94M showed little to no function, and p.R51W and p.D103G showed similar intermediate function. For the interconversion of UDP-galNAc to UDP-glcNAc, GALE p.V94M again showed little function but was more effective at converting UDP-galNAc to UDP-glcNAc than converting UDP-gal to UDP-glc. p.R51W and p.D103G both showed intermediate function, but the activity of p.R51W was significantly lower than the activity of p.D103G. With respect to protein stability, p.R51W had a significantly lower melting point than any of the other proteins tested, suggesting p.R51W is the most unstable of these alleles. Based on these observations, we posit that the effect of p.R51W on UDP-gal/UDP-glc interconversion is not great enough to cause classical galactosemia symptoms, but that the effect of p.R51W on UDP-galNAc/UDP-glcNAc interconversion, coupled with instability of the protein, is uniquely detrimental to GALE function in hematopoietic cells.

GALE mutations have not previously been associated with defects in hematopoiesis, but one reported case offers an intriguing possibility. A female patient, age 4, of European-American ancestry, was diagnosed with very similar clinical features to the affected individuals our family: severe thrombocytopenia with dysplastic megakaryocytes, occasional anemia, and febrile neutropenia.[92] She also had peripheral galactosemia, detected by the absence of GALE

enzyme activity from her red blood cells. Like the affected relatives of our family, this patient did not present with acute galactosemia. Her GALE genotype was not determined and she is unfortunately not available.

Other defects in sugar metabolism and glycosylation have been reported to be associated with bone marrow failure (Figure 5.8). Mutations in *ALG8*, encoding alpha-1,3-glucosyltransferase, cause congenital disorder of glycosylation type 1K (CDG1K), which can include anemia and thrombocytopenia.[75,76] Mutations in *SLC37A4*, encoding the glucose-6-phosphate transporter, cause glycogen storage disease type 1B (GSD1B), which includes neutropenia.[104] Mutations in *G6PC3*, glucose-6-phosphatase, cause severe congenital neutropenia type 4 (SCN4).[105] Mutations in *SLC35A1*, encoding the CMP-sialic acid transporter, cause congenital disorder of glycosylation type 2F (CDG2F), which can lead to thrombocytopenia and the presence of giant platelets.[77]

GALE function has been modeled in worm (*C. elegans*) and fruit fly (*D. melanogaster*) and has shown to be essential for normal development.[106,107] In. *C. elegans*, complete knockout of *gale-1*, the single homolog of human *GALE*, is embryonic lethal.[107] Partial loss-of-function leads to abnormal levels of UDP-gal and UDP-galNAc. Mutant worms develop more slowly than wildtype, display misshapen gonads, and have activated unfolded protein response. In *D. melanogaster*, complete loss of *dGALE*, the homolog of human *GALE,* is embryonic lethal.[106] Complementation with *E. coli* GALE (*eGALE*), which only interconverts UDP-gal and UDP-glc, along with *P. shigelloides* GALE (*wbgU*), which only interconverts UDP-galNAc and UDP-glcNAc, rescues GALE function and restores organism viability. However, single complementation with either *eGALE* or *wbgU* leads to reduced fecundity. Loss of UDP-gal/UDP-glc interconversion results in reduced fecundity in both males and females, whereas loss of UDP-galNAc/UDP-glcNAc interconversion leads to reduced fecundity in only females. Taken

together, these observations suggest that complete loss-of-function of GALE in humans is likely to be lethal, and the consequences of glycosylation defects may be broad and yet to be fully appreciated.

Emerging evidence has begun to shed light on the importance of glycosylation for hematopoiesis, in particular for homing of hematopoietic cells[108,109] and platelet biogenesis.[110–112] Glycosylation products with sialylated N-acetyl lactosamine (LacNAc) help regulate platelet lifespan and production of hepatic thrombopoietin. These products are found in abundance on β1-integrin, a membrane protein enriched on platelets that is important for interactions with the extracellular matrix and homing.[113–116] LacNAc is a dimer of galactose and glcNAc, and its synthesis is regulated by β1,4-galactosyltransferase 1 (*β4GALT1)*. Deletion of *β4GALT1* in mice results in severe thrombocytopenia, with defective megakaryocyte localization to bone marrow.[117] In another study, although genetic ablation of *GALE* and *β4GALT1* with CRISPR/Cas9 in mice did not affect *in vitro* growth rates of hematopoietic progenitors, engraftment was noticeably reduced.[118] Loss of function of *GALE* leads to glycosylation defects, as modeled in a mutant Chinese hamster ovary (CHO) cell line that did not properly express low density lipoprotein (LDL) receptor.[95] It was shown that these cells have nearly undetectable GALE enzyme activity, leading to impaired N-linked and O-linked glycosylation, and hence to impaired LDL receptor glycosylation and increased degradation. These observations suggest that *GALE* may play an important role in the glycosylation of molecules involved in cell adhesion and homing. Defects in GALE enzyme function may lead to impairment of megakaryocyte and platelet progenitors with respect to either homing into their niches or interacting properly with the extracellular matrix.

In conclusion, we have identified GALE p.R51W, a homozygous mutation in a large Bedouin family with affected individuals presenting with severe thrombocytopenia and intracranial bleeding. GALE p.R51W causes decreased enzyme function and is unstable. The unique combination of mutation homozygosity, decreased enzyme function, and protein instability may help explain why the family's phenotype is severe, and different than individuals with other mutations in GALE. Future *in vitro* and *in vivo* studies with allele-specific perturbations in GALE may help elucidate the mechanisms of bone marrow failure in patients with GALE mutations and further reveal the importance of glycosylation on both platelet biogenesis and hematopoiesis.

5.6 FIGURES

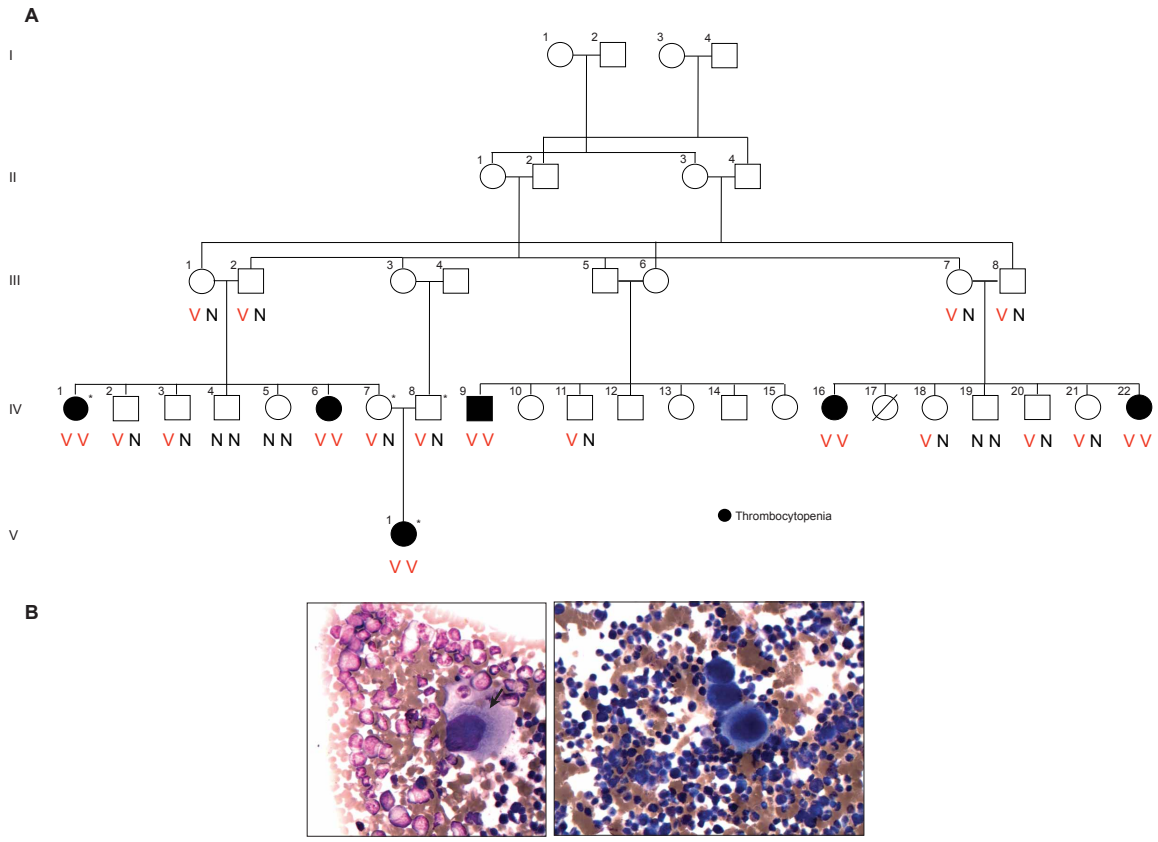


Figure 5.1. Pedigree and phenotype.

A) Family pedigree. Black indicates affected individual. V indicates variant allele (GALE c.C151T, p.R51W). N indicates reference allele. Asterisk indicates individual evaluated with whole exome sequencing. B) Histology of bone marrow aspirate in affected family member V-1. Samples shown with Wright-Giemsa stain. Arrow indicates megakaryocyte with hypogranular cytoplasm and hypolobated nuclei.

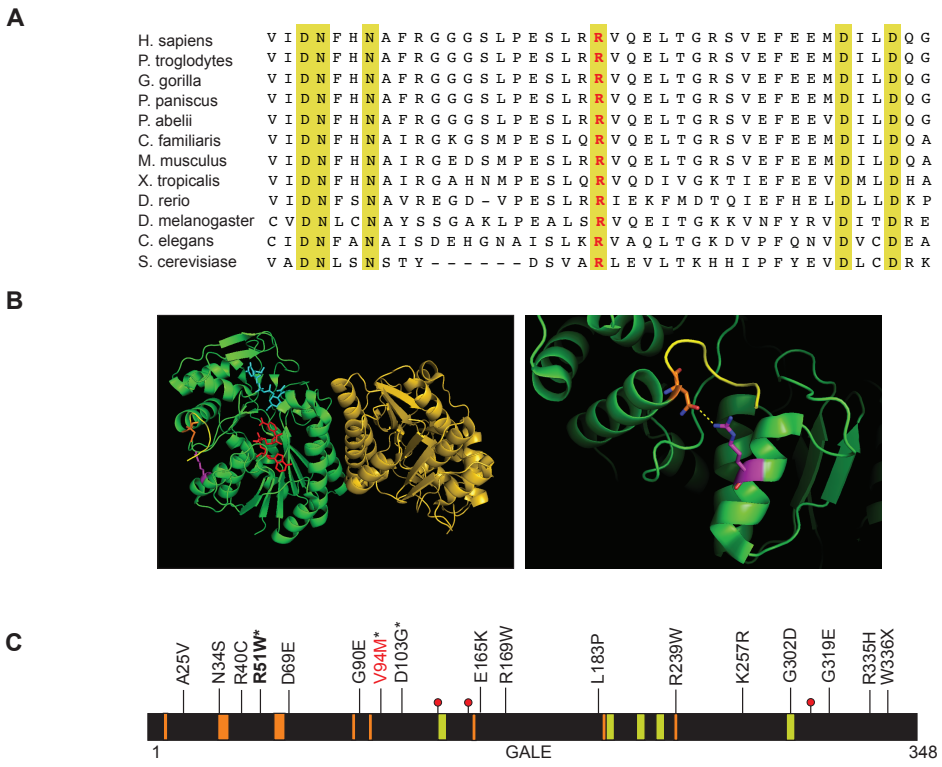


Figure 5.2. GALE mutations, conservation, and crystal structure.

A) Sequence alignment of multiple species showing evolutionary conservation of flanking region around Arg51. Gold indicates complete conservation through *S. cerevisiae*. Arg51 in red text. B) GALE crystal structure (PDB ID: 1HZJ). Crystal structure viewed in PyMOL.[24] *Left:* GALE homodimer shown (green and gold). Magenta = Arg51, orange = Asn340, blue = UDP-glcNAc, red = NAD cofactor, yellow = C-terminal end of GALE. *Right*: Close-up view of Arg51 in close proximity to the C-terminal end of GALE. Dotted yellow line indicates distance between Arg51 and Asn340 (3Å). C) Alleles described in GALE. Asterisk indicates homozygous mutations. Most alleles have been described as compound heterozygotes, or the other alleles were not described. p.R51W shown in bold. Red text indicates mutation with severe generalized galactosemia phenotype. Gold = substrate-binding site, orange = NAD-binding site, red dot = catalytic (residues 132, 157) and substrate specificity (residue 307) sites.

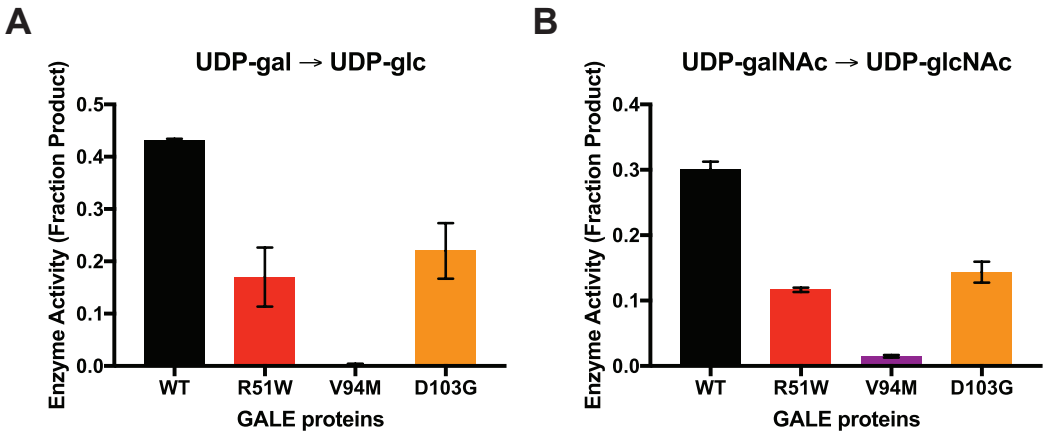


Figure 5.3. *In vitro* enzyme assays.

Enzyme assays performed with 10 ng protein. Substrates are either UDP-gal or UDP-galNAc as indicated. Mean fraction product is plotted per allele with ± SD. Experiment performed in triplicate.[82–84]

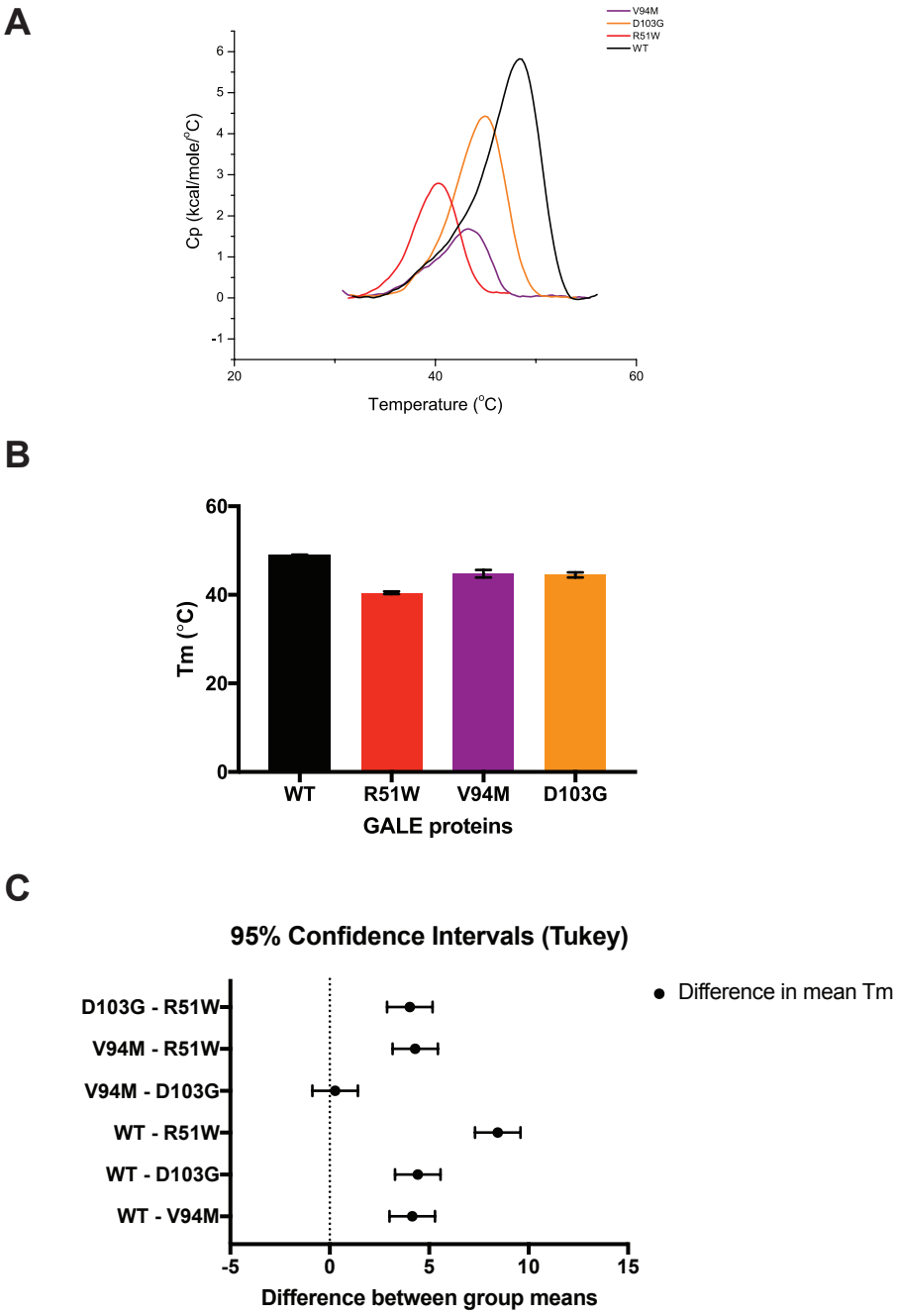


Figure 5.4. Differential scanning calorimetry.

A) Representative endotherm plots for each allele. Endotherm data has been normalized for concentration, and plots are shown with intrascan baseline correction. Black indicates wildtype protein (WT), red indicates p.R51W, purple indicates p.V94M, and orange indicates p.D103G. B) Melting point graph of each protein. Values represent mean melting points ± SD of four replicates across two independent experiments. C) 95% confidence interval plots from the *post hoc* Tukey test. Plots represent the difference of melting point (Tm) means between each protein.

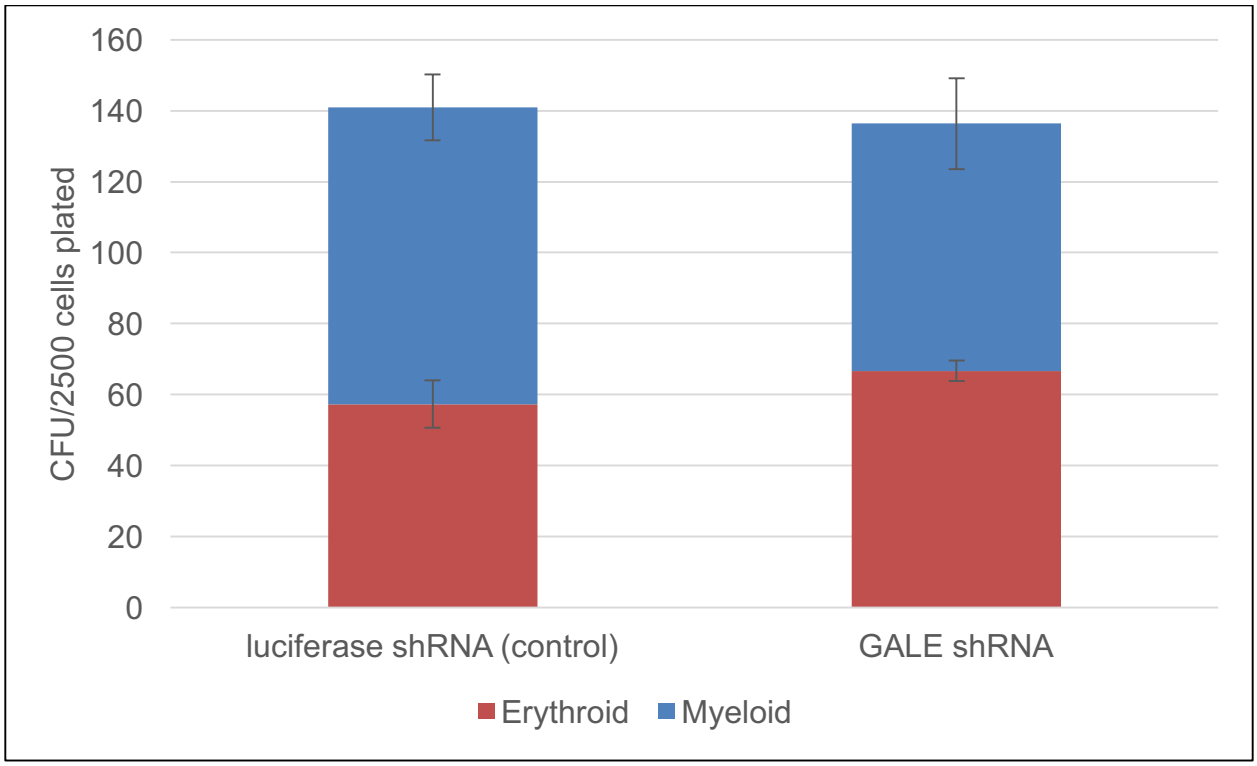


Figure 5.5. Erythroid and myeloid colony formation assay.

Methylcellulose colony formation assays. 2500 cells were plated and counted in triplicate. CFU, colony formation unit. Data shown is a representative graph of two independent experiments.

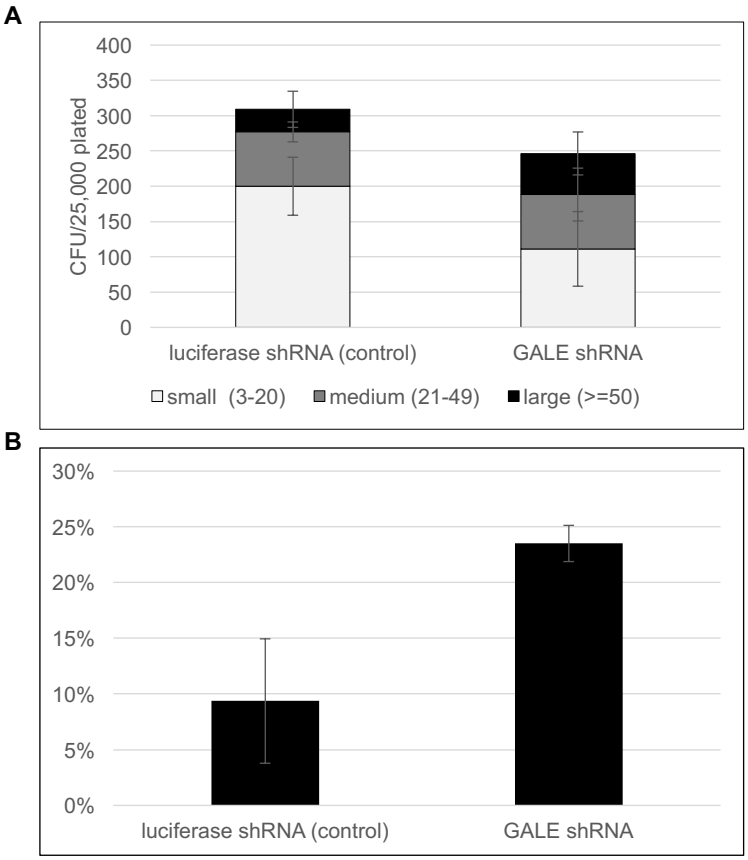


Figure 5.6. Megakaryocyte colony formation assay.

Collagen-based colony formation assays with megakaryocyte growth conditions. A) Colony counts binned to colony size. B) Percent of large colonies. CFU, colony formation unit. 25,000 cells were plated and counted in duplicate. Data is an average of two independent experiments.

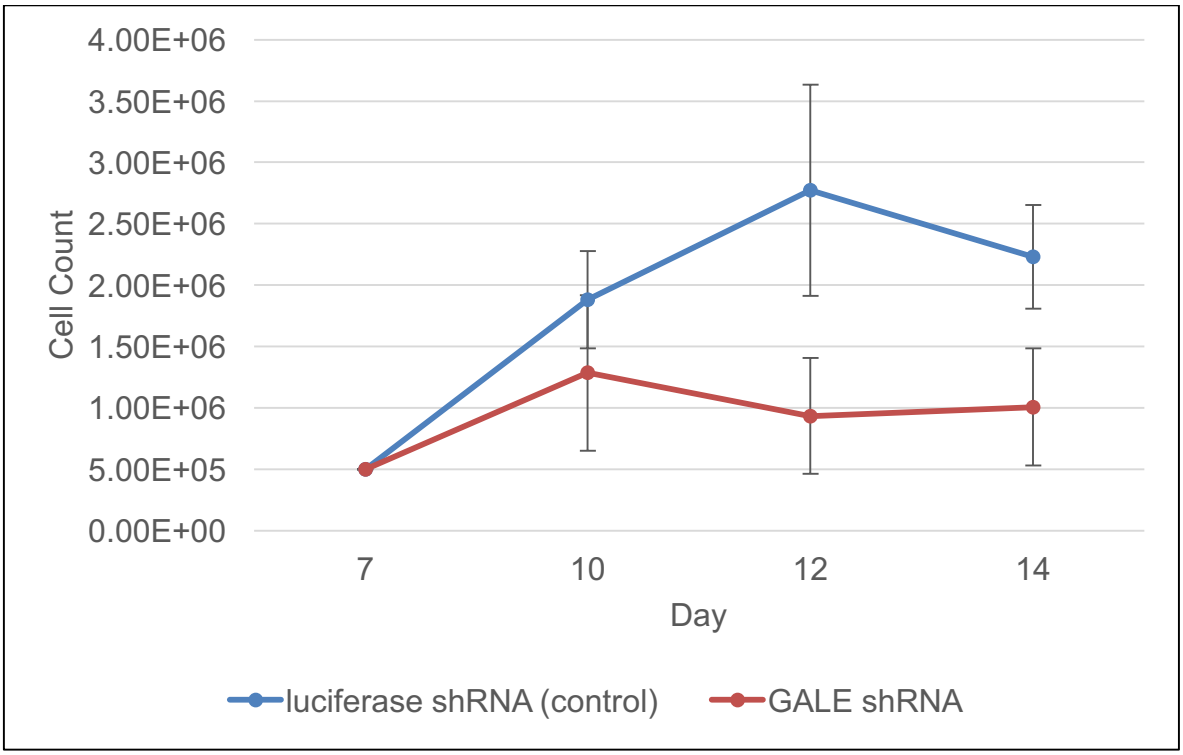


Figure 5.7. Megakaryocyte liquid culture proliferation.

Total proliferation cell counts of CD34+ cells grown under megakaryocyte liquid culture conditions. Data is an average of two independent experiments.

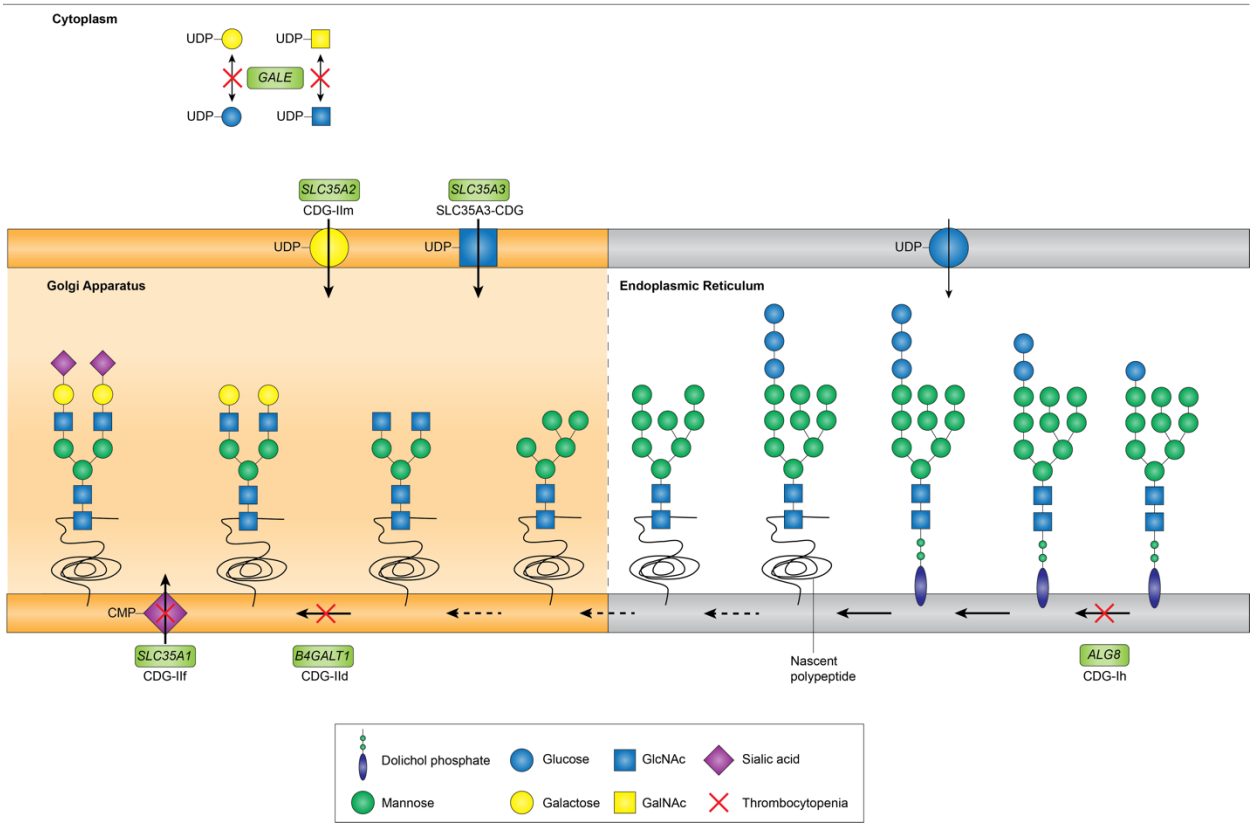


Figure 5.8. Glycosylation pathway and thrombocytopenias.

Schematic representation of a subset of the *N*-linked glycan biosynthesis pathway with genes associated with thrombocytopenia. Glycan biosynthesis begins in the cytoplasm, with subsequent modifications made in the endoplastic reticulum (ER) and Golgi apparatus (Golgi). Membrane transporters allow building-block sugars to enter the ER and Golgi. Enzymes in the pathway can control the levels of these sugars, add sugars to the glycan chain, or subtract sugars from the glycan chain. Mutations in this pathway cause congenital disorders of glycosylation (CDG). A subset of CDGs have been associated with thrombocytopenia. The steps in the pathway where mutations have been associated with thrombocytopenia are indicated with a red X. Genes responsible for a specific step in the pathway are indicated in green boxes. Congenital disorders of glycosylation (CDG) are annotated with the disorder name/CDG designation under the gene name. Figure design is adapted from Freeze 2006.[119]

Chapter 6. DISCUSSION

6.1 SUMMARY OF RESEARCH

This dissertation focused on uncovering new insights into the genetic basis of inherited bone marrow failure. To begin this dissertation, I started with two hypotheses. First, I hypothesized that patients with inherited bone marrow failure of previously unidentified genetic cause harbor mutations in as-yet-uncharacterized genes that are important for hematopoiesis and/or global cellular mechanisms of growth or maintenance. I also hypothesized that new mutations in genes known to be mutated in patients with bone marrow failure or other hematopoietic disorders may reveal new and important insights into gene function and biology.

The observations and results from the genetic studies in this dissertation have led to intriguing new insights into a wide range of genes, from genes with little-to-no known function to genes whose functions have been well-studied and established.

6.2 FAM111B AS A SHWACHMAN-DIAMOND SYNDROME-LIKE GENE

Clinical signs of bone marrow failure and exocrine pancreatic dysfunction coupled together are often associated with Shwachman-Diamond Syndrome, caused by biallelic loss-of-function of the *SBDS* gene.[9,13] We identified a young patient with clinical features overlapping Shwachman-Diamond Syndrome, although this patient did not have clinical evidence of severe cytopenias but rather hypocellular bone marrow. The proband presented with other clinical features rarely found in Shwachman-Diamond Syndrome: facial rash, sparse hair, hypohidrosis, and swelling of the extremities. This latter phenotype was observed in a dominant fashion in the family across multiple generations. Whole exome sequencing identified a heterozygous single-amino acid deletion in FAM111B (p.K421del) that co-segregates perfectly with the phenotype.

Previous heterozygous mutations in *FAM111B* have been associated with a disease resembling the facial rash phenotype in our family.[19] Though the early reports do not indicate that these patients have marrow failure or exocrine pancreatic dysfunction, these effects were not ascertained; it remains to be determined whether these patients have some deficits in the bone marrow or in exocrine pancreatic dysfunction at the enzyme level even without clinical symptoms. Indeed, another FAM111B mutation has now been reported with patients presenting with exocrine pancreatic dysfunction.[120]

The function of FAM111B remains unknown. Sequence homology shows that FAM111B contains a trypsin-like peptidase domain, suggesting that FAM111B may have enzymatic roles. Because FAM111B mutations result in syndromic disease affecting many organ systems, FAM111B likely plays a fundamental role in cellular homeostasis, growth, and maintenance. Identification and characterization of more patients with FAM111B mutations will likely aid in the understanding of the pathobiology and function of FAM111B in normal and disease states.

6.3 *THPO* MUTATIONS AND EXTRINSIC BONE MARROW FAILURE

Thrombopoietin, encoded by the gene *THPO*, is a cytokine important for hematopoietic stem cell maintenance and is the defining signal for megakaryocyte and platelet development.[121,122] Prior to our study, *THPO* mutations were generally associated with thrombocythemia, based on the characterization of a few gain-of-function mutations.[36,37] We identified two loss-of-function mutations in *THPO* (p.R99W, p.R157X) that when homozygous lead to trilineage bone marrow failure unresponsive to bone marrow transplant. Homozygosity for either mutation leads to absence of THPO in patients' blood. The underlying mechanism causing decreased serum THPO levels in patients with homozygous THPO p.R99W and p.R157X mutations remains

unclear. *In vitro* expression experiments of p.R99W show that intracellular production is comparable to normal levels, but that the secreted levels are significantly lower. It remains to be seen if secretion mechanisms are indeed impaired or if p.R99W is unstable, resulting in quickly degraded protein once secreted.

Because THPO is primarily produced in the liver and kidneys and not produced in the marrow, bone marrow transplant is not likely to cure disease. Loss of hepatocyte THPO production in mice is sufficient to cause bone marrow failure.[123] In our patients, treatment with the thrombopoietin receptor agonist romiplostim was able to restore trilineage hematopoiesis. This finding has profound implications for the management of suspected inherited bone marrow failure in that accurate and timely diagnosis is critical for proper treatment. Serum THPO levels should be measured in marrow failure patients with thrombocytopenia to rule out THPO defects as a hematopoietic-extrinsic cause of bone marrow failure. Further sequencing of patients with inherited bone marrow failure may reveal other loss-of-function mutations in *THPO*. Indeed, a subsequent study showed homozygous p.R119C mutations in three children with congenital amegakaryocytic thrombocytopenia that led to low serum THPO levels with romiplostim used as curative treatment.[124] Studying new alleles may help elucidate the importance of specific amino acid residues in proper THPO production, secretion, and function.

6.4 SURVIVING HOMOZYGOUS NONSENSE *BRCA1* MUTATIONS

BRCA1 is an essential gene for embryonic development; *BRCA1* aids in DNA repair of double strand breaks via homologus recombination. Though many breast and ovarian cancer patients carry inherited mutations in *BRCA1*, patients generally do not have two inherited mutations in *BRCA1*. Biallelic mutations in *BRCA1* have been previously shown to lead to a Fanconi Anemia-like phenotype. Individuals with bilallelic *BRCA1* mutations were hypothesized to have survived

because one allele is hypomorphic. We discovered two homozygous nonsense *BRCA1* mutations (p.W372X and p.L431X) in four patients across two families with a Fanconi Anemia-like phenotype characterized by chromosomal fragility, congenital anomalies, and childhood cancer predisposition. Both mutations lie in the same region of *BRCA1*, in the long exon 11, just 3-prime of an alternative splice site. The del11q transcript is a naturally occurring alternative transcript that utilizes this alternative splice site to create a shorter in-frame isoform. Del11q transcripts do not contain the deleterious nonsense mutations and thus are impervious to the effects of these mutations. We showed that del11q transcripts are enriched in patient cells as a result of nonsense-mediate decay of the full-length transcripts. We also showed that patient cells do not express full-length protein but rather a smaller isoform that corresponds in size to the del11q isoform.

Our results show that contextual interpretation of variants is important to evaluate functional consequence of mutations and to provide accurate genetic counseling to individuals. Our study adds further evidence that *BRCA1* is a Fanconi Anemia gene. Future sequencing evaluation of other consanguineous families with Fanconi Anemia, young onset breast/ovarian cancer, or childhood cancer may reveal other *BRCA1* mutations that are compatible with life, perhaps by mechanisms different from the one we described.

6.5 A GLYCOSYLATION GENE NEWLY ASSOCIATED WITH INHERITED THROMBOCYTOPENIA

Glycosylation is an important modification in hematopoiesis. Mutations in genes that affect glycosylation have been shown to cause thrombocytopenia and other forms of bone marrow failure. We identified a homozygous missense mutation in *GALE* (p.R51W) in a large and consanguineous Bedouin family whose affected individuals present with severe intracranial

bleeding and thrombocytopenia, mild anemia, and febrile neutropenia. We showed that p.R51W impairs the two enzyme functions of GALE: the interconversions of UDP-gal with UDP-glc and UDP-galNAc with UDP-glcNAc. We also showed that the p.R51W mutant protein is unstable, as evidenced by the significant decrease in melting point compared to wildtype.

Mutations in *GALE* have been previously associated with epimerase deficiency galactosemia. Enzyme deficiency is generally identified via newborn screening. Most individuals with enzyme deficiency at the peripheral blood level do not have clinical symptoms of galactosemia. One GALE allele (p.V94M) has been characterized when homozygous as a cause of severe galactosemia with symptoms of vomiting, hypotonia, jaundice, galactosuria, and hepatomegaly.[84] There is precedent for the association of GALE defects and thrombocytopenia. One patient has been described with GALE enzyme deficiency with a hematopoietic phenotype (thrombocytopenia)[92] similar to that of our family.

The importance of GALE function in hematopoiesis remains to be elucidated. One possible contributing mechanism is that GALE impairment leads to abnormal intracellular glycosylation processing, leading to ER stress. Other congenital disorders of glycosylation involving neutropenia have been shown to have increased ER stress response.[105,125] Also, a *C. elegans* *GALE* mutant model has been shown to activate the ER stress response pathway.[107] Future studies modeling GALE knockout or knockdown in hematopoietic cells may reveal whether or not GALE enzyme impairment leads to activation of the ER stress response pathway.

Preliminary *in vitro* shRNA knockdown studies in cord blood CD34+ hematopoietic progenitor cells have not yielded a striking defect in hematopoietic colony formation potential, though substantial knockdown of *GALE* with a single shRNA hairpin sequence causes impaired megakaryocyte proliferation. In addition, there is emerging evidence that *in vivo* mouse

transplantation of hematopoietic cells with CRISPR/Cas9-perturbation of *GALE* are characterized by decreased engraftment, suggesting that decreased GALE enzyme function may impair the glycosylation of cell surface proteins important for homing, thus impairing the ability of hematopoietic cells either to home into the bone marrow compartment or to progress to their final destination.

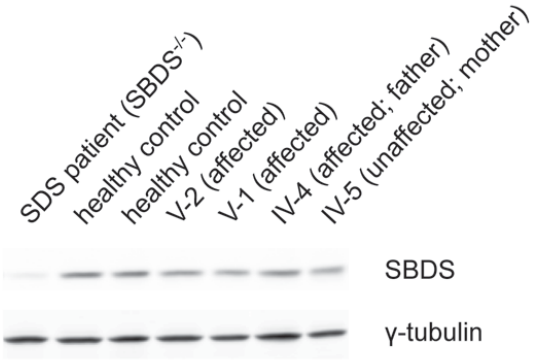
Future studies of GALE function, and of mutations that alter this function, may reveal important new insights into the role of glycosylation in hematopoiesis, specifically megakaryocyte development and platelet biogenesis. One challenge will be to model the right kind of GALE defect since the clinical consequence of GALE mutations is heterogeneous, and not all mutations lead to hematopoietic defects. Therefore, it may behoove future studies of GALE function to focus on the p.R51W allele, the only specific homozygous mutation that has been associated with severe thrombocytopenia.

6.6 FINAL THOUGHTS

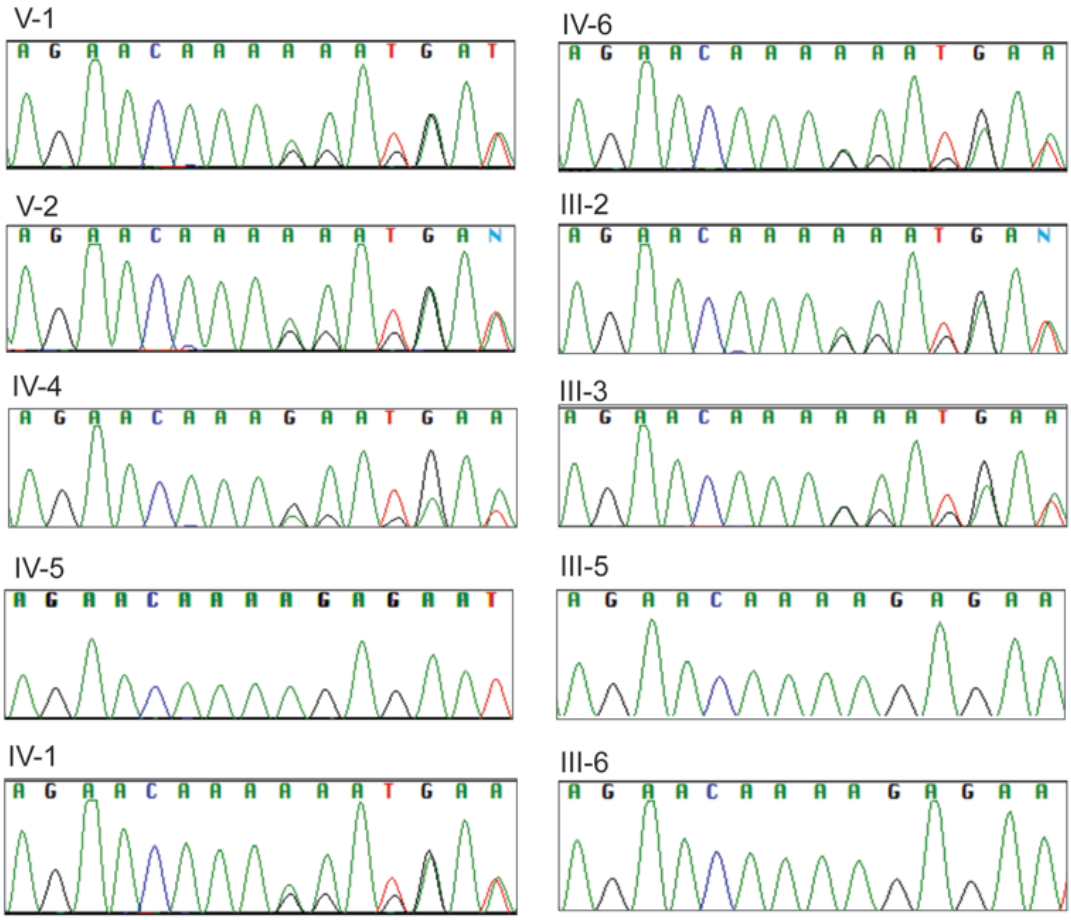
The goals of this dissertation were to identify and characterize new mutations and genes as causes of inherited bone marrow failure by sequencing the genomes of patients and families with likely inherited bone marrow failure, for which the genetic diagnosis was not readily apparent. This patient-oriented basic research approach has revealed a number of intriguing stories and explanations that we did not necessarily anticipate. The results have had significant implications in clinical management, genetic counseling, and our understanding of biology. This dissertation highlights the importance of studying rare genetic diseases. As sequencing becomes more and more accessible to clinicians and researchers alike, we anticipate that elucidating the genetic basis of rare diseases will reveal more key insights into cellular and developmental processes with broad consequence. In the long term, it is my hope that

characterization of new genes involved in human blood cell development will promote the development of rationally designed, targeted therapies.

APPENDIX A – SUPPLEMENTAL MATERIALS FOR CHAPTER 2.



Supplementary Figure 1. SBDS Immunoblot. A western blot of protein lysates from patient-derived lymphoblastoid cell lines was analyzed with antibodies against SBDS as well as γ-tubulin as an internal loading control. The patient with Shwachman-Diamond Syndrome (SDS) had biallelic damaging mutations in SBDS (homozygous c.258+2T>C)



Supplementary Figure 2. Sanger Sequencing of *FAM111B* variant. Sanger Sequencing traces confirming genotypes of individuals for the *FAM111B c.1261_1263delAAG* (p.Lys421del) variant.

APPENDIX B – SUPPLEMENTAL MATERIALS FOR CHAPTER 3

Supplemental Materials

Clinical Features

A Bedouin boy presented at the age of 9 months with isolated thrombocytopenia (Family A III-1; Figure 1A, Supplementary Table 2). His platelet (PLT) count was $40x10^9$/L. Physical examination was unremarkable and no physical anomalies were observed. His parents were first cousins. BM aspiration revealed normal cellularity and trilineage hematopoiesis with paucity of megakaryocytes. At 3 years of age macrocytosis [mean corpuscular volume (MCV) of 100fl] was first noticed. A year later, he presented with pancytopenia. His white blood cell count (WBC) was $2.45x10^9$/L, absolute neutrophil count (ANC)- $0.5x10^9$/L, hemoglobin (Hb)- 99g/L, MCV- 98fl, and platelet (PLT) count was - $9x10^9$/L. The BM appeared hypocellular (5-10% cellularity). Megakaryocytes were absent. Bone marrow colony cultures yielded no erythroid (BFU-Es, CFU-Es) myeloid (CFU-Cs) or megakaryocytic (CFU-MK) colonies. There was no increase in chromosomal breakage with diepoxybutane. Telomere length by flow-and fluorescent in-situ hybridization (FISH) was normal. He soon became transfusion-dependent and underwent bone marrow transplantation from an HLA-A allele mis-matched unrelated female donor following a preparative conditioning regimen with fludarabine, cyclophosphamide and rabbit anti-thymocyte globulin. Prophylaxis against graft vs. host disease (GvHD) consisted of methotrexate and cyclosporine A. Despite the administration of G-CSF, the patient's WBC did not rise above 1.3 $x10^9$/L. FISH studies demonstrated the presence of the Y chromosome in 100% of bone marrow cells. Due to failure of engraftment, a second transplant using G-CSF mobilized peripheral blood derived stem cells from the same donor was performed after an interval of 60 days from the first procedure. Myeloablative conditioning included busulfan (with targeted dose monitoring), cyclophosphamide and rabbit anti-thymocyte globulin. The same GvHD prophylaxis was administered as in the first procedure. After transient improvement in the

neutrophil count (ANC>1x10^9/L), the patient again became profoundly pancytopenic. A bone marrow aspirate on day +70 revealed empty spicules but 100% XX cells by FISH. Immune suppression was discontinued without improvement in the blood counts. Four months after the second transplant, severe gastrointestinal and pulmonary hemorrhage led to his demise.

The younger brother (Family A III-6; Figure 1A, Supplementary Table 2) of the patient III-1 in Family A was thrombocytopenic at birth (PLT 21x10^9/L). BM aspirate demonstrated only mild decrease in megakaryocytes (not shown). Around 3 years of age, macrocytosis with pancytopenia developed (Hb-65g/L, WBC-4.5x10^9/L, ANC-0.51x10^9/L, PLT-3x10^9/L). THPO levels in the serum were undetectable. Empiric therapy with romiplostim (5 µg/kgBW/week) was followed by improved trilineage hematopoiesis with increased PLT, WBC, ANC and Hb levels (Figure 2).

A 2.5 year old boy (Family B II-1; Figure 1A, Supplementary Table 2) from a second family was referred to our clinic on April 1996 for evaluation of cutaneous bleeding. He was born after normal pregnancy and delivery to Arabic parents who were first-degree cousins. No other abnormalities were noted on physical examination. His blood count revealed a platelet count of 18x10^9/L, Hb of 104 g/L and MCV of 92.3fl. Bone marrow aspiration disclosed normal cellularity with paucity of megakaryocytes. He was lost to follow up for a year and a half but subsequently returned with pancytopenia and an aplastic marrow. As there was no HLA-matched related donor, two courses of anti-thymocyte globulin and cyclosporine A were administered with no improvement. Therefore, at age 4.5 years he underwent an unrelated HLA-matched (low resolution HLA-A and HLA-B, high resolution HLA-DRB1) cord blood transplant with a conditioning regimen consisting of busulfan and cyclophosphamide followed by rabbit anti-thymocyte globulin. Twenty-five million nucleated cells/kgBW containing 0.75 x10^5 CD34+ cells/kgBW (values at collection) were infused. The patient remained pancytopenic with no

hematopoietic recovery for 60 days. There was no evidence of donor cell engraftment by Restriction Fragment Length Polymorphism (RFLP) analysis of the patient's blood. A subsequent stem cell infusion from the patient's haploidentical father was administered following conditioning with fludarabine and rabbit anti-thymocyte globulin. Nineteen million positively-selected CD34+ cells/kgBW were administered, but despite a transient rise in the patient's neutrophil count, he remained pancytopenic. The patient's post-transplant course was complicated by recurrent fevers, refractory CMV retinitis, and hemorrhagic cystitis. Multi-organ failure ensued, and the patient died three months after the first stem cell graft.

During the search for an HLA-matched family donor, the patient's 2 year old brother (Family B II-2) was found to be thrombocytopenic (PLT $22x10^9$/L) with mild neutropenia (ANC $1.4x10^9$/L). Physical examination was unremarkable except for several hematomas. Bone marrow cellularity was normal however only scant megakaryocytes were present. He was followed for 2 years over which he progressively developed pancytopenia with hypoplastic marrow. Bone marrow cultures failed to yield any erythroid (BFU-Es, CFU-Es), myeloid (CFU-Cs) or megakaryocytic (CFU-MK) colonies. No HLA-matched donor was identified in the extended family. Over the next 3 years, pancytopenia developed and progressively reduced bone marrow cellularity was observed. At the age of 5.5 years, the patient received myeloablative doses of busulfan and cyclophosphamide together with orthoclone (OKT3) antibody, followed by infusion of CD34+ cell purified haploidentical paternal peripheral blood derived stem cells (10 million CD34+ cells/kgBW). The patient developed *pneumocystis jirovecii* pneumonia followed by CMV pneumonia. Two months later, in the absence of any sign of hematopoietic engraftment, he was treated with fludarabine and anti-thymocyte globulin followed by CD34+ enriched maternal haploidentical stem cells, but a disseminated infection with *pseudallescheria boydii* led to muti-organ failure and death after three weeks.

In a third family, the patient (Family C II-5; Figure 1A, Supplementary Table 2) was the fifth child of a consanguineous family of Saudi-Arabian origin. Following spontaneous vaginal delivery, the girl presented with a gluteal petechial rash attributed to isolated thrombocytopenia ($16x10^9$/L). Maternal allo-immune thrombocytopenia was ruled out. IVIG and steroid therapy showed no effect. She required platelet transfusions 1-2 times per week but had no severe bleeding complications. At the age of 3 months she developed moderate neutropenia with ANC 0.5-$0.9x10^9$/L but no recurrent or severe infections. At the age of 10 months, a bone marrow biopsy revealed a hypocellular marrow. Congenital amegakaryocytic thrombocytopenia (CAMT) and Shwachman-Diamond Syndrome (SDS) were ruled out by genetic analysis of *MPL* and *SBDS*. Fanconi Anemia (FA) was excluded by a normal chromosomal breakage assay. Cytogenetic analysis of the bone marrow did not show any clonal aberrations. A second bone marrow biopsy at the age of 18 months revealed low cellularity. The patient's father and most of her uncles suffer from complete albinism but no cases of bone marrow failure were reported in the family. A brother, 9 years of age, had a history of a bilateral cataract on the basis of a galactokinase deficiency but had a normal blood count. He was otherwise healthy and HLA-identical to the patient. The other two brothers and one sister are healthy. In the context of recurrent platelet transfusions, the patient developed severe allo-immune thrombocytopenia with an insufficient increment of her platelet counts after transfusions. Allo-antibodies were successfully depleted by plasmapheresis and a therapeutic course with rituximab. Bone marrow transplantation was performed at the age of 19 months with a graft from the above mentioned HLA-identical brother. Conditioning included treosulfan [30 g/m^2], fludarabine [160 mg/m^2], and thiotepa [15 mg/kg]. She received cyclosporine A (CsA) and mycophenolatemofetil (MMF) for GvHD prophylaxis. Although WBC reached >$1x10^9$/L on day +13 and ANC >$0.5x10^9$/L on day +16 with full donor chimerism, the patient remained dependent on transfusions with platelets and erythrocytes. MMF was stopped on day +37, and CsA was stopped on day +146. Approximately two months after transplantation, the WBC fell below $1x10^9$/L and was unresponsive to stimulation with G-

CSF over more than a week (maximal dose of 10 µg/kgBW/day). The bone marrow biopsy performed on day+88 after BMT showed a hypocellular bone marrow with mildly dysplastic erythropoiesis and granulopoiesis and absent megakaryopoiesis. Persistent complete donor chimerism was documented in her bone marrow and peripheral blood. Serum levels of thrombopoietin were undetectable in spite of low platelet counts. Blood thrombopoietin levels of the donor were normal. Whole exome sequencing identified a homozygous mutation in the *THPO* gene. We therefore started treating the patient with the thrombopoietin agonist romiplostim. A starting dose of 1 µg/kgBW per week was administered subcutaneously and well tolerated. The dose was increased weekly until stable platelet counts of more than $100x10^9$/L were achieved. The final dose was 5µg/kgBW/week of romiplostim. Under this treatment thrombocyte counts rose above $50x10^9$/L and reticulocytes above 20% after 6 weeks; ANC rose above $1x10^9$/L without further G-CSF therapy after 9 weeks (Figure 2). Bone marrow biopsy now showed regeneration of all three cell lines with only mild morphologic signs of dysplasia. The last follow-up was 13 months after BMT with romiplostim being administered every 2 weeks with a normal full blood count.

Methods

Cell Culture & Transfections: HEK-293T cells (ATCC; CRL-1573) were cultured with DMEM supplemented with 10% FBS, 2 mM glutamine, and penicllin/streptomycin. BaF3-Mpl cells (kind gift of Dr. Kenneth Kaushansky, Stony Brook University) were cultured in RPMI-1640 (Life Technologies, Carlsbad, CA) supplemented with 10% FBS, 2 mM glutamine, penicillin/streptomycin, and 20 ng/mL murine IL-3 (Peprotech, Rocky Hill, NJ).

Results

Genomic analysis: The THPO R99W variant was absent in public databases (see Supplemental Methods) dbSNP141, the Exome Variant Server, the 1000 Genomes Project, the ExAC database of 60,706 exomes, and has not been seen in our in-house databases.

Structural Modeling: Since the amino terminal sequences of EPO and THPO are highly homologous,[126] the crystal structure of EPO complexed with its receptor was compared to that of THPO to inform how the THPO R99W mutation might affect structure and/or receptor-binding. Comparison of the crystal structure of the N-terminal domain of THPO (PDB ID: 1V7M) to the EPO and EPO-receptor complex (PDB ID: 1CN4) revealed close similarity between the tertiary structure of THPO with that of EPO (Figure S1). The THPO R99 residue is structurally equivalent to EPO Q105, and both residues lie in alpha helices and fall outside of the receptor-interaction domain. Replacement of a charged basic arginine residue with a hydrophobic tryptophan residue may potentially alter THPO tertiary structure.

Supplementary Table 1. Gene Panel for high-throughput sequencing of inherited bone marrow failure and myelodysplastic syndrome genes.

ABCB7	*CTC1*	*GATA1* + UTR5	*MSH6*	*RPL31*	*SUZ12*
ABL1	*CXCR2*	*GATA2* + intron	*MYSM1*	*RPL35A*	*TAZ*
AK2	*CXCR4*	*GFI1*	*NBN*	*RPL5*	*TCIRG1*
ALAS2	*DKC1 + UTR5*	*HAX1*	*NHP2*	*RPS10*	*TERC*
ANKRD26 + UTR5	*DNMT3A*	*IDH1*	*NOP10*	*RPS14*	*TERT*
ASXL1	*ELANE*	*IDH2*	*NPM1*	*RPS17*	*TET2*
ATM	*ERCC1*	*IL2RG*	*NRAS*	*RPS19*	*THPO*
ATR	*ERCC4/FANCQ*	*JAGN1*	*PALB2*	*RPS24*	*TINF2*
ATRX	*EZH2*	*JAK2*	*PAX5*	*RPS26*	*TPP1*
BCL2L11	*FAM111B*	*JAK3*	*PML*	*RPS29*	*TP53*
BCOR	*FANCA*	*JARID2*	*PMS2*	*RPS7*	*U2AF1*
BCR	*FANCB*	*KIF23*	*PRKDC*	*RTEL1*	*U2AF2*
BLM	*FANCC*	*KIT*	*PRPF40B*	*RUNX1*	*USB1*
BRAF	*FANCD2*	*KLF1*	*PTPN11*	*SBDS*	*VPS45*
BRCA1	*FANCE*	*KMT2A*	*RAB27A*	*SEC23B*	*WAS*
BRCA2	*FANCF*	*KRAS*	*RAD50*	*SETBP1*	*WRAP53*
BRCC3	*FANCG*	*LIG4*	*RAD51C*	*SF1*	*WT1*
BRIP1	*FANCI*	*LUC7L2*	*RARA*	*SF3A1*	*XRCC2*
C15ORF41	*FANCL*	*LYST*	*RBM8A*	*SF3B1*	*ZRSR2*
CBL	*FANCM*	*MCM2*	*RMRP*	*SLX4*	
CDAN1	*FERMT1*	*MET*	*RNF168*	*SMC1A*	
CDKN2A	*FLI1*	*MLH1*	*RPL10*	*SMC3*	
CEBPA	*FLNB*	*MPL*	*RPL11*	*SRP72*	
CHEK2	*FLT3*	*MRE11A*	*RPL15*	*SRSF2*	
CSF3R	*G6PC3*	*MSH2*	*RPL26*	*STAG2*	

Supplementary Table 2. Hematological features and clinical course.

Family	Patient	Age (mo)	WBC $x10^9$/L	ANC $x10^9$/L	Hb g/L	MCV fl	PLT $x10^9$/L	BM morphology	Treatment	Outcome
A	III-1	9	NA	NA	NA	NA	NA	Normal		
		21	2.5	0.5	99	98	9	5-10% cellularity, no megakaryocytes	URD HSCT X2 (same donor)	Poor graft function despite full donor engraftment after 2nd Tx : Died
	III-6	Birth	7	3	157		21	Paucity of megakaryocytes		
		36	4.5	0.5	65		3			
		42							Romiplostim	Alive
B	II-1	30	7.1	1.5	112	93	18	Normal cellularity, Paucity of megakaryocytes		
		54	1. 8	0.6	74	NA	10	Severely Hypoplastic	Haploidentical HSCT	No engraftment: Died
	II-2	26	6.4	1.9	110	NA	15	Paucity of megakaryocytes		
		48	3.6	0.7	82	92	11	Hypoplastic marrow	Cord blood HSCT, haploidentical HSCT	Engraftment failure both transplants: Died
C	II-5	birth					16			
		19	6.9	0.5	70		3	Hypoplastic marrow with low cellularity	HLA-identical MSD BMT	Poor graft function despite full donor engraftment
		32	6.5	3.2	119		394		Romiplostim	alive

BMT: bone marrow transplant; HSCT: hematopoietic stem cell transplant; MSD: matched sibling donor; NA: not available; Tx: treatment; URD: unrelated donor

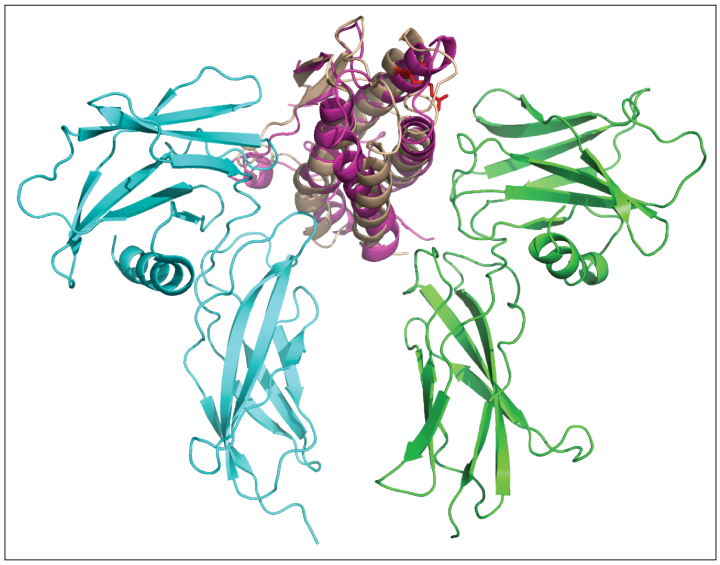


Figure S1. Crystal structure modeling. THPO (wheat, PDB ID: 1V7M) is superimposed using PDBeFold SSM on EPO (magenta) and EPO-receptor (cyan and green) complex (PDB ID: 1CN4). The R99 position in THPO is highlighted (red). The structures were modeled using PyMOL.

Figure S2

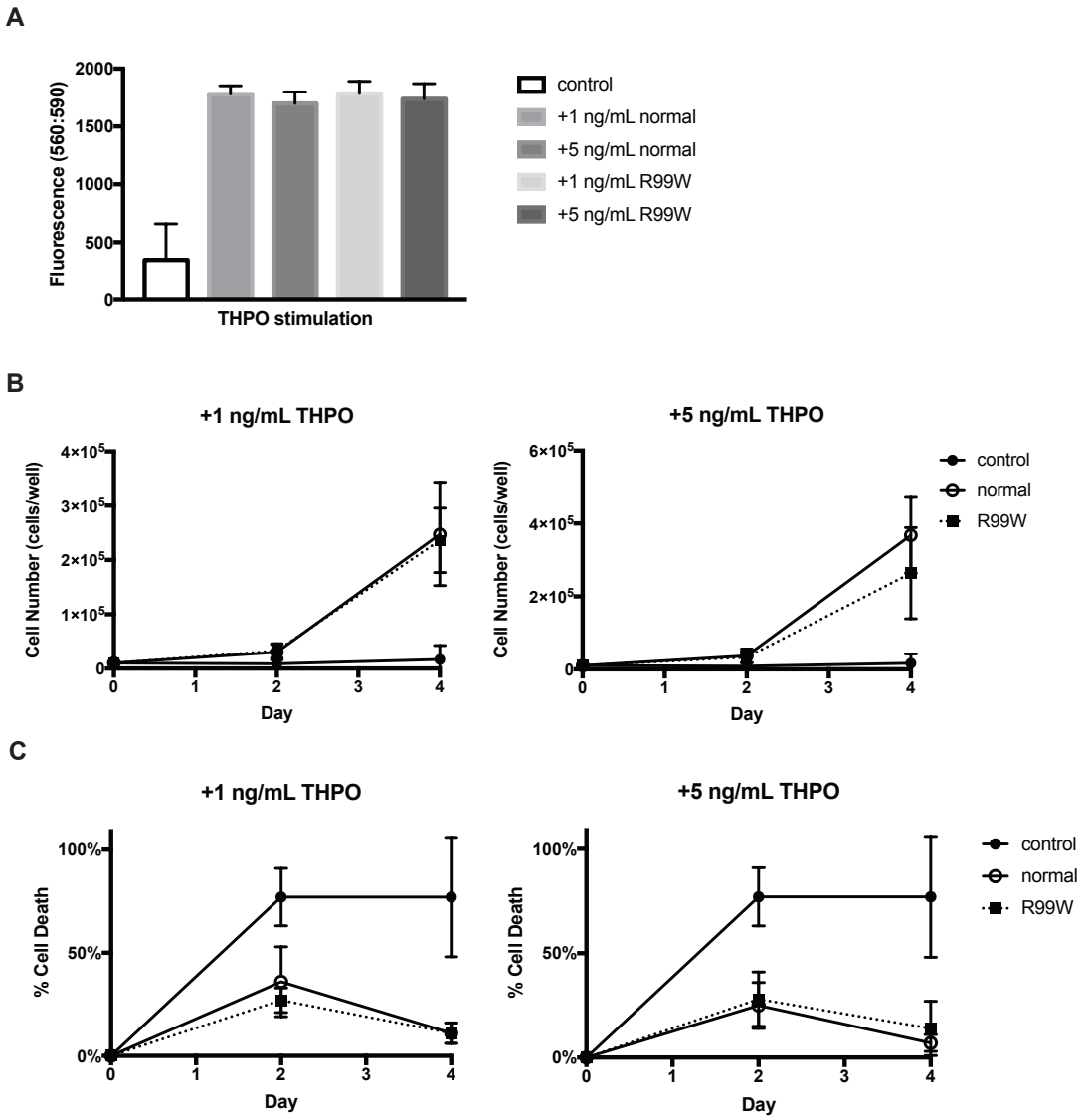
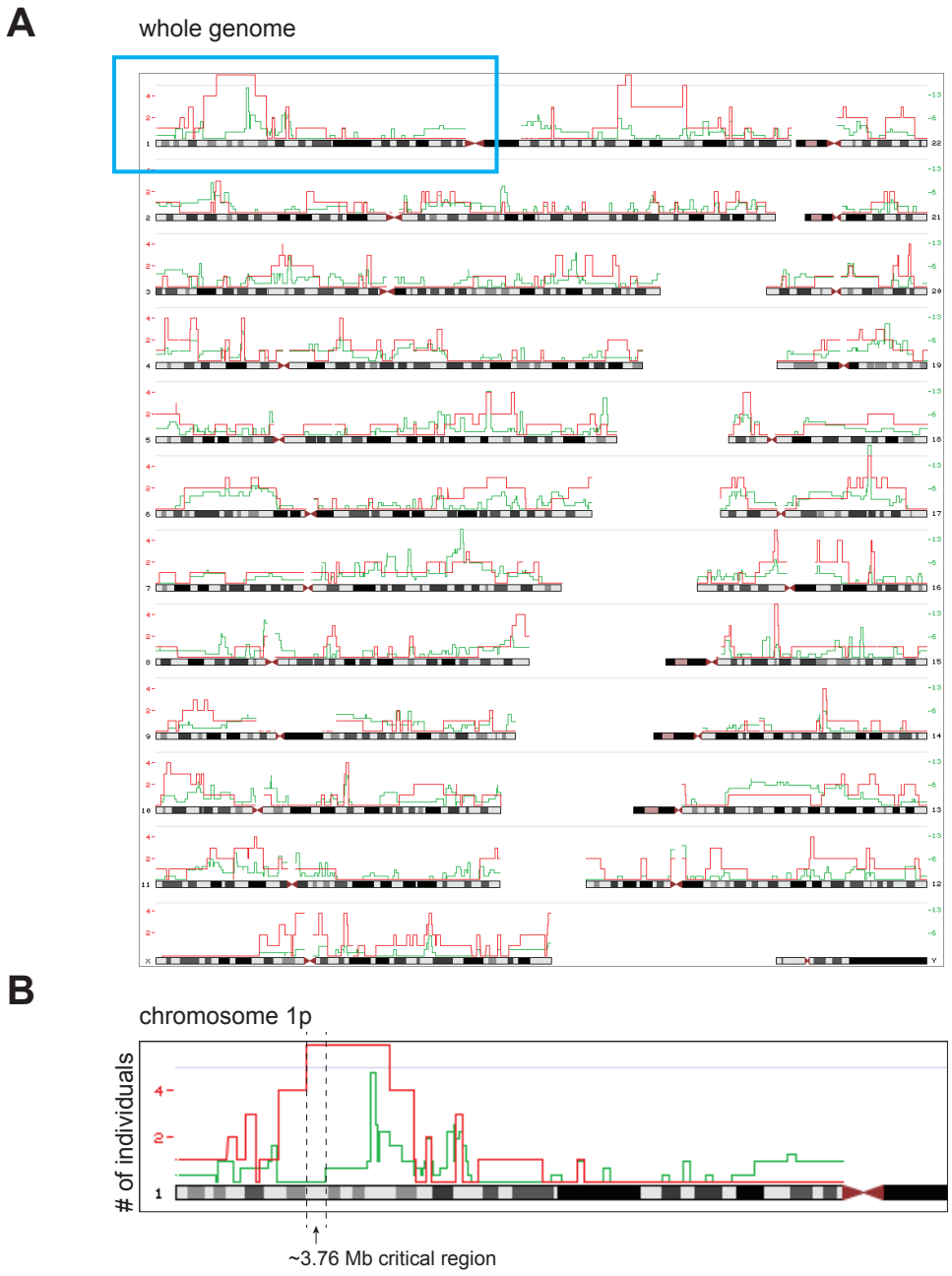
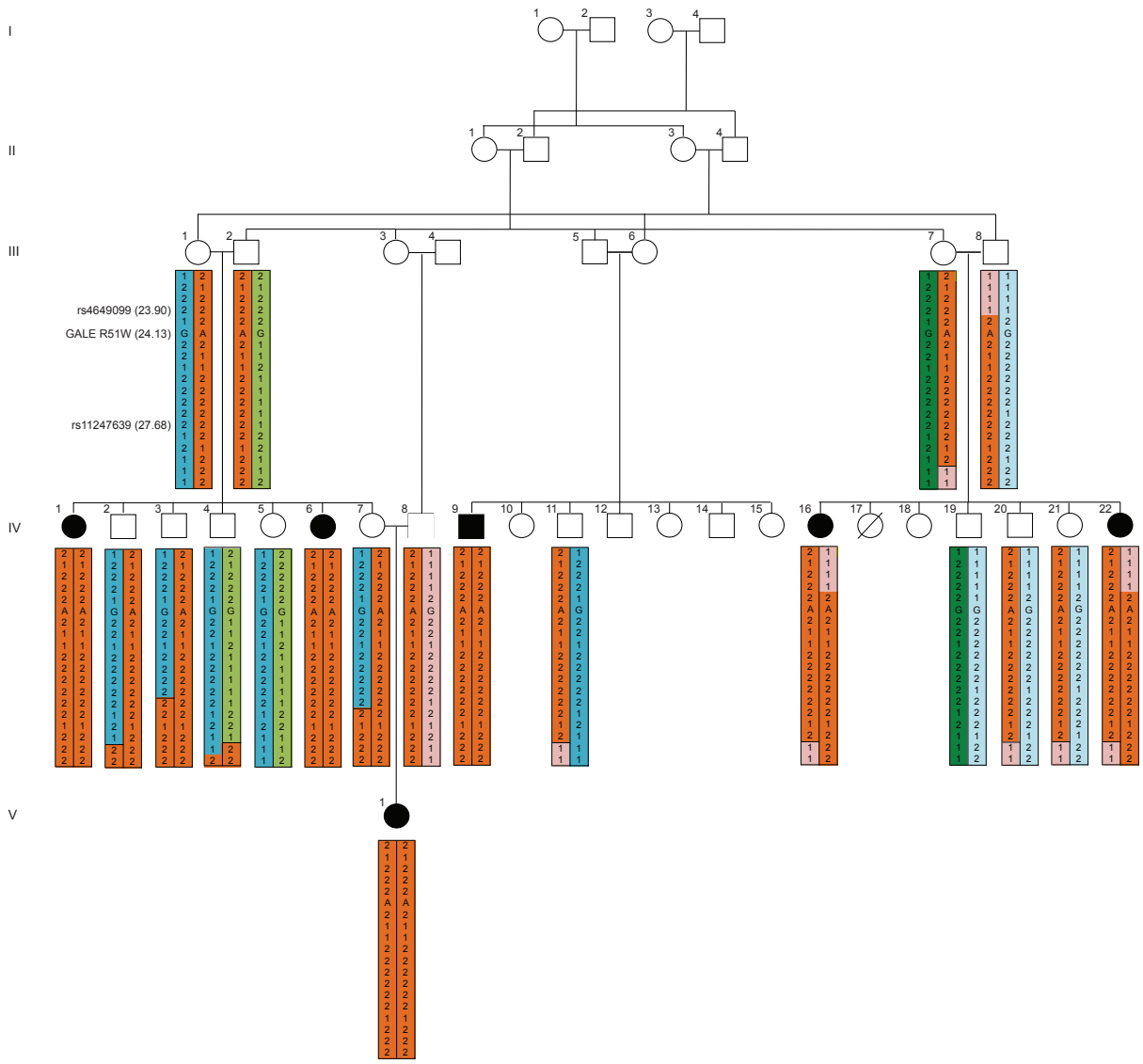


Figure S2. Functional characterization of THPO R99W. (A) Cell Viability: Baf3-Mpl cells were grown in the presence of normal or R99W THPO, and cell viability was measured with Almar Blue. Fluorescence was measured with wavelengths of 560nm for emission and 590nm for excitation. Control indicates no THPO. Figure represents the average of three experiments. (B) Cell proliferation: Baf3-Mpl cells were grown in the presence of 1ng/ml (left panel) or 5ng/ml (right panel) of normal or R99W THPO and stained with trypan blue at the indicated times. Cells were counted in triplicate for each experiment. Control indicates no THPO added. The average of three independent experiments is shown. (C) Cell death: Baf3-Mpl cells were grown in 1ng/ml (left panel) or 5ng/ml (right panel) of normal or R99W THPO and stained with trypan blue at the indicated times. The percentage of trypan blue-positive cells was determined in triplicate for each experiment. Control indicates no THPO added. The average of three independent experiments is shown.

APPENDIX C – SUPPLEMENTAL MATERIALS FOR CHAPTER 5

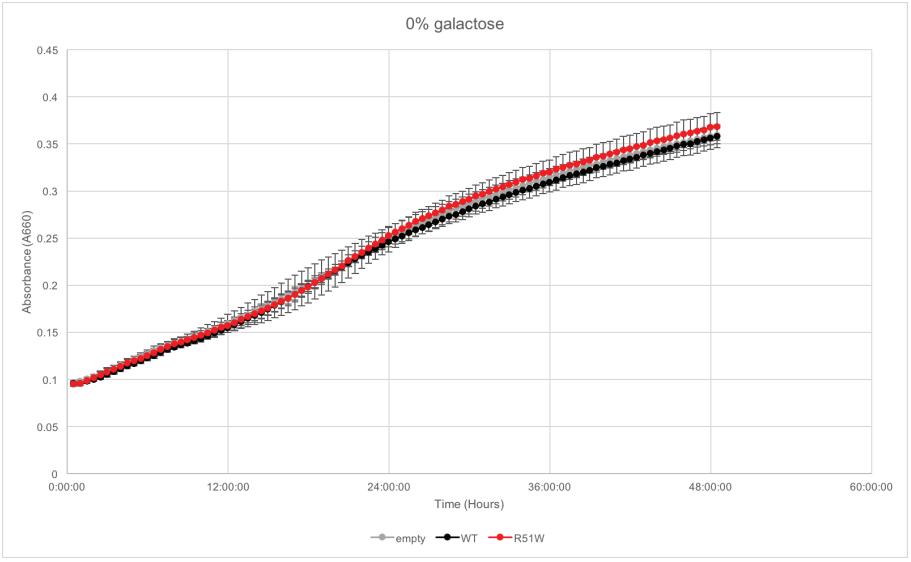


Supplementary Figure 1. Homozygosity Mapping.
Histogram plot overview of runs of homozygosity. Red and green indicates affected and unaffected individuals genotyped, respectively. A) Whole genome overview. Chromosomes 1-23 numbered left to right then top to bottom. Blue box indicates zoomed in view in Supplementary Figure 1B. B) Zoomed in view of chromosome 1p. Region between dashed lines indicate critical region of ~3.76 Mb where all affected individuals share homozygosity but unaffected individuals do not.

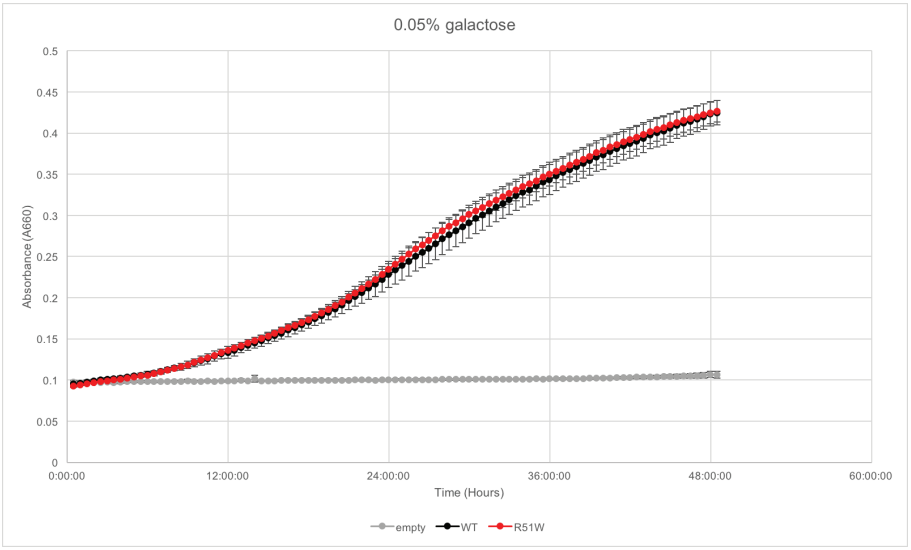


Supplementary Figure 2. Haplotypes.
Haplotypes determined with dense SNP chip data. GALE p.R51W (c.C151T) indicated as A (reference genomic alelle = G). Other SNP alleles marked as either allele 1 or 2. The boundaries of shared homozygosity amongst affected individuals are annotated next to the haplotypes in individual III-1 (approximate Mb coordinate in parantheses).

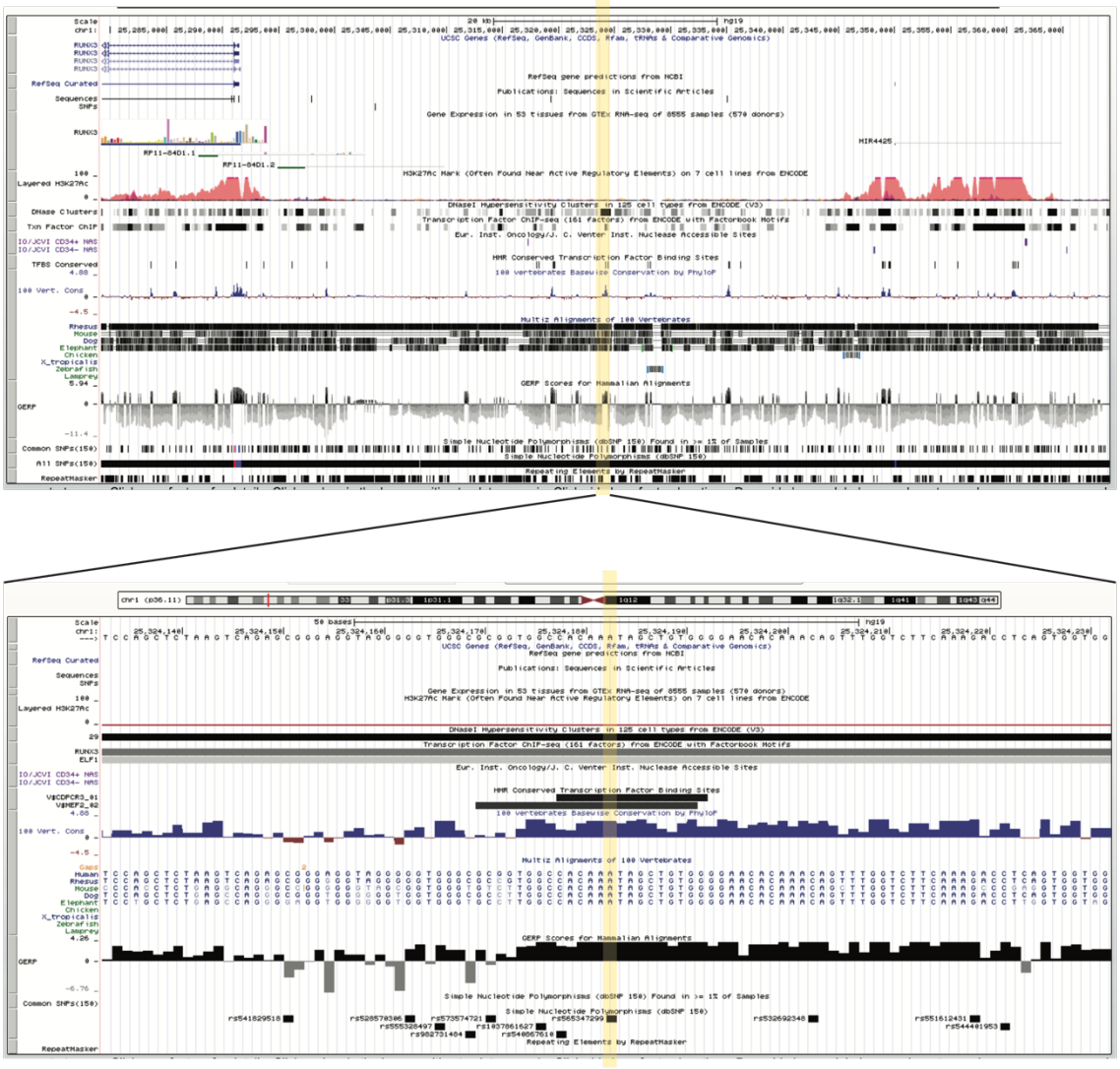
A



B



Supplementary Figure 3. Complementation Yeast Growth Assay with Galactose Challenge. *S. cerevisiae* with knockout of endogenous *GALE* locus was complemented with either empty vector, human GALE wildtype (WT), or human GALE R51W. Absorbance at 660 nm was measured every 30 min. Experiment performed in biological triplicate. A) Growth assay with no galactose in media B) Growth assay with 0.05% galactose in media.

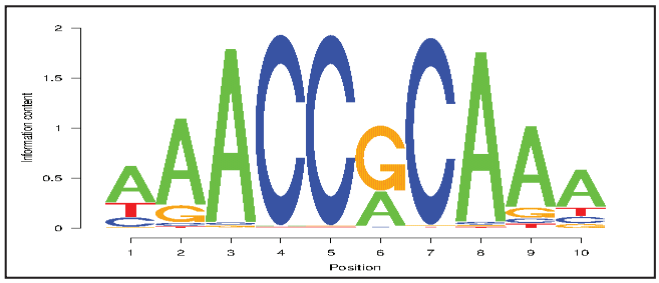


Supplementary Figure 4. UCSC Genome Browser view of intergenic variant upstream of *RUNX3.*
UCSC Genome Browser view of the intergenic variant upstream of RUNX3. Variant highlighted in yellow. Top: ~33kb flanking region around the variant showing RUNX3. Bottom: zoomed in view showing 100 bp flanking the variant. Browser tracks show potential transcription factor binding and sequence conservation across multiple vertebrate species.

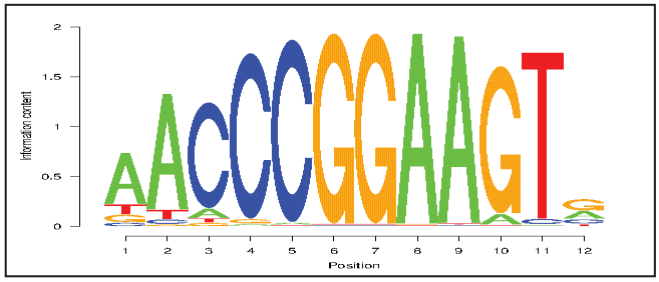
7 bp sequence
flanking intergenic variant

5′ GCCACAA**A**TAGCTGT 3′

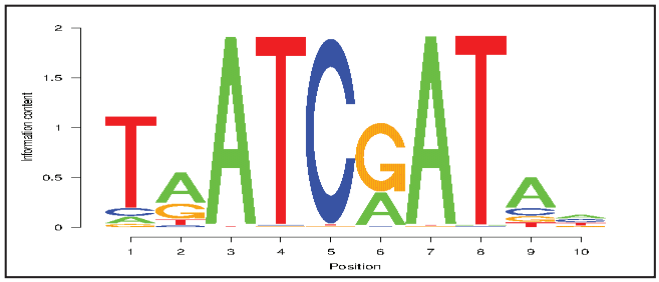
RUNX3



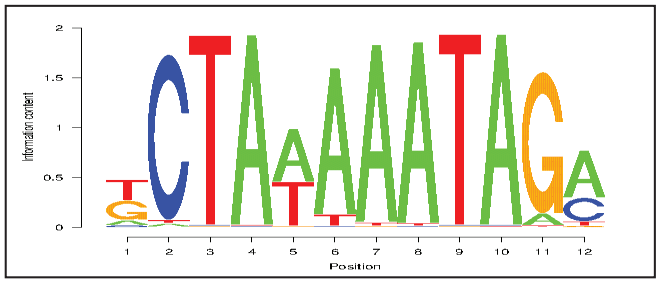
ELF1



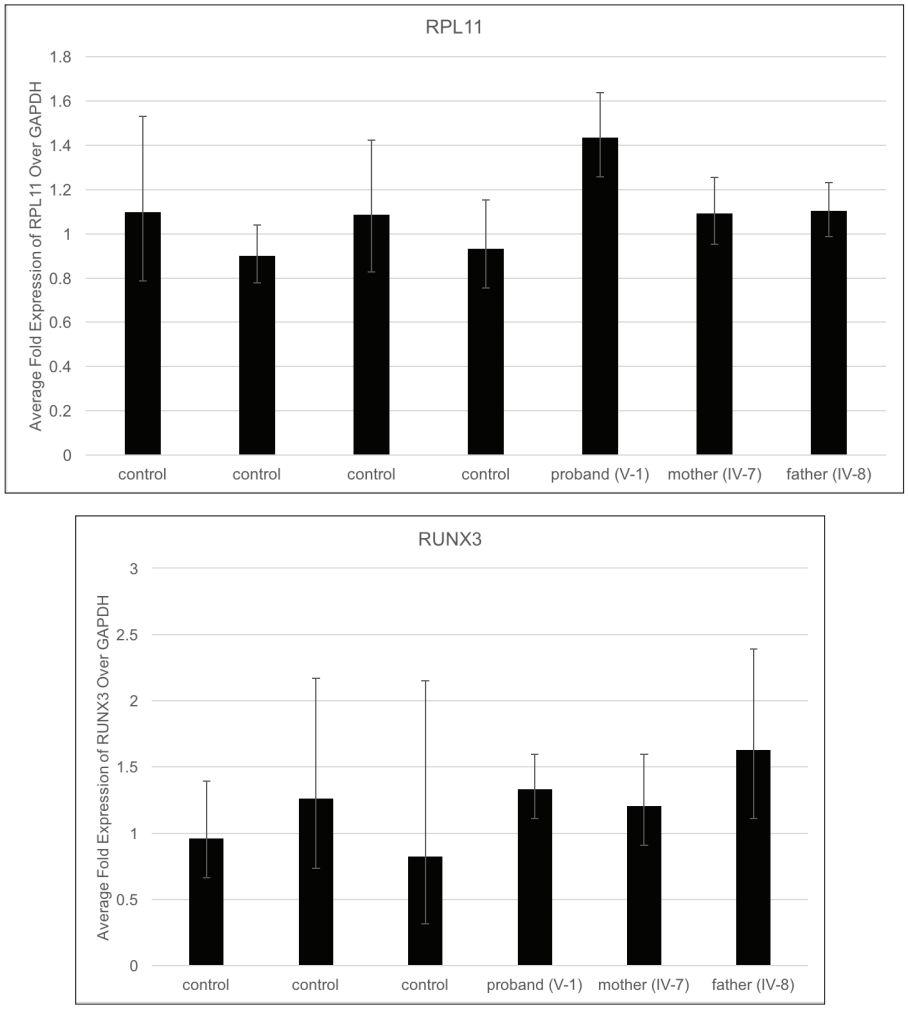
CUX1



MEF2A



Supplementary Figure 5. Transcription factor binding position weight matrices.
Position weight matrices for transcription factors prediced to bind the region surrounding the intergenic variant upstream of RUNX3 (rs565347299). Position weight matrices based on the JASPAR CORE 2018 database.



Supplementary Figure 6. Evaluation of noncoding variants by quantitative PCR.
Quantitative PCR (qPCR) was performed with TaqMan probes specific to *RPL11* and *RUNX3* using RNA extracted from control and patient-derived EBV-immortalized lymphoblastoid cell lines. Top: qPCR results for *RPL11* Bottom: qPCR results for *RUNX3*.

Alternative genomic causes of disease: whole genome sequencing

To assess the non-coding mutational landscape in the shared region of homozygosity in the affected individuals, we performed whole genome sequencing on one affected individual (IV-22), her unaffected parent (III-7), and an unaffected sibling (IV-19). There were 63 rare variants identified in the critical region that segregated with the phenotype in the sequenced individuals (Supplementary Table 1). GALE p.R51W was the only variant causing a coding change, confirming our whole exome sequencing results. Of these 63 variants, two variants were near genes associated with bone marrow failure or likely to be associated with bone marrow failure: an intronic variant in *RPL11* and an intergenic variant approximately 33 kb upstream of *RUNX3.* The intronic variant in *RPL11* is not well-conserved (GERP score < 2), whereas the intergenic variant upstream of *RUNX3* is well-conserved (GERP score = 4.3).

Mutations in *RPL11* are associated with Diamond-Blackfan Anemia.[127–129] Mutations in *RUNX1* are associated with a familial platelet disorder with leukemia predisposition. *RUNX3* is a runt-related transcription factor with high sequence similarity to *RUNX1*. Double knockout of *RUNX1* and *RUNX3* in mice leads to a bone marrow failure phenotype.[130] Knockdown of *RUNX3* in human hematopoietic progenitor cells leads to a defect in megakaryopoiesis and erythropoiesis.[131] The intergenic variant that lies 33 kb upstream of *RUNX3* has been seen previously, with an allele frequency of 0.06%, and has been reported in the dbSNP database as rs565347299 (Supplementary Figure 4). ENCODE ChIP-Seq data shows transcription factor binding with RUNX3 and ELF1. HMM predictions of transcription factor binding include CUX1 and MEF2A. However, comparison of the genomic region flanking rs565347299 to transcription factor positional weight matrices based does not reveal strong matches (Supplementary Figure 5).[132] Mutations in these transcriptions have not been reported to be associated with bone marrow failure.

To evaluate whether or not the intronic variant in *RPL11* and rs565347299 upstream of *RUNX3* affect gene expression levels of these respective genes, we performed quantitative PCR on RNA extracted from EBV-immortalized lymphoblastoid cell lines of the proband (V-1), mother (IV-7), and father (IV-8). qPCR analysis did not show any significant gene expression knockdown of *RPL11* or *RUNX3* (Supplementary Figure 6).

Supplementary Table 1. Rare variants from whole genome sequencing.

Chr	Coordinate	Ref	Var	Type	Gene	Gerp	Conserved TFBS	Encode HMM	Chipseq	DNaseI
1	23,926,559	A	-	intergenic	ID3(dist=40274),MDS2(dist=27265)	-	-	05_Strong_Enhancer	-	-
1	24,020,197	G	A	intronic	RPL11	-	-	04_Strong_Enhancer	HEY1	5
1	24,125,191	G	A	missense	GALE	4.8	-	08_Insulator	CTCF_(SC-5916),CTCF	20
1	24,284,591	C	A	intergenic	MIR378F(dist=28954),PNRC2(dist=1710)	-	-	04_Strong_Enhancer	-	-
1	24,303,239	-	AG	intronic	SRSF10	-	-	02_Weak_Promoter	-	-
1	24,305,547	-	TT	intronic	SRSF10	-	-	01_Active_Promoter	-	19
1	24,515,681	TTCTTTCCTTCCTTCC	-	intergenic	IFNLR1(dist=1916),LOC284632(dist=11034)	-	-	05_Strong_Enhancer	-	-
1	24,527,525	C	T	ncRNA_intronic	LOC284632	-	-	101_Weak_Txn	-	-
1	24,589,456	T	C	intergenic	LOC284632(dist=51276),GRHL3(dist=56356)	-	-	102_Repressed	-	-
1	24,599,126	G	A	intergenic	LOC284632(dist=60946),GRHL3(dist=46686)	-	-	103_Heterochrom/lo	-	-
1	24,644,270	T	G	intergenic	LOC284632(dist=106090),GRHL3(dist=1542)	-	-	102_Repressed	-	-
1	24,698,825	C	T	intronic	STPG1	-	-	07_Weak_Enhancer	-	-
1	24,804,284	A	G	intergenic	NIPAL3(dist=4811),RCAN3AS(dist=18539)	-	-	103_Heterochrom/lo	-	-
1	24,815,413	G	A	intergenic	NIPAL3(dist=15940),RCAN3AS(dist=7410)	-	-	103_Heterochrom/lo	-	-
1	24,849,840	G	A	intronic	RCAN3	-	-	102_Repressed	-	-
1	24,854,540	C	T	intronic	RCAN3	-	-	101_Weak_Txn	-	-
1	24,892,064	TG	-	intronic	NCMAP	-	-	103_Heterochrom/lo	-	-
1	24,985,490	G	A	intronic	SRRM1	-	-	10_Txn_Elongation	-	-
1	25,086,850	-	A	intronic	CLIC4	-	-	101_Weak_Txn	-	-
1	25,184,753	A	G	intergenic	CLIC4(dist=13938),RUNX3(dist=41249)	-	-	101_Weak_Txn	-	-
1	25,186,263	-	CACACACACACACACA	intergenic	CLIC4(dist=15448),RUNX3(dist=39739)	-	-	101_Weak_Txn	-	-

			CACATATATT							
1	25,187,920	C	T	intergenic	CLIC4(dist=17105),RUNX3(dist=38082)	-	-	101_Weak_Txn	-	-
1	25,242,297	G	A	intronic	RUNX3	-	-	09_Txn_Transition	-	13
1	25,250,541	C	T	intronic	RUNX3	-	-	04_Strong_Enhancer	-	-
1	25,252,493	G	A	intronic	RUNX3	-	-	04_Strong_Enhancer	Pol2-4H8	-
1	25,324,183	A	G	intergenic	RUNX3(dist=32682),MIR4425(dist=25811)	4.3	CDPCR3_01(865)	07_Weak_Enhancer	-	7
1	25,366,918	G	A	intergenic	MIR4425(dist=16841),SYF2(dist=181849)	-	-	101_Weak_Txn	-	-
1	25,398,734	G	A	intergenic	MIR4425(dist=48657),SYF2(dist=150033)	-	-	103_Heterochrom/lo	-	-
1	25,455,052	G	T	intergenic	MIR4425(dist=104975),SYF2(dist=93715)	-	-	05_Strong_Enhancer	Egr-1,EBF,EBF1_(C-8)	-
1	25,893,961	C	T	UTR3	LDLRAP1	-	-	101_Weak_Txn	-	-
1	25,905,634	C	T	intergenic	LDLRAP1(dist=10257),MAN1C1(dist=38325)	2.2	-	103_Heterochrom/lo	-	-
1	25,913,308	T	G	intergenic	LDLRAP1(dist=17931),MAN1C1(dist=30651)	-	-	07_Weak_Enhancer	-	2
1	25,916,449	C	T	intergenic	LDLRAP1(dist=21072),MAN1C1(dist=27510)	-	-	103_Heterochrom/lo	-	-
1	25,927,260	-	T	intergenic	LDLRAP1(dist=31883),MAN1C1(dist=16699)	-	-	06_Weak_Enhancer	-	-
1	25,972,031	A	C	intronic	MAN1C1	-	-	102_Repressed	-	-
1	26,020,117	A	G	intronic	MAN1C1	-	-	102_Repressed	-	-
1	26,044,169	C	T	intronic	MAN1C1	-	-	103_Heterochrom/lo	-	-
1	26,068,576	G	A	intronic	MAN1C1	2.1	-	102_Repressed	-	-
1	26,098,251	C	T	silent	MAN1C1	-	-	07_Weak_Enhancer	-	26
1	26,258,357	T	C	intergenic	STMN1(dist=24989),PAFAH2(dist=27901)	-	-	103_Heterochrom/lo	-	-
1	26,376,537	G	A	intergenic	SLC30A2(dist=3908),TRIM63(dist=1259)	-	-	102_Repressed	E2F6_(H-50),eGFP-JunD,USF2,USF-1,E2F6	18
1	26,395,991	-	TT	intergenic	TRIM63(dist=1866),PDIK1L(dist=41665)	-	-	103_Heterochrom/lo	-	-
1	26,580,028	G	A	intronic	CEP85	-	-	10_Txn_Elongation	-	-
1	26,584,502	C	G	intronic	CEP85	-	-	10_Txn_Elongation	-	-

1	26,586,360	T	A	intronic	CEP85	-	-	10_Txn_Elongation	-	14
1	26,627,597	G	A	intronic	UBXN11	-	-	103_Heterochrom/lo	AP-2gamma,AP-2alpha	-
1	26,682,528	C	G	intergenic	AIM1L(dist=1907),ZNF683(dist=5597)	-	-	103_Heterochrom/lo	-	-
1	26,689,759	C	T	intronic	ZNF683	-	-	102_Repressed	-	-
1	26,728,707	C	T	intergenic	ZNF683(dist=29441),LIN28A(dist=8562)	-	-	103_Heterochrom/lo	STAT3	-
1	26,736,869	GTGTGTGTGT	-	upstream	LIN28A	-	RREB1_01(869)	102_Repressed	BCL11A,SP1,ZBTB7A_(SC-34508),Egr-1,USF2,USF-1	-
1	26,884,689	C	T	intronic	RPS6KA1	-	-	10_Txn_Elongation	-	-
1	26,959,592	-	A	intergenic	RPS6KA1(dist=58072),ARID1A(dist=62930)	-	-	07_Weak_Enhancer	-	-
1	27,056,882	C	A	intronic	ARID1A	-	-	09_Txn_Transition	-	-
1	27,084,517	T	C	intronic	ARID1A	3.8	-	10_Txn_Elongation	-	-
1	27,142,745	GAAA	-	intergenic	PIGV(dist=17851),ZDHHC18(dist=10453)	-	-	101_Weak_Txn	-	-
1	27,232,334	C	G	intergenic	GPATCH3(dist=5372),NR0B2(dist=5641)	-	-	103_Heterochrom/lo	-	-
1	27,304,585	T	C	intergenic	KDF1(dist=17684),TRNP1(dist=1561 0)	-	-	103_Heterochrom/lo	-	-
1	27,320,968	C	T	UTR3	TRNP1	2.3	-	03_Poised_Promoter	Pol2,GABP,TAF1,ZBTB7A_(SC-34508),Pol2-4H8	-
1	27,343,346	C	A	intergenic	FAM46B(dist=4013),SLC9A1(dist=81954)	-	-	103_Heterochrom/lo	-	-
1	27,354,979	-	AAA	intergenic	FAM46B(dist=15646),SLC9A1(dist=70321)	-	-	103_Heterochrom/lo	-	-
1	27,440,503	G	A	silent	SLC9A1	-	-	10_Txn_Elongation	-	-
1	27,564,845	C	T	intronic	WDTC1	-	-	101_Weak_Txn	-	-
1	27,625,246	C	T	intronic	WDTC1	3.0	-	101_Weak_Txn	-	5

Supplementary Methods

Yeast Growth Assays.

Yeast growth assays were performed as previously described[84,97,98]. The yeast strain used was JFy3835 a *gal10-*, *gal80-* haploid strain derived from W303 (MATa *ade2-1 his3-11,15 leu 2-3,112 ura3-1 trp1-1 can1-100 RAD5+*) (kind gift of Dr. Judy Fridovitch-Keil, Emory University, Atlanta, GA). Low copy number (*CEN*) plasmids (pMM33) containing GALE cDNA sequence were transformed into JFy3835 and grown with and without 0.05% galactose. pMM33-hGALE-R51W was created via Q5 site-directed mutagenesis (New England Biolabs) using pMM33-hGALE-WT (kind gift of Dr. Judith Fridovich-Keil, Emory University, Atlanta, GA). Experiment was performed in the Synergy H1 microplate reader (Biotek) over 48 hours, with measurements taken every 30 min.

UPLC-MS/MS Instrument and Parameters.

The Waters Acquity LC, Waters Xevo TQ MS/MS system was used with the following UPLC column: CORTECS UPLC HILIC Column, 90Å, 1.6um, 2.1mm x 100mm (15 cm column) (Waters). Column temperature was 30°C. Chromatographic pump conditions and gradients are listed below. Peak areas were integrated with TargetLynx software (Waters).

Chromatographic Pump Gradient:
Run time: 9.0 min
Seal wash: 5 min
Pressure limits: low: 0 psi, high: 15,000 psi

Step	Time (min)	Flow (mL/min)	%A	%B	Curve
1	(initial)	0.400	10.0	90.0	(initial)
2	5.00	0.400	20.0	80.0	6
3	5.50	0.400	40.0	60.0	6
4	6.50	0.400	40.0	60.0	6
5	9.00	0.400	10.0	90.0	1

- Solvent A = H_2O + 10mM Ammonium Formate + 0.125% (v/v) Formic Acid (FA)(Fisher UPLC-grade)
- Solvent B = 95:5 Acetonitrile:H_2O + 10mM Ammonium Formate + 0.125% (v/v) FA
- Strong Needle Wash = 95:5 ACN: H_2O
- Weak Needle Wash = 90:10 H_2O:ACN

Mass Spectrometer Tuning Parameters:
Negative ionization mode

Collision energy for different multiple-reaction monitoring (MRM) transitions

Compound Name	Parent (m/z)	Daughter (m/z)	Cone (V)	Collision (V)
UDP-gal-241	564.9000	241.0000	35	29
UDP-glc-241	564.9000	241.0100	35	29
UDP-gal-280	564.9000	279.9000	35	27
UDP-glc-280	564.9000	279.9100	35	27
UDP-gal-403	564.9000	403.0000	35	22
UDP-glc-403	564.9000	403.0100	35	22
UDP-galNAc-159	606.0000	159.0000	39	44
UDP-glcNAc-159	606.0000	159.0100	39	44
UDP-galNAc-273	606.0000	272.9000	39	32
UDP-glcNAc-273	606.0000	272.9100	39	32
UDP-galNAc-282	606.0000	282.0000	39	27
UDP-glcNAc-282	606.0000	282.0100	39	27
UDP-galNAc-385	606.0000	384.9000	39	24
UDP-glcNAc-385	606.0000	384.9100	39	24
UDP-galNAc-403	606.0000	403.0000	39	23
UDP-glcNAc-403	606.0000	403.0100	39	23

Sugar Retention Times:

Sugar	Retention Time (min.)
UDP-glc	4.37
UDP-gal	4.60
UDP-glcNAc	4.76
UDP-galNAc	4.98

Immunoblotting.

Whole cell lysates from all cell lines were extracted using RIPA lysis buffer as previously described.[80,133] 15-20μg lysates were resolved on a 4-12% Bis-Tris gel (Thermo Fisher Scientific) under denaturing conditions. Wet transfer was performed at 400 mA for 2 hours at 4°C. Membranes were blocked in LI-COR Odyssey Blocking Buffer. Primary antibody incubation was performed in the cold room overnight. Antibody to detect GALE was ab155277, diluted 1:2,500 (Abcam). Other antibodies were against β-actin, 1:10,000 (Sigma), αRb-800nm, and αMs-700nm (LI-COR). Membranes were developed using the LI-COR Odyssey Imager (LI-COR).

BIBLIOGRAPHY

1. Hsu AP, Sampaio EP, Khan J, et al. Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. *Blood*. 2011;118(10):2653–2655.
2. Zhang MY, Churpek JE, Keel SB, et al. Germline *ETV6* mutations in familial thrombocytopenia and hematologic malignancy. *Nat. Genet.* 2015;47(2):180–185.
3. Shimamura A, Alter BP. Pathophysiology and management of inherited bone marrow failure syndromes. *Blood Rev.* 2010;24(3):101–122.
4. Teo JT, Klaassen R, Fernandez C V, et al. Clinical and genetic analysis of unclassifiable inherited bone marrow failure syndromes. *Pediatrics*. 2008;122(1):e139-48.
5. Ng SB, Buckingham KJ, Lee C, et al. Exome sequencing identifies the cause of a mendelian disorder. TL - 42. *Nat. Genet.* 2009;42 VN-r(1):30–35.
6. Zhang MY, Keel SB, Walsh T, et al. Genomic analysis of bone marrow failure and myelodysplastic syndromes reveals phenotypic and diagnostic complexity. *Haematologica*. 2015;100(1):42–48.
7. Walsh T, Casadei S, Lee MK, et al. Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. *Proc. Natl. Acad. Sci. U. S. A.* 2011;108(44):18032–7.
8. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* 2016;17(6):333–351.
9. PR D. Inherited causes of exocrine pancreatic dysfunction. *Can. J. Gastroenterol.* 1997;11(2):145–152.
10. Riordan JR, Rommens JM, Kerem B, et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science*. 1989;245(4922):1066–73.
11. Kerem B, Rommens JM, Buchanan JA, et al. Identification of the cystic fibrosis gene: genetic analysis. *Science*. 1989;245(4922):1073–1080.
12. Rommens JM, Iannuzzi MC, Kerem B, et al. Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science*. 1989;245(4922):1059–1065.
13. Boocock GRB, Morrison JA, Popovic M, et al. Mutations in SBDS are associated with Shwachman-Diamond syndrome. *Nat. Genet.* 2003;33(1):97–101.
14. Austin KM, Leary RJ, Shimamura A. The Shwachman-Diamond SBDS protein localizes to the nucleolus. *Blood*. 2005;106(4):1253–1258.
15. Nord AS, Roeb W, Dickel DE, et al. Reduced transcript expression of genes affected by

inherited and de novo CNVs in autism. 19(6):727–731.

16. Ip WF, Dupuis A, Ellis L, et al. Serum pancreatic enzymes define the pancreatic phenotype in patients with Shwachman-Diamond syndrome. *J. Pediatr.* 2002;141(2):259–265.
17. Toiviainen-Salo S, Durie PR, Numminen K, et al. The natural history of Shwachman-Diamond syndrome-associated liver disease from childhood to adulthood. *J. Pediatr.* 2009;155(6):807–811.e2.
18. Myers KC, Davies SM, Shimamura A. Clinical and Molecular Pathophysiology of Shwachman-Diamond Syndrome. An Update. *Hematol. Oncol. Clin. North Am.* 2013;27(1):117–128.
19. Mercier S, Küry S, Shaboodien G, et al. Mutations in FAM111B cause hereditary fibrosing poikiloderma with tendon contracture, myopathy, and pulmonary fibrosis. *Am. J. Hum. Genet.* 2013;93(6):1100–1107.
20. Ritchie DS, Angus PW, Bhathal PS, Grigg AP. Liver failure complicating non-alcoholic steatohepatitis following allogeneic bone marrow transplantation for Shwachman-Diamond syndrome. 2002.
21. Sakamoto KM, Shimamura A, Davies SM. Congenital disorders of ribosome biogenesis and bone marrow failure. *Biol. Blood Marrow Transplant.* 2010;16(1 Suppl):S12–S17.
22. Parikh S, Bessler M. Recent insights into inherited bone marrow failure syndromes. *Curr. Opin. Pediatr.* 2012;24(1):23–32.
23. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 2004;60(12 I):2256–2268.
24. Schrodinger LLC. The PyMOL Molecular Graphics System, Version 1.8. 2015;
25. Bartley TD, Bogenberger J, Hunt P, et al. Identification and cloning of a megakaryocyte growth and development factor that is a ligand for the cytokine receptor Mpl. *Cell*. 1994;77(7):1117–1124.
26. Linden HM, Kaushansky K. The glycan domain of thrombopoietin enhances its secretion. *Biochemistry*. 2000;39(11):3044–3051.
27. Muto T, Feese MD, Shimada Y, et al. Functional analysis of the c-terminal region of recombinant human thrombopoietin. C-terminal region of thrombopoietin is a “shuttle” peptide to help secretion. *J. Biol. Chem.* 2000;275(16):12090–12094.
28. Linden HM, Kaushansky K. The glycan domain of thrombopoietin (TPO) acts in trans to enhance secretion of the hormone and other cytokines. *J. Biol. Chem.*

2002;277(38):35240–35247.

29. Wolber EM, Dame C, Fahnenstich H, et al. Expression of the thrombopoietin gene in human fetal and neonatal tissues. *Blood*. 1999;94(1):97–105.
30. Sungaran R, Markovic B, Chong BH. Localization and regulation of thrombopoietin mRNa expression in human kidney, liver, bone marrow, and spleen using in situ hybridization. *Blood*. 1997;89(1):101–7.
31. Foster DC, Sprecher CA, Grant FJ, et al. Human thrombopoietin: gene structure, cDNA sequence, expression, and chromosomal localization. *Proc. Natl. Acad. Sci. U. S. A.* 1994;91(26):13023–7.
32. Sungaran R, Chisholm OT, Markovic B, et al. The role of platelet alpha-granular proteins in the regulation of thrombopoietin messenger RNA expression in human bone marrow stromal cells. *Blood*. 2000;95(10):3094–3101.
33. Dasouki MJ, Rafi SK, Olm-Shipman AJ, et al. Exome sequencing reveals a thrombopoietin ligand mutation in a Micronesian family with autosomal recessive aplastic anemia. *Blood*. 2013;122(20):3440–9.
34. Mandrile G, Dubois A, Hoffman JD, et al. 3q26.33-3q27.2 microdeletion: A new microdeletion syndrome? *Eur. J. Med. Genet.* 2013;56(4):216–221.
35. Zeigler FC, de Sauvage F, Widmer HR, et al. In vitro megakaryocytopoietic and thrombopoietic activity of c-mpl ligand (TPO) on purified murine hematopoietic stem cells. *Blood*. 1994;84(12):4045–4052.
36. Zhang B, Ng D, Jones C, et al. A novel splice donor mutation in the thrombopoietin gene leads to exon 2 skipping in a Filipino family with hereditary thrombocythemia. *Blood*. 2012;118(26):6988–6990.
37. Kondo T, Okabe M, Sanada M, et al. Familial essential thrombocythemia associated with one-base deletion in the 5'-untranslated region of the thrombopoietin gene. *Blood*. 1998;92:1091–1096.
38. Liu K, Kralovics R, Rudzki Z, et al. A de novo splice donor mutation in the thrombopoietin gene causes hereditary thrombocythemia in a Polish family. *Haematologica*. 2008;93(5):706–714.
39. Gotlib J, Zhang B, Jones CD, et al. Identification of a novel splice donor mutation in the thrombopoietin gene in a philippine family with hereditary thrombocythemia. *Blood*. 2010;116(21):.
40. Ihara K, Ishii E, Eguchi M, et al. Identification of mutations in the c-mpl gene in congenital amegakaryocytic thrombocytopenia. *Proc. Natl. Acad. Sci. U. S. A.* 1999;96(6):3132–

3136.

41. Ballmaier M, Germeshausen M, Schulze H, et al. c-mpl mutations are the cause of congenital amegakaryocytic thrombocytopenia. *Blood*. 2001;97(1):139–146.
42. Walne AJ, Dokal A, Plagnol V, et al. Exome sequencing identifies MPL as a causative gene in familial Aplastic anemia. *Haematologica*. 2012;97(4):524–528.
43. Ballmaier M, Germeshausen M. Congenital amegakaryocytic thrombocytopenia: Clinical presentation, diagnosis, and treatment. *Semin. Thromb. Hemost.* 2011;37(6):673–681.
44. Gurney a L, Carver-Moore K, de Sauvage FJ, Moore MW. Thrombocytopenia in c-mpl-deficient mice. *Science*. 1994;265(5177):1445–7.
45. Alexander WS, Roberts a W, Nicola N a, Li R, Metcalf D. Deficiencies in progenitor cells of multiple hematopoietic lineages and defective megakaryocytopoiesis in mice lacking the thrombopoietic receptor c-Mpl. *Blood*. 1996;87(6):2162–2170.
46. Carver-Moore K, Broxmeyer HE, Luoh SM, et al. Low levels of erythroid and myeloid progenitors in thrombopoietin-and c-mpl-deficient mice. *Blood*. 1996;88(3):803–808.
47. Russell ES, Bernstein SE, Lawson FA, Smith LJ. Long-Continued Function of Normal Blood-Forming Tissue Transplanted Into Genetically Anemic Hosts. *JNCI J. Natl. Cancer Inst.* 1959;23(3):557–566.
48. Dexter TM, Moore MAS. In vitro duplication and “cure” of haemopoietic defects in genetically anaemic mice. *Nature*. 1977;(269):412–414.
49. Bernstein SE. Tissue transplantation as an analytic and therapeutic tool in hereditary anemias. *Am. J. Surg.* 1970;119(4):448–451.
50. Wolf NS. Dissecting the Hematopoietic Microenvironment: III. Evidence for a Positive Short Range Stimulus for Cellular Proliferation. *Cell Prolif.* 1978;11(4):335–345.
51. Ceccaldi R, Sarangi P, D’Andrea AD. The Fanconi anaemia pathway: new players and new functions. *Nat. Rev. Mol. Cell Biol.* 2016;17(6):337–349.
52. King M-C, Marks JH, Mandell JB. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science*. 2003;302(5645):643–646.
53. Norquist BM, Harrell MI, Brady MF, et al. Inherited mutations in women with ovarian carcinoma. *JAMA Oncol.* 2016;2(4):482–490.
54. Antoniou AC, Casadei S, Heikkinen T, et al. Breast-cancer risk in families with mutations in PALB2. *N Engl J Med*. 2014;371(6):497–506.
55. Pritchard CC, Mateo J, Walsh MF, et al. Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. *N Engl J Med*. 2016;375(5):443–453.
56. Gowen LC, Johnson BL, Latour AM, Sulik KK, Koller BH. *Brca1* deficiency results in early

embryonic lethality characterized by neuroepithelial abnormalities. *Nat. Genet.* 1996;12(2):191–194.

57. Hakem R, De La Pompa JL, Sirard C, et al. The tumor suppressor gene Brca1 is required for embryonic cellular proliferation in the mouse. *Cell*. 1996;85(7):1009–1023.
58. Ludwig T, Chapman DL, Papaioannou VE. Targeted mutations of breast cancer susceptibility gene homologs in mice: lethal phenotypes of *Brca1, Brca2, Brca1/Brca2, Brca1/p53, and Brca2/p53* nullizygous embryos. *Genes Dev.* 1997;1226–1241.
59. Domchek SM, Tang J, Stopfer J, et al. Biallelic deleterious *BRCA1* mutations in a woman with early-onset ovarian cancer. *Cancer Discov.* 2013;3(4):399–405.
60. Sawyer SL, Tian L, Kähkönen M, et al. Biallelic mutations in *BRCA1* cause a new Fanconi anemia subtype. *Cancer Discov.* 2015;5(2):135–142.
61. Wang Y, Bernhardy AJ, Cruz C, et al. The *BRCA1*-Δ11q alternative splice isoform bypasses germline mutations and promotes therapeutic resistance to PARP inhibition and cisplatin. *Cancer Res.* 2016;76(9):2778–2790.
62. Castella M, Pujol R, Callén E, et al. Chromosome fragility in patients with Fanconi anaemia: diagnostic implications and clinical impact. *J. Med. Genet.* 2011;48(4):242–50.
63. Raponi M, Douglas AGL, Tammaro C, Wilson DI, Baralle D. Evolutionary constraint helps unmask a splicing regulatory region in *BRCA1* exon 11. *PLoS One*. 2012;7(5):e37255.
64. DelloRusso C, Welcsh PL, Wang W, et al. Functional Characterization of a Novel BRCA1-Null Ovarian Cancer Cell Line in Response to Ionizing Radiation. *Mol. Cancer Res.* 2007;5(1):35–45.
65. Fargo JH, Rochowski A, Giri N, et al. Comparison of chromosome breakage in non-mosaic and mosaic patients with Fanconi anemia, relatives, and patients with other inherited bone marrow failure syndromes. *Cytogenet. Genome Res.* 2014;144(1):15–27.
66. Stary J, Zimmermann M, Campbell M, et al. Intensive Chemotherapy for Childhood Acute Lymphoblastic Leukemia: Results of the Randomized Intercontinental Trial ALL IC-BFM 2002. *J. Clin. Oncol.* 2014;32(3):174–184.
67. ClinVar.
68. Lu M, Conzen SD, Cole CN, Arrick BA. Characterization of functional messenger RNA splice variants of *BRCA1* expressed in nonmalignant and tumor-derived breast cells. *Cancer Res.* 1996;56(20):4578–81.
69. Freire BL, Homma TK, Funari MFA, et al. Homozygous loss of function BRCA1 variant causing a Fanconi-anemia-like phenotype, a clinical report and review of previous patients. *Eur. J. Med. Genet.* 2018;61(3):130–133.

70. Hirsch B, Shimamura A, Moreau L, et al. Association of biallelic *BRCA2/FANCD1* mutations with spontaneous chromosomal instability and solid tumors of childhood. *Blood*. 2004;103(7):2554–2560.
71. Alter BP, Rosenberg PS, Brody LC. Clinical and molecular features associated with biallelic mutations in *FANCD1/BRCA2*. *J. Med. Genet.* 2007;44(1):1–9.
72. Vasanthakumar A, Arnovitz S, Marquez R, et al. *Brca1* deficiency causes bone marrow failure and spontaneous hematologic malignancies in mice. *Blood*. 2015;127(3):310–13.
73. Johnson B, Fletcher SJ, Morgan N V. Inherited thrombocytopenia: novel insights into megakaryocyte maturation, proplatelet formation and platelet lifespan. *Platelets*. 2016;27(6):519–525.
74. Izumi R, Niihori T, Suzuki N, et al. GNE myopathy associated with congenital thrombocytopenia: A report of two siblings. *Neuromuscul. Disord.* 2014;24(12):1068–72.
75. Höck M, Wegleiter K, Ralser E, et al. ALG8-CDG: novel patients and review of the literature. *Orphanet J. Rare Dis.* 2015;10(1):73.
76. Schollen E. Clinical and molecular features of three patients with congenital disorders of glycosylation type Ih (CDG-Ih) (ALG8 deficiency). *J. Med. Genet.* 2004;41(7):550–556.
77. Jones C, Denecke J, Strter R, et al. A novel type of macrothrombocytopenia associated with a defect in α2,3-sialylation. *Am. J. Pathol.* 2011;179(4):1969–1977.
78. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297–303.
79. Gulsuner S, Walsh T, Watts AC, et al. Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell*. 2013;154(3):518–29.
80. Seo A, Walsh T, Lee M, et al. FAM111B Mutation Is Associated With Inherited Exocrine Pancreatic Dysfunction. *Pancreas*. 2015;0(0):1–5.
81. Purcell S, Neale B, Todd-Brown K, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* 2007;81(3):559–575.
82. McCorvie TJ, Wasilenko J, Liu Y, Fridovich-Keil JL, Timson DJ. In vivo and in vitro function of human UDP-galactose 4′-epimerase variants. *Biochimie*. 2011;93(10):1747–1754.
83. McCorvie TJ, Liu Y, Frazer A, et al. Altered cofactor binding affects stability and activity of human UDP-galactose 4'-epimerase: implications for type III galactosemia. *Biochim.*

Biophys. Acta. 2012;1822(10):1516–26.

84. Wohlers TM, Christacos NC, Harreman MT, Fridovich-Keil JL. Identification and characterization of a mutation, in the human UDP-galactose-4-epimerase gene, associated with generalized epimerase-deficiency galactosemia. *Am. J. Hum. Genet.* 1999;64(2):462–470.
85. Tamary H, Yaniv I, Stein J, et al. A clinical and molecular study of a Bedouin family with dysmegakaryopoiesis, mild anemia, and neutropenia cured by bone marrow transplantation. *Eur. J. Haematol.* 2003;71(3):196–203.
86. Levin J, Peng JP, Baker GR, et al. Pathophysiology of thrombocytopenia and anemia in mice lacking transcription factor NF-E2. *Blood*. 1999;94(9):3037–3047.
87. Shivdasani RA. Molecular and transcriptional regulation of megakaryocyte differentiation. *Stem Cells*. 2001;19(5):397–407.
88. Nichols KE, Crispino JD, Poncz M, et al. Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA 1. *Nat. Genet.* 2000;24(3):266–270.
89. Saleque S, Cameron S, Orkin SH. The zinc-finger proto-oncogene Gfi-1b is essential for development of the erythroid and megakaryocytic lineages. *Genes Dev.* 2002;16(3):301–306.
90. Spyropoulos DD, Pharr PN, Lavenburg KR, et al. Hemorrhage, Impaired Hematopoiesis, and Lethality in Mouse Embryos Carrying a Targeted Disruption of the Fli1 Transcription Factor. *Mol. Cell. Biol.* 2000;20(15):5643–5652.
91. Thoden JB, Wohlers TM, Fridovich-Keil JL, Holden HM. Molecular Basis for Severe Epimerase Deficiency Galactosemia: X-RAY STRUCTURE OF THE HUMAN V94M-SUBSTITUTED UDP-GALACTOSE 4-EPIMERASE. *J. Biol. Chem.* 2001;276(23):20617–20623.
92. Rosoff PM. Myelodysplasia and deficiency of uridine diphosphate-galactose 4-epimerase. 1995;127(4):5–8.
93. Maceratesi P, Daude N, Dallapiccola B, et al. Human UDP-galactose 4' epimerase (GALE) gene and identification of five missense mutations in patients with epimerase-deficiency galactosemia. *Mol. Genet. Metab.* 1998;63(1):26–30.
94. Piller F, Hanlon MH, Hill RL. Co-purification and characterization of UDP-glucose 4-epimerase and UDP-N-acetylglucosamine 4-epimerase from porcine submaxillary glands. *J. Biol. Chem.* 1983;258(17):10774–8.
95. Kingsley DM, Kozarsky KF, Hobbie L, Krieger M. Reversible defects in O-linked

glycosylation and LDL receptor expression in a UDP-Gal/UDP-GalNAc 4-epimerase deficient mutant. *Cell*. 1986;44(5):749–759.

96. Park H-D, Park KU, Kim JQ, et al. The molecular basis of UDP-galactose-4-epimerase (GALE) deficiency galactosemia in Korean patients. *Genet. Med.* 2005;7(9):646–649.
97. Quimby BB, Alano a, Almashanu S, et al. Characterization of two mutations associated with epimerase-deficiency galactosemia, by use of a yeast expression system for human UDP-galactose-4-epimerase. *Am. J. Hum. Genet.* 1997;61(3):590–598.
98. Wasilenko J, Lucas ME, Thoden JB, Holden HM, Fridovich-Keil JL. Functional characterization of the K257R and G319E-hGALE alleles found in patients with ostensibly peripheral epimerase deficiency galactosemia. *Mol. Genet. Metab.* 2005;84(1):32–8.
99. Timson DJ. Functional analysis of disease-causing mutations in human UDP-galactose 4-epimerase. *FEBS J.* 2005;272(23):6170–7.
100. Alano a, Almashanu S, Chinsky JM, et al. Molecular characterization of a unique patient with epimerase-deficiency galactosaemia. *J. Inherit. Metab. Dis.* 1998;21(4):341–50.
101. Chhay JS, Vargas C a, McCorvie TJ, Fridovich-Keil JL, Timson DJ. Analysis of UDP-galactose 4'-epimerase mutations associated with the intermediate form of type III galactosaemia. *J. Inherit. Metab. Dis.* 2008;31(1):108–16.
102. Schulz JM, Watson AL, Sanders R, et al. Determinants of function and substrate specificity in human UDP-galactose 4'-epimerase. *J Biol Chem*. 2004;279(31):32796–32803.
103. Fridovich-keil JL. Galactosemia: The Good, the Bad, and the Unknown. *J. Cell. Physiol.* 2006;209:701–705.
104. Klein C. Congenital neutropenia. *Hematology Am. Soc. Hematol. Educ. Program*. 2009;344–350.
105. Boztug K, Appaswamy G, Ashikov A, et al. A syndrome with congenital neutropenia and mutations in G6PC3. *N. Engl. J. Med.* 2009;360(1):32–43.
106. Daenzer JMI, Sanders RD, Hang D, Fridovich-Keil JL. UDP-galactose 4'-epimerase activities toward UDP-Gal and UDP-GalNAc play different roles in the development of Drosophila melanogaster. *PLoS Genet.* 2012;8(5):e1002721.
107. Brokate-Llanos a. M, Monje JM, Murdoch PDS, Munoz MJ. Developmental Defects in a Caenorhabditis elegans Model for Type III Galactosemia. *Genetics*. 2014;198(4):1559–1569.
108. Dimitroff CJ, Lee JY, Rafii S, Fuhlbrigge RC, Sackstein R. CD44 is a major E-selectin ligand on human hematopoietic progenitor cells. *J. Cell Biol.* 2001;153(6):1277–86.

109. Sackstein R, Merzaban JS, Cain DW, et al. Ex vivo glycan engineering of CD44 programs human multipotent mesenchymal stromal cell trafficking to bone. *Nat. Med.* 2008;14(2):181–187.
110. Wang Y, Jobe SM, Ding X, et al. Platelet biogenesis and functions require correct protein O-glycosylation. *Proc. Natl. Acad. Sci.* 2012;109(40):16143–16148.
111. Ju T, Cummings RD. Protein glycosylation: chaperone mutation in Tn syndrome. *Nature*. 2005;437(7063):1252.
112. Wandall HH, Rumjantseva V, Louise A, et al. The origin and function of platelet glycosyltransferases The origin and function of platelet glycosyltransferases. 2014;120(3):626–635.
113. Hemmoranta H, Satomaa T, Blomqvist M, et al. N-glycan structures and associated gene expression reflect the characteristic N-glycosylation pattern of human hematopoietic stem and progenitor cells. *Exp. Hematol.* 2007;35(8):1279–1292.
114. Nasirikenari M, Veillon L, Collins CC, Azadi P, Lau JTY. Remodeling of marrow hematopoietic stem and progenitor cells by non-self ST6Gal-1 sialyltransferase. *J. Biol. Chem.* 2014;289(10):7178–7189.
115. Grozovsky R, Giannini S, Falet H, Hoffmeister KM. Novel mechanisms of platelet clearance and thrombopoietin regulation. *Curr. Opin. Hematol.* 2015;22(5):445–451.
116. Grozovsky R, Giannini S, Falet H, Hoffmeister KM. Regulating billions of blood platelets: Glycans and beyond. *Blood*. 2015;126(16):1877–1884.
117. Giannini S, Lee-Sundlov MM, Adelman M, et al. β1,4-Galactosyltransferase 1 Is a Key Regulator of Thrombopoiesis [abstract]. *Blood*. 2017;130:Supplemental 1018.
118. Mercier F, Shi J, Sykes DB, et al. A Genome-Wide, In Vivo, Dropout CRISPR Screen in Acute Myeloid Leukemia Identifies an Essential Role for Beta-Galactosylation in Leukemic Cell Homing [abstract]. *Blood*. 2017;130:Supplemental 2493.
119. Freeze HH. Genetic defects in the human glycome. 2006;7(JULY):537–551.
120. Kazlouskaya V, Feldman EJ, Jakus J, Heilman E, Glick S. A Case of Hereditary Fibrosing Poikiloderma with Tendon Contractures, Myopathy, and Pulmonary Fibrosis (POIKTMP) with the Emphasis on Cutaneous Histopathological Findings. *J. Eur. Acad. Dermatology Venereol.* 2018;
121. Kaushansky K. Molecular mechanisms of thrombopoietin signaling. *J. Thromb. Haemost.* 2009;7(SUPPL. 1):235–238.
122. Geddis AE, Linden HM, Kaushansky K. Thrombopoietin: A pan-hematopoietic cytokine. *Cytokine Growth Factor Rev.* 2002;13(1):61–73.

123. Decker M, Leslie J, Liu Q, Ding L. Hepatic thrombopoietin is required for bone marrow hematopoietic stem cell maintenance. *Science*. 2018;360(6384):106–110.
124. Pecci A, Ragab I, Bozzi V, et al. Thrombopoietin mutation in congenital amegakaryocytic thrombocytopenia treatable with romiplostim. *EMBO Mol. Med.* 2018;10(1):63–75.
125. Grenda DS, Murakami M, Ghatak J, et al. Mutations of the ELA2 gene found in patients with severe congenital neutropenia induce the unfolded protein response and cellular apoptosis. *Blood*. 2007;110(13):4179–4187.
126. Feese MD, Tamada T, Kato Y, et al. Structure of the receptor-binding domain of human thrombopoietin determined by complexation with a neutralizing antibody fragment. *Proc. Natl. Acad. Sci. U. S. A.* 2004;101(7):1816–1821.
127. Cmejla R, Cmejlova J, Handrkova H, et al. Identification of mutations in the ribosomal protein L5 (RPL5) and ribosomal protein L11 (RPL11) genes in Czech patients with diamond-blackfan anemia. *Hum. Mutat.* 2009;30(3):321–327.
128. Quarello P, Garelli E, Carando A, et al. Diamond-blackfan anemia: Genotype-phenotype correlations in Italian patients with RPL5 and RPL11 mutations. *Haematologica*. 2010;95(2):206–213.
129. Zhang Z, Jia H, Zhang Q, et al. Assessment of hematopoietic failure due to Rpl11 deficiency in a zebrafish model of Diamond-Blackfan anemia by deep sequencing. *BMC Genomics*. 2013;14(1):.
130. Wang CQ, Krishnan V, Tay LS, et al. Disruption of Runx1 and Runx3 leads to bone marrow failure and leukemia predisposition due to transcriptional and DNA repair defects. *Cell Rep.* 2014;8(3):767–782.
131. Balogh P, Goldfarb A. RUNX3 is required for megakaryocyte-erythroid specification in primary human progenitors. *Exp. Hematol.* 2017;53:S130.
132. Khan A, Fornes O, Stigliani A, et al. JASPAR 2018: Update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* 2018;46(D1):D260–D266.
133. Austin KM, Gupta ML, Coats SA, et al. Mitotic spindle destabilization and genomic instability in Shwachman-Diamond syndrome. *J. Clin. Invest.* 2008;118(4):1511–1518.

CURRICULUM VITAE

Aaron Seo

Department of Genome Sciences
University of Washington
Seattle, WA 98195
(206) 606-4294

5110 Ravenna Pl NE #101
Seattle, WA 98105
(281) 455-4937
aaronseo@uw.edu

EDUCATION

University of Washington, Seattle, WA, 2011-present
Medical Scientist Training Program (M.D./Ph.D.) trainee
Graduate Student in the lab of Dr. Mary-Claire King

University of Texas at Austin, Austin, TX, 2007-2011
B.S. in Biology, with highest honors
B.A. in Plan II, with highest honors

AWARDS and HONORS

- NIH NRSA F30 Fellowship, 2016-present
- Northwest Developmental Biology Meeting Poster Presentation Award, 2015
- NIH Developmental Biology Training Grant, 2014-2016
- ARCS Foundation Fellowship, 2013-2015
- Dean's Honored Graduate, 2011
- Beckman Scholar, 2010-2011
- Phi Beta Kappa, 2010

ACTIVITIES

- Leader, UW Sun Protection Outreach Teaching by Students, 2017-present.
- Student Interviewer, UW School of Medicine Admissions Committee, 2012-2013.
- Co-Chair, UW Asian Pacific American Medical Student Association, 2012-2013.
- Co-Leader, Christian Medical/Dental Student Association, 2012-2013.
- Member, Pediatrics Interest Group, 2011-present.

TEACHING EXPERIENCE

Teaching Assistant: Advanced Human Genetics, University of Washington, Winter 2015

- Led two hours of discussion section per week teaching supplemental and complementary material to lecture and reviewing problem sets and lecture material
- Held multiple office hours per week
- Developed problem sets, grading rubrics, and grading all course material
- Gave lecture in class on hematopoietic malignancies

PUBLICATIONS

Note: Previous name "Kyunghyun Seo"
*contributed equally to the work
**co-corresponding authors

Seo A, Ben-Harosh M, Walsh T, Lee MK, Gulsuner S, Tamary H, Shimamura A, King MC. Mutation of UDP-Galactose-4-Epimerase (*GALE*) Associated with Inherited Thrombocytopenia. *Manuscript in preparation.*

Seo A, Steinberg O, Unal S, Casadei S, Walsh T, Gumruk F, Shimamura A, Akarsu NA, Tamary H, King MC. A Mechanism for Survival of Homozygous BRCA1 Nonsense Mutations. Proc. Natl. Acad. Sci. USA. Published ahead of print April 30, 2018. https://doi.org/10.1073/pnas.1801796115

Seo A*, Ben-Harosh M*, Sirin M*, Stein J, Dgany O, Kaplelushnik J, Hoenig M, Pannicke U, Lorenz M, Schwarz K, Stoccklausner C, Walsh T, Gulsuner S, Lee MK, Sendamarai A, Sanchez-Bonilla M, King MC, Cario H, Kulozik AE, Debatin KM, Schulz A**, Tamary H**, Shimamura A**. Bone Marrow Failure Unresponsive to Bone Marrow Transplant is Caused by Mutations in *thrombopoietin*. Blood. 2017 Aug 17; 130(7):875-880. PMCID: PMC5561901

Seo A, Walsh T, Lee M, Ho P, Hsu E, Sidbury R, King MC, and Akiko Shimamura. FAM111B Mutation is Associated with Inherited Exocrine Pancreatic Dysfunction. Pancreas. 2015 Oct 22. PMCID: PMC4841754

Paulson K, Tegeder A, Willmes C, Iyer JG, Afanasiev OK, Schrama D, Koba S, Thibodeau R, Nagase K, Simonson WT, **Seo A**, Koelle DM, Madeleine M, Bhatia S, Nakajima H, Sano S, Hardwick JS, Disis ML, Cleary MA, Becker JC, and Paul Nghiem. Downregulation of MHC-1 expression is prevalent but reversible in Merkel cell carcinoma. Cancer Immunology Research. 2014 Nov; 2(11):1071-9. PMCID: PMC4221542

Hall B, Arshad S, **Seo K**, Bowman C, Corley M, Jhaveri SD, Ellington AD. In vitro selection of RNA aptamers to a protein target by filter immobilization. Curr Protoc Nucleic Acid Chemistry. 2010 Mar; Chapter 9:Unit 9.3.1-27.

Hall B, Arshad S, **Seo K**, Bowman C, Corley M, Jhaveri SD, Ellington AD. In vitro selection of RNA aptamers to a protein target by filter immobilization. Curr Protoc Mol Biol. 2009 Mar; Chapter 24:Unit 24.3

ABSTRACTS and PRESENTATIONS

Seo A, Walsh T, Rosenfeld N, Lee M, Dgany O, Levy-Lahad E, Shimamura A, King MC, Tamary H. Homozygous *BRCA1* Truncation Causes Fanconi Anemia; (Abstract/Talk). Presented at the 66th Annual Meeting of The American Society of Human Genetics, October 22, 2016 in Vancouver, BC, Canada

Seo A, Walsh T, Gulsuner S, Lee M, Tamary H, Shimamura A, King MC, Evaluation of a *TFDP2* Mutation as a Cause of an Inherited Bone Marrow Failure and Craniosynostosis Syndrome; (Abstract/Talk). Presented at the 2016 Northwest Developmental Biology Meeting, March 18, 2016 in Friday Harbor, WA

Seo A, Walsh T, Lee M, Ho P, Hsu E, Sidbury R, King MC, Shimamura A, *FAM111B* Mutation Presenting with Features of Shwachman-Diamond Syndrome; (Abstract/Poster). Presented at the 2015 Northwest Developmental Biology Meeting, March 20, 2015 in Friday Harbor, WA

Seo A, Walsh T, Lee MK, King MC, Shimamura A, Expansion of the fibrosing poikiloderma phenotype caused by FAM111B to include cytopenia and pancreatic dysfunction; (Abstract/Program#2948M). Presented at the 64th Annual Meeting of The American Society of Human Genetics, October 20, 2014 in San Diego, CA