Investigating the Genetic Modifiers of Thrombosis Using the Zebrafish Model

by

Steven Grzegorski

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Cell and Molecular Biology)
In The University of Michigan
2020

Doctoral Committee:

Associate Professor Jordan Shavit, Chair
Professor David Ginsburg
Associate Professor Mike Holinstat
Professor Steven Pipe
Associate Professor Cristen Willer

Steven Grzegorski

sgrzegor@umich.edu

ORCID iD: 0000-0003-2221-3325

Dedication

To my family, for being a living reminder that I could reach my dreams without having my feet leave the ground.

Acknowledgements

I'd like to thank my dissertation committee (Jordan Shavit, David Ginsburg, Mike Holinstat, Steve Pipe, and Cristen Willer) for your time, support, and input during the many meetings and discussion we have had during my training. A special thanks to my mentor Jordan Shavit for his countless hours of guidance and mentorship during the progression of my research and towards my personal and scientific development.

Thank you to the University of Michigan Medical Scientist Training Program both past and present (Ron Koenig, Ellen Elkin, Hilkka Ketola, Laurie Koivupalo, Justine Hein, Gretchen Aland and Liz Bowman) for your support and for the work you do to develop and foster a phenomenal MSTP training environment.

Thank you to the Cellular and Molecular Biology program, especially Bob Fuller, Pat Ocelnik and Lauren Perl for maintaining the infrastructure that gave me opportunities to stay involved and the resources to present my work both locally and externally over the course of my training.

Thanks to the National Institute of Health, the National Heart, Lung and Blood Institute and the American Heart Association for providing funding during my graduate work. A special thanks goes to the Cardiovascular Translational Research and Entrepreneurship training program, especially Dan Michele and José Jalife for providing a wonderful opportunity to expand my training during the beginning of my candidate years.

Finally, I'm grateful beyond words for my friends, lab mates and especially my family for supporting me emotionally and intellectually through the last few years and beyond. Mom, Dad, Amy, Cindy and Zee. I could never thank you enough!

Table of Contents

Dedication ..... **ii**
Acknowledgements ..... **iii**
List of Tables ..... **vi**
List of Figures ..... **vii**
List of Abbreviations ..... **ix**
Abstract ..... **xi**
Chapter I: Introduction **1**
Hemostasis 1
Genetics of Thrombosis 3
The Coagulation Cascade in Zebrafish 6
Forward Genetic Screens 9
Reverse Genetics 10
Remaining Questions for the Genetics of Thrombosis 11
Chapter II: Disruption of the Kringle 1 Domain of Prothrombin Leads to Late Onset Mortality in Zebrafish **14**
Introduction 14
Materials and Methods 16
Results 22
Discussion 26
Chapter III: A Bioinformatic Pipeline for Identifying Dominant Mutations in the Zebrafish Model Using High Throughput Sequencing **38**
Introduction 38
Methods 41
Results 46
Discussion 48

Chapter IV: Dissecting Genetic Modifiers of a Sensitized Zebrafish Model of Thrombosis 55
Introduction 55
Methods 57
Results 59
Discussion 63
Chapter V: Genetic Duplication of Tissue Factor in Teleost Reveals Partial Subfunctionalization 73
Introduction 73
Materials and Methods 75
Results 78
Discussion 84
Chapter VI: Conclusions and Future Directions 98
Conclusions 98
Future Directions 101
References 109

List of Tables

Table 2-1. Oligonucleotide sequences used in methods 37
Table 4-1. Oligonucleotide sequences used in methods 71
Table 4-2. Candidate suppressor mutations for survival benefit 72
Table 5-1. Oligonucleotides used in methods 96
Table 5-2. Synthesized recombinant TF gBlocks 97

List of Figures

Figure 1-1. Diagram of hemostasis demonstrates slight differences between human and zebrafish coagulation cascades. 13
Figure 2-1. Peptide sequence alignment shows strong conservation of prothrombin across a broad range of species. Sequences shown include the conserved domains. 30
Figure 2-2. Direct oral anticoagulants prevent occlusive thrombus formation in zebrafish. 31
Figure 2-3. Genome editing creates a 14 bp genomic deletion with a resulting decrease in mRNA expression. 32
Figure 2-4. Single molecule real time sequencing of *f2*$^{+/\Delta 14}$ mRNA demonstrates altered splicing in *f2* following a deletion in exon 6. 33
Figure 2-5. Loss of prothrombin results in early lethality due to hemorrhage. 34
Figure 2-6. Loss of thrombin activity results in defects in secondary hemostasis. 35
Figure 2-7. Expression and activation of prothrombin variants and resulting activity of thrombin upon activation. 36
Figure 3-1 Workflow diagram for the processing of zebrafish samples for high-throughput sequencing. 51
Figure 3-2. ENU mutagenesis creates a robust increase in SNVs in the offspring of chemically-treated males. 52
Figure 3-3. Variant sites identified from high-throughput sequencing provide evidence of positively selected loci on chromosomes 2,7 and 16. 53
54
Figure 3-4. SNV distribution across chromosomes confirms a known modifier gene and identifies candidate loci for future analysis. 54

Figure 4-1. *f2, f10 and plg* heterozygosity protect against early lethality due to *at3* deficiency. 66
Figure 4-2. Overview of ENU mutagenesis. 67
Figure 4-3. Short-read sequencing identifies a missense mutation in the codon for a highly conserved cysteine of *f2.* 68
Figure 4-4. The *frost* allele is a loss-of-function variant in *f2* that partially rescues *at3* deficiency. 69
Figure 4-5. C507F mutation leads to reduced biosynthesis, activation, and activity of human prothrombin 70
Figure 5-1. Genomic engineering creates large loss-of-function deletions in both copies of zebrafish *TF*....89
Figure 5-2. Loss of TF results in early lethality due to hemorrhage. 90
Figure 5-3. Vascular development is unaffected in the absence of TF. 91
Figure 5-4. Laser-induced endothelial injury demonstrates a partial subfunctionalization of TFa in the venous system and TFb in the arterial system....92
Figure 5-5. TFa has increased procoagulant activity compared to TFb in an *ex vivo* clotting assay. 93
Figure 5-6. TFb acts through FIXb to trigger a consumptive coagulopathy. 94
Figure 5-7. Loss of TF contributes to a risk of cardiac tamponade through hemostatic and nonhemostatic effects. 95

List of Abbreviations

APC	activated protein C
at3	Antithrombin (gene)
AT3	Antithrombin (protein)
dpf	Days post fertilization
DVT	Deep-vein thrombosis
eGFP	Enhanced Green Fluorescent Protein
f10	Coagulation factor X (gene)
f2	Prothrombin (gene)
f5	Coagulation factor V (gene)
f7	Coagulation factor VII (gene)
F7	Coagulation factor VII - zebrafish (protein)
F7i	Coagulation factor VII-inhibitor - zebrafish (protein)
F7L	Coagulation factor VII-like - zebrafish (protein)
f9	Coagulation factor IX (gene)
FIXa	Coagulation factor IX A - zebrafish (protein)
FIXb	Coagulation factor IX B - zebrafish (protein)
fga	Fibrinogen alpha (gene)
FIX	Coagulation factor IX (protein)
FV	Coagulation factor V (protein)
FVII	Coagulation factor VII (protein)
FVIII	Coagulation factor VIII (protein)
FX	Coagulation factor X (protein)
FXI	Coagulation factor XI (protein)
FXII	Coagulation factor XII (protein)
Gla	Carboxylation/gamma-carboxyglutamic acid domain

HPC4	Human protein C epitope tag
hpf	Hours post fertilization
PAR	Protease-activated receptor
PCR	polymerase chain reaction
PE	Pulmonary embolism
plg	Plasminogen (gene)
TF	Tissue factor (protein)
TFa	Tissue factor A (protein)
TFa	Tissue factor A (gene)
TFb	Tissue factor B (protein)
TFb	Tissue factor B (gene)
TTO	time to occlusion
VTE	Venous thromboembolism
VWF	von Willebrand Factor (protein)
**	Zebrafish genes are lowercase
**	Human genes are uppercase

Abstract

Thrombosis is a leading cause of morbidity and mortality. Although 50-60% of thrombotic risk is estimated to be due to genetics, only 35% of individuals with a thrombotic event carry one of the 20-30 known genetic modifying risk factors. Antithrombin (AT3) is the primary endogenous inhibitor of the coagulation cascade, but deficiency of AT3 leads to an incompletely penetrant risk of thrombosis due to presumed genetic modifiers. The zebrafish is a small aquatic vertebrate with a well-characterized and highly conserved hemostatic system, but previous work from our laboratory suggests species specific factors allow zebrafish to survive severe hemostatic defects. This dissertation focuses on multiple areas: understanding the influence of known factors on thrombosis, developing a framework to identify novel factors and characterizing the conservation of hemostasis in zebrafish.

Zebrafish are able to survive into adulthood with a severe genetic deficiency of *factor V (f5)* or *factor X (f10)*, two factors upstream of prothrombin (*f2*), raising the possibility of alternative pathways for thrombin generation. To rule out this possibility, a *f2* knockout phenocopied the life expectancy of the *f5* and *f10* knockouts, confirming canonical function. The knockout led to an unexpected deletion in the kringle 1 domain and revealed a potential maturation and functional roles for (pro)thrombin. This model allows the study of the broad *in vivo* roles of thrombin such as cell signaling and response to inflammation that are impossible to explore in mice due to lethality.

Loss of At3 leads to larval consumptive coagulopathy followed by lethal adult thrombosis. Intercrossing knockout alleles of *f2*, *f10* and *plasminogen* (*plg*) onto the At3-deficient zebrafish genetic background demonstrated that heterozygosity for each individual mutation improved survival in the thrombotic population. After demonstrating the feasibility of genetically suppressing thrombosis, an unbiased chemical mutagenesis screen was performed to identify

novel suppressors. Four lines were established that demonstrated stable improvement in survival. A candidate approach identified *frost,* a point mutation in prothrombin. Heterozygosity for *frost* in *at3*$^{-/-}$ fish rescued both survival and the larval consumptive coagulopathy. Biochemical studies of the *frost* mutation demonstrated decreased prothrombin production and an inability to cleave fibrinogen. In parallel, the remaining lines were sequenced and a modern genomics pipeline confirmed the mutagenesis protocol was a successful unbiased survey, but did not achieve functional saturation of the genome. This indicates the strong potential of this protocol to identify new thrombosis biology.

Finally, loss of tissue factor (TF), a key initiating factor of the coagulation cascade, is incompatible with murine embryonic survival and studies in humans are lacking. Zebrafish have 2 copies of TF (TFa and TFb) with unknown conservation of function, and loss of function alleles for both copies were generated. Loss of both resulted in early lethality, but single copies were sufficient for normal survival. TFa and TFb were shown to have predominant roles in the arterial and venous systems, respectively. Finally, although TFb has lower *in vitro* procoagulant activity, it may play a role in factor VII mediated activation of the intrinsic pathway.

Overall, the work in this dissertation provides a strong foundation for studying and discovering modifiers of hemostasis and thrombosis using the zebrafish. Throughout this study, the utility of zebrafish was confirmed through the discovery of unexpected biology, demonstration of hemostatic rebalancing through thrombotic suppression and identification of promising novel alleles and pathways for informing human biology.

Chapter I:
Introduction

HEMOSTASIS

In all vertebrates, the distribution of oxygen and nutrients to tissues is facilitated by the circulation of blood through a close-looped cardiovascular system. Under normal physiology, the proteinaceous plasma and cellular component remain free-flowing through a network of endothelial cell coated blood vessels.[1] In response to vessel damage, a series of reactions occur at the site of injury to prevent excessive loss of blood.[2,3] Following injury, the process that stops blood flow and prevents further bleeding is known as hemostasis and is divided into a primary and a secondary phase.[4]

Primary hemostasis

Initial exposure of collagen in the subendothelial matrix to the circulating blood reveals sites for the attachment of the large glycoprotein von Willebrand factor (VWF).[5] Sheer force then causes unravelling of the VWF multimer and exposes sites for the binding of circulating platelets.[6] Attachment of platelets induces their activation, facilitating a shape change and the release of platelet agonists.[7] These agonists recruit and activate additional platelets which begin to adhere through GPIIb/IIIa mediated fibrinogen attachment.[8] The aggregation of platelets results in the formation of a plug that stabilizes the site of injury as illustrated on the endothelial surface in Figure 1-1.[9] The role of platelets is especially critical in high shear conditions, such as those found in the arterial system.[10] Under high shear, aggregation of lipid rich surfaces provides a

foundation for the assembly of the factors required to greatly enhance the efficiency of secondary hemostasis.[11,12]

Secondary hemostasis

In parallel to platelet plug formation, a secondary series of proteolytic reactions known as the coagulation cascade begins. This cascade is initiated through the binding of the transmembrane glycoprotein tissue factor (TF) on the surface of exposed subendothelial cells to the circulating serine protease factor VIIa (FVIIa) or its zymogen form factor VII (FVII). This interaction is stabilized through the interaction of the vitamin K-dependent carboxylation/gamma-carboxyglutamic (Gla) domain on FVIIa with the phospholipid surface containing TF.[13] Once bound to TF, FVIIa is able to cleave and activate the Gla-containing zymogen factor X (FX) to the serine protease factor Xa (FXa). This pathway is known as the extrinsic pathway of coagulation.[14] In the presence of the cofactor factor Va (FVa) and a phospholipid surface, FXa is able to convert prothrombin to the active serine protease thrombin, shown in blue in Figure 1-1.[15] This results in the end reaction of secondary hemostasis which is the thrombin-mediated cleavage of the soluble protein fibrinogen to the insoluble protein fibrin. The insoluble fibrin then self assembles into strands that are able to stabilize the platelet plug and results in a durable blood clot at the site of injury.[16] The FXa-FVa-thrombin-fibrin axis is referred to as the common pathway of coagulation (Figure 1-1 top).

In addition to the cleavage of fibrinogen, thrombin is able to feedback activate the serine protease factor XIa (FXIa) along with the cofactors factor Va and factor VIIIa (FVIIIa). FXIa converts factor IX (FIX) to factor IXa (FIXa) which subsequently generates FXa with the assistance of the cofactor FVIIIa. This portion of coagulation cascade is referred to as the intrinsic pathway (Figure 1-1 top right).[17] This pathway serves to amplify the generation of thrombin through the increased production of FXa.[18] The intrinsic pathway can also be activated by the protease factor XIIa *ex vivo*, but this appears to be inconsequential for normal hemostasis *in vivo* as deficient patients do not bleed.[19]

GENETICS OF THROMBOSIS

In the absence of vascular injury, the procoagulant process of hemostasis is kept in check by a series of anticoagulant factors.[20] These factors allow the undisrupted flow of blood through the cardiovascular system. Disruption of this balance can lead to a risk of bleeding, when the anticoagulant factors dominate, or a risk of pathological clotting known as thrombosis, when the procoagulant factors dominate. Thrombosis is defined as the formation of a blood clot within the vessel and can disrupt delivery of an adequate blood supply to sites past the point of thrombus formation.[21,22]

In the arterial system, thrombosis usually occurs due to abnormal platelet activation, e.g. through disruption of an atherosclerotic plaque in the wall of the vessel and can result in an ischemic injury such as a stroke or myocardial infarction.[23] Thrombi that form in the arterial system are classically referred to as "white thrombi" in reference to the histological appearance of platelet rich clots.[24] As of 2013, the Global Burden of Disease project estimated that there are 8.5 million myocardial infarctions and close to 7 million ischemic strokes globally each year.[25]

In contrast, venous thromboembolisms (VTEs), or thrombi in the venous system, present as either deep-vein thrombosis (DVT) or, when mobilized and trapped in the lung vasculature, as pulmonary embolisms (PE).[26] These are referred to as "red thrombi" due to their histology and high levels of fibrin and red blood cells relative to platelets.[24] In the United States it is estimated that 1-2 in 1000 individuals have a VTE each year, resulting in roughly 900,000 cases/year and 60,000-100,000 deaths.[27,28]

Individuals with a genetic predisposition for forming thrombi are said to have inherited thrombophilia. The most common known causes typically result in an increased risk of VTE without an apparent influence on arterial thrombosis although common environmental risk factors may contribute to both.[29,30]

Factor V Leiden

The most common of prothrombotic genetic risk factors is a missense mutation leading to an arginine to glutamine substitution at position 506 (R506Q) in FV referred to as Factor V Leiden. During normal physiology, activated protein C (APC) inactivates FV by cleaving the mature protein at multiple residues including R506.[31] The Leiden allele results in resistance of FV to APC inactivation and subsequent prolonged *in vivo* activity. In heterozygotes the risk of a first VTE is estimated to be 3 to 8-fold higher than an unaffected patient. In homozygotes this risk increases to 10 to 80-fold.[32] Furthermore, the Leiden allele is present in 3-8% of the US and European Caucasian populations. Combined, the prevalence and odds ratio mean that roughly 25% of first time idiopathic VTEs occur in carriers of the Leiden allele.[33] This increases to 40-50% allelic frequency in patients with recurrent VTE.[34]

Prothrombin G20210A

The second most common inherited thrombophilia is the result of a guanine to adenine transition in the 3' untranslated region at position 20210 of the prothrombin gene. Potentially due to increased pre-messenger RNA (mRNA) stability, this mutation leads to an increase in circulating prothrombin levels.[35] The mutation primarily occurs in Caucasians at an estimated frequency of 2%.[36] Heterozygosity increases the risk of unprovoked VTE by 2.5-fold while homozygosity leads to a 20-fold increase.[37]

Protein C Deficiency

Protein C is a vitamin K-dependent protease that acts to inactivate FVa and FVIIIa to endogenously inhibit coagulation. Inherited defects in protein C are estimated to occur at a rate as high as 1 in 500 in healthy populations.[38] In individuals with an unprovoked VTE, the prevalence of protein C deficiency is 2-5% and 10-15% in those with recurrent VTE.[34,39] Homozygosity is very rare and leads to a typically lethal condition called neonatal purpura fulminans that presents with widespread small vessel thrombosis and necrosis of the surrounding

cutaneous and subcutaneous tissues.[40] Overall, the relative risk of VTE is 8 to 10-fold higher when heterozygous for protein C deficiency.[41] Over 160 mutations have been identified to date and can lead to either a quantitative defect in protein C levels (Type I) or a qualitative defect in protein C activity (Type II).[42]

Protein S Deficiency

Protein S is the vitamin K-dependent nonenzymatic cofactor of protein C that aids in the inactivation of FVa and FVIIIa. Protein S deficiency is estimated to occur in 1 out of 700 patients and approximately 2% of those presenting with VTE.[43,44] This rate increases to 3-6% in populations with recurrent VTE. Similar to protein C, homozygous mutation leads to severe thrombophilia in early life with occasional purpura fulminans and high mortality.[45] Protein S deficiency leads to a relative risk of thrombosis of 2.4.[46]

Antithrombin Deficiency

Antithrombin III (AT3) is a serine protease inhibitor with a particular affinity to thrombin and FXa. AT3 acts as a suicide inhibitor, meaning that it forms a 1:1 covalent complex with active proteases to block their activate site and irreversibly inactivate them.[47] Deficiency of AT3 is the most severe of inherited thrombophilias with a relative risk of thrombosis of 5-50 compared to non-carrier controls.[48] The frequency of deficiency is estimated at 1 in 2000 but present at 0.5-5% of patients who develop thrombosis.[39,44] Patients with inherited AT3 deficiency have a poorly defined lifetime risk of developing a VTE with estimates of cumulative risk spanning from 20-80%.[49,50] Homozygous loss of AT3 activity is thought to be incompatible with life.[51]

Blood Type

The agglutinogens A and B on the surface of human erythrocytes are the result of the histo-blood group ABO system transferase catalyzing the modification of saccharides on the H antigen.[52] The presence of A antigen, B antigen, neither or both is the foundation of the ABO blood group system.[53] It has been found that

the ABO system is a potent genetic modifier. O blood type, lacking both antigens, is associated with lower levels of circulating VWF and FVIII.[54,55] In a study of hypertensive adults and post-menopausal women, it was found that blood types A, B and AB have an increased risk of thrombosis over type O with odds ratios of 1.79, 1.82 and 2.70 respectively.[56] The risk of non-O blood types has been confirmed in several large studies.[57]

Current Therapeutic Options for Treatment of Thrombosis

There are 3 main classes of anticoagulant therapy. Warfarin, a vitamin K epoxide reductase inhibitor, blocks maturation of the many vitamin K-dependent coagulation factors (prothrombin, FVII, FIX, FX, protein C and protein S).[58] Heparin (and its derivatives) is a naturally occurring glycosaminoglycan that substantially increases the affinity of AT3 for thrombin and FXa.[59-61] Finally, the direct oral anticoagulants (DOACs) are relatively new small molecules that act to directly block the active sites of either thrombin or FXa.[62,63] All three classes lead to an increased risk of bleeding at therapeutic levels.[64-66] This high risk of adverse events with the low penetrance of many of the common inherited thrombophilias means that thrombophilic patients are generally not anticoagulated until they present with a VTE.[67] This ultimately means that the clinical management of many patients with inherited thrombophilia is similar to patients with idiopathic VTE, and thus there is no benefit to genetic testing. Following initiation of treatment, patients must be monitored for adverse effects, and in the case of warfarin, routinely tested to maintain therapeutic levels.[68]

THE COAGULATION CASCADE IN ZEBRAFISH

The last 20 years have seen an explosion in the characterization of the coagulation cascade in fish broadly and zebrafish specifically. Genetic conservation has been shown for many of the major components including fibrinogen, prothrombin, FV, FX, FIX, FVIII, FVII, tissue factor (TF), AT3, protein C, and protein S.[69-71] Furthermore, functional conservation of fibrinogen,

prothrombin, FV, FX and AT3 has been demonstrated through genetic knockout or transcriptional knockdown studies.[72-77] For this work, the focus is on the coagulation cascade, but it is worth nothing that the cellular component is thought to be conserved as well.[78,79] Notably, zebrafish lack megakaryocytes but their blood contains nucleated thrombocytes as the presumed functional equivalent of platelets.[80-82] When compared to secondary hemostasis in humans, the differences in the zebrafish coagulation system can be classified as either absent factors or duplicated factors (Figure 1-1 bottom).

Absent Factors

A few mammalian coagulation factors are absent in zebrafish. FXII is an endogenous activator of the contact pathway of coagulation in humans. In human plasma, FXII is activated by polyanionic surfaces and subsequently activates FXI. Zebrafish do not appear to have a homologue of FXII. Interestingly, it appears that FXII is either absent or has been lost in all aquatic species, indicating a potential role for a terrestrial specific function for FXII.[83,84]

The immediate substrates of FXII, FXI and prekallikrein (PK), also appear to be absent in zebrafish and the broader teleost family. It is thought that all 3 genes arose via gene duplication in the tetrapod lineage with FXI and PK originating from a common precursor.[71,83]

Duplicated Factors

When comparing the human and zebrafish genome, >70% of human genes have a zebrafish orthologue but approximately 1/3 of the zebrafish genome is duplicated.[85] Duplicated genes in zebrafish appear in two forms: tandem or as a result of whole genome duplications. The former is the result of small regions of the genome being inadvertently duplicated, likely due to errors in homologous recombination.[86] This leads to 1 or more adjacent copies of a gene. There is evidence of this affecting 2 zebrafish coagulation proteins: FVII and protein C.[87]

FVII appears in a cluster of 3 tandem copies referred to as FVII (in this work F7 is used to reference the zebrafish protein aid in differentiating the various

copies), factor VII-like (F7L) and factor VII inhibitor (F7i). All three copies appear to be expressed during development and in adult tissue.[88,89] They all maintain the conserved F7 structure of a putative Gla domain, 2 EGF domains and a predicted serine protease domain.[90] F7i appears to have acquired multiple deleterious mutations including loss of the reactive serine in the catalytic triad of the protease domain and loss of the canonical site of cleavage required for activation. Functional studies confirm that F7i acts as a competitive inhibitor of tissue factor activated coagulation *in vitro*, and possibly *in vivo*.[91] To date, there are no studies comparing the binding preferences or activity of F7 and F7L.

Protein C also underwent a tandem duplication resulting in 2 copies. The amino acid sequence is highly conserved between these 2 copies, suggesting a recent event. Sequence analysis and unpublished data (Shavit laboratory) suggest that one copy is non-functional.

When the ancestor of the teleost lineage evolutionarily diverged 320-350 million years ago, it came with a whole genome duplication event.[92] During a whole genome duplication event, there is an increase in ploidy due to mitotic or meiotic errors leading to extra chromosomes. Over time, mutations cause a divergence of sequence leading to the loss of many of the initially duplicated genes.[93] However, many genes maintain multiple copies due to subfunctionalization, neofunctionalization or dosage compensation. Due to this historical event, roughly 20-25% of the genes conserved from humans to zebrafish exist in duplicate, typically referred to as a and b copies.[85,94] In contrast to tandem duplicates, whole genome duplicates exist on separate chromosomes and are often identified by both sequence conservation but also surrounding chromosomal gene synteny.[95] Two coagulation genes exemplifying this type of duplication exist in zebrafish: FIX (FIXa and FIXb) and TF (TFa and TFb). Currently, there are no published data regarding FIXa, FIXb or TFa and only a single report on TFb.[96] It does appear at a sequence level that all 4 genes are functional and expressed during both development and in adult tissues.[88,89]

FORWARD GENETIC SCREENS

Since the dawn of nucleic acid-based life, natural variation in sequence has provided the substrate for the process of natural selection. Before science understood how to create targeted mutations, methods to increase the random mutation rate in a population were discovered.[97,98] Forward genetics is the principle of using the observed phenotypes from natural or induced genetic variation to discover the genes influencing those phenotypes.[99] One of the most common ways to create random point mutations is through the use of a chemical mutagen such as N-ethyl-N-nitrosourea (ENU). ENU is an alkylating agent that preferentially modifies nitrogenous bases in nucleosides (primarily thymine) in spermatogonial stem cells.[100,101] Addition of ENU is estimated to introduce a deleterious mutation in 1 out of every 700 genetic loci or roughly 30 genes per sperm in mice, but the rate is highly dependent on the treatment conditions.[102,103]

In addition to a potent mutagen, a forward genetic screen requires readily observable phenotypes. Zebrafish grew rapidly as a model system in the early 1990s due to their unique strength as a vertebrate model suitable for large forward genetic screens. A single mating pair of zebrafish can produce hundreds of offspring each week. These offspring develop transparently and externally making feasible the rapid observation of their entire developmental time course. The initial genetic screens in zebrafish generated thousands of offspring from the inbreeding of the second generation of mutagenized founders.[104,105] This allowed for discovery of changes to development milestones that were transmitted in a recessive pattern.[106] Many of these initial mutants were described in a special zebrafish issue of Development in December 1996.[107] Over the next decade, hundreds of these mutations were isolated, and the causative genes discovered through the use of linkage of phenotypes with known genetic markers in large horizontal pedigrees.[108]

Previously, the Shavit laboratory completed a chemical mutagenesis screen in zebrafish with several modifications to these initial screens. First, the screen was performed as a sensitized suppressor screen. The screen was sensitized by

being performed on zebrafish lacking the *at3* gene which results in early lethality due to increased thrombosis.[75] The phenotype of interest was the *suppression* of thrombosis induced mortality. Secondly, the screen was completed in a dominant manner, meaning any modifiers improving survival would be present in the heterozygous state. This both increases the throughput of the screen but also enhances the relevance for common naturally occurring variants in human populations. Finally, the screen was done on adult survival which has the caveat of increasing false negatives due to non-specific lethality, and false positives due to stochastic survivors.

REVERSE GENETICS

Reverse genetics is the process of understanding the function of a gene through manipulation and observation for phenotypic alterations.[109] Historically, this was done in zebrafish through the use of antisense oligonucleotide technologies such as morpholinos that transiently suppressed translation or through the transient overexpression of wild-type and mutant proteins.[110-113] With the advent of zinc finger nucleases, it became possible for the first time to create transmittable targeted genetic mutations.[114,115] This ability improved with the introduction of transcription activator-like effector nucleases.[116,117] Finally, in the last decade, the ease and access of genetic engineering has exploded with the development of CRISPR (clustered regularly interspaced short palindromic repeats)/Cas9 technology.[118-120]

CRISPR/Cas9 is a two-part system discovered as a form of adaptive immunity in prokaryotes that has been modified to introduce targeted mutations in eukaryotes.[121] For use in zebrafish, the system works as briefly follows. The DNA nuclease Cas9 with nuclear localization signals is either expressed via plasmid DNA or directly injected as mRNA or protein into the single cell zebrafish zygote. In addition, a single guide RNA (sgRNA) is co-delivered with the Cas9. The sgRNA consists of a domain that complexes with the mature Cas9 protein (the tracrRNA portion) and a second gene specific domain (the crRNA) that allows targeting of

the Cas9 to the site of interest. The only requirement for targeting is the presence of a protospacer adjacent motif (PAM) sequence at the site of interest. For Cas9, the PAM sequence fits the pattern of NGG.[122] Once the assembled complex is delivered to the nucleus, it is able to make double stranded breaks (DSB) at the site of interest. Multiple sgRNAs can be delivered and multiple DSBs induced simultaneously. Generally, this process of delivery at the single cell stage results in chimeric founder fish carrying dozens of different alleles at the site due to errors during repair.[123] Offspring of these founder fish are screened for ideal alleles that disrupt the coding sequence of the targeted gene. Previous studies have demonstrated that co-delivery of repair templates facilitates the targeted introduction of point mutations or reporter genes.[124-126]

REMAINING QUESTIONS FOR THE GENETICS OF THROMBOSIS

The explosion of high throughput sequencing data availability and integration with large cohorts of patients has greatly expanded our ability to perform observational studies on complex diseases.[127,128] Despite this and the fact that 50-60% of VTE risk is attributed to genetics, only 35% of VTE patients contain a polymorphism in one of the 20-30 genes known to contribute to thrombophilia.[129,130] Additionally, the complexity of concurrent chemical reactions in the blood means that the disruption of known factors can have unexpected compensatory consequences. For example, mathematical modeling suggests that the bleeding risk of hemophilia A (FVIII deficient) patients can be unexpectedly attenuated by lowering the plasma levels of the procoagulant protein FV.[131]

Zebrafish offer an integrated framework both for the high-throughput discovery of novel factors and lower-throughput characterization of the complex multigenic environment of known factors. Using forward genetics facilitates the unbiased discovery of factors outside the canonical pathway of coagulation and may uncover diagnostic markers for VTE risk stratification and therapeutic targets for risk minimization. Incorporating genome engineering with the large horizontal pedigrees of zebrafish will provide in depth knowledge of how multiple genes

interact to either exacerbate or reestablish the hemostatic balance induced by known genetic deficiencies. The long-term goal is to help identify which at-risk patients should receive robust therapeutic intervention and which patients can be safely monitored without unnecessary anticoagulant therapy.

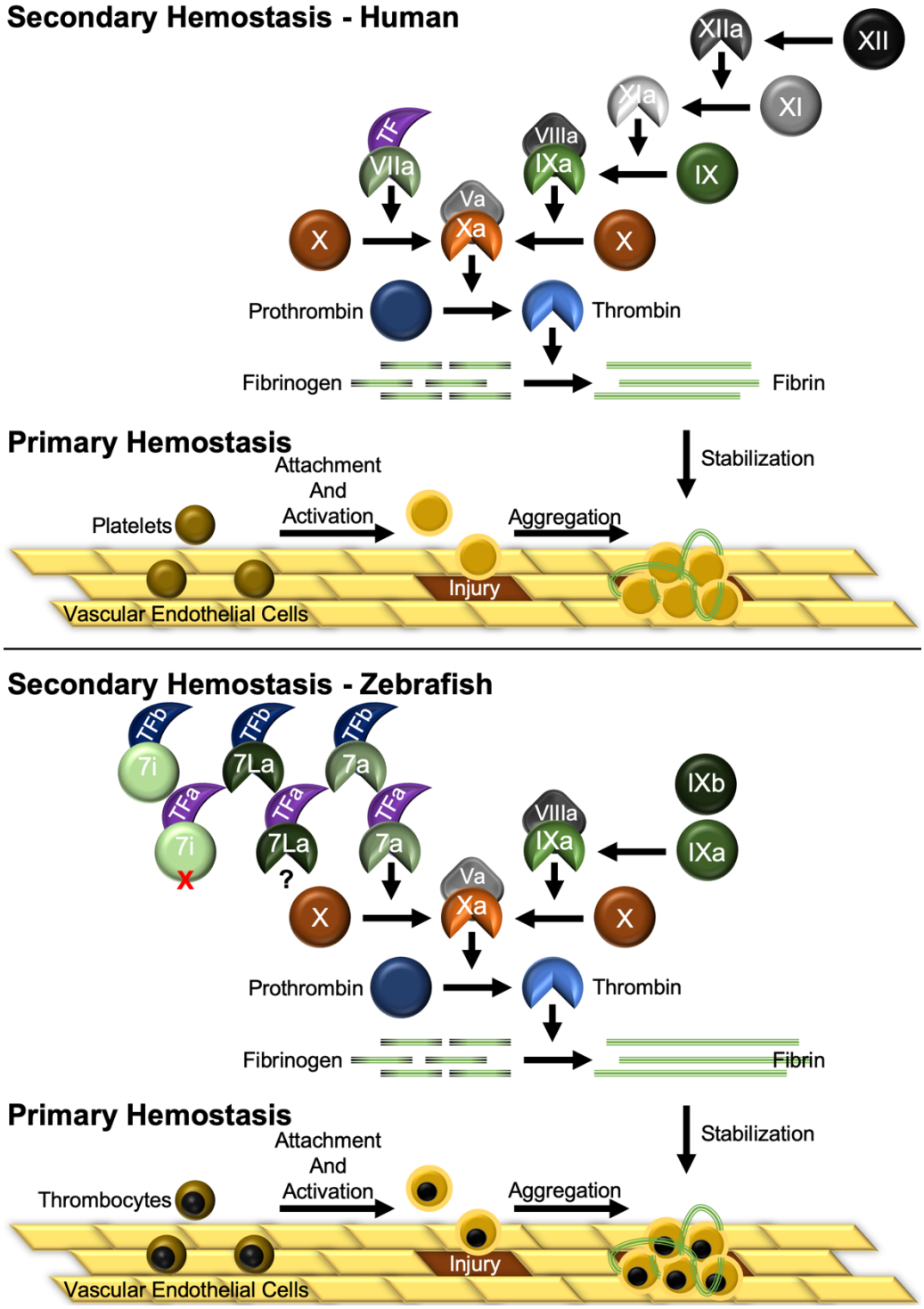


Figure 1-1. Diagram of hemostasis demonstrates slight differences between human and zebrafish coagulation cascades.

Human (top) and zebrafish (bottom) have a similar process of primary hemostasis carried out by platelets in humans and nucleated thrombocytes in zebrafish. Secondary hemostasis differs due to the additional copies of factor IX (f9), tissue factor (TF), and factor 7 (f7) and the absence of factor XI (F11) and factor XII (F12).

Chapter II:

Disruption of the Kringle 1 Domain of Prothrombin Leads to Late Onset Mortality in Zebrafish

*Note: Chapter 2 was published in **Scientific Reports** as a regular article in the March 04, 2020 issue (Co-Authors: Steven J. Grzegorski, Zhilian Hu, Yang Liu, Xinge Yu, Allison C. Ferguson, Hasam Madarati, Alexander P. Friedmann, Deepak Reyon, Paul Y. Kim, Colin A. Kretz, J. Keith Joung and Jordan A. Shavit).*

INTRODUCTION

Maintaining blood flow in a closed circulatory system requires a delicate balance between pro- and anticoagulant factors. Disequilibrium of these factors in either direction can lead to pathology. In response to vascular injury, the balance shifts towards coagulation in an effort to stabilize blood clots and prevent exsanguination. Critical to this clot stabilization is the activation of prothrombin, a vitamin K-dependent clotting factor, to form the central clotting enzyme, thrombin.[132] Thrombin cleaves soluble fibrinogen into fibrin monomers, which then polymerize to form the insoluble fibrin clot.[133] Additionally, thrombin interacts with protease activated receptors (PARs) on the surface of various cells, including platelets, which makes it the most potent endogenous agonist of primary hemostasis.[134] Human prothrombin (F2*)* variants have been linked to both thrombophilia and bleeding diatheses. The most common variant (~2% in European populations)[135] is a guanine to adenine transition in the 3' untranslated region at position 20210 that leads to increased plasma prothrombin levels and a 2-3 fold elevated risk of deep vein thrombosis.[4,35] Congenital prothrombin deficiencies result in a bleeding diathesis, but are rare. Acquired deficiencies due to liver failure or vitamin K deficiency are more common.[136,137]

Structurally, prothrombin consists of six domains: signal peptide, propeptide, Gla, kringle 1, kringle 2, and the heavy chain and light chain forming the serine protease.[132] Following translation, the propeptide targets the protein for post-translational gamma-carboxylation of the glutamic acid residues within the Gla domain. This vitamin K-dependent process is necessary for proper localization of the mature zymogen to the activated membrane surface.[138] While relatively understudied, the kringle domains are thought to interact with activated factor V (FVa) during assembly of the prothrombinase complex (FVa and activated factor X (FXa)).[139-141] Cleavage of prothrombin by the prothrombinase complex results in fragment 1.2 (the Gla and 2 kringle domains) and thrombin (light and heavy chains).

Targeted mutagenesis of *F2* in mice demonstrates 50% mortality by embryonic day 10.5 with hemorrhage associated with suspected defects in yolk sac vasculature. Mutants surviving to birth succumb to hemorrhage by postnatal day 1.[142,143] Loss of other common pathway factors, including FV and FX, show a similar pattern of bimodal lethality.[144,145] These data are suggestive of a secondary role beyond the coagulation system during development. More recent studies reveal that conditional knockout of prothrombin in adult mice leads to mortality due to hemorrhagic events within 5-7 days, although there remains a residual level of prothrombin.[146] Combined with the long half-life of prothrombin (60-72 hours)[147] it is likely that the adult mice do not achieve complete deficiency prior to lethality.

Due to early lethality of complete genetic knockouts of prothrombin in mice, the nuances of thrombin's role in *in vivo* hemostasis and development are difficult to study. Zebrafish (*Danio rerio*) has advantages for the investigation of early development because of its high fecundity, external fertilization, and transparent development. Embryonic and adult studies have demonstrated the benefits of zebrafish for the study of hemostatic and other human diseases.[69,148] Conservation of coagulation factors in zebrafish is well characterized with a variety of techniques and genetic models.[19,75,76,91,149-151] FX knockout (*f10*$^{-/-}$) zebrafish survive several months into adulthood before succumbing to lethal hemorrhage, suggesting that the model is more resistant to extreme disturbances in hemostasis than

mammals.[122] Notably, evaluation of *f10*$^{-/-}$ fish vasculature shows no abnormalities in the embryonic period.

Complete prothrombin deficiency has never been seen in humans and is incompatible with life in mice, limiting the ability to understand the entirety of prothrombin's *in vivo* functions. Therefore, we have now used genome editing TALENs (transcription activator-like effector nucleases) to produce a genetic knockout of the prothrombin gene (*f2*) in zebrafish. We show here that loss of prothrombin in zebrafish does not result in severe developmental defects but does prevent the formation of clots in response to endothelial injury and leads to early mortality at 2-3 months of age. Furthermore, the mutation generated lends insight into the role of the kringle 1 domain in prothrombin biosynthesis and activation.

MATERIALS AND METHODS

Animal Care

Zebrafish were maintained according to protocols approved by the University of Michigan Animal Care and Use Committee. All wild-type fish were a hybrid line generated by crossing AB and TL fish acquired from the Zebrafish International Resource Center. A hybrid line was selected to minimize strain specific effects and for improved reproductive success.[152] *Tg(cd41:egfp)* fish were used for tracking of fluorescently labeled thrombocytes.[60] Tris-buffered tricaine methanosulfate (Western Chemical) was used for anesthesia during all live animal experimental manipulation and for euthanasia.

Sequence analysis

Prothrombin protein sequences (NP_000497.1, NP_034298.1, NP_989936.1, NP_998555.1, NP_001015797.1) were downloaded from the National Center for Biotechnology Information RefSeq database.[153] Homology predicted propeptides were removed and the prothrombin numbering scheme[154] was used relative to the human prothrombin sequence unless otherwise noted.

Zebrafish sequence was modified to include an incompletely annotated exon. Sequences were aligned using MUSCLE[155] and shaded using the Boxshade server (https://embnet.vital-it.ch/software/BOX_form.html).

Targeted mutagenesis using TALENs

The *f2* coding sequence was identified in the zebrafish reference genome (Gene ID: 325881).[85] Factoring in technical design limitations and predicted efficacy, an optimal target site 5' to the protease domain was selected in exon 6 for mutagenesis. TALEN constructs were created using fast ligation-based solid-phase high-throughput (FLASH) assembly targeted to exon 6 of the zebrafish genomic *f2* locus and validated as described.[114,156] mRNA was transcribed from plasmids and injected into single cell embryos. The resulting chimeric founders were raised, outcrossed, and offspring screened for deleterious mutations. A single founder heterozygous for a 14 base pair deletion ($f2^{\Delta 14}$) in the coding sequence was identified and crossed to wild-type fish to establish the mutant line.

Genotyping of mutant offspring

Whole embryos or adult fin biopsies were lysed in buffer (10 mM Tris-HCl, pH 8.0, 2 mM EDTA, 2% Triton X-100, 100 µg/mL proteinase K) for 2 hours at 55°C followed by incubation at 95°C for 5 minutes. One microliter of the resulting solution was used as template for gene specific polymerase chain reaction (Table 1-1) and analyzed by gel electrophoresis.

Histochemical analysis

For *in situ* hybridization, DIG-labeled riboprobes (DIG RNA-labeling kit, Roche) were synthesized using 2-day old wild type embryonic cDNA and gene specific primers with T7 overhangs (Table 1-1). Embryos were fixed overnight at 4°C in 4% paraformaldehyde in phosphate buffered saline prior to dehydration. Permeabilization and staining were performed as described.[157] Stained samples were evaluated by phenotype prior to genotyping.

For hematoxylin and eosin staining, juvenile zebrafish (30-89 days) were fixed overnight in 4% phosphate buffered paraformaldehyde and embedded in paraffin. Sagittal sections (3 μm) were collected every 50 μm and stained.

Single molecule real-time sequencing of RNA

Three *f2*$^{+/\Delta 14}$ larvae were homogenized at 5 days post fertilization (dpf) in lysis buffer using a 21-gauge syringe. Total RNA was purified from the lysate using the PureLink RNA Mini Kit (Life Technologies) followed by DNAse I treatment (Invitrogen) and cDNA synthesis using the Superscript III First Strand cDNA kit (Invitrogen). Primers (Table 1-1) were used to amplify a 918-base pair (bp) region surrounding the predicted deletion. This region included several non-deleterious single nucleotide polymorphisms (SNPs) in downstream exons known to be allelic with either the wild-type or mutant alleles. Purified products from the 3 samples were pooled and sent to the University of Michigan Sequencing Core for library preparation and single molecule real-time high throughput sequencing (SMRT, Pacific Biosciences). Circular consensus reads with at least 5x coverage were filtered for full-length single inserts. After filtering for quality, the resulting 3826 reads were sorted by haplotype and splice variation.

Quantitative real-time PCR

Total RNA was extracted from 7 dpf larvae using the RNeasy Mini Plus kit (Qiagen) and transcribed using oligo(dT)$_{12\text{-}18}$ primer and Superscript III (Invitrogen). Three pools of three whole larvae were used per genotype. The resulting cDNA was used as template for qPCR (StepOnePlus, Applied Biosystems) using gene specific PrimeTime© probe-based qPCR Assays (Integrated DNA Technologies) (Table 1-1). The expression level of *f2* was normalized to the *actb2* gene and significance analyzed using the double delta Ct method as described.[158] Transcript flanking the deletion was amplified using allele unbiased primers and separated by capillary gel electrophoresis (Table 1-1). Molar ratios were calculated from standardized relative band intensities using QIAxcel ScreenGel software (Qiagen).

Laser-induced endothelial injury

Anesthetized zebrafish larvae were embedded in 0.8% low melting point agarose and oriented in the sagittal position on a glass coverslip. Larvae were then positioned under an Olympus IX73 inverted microscope with an attached pulsed nitrogen dye laser (Micropoint, Andor Technology). For venous injury, ninety-nine laser pulses were administered to the luminal surface of the endothelium on the ventral side of the posterior cardinal vein (PCV) 5 somites posterior to the anal pore of 3 dpf larvae as previously described.[159] For arterial injury, pulses were administered to the endothelial surface of the dorsal aorta 3 somites posterior to the anal pore of 5 dpf larvae.[159] Following injury, time to occlusion and/or time to thrombocyte attachment were monitored for 2 minutes by a blinded observer. Larvae were manually extracted from agarose for genotyping.

Chemical treatment

Dabigatran etexilate, apixaban, and rivaroxaban were dissolved in DMSO and diluted in embryo water to final concentrations of 50, 100, and 250 μM, respectively. At 5 dpf larvae were treated for 24 hours prior to laser-mediated endothelial injury on day 6. For ε-aminocaproic acid (Sigma) treatment, zebrafish embryos were incubated in 100 mM at 1 dpf and the time to occlusion was evaluated at 3 dpf following laser-induced endothelial injury.

O-dianisidine staining

Anesthetized larvae were incubated in the dark at 4 dpf for 30 minutes in o-dianisidine staining solution as previously described.[160,161] Larvae were subsequently fixed overnight at 4°C in 4% paraformaldehyde, and pigment was removed using bleaching solution (1% KOH, 3% H_2O_2).

Human prothrombin expression vector construction and injection

Plasmid pF2-bio-His was obtained from Addgene (plasmid #52179) and human *F2* cDNA amplified using primers containing vector sequence homology (Table 1-1). pcDNA3.1 was digested with BstBI and KpnI, followed by gel purification of the linearized backbone. The NEBuilder kit was used to fuse the *F2* cDNA with linearized vector and transformed into 10beta competent cells (NEB). Site directed mutagenesis was used for missense mutation introduction.[75] and NEBuilder end homology joining was used to create internal deletions. Replacement of residues 100 to 115 of human F2 with a glutamic acid resulted in a homologous 15 amino acid deletion in human prothrombin (Δ15). As the deletion includes the cysteine residue at position 114, the effect of introducing a free-cysteine at position 138 due to the loss of its binding partner was investigated by generating two additional vectors that included an alanine substitution without (C138A) or with the 15-residue deletion (C138A/Δ15). For *in vivo* rescue assays, plasmid DNA (90 pg) in 0.1M KCl was injected into one-cell stage embryos generated from heterozygous *f2*$^{\Delta 14}$ incrosses using pulled glass capillary needles.

Human prothrombin protein expression and isolation

F2 mutant vectors were transfected into human embryonic kidney 293T (HEK293T) cells in 6-well plates using Lipofectamine 3000 (ThermoFisher Scientific; L3000008) according to manufacturer's instructions. The media and cells were collected after 3 days. The cells were washed with DPBS and lysed with 50 µL RIPA buffer on ice for 15 minutes. Prothrombin concentration in the expression media and cell lysate were quantified by ELISA using a matched-pair antibody set (Affinity Biologicals Inc., Ancaster, Canada), according to manufacturer's instructions.

To characterize the functional characteristics of the prothrombin derivative, the plasmid containing the cDNA for prothrombin C138A/Δ15 was transfected and selected in HEK293 cells to generate a stable cell line. Using sustained rolling incubation at 37°C, expression was induced in D-MEM supplemented with vitamin K_1, and collected every 2-3 days.[162] The media were then centrifuged at 2,000x g

to remove cell debris, and the supernatant was filtered through a 0.22 μm filter (Millipore). The media was loaded onto tandem XAD_2 and Q-Sepharose fast flow columns pre-equilibrated with 0.02M Tris, 0.15M NaCl, pH 7.4 (TBS) and eluted with 0.02M Tris, 0.5M NaCl, pH 7.4 as described previously.[163]

Prothrombin activation and activity

Prothrombin activation was quantified as described previously.[163] Briefly, prothrombin (0.1 μM) was incubated with PCPS vesicles (50 μM; 75% phosphatidylcholine, 25% phosphatidylserine), FVa (20 nM), $CaCl_2$ (5 mM), and dansylarginine *N*-(3-ethyl-1,5-pentanediyl)amide (1 μM, DAPA) in 0.02 M HEPES, 0.15 M NaCl, pH 7.4 with 0.01% Tween80 (HBST). The reactions were initiated by the addition of FXa (70 pM) and activation was monitored for fluorescence change at 1-minute intervals using a SpectraMax M2 plate reader (Molecular Devices). Excitation and emission spectra were 280 nm and 540 nm, respectively, with an emission cutoff filter set at 530 nm. The quantum yield of the thrombin-DAPA complex was determined by plotting the total signal change observed with respect to known concentrations of the thrombin-DAPA complex.[141]

To determine the resulting thrombin activity, prothrombin was activated to thrombin by prothrombinase as described above. After a 30-minute incubation, complete prothrombin activation was verified by SDS-PAGE (not shown). The resulting thrombin (5 nM) was simultaneously added to either a thrombin-specific substrate S-2238 (400 μM; DiaPharma, West Chester, OH) or fibrinogen (3 μM), all in the presence of 5 mM $CaCl_2$ in HBST. Both reactions were monitored at 405 nm using SpectraMax M2 plate reader (Molecular Devices) to quantify the amidolytic (S-2238) or clotting (fibrinogen) activity of thrombin.

Statistical analysis

The occlusion data were analyzed using Mann-Whitney *U* or two-tailed Student *t* tests. Survival was evaluated by log-rank (Mantel-Cox) testing. Significance testing, graphs, and survival curves were made using GraphPad

Prism (Graphpad Software, La Jolla, California). P-values ($p<0.05$ or $p<0.0001$) were used to evaluate statistical significance.

RESULTS

Zebrafish prothrombin demonstrates a high degree of sequence conservation

Mature zebrafish prothrombin is a 579 amino acid protein that shares 55% identity and 72% homology with human prothrombin, including 68% identity and 86% homology when comparing the regions corresponding to thrombin. Zebrafish prothrombin is detected within the first day of development and leads to measurable fibrin forming activity by 48 hours.[82] Overall, the prothrombin sequence is highly conserved across vertebrates, with complete conservation of cysteine residues, suggesting the conservation of major structural features such as the kringle domains (Figure 2-1).

Zebrafish thrombin activity is inhibited by direct oral anticoagulants

To test the relative conservation between zebrafish and human thrombin, larvae were treated with three direct oral anticoagulants: direct thrombin inhibitor (dabigatran) as well as direct FXa inhibitors (rivaroxaban and apixaban). Following 24 hours of treatment, 6 dpf larvae exposed to dabigatran etexilate, rivaroxaban, or apixaban were unable to form occlusive thrombi in response to injury (Figure 2-2). These data show that the carboxylesterase function required to convert dabigatran etexilate to dabigatran is conserved in zebrafish, as is the ability of dabigatran to inhibit zebrafish thrombin. Apixaban and rivaroxaban also demonstrate conserved inhibition of FXa, which inhibits prothrombin activation. In humans, mean plasma concentrations of all 3 inhibitors are in the range of 100-600 nM.[164,165] Similar concentrations of dabigatran and 10-100 fold higher concentrations of rivaroxaban have been shown to impair clot formation following tail bleeds in mice.[166,167] Although difficult to predict, the low polarity of the 3

molecules leads us to estimate that the molecules will be poorly absorbed from the media and result in similar tissue levels to those seen in humans.[168]

Targeted genome editing induces a deletion resulting in decreased f2 expression

Having established the functional conservation of prothrombin, we sought to analyze the long-term effects of thrombin deficiency using a genetic model. Utilizing TALEN-mediated genome editing, exon 6 of *f2* was targeted with the aim of creating a frameshift and subsequent nonsense mutation prior to the protease domain. Sequencing data showed a 14 bp deletion within the genomic region homologous to the human prothrombin kringle 1 domain (Figure 2-3A). *In situ* hybridization demonstrated decreased, but not absent *f2* mRNA in homozygous mutants at 3 and 5 dpf compared to wild-type siblings (Figure 2-3B). This is further supported by quantitative RT-PCR data demonstrating a 45% reduction in mRNA transcript in homozygous mutants (Figure 2-3C). To characterize the residual mutant transcript, semi-quantitative RT-PCR was performed using primers flanking the mutation site. Only wild-type and mutant bands were seen in pools of wild-type and homozygous mutant embryos, respectively. In heterozygous embryos, the computed molar amount of the mutant band was only 26% of the total, with the remainder being wild-type (Figure 2-3D). Notably, the homozygous mutant transcript was roughly 30 base pairs smaller than expected.

Genomic deletion reveals a cryptic splice site that creates an alternative splice variant

To identify potential splice variants, full length *f2* cDNA was sequenced from $f2^{+/+}$ and $f2^{\Delta 14/\Delta 14}$ larvae (Figure 2-4, top). This revealed that there is a cryptic splice acceptor 3' to the 14 bp genomic deletion in exon 6. In contrast to the canonical splicing which forms a transcript containing the 14 bp deletion, alternative splicing from the exon 5 splice donor site to the cryptic splice site generates a 45 bp deletion in kringle 1 of prothrombin (Figure 2-4). Notably, this maintains the reading frame and replaces DEFNVTQLKLQENFCR with a single glutamic acid for a net

loss of 15 amino acids. This deletion includes a highly conserved cysteine (C119 in the zebrafish peptide) predicted by homology to form a disulfide bond with zebrafish C143 (Figure 2-1). The protein product of this alternatively spliced transcript is referred to as Δ15. To understand the distribution of splicing, we performed deep SMRT sequencing of *f2* cDNA from *f2*$^{+/\Delta 14}$ mutant larvae at 3 dpf. Overall, 3842 consensus sequences were analyzed. Using downstream allelic polymorphisms, the reads were sorted into wild-type and *f2*$^{\Delta 14}$ haplotypes. Of the 2384 reads derived from the wild-type locus, 100% underwent canonical splicing (Figure 2-4, middle). In contrast, only 17/1458 (1.2%) mutant locus reads displayed canonical splicing (Figure 2-4, bottom) and the remainder made use of the cryptic splice site to form the Δ15 transcript (1441/1458, 98.8%).

Disruption of the prothrombin kringle 1 domain results in adult lethality due to internal hemorrhage

To examine the functional consequence of kringle 1 disruption, we evaluated homozygous mutant fish at various stages of development. Beginning around 1 month of age, *f2*$^{\Delta 14/\Delta 14}$ fish showed signs of overt internal hemorrhage and the majority (23/24) died by adulthood, defined as 90 days of age (Figure 2-5A-B). Gross intracranial and fin hemorrhage was visible, and histological analysis demonstrated the occurrence of bleeds within the head, jaw, muscle, fins, and around the heart and abdomen (Figure 2-5C).

Homozygous mutant larvae fail to form thrombi in response to endothelial injury

Given that 80% of *f2*$^{\Delta 14/\Delta 14}$ fish were able to survive development to the juvenile stage (30-90 days of age), we assessed the possibility that there could be residual detectable thrombin activity in early embryos and larvae. Thrombin has roles in both primary and secondary hemostasis, platelet activation and fibrin formation, respectively. These roles were evaluated using laser-mediated endothelial injury models of arterial and venous thrombus formation (Figure 2-6A). Homozygous mutant larvae were unable to form occlusive thrombi in the venous

system in response to injury. This was refractory to treatment with the fibrinolytic inhibitor ε-aminocaproic acid, further confirming a lack of fibrin generation (Figure 2-6B). To confirm that defects in hemostasis were not due to off target effects of genome editing, embryos were injected at the one-cell stage with plasmid expressing human *F2* cDNA driven by the constitutively active cytomegalovirus promoter.[169,170] Endothelial injury at 3 dpf induced clot formation within 2 minutes in 50% of homozygous embryos in contrast to uninjected controls (Figure 2-6C). To assess the role of thrombin in the zebrafish arterial system, the $f2^{\Delta 14}$ mutation was bred into the *Tg(cd41-egfp)* background in which circulating thrombocytes express GFP. At 5 and 6 dpf, $f2^{\Delta 14/\Delta 14}$ larvae had a decreased ability to form occlusive arterial thrombi (Figure 2-6D). The time to thrombocyte attachment was not statistically different between groups (Figure 2-6E) while the number of attached thrombocytes appeared to be slightly increased in the $f2^{\Delta 14/\Delta 14}$ larvae, but only at 6 dpf (Figure 2-6F). Overall these data reveal intact primary hemostasis but a loss of secondary hemostasis. Despite this, erythrocyte staining with o-dianisidine revealed no overt signs of hemorrhage in 1-week old $f2^{+/+}$ or $f2^{\Delta 14/\Delta 14}$ larvae (data not shown).

An intact kringle 1 domain is important for normal prothrombin levels

To interrogate how partial loss of the kringle 1 domain alters prothrombin function, a Δ15 version of human prothrombin was generated, corresponding to the zebrafish deletion. As this deletion removes C114 and results in a free-thiol at C138, two additional prothrombin variants were generated (C138A and C138A/Δ15). Transient expression in HEK293T cells led to wild-type secretion >100 ng/mL while all mutants were <3 ng/mL and not significantly different from the non-transfected control (Figure 2-7A). Furthermore, measurement of prothrombin in the cell lysate revealed >10 ng/mL for wild-type expression while two of the three mutants (C138A and C138A/Δ15) were < 5 ng/mL and not significantly different from control. The remaining mutant, Δ15, was significantly different than control but on average produced less protein than the wild-type product (Figure 2-7B). Overall, this suggests that the loss of these residues within

kringle 1 impaired biosynthesis or stability and led to a significant reduction in overall secretion.

The kringle 1 mutant has impaired endogenous activity despite an intact protease domain

The C138A/Δ15 prothrombin was expressed from a stably transfected HEK293 cell line, purified, and its activation by prothrombinase was quantified. The quantum yield of fluorescence signal change observed upon thrombin-DAPA complex formation was comparable between wild-type (3262 ± 347 RFU/μM) and C138A/Δ15 (3354 ± 180 RFU/μM), consistent with the protease domain remaining intact. However, the initial rate of activation, defined as the linear phase of steady state kinetics during the first 15-20% of the reaction, was 3.3-fold faster for wild-type (19.5 ± 1.4 s^{-1}) compared with C138A/Δ15 (5.8 ± 0.5 s^{-1}) (Figure 2-7C).

The functionality of thrombin generated from the mutant prothrombin variant was then investigated. The amidolytic activity of thrombin towards S-2238 was similar between wild-type (82.7 ± 6.5 mOD/min) and C138A/Δ15 (89.0 ± 6.4 mOD/min) (Figure 2-7D). However, C138A/Δ15-thrombin had a significantly prolonged clot time (200.6 ± 5.3 s) compared to wild-type (67.7 ± 3.1 s) (Figure 2-7E), mostly due to delayed clot initiation and 1.5-fold higher total turbidity change (Figure 2-7F) compared with wild-type. The rate of clot formation was also slower (82.4 ± 6.3 mOD/min) compared with the wild-type (135.9 ± 12.3 mOD/min) (Figure 2-7G). These data suggest that thrombin derived from C138A/Δ15 prothrombin has normal catalytic properties towards small peptidyl substrates, but impaired activity towards macromolecular substrates like fibrinogen.

DISCUSSION

The zebrafish model has been developed as a useful tool for understanding coagulation, especially during early development. In contrast to mammals, in which a number of coagulation factors are necessary for embryonic and/or neonatal viability, zebrafish are able to survive the loss of many aspects of the canonical

cascade at least until early adulthood. Loss of antithrombin, fibrinogen, FV, and FX are all compatible with development to adulthood.[19,75,77,122,151] Targeting of *f10* in fish results in the absence of larval hemostasis, but is accompanied by extended survival.[122] These data suggest that there could be residual thrombin activity present. Previous studies have utilized transient knockdown and chemical inhibition to study the loss of prothrombin in zebrafish,[149,171-173] providing valuable insight into the conserved function of prothrombin and suggesting a potential role for the coagulation cascade in development. Unfortunately, these technologies are susceptible to toxicity and off target effects. Additionally, specific chemical inhibition targets the proteolytic activity of thrombin without addressing the potential for exosite mediated binding of other factors. We sought to leverage knockout technology to create a specific genetic knockout of *f2*. We produced a genomic deletion in exon 6 with the intent of creating a nonsense mutation. Unexpectedly, we found reduced expression of an alternatively spliced version of prothrombin lacking a conserved cysteine in the kringle 1 domain. *In vitro* biochemical studies demonstrated that this protein is synthesized and secreted at a low level, and the secreted protein has decreased activation and activity. *In vivo*, the Δ15 mutant results in the loss of the ability to form fibrin rich clots and lethality by 2 months of age due to spontaneous hemorrhage. However, thrombocyte attachment was unaffected, consistent with the dogma that platelet adherence is mediated primarily via von Willebrand factor and collagen binding.[174-176] Overall, our data suggest that this is likely a severe hypomorph, if not functionally a complete null. Given that $f2^{\Delta 14/\Delta 14}$ phenocopies the previously described *f10* knockout, it appears that any residual thrombin activity is likely to be physiologically negligible

Surprisingly, no obvious developmental defects were observed, including grossly normal vascular development, with survival through early adulthood. Previous work using antisense technology has demonstrated a variety of developmental malformations including hemorrhage, and circulatory, brain, and tailbud malformations.[171] These defects were not observed across multiple clutches in our knockout. This inconsistency could be due generally to off target

effects of antisense knockdown or genetic compensation that can occur in response to genomic editing.[177] Conversely, the discrepancy may be a result of the fibrinogen-independent residual activity of the mutant thrombin, as detected in the amidolytic assay. Notably, the *victoria* mutant, which mapped close to *f2* and is likely a mutant allele, similarly affects induced thrombus formation without a described developmental defect,[70] consistent with our data.

Kringle 1 is an understudied domain of prothrombin, but it is thought to interact directly with FVa and regulate prothrombin activation.[140] For the first time, we demonstrate that a defect in the kringle 1 domain may lead to decreased levels of prothrombin in circulation as well as decreased thrombin generation, thus contributing to the observed coagulopathy. The impairment of overall thrombin activity occurs at multiple levels. First, presumed inefficiency of the cryptic splice site in mRNA maturation leads to a roughly 45% reduction in transcript levels due to predicted nonsense mediated decay of the mutant transcript making use of the canonical splice site. Notably, splicing to the cryptic site is not seen in wild-type fish and a cryptic AG splice acceptor is not present in the human transcript. Second, when introduced into human prothrombin, the deletion resulted in a substantial decrease in secretion. It is possible that a large amount of the mutant prothrombin misfolds, aggregates, and is degraded intracellularly. Additionally, low secretion may be indicative of a more involved role of the kringle 1 domain in regulating prothrombin secretion as has been previously suggested.[178] The C138A/Δ15 prothrombin derivative (controlling for the unpaired C138 resulting from Δ15) showed a 3-fold reduction in its rate of activation by prothrombinase when compared with the wild-type prothrombin. In addition, thrombin generated from the C138A/Δ15 variant had a decreased procoagulant function (*i.e.* macromolecular substrates) while maintaining a functional active site as demonstrated by the unaltered S-2238 hydrolysis (*i.e.* small synthetic substrates) when compared with wild-type thrombin. Although similar thrombin functionality towards S-2238 was expected since the mutations did not involve the protease domain of prothrombin (i.e. thrombin), the delineation of substrate specificity towards fibrinogen was unexpected. This variation in substrate specificity could be

explained by incomplete cleavage leading to an accumulation of a meizothrombin intermediate.[179] However, we did not observe the associated increase in DAPA-thrombin fluorescence predicted by increased meizothrombin presence. Conversely, our findings propose the potential role of F1.2 in the substrate recognition and specificity of thrombin, whereby the disruption of the kringle 1 domain leads to reduced fibrinogen recognition and/or cleavage by the F1.2:thrombin complex. Therefore, our data suggest that F1.2 is more than just a peptide fragment that results in product-inhibition of thrombin.[180] Overall, the compounding reductions in transcription, secretion, activation, and activity shows that this genomic deletion results in a severe hypomorph. At the same time our data propose a role for the kringle 1 domain in the activation kinetics and subsequent substrate recognition of thrombin.

The ability of zebrafish to survive severe hemostatic imbalance may be due to a combination of influences, including the absence of birthing trauma, limited hemostatic challenges in an aqueous laboratory environment, a relatively low systolic blood pressure, and possible species-specific genetic differences. Nevertheless, the phenotypes of all zebrafish hemostatic mutants eventually converge with their mammalian counterparts in adulthood. Overall, this study further demonstrates the conservation of the coagulation cascade in fish while leveraging unique physiologic differences to build our understanding of the complex biological functions of thrombin. Furthermore, understanding how fish tolerate such a severe bleeding diathesis could provide new insights into managing patients with congenital bleeding disorders or acquired hemorrhage, with future studies leveraging the model system to develop new diagnostic markers and therapies.

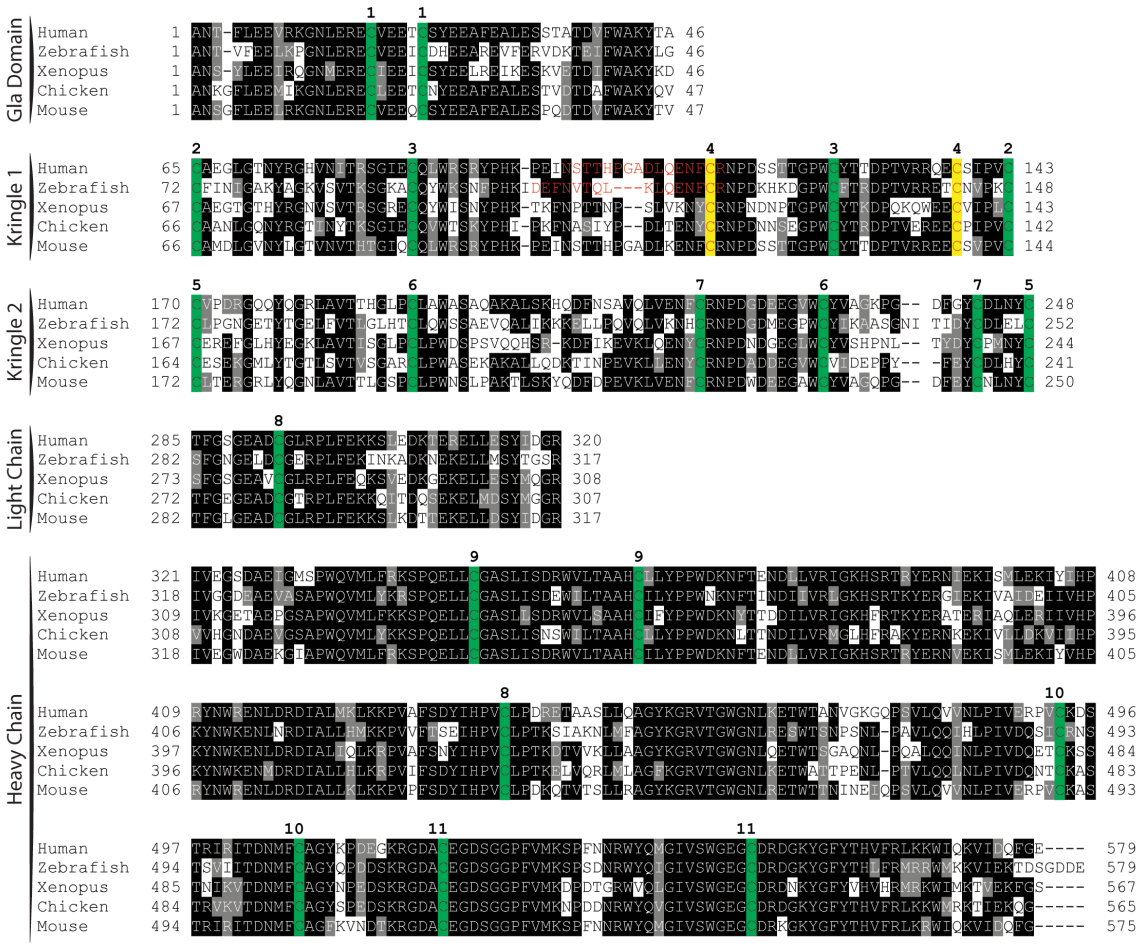


Figure 2-1. Peptide sequence alignment shows strong conservation of prothrombin across a broad range of species. Sequences shown include the conserved domains.

Numbering begins at the first residue after the propeptide according to the prothrombin numbering scheme. Sequences are shaded to indicate degree of conservation. Green colored residues indicate conserved cysteines. Red colored residues represent amino acids altered by mutagenesis (Δ15; C138A) with the affected paired cysteines highlight in yellow. Numbers above alignment indicate paired cysteines in human prothrombin.

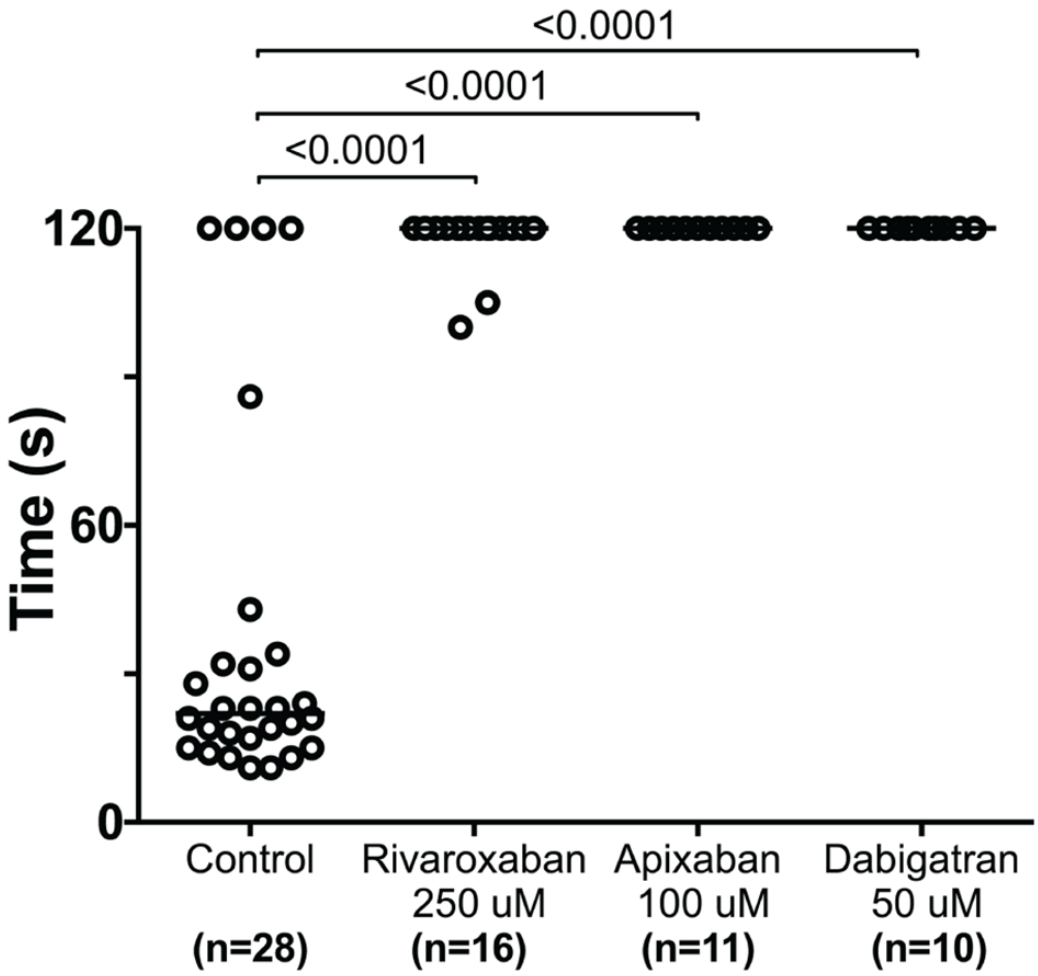


Figure 2-2. Direct oral anticoagulants prevent occlusive thrombus formation in zebrafish.

5 dpf larvae were exposed to chemical inhibitors of thrombin (dabigatran) and FXa (apixaban, rivaroxaban) for 24 hours. This resulted in the inability to form occlusive venous thrombi at 6 dpf in wild-type larvae following laser-mediated endothelial injury.

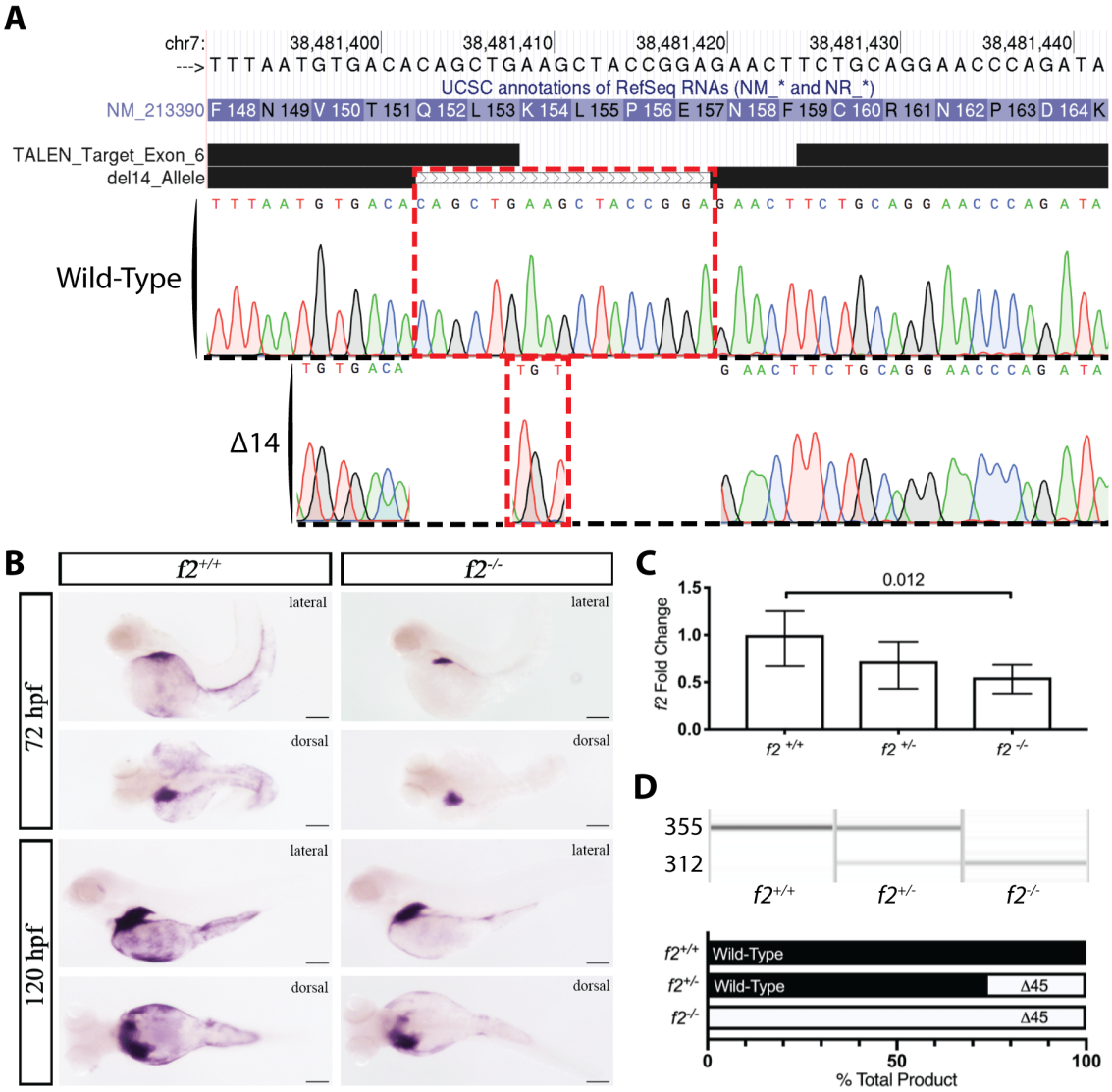


Figure 2-3. Genome editing creates a 14 bp genomic deletion with a resulting decrease in mRNA expression.

(A) Alignment of Sanger sequencing with the chromosome 7 genomic region showed an overall 17 bp genomic deletion replaced with a 3 bp insertion; outlined in red, resulting in a net 14 bp deletion. (B) *in situ* hybridization demonstrated reduction of transcript at 72 and 120 hours post fertilization in homozygous mutants compared to control siblings. Spatial regulation remained intact with expression restricted primarily to the liver. (C) qPCR data of *f2* expression reveals significant decrease of 45% in the homozygous mutant embryos. (D) Semi-quantitative RT-PCR of embryos shows a mutant band ~30 bp smaller than expected (later shown to be a 45 bp deletion, Figure 4). Quantitation of the bands reveal that the mutant band is only 26% of the total in heterozygotes.

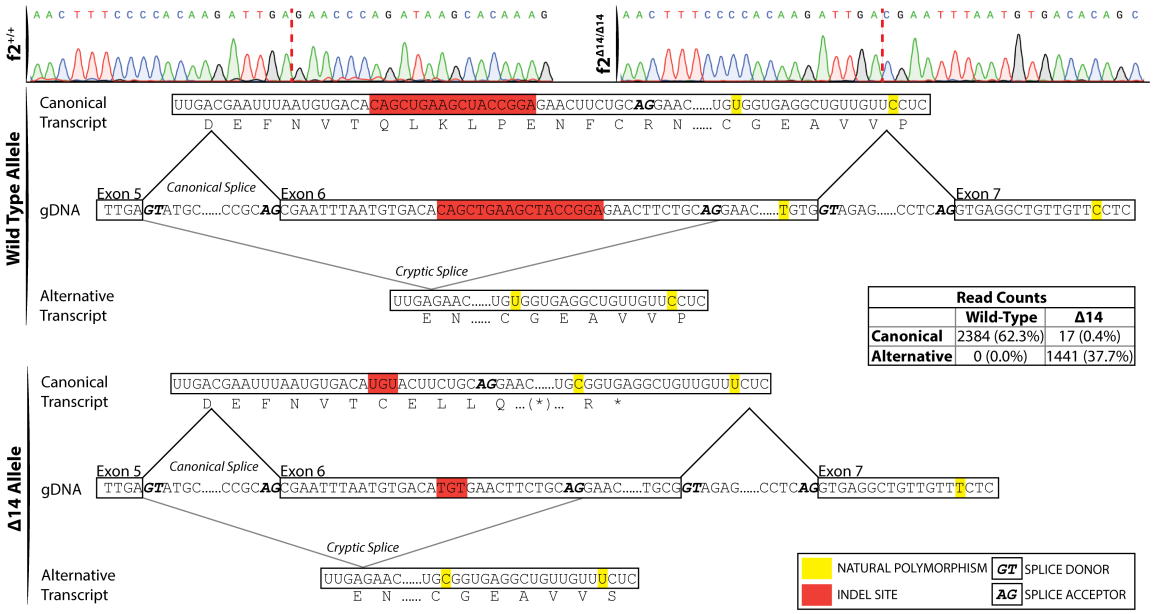


Figure 2-4. Single molecule real time sequencing of *f2*$^{+/\Delta 14}$ mRNA demonstrates altered splicing in *f2* following a deletion in exon 6.
(top) Sanger sequencing of f2$^{+/+}$ and f2 $^{\Delta 14/\Delta 14}$ cDNA reveals a canonical splice donor but alternative splice acceptor (red dashed lines indicate splice junctions). Allelic SNPs in exon 7 allowed the sorting of transcripts by their genomic allele in mRNA extracted from heterozygotes. Wild-type transcripts (middle) solely demonstrated canonical splicing. Mutant transcripts (bottom) primarily exhibited alternative splicing to a cryptic splice site with a resulting 45 bp inframe deletion, as well as low frequency canonical splicing with a resulting 14 bp out of frame deletion. gDNA, genomic DNA.

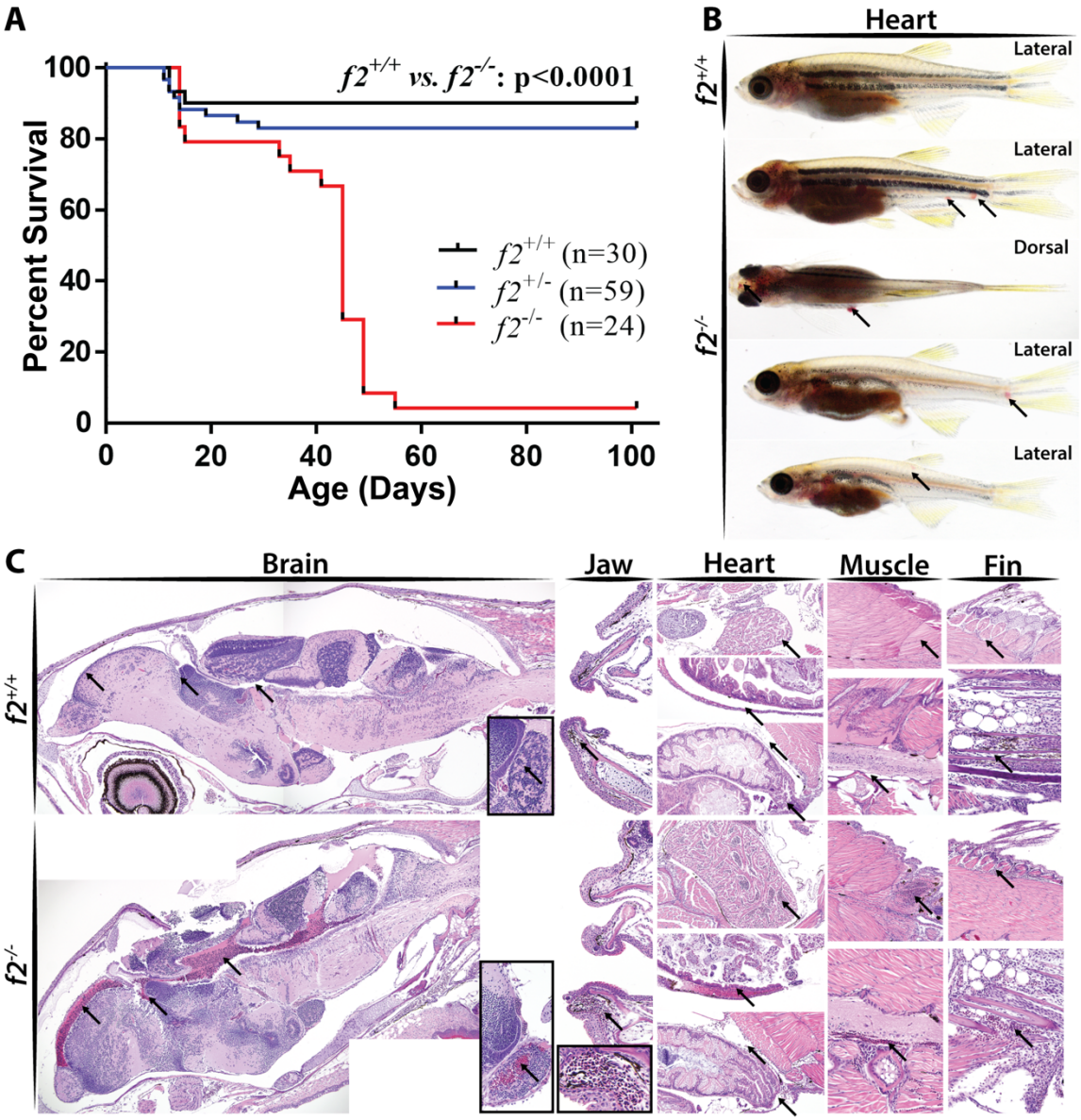


Figure 2-5. Loss of prothrombin results in early lethality due to hemorrhage. (A) Survival curve demonstrating significant mortality by 2 months of age in *f2*$^{-/-}$ siblings, log-rank (Mantel-Cox) analysis. (B) Examples of grossly visible intracranial, intramuscular, and fin bleeds (arrows). (C) Histological sections of wild-type and *f2*$^{-/-}$ siblings demonstrated microscopic bleeds in the brain, jaw, heart, muscle, and fins. Arrows point to pools of erythrocytes and unaffected comparable tissue in the control.

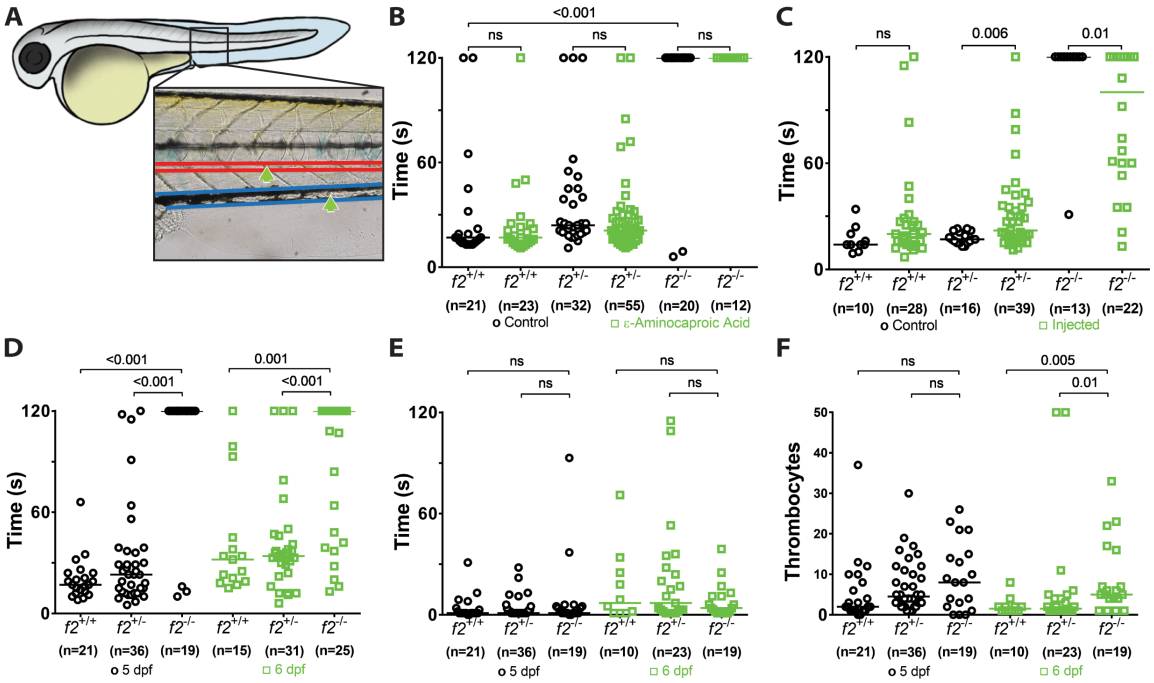


Figure 2-6. Loss of thrombin activity results in defects in secondary hemostasis.

(A) Larvae were immobilized in agarose, subjected to laser-mediated endothelial injury (green arrow) of the venous (PCV, blue) or arterial (dorsal aorta, red) circulation, and followed for 2 minutes by a blinded observer. (B) Genetic ablation of *f2* resulted in the inability to form induced PCV thrombi at 3 dpf and was not influenced by inhibiting fibrinolysis (ε-aminocaproic acid treatment, green). (C) Overexpression of human *F2* cDNA (green) rescued the ability to form thrombi in the PCV at 3 dpf. (D) Homozygous mutant larvae demonstrated a significant impairment in arterial thrombus formation at 5 and 6 dpf without any changes in the time to initial thrombocyte attachment (E). (F) The number of thrombocytes attached to the site of injury in 2 minutes was significantly increased at 6 dpf in *f2* homozygous mutants. Statistical significance assessed by Mann-Whitney *U* testing.

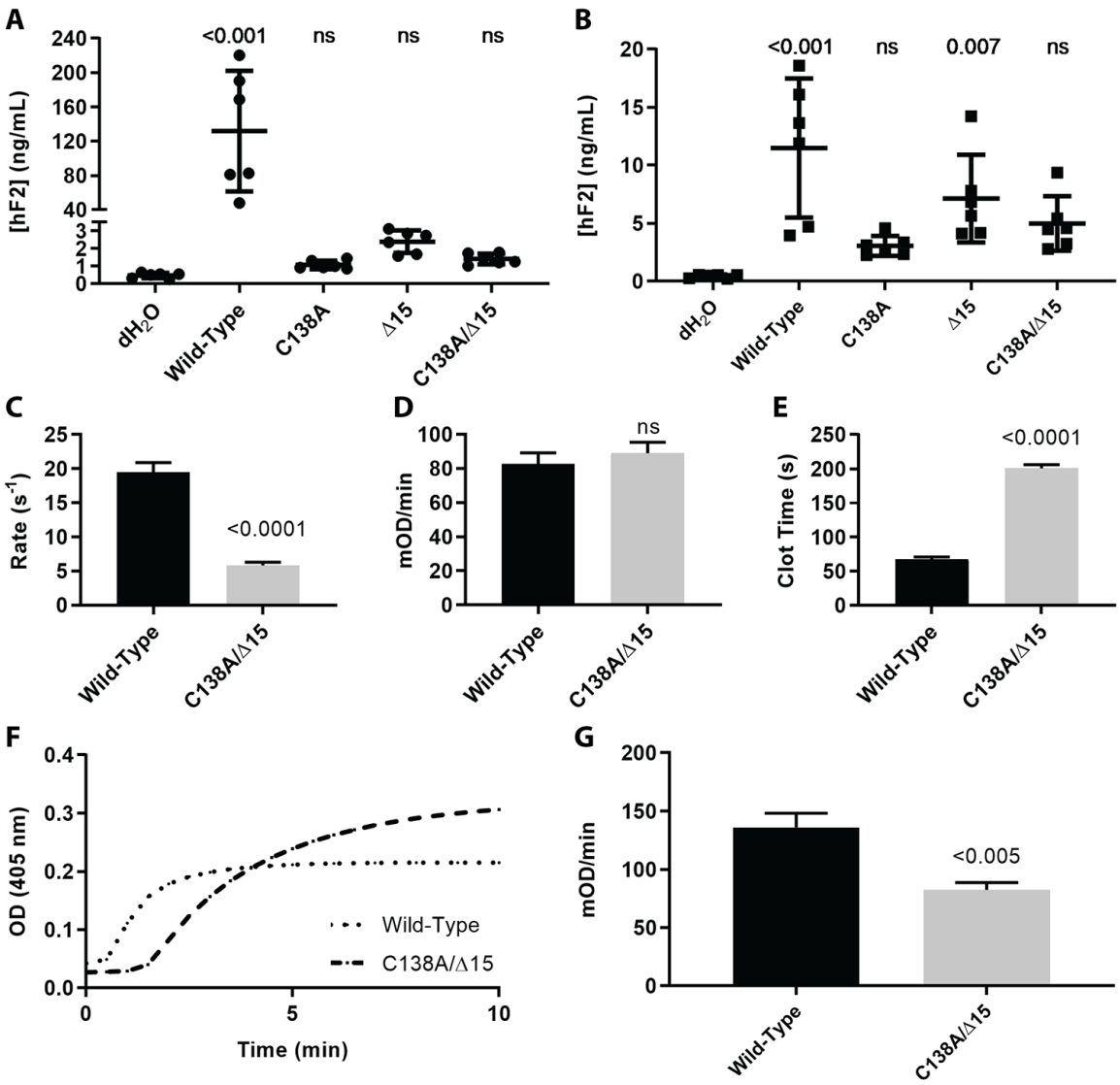


Figure 2-7. Expression and activation of prothrombin variants and resulting activity of thrombin upon activation.

(A) Measurement of prothrombin in the cell media by ELISA demonstrated that expression of prothrombin variants in HEK293T cells resulted in reduced secretion levels that were indistinguishable from control and (B) corresponding measurement of prothrombin in the cell lysate demonstrated decreased biosynthesis. (C) Rates of prothrombin activation by prothrombinase measured using DAPA. Once prothrombin variants were fully activated by prothrombinase, their ability to cleave synthetic substrate S-2238 (D) or fibrinogen (E) were monitored, with significant differences in the latter, but not the former. For both S-2238 and fibrinogen, the rate was calculated during the initial reaction phase when the substrate is non-limiting and the conditions are presumed to be steady-state. Rate was then normalized to enzyme concentration. (F) Clotting profile between wild-type and C138A/Δ15 demonstrated delayed clot initiation and altered turbidity. (G) The rates of clot formation were determined from the clotting profile and reduced in the mutant.

Table 2-1. Oligonucleotide sequences used in methods

Primer Description	Primer Sequence
f2 qPCR Forward	GAGGAAGCTCGAGAAGTGTTT
f2 qPCR Reverse	CCCTCTGTAGTCGCACATATTC
f2 qPCR Probe	FAM/TGGACAGGG/ZEN/TTGTTCCTTCACAGC/IABkFQ
actb2 qPCR Forward	ATGAAGATCCTGACCGAGAGA
actb2 qPCR Reverse	TCAAAGTCAAGGGCCACATAG
actb2 qPCR Probe	FAM/ACCACCACA/ZEN/GCTGAGAGGGAAATT/IABkFQ
f2 delta Flanking Forward	GAGGAAGCTCGAGAAGTGTTT
f2 delta Flanking Reverse	AAAACACCAAGGGCCATCTTT
f2 genotyping Forward	TGCCTTTAGTGATGTTCCTCTG
f2 genotyping Reverse	CTGACAGTCGGGTCTCTGGT
f2 SMRT sequencing Forward	AGGCTGTGAAGGAACAACC
f2 SMRT sequencing Reverse	CTGATCAGACTGGCTCCAC
human *F2* with vector homology Forward	TGGCTAGTTAAGCTTGGTACATGGCTCATGTGCGGGGCC
human *F2* with vector homology Reverse	TTAGGGATAGGCTTACCTTCTCCGAACTGGTCGATCAC
human *F2* Δ45 mutagenesis Forward	CGAGATCGAGAACCCCGATAGCAGCACCACC
human *F2* Δ45 mutagenesis Reverse	CGGGGTTCTCGATCTCGGGCTTGTGGGGGT
human *F2* C138A mutagenesis Forward	GAGACAGGAAGCATCGATCCCCGTCTGCGGCCAGGATCAGG
human *F2* C138A mutagenesis Reverse	AGACGGGGATCGATGCTTCCTGTCTCCGCACGGTGGGGTCGG
Antisense Riboprobe + T7 Forward	TAATACGACTCACTATAGGGGTGATGATGACCGAGGTGGA
Antisense Riboprobe Reverse	CCGACTGTCAGGAGGGAGAC

Chapter III:
A Bioinformatic Pipeline for Identifying Dominant Mutations in the Zebrafish Model Using High Throughput Sequencing

INTRODUCTION

In the early 1990s, zebrafish were firmly established amongst model organisms after the successful completion of parallel mutagenic screens by Christiane Nüsslein-Volhard and Wolfgang Driever, which identified countless developmental mutants.[104,181] At the time, identifying the causative genetic mutation underlying the phenotypic observations was a painstaking, time-intensive process dependent on large numbers of mutant individuals and an incomplete set of genetic markers.[182-185] With the completion of the zebrafish genome and the rapidly declining cost of high-throughput sequencing, many of the limiting factors on the phenotype to genotype pathway for monogenic traits have been removed.[85] In fact, pipelines are now available for high resolution mapping of recessive traits with as few as 5-20 affected animals.[186,187] Unfortunately, there is no automated pipeline for the same efficiency when the causative mutations are dominant in inheritance. Additionally, zebrafish are still a relatively niche field in the genomics space without strong standardization of resources.

As broadly stated in Chapter 1, the Shavit laboratory previously completed a dominant suppressor screen through mutagenesis to rescue lethality of a prothrombotic phenotype. N-ethyl-N-nitrosourea (ENU) was used at a concentration that has previously shown to induce approximately 30 deleterious mutations per sperm, and enough fish were raised to achieve an estimated genomic coverage of 1.5x. From that screen, four separate lines transmitting protective mutations were identified; referred to as *frost*, Q, L and F. More recently, a second mutagenesis experiment was performed. Although the details of the

second mutagenesis are not relevant for this study, it provides a dataset for validating the mutagenic efficiency of ENU. This current study aims to lay out a framework for processing high throughput genomic sequence from zebrafish with the goal of identifying ENU mutations and secondary genetic markers and using these for genetically mapping dominant traits.

Fundamentals of positional cloning

The premise of positional cloning is to identify and isolate the DNA sequence of a disease-causing gene. This traditionally relied on performing linkage analysis using known genetic markers to define a small region of the genome that associated with the phenotype of interest.[188]

As an example, the striped pattern of zebrafish at the *leo* locus on chromosome 1 is a well-known dominant phenotype ($leo^{+/+}$ or $leo^{+/-}$) to the naturally occurring spotted pattern ($leo^{-/-}$).[189] If 100 offspring were produced from a $leo^{+/-}$ and $leo^{-/-}$ mating pair, 50 offspring would be striped and 50 spotted. A heterozygous marker on an unlinked chromosome in the striped parent would be present in roughly 25 fish in each of the striped and spotted populations. In contrast, a linked genetic polymorphism in cis with the *leo* locus on chromosome 1 would be present at a higher rate in the striped population than the spotted population. The frequency of the polymorphism in the striped population would increase the closer the marker is to the *leo* locus to a maximum of 100% for a marker within the *leo* locus.

With a population of 50 striped fish, the ability to narrow down the chromosomal location is related to the meiotic recombination rate surrounding that location. Performing 10,000 simulations of a hypothetical 122 cM long chromosome (the approximate length of chromosome 1) with 1 recombination event per chromosome per meiosis and 2 markers per cM predicts that with 50 informative individuals the region of interest would be on average 5.3 cM with a standard deviation of 3.3 cM. In zebrafish, an average centimorgan (cM) is around 700,000 bp which means the previous example narrows the interval on average to 3.7 Mbp.[185] This simulation comes with many caveats and is meant to illustrate

that the target region is quite large and could still potentially contain dozens of genes.

Classically, once the region of interest is defined, cDNA sequences would be isolated and potential mutations identified. These various mutations would then be tested functionally for phenotypic effects. This study proposes using high-throughput sequencing to identify high-density genomic markers, calculate conserved areas and predict causal mutations from published annotations.

Reference genome

Sequencing of the first zebrafish reference genome began at the Sanger Institute in 2001 (https://www.sanger.ac.uk/data/zebrafish-genome-project/). This project culminated with the release of Zv9 in 2010 and subsequent publication.[85] Maintenance of the genome was then transferred to the Genome Reference Consortium which has released two updates. GRCz10/danRer10 was released in 2014 and the latest version, GRCz11/danRer11 was released in 2017 (https://www.ncbi.nlm.nih.gov/grc). GRCz11 brought continued consolidation of previously unmapped scaffolds with an overall placed scaffold length of 1,345,101,833 bp on 25 chromosomes. Additionally, GRCz11 is the first non-human genome release to incorporate alternate loci scaffolds for the representation of variant sequences (Genbank: 4482478). The use of alternate scaffolds during alignment[190] can help to reduce the false positives that result from the large genetic heterogeneity of zebrafish strains.[191,192] For this reason, we decided to base all work in this study on the GRCz11 reference.

Known genetic variability

Few databases exist for known genetic variability in zebrafish. The largest dataset is the dbSNP database hosted by Ensembl.[87] Unfortunately, the dbSNP dataset is not curated. Furthermore, following the release of version 150, it was determined that dbSNP would no longer be updated for non-human species. However, dbSNP v150 is useful as a starting point. In addition to the dbSNP, an NIH hosted dataset referred to as SNPFisher also provides a variety of known variation across several zebrafish lines.[193] However, SNPFisher is based on a

small sample size and is only available for Zv9. Using a bootstrapping strategy, this work aims to develop a high-quality reference set of variation for future analysis.

METHODS

Computational resources

Processing for the fastq to vcf pipeline was done on the Great Lakes High Performance Computing (HPC) cluster at the University of Michigan. All jobs were processed on standard nodes featuring Intel Xeon Gold 6154 processors with 36 cores and 180 GB requestable RAM per node. Reference files were maintained on user specific HPC home directories. Large files were manipulated on local server scratch space. Permanent large files were maintained on the high capacity "Turbo Research Storage" infrastructure at the University of Michigan. Variant annotation and small-scale file manipulation were performed locally on a Dell Optiplex 780 running Ubuntu 16.04 LTS.

Software Packages

Standard packages were selected to be consistent with the Broad Institute's Genomic Analysis Toolkit (GATK) Best Practices[194] Alignment was done using the burrow-wheelers aligner (bwa v0.7.15; bwa-mem algorithm) and converted to the bam file format using samtools (v1.9; view tool).[195,196] Optical duplicates were identified and marked using the GATK (v4.1.0.0) tool MarkDuplicatesSpark which also coordinate sorts the output. Notably, MarkDuplicateSparks is a Spark implementation for multi-core parallel processing with processing time improvements that scale linearly to 16 local cores. For this reason, major portions of this pipeline were optimized to run on 16 cores. Base qualities were recalibrated using the GATK(v4.1.0.0) tools BQSR and ApplyBQSR. Variants were called using a two-phase approach with the GATK (v4.1.0.0) tools HaplotypeCaller and GenotypeGVCFs. Additional tools included GATK(v4.1.0.0) SelectVariants, SnpSift, SnpEff, Picard(v2.8.1) mergeVCFs, and VariantEffectPredictor (VEP).[197]

Reference data sets

The reference genome utilized was danRer11 acquired from the University of California – Santa Cruz repository (UCSC; https://genome.ucsc.edu) and indexed using bwa. A custom alternate loci index (*.alt) was generated to make use of the bwa mem alternate contig aware alignment feature. dbSNP (v150) was acquired from Ensembl and used as a low-quality set of reference SNPs (single-nucleotide polymorphisms). Gene annotations were additionally acquired from Ensembl. A custom set of regions was defined for analysis based on the following parameters: defined as a gene within Ensembl minus regions marked as repeats > 500 bp by RepeatMasker plus any region defined as an exon by Ensembl.[198] This was done to minimize the processing time spent on the computationally intensive repeat regions and on the uninformative intergenic regions with a high false positive rate and low testability rate. These regions were divided into 16 roughly equal sets of nonoverlapping intervals for subsequent parallel processing. The total area of the intervals is 650,565,514 bp.

ENU Mutagenesis Protocol

Fish were maintained and chemically treated in accordance with a protocol approved by the University of Michigan Animal Care and Use Committee. Briefly, ENU was resuspended in 10 mM acetic acid to a concentration of 85.5 mM and diluted in 10mM sodium phosphate buffer to a final concertation of 3.3 mM. Male fish were treated for 1 hour in 500 mL of solution. Fish were allowed to recover for 1 hour in phosphate buffer and then transferred to fish water with tricaine for another hour followed by a final 3-hour wash in fish water and tricaine. Fish were then transferred to clean fish water and left undisturbed overnight. This procedure was repeated for 3 cycles with 1 week of rest between periods with breeding beginning 1 week after the final treatment. ENU solutions were inactivated post treatment with a sodium thiosulfate solution.

Experimental data sets

Tail DNA was extracted from 85 adult zebrafish and purified using the Qiagen DNeasy Kit. The purified DNA was sent to the University of Michigan Sequencing Core or Novogene for library preparation and sequencing on the Illumina short read platform.

Pipeline

Alignment:

```
bwa mem -t 16 -R "@RG\\tID:${samplename}\\tPL:Illumina\\tSM:
   ${samplename}\\tPU:${flowcell}" \
   ~/REFERENCE/danRer11.fa \
   ${samplename}.PE1.fastq.gz ${samplename}.PE2.fastq.gz \
   |samtools view -hb - -o ~/BAM/${samplename}.bam
```

Duplicate Marking:

```
gatk --java -options "-DsamJdk.compression_level=5"
   MarkDuplicatesSpark \
   -I ~/BAM/${samplename}.bam \
   -O ~/BAM/${samplename}.mark.bam \
   -OBI false

samtools index ~/BAM/${samplename}.mark.bam
```

Base Recalibration:

```
gatk --java-options "Xmx4g" BaseRecalibrator \
   -R ~/REFERENCE/danRer11.fa \
   -I ~/BAM/${samplename}.mark.bam \
   -L ~/INTERVALS/total.intervals \
   -O ~/BASE_RECALIBRATION/${samplename}.table \
   --known-sites ~/REFERENCE/VARIANTS/Shavit_HQ_SNP.vcf.gz \
   --known-sites ~/REFERENCE/VARIANTS/Shavit_HQ_INDEL.vcf.gz

gatk --java-options "Xmx4g" ApplyBQSR \
   -R ~/REFERENCE/danRer11.fa \
   -I ~/BAM/${samplename}.mark.bam \
   --bqsr-recal-file ~/BASE_RECALIBRATION/${samplename}.table \
   -O ~/BAM/${samplename}.recal.bam
```

Variant Calling:

The following command is run once per interval file (16) in parallel:

```
gatk --java-options "Xmx4g" HaplotypeCaller \
   -R ~/REFERENCE/danRer11.fa \
   -I ~/BAM/${samplename}.recal.bam \
   -O ~/GVCF/${samplename}/${samplename}.interval_$.g.vcf.gz \
   -ERC GVCF \
   -L ~/INTERVALS/interval_$_scattered.intervals \
   --native-pair-hmm-threads 1 \
```

```
  -new qual
```

The following command is run to merge the 16 generated files:

```
java -jar ~/JAVA/picard.jar MergeVcfs \
  I=~/GVCF/${samplename}/${samplename}.interval_1.g.vcf.gz \
  ...
  I=~/GVCF/${samplename}/${samplename}.interval_16.g.vcf.gz \
  O=~/GVCF/${samplename}.g.vcf.gz \
  D=~/REFERENCE/danRer11.dict
```

The following commands are run once per chromosome after a gvcf file has been generated for all samples:

```
gatk --java-options "Xmx16g" GenomicsDBImport \
  --genomicsdb-workspace-path ~/GenomicsDB/chr$/database \
  --sample-name-map ${List_of_all_GVCF_files} \
  -L chr$
  --reader-threads 4

gatk --java-options "-Xmx16g" GenotypeGVCFs \
  -R ~/REFERENCE/danRer11.fa \
  -V gendb://~/GenomicsDB/chr$/database \
  -O ~/VCF/danRer11_chr$.vcf.gz \
  -new-qual
```

The following command is run once to merge all chromosomes into one file:

```
java -jar ~/JAVA/picard.jar MergeVcfs \
  I=~/VCF/danRer11_chr1.vcf.gz \
  ...
  I=~/VCF/danRer11_chr25.vcf.gz \
  O=~/VCF/danRer11_final.vcf.gz \
  D=~/REFERENCE/danRer11.dict
```

Commands were submitted to the Great Lakes cluster using the slurm workload manager job scheduler, or slurm for short (https://slurm.schedmd.com), as outlined in Figure 3-1. Alignment and duplicate marking were accomplished with 1 script requesting 16 cores. A second script accomplished the base recalibration using 1 core. A third script was used for intermediate variant calling on 16 cores. Following the processing of each sample through the first 3 scripts, a fourth script requesting 4 cores was run once per chromosome to complete final variant calling and genotyping. The per chromosome files were merged using a final single-core script. The first 3 scripts can be merged into 1 script with no change in results; however, the single core nature of the base recalibration significantly decreases the computational resource utilization efficiency.

Identification and bootstrapping of high-quality variants

A high-quality variant set is required for the proper recalibration of aligned base pairs. The dbSNP database is uncurated and thus does not meet the required standard. To resolve this conflict, a bootstrapping approach was required. Briefly, the data processing pipeline was run without the base quality score recalibration step. Variants that were seen in at least 3 samples were divided into 2 subsets based on whether or not they were also seen in the dbSNP database. Variants that were not seen in the dbSNP database were further evaluated to exclude variants that were only seen in 1 mutagenized family. A hard filter was applied to single-nucleotide variants (SNVs) and INDELs based on prior Broad Institute recommendations with 1 minor change: SNVs were excluded if any of the conditions were not met: QualByDepth (QD) > 2.0, FisherStrand (FS) < 60.0, RMSMappingQuality (MQ) > 40.0, MappingQualityRankSumTest (MQRankSum) > -12.5, ReadPosRankSum > -8.0. QualByDepth was increased to 10.0 for the SNV set not present in the dbSNP database based on the hard filter optimization suggestions. For INDEL variants the following thresholds were required: QualByDepth (QD) > 2.0, FisherStrand (FS) < 200.0, ReadPosRankSum > -20.0. Once again, the QD metric was raised to 10.0 for variants not present in the dbSNP database. These “round 1” high quality variants were then used to recalibrate base scores of the initial BAM files and the pipeline was run a second time. After the second run, variants were filtered with the standard hard filter recommendations and final sets of high-quality variants were generated, *Shavit_HQ_SNP.vcf.gz,* and *Shavit_HQ_INDEL.vcf.gz*. The original BAM files were recalibrated using this set of variants and the pipeline was run a final time to generate a final variant call set for further analysis.

Variant annotation and manipulation

Four tools were used for the manipulation of the final VCF file. GATK SelectVariants was used to both apply hard filters to the data set and to subset the dataset by sample, variant type or location.[194] SnpSift was used to manipulate

variants at a finer scale than GATK.[197] The snpEff tool was used to annotate variants by their effect on coding sequences.[199] Variant Effect Predictor (VEP) was used to annotate missense mutations with SIFT scores.[200,201]

RESULTS

Zebrafish exhibit high genetic heterogeneity

The 85 fish sequenced came from multiple generations but were primarily derived from the *at3* knockout background maintained as an open colony by outcross to hybrid AB x TL wild-type fish. Six fish came from a *f5* knockout background. Focusing only on SNVs, 16,657,177 sites of variation were found after filtration. The absolute number of transitions in each fish ranged from 5,441,112 to 6,857,548 and transversions ranged from 4,822,582 to 6,094,830 (not controlled for coverage). This yielded a Ti/Tv ratio from 1.125 to 1.132 which is consistent with previously published ratios for zebrafish.[202]

The baseline ENU mutation rate is roughly 8 per $1x10^6$ bp

To calculate the rate of transmitted ENU mutations, a mutagenized male was crossed to a control female and 4 offspring were raised to adulthood. These 6 fish were sequenced as part of the total set. Filters were applied to remove any SNVs present in dbSNP v150 or in-house non-mutagenized controls. ENU-induced variants were identified as SNVs that met the following 3 conditions: the parents were wild-type, the parental depth of coverage was 14 or greater, and only 1 of the offspring carried the variant. In total 21,244 variants were identified across the 4 offspring for an average of 5,311 variants per individual. With a total region of ~650 Mbp, the mutation rate is roughly 1 per 122,493 bp. This is approximately 8-fold higher than the expected rate; however, the predominance of transitions and A⇔T transversions is consistent with previously published ENU mutation distribution (Fig. 3-2).[203,204]

When considering exonic mutations, there were 1265 SNVs of which 472 are predicted as synonymous. Of the remaining mutations, there were an average

of 8.5 nonsense mutations, 0.5 lost start codons, 0.66 stop lost, and 4.5 splice site mutations per fish. Additionally, there were another 770 missense mutations found across the 4 offspring. Of these, 234 occurred in unflagged Ensembl transcripts and had predicted deleterious SIFT scores (<0.05). Overall, this suggests approximately 73 predicted deleterious mutations per gamete.

Given the higher than expected rate of mutation, a separate control set of 12 offspring from non-mutagenized parents was processed through an identical protocol. This revealed 25,900 mutations with 2,362 being in the coding region and 774 being non-synonymous. This is 3-fold lower per fish than the ENU samples but much higher than expected for *de novo* mutations. This unexpected mutation rate indicates false positives resulting from technical artifacts.[205]

High throughput sequencing facilitates the mapping of a known modifier

Two *at3*$^{-/-}$ fish from the Q lineage fifth generation were generated carrying a known loss of function deletion in *f2* with the goal of identifying a possible synergy in survival benefit. Eleven *at3*$^{-/-}$ offspring living longer than 1 year were identified from these fish and the 4 parents (2 Q line and 2 control) had their whole genome sequenced. SNVs were filtered for those that were wild-type in the control parents and heterozygous in the Q line parents. Additionally, only SNs with coverage of 10 or more in the control parents were used to minimize missed calls. This subset was considered mappable markers and contained 257,974 SNPs.

To minimize false negatives, only 70,792 SNVs with a minimum depth of coverage of 6 in all samples were considered. Each chromosome had at least 285 SNVs with chromosome 6 having over 10,000 SNVs. For each SNV, the number of heterozygous offspring (out of 11) was calculated. These were then graphed by chromosome. The majority of highly shared SNVs (shared in 10 or more fish) appeared on chromosomes 2, 7 and to a lesser extent 16 (Fig. 3-3).

A frequency histogram of these highly shared SNVs by chromosome position demonstrates improved spatial resolution. Chromosome 2 contains a large shared peak from 10-24 Mbp. Chromosome 16 has a smaller peak near the 5’ end of the chromosome. Finally, as a proof of principle, chromosome 7

demonstrates a conserved peak from 24-45 Mbp. The gap in the middle is consistent with the low genetic diversity surrounding centromeres precluding the existence of informative genetic markers. This peak is notable because it confirms the known survival advantage provided by the introduced *f2* mutation and validates the success of mapping (Fig. 3-4). The presence of the *f2* deletion was confirmed via PCR in all 10 of the predicted carriers from the whole-genome data.

Surprisingly, when the mappable variants were filtered against the dbSNP database and in-house controls, no unique mutations were found on chromosome 2 or 16. This suggests that the enriched regions were either a result of coincidence or a selective pressure for a background variant or haplotype.

DISCUSSION

Improvements in whole genome sequencing have revolutionized biology. Consistent application of these advancements to new model systems is crucial for building a foundation for future studies. This study lays out a framework for the efficient data processing of zebrafish short-read genomic sequencing using the newest reference genome, GRCz11. This study describes a computational pipeline using readily available software, establishes a functional set of reference files and provides a proof of functionality.

Although not discussed, the depth of sequencing coverage is a critical value for successful experiments. The sequencing of individuals here ranged from 5x coverage to 35x average coverage. Using an earlier version of the pipeline, we saw that although average coverage did vary across the genome, this variation scaled with coverage across samples, e.g. areas with 3x average coverage in 10x mean coverage samples had roughly 6x average coverage in the 20x mean coverage samples. We therefore determined that future studies would utilize 20x coverage for high-depth samples and 10x coverage for mapping data.

A second area not covered is the application of alternate contigs. The representation of alternate haplotypes is an area of active research in the genomics field. Currently, only the bwa aligner is able to make use of the alternate

contigs that exist for the human and zebrafish genomes.[190] This has been shown to help in the ultimate accuracy of the called SNVs; however, analyzing variants on alternate contigs is still not streamlined. For this reason, it is important to remember the possibility of variants on alternate contigs even if they are not a part of the primary analysis.

Traditionally, work to identify ENU mutation rates relied on test crosses and highly visible phenotypes. The ratios from these studies have proven fairly consistent. In zebrafish, the ENU protocol has been optimized to induce a 0.5-2 mutations per 1 million base pairs and 30 loss of function mutations per gamete. Our sequencing work finds a rate 4 to 16-fold higher. When looking only at exonic regions, this rate dropped significantly with ~200 nonsynonymous mutations found per fish across an estimated 90 Mbp exome. This suggests that non-coding regions are either more susceptible to the effects of ENU or more susceptible to false positive sequencing artifacts. Interestingly, the characteristics of the induced mutations are remarkably consistent with the previously described ENU profile in mice, validating the reliability of this work.[203,204] Surprisingly, the *de novo* mutation rate was higher than the ratio expected of 2-3 variants per exome but consistent with recent work in which manual validation of variants revealed a majority of *de novo* variants were false positives.[205] This does point to an unidentified source of artifacts such as miscalling due to repeat regions, segmental duplications or errors around insertion/deletion sites. Controlling for the unmutagenized *de novo* rate indicates the actual ENU mutation rate is likely around 30% lower than the sequence-based prediction. When subsetting for predicted deleterious mutants and factoring in the overestimation bias, the rate of 73/gamete is comparable to the 30 expected mutants in mice with a smaller exome.

Introducing a known beneficial mutation into the mutagenized background provided an internal control for genetic mapping. As expected, we were able to identify the *f2* locus, while establishing several guidelines. Although complete At3 deficiency generally leads to lethality by 7 months of age, several generations of selection have possibly improved baseline survival. Due to this, we extended our selection age to 1 year and still believe we have an observable stochastic survival

rate. Therefore, for this and future studies, we assume a conservatively high rate of false positive survivors. In this work, this was done by looking at SNVs conserved in 10 or 11 fish but also by tracking SNVs present in 8 and 9 fish. Dropping the threshold also afforded the inclusion of SNVs lost in some samples due to low coverage.

The lack of an identifiable ENU induced mutant in the Q line mapping regions could be due to multiple reasons. First, as stated, the chromosome 2 and 16 peaks may be coincidental. It's possible that the selective benefit of the *f2* locus removed the selective pressure on additional loci. Second, the other peaks may include naturally beneficial background mutations that will not be identified as novel. This does make identification more difficult but not less informative. Notably, the chromosome 2 peak surrounds the *TFb* locus that has been shown to offer some protection against At3 deficiency (Chapter 5). Finally, the true mutation may be obscured due to technical limitations. As suggested, the true mutation could be on an alternate contig that has not been properly analyzed. Conversely, the mutation may be on the reference scaffold, but filtered out due to low coverage. Finally, we focused on coding mutants and, more broadly, known genes. Unannotated genes and noncoding mutations may explain the survival benefit but would require additional studies to identify and validate.

Whole genome sequencing has provided incredible advances in the identification of disease-causing mutations. By applying the technology to model organisms, it has streamlined the labor-intensive work of forward genetics. As zebrafish are the most accessible vertebrate model for high coverage screens and dominant mutations are often more representation of multigenic disease traits, it seems obvious to create a modern standardized framework. This study demonstrates a state-of-the-art pipeline for zebrafish genomics, establishes necessary reference files and validates the application to mutagenesis and genetic mapping. We believe this will greatly benefit the zebrafish community by increasing the accessibility of the zebrafish genome.

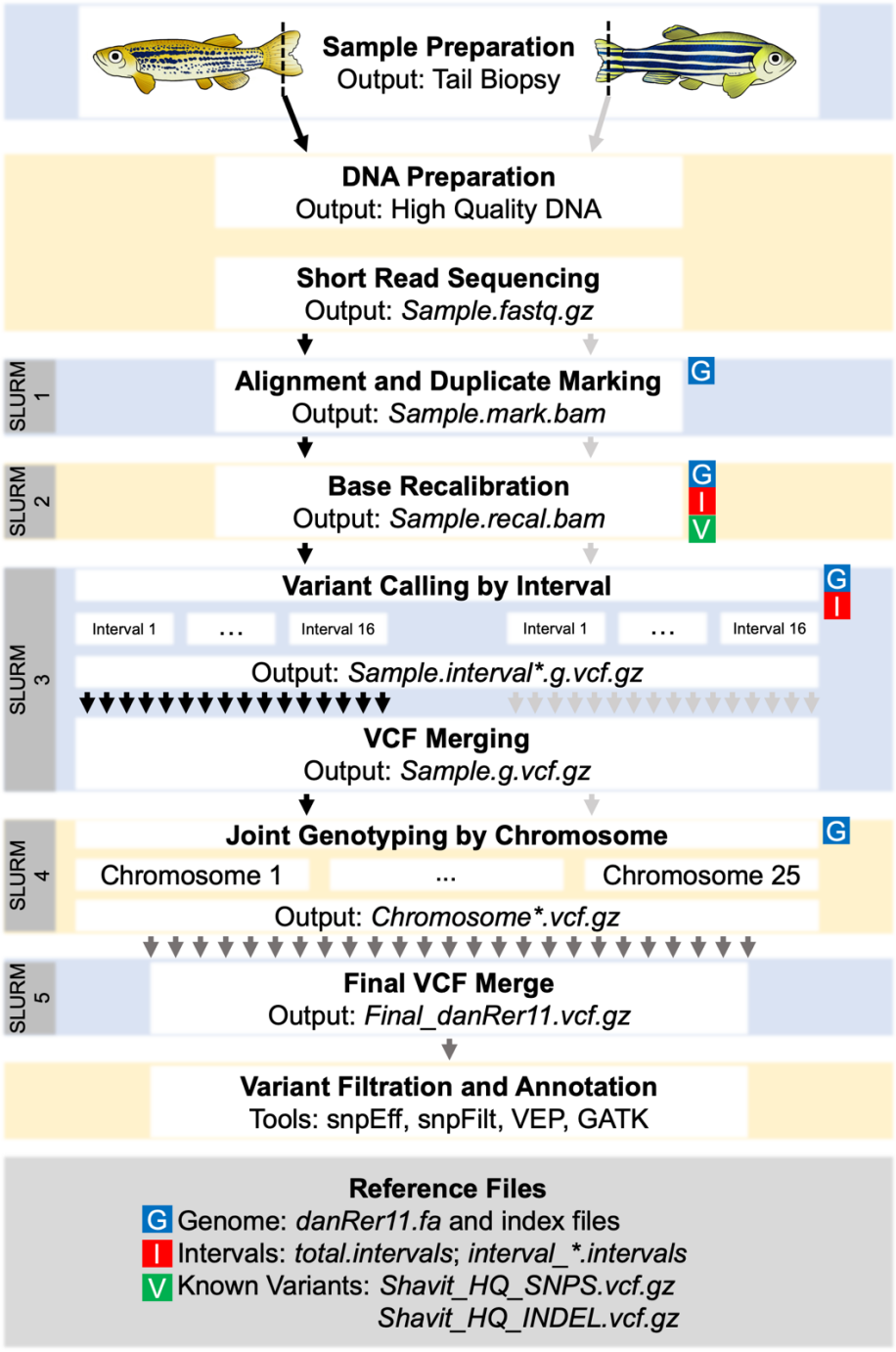


Figure 3-1 Workflow diagram for the processing of zebrafish samples for high-throughput sequencing.

Beginning with tail biopsies, 2 samples are processed in parallel through the first 3 Slurm scripts. Slurm script 4 aggregates all samples and joint genotypes them by chromosome, resulting in 25 intermediate files. These 25 files are optionally merged with script 5 and further analyzed with a series of tools. The reference files listed at the bottom are used throughput the workflow as indicated. VCF merging can optionally use a sequence dictionary not listed.

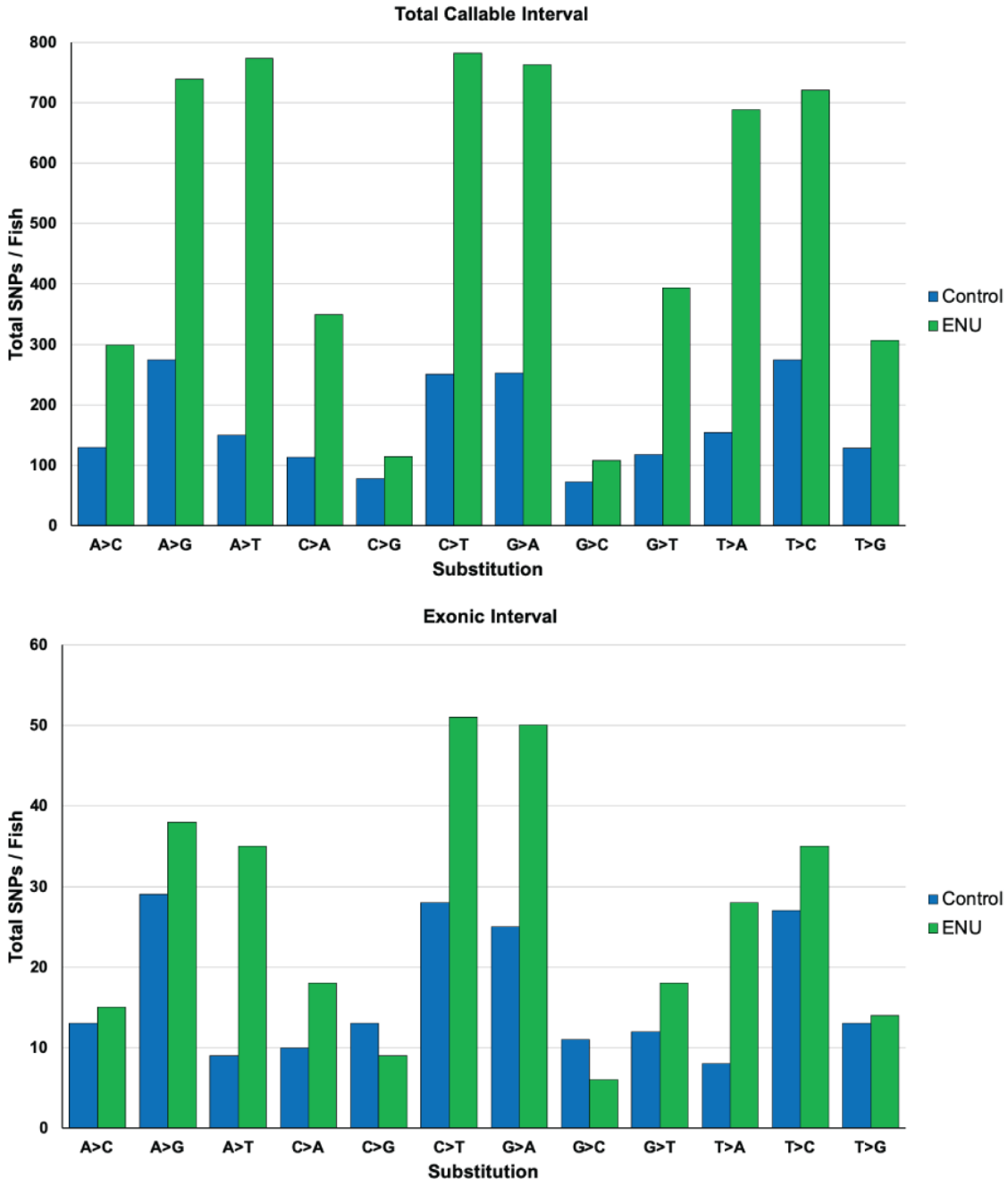


Figure 3-2. ENU mutagenesis creates a robust increase in SNVs in the offspring of chemically-treated males.
(Top) Total interval and (Bottom) exonic subset both demonstrate a drastic increase in the raw number of SNVs per fish. There is a predominance for all transitions (A>G, G>A, C>T, T>C) and A>T and T>A transversions. In contrast, C>G and G>C transversions are not substantially different than the baseline mutation rate.

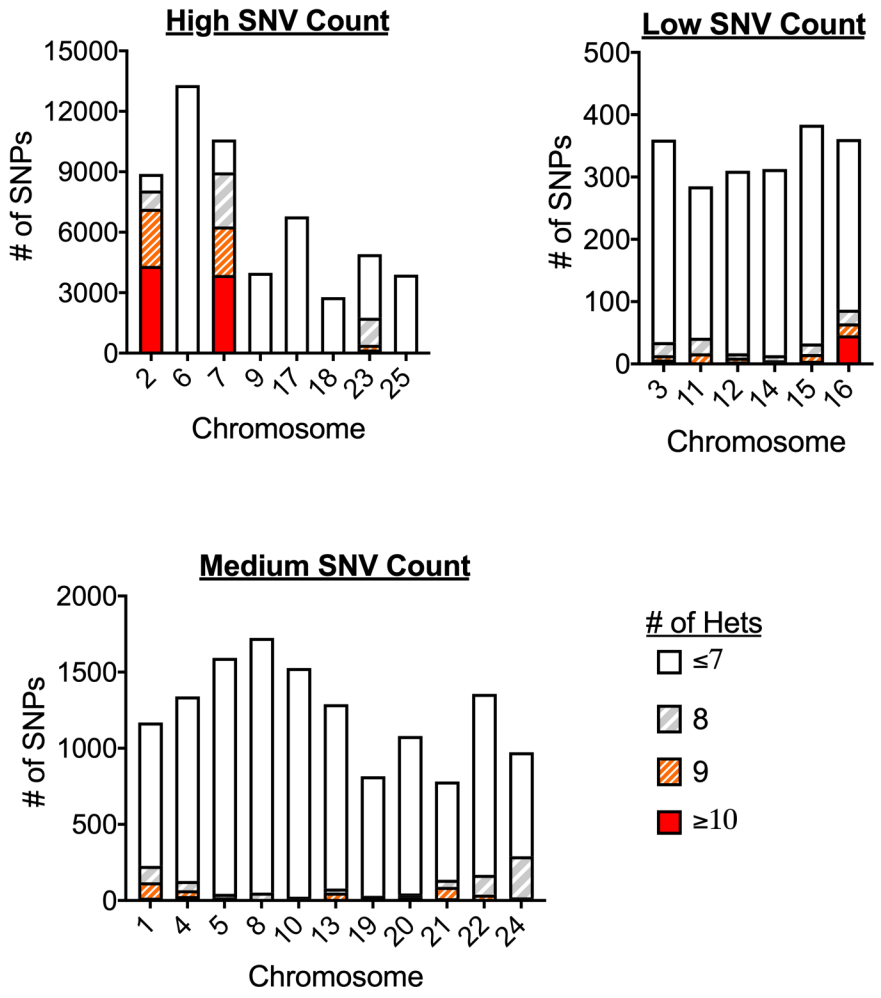


Figure 3-3. Variant sites identified from high-throughput sequencing provide evidence of positively selected loci on chromosomes 2,7 and 16.
70,792 informative SNVs identified as heterozygous in a carrier parent and wild-type in the breeding partner were tracked in 11 offspring. The raw number of SNVs shared by ≥10, 9, 8 or ≤7 fish were plotted by chromosome. Chromosomes were separated into High, Medium and Low groups based on the absolute number of informative variants per chromosomes. The vast majority of highly shared SNVs clustered onto chromosome 2, 7 and to a lesser extent 16.

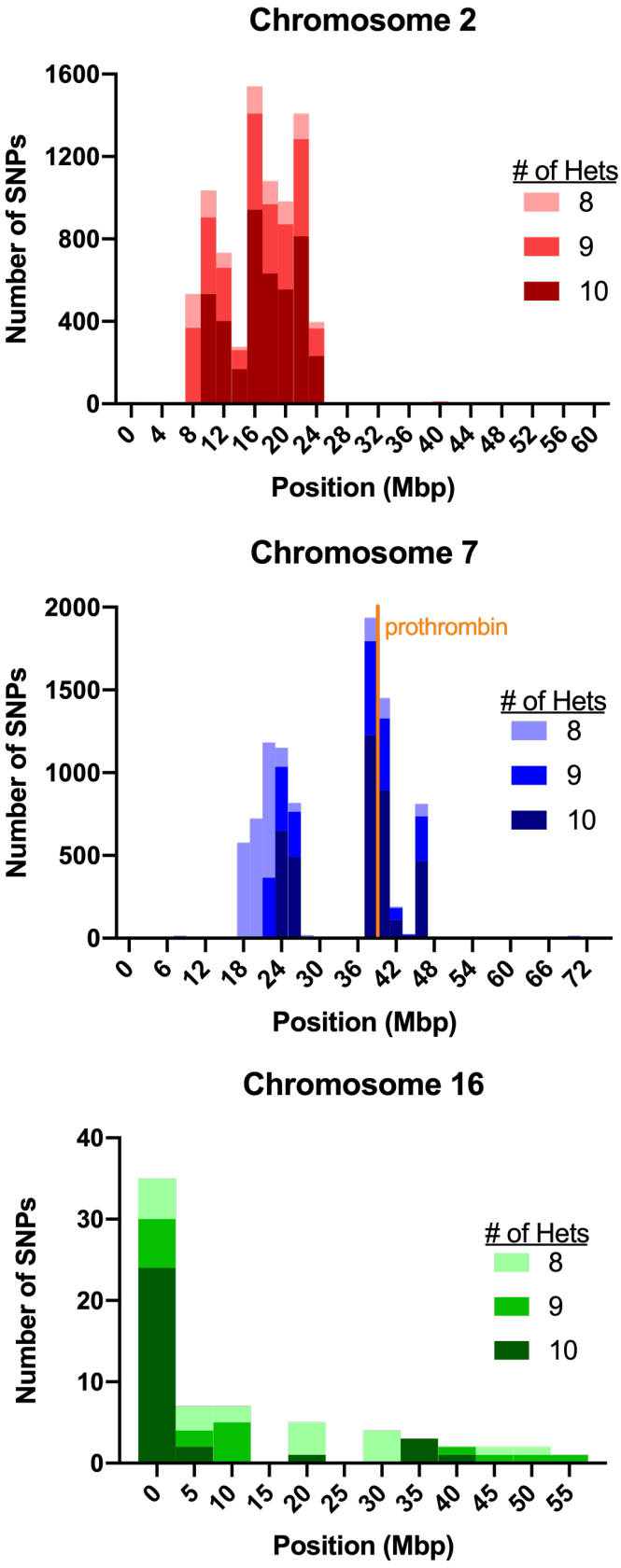


Figure 3-4. SNV distribution across chromosomes confirms a known modifier gene and identifies candidate loci for future analysis.
Based on the positive hits on chromosomes 2, 7 and 16, frequency histograms along the length of the chromosomes were generated. SNVs occurring in 8, 9 or 10 offspring were aggregated into bins of 2 Mbp (chromosomes 2 and 7) or 5 Mbp (chromosome 16) and absolute counts were plotted by the center position of the bin. The majority of SNVs sequestered into a 16 Mbp region on chromosome 2, a 25 Mbp region on chromosome 7 and at the 5' end of chromosome 16. Notably, chromosome 7 lacks informative markers from 30-36 Mbp consistent with a low diversity centromere. The location of the known modifier, *f2*, is indicated in orange on the chromosome 7 map.

Chapter IV:
Dissecting Genetic Modifiers of a Sensitized Zebrafish Model of Thrombosis

Note: Chapter 4 is being prepared for submission as a regular article (Authors: Steven J. Grzegorski, Yang Liu, Catherine Richter, Hasam Madarati, Alexander P. Friedmann, Deepak Paul Y. Kim, Colin A. Kretz, and Jordan A. Shavit).

INTRODUCTION

Venous thromboembolism (VTE) is a disease process defined by the formation of blood clots in deep veins and potential subsequent embolization to the pulmonary vasculature that is estimated to cause 60,000-100,000 American deaths per year.[27,28] Antithrombin III (AT3) is the primary endogenous inhibitor of coagulation, but due to unknown genetic modifiers, deficiency leads to a highly variable lifetime risk of VTE with estimates ranging from 20-80%.[49,50]

The pathology of VTE centers around disruption of the normal processes underlying hemostasis including formation, stabilization and dissolution of blood clots at sites of vascular injury. Under physiologic conditions, vascular injury results in the activation of the common pathway of coagulation.[4] Briefly, the serine protease factor Xa (FXa) along with the cofactor factor Va (FVa) converts the zymogen prothrombin (F2) to the protease thrombin. The now active protease cleaves soluble fibrinogen to insoluble fibrin strands, which stabilize clots to minimize blood loss at the site of injury. Following resolution of the injury, the fibrinolytic system facilitates the breakdown of these clots through proteolytic degradation mediated by the enzyme plasmin.

At each step, environmental and genetic factors contribute to the risk of thrombosis. Genetically, increased procoagulant activity (*F2* 20210G-A, FV Leiden (R506Q)) or decreased anticoagulant activity (AT3 deficiency, protein C

deficiency, and protein S deficiency) can lead to ectopic and/or excess clot formation resulting in intravascular thrombus formation with the risk of subsequent embolization.[30,33,206] From the resolution perspective, plasminogen (*PLG*) deficiency in both humans and mice has been associated with recurrent thrombotic lesions.[207] Clinically, catheter-directed thrombolysis through upstream activation of PLG leads to improved outcomes in some patient populations, but is associated with an increased risk of bleeding.[208]

This entire process occurs in the complicated milieu that is blood and is modified by dozens of known and likely unknown factors. The complex genetic and environmental landscape that modifies VTE risk necessitates a high throughput model for characterizing expected multigenic interactions and for identifying novel modifiers.[127]

Zebrafish are an established model for studying the basic science of hemostasis and thrombosis with a high degree of conservation of the coagulation cascade.[69,91] Previous work demonstrated a conserved physiologic role of AT3 in zebrafish through genome editing.[75] Complete deficiency leads to both a consumptive coagulopathy with spontaneous venous thrombosis in embryos and early mortality due to intracardiac thrombi in adults. This clot burden results in 90% lethality by 7 months of age. With a strong history of forward genetics and a rapidly developing toolkit of reverse genetic technology, the zebrafish offers an excellent *in vivo* test tube for interrogating complex modifiers of thrombosis.

Using the *at3*$^{-/-}$ thrombophilic zebrafish strain, we pursued a two-pronged approach to understanding thrombotic modifiers. We first utilized genome editing technology to knock out suspected modifiers that might enhance or exacerbate survival. We then utilized a non-biased classical mutagenesis protocol to isolate new contributors. In the process we improved our understanding of thrombin biology, uncovered an unexpected risk mitigator, and identified candidate suppressor loci independent of known coagulation factors.

METHODS

Animal Care

Zebrafish were maintained according to protocols approved by the University of Michigan Animal Care and Use Committee. All wild-type fish were a hybrid line generated by crossing AB and TL fish acquired from the Zebrafish International Resource Center. Tris-buffered tricaine methanosulfate (Western Chemical) was used for anesthesia during all live animal experimental manipulation and for euthanasia. Knockout backgrounds of *at3*, *f2* (Δ14) *and f10* were previously described.[74,75,209]

Targeted mutagenesis using genome engineering

Mutagenesis was performed as described previously.[77] Briefly reviewed, the *plg* coding sequence was identified in the zebrafish reference genome (Gene ID: 322691).[85] A CRISPR/Cas9 target site was selected in exon 5 (GGGAGTAC TGCAATATTG AG) using the ZiFit Targeter program (http://zifit.partners.org).[210] The targeting sequence was synthesized (Integrated DNA Techonologies, Coralville, IA) and inserted into the linearized pDR274 vector.[211] The sgRNA was transcribed and purified from this template and 12.5 ng were coinjected with 300 ng of Cas9 mRNA into embryos at the 1-cell stage. Chimeric founders were raised from injected single-cell embryos. The offspring of these founders were screened, and a frameshift allele was identified in a first-generation offspring (*plg*$^{+/-}$*).* This heterozygous fish was outcrossed to establish this knockout line.

Genotyping of mutant offspring

As described in chapter 2, whole embryos or adult fin biopsies were lysed in buffer (10 mM Tris-HCl, pH 8.0, 2 mM EDTA, 2% Triton X-100, 100 μg/mL proteinase K) for 2 hours at 55°C followed by incubation at 95°C for 5 minutes. One microliter of the resulting solution was used as template for gene specific polymerase chain reaction (Table 4-1) and analyzed by gel electrophoresis. For *frost* genotyping, the amplified product was digested for 1 hour with ApaLI prior to

electrophoresis. Additionally, an rhAmp Assay (IDT) was designed to rapidly genotype for the *frost f2* variant (Design ID: CD.GT.RQFP5275.2).

High-throughput sequencing

Tail DNA was extracted from adult zebrafish and purified using the Qiagen DNeasy Kit. The purified DNA was sent to the University of Michigan Sequencing Core or Novogene for library preparation and sequencing on the Illumina short read platform or sent to Otogenetics for whole exome capture and subsequent library preparation and Illumina short read sequencing. Exome data were aligned to the GRCz10 reference genome while whole genome data were aligned to GRCz11. Variants were called using the protocol outlined in Chapter 3.

Laser-induced endothelial injury

Laser injury was performed as previously described in Chapter 2. Briefly, larvae were mounted in 0.8% low melt agarose and placed on an Olympus IX73 with attached Andor Micropoint pulsed-dye laser. 99 pulses were administered to the ventral surface of the posterior cardinal vein 5 somites caudal to the animal pore at 3 dpf. At 2 minutes, larvae were checked for formation of an occlusive thrombus and scored by a blinded observer. Larvae were subsequently extracted from agarose for genotyping.

Human prothrombin protein expression and isolation

The previously published human *F2* expression vector was altered using site directed mutagenesis to create the homologous C507F mutation or a cysteine to alanine mutation at the partner residue C493 (Table 4-1). Vectors were then expressed in HEK293T cells and the protein purified as previously described in Chapter 2.

Prothrombin activation and activity

Assays on prothrombin were done as previously described in Chapter 2. Purified prothrombin was activated using FXa and quantified by formation of

DAPA-thrombin complexes. Thrombin activity was assayed using the substrates S-2238 and fibrinogen to check for active site presence and macromolecular substrate recognition, respectively.

Statistical Analysis

The occlusion data were analyzed using Mann-Whitney *U* testing. Protein expression and activity levels were compared using a two-tailed Student *t* test. Survival was evaluated by log-rank (Mantel-Cox) testing. Significance testing, graphs, and survival curves were made using GraphPad Prism (Graphpad Software, La Jolla, California). P-values ($p<0.05$ or $p<0.0001$) were used to evaluate statistical significance.

RESULTS

Genetic insufficiency of common pathway factors rescues antithrombin deficiency

Complete genetic deficiency of *f2* or *f10* in zebrafish has previously been shown to result in lethality by 3 months of age due to hemorrhage. These genetic knockouts were bred separately to the At3-deficient strain to assess whether attenuation of the common pathway would improve thrombosis-induced lethality. When the fish were genotyped at two months of age, heterozygosity for *f2* but not *f10* provided a significant protective effect versus wild-type in $at3^{-/-}$ fish (Figure 4-1A). Survival of these fish was tracked for 450 days, and both $f2^{+/-}$ and $f10^{+/-}$ demonstrated greater survival than the corresponding wild-type and homozygous siblings (Figure 4-1B-C). This supports that partial common pathway deficiency is able to improve survival in a prothrombotic background.

Plasminogen deficiency relieves consequences of a consumptive coagulopathy

Although $at3^{-/-}$ adults are believed to succumb primarily due to thrombosis, there is a potential risk of bleeding due to the previously published

hypofibrinogenemia in the first week of life. To interrogate the hemodynamic balance, *plg* was genetically ablated in the *at3* deficient zebrafish background using CRISPR-mediated genome engineering. At 2 months of age amongst *at3*$^{-/-}$ animals, a significant decrease of *plg*$^{-/-}$ animals was seen relative to *plg*$^{+/-}$ *and plg*$^{+/+}$ (Figure 4-1A). When tracked for survival, *at3*$^{-/-}$*;plg*$^{+/-}$ *fish* have a modest survival advantage compared to *at3*$^{-/-}$*;plg*$^{+/+}$ fish (Figure 4-1D). Both live significantly longer than *at3*$^{-/-}$*;plg*$^{-/-}$ fish, demonstrating a compounding effect of loss of fibrinolysis. Complete *plg* deficiency on its own does not alter survival (data not shown).

frost, a novel *f2* mutation, is a substitution of a highly conserved cysteine residue

Chemical mutagenesis of *at3*$^{+/-}$ males and subsequent breeding to *at3*$^{+/-}$ females yielded 4030 surviving offspring at the age of genotyping (2-5 months). Of these there were 1248 *at3*$^{+/+}$, 2516 *at3*$^{+/-}$, and 266 *at3*$^{-/-}$. Based on the work in chapter III, the ENU mutagenesis procedure is projected to yield 50-70 coding mutations per spermatogonia. With a Mendelian expectation of an initial ~1250 *at3*$^{-/-}$ offspring, the screen is calculated to have achieved 2.4-3.3x coverage of the genome. 53 of the homozygotes survived to 7 months, 23 were able to produce offspring and 4 lines stably transmitted a survival benefit and established permanent lines (Figure 4-2). The most promising line, *frost,* was further characterized by performing whole-genome sequencing on a single breeding pair and 3 adult offspring. An initial candidate-based approach found 1 missense mutation in a known clotting factor gene resulting in a cysteine to phenylalanine substitution at the highly conserved residue 505 (based on the prothrombin numbering scheme with numbering beginning after the propeptide) of prothrombin (Figure 4-3).

Survival benefit of prothrombin deficiency is recapitulated by *frost*

An initial pool of 47 *at3*$^{-/-}$ fish greater than 2 years of age from the *frost* lineage were genotyped for the prothrombin C504F allele and 45 were found to be carriers. A separate group of fish were genotyped at 2 months and followed for

survival. This group again demonstrated a survival advantage when carrying the C504F allele (Figure 4-4A).

frost rescues embryonic consumption in a time dependent manner

To further confirm the loss of procoagulant function, larvae were generated from *at3*$^{+/-}$; *f2*$^{+/C504F}$ and *at3*$^{+/+}$; *f2* $^{+/\Delta 14}$ parents and assayed using a laser-induced endothelial injury model at 3 dpf. *f2* $^{C504F/\Delta 14}$ larvae were unable to form occlusive thrombi regardless of *at3* genotype (0/4 occluded vs 31/31 control). Knowing that *at3*$^{-/-}$ larvae are unable to form occlusive clots due to a consumptive coagulopathy, clot formation was assayed in an incross of *at3*$^{+/-}$; *f2*$^{+/C504F}$. As expected, loss of *f2* prevented occlusive thrombus formation independent of *at3* status. In *at3*-deficient larvae clot formation was reduced as previously shown. However, loss of 1 copy of *f2* rescued the consumptive coagulopathy in 20% of fish (Figure 4-4B). To confirm and further characterize the rescue, clot formation was analyzed at 3 time points (57 hpf, 78 hpf and 101 hpf) in *at3*$^{-/-}$ larvae with differing *frost* genotypes. At all three time points, roughly 10% of *f2*$^{+/+}$ larvae and 0% of *f2*$^{C504F/C504F}$ larvae demonstrated complete occlusion. In contrast, complete occlusion was seen in 13% of *f2*$^{+/C504F}$ larvae at 57 hpf. This rate increased to 24% at 83 hpf and 74% at 101 hpf (Figure 4-4C). Overall this suggests that a single copy of *frost* is able to restore the ability to form clots in a time dependent manner in *at3*$^{-/-}$ larvae due to a reduction of prothrombin mediated fibrinogen consumption.

Prothrombin secretion is reduced by the *frost* mutation

To interrogate the mechanism behind the reduced prothrombin activity, the orthologous cysteine (C507) to the *frost* variant was mutated in a human prothrombin expression vector. As this mutation results in a free-thiol at C493, two additional prothrombin variants were generated; C493A and C493A/C507F. Expression of wild-type prothrombin in HEK293T cells resulted in secreted protein levels >100 ng/mL in the media. In contrast, all 3 cysteine mutants yielded levels <2 ng/mL in the media (Figure 4-5A). In the corresponding cellular lysate, wild-type prothrombin transfected cells contained approximately 12 ng/mL prothrombin. All

cells transfected with cysteine variants yielded levels lower than 5 ng/mL prothrombin in the lysate (Figure 4-5B). This suggests that loss of the C493/C507 disulfide bond results in decreased secretion due to impaired protein production and/or increased degradation.

Loss of the C493/C507 bond has impaired activation and activity.

The C493A/C507F prothrombin was purified from a stably transfected HEK293 line. Thrombin-DAPA complex formation demonstrated a decreased fluorescence signal change in mutant (1613 ± 91 RFU/uM) versus wild-type (3262 ± 347 RFU/uM), consistent with a disrupted protease domain. Additionally, the initial rate of activation was 2.2-fold slower in the mutant (8.4 ± 1.9 s^{-1}) compared to the wild-type (19.5 ± 1.4 s^{-1}) (Figure 4-5C).

The generated thrombin was assayed for proteolytic activity using S-2238 and fibrinogen as substrates. The amidolytic activity towards S-2238 was significantly decreased in the mutant thrombin (49.2 ± 0.2 mOD/min) compared to wild-type thrombin (82.7 ± 6.5 mOD/min), confirming a partially disrupted protease domain (Figure 4-5D). Furthermore, C493A/C507F failed to cleave fibrinogen or alter the turbidity of the solution (Figure 4-5E-G). Overall, these data suggest a partially disrupted protease domain with reduced activity towards small peptidyl substrates and no activity towards macromolecular substrates like fibrinogen.

Additional suppressor lines do not contain candidate mutations in known coagulation genes

Long lived survivors from the additional lines were processed for high throughput whole genome sequencing. Variants that were unique to each family (absent in control fish, dbSNP v150, SNPFisher and other mutagenesis lines) and present in at least 10 members of that family were annotated and compared to genes in the International Society of Thrombosis and Hemostasis Gold Variant list (Table 4-2, predicted deleterious mutants are bolded).[212] No known coagulation genes were present. Although *f2, f5* and *plg* are all able to rescue the *at3*$^{-/-}$ mortality, only *f2* was found in this screen. This suggests that the screen did not

reach saturation and that several novel thrombosis modifying mutations were introduced by ENU.

DISCUSSION

Although thrombosis is a leading cause of death worldwide, very few of the genetic risk factors are known. The increasing use of big data to develop risk scores for thrombotic risk offers great potential but is limited to observational data.[127,128] We aim to marry the discovery of new genetic factors with the genetic accessibility of the zebrafish model to understand multigenic interactions.

We've previously shown that complete deficiency of *at3* leads to early death by roughly 7 months of age in zebrafish.[75] We incorporated both forward and reverse genetic approaches to try to understand how hemostasis can be rebalanced in the context of pathology. Using a reverse genetics approach, we knocked out 2 members of the common pathway, *f2* and *f10,* that are endogenous targets of At3.[74,209] As expected, insufficiency was able to partially rescue survival, but complete deficiency led to greater mortality independent of *at3* status. This is consistent with current clinical interventions that target F2 and F10 inhibition for treatment of recurrent thrombosis.[213]

We then explored the fibrinolysis system to understand how clot stability affects survival. As expected, complete *plg* deficiency greatly increased mortality in combination with loss of At3. Surprisingly, heterozygosity led to improved survival. This suggests that there may be an underappreciated risk of hemorrhage in the *at3* knockout line which is ameliorated by *plg* deficiency. This may explain reports suggesting the use of combination epsilon-aminocaproic acid (a fibrinolysis inhibitor) and low dose heparin (an anticoagulant) as treatment for cases of disseminated intravascular coagulation with hyperfibrinolysis.[214,215] In addition, median survival of the wild-type controls in the *plg* survival study was around 1 year, which is longer than we've traditionally seen in the *at3* knockout line. This is likely due to the selection over time of background haplotypes that benefit the *at3*

survival in addition to environmental effects such as housing density and temperature that can greatly affect zebrafish growth. Regardless, the use of siblings gives us confidence in the accuracy of the results.

With evidence that dominant suppressors of the early mortality are possible, we completed a chemical mutagenesis screen. Having established 4 lines capable of improving survival, we performed high throughput sequencing and looked for known coagulation genes. Our initial analysis uncovered a *f2* mutation at a conserved cysteine that we refer to as *frost.* This provided a proof-of-principle for the screen and allowed us the opportunity to improve our understanding of thrombin. When looking specifically at mature C493A/C507F thrombin, the site of proteolysis is still maintained, albeit with reduced activity, as evidenced by S-2238 cleavage. In contrast, the ability to cleave fibrinogen is completely lost, suggesting a role in exosite mediated recognition or stabilization of the interaction with macromolecular substrates. This may be explained by lower conformational stability around Tyr 510 due to the lost disulfide bond resulting in decreased Na+ binding and indirectly altering the exosite I.[132,216]

Having rescued *at3*$^{-/-}$ dependent lethality with 2 independent *f2* alleles, we tested the effect on larval consumptive coagulopathy. While our endothelial-injury assay is normally performed at 3 dpf, initial studies yielded mixed results. Preliminary results showed an incompletely penetrant rescue with an unpredictable inheritance pattern. After efforts to identify a second genetic modifier influencing these results were inconclusive, a time course was performed to understand environmental confounding variables. This time course revealed an age-dependent rescue of the consumptive coagulopathy with older fish having an increased penetrance of rescue. This suggests that as the larval liver develops, the produced fibrinogen is consumed rapidly by basal thrombin activity in the absence of antithrombin. However, over time, the amount of fibrinogen produced is able to overcome this consumption and reestablish hemostasis, and partial loss of *f2* enables this to occur at an earlier timepoint. This is further supported by the fact that hemorrhage leads to early mortality around 2-3 months, while the thrombophilic At3 deficient fish survive for greater than 6 months. This difference

implies that a persistent hemorrhagic defect due to consumption is not the primary cause of death for *at3*$^{-/-}$ fish. Therefore, the early inability to clot must be overcome at some point in development in order to facilitate survival beyond 3 months.

More interesting for the long-term implications of this study are the additional 3 lines identified. None of these lines carry a mutation in a known clotting gene. Based on our targeted knockouts, we have shown that loss of *f10* or *plg* can rescue the phenotype. The fact that we did not observe either gene suggests that the mutagenesis did not achieve enough genome coverage to identify all potential modifier genes. It also indicates the presence of suppressor mutations outside of the canonical coagulation cascade in these lines. Interestingly, preliminary data point to multigenic rescue in both the Q and F Lines. While it is unlikely that ENU mutagenesis would induce 2 modifiers in the same gamete, it is not farfetched to imagine preexisting background mutations that benefit *at3* deficient survival as has already been suggested for the chromosome 2 peak in Chapter 3. Although this does make identification more challenging, it also demonstrates the utility of the large horizontal zebrafish pedigree in understanding complex genetic interactions.

Our work continues to demonstrate the use of zebrafish in studying the genetics of human thrombosis. Not only were we able to prove the conservation exhibited by rebalancing hemostasis through *f2* and *f10* modification, but we were able to confirm an unexpected role of the fibrinolytic system. Ultimately, while we have yet to discover a novel thrombotic modifier, we have proof that our screen is technically feasible with strong evidence for the existence of novel modifiers. Further work to map these novel loci offers great potential in advancing the diagnostic and therapeutic opportunities in the millions of people at risk of and affected by thrombosis.

A

Modifier Gene	+/+	+/-	-/-	p-value
prothrombin *(f2)*	3	29	3	<0.001
factor X *(f10)*	22	69	32	0.178
plasminogen *(plg)*	40	74	8	<0.001

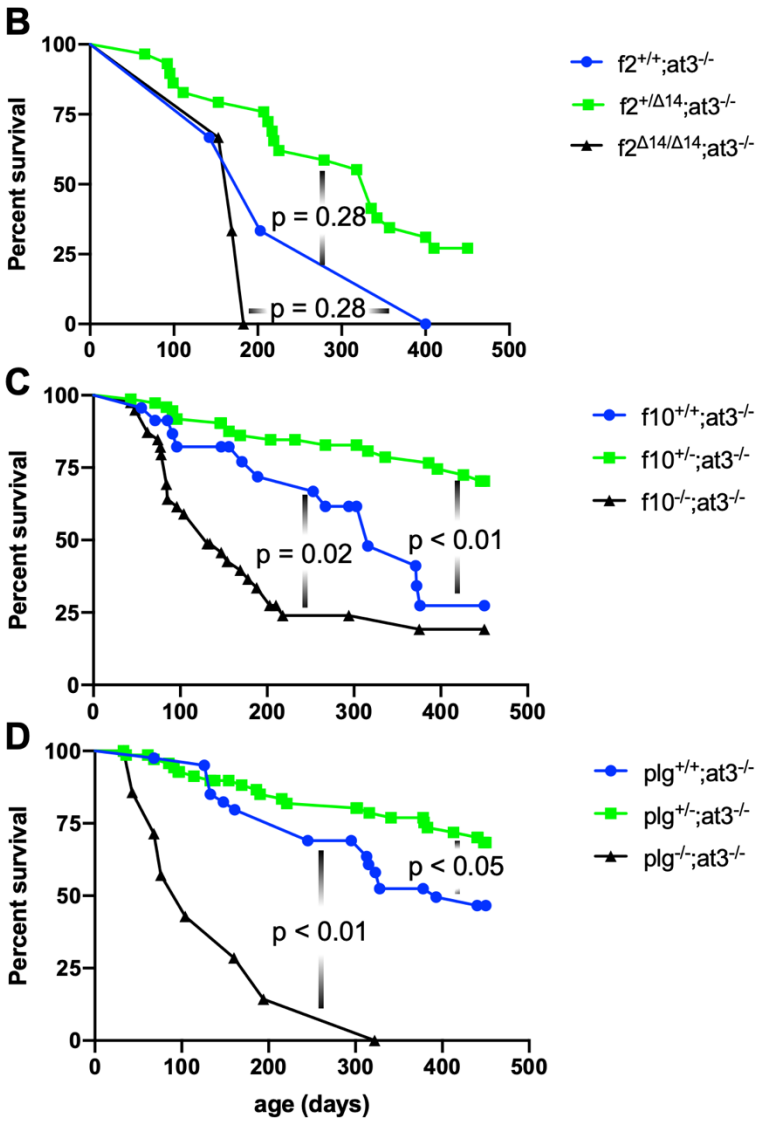


Figure 4-1. *f2, f10 and plg* heterozygosity protect against early lethality due to *at3* deficiency.

(A) At the age of genotyping, fish with complete deficiency of *f2* or *plg,* or wild-type for *f2* are underrepresented in their populations (p-value <0.001; chi-square analysis). Following *f2* (B), *f10* (C) and *plg* (D) populations over time demonstrates that heterozygosity for these genes on an At3-deficient background results in improved survival compared to wild-type siblings, but that homozygosity for the modifier in each family results in increased mortality.

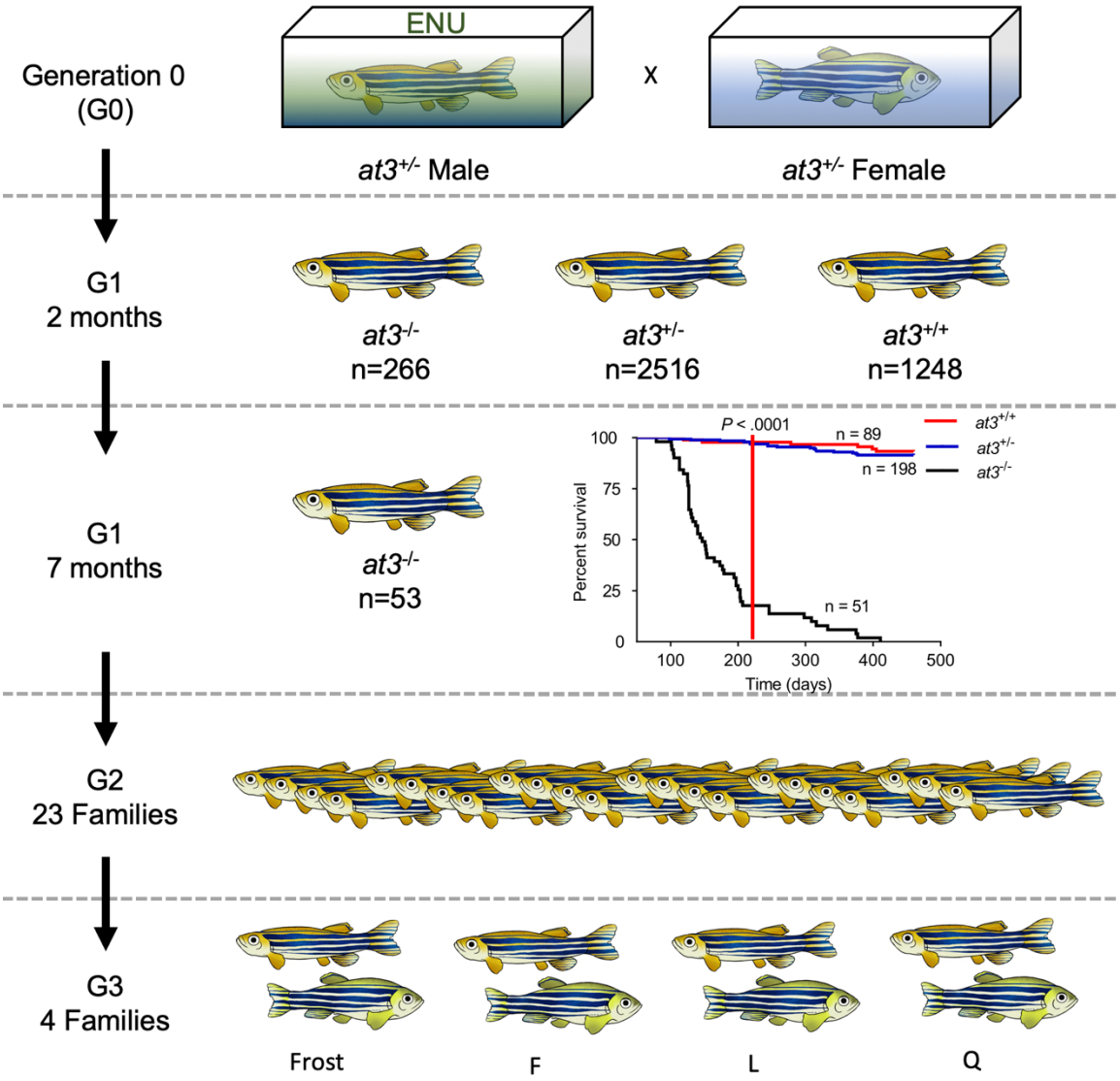


Figure 4-2. Overview of ENU mutagenesis.
Mutagenized *at3*$^{+/-}$ males were crossed to non-mutagenized females and 4030 offspring were produced with a predicted genomic coverage of 2.4-3.3x. At 3 months, *at3* deficient offspring were significantly underrepresented. By 7 months, 53 surviving fish were selected for outcrossing based on the published *at3* survival curve (survival curve inset, red line indicating 7 month time point). 23 G1 fish produced viable offspring that established 23 initial families, and from these 4 families were established with a continuing transmittable survival benefit.

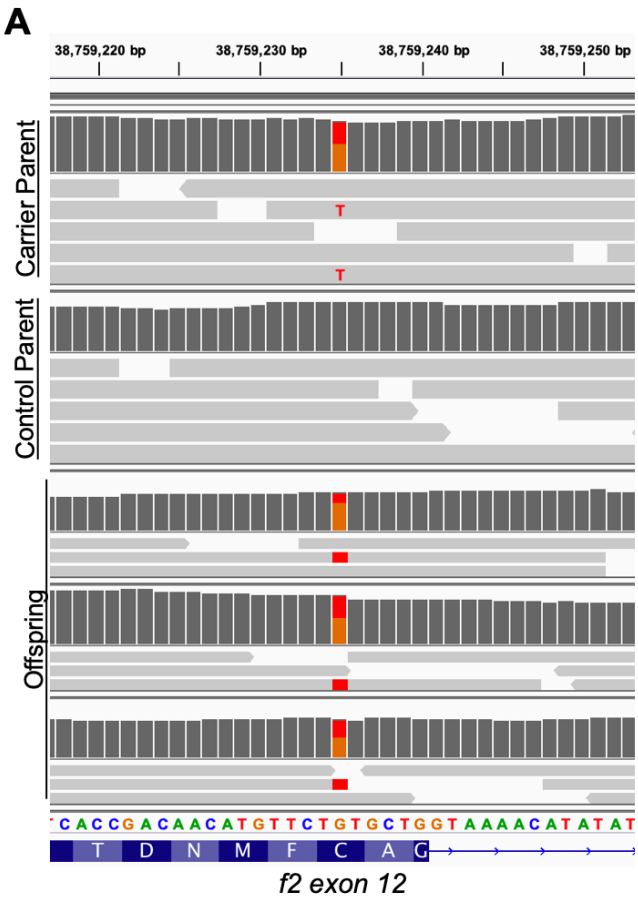


Figure 4-3. Short-read sequencing identifies a missense mutation in the codon for a highly conserved cysteine of *f2*.
(A) Integrative Genomics Viewer display of the short-read sequencing around the identified variant site in exon 12 of *f2*. Carrier and control parents and 3 surviving offspring were sequenced. Dark grey bars display coverage for each sample scaled to 35 reads. Light grey bars correspond to individual reads with single G>T mutations indicated by red labeling. (B) Amino acid sequence alignment of prothrombin residues surrounding the conserved cysteine in multiple organisms. Complete identity is shaded in dark grey and bolded. Homology is shaded in light grey. The cysteine of interest is highlighted in yellow and the disulfide bond annotated above.

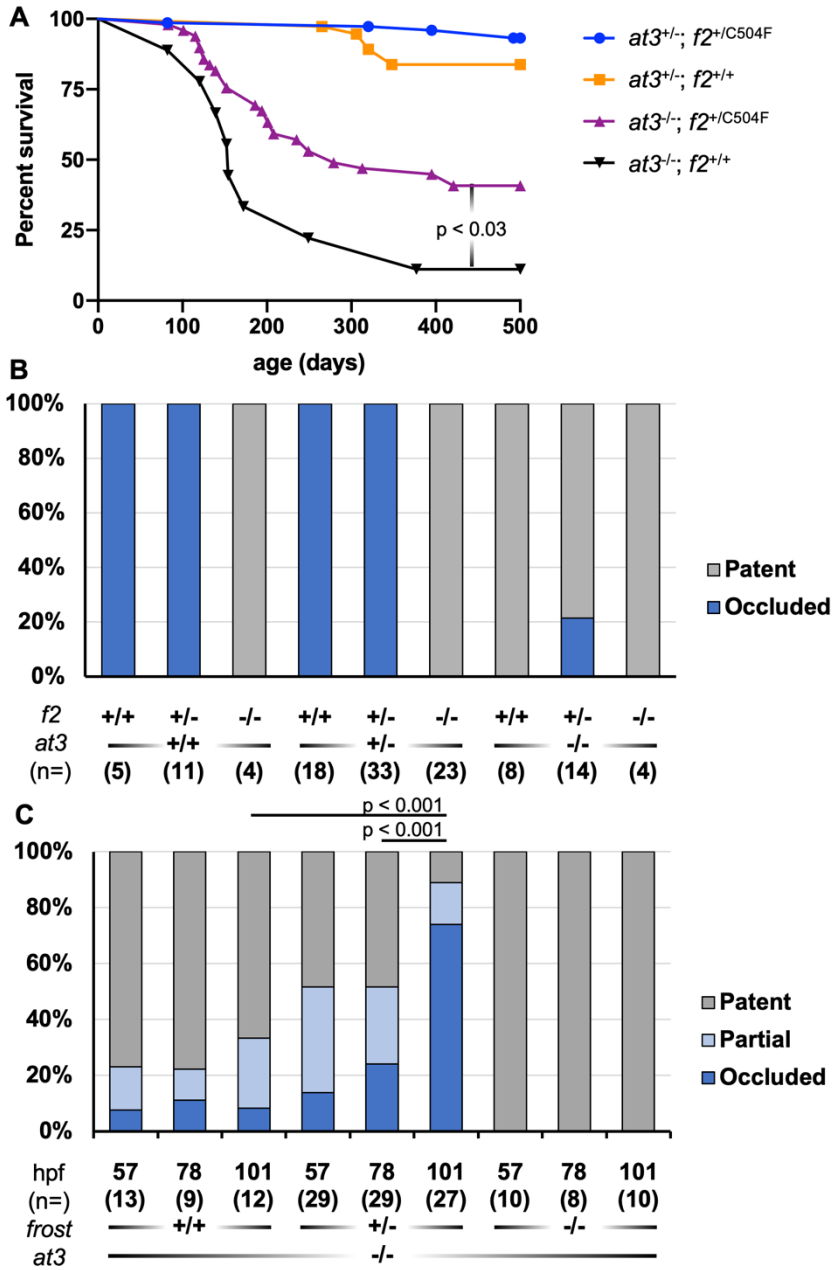


Figure 4-4. The *frost* allele is a loss-of-function variant in *f2* that partially rescues *at3* deficiency.

(A) A single copy of *frost* (purple) significantly improves survival in *at3* deficient fish compared to wild-type (black) but does not return survival to the levels of heterozygous *at3* controls (blue and orange). (B) Loss of either *at3* or *f2* prevents vessel occlusion at 3 dpf following laser-mediated endothelial injury. Heterozygosity for *f2* partially rescues the ability of $at3^{-/-}$ offspring to occlude. (C) A time course of venous laser injury of $at3^{-/-}$ offspring was generated from a $f2^{+/C504F}$ incross. Homozygosity for *frost* prevented clot formation while heterozygosity rescued the low occlusion rate of *at3* deficient larvae in a time dependent manner. Partial occlusion indicates the presence of vascular deposits or transient alterations in turbidity without the formation of an occlusive clot.

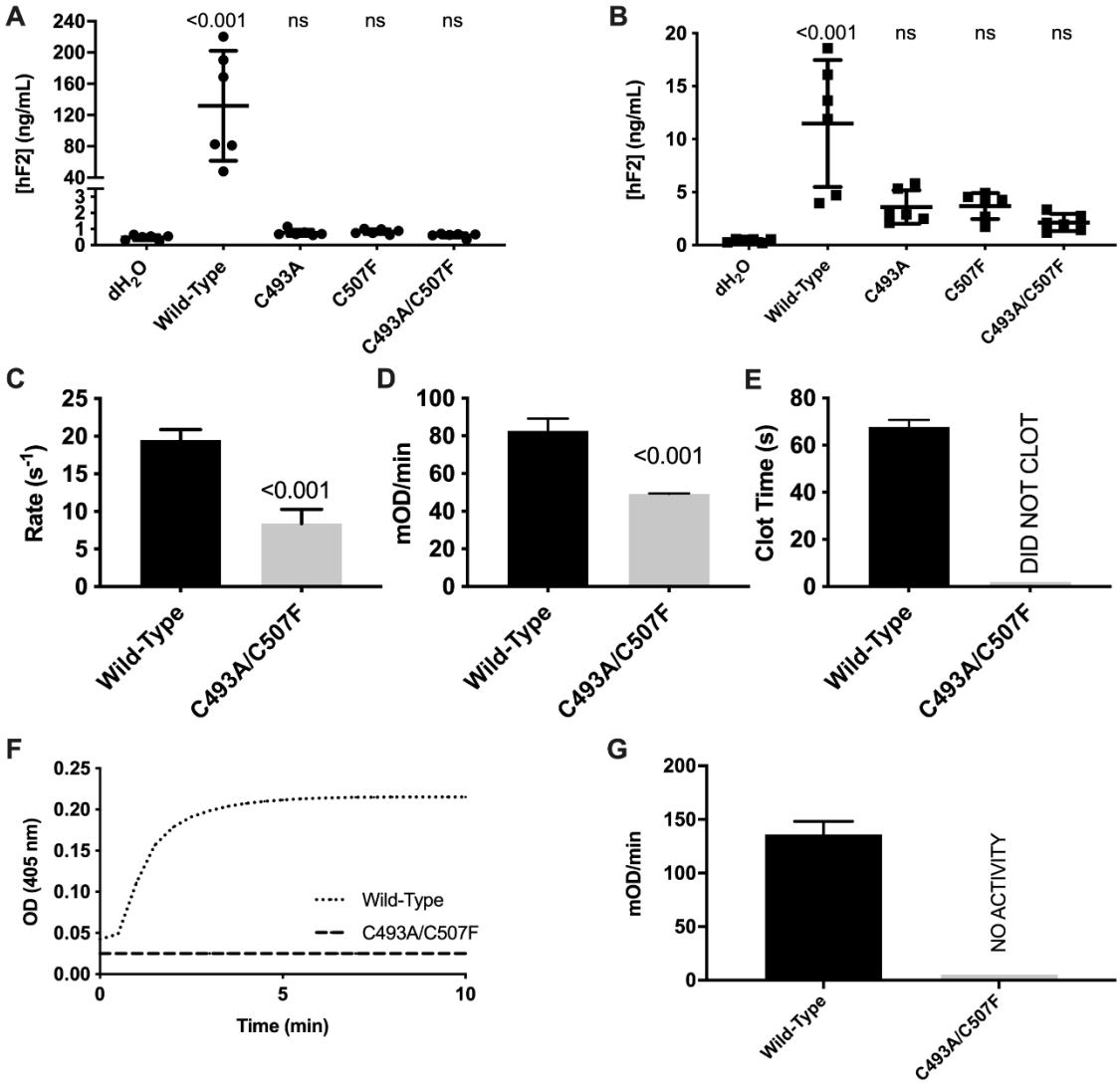


Figure 4-5. C507F mutation leads to reduced biosynthesis, activation, and activity of human prothrombin .

The homologous mutation to *frost* was made in human prothrombin (C507F) and expressed in HEK292T cells. Measurement by ELISA revealed protein levels indistinguishable from water controls in both the media (A) and cell lysate (B). (C) Prothrombin activation by prothrombinase was significantly reduced when measured by DAPA fluorescent change. Fully activated prothrombin had reduced ability to cleave the synthetic substrate S-2238 (D) and was unable to cleave fibrinogen (E). The rate was calculated during steady-state conditions and normalized to enzyme concentration. (F) Turbidity measurements showed no changes in the presence of mutant prothrombin. (G) The rate of clot formation could not be calculated for prothrombin C493A/C507F due to the inability to cleave fibrinogen.

Table 4-1. Oligonucleotide sequences used in methods

Primer Description	Primer Sequence
f2 genotyping Forward	TGCCTTTAGTGATGTTCCTCTG
f2 genotyping *Reverse*	CTGACAGTCGGGTCTCTGGT
frost genotyping Forward	GCAGGTTATAAGGGCCGTGCACTGGC
frost genotyping Reverse	AAGAATATGAAGATATATGTTTTACGTGCA
at3 genotyping Forward	ACACGGAAACGAGGAATCTG
at3 genotyping Reverse	TGCAAAAATTCCTGAGGACAA
human *F2* with vector homology Forward	TGGCTAGTTAAGCTTGGTACATGGCTCATGTGCGGGGCC
human *F2* with vector homology Reverse	TTAGGGATAGGCTTACCTTCTCCGAACTGGTCGATCAC
human *F2* C493A Reverse	GCTGTCCTTGGCCACGGGCCG
human *F2* C507F Reverse	ATAGCCGGCGAAGAACATGTTG

Table 4-2. Candidate suppressor mutations for survival benefit

POSITION	ALLELE	LINE	ENSDART	SYMBOL	AA	SIFT
5:8710859	A	F	00000062957.6	paip1	A21V	0.11
5:9824349	C	F	00000169698.2	zgc:158343	L173W	0.14
5:9824444	C	F	00000169698.2	zgc:158343	D141E	1
5:9824544	C	F	00000169698.2	zgc:158343	S108C	0.74
5:9908000	T	F	00000077245.6	tor4ab	A55T	0.61
POSITION	**ALLELE**	**LINE**	**ENSDART**	**SYMBOL**	**AA**	**SIFT**
2:30180943	**T**	**L**	**00000019149.10**	**rpl7**	**A134T**	**0**
2:30180949	**A**	**L**	**00000019149.10**	**rpl7**	**R132***	**-**
3:14170471	A	L	00000156766.2	si:ch211-108d22.2	G715D	0.23
3:16138950	G	L	00000016616.8	arl5c	K58N	0.14
3:17600113	T	L	00000175485.2	kcnh4a	S896F	0.13
3:17863112	A	L	00000181559.1	BX901973.1	I11N	NA
8:10425103	**A**	**L**	**00000098009.6**	**tbc1d22b**	**V347M**	**0.05**
11:8271311	T	L	00000168253.3	pimr205	V22I	0.74
11:43114391	A	L	00000013642.7	foxg1b	E66K	0.88
20:1216834	**G**	**L**	**00000140650.3**	**ankrd6b**	**L81S**	**0**
20:5084855	**G**	**L**	**00000142518.2**	**ccnk**	**E237G**	**0**
20:6372328	T	L	00000077095.6	trappc8	F944I	0.44
20:16843593	**T**	**L**	**00000027582.9**	**brms1lb**	**Y133***	**-**
20:18414178	T	L	00000049437.6	cdc42bpb	V1344M	0.62
20:19143768	**T**	**L**	**00000151881.2**	**rp1l1b**	**E2367***	**-**
20:26273618	T	L	00000139350.4	syne1a	A3576S	NA
20:30897882	C	L	00000062536.6	hebp2	D117G	0.17
20:31436056	**A**	**L**	**00000046841.9**	**sash1a**	**H1048Y**	**0.01**
20:34382803	**C**	**L**	**00000133593.4**	**swt1**	**I301V**	**0**
20:52404591	**T**	**L**	**00000166651.2**	**arhgap39**	**P290S**	**0**
21:122571	G	L	00000160005.2	si:ch1073-398f15.1	L506P	0.12
24:10906980	**G**	**L**	**00000145466.3**	**asap1b**	**I1086T**	**0**
24:19118173	A	L	00000106186.6	prex2	L1247M	0.06
POSITION	**ALLELE**	**LINE**	**ENSDART**	**SYMBOL**	**AA**	**SIFT**
2:10350190	**T**	**Q**	**00000016369.8**	**wls**	**E458K**	**0.02**
2:10350196	A	Q	00000016369.8	wls	V456L	0.07
2:10354682	G	Q	00000016369.8	wls	D293H	0.07
2:10354691	A	Q	00000016369.8	wls	V290L	1
14:32015583	**A**	**Q**	**00000105761.4**	**zic3**	**V262F**	**0**
14:36332517	**A**	**Q**	**00000077823.6**	**lrit3a**	**G290***	**-**
15:47492848	**C**	**Q**	**00000122372.3**	**sik3**	**K4T**	**0.01**
23:43467836	**C**	**Q**	**00000169726.2**	**e2f1**	**D418G**	**0**
23:46115544	G	Q	00000030004.9	CABZ01071903.1	A89P	0.2
23:46115549	T	Q	00000030004.9	CABZ01071903.1	L87Q	0.04

Chapter V:

Genetic Duplication of Tissue Factor in Teleost Reveals Partial Subfunctionalization

Note: Chapter 5 is being prepared for submission as a regular article (Authors: Steven J. Grzegorski, Queena Zhou, Divyani Paul, Jim Morrissey and Jordan A. Shavit).

INTRODUCTION

Preventing pathological blood clotting in the setting of normal physiology requires a well-regulated system to minimize and inhibit sites of ectopic procoagulant activity. One primary mechanism to prevent activation of the coagulation cascade is the spatial separation of the potent activator tissue factor (TF) from the plasma localized procoagulant proteins, including the protease factor VII (FVII). Under normal conditions, TF is a transmembrane glycoprotein expressed on the surface of cells surrounding the vasculature. Following disruption of the endothelial layer, TF is exposed and able to act as a receptor for circulating FVII.[217] Once bound, the TF/FVII complex serves as an initiator of coagulation via the activation of factor X and factor IX.[218,219] Outside of vascular damage, several conditions can trigger the intravascular expression of TF. During infection, circulating proinflammatory cytokines are known to trigger the expression of TF on the surface of mononuclear and endothelial cells.[220] Separately, many neoplasms express high levels of TF which is subsequently incorporated into shed microvesicles that enter circulation.[221]

Beyond its procoagulant activity, the TF/FVII complex has a role in cell signaling. The complex has been shown to cleave several of the g-protein coupled protease activated receptors (PARs).[222] This PAR-mediated signaling may be responsible for the observed angiogenic role that TF plays in cancer.[223,224] Additionally, it has been shown that TF/FVII is able to interact with β_1 integrins to

modulate intracellular signaling.[225,226] Understanding the regulatory controls over TF expression and the secondary roles of TF, including crosstalk with inflammation, could be crucial to minimizing the thrombotic risk of various pathologies.

Several global and tissue specific knockouts of TF in mice have been generated. Complete TF deficiency leads to lethality by E6.5-E8.5 making it the most severe of procoagulant knockouts.[227,228] In contrast, loss of FVII leads to neonatal lethality. This differential phenotype is thought to be due to the ability of small amounts of FVII to leak from maternal to embryonic circulation, although this has not been proven.[229,230] Tissue specific knockouts of a floxed *TF* allele yield a variety of phenotypes. Notably, cardiomyocyte (Mlc2v) specific loss results in cardiac fibrosis.[231] Myeloid (LysM) or endothelial and hematopoietic loss (Tie2) specific knockout leads to reduced lipopolysaccharide induced coagulation.[232] Neuroectodermal (Nestin) targeted loss causes an increase in intracerebral hemorrhage after collagenase infusion.[233] Finally, keratinocyte promoter (K14) driven knockout results in delayed wound healing.[220]

Due to the early lethality of complete deficiency, a transgenic approach was used to express a human *TF* minigene in $TF^{-/-}$ mice. This system resulted in levels at 1-2% of normal TF. Although the minigene rescues survival into adulthood, low TF mice exhibit prolonged bleeding after hemostatic challenge and have increased rates of hemorrhage.[234] Interestingly, while human TF is able to activate coagulation in mice, mouse TF is unable to effectively activate coagulation in human plasma. This is likely due to a species-specific interaction between TF and FVII as validated by the fact that mouse TF/FVII complexes are able to fully activate human FX (correspondence with JH Morissey). This demonstrates that although the function of TF is highly conserved, the primary amino acid sequence is not. Therefore, it is important to develop an accessible model that does not rely on the complications of xenogeneic proteins.

Zebrafish are a small aquatic vertebrate with a hemostatic system that is highly conserved with humans. A single mating pair can produce hundreds of offspring each week with transparent, external development. Circulating blood is visible by 36 hours of life and assays for assessing coagulation *in vivo* are readily

performed between 72 and 120 hours post fertilization (hpf).[70,150] Importantly, zebrafish are able to survive severe hemostatic defects and develop into adulthood.[73-75,77,209] Notably, zebrafish have two copies of *TF* referred to as *TFa* and *TFb* due to an ancestral whole genome duplication event. Previously performed antisense knockdown of *TFb* suggested a role in early vascular development, consistent with mouse models.[96] The function of TFa is currently unknown, and to date no studies have been done to completely block TF activity in zebrafish.

TF deficiency has never been seen in humans and complete deficiency is difficult to study in mice. The genetic duplication of *TF* in zebrafish provides an opportunity to understand the endogenous roles of TF due to the potential for neofunctionalization or subfunctionalization. This study makes use of CRISPR/Cas9 to generate loss-of-function alleles in both copies of *TF*. Loss of both is compatible with development but leads to early lethality. A single copy of either *TFa* or *TFb* is sufficient to rescue survival. Interestingly, this study shows that TFa has higher procoagulant activity and plays a predominant role in venous hemostasis. In contrast, TFb has a greater role in arterial coagulation and this difference appears to be mediated by the intrinsic pathway. Understanding the spatial and temporal control of *TFa* and *TFb* and increasing our understanding of their functional differences could prove valuable for modifying thrombotic pathologies.

MATERIALS AND METHODS

Animal Care

Zebrafish were maintained according to protocols approved by the University of Michigan Animal Care and Use Committee. All wild-type fish were a hybrid line generated by crossing AB and TL fish acquired from the Zebrafish International Resource Center. Tris-buffered tricaine methanosulfate (Western Chemical) was used for anesthesia during all live animal experimental manipulation and for euthanasia.

Targeted mutagenesis using genome engineering

The ChopChop2 server was used to identify predicted high efficiency target sites for CRISPR/Cas9 engineering.[235] Three guides each were selected for TFa and TFb. DNA oligos for the target sequence were synthesized with homology to a second common backbone oligo (Table 5-1). The pieces were assembled via PCR and purified using phenol:chloroform. The sgRNA was synthesized using the Maxiscript T7 kit and purified via ethanol/ammonium acetate precipitation. 600 ng gRNA were mixed with 0.5 uL EnGen Cas9 protein (New England BioLabs) at a 1.5:1 molar ratio in 300 mM KCl. The solution was incubated at 37°C for 10 minutes and 2 nL of the solution was injected into the yolk of single cell embryos. Test embryos were lysed at 24 hpf and flanking primers were used to amplify the target site. High resolution gel electrophoresis was performed to validate efficient cleavage. Larvae injected with 2 known successful guides were raised, bred, and their offspring screened for single large deletions. Lines were established from single heterozygous founders.

Genotyping of mutant offspring

As described in chapter 2, whole embryos or adult fin biopsies were lysed in buffer (10 mM Tris-HCl, pH 8.0, 2 mM EDTA, 2% Triton X-100, 100 µg/mL proteinase K) for 2 hours at 55°C followed by incubation at 95°C for 5 minutes. One microliter of the resulting solution was used as template for gene specific polymerase chain reaction (Table 5-1) and analyzed by gel electrophoresis.

Real time quantitative polymerase chain reaction

RT-qPCR was performed as described in chapter 2. Briefly, total RNA was extracted from 4 dpf larvae using the RNeasy Mini Plus kit (Qiagen) and transcribed using oligo$(dT)_{12\text{-}18}$ primer and Superscript III (Invitrogen). Three equal pools of whole larvae were used per genotype. The resulting cDNA was used as template for qPCR (StepOnePlus, Applied Biosystems) using gene specific PrimeTime© probe-based qPCR Assays (Integrated DNA Technologies) (Table 5-

1). The expression levels of *TFa and TFb* were normalized to the *actb2* gene and significance analyzed using the double delta Ct method as described.[158]

Laser-induced endothelial injury

Laser injury was performed as previously described in Chapter 2. Briefly, larvae were mounted in 0.8% low melt agarose and placed on an Olympus IX73 with attached Andor Micropoint pulsed-dye laser. 99 pulses were administered to the ventral surface of the posterior cardinal vein 5 somites caudal to the anal pore at 2 or 3 dpf or 3 somites caudal to the anal pore on the dorsal surface of the cardinal artery at 2, 3 or 5 dpf. At 2 minutes, larvae were checked for formation of an occlusive thrombus and scored as either occluded or non-occluded by a blinded observer prior to genotyping.

Membrane bound recombinant TF production

The nucleic acid sequences for zebrafish *TFa* and *TFb* were identified and confirmed. The propeptides and cytoplasmic tails of the sequences were removed and replaced with a periplasmic localization signal and an HPC4 tag. The sequence was codon optimized for *E. coli* expression, synthesized in gBlocks (IDT) and subcloned into pJH698 in place of the previously published human TF sequence (Table 5-2). The validated plasmids were transformed into BL21(DE3) competent cells. Single colonies were grown up in minimal MDG media to OD_{600} = 6. Cells were pelleted, resuspended in BYE to an OD_{600} of 4.5 and shaken at 25° for 2 hours. IPTG was added to a concentration of 100 uM and allowed to shake for another 4 hours. The cells were then pelleted and stored frozen at -80° until the next step.

Frozen pellets were resuspended in in lysis buffer (30mM HEPES, 50 mM NaCl, 0.1% NaN_3, 1% Triton-X 100 and 0.1% lysozyme) and briefly sonicated on ice. The bottle was rocked for 15 minutes at room temperature and then $CaCl_2$ was added to a final concentration of 2 mM and incubated on ice for 30 minutes. The remaining debris was pelleted, and the supernatant transferred to a new container. Equilibrated Q-Sepharose beads were added to the supernatant (1g / 10 mL) and rocked for 1 hour at 4°. Supernatant was then passed through an equilibrated

HPC4 affinity column. The column was washed (30mM HEPES, 1% triton, 1 M NaCl, 1 mM $CaCl_2$, and 0.1% NaN_3) and eluted (30mM HEPES, 1% triton, 100 mM NaCl, 5 mM EDTA, and 0.1% NaN_3). The resulting eluate was dialyzed against HBS-Triton (30mM HEPES, 1% triton, 100 mM NaCl, and 0.1% NaN_3).

The purified TF was incorporated into 80:20 phosphotidylcholine:phosphotidylserine liposomes at a ratio of 15000:1 using deoxycholate and Biobeads SM2 as previously described.

Ex vivo clotting assays

Lake trout blood was collected from a venous puncture and anticoagulated using 9:1 plasma:3.2% sodium citrate and centrifuged at 1500g for 15 minutes to isolate thrombocyte poor plasma. Plasma, TF liposomes and 25 mM $CaCl_2$ were mixed at a 1:1:1 ratio and placed on a rocker at 25°C. The tubes were observed through repeated tilting until the solution became solid and a clot was apparent. Clots were scored by a blinded observer for 5 minutes and then checked a final time at 15 minutes.

Chemical treatment

Larvae were placed in a light protected plate and stock solutions of epinephrine and hydrocortisone in DMSO were added to a final concentration of 25 uM epinephrine, 0.01% hydrocortisone and 0.1% DMSO and incubated in the dark at 28.5°C.

Statistical Analysis

The occlusion data and *ex vivo* clotting assays were analyzed using Mann-Whitney *U* or binomial proportion tests. Significance testing, graphs, and survival curves were made using GraphPad Prism (Graphpad Software, La Jolla, California).

RESULTS

CRISPR/Cas9 facilitates the generation of null alleles for both *TFa* and *TFb*

Large deletions were made in each copy of TF by inducing 2 double stranded breaks in each coding region using CRISRP/Cas9. Stably transmitted deletions from exon 1 to exon 3 in *TFa* and from exon 4 to exon 5 in *TFb* were identified in the offspring of injected fish (Figure 5-1A-B). The selected *TFb* allele contained an inverted portion of a neighboring intron inserted at the site of editing. Quantitative polymerase chain reaction (qPCR) at 4 dpf demonstrated that homozygosity of *TFa* resulted in reduction of *TFa* mRNA consistent with nonsense mediated decay. Conversely, qPCR of homozygous *TFb* larvae shows constant levels of residual transcript (Figure 5-1C). When the full length *TFb* transcript was amplified and analyzed via agarose gel electrophoresis, multiple smaller transcripts were identified (Figure 5-1D). The most prevalent involved splicing from exon 2 to exon 6 resulting in a stable transcript generating a 61 amino acid truncated protein. Overall, both alleles appear to result in complete loss of functional TF protein.

A single copy of TF is both necessary and sufficient for long term survival

A large group of offspring was generated from a double heterozygous incross to determine the influence of TF on survival. For simplicity, in this thesis when discussing both *TF* genotypes wild-type alleles will be referred to as “A” and “B” while null-alleles will be referred to as “a” and “b,” e.g., double heterozygotes are *Aa/Bb*. At the age of genotyping (8 weeks), there was no deviation from the expected Mendelian ratios. However, over the next week, all 9 double homozygous fish succumbed to signs of overt hemorrhage (Figure 5-2A). Notably, the remaining genotypes, including *Aa/bb* and *aa/Bb* saw equivalent survival over the following year. To determine whether these results were due to the stress of genotyping, a second group of 149 fish was raised to 6 months of age before genotyping (Figure 5-2A). No double homozygous fish were identified in this clutch consistent with complete TF deficiency being incompatible with long-term survival. A final group of fish was genotyped 1 month of age and the 7 identified homozygotes were tracked. 6/7 succumbed to hemorrhage by 2 months of age while the last fish was

able to reach sexual maturity before dying at 120 days of age (Figure 5-2B). The cause of death across clutches appeared to be overt hemorrhage into the head or pericardial space (Figure 5-2C) To determine whether maternal contribution of TFa or TFb could be responsible for survival through early development, groups of offspring were generated from *Aa/bb* males crossed to *aa/Bb* females and from *Aa/bb* females crossed to *aa/Bb* males. At 1 month of age, all possible genotypes were observed in the assayed offspring, suggesting the ability to complete development in the complete absence of TF (data not shown).

Loss of TF does not alter vascular development

Given TF's suspected role in angiogenesis, two transgenic lines were utilized to look at endothelial development and vascular perfusion. The *gata1a:DsRed*[236] transgenic line displays red fluorescent erythrocytes, while the *kdrl:eGFP*[237] transgenic labels vascular endothelial cells with green fluorescent protein. At 30 hpf, the number of vascular sprouts and the growth progress were scored prior to genotyping and no significant difference was observed across genotypes (Figure 5-3A-C). At 54 hpf, there were no obvious defects in endothelial development or in erythrocyte perfusion in double homozygous larvae (Figure 5-3D). Finally, through the first week of development, there were no obvious morphological defects under bright field examination (data not shown). Overall, this suggests that TF is not necessary for the differentiation, proliferation or function of the vascular endothelium.

TFa plays a dominant role in venous hemostasis

A laser-induced endothelial injury assay was used to trigger clot formation in the venous system at 3 dpf in *TF* mutants. Loss of TFb alone did not lead to any alterations in the time to occlusion (TTO). Loss of TFa alone resulted in a mild delay in TTO from ~20 to ~30 seconds. This delay was exacerbated in a dose dependent manner with loss of TFb. Complete deficiency of TFa and TFb resulted in the inability to form occlusive thrombi (Figure 5-4A). At 5 dpf, a binary (clot vs. no clot) version of the injury assay was used to assess developmental changes.

At this stage, the ability to form occlusive thrombi again appeared to be dependent on the amount of TF activity present with loss of TFa or TFb causing a decrease in the ability to form occlusive thrombi. Notably, TFa still appears to play a predominant role (Figure 5-4B). This suggests that both TFa and TFb are present surrounding the venous bed and that TTO is dependent on the amount of total TF procoagulant activity present, with TFa providing enough activity to fully compensate in the absence of TFb at 3 dpf but not vice versa.

TFb plays a dominant role in arterial hemostasis

To understand differences between arterial and venous circulation, the laser injury assay was performed on the caudal artery at 5 dpf. Unlike the venous system, loss of TFa alone did not appear to alter the TTO. Surprisingly, while loss of TFb did not alter the TTO in those that clotted, TFb deficiency did lead to an increased failure to occlude with phenotypic penetrance varying from 30% to 90% in various groups of offspring. Loss of both TFs did result in a failure to form occlusive thrombi (Figure 5-4C).

At 5 dpf thrombocytes play a major role in arterial occlusion. To remove their influence in the clotting process, arterial injury was performed at 2 dpf, prior to the robust expansion of thrombocytes in circulation. In contrast to previous experiments, this one was done in 2 steps. First, arterial injury was performed on all fish. Subsequently, all fish that had not occluded within 120 seconds were injured in the venous system. Once again, the ability to occlude in the arterial system appeared to be entirely dependent on the presence of TFb with occlusions forming in greater than 95% of *TFb* heterozygotes. Larvae containing only TFa (*Aa/bb)* were able to form occlusive thrombi only 10% of the time. When the subsequent venous injury was performed on larvae that failed to occlude in response to arterial injury at 2 dpf (predominately *Aa/bb* and *Aa/bb* larvae), the vast majority (>90%) of *TFa* containing larvae (*Aa/bb* but not *aa/bb*) successfully formed an occlusive venous clot (Figure 5-4D). This suggests a tissue specific distribution of tissue factor.

Unexpectedly, at 3 dpf a single copy of either *TF* was sufficient for arterial clot formation (19/19 *aa/BB* ; 13/15 *AA/bb*). Overall, these data suggest that *TFb* is the primary TF expressed around the arterial vessels at 2 dpf but that due to either normal expression or a compensatory mechanism *TFa* is expressed around the arterial vessels at 3 dpf in *TFb* mutants. Furthermore, although sufficient at 3 dpf, the presence of *TFa* in *TFb* mutants is unable to completely rescue arterial hemostasis at 5 dpf indicating a disruption in the ability to generate a sustained thrombin burst.

TFa has increased *in vitro* procoagulant activity

Although the spatial differentiation of TF roles is likely due to differences in expression patterns, the inability to achieve complete compensation of TFa by TFb or TFb by TFa suggests a possible functional difference. To answer this, the coding sequences for the extracellular and transmembrane domain of both *TFs* were identified, codon optimized, modified for periplasmic secretion, epitope tagged and synthesized prior to being inserted into an inducible *E. coli* expression vector. The resulting proteins were expressed, affinity purified and incorporated into 80:20 phosphotidylcholine:phosphotidylserine liposomes to a final ratio of 15000:1 TF to lipid and final TF concentration of 173 nM. The TF containing liposomes were then used to trigger coagulation at a final concentration of 17 nM in citrated lake trout plasma following recalcification. Using a manual tilt test at 27°C, plasma clotted in roughly 120 seconds following addition of TFa-containing liposomes. TFb containing liposomes were able to form clots by 15 minutes (precise timing ended at 300 seconds). Empty control liposomes only formed clots in 2 of 6 plasma samples (Figure 5-5A). A titration of liposomes at a lower sodium concentration demonstrated a roughly 10 to 100-fold increase in procoagulant activity of TFa containing liposomes (Figure 5-5B). This supports the venous injury results and suggests a divergent role for TFb in the arterial circulation.

TFb mediated arterial clotting relies on intrinsic amplification

The TF/FVII complex is able to activate both FX and FIX in human assays. Additionally, zebrafish lack FXI, the endogenous mammalian activator of FIX. To determine whether the difference in TFa and TFb activity is due to altered ability to activate FIX, laser injury was performed in the arterial system at 2 dpf in larvae from a *f9b*$^{+/-}$ incross. Homozygosity for the *f9b* null allele resulted in a significant decrease in the ability to form occlusive thrombi (Figure 5-6A). However, loss of F9b does not alter the time to occlusion in the venous system at 3 dpf (Shavit laboratory, unpublished data). This pattern of clot formation in response to injury is similar to what was observed in the TFb deficient larvae. This alludes to TFb and FIXb operating through a similar pathway that is distinct from TFa. This would be consistent with subfunctionalization of TFa and TFb, resulting in TFb activating FIXb while TFa maintains canonical extrinsic tenase function.

Loss of TFb protects against inherited thrombophilia

Previous work demonstrates that the loss of antithrombin (At3) in zebrafish leads to an inability to form occlusive thrombi at 3 dpf due to excessive thrombin activity leading to the consumption of fibrinogen, followed by early adult lethality. To test whether concomitant loss of TF in At3 deficient offspring, three separate crosses were performed: (*at3*$^{+/-}$; *Aa/BB* x *at3*$^{-/-}$; *Aa/BB*), (*at3*$^{+/-}$; *AA/Bb* x *at3*$^{-/-}$; *AA/Bb*), and (*at3*$^{-/-}$; *Aa/BB* x *at3*$^{-/-}$; *AA/Bb*). No obvious trends were seen in the low number of *at3*$^{-/-}$ offspring produced from the *TFa* incross. The *TFb* incross demonstrated a significant overrepresentation of *AA/bb* genotypes with the *at3*$^{-/-}$ offspring. Finally, the *TFa* to *TFb* cross did lead to a significant deviation from normal expected numbers with double heterozygous offspring being overrepresented in the *at3*$^{-/-}$ population (Figure 5-6B). This shows that a loss of a single TF allele does not affect survival in the context of At3 deficiency and may make it worse in the case of TFb. In contrast, loss of 2 copies of TF, either TFb deficiency or double *TF* heterozygosity, does appear to improve survival in the context of At3 deficiency.

Larvae were generated from the crossing of *at3*$^{-/-}$; *Aa/Bb* fish and assayed for clot formation following endothelial laser injury at 3 dpf. In samples that were

wild-type or heterozygous for *TFa*, the loss of one copy of *TFb* did not seem to alter the ability to form clots. However, the complete loss of TFb resulted in all offspring regaining the ability to clot indicating that the TFb is a primary trigger of the *at3*$^{-/-}$ consumptive coagulopathy. Conversely, progressive loss of TFa appeared to cause a decrease in the rate of occlusion (Figure 5-6C). This is likely due to impairment of the thrombin burst necessary to form a clot under hypofibrinogenemic conditions present during a consumptive coagulopathy. As expected, complete loss of TF resulted in an inability to occlude. This is presumably independent of the *at3* deficiency.

Loss of TF predisposes to cardiac tamponade with hemostatic and nonhemostatic components

The aquatic environment, low blood pressure and lack of birthing trauma create a low stress early developmental environment for zebrafish compared to mammals. To understand if this low stress environment explains the ability of zebrafish to survive to 2 months of age without TF, a chemical stress challenge was performed, and the hearts of treated fish were observed (Figure 5-7A). Following 16 hours of treatment with cortisol and epinephrine beginning at 3 dpf, zebrafish with overall TF deficiency saw a significant increase in erythrocytes in the pericardial space consistent with a cardiac tamponade (Figure 5-7B Left). When the same assay was applied to prothrombin deficient fish, cardiac tamponade was additionally seen (Figure 5-7B Center). However, when the stress assay was applied to fibrinogen deficient larvae, the rate of tamponade was reduced. This suggests that the risk of tamponade is partially a result of the loss of fibrin formation (Figure 5-7B Right) but more due to an extra hemostatic role of TF and prothrombin.

DISCUSSION

TF is a fascinating protein because it has a highly conserved function that is essential for vertebrate life while having low sequence level conservation (~60%

between human and mouse and ~40% between human and fish). This low sequence conservation has been a challenge, necessitating artificial contrived systems and specialized reagents. While loss of TF has been linked to developmental abnormalities, angiogenesis, inflammation and regeneration it has been difficult to validate these phenotypes and eliminate artifacts in whole-organism models.[220,228,238,239] Zebrafish offer a model to address many of these concerns due to their easily visualized development, conserved coagulation cascade, resistance to hemostatic imbalance and finally due to the presence of 2 unique copies of TF in their genome.

The limited work on TF in zebrafish has focused on the use of morpholino antisense technology to transiently knock down translation in larvae. While knockdown of either *TFa* or *TFb* using morpholinos appears to cause mild vascular defects and edema, they have never been knocked down together.[96] In contrast to these separate TF studies, our study uses genetic mutants to study complete, permanent deficiency of TF. Following genetic *TF* deficiency, zebrafish typically survive to 8 weeks of age without gross defects in vascular morphology or endothelial development. This suggests that the morpholino-induced phenotypes are non-specific or that the genetic knockouts have a compensatory mechanism to avoid the vascular defect. Furthermore, the lack of vascular defects indicates that the weakened yolk sac vasculature in developing *TF*-null mice is specific to mammalian gestation and the unique environment of the mammalian yolk sac.

Importantly, complete TF deficiency typically causes lethality by 2 months, but a single copy of either *TFa* or *TFb* is sufficient for normal survival and sexual reproduction. This supports the use of zebrafish for studying the broad biological functions of TF that are difficult to test due to mammalian lethality. Additionally, these data confirm that although TFa and TFb share only 50% sequence identity, they both maintain the ability to activate the coagulation cascade. Building upon this baseline functional conservation, this study was able to identify subfunctionalization. Namely, TFa has higher procoagulant activity and a dominant role in the venous system while TFb contributes to a binary phenotype in the arterial system. Unfortunately, efforts to look at expression have been

unsuccessful presumably due to the low absolute levels of TF mRNA and the lack of available antibodies. However, multiple large datasets indicate that *TFa* does appear to have higher expression in the ventral aspect of the embryo near the venous circulation while *TFb* is more broadly expressed.[240,241] Interestingly, evidence for *TFb* expression in hemangiogenic progenitors points to a potential function in cell differentiation that may partially explain the observed arterial phenotype.[240] Future work to understand where and how the expression of *TF* is regulated in zebrafish could prove valuable for risk identification and thrombogenic triggers in mammals.

Beyond the potential for unique spatiotemporal expression patterns, this study provides strong evidence of a functional difference. Recombinant protein work reveals that TFa has up to 100-fold greater procoagulant activity with the caveat that lake trout plasma was used due to ease of obtainability. The mild differences in venous clotting time suggest that there likely is an absolute difference in procoagulant activity that is exaggerated by species-specific incompatibilities between zebrafish and lake trout.

In addition to its extrinsic tenase activity, it is well established that the TF:F7 complex in humans is able to activate FIX.[218] Classically, FIX is activated by FXIa in response to FXII activity of the thrombin feedback loop.[242,243] Zebrafish do not have homologues for FXI or FXII suggesting that TF is the primary activator of zebrafish FIX.[71] To explore this possibility, the *f9b* knockout line was analyzed in parallel to TF work. Separate unpublished work in the Shavit laboratory has demonstrated that *f9b* knockouts do not have an observable effect on laser-induced clot formation in the venous system at 3 dpf. Furthermore, that work also showed that *f9b* deficiency protects against the early consumptive coagulopathy present in the *at3* knockout fish. This study extended the *f9b* work to the arterial system at 2 dpf and showed a similar, but milder phenotype to the TFb arterial injury results. This is consistent with TFb preferentially activating FIXb in the arterial system during normal hemostasis and basally in the context of AT3 deficiency. Inconsistent with this, is the limited evidence that *TFb* heterozygosity does not have an apparent benefit for the laser injury or survival defects in the

thrombophilic background. This may be explained due to the compensatory upregulation of TFa in response to TFb deficiency exacerbating the thrombophilic state. This warrants further study into the mechanism of expression changes and the establishment of TFa and TFb localization. Furthermore, it suggests that understanding the substrate specificity of TFa vs. TFb may provide important insight into arterial vs. venous hemostasis and provide a model for studying intrinsic vs. extrinsic coagulation.

The ability of zebrafish to survive complete TF deficiency is compatible with previous studies of common pathway knockouts. The resistance to developmental lethality is likely due to the low impact aquatic environment, the lack of birthing trauma, and the low-pressure vasculature. By chemically inducing stress, this study was able to bring out a phenotype consistent with cardiac tamponade. Surprisingly, this same phenotype was inducible in the *f2* knockout line but not in *fga* knockout fish. This implicates secondary functions of TF and prothrombin that are separate from fibrin polymerization. This is consistent with the fact that *fga* knockout fish have a very mild survival defect, but see significantly increased mortality in the context of genetic thrombocytopenia.[73] Beyond the role of thrombocytes, TF:F7 may also protect against tamponade through its less prominent roles in PAR signaling and integrin binding. Expanding on this tamponade, humanized low-TF mice present with significant cardiac fibrosis in adulthood.[234] This is thought to be a symptom of chronic hemorrhage but in the context of these fish data may be a sign of additional underlying dysfunctions such as structural abnormalities.

Overall, the duplication of *TF* in zebrafish and the evidence for subfunctionalization offers multiple opportunities to explain the subtleties of TF's biological role. The work presented is consistent with the expectations from previous fish and mouse knockouts. However, the novel differentiation of arterial and venous function and the potential substrate specialization facilitates asking narrower questions than previously answerable. Furthermore, the presented role in thrombosis suggests a novel platform for interrogating modifiers. Ultimately, this work leaves more questions than it answers but also offers the long-term promise

of expanding our knowledge of the switches controlling hemostasis and thrombosis.

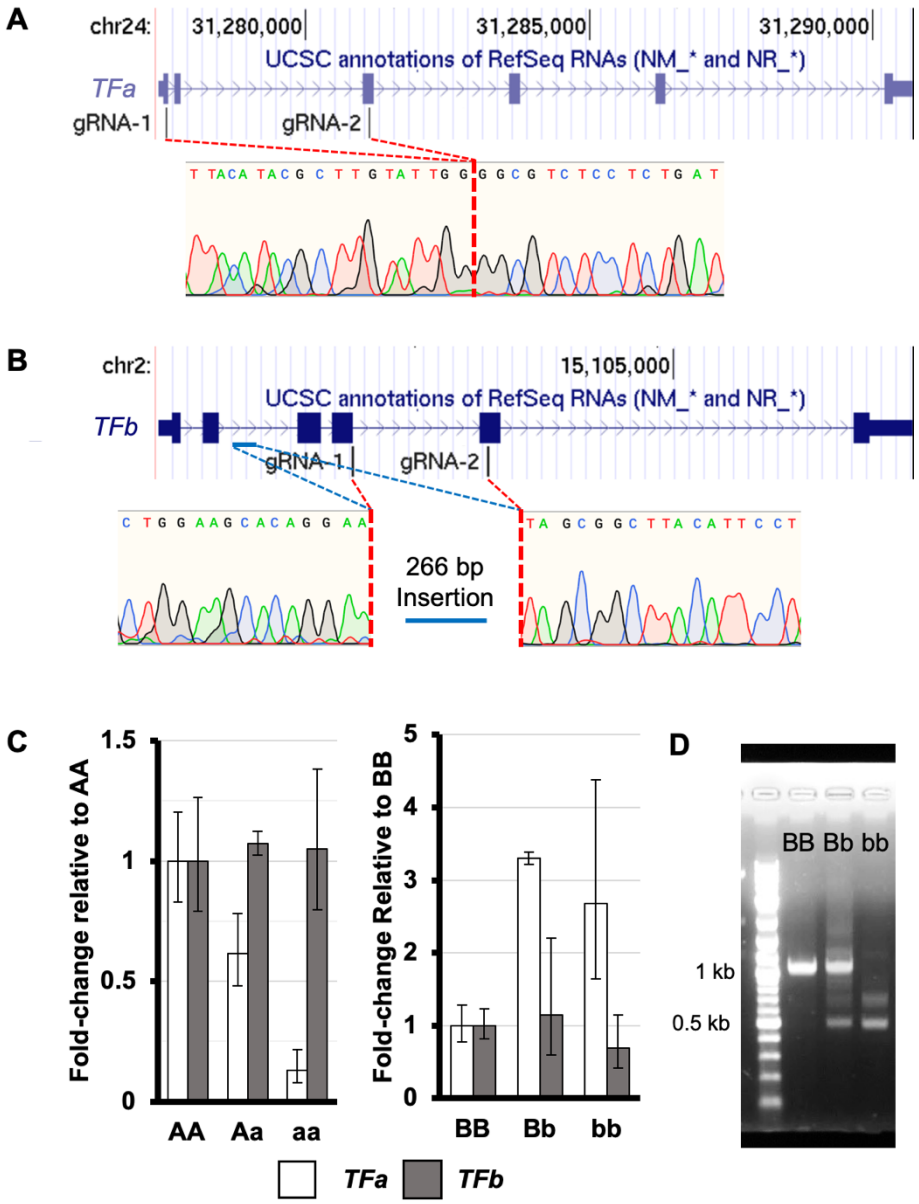
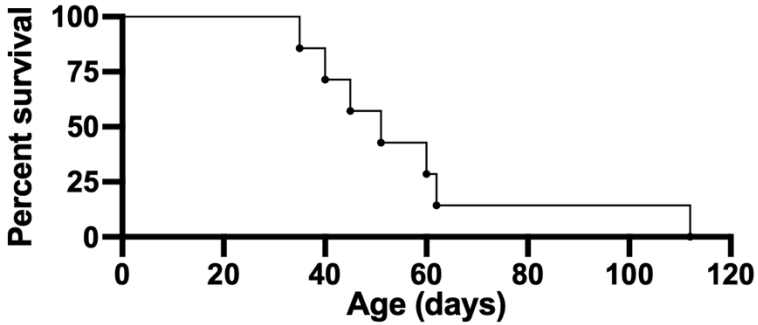


Figure 5-1. Genomic engineering creates large loss-of-function deletions in both copies of zebrafish *TF*.
Two gRNA targets were injected for both *TFa* and *TFb*. (A) A large deletion was identified between exon 1 and exon 3 in *TFa* resulting in an early stop codon. (B) A deletion was introduced between exon 4 and exon 5 of *TFb*. During endogenous repair, a section of intronic DNA (blue bar) was inverted and inserted at the site of the deletion. (C) RT-qPCR at 4 dpf demonstrates that the mutant *TFa* allele results in nonsense mediated decay of the residual transcript without affecting *TFb* levels. Conversely, *TFb* transcript is not degraded and loss of *TFb* induces upregulation of *TFa*. (D) Full-length *TFb* cDNA was isolated from pools of 4 dpf larvae. The expected wild-type band is visible in *BB* and *Bb.* Two smaller bands are evident in the *Bb* and *bb* pools consistent with alternative splicing of the mutant transcript.

A

Age(Weeks)		AABB	AABb	AAbb	AaBB	AaBb	Aabb	aaBB	aaBb	aabb	Total
8	n=	10	17	7	15	36	20	10	13	7	135
	% of expected	119%	101%	83%	89%	107%	119%	119%	77%	83%	
9	n=	10	17	7	15	36	20	10	13	0	128
	% of expected	125%	106%	88%	94%	113%	125%	125%	81%	0%	
56	n=	4	6	6	9	26	13	7	7	0	78
	% of expected	82%	62%	123%	92%	133%	133%	144%	72%	0%	
26	n=	6	14	12	27	30	27	9	21	0	146
	% of expected	66%	77%	132%	148%	82%	148%	99%	115%	0%	

B



C

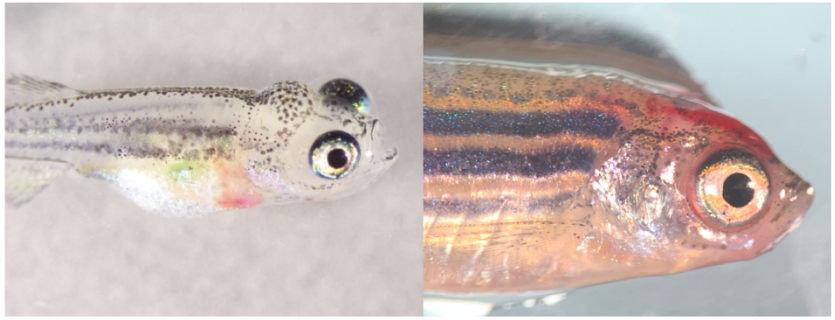


Figure 5-2. Loss of TF results in early lethality due to hemorrhage.
(A) Two groups of fish were produced from double heterozygous *TF* incrosses. The first group had normal Mendelian ratios at 8 weeks but saw complete die off of the double homozygous fish (*aa/bb*) by 9 weeks. By 56 weeks, the remaining genotypes were distributed evenly showing that a single copy of either *TF* is sufficient for survival. The second group (below solid line) was left undisturbed until genotyping at 6 months to minimize the early stress of the tailing procedure. (B) A group of 7 double homozygous fish were identified at 1 month of age and followed. 1 fish survived past 2 months and reached sexual maturity before dying by 4 months. (C) Early lethality was due to signs of hemorrhage in the pericardial space (left) or head (right).

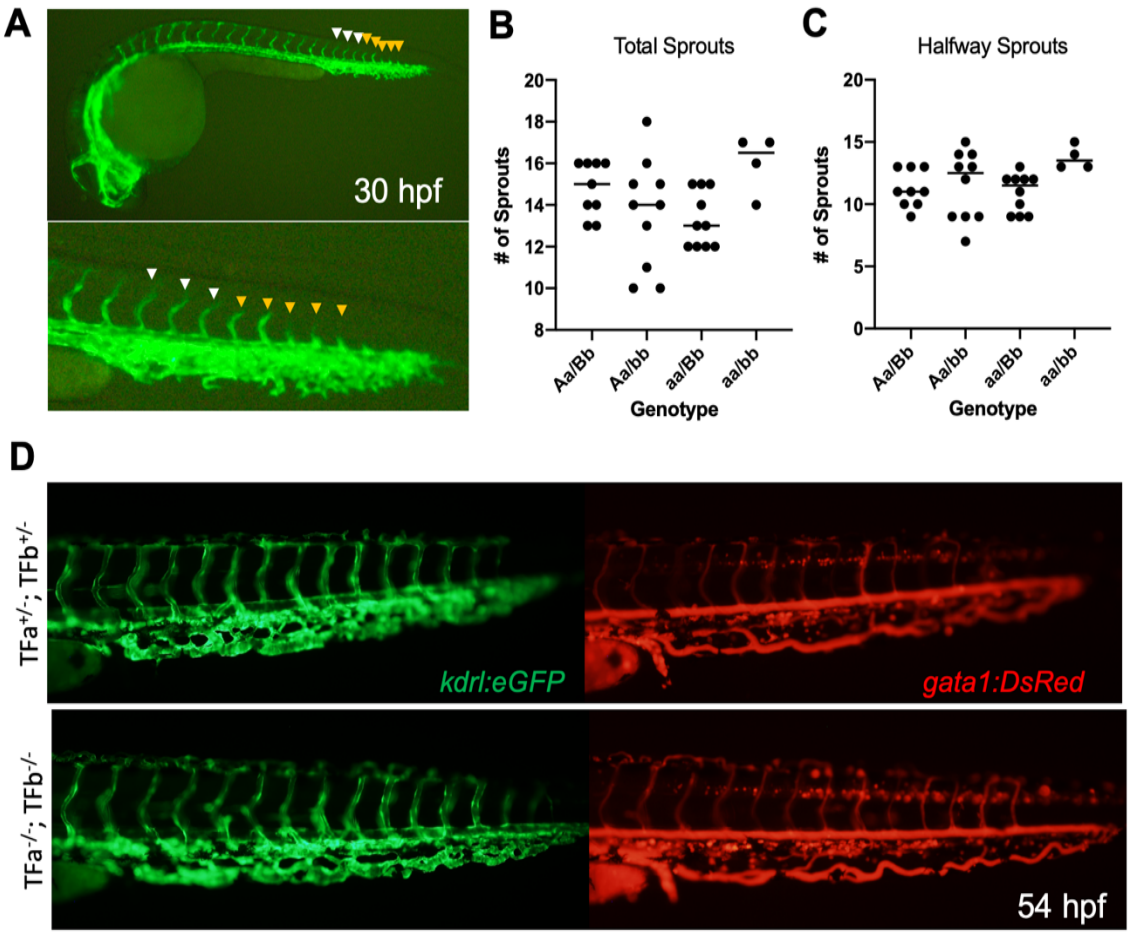


Figure 5-3. Vascular development is unaffected in the absence of TF.
(A) A 30 hpf larvae carrying the *kdrl:eGFP* transgene marking vascular endothelial cells (top) with a close-up (below). White arrows mark examples of vessel sprouts that have made it at least halfway through the expected migration. Orange arrows represent sprouts that have not made it halfway. No statistically significant changes are seen in total vessel sprouts (B) or sprouts that have made it past the halfway point (C). (D) At 54 hpf, no obvious changes are seen in endothelial cell development or in vascular perfusion as evidenced by long exposure of *gata1:DsRed*, a marker of erythrocytes.

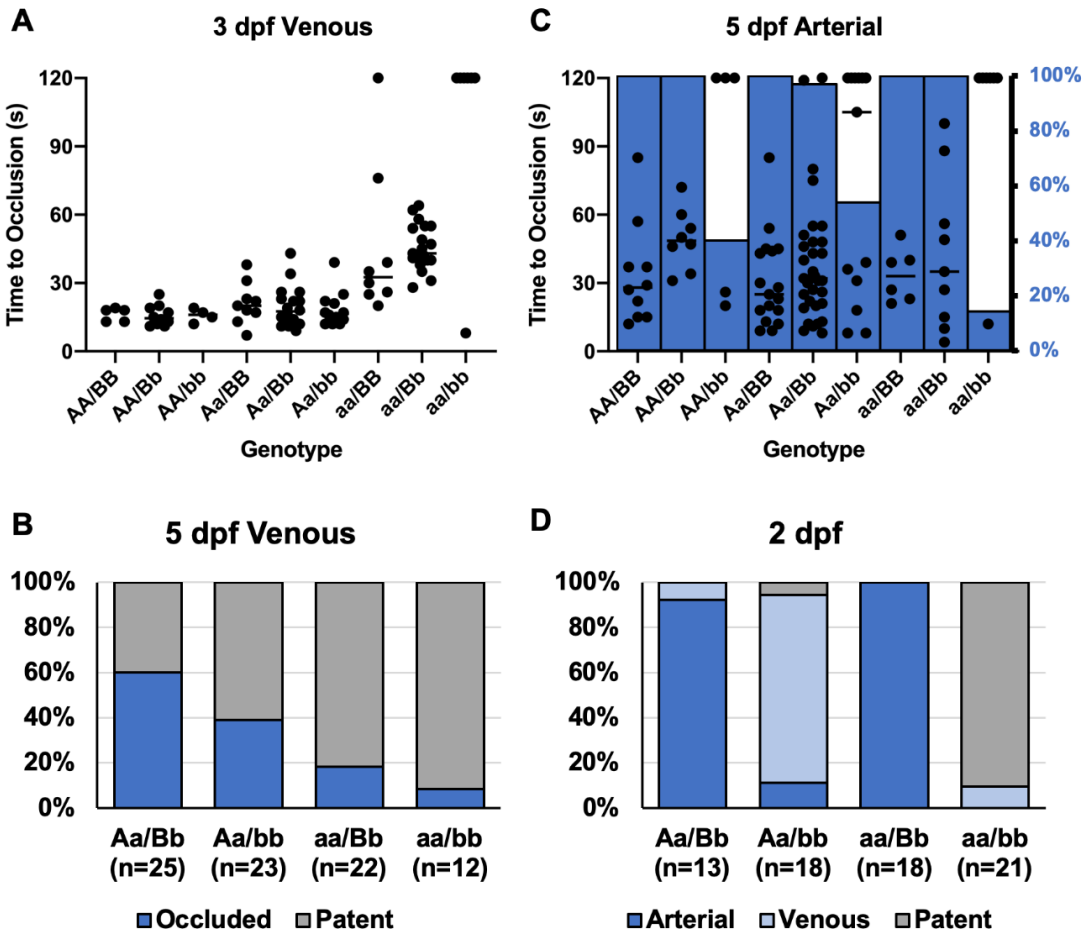


Figure 5-4. Laser-induced endothelial injury demonstrates a partial subfunctionalization of TFa in the venous system and TFb in the arterial system.

(A) Venous injury at 3 dpf shows that time to occlusion (TTO) is not altered by loss of TFb (bb) but is delayed in the complete absence of TFa (aa). This delay is exacerbated by the loss of one copy of TFb (Bb) and clots fail to form after complete TF loss (*aa/bb*). (B) Performing venous laser injury at 5 dpf confirms the predominant role of TFa in the venous system. While loss of TFb (*Aa/bb*) causes a slight decrease in occlusion rate compared to Aa/Bb, loss of TFa (*aa/Bb*) causes a significant reduction in the ability to occlude ($p = 0.01$) compared to *Aa/Bb*. (C) Arterial injury at 5 dpf shows that TTO is not altered by TF genotype. However, the ability to form occlusive thrombi is significantly reduced in fish lacking TFb (*AA/bb*, *Aa/bb, aa/bb*) when compared to *Aa/Bb* ($p < 0.001$; all 3 comparisons). (D) Laser injury at 2 dpf confirms the role of TFa in the venous and TFb in the arterial system. Larvae were injured in the arterial system first and a single copy of TFb was sufficient for occlusive arterial clot formation. Fish that failed to occlude were subsequently injured in the venous system and those with functional TFa were able to form occlusive venous thrombi. Those without TFa or TFb (*aa/bb*) did not occlude in the arterial system and had nearly absent occlusion in the venous system.

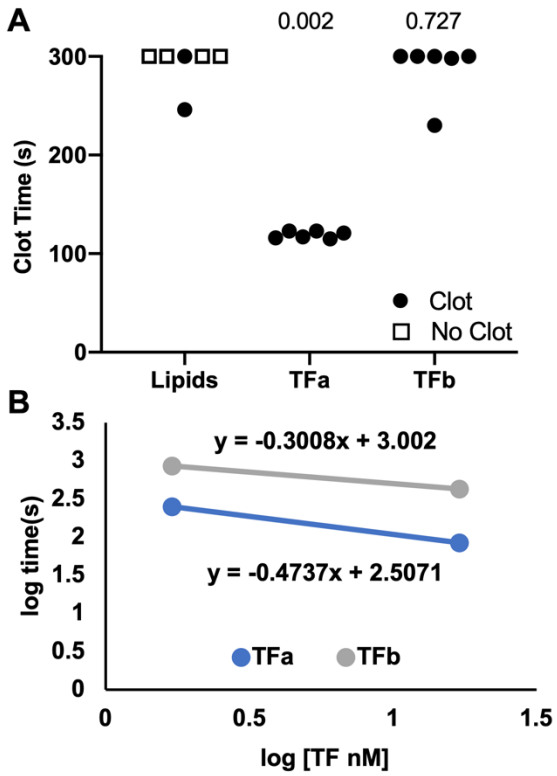


Figure 5-5. TFa has increased procoagulant activity compared to TFb in an *ex vivo* clotting assay.
Recombinant TFa and TFb were produced in *E.coli* and incorporated into 80:20 phosphotidylcholine:phosphotidylserine liposomes at a ratio of 1:15000. (A) Liposomes were mixed with citrated trout plasma and recalcified for a final concentration of 17 nM TF. TFa containing liposomes induced clot formation during a tilt test significantly faster than TFb and lipid controls. The assay was observed for 300 seconds and a final check was performed at 15 minutes to look for delayed clot formation. (B) TF solutions were diluted in HEPES-buffered saline with a lower (10mM) sodium concentration and a log-log curve plotted demonstrating a 10 to 100-fold increase of TFa procoagulant activity over TFb.

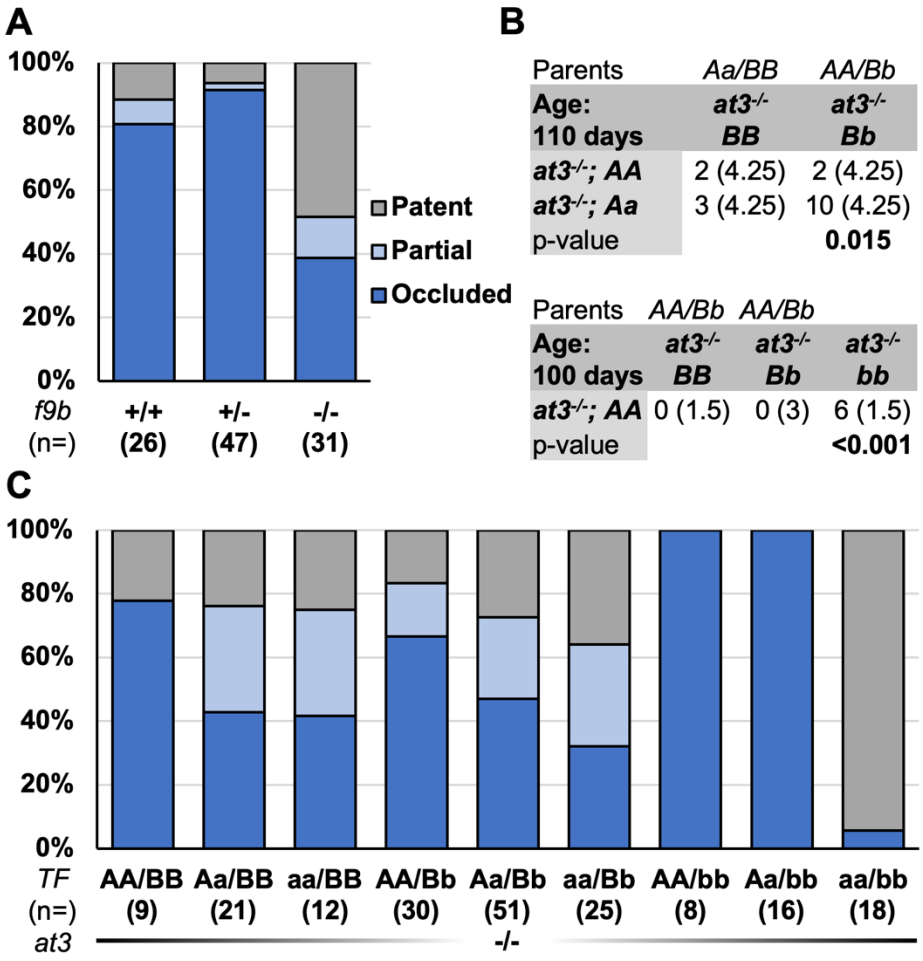


Figure 5-6. TFb acts through FIXb to trigger a consumptive coagulopathy. (A) Arterial laser injury at 2 dpf reveals that *f9b*$^{-/-}$ larvae have a significantly reduced likelihood of forming occlusive thrombi compared to wild-type and heterozygous larvae ($p < 0.005$). (B) Double heterozygosity for TFa and TFb leads to an improvement in survival of fish completely deficient of AT3 at 110 days (expected numbers in parentheses). Heterozygosity for TFb does not appear to benefit survival of thrombotic fish but complete deficiency of TFb significantly improves survival by 100 days. (C) Loss of TFb rescues the consumptive coagulopathy following venous laser injury at 3 dpf while loss of TFa exacerbates the inability to form occlusive thrombi.

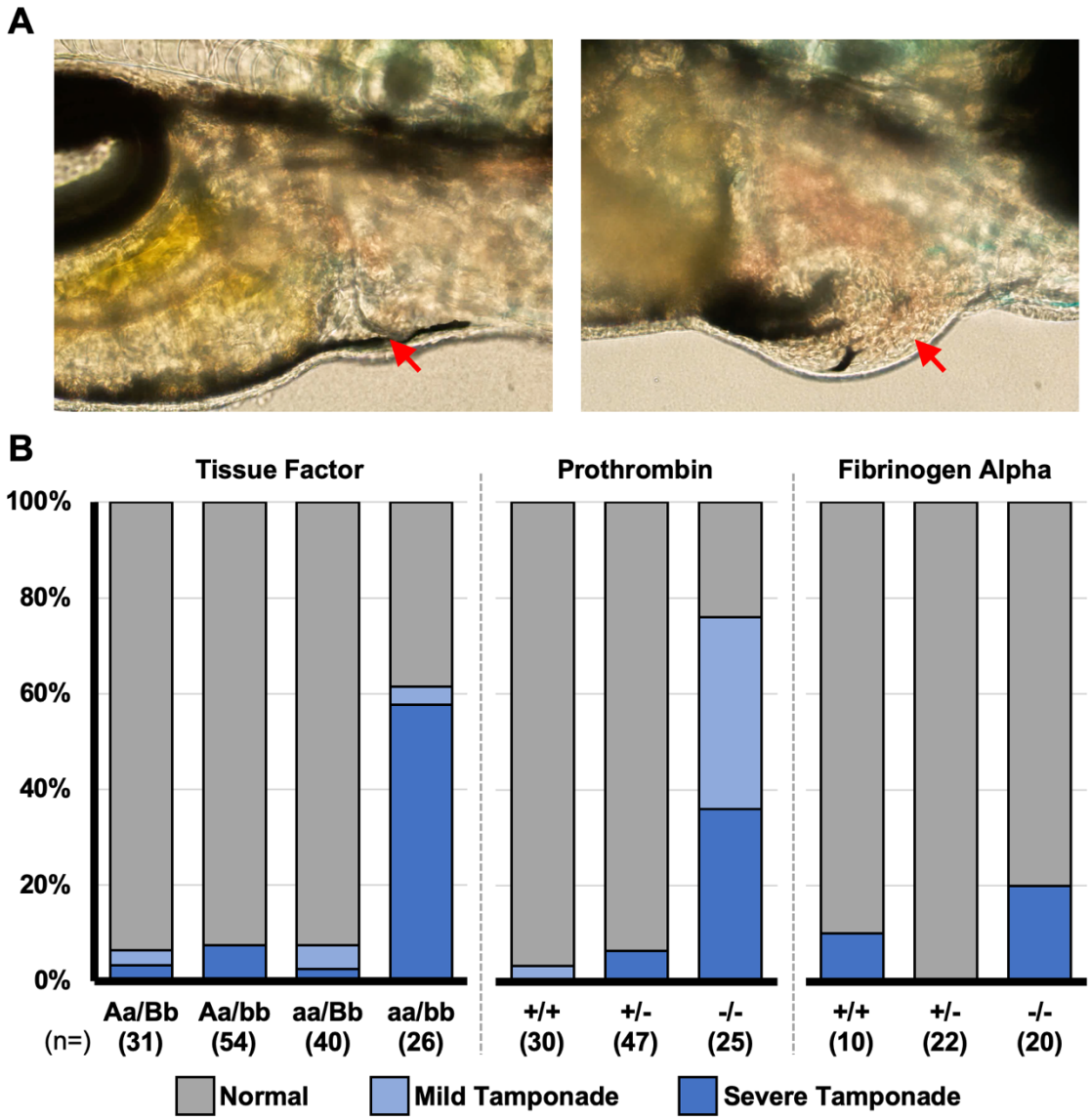


Figure 5-7. Loss of TF contributes to a risk of cardiac tamponade through hemostatic and nonhemostatic effects.

(A) Stress was chemically induced in 3 dpf by exposure for 16 hours to 25 uM epinephrine and 0.01% hydrocortisone. Larvae were score for the presence of erythrocytes in the pericardial space marked by red arrows in the control (left) and tamponade positive fish (right). (B) TF-deficient and prothrombin-deficient larvae have a significantly increased ($p < 0.01$) occurrence of cardiac tamponade compared to controls. TF deficient larvae have a stronger but not significantly different rate of tamponade. FGA-deficient larvae do not have a significantly increased rate of tamponade, suggesting hemostatic independent compensation.

Table 5-1. Oligonucleotides used in methods

Primer Description	Primer Sequence
f2 qPCR Forward	GAGGAAGCTCGAGAAGTGTTT
f2 qPCR Reverse	CCCTCTGTAGTCGCACATATTC
f2 qPCR Probe	FAM/TGGACAGGG/ZEN/TTGTTCCTTCACAGC/IABkFQ
actb2 qPCR Forward	ATGAAGATCCTGACCGAGAGA
actb2 qPCR Reverse	TCAAAGTCAAGGGCCACATAG
actb2 qPCR Probe	FAM/ACCACCACA/ZEN/GCTGAGAGGGAAATT/IABkFQ
TFa qPCR Forward	GCCAAGAGCACAGGAAAGAA
TFa qPCR Reverse	GCAGCACGTGAAGGTAGATAAG
TFa qPCR Probe	/56-FAM/ATTGAGATG/ZEN/CCGGATTTGGACCGA/3IABkFQ/
TFb qPCR Forward	GACCCAAACCTGTCAACTACA
TFb qPCR Reverse	ACTCTGTCTCTGTGCTTCTAATG
TFb qPCR Probe	/56-FAM/TCGTGACAA/ZEN/ACAGAGAAACCCTCACTG/3IABkFQ/
gRNA T7 Universal Forward	GCTAATACGACTCACTATAGG
gRNA Universal Reverse	AAAAGCACCGACTCGGTGCC
gRNA Backbone	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT
TFa gRNA #1	TAATACGACTCACTATAGGATACGCTTGTATTGGCGCGTTTTAGAGCTAGAAATAGC
TFa gRNA #2	TAATACGACTCACTATAGGCCTGTCTGAACCCATGCGGTTTTAGAGCTAGAAATAGC
TFb gRNA #1	TAATACGACTCACTATAGGCACAGGAAAAGTAAGTAAGTTTTAGAGCTAGAAATAGC
TFb gRNA #2	TAATACGACTCACTATAGGGTTACTGCTTCAGTGTAGGTTTTAGAGCTAGAAATAGC
TFa genotyping primer 1	AGACTCAGCCAGGACAGAGAAC
TFa genotyping primer 2	GAGAATTTCCCAGAGCTGACAT
TFa genotyping primer 3	CAGTATTGTGGACTTGGGACAA
TFb genotyping primer 1	TGACCATCCAAGACCCCATCA
TFb genotyping primer 2	CATCCCTGCGGTACATTTATTT
TFb genotyping primer 3	TAGGGGTTCAGCTTACCTTCAA
f9b genotyping primer 1	CCTCAAAGTAGCACCACTCC
f9b genotyping primer 2	CAATAGGTGAACATGACGTC
f9b genotyping primer 3	CCAGGCAGATAGGTACAGCG
at3 genotyping Forward	ACACGGAAACGAGGAATCTG
at3 genotyping Reverse	TGCAAAAATTCCTGAGGACAA

Table 5-2. Synthesized recombinant TF gBlocks

>Zebrafish Membrane Localized TFa for E.coli

GGAGATATACATATGAAATATCTGCTGCCGACGGCGGCCGCGGGATTACTGTTGTTAG
CCGCTCAGCCCGCGATGGCTGCAGAAGACCAGGTGGATCCTCGGTTGATAGATGGTAA
GGTCTTTCCCAAAGCTAAGAACGTCTCTTGGAGTAGCGTTAACTTCAAATCTATGCTG
ACATGGTCTCCTAAGCCAACAAATTACTCGTACACGGTGGAGTTCTCGGAGCTGGGTC
AGGACCGCGAGCGGACTCCCTACTGTATTCGGACAATGGATACCGAGTGTGATCTTAC
GGCGGTCTTAAAGAACTTAAAGGCCTATTATTCCGCTGATGTACTTAGCGAGCCGATG
AGAGGTGTATCAAGCGATCTTGTTGAGTTTCCCCACGTCTCTTCTGGAAAGTTCAGTC
CCTATCACGATACTGACATCGGACGTCCAGAGTTTAAGATTGAGGTCTCCTCCGATAA
ACGGATGACCAAACTGCATGTCACCGACGTACCTACGGCGTTATTTGATGATCAAAAA
AAACGGCTTAACATCAGAGATGTTTTCGGTGACGAGTTGCAGTATAAAGTCATTTACC
GTAAAGCCAAAAGCACCGGCAAGAAGGAAATGCTGTCCAAACAATCCATCATAGAGAT
GCCGGATTTAGACAGAGGTGTGGGATATTGTTTTAACGTGCAGGCCTATCTGCCTAGT
CGCGCAGCAAACAAGCAATTTGGGGAGCTGTCCTCGGTTCATTGCTCCACCGAGGAAA
ATACGACGGTCTTTGAAGAATATGGGACTGGAGTGATCACCGGTGTTATCGTATTCAT
ATTATCGGCCATAATCGTCATTGTTTTAGTAATTGTGCACAAGTGATAAGCTTTCCGA
CATCACCATCACCATCACTGAGATCCGGCTGCTA

> Zebrafish Membrane Localized TFb for E.coli

GGAGATATACATATGAAATATCTGCTGCCGACGGCGGCCGCGGGATTACTGTTGTTAG
CCGCTCAGCCCGCGATGGCTGCAGAAGACCAGGTGGATCCTCGCCTTATTGATGGGAA
AGCTAGCATGGACGTTGGCAAGTTAACTAAAGCAACGAATGTAAGCTGGACATCCTAC
AATTTCAAGACGATCTTGAGCTGGGGCCCGAAGCCTGTCAACTATACTTACACTGTCG
AGTTTTCACGGACAAACCGCGATAAGCAGCGCAACCCTCATTGTATCAGAAGCACCGA
AACTGAATGCGATCTGACGAACGACCTGGATATCAATGAGGTCTATTCCGCTGAAGTG
TTATCTGAGCCTTTACCGAGCATGAATATTGATCAAGTCGAACCTCCCTATAGTCGTT
CAAAGATATTCAGACCATATGACGATACTTTGATCGGACGGCCACAATTTACGCTTAC
AGTTAGCGTGGACAAAAAATTAGTATTGACTATACAAGATCCAATAACTGCGCTTCAC
AAGGATAATAGATCGCTGAACATCAGAGACATCTTTAAGAAGAATTTGAAATATAAAG
TGGCTTATTCTAAGGCGGGGAGCACAGGGAAGAAGATCAAAGTTGTAGAAGAAAGCCG
CGCCGAGTTCAATAGATTAGATGAGGATCAATCCTACTGTTTTTCAGTAGCGGCTTAT
ATCCCGAATCGCAAAGGGGATAAACGGTTGGGGGAATGGAGCCTGCCAAAGTGTTCGC
CACAAGAATCGAAATCACTGTTTGAGGAGTATGGCTTAGCCGTAATAGGCGGCGCAGC
ATTAGCGACCCTTGCATTTGTAATAGCCGTTATAGTCCTTATAGTGGTCCATAAATGA
TAAGCTTTCCGACATCACCATCACCATCACTGAGATCCGGCTGCTA

Chapter VI:
Conclusions and Future Directions

CONCLUSIONS

Disruption of the kringle 1 domain of prothrombin leads to late onset mortality in zebrafish

Complete prothrombin deficiency is incompatible with life in mammals. Previous work demonstrated that zebrafish were able to tolerate the loss of common pathway factors (FV and FX) and survive to several months of age. To understand if this survival was due to low level thrombin activity or whether zebrafish would facilitate the exploration of secondary biological roles of prothrombin, a targeted knockout was generated. Surprisingly, the 14-bp genomic deletion resulted in an alternatively spliced transcript with a 15 amino acid deletion in the kringle 1 domain. Analysis of the corresponding deletion in human prothrombin revealed a reduction in synthesis, a reduced activation rate, and a decrease in activity of the mature thrombin. Overall, theses compounding deficiencies suggest a >99% reduction of prothrombin procoagulant activity while demonstrating a novel role for kringle 1 in the maturation and activity of thrombin. Returning to the zebrafish model, fish homozygous for the *f2* deletion had an inability to form occlusive thrombi in response to injury and suffered mortality due to hemorrhage within the first 3 months of age. This confirms the conserved role of prothrombin in zebrafish, counters the suggestion that prothrombin has an extrahemostatic role in development and offers a model for studying potential secondary functions of thrombin activity.

A bioinformatic pipeline for identifying dominant mutations in the zebrafish model using high throughput sequencing

Rapid technological advances have made obtaining base pair level genomic resolution readily available for countless model systems. With the release of the 11th version of the zebrafish genome, GRCz11/danRer11, the zebrafish community has a high-quality reference with the addition of alternative contigs for the first time in a non-human model. Merging the availability of sequencing data with the strengths of mutagenic screens in zebrafish promises to rapidly advance our ability to identify new biology. Unfortunately, the limited availability of high-quality resources and standardized protocols within the community limit the potential of this technology.

Chapter III established a standard protocol and created necessary index files and variant datasets for processing high-throughput sequencing. Applying this protocol defined baseline variation across a population of 85 fish with roughly 10-12 million SNVs per fish and a Ti/Tv ratio of 1.12. Furthermore, the pipeline was used to confirm an ENU mutagenesis rate slightly higher than previously described with roughly 70 loss of function mutations per gamete. Finally, an unbiased approach was able to identify mappable genetic markers and confirm the chromosomal location of a known modifier of thrombosis as well as provide evidence of a secondary locus contributing to improved survival. This pipeline should provide a streamlined approach for identifying dominant mutations underlying phenotypic variation.

Dissecting genetic modifiers of a sensitized zebrafish model of thrombosis

Thrombosis is one of the leading causes of morbidity and mortality, but the landscape of heritable risk factors is incompletely defined. Zebrafish lacking AT3 suffer from a consumptive coagulopathy and succumb to early mortality due to thrombosis. Using this as a sensitized model, chapter IV aimed to identify how hemostasis could be restored with known and novel factors. Complete genetic deficiency of *f2*, *f10 and plg* resulted in increased mortality in the thrombophilic At3 deficient fish. However, heterozygosity of each of these genes led to improved

survival. Loss of *f2* and *f10* likely reduces the thrombotic burden by reducing the procoagulant content of the plasma. In contrast, *plg* deficiency likely stabilizes physiologically appropriate clots and alleviates a previously unrecognized risk of hemorrhage.

Chemical mutagenesis was performed, and 4 lines were established with a heritable survival benefit in the AT3 deficient background. Initial analysis identified the *frost* allele due to in a missense mutation at a highly conserved cysteine in prothrombin. When engineered into human prothrombin, this mutation resulted in lower synthesis and a compete inability to cleave fibrinogen. *In vivo*, the *frost* allele rescued adult survival and the larval consumptive coagulopathy of *at3*$^{-/-}$ fish in a time dependent manner. Continued analysis failed to identify known coagulation mutations in the additional lines, which points to the existence of novel modifiers. Additionally, given that *plg* and *f10* mutations are known to rescue *at3*$^{-/-}$ survival, their absence in the candidate analysis implies the experiment did not reach saturation of the genome. Further genetic characterization of these lines is likely to reveal new biology underlying thrombosis.

Genetic duplication of tissue factor in teleost reveals partial subfunctionalization

TF deficiency is the most severe of procoagulant mutations in mice with lethality by E6.5-E8.5 and severe deficiency has never been observed in humans. This suggests a secondary role for TF outside canonical fibrin generation. Zebrafish have 2 copies of TF (TFa and TFb) that may provide insight into subfunctions of TF. Loss of both TFs in zebrafish leads to normal development but results in lethality by 8 weeks. A single copy of either *TFa* or *TFb* is sufficient to rescue normal survival. Interestingly, TFa appears to have greater *ex vivo* procoagulant activity based on recombinant protein work and is the predominant provider of *in vivo* activity in the venous system. In contrast, TFb has little *ex vivo* procoagulant activity but deficiency results in a failure to form arterial clots at 2 dpf and a significant decrease in the ability to form occlusive arterial thrombi at 5 dpf. Coincidentally, although loss of *f9b* does not alter venous clot formation, it does

lead to a reduction in arterial clot formation at 2 dpf. Furthermore, complete deficiency of TFb or FIXb is able to rescue the prothrombotic phenotypes resulting from AT3 deficiency. This is consistent with the hypothesis that TFb is the predominant activator of FIX in zebrafish while TFa acts primarily through canonical extrinsic FX activation. Finally, loss of TF leads to an increased risk of cardiac tamponade in response to stress. This risk is phenocopied by thrombin deficiency but not by loss of fibrinogen. These data suggest unique biology differentiating arterial and venous hemostasis and provide an avenue for isolating the influence of intrinsic and extrinsic tenase activity in normal hemostasis and thrombosis.

FUTURE DIRECTIONS

Characterize secondary biological roles of prothrombin

The ability of zebrafish to survive severe hemostatic challenge has been consistent across studies, but nevertheless surprising. One explanation for this resilience is the lack of environmental risk factors relative to terrestrial models. A second explanation relates to the nature of the zebrafish knockouts used. *f5, f10,* and *TF* knockouts do not eliminate the possibility of low-level thrombin activity. Furthermore, while the *frost* and engineered *f2* alleles have >99% reduction in realized procoagulant activity, they both still maintain some catalytic activity. It is entirely possible that miniscule levels of prothrombin are capable of providing the secondary activation of pathways necessary to prevent the defects seen in chemical and antisense knockdown of prothrombin.

To definitively answer this possibility, a third knockout could be made with the complete deletion of the *f2* coding sequence. Genomic engineering technology has advanced sufficiently to make this feasible and avoid the pitfalls of alternative-splicing previously seen. Once this genetic background is established, several questions regarding prothrombin activity in vascular development/integrity, inflammation and cell migration could be addressed.

First, if the complete knockout recapitulates the developmental defects seen in previous mouse work, it would be imperative to do RNAseq comparing the complete knockout to the previous loss-of-function alleles to understand which transcriptional pathways are being activated extrahemostatically. In contrast, if those same developmental defects are not seen, then a comparison could be done between genetic knockouts and the morpholino knockdowns to determine if the phenotypic difference is due to non-specific effects or as a result of genome specific compensation. Presumably, performing morpholino knockdown in *f2*-null fish would cause developmental defects if the phenotype is due to toxicity or result in no change if there is specific compensation.

Beyond the developmental defects, thrombin has been implicated in inflammation and immune response. Utilizing fluorescent angiography and neutrophil staining following exposure to LPS and/or physical injury would provide an *in vivo* platform for understanding secondary thrombin roles. Along with the analysis of developmental defects, this understanding would be vital for the improved management of patients receiving anticoagulant therapy.

Automate the bioinformatics pipeline and expand on the use of alternate contigs

The key to any new technology is lowering the barrier to accessibility. While high-throughput sequencing has improved exponentially in the last decade, niche models still lag behind in custom resources. Although this work has expanded on the availability of resources, the computational processing requires substantial technical knowledge. Moving forward, a framework for automating the dynamic modification of the genomics pipeline would greatly increase the throughput and accessibility. As an initial step, a custom python script for local modification and automatic generation job scheduling scripts is achievable with relatively little investment. Further plans would involve creating a platform agnostic framework for use by the broader community.

From the technical aspect of the pipeline, the use of alternative contigs is increasingly important. For a model like zebrafish where hybrid strains are

commonly used, the alternative contigs provide a valuable reduction in technical artifacts. Unfortunately, the current pipeline ignores many of these secondary variants. Long term, the reincorporation of these variants is critical. In the context of mutagenesis screens, an initial approach would involve aggregating alternate contigs with their standard reference locations and computationally determining broad heterozygosity or homozygosity. This would provide mapping data and then the alternate contigs could be closely analyzed for novel candidate mutations. This secondary analysis would require the manual curation of coding sequences on alternate contigs to confirm the reliability.

Explore progeny testing for minimization of stochastic false positives in suppressor mapping

The single largest issue with the genetic mapping of a survival benefit attempted in this work is the existence of stochastic survivors producing false positive information. A secondary issue is the interaction of background mutations and the potential for multigenic effects on survival. Limiting this requires a significant expansion of the pedigree that is uniquely possible in zebrafish. Instead of sequencing long lived survivors, it would be valuable to use a progeny testing approach. Progeny testing involves testing a phenotype in the offspring of a sample of interest.

Beginning with the unmapped Q, F or L line, 20 candidate *at3*$^{-/-}$ fish would be crossed to unmutagenized *at3*$^{+/-}$ partners. The goal would be to generate roughly 80-100 offspring surviving >3 months from each pair. Comparing the ratio of homozygous to heterozygous fish at 6 months would provide a penetrance score for each parent. For example, 80 survivors with a ratio of 60:20 heterozygous to homozygous genotypes calculates to a 66% rescue penetrance. (20 *at3*$^{-/-}$ */ (*60 expected *at3*$^{-/-}$ x 50% suppressor carriers) = 2/3). The penetrance scores across the population would presumably cluster into 2 groups for a monogenic trait and 3 or more for multigenic traits. Carriers and controls would then be sent for whole genome sequencing based on the predicted suppressor status. This would provide

significantly higher confidence in the analysis and lead to improved candidate reliability.

Uncover the origin of arterial and venous tropisms of tissue factor duplicates

Previous work shows that exposing venous grafts from patients to coronary arterial flow induces *TF* expression.[244] The different phenotypes of TFa and TFb deficiencies in the arterial and venous systems in chapter V implies likely differences in spatial and/or temporal expression patterns. This regulatory landscape may overlap with the single TF landscape in humans. This leads to 3 unanswered questions from the perspective of expression. First, where are *TFa* and *TFb* endogenously expressed? Attempts to perform *in situ* hybridization have been unsuccessful, potentially due to extremely low endogenous levels. This could be circumvented through one of three options. CRISPR/Cas9 has made possible the potential to make targeted knock-ins in zebrafish. This technology could be applied to fuse a fluorescent reporter to the endogenous TF and view expression with epifluorescence. A secondary approach would be to create a transgenic reporter using a predicted promoter region, but this would be limited by an inability to confirm the accuracy of the expression pattern. Similarly limited, single cell RNAseq could be performed on disassociated developing larvae to look for differences in expression, but as seen in a previous study, correlating transcriptional patterns back to known cell populations is challenging..[241] Finally, antibodies could be generated from recombinant TFa and TFb and protein expression visualized.

Second, assuming a difference in baseline expression, the next question would be how is the expression pattern established between arterial and venous circulation. This is especially interesting due to the fact that TFb is deposited in the oocyte transcriptome and present before zygotic transcription takes place. An initial hypothesis is that low level leakage of F7 and prothrombin from the circulation creates a gradient of TF-mediated PAR signaling proportional to the flow of the proximal vessel. Differences in this gradient may create an “arterial” or “venous” expression pattern that leads to differing and overlapping *TFa* and *TFb*

patterns. This hypothesis has the flexibility to incorporate maternally deposited *TFb* as the seed for initializing transcriptional programming. To begin testing this idea, expression of TF would be analyzed in offspring generated from *TFb*-null female parents to check for changes in the initial establishment of the pattern. In parallel to this, chimeric CRISPR knockdowns of the various PAR receptors, prothrombin and F7 would be used to check if those genes mediate the distribution of *TF*. Finally, morphants with altered circulation such as *silent heart* could be analyzed for changes in TF expression to understand if the pattern relies on flow.

Finally, the upregulation of *TFa* in response to *TFb* loss suggests a feedback loop. The first task would be simply to characterize whether the upregulation occurs in cells that already express *TFa* or if the expression of *TFa* is turned on in cells that usually only express *TFb*. The latter is supported given the fact that *TFb*-null larvae have an arterial phenotype at 2 dpf but not 3 dpf, suggestive of a delayed compensatory mechanism. The return of the phenotype at 5 dpf may be a result of incomplete functional compensation.

Regardless of the identity of the overexpressing cells, the next task is to understand the mechanism. Compensation may be the result of genomic compensation due to inappropriate mRNA products and independent of TF function.[245] This could be answered by overexpressing functional TFb to see if it reverses the induction of *TFa* expression as seen via qPCR. If overexpression does prevent the compensation, then there is likely a protein level feedback loop through extracellular signaling or through signaling mediated by the cytoplasmic tail of TF. The first step to detect this would be performing RNAseq on TFb-deficient larvae to determine candidate pathways for further analysis. Ultimately, understanding how the transcriptional profile of TFa and TFb is established and maintained would provide an excellent model for determining pathologies that might predispose to ectopic expression and increase the risk of thrombosis.

Detail the substrate specificity of TF:F7

In addition to expression differences, there appears to be a functional difference between TFa and TFb as evidenced by the *ex vivo* clotting assay and

the FIXb mediated thrombophilic contribution. In humans, it is well established that the TF:F7 complex is able to cleave FX as well as FIX but that the primary activation of FIX occurs via FXI. Zebrafish are lacking FXI and thus TFb may take on a more critical role in the activation of FIXb. Unfortunately, in zebrafish the story is cloudier due to the tandem duplication of F7 resulting in F7L, F7i and F7. Thus, the substrate specificity of TF may be reliant not only on the binding of substrates but modified by affinity for the various F7 copies. First, to confirm a functional physiologic difference, TFa and TFb should be overexpressed in the absence of the other. For example, overexpression of TFa in TFb-null larvae will likely rescue the 2 dpf and possibly rescue the 5 dpf arterial phenotype. This would confirm the functional overlap and indicate a primarily expression mediated phenotype. Similarly, overexpression of TFb in the absence of TFa may be able to increase the venous TTO. Second, overexpressing TFa in the FIXb knockout background would reveal the degree of functional overlap between TFa and TFb.

For a more definitive answer, a substantial set of *in vitro* assays would need to be performed. This would require the recombinant production not only of TFa and TFb but also of F7, F7L, F7i, FIXa, FIXb, and FX. The first set of assays would be to calculate the binding affinities of the TF:F7 permutations using surface plasmon resonance. The sequence divergence between the protein families suggests there would be significant differences in affinity.

Second, using chromogenic substrates specific to FIX and FX, the 18 combinations of TF:F7:FIX/FX would need to be assayed. This would definitively prove whether the TF:F7 complexes have substrate preferences and the degree to which F7 modifies this preference. Having established the preferences, the previously performed clotting assays could be slightly modified through the addition of the relevant F7 copy to understand if the absolute difference in procoagulant activity is artifactual or a physiologic reality. From this information, structural studies could be undertaken to determine exactly how these differences are manifested and whether or not they can be applied to potential new hemostatic modifying therapies.

Identify potential tissue factor dependent nonhemostatic factors stabilizing the cardiac structure

TF-deficient mice carrying low levels of human TF develop hemosiderosis and cardiac fibrosis by 3 months of age.[231] Consistent with this, TF-deficient zebrafish larvae develop a cardiac tamponade in response to chemical stress. Surprisingly, this phenotype is minimal in *fga*$^{-/-}$ and points to extrahemostatic stabilizers. Two future avenues of research may prove beneficial. First, as previously suggested, laboratory zebrafish live in an environment with low hemostatic challenge. Their ability to survive into adulthood without critical clotting factors could be due to the absence of stress. It would be informative to expose developing zebrafish to chronic cortisol and epinephrine with the goal of narrowing the prolonged survival window. This would have the benefit of confirming the low stress hypothesis while establishing a more predictable phenotype for future studies.

Returning to the cardiac tamponade, the normal presence of TF likely protects against tamponade through 3 pathways. The most obvious is TF's well-established role in fibrin generation. If this was the only factor, then loss of *fga* would have a similar phenotype. The second factor is TF's role in FVII mediated PAR signaling. In humans, FVII can directly activate PAR2 but can also lead to the downstream activation of prothrombin to signal through PAR1, PAR3 and PAR4. The presence of tamponade in *f2*$^{-/-}$ fish suggests that PAR signaling does play a role. To further this hypothesis, it would be valuable to analyze the formation of tamponade in *f10*$^{-/-}$ larvae and in a not yet established *f7*$^{-/-}$ line. Determining which coagulation factors replicate the tamponade phenotype would narrow down potential candidates for extrahemostatic signaling. Furthermore, there are 5 PAR receptors in zebrafish and it would be possible to transiently knock each of these down in *fga*$^{-/-}$ larvae and determine if this is able to recapitulate the phenotype. If a combination of PAR knockdown with known coagulation factors fully explains the tamponade, then further studies could be pursued to determine how PAR signaling stabilizes the cardiac structure and whether this has potential implications in aging populations.

In contrast, if PAR signaling does not contribute to the tamponade phenotype, the recently described binding of TF to integrins and complementary intracellular signaling may implicate new biology. The low likelihood of this pathway and incomplete picture regarding TF:integrin binding precludes the aggressive physiologic pursuit of this hypothesis. However, preliminary RNAseq would be performed on TF-deficient hearts and the molecular pathways compared to control. This would help to identify any transcriptional changes that would implicate known pathways. The broad role of integrins in structural interactions would preclude knockdown studies. However, targeting these intracellular mediators through knockdown or targeted overexpression to modify cardiac morphology and tamponade frequency would provide downstream evidence of this pathway and encourage future pursuit.

Ultimately, improved understanding of the crosstalk between tissue structure and hemostasis provides fundamental knowledge for understanding closed-system circulation and vertebrate survival.

References

1. Michiels C. Endothelial cell functions. *J Cell Physiol*. 2003;196(3):430-443.
2. Brass S. Cardiovascular biology. Platelets and proteases. *Nature*. 2001;413(6851):26-27.
3. Davidson CJ, Hirt RP, Lal K, et al. Molecular evolution of the vertebrate blood coagulation network. *Thromb Haemost*. 2003;89(3):420-428.
4. Shavit JA, Ginsburg D. Hemophilias and Other Disorders of Hemostasis. In: Rimoin D, Pyeritz R, Korf B eds. Emery and Rimoins Principles and Practice of Medical Genetics; 2013:1-33.
5. Farndale RW, Sixma JJ, Barnes MJ, de Groot PG. The role of collagen in thrombosis and hemostasis. *J Thromb Haemost*. 2004;2(4):561-573.
6. Savage B, Saldivar E, Ruggeri ZM. Initiation of platelet adhesion by arrest onto fibrinogen or translocation on von Willebrand factor. *Cell*. 1996;84(2):289-297.
7. Ren Q, Ye S, Whiteheart SW. The platelet release reaction: just when you thought platelet secretion was simple. *Curr Opin Hematol*. 2008;15(5):537-541.
8. Bennett JS. Platelet-fibrinogen interactions. *Ann N Y Acad Sci*. 2001;936:340-354.
9. Holinstat M. Normal platelet function. *Cancer Metastasis Rev*. 2017;36(2):195-198.
10. Cosemans JM, Angelillo-Scherrer A, Mattheij NJ, Heemskerk JW. The effects of arterial flow on platelet activation, thrombus growth, and stabilization. *Cardiovasc Res*. 2013;99(2):342-352.
11. Morrissey JH, Pureza V, Davis-Harrison RL, et al. Protein-membrane interactions: blood clotting on nanoscale bilayers. *J Thromb Haemost*. 2009;7 Suppl 1:169-172.
12. Mann KG, Nesheim ME, Church WR, Haley P, Krishnaswamy S. Surface-dependent reactions of the vitamin K-dependent enzyme complexes. *Blood*. 1990;76(1):1-16.
13. Morrissey JH, Tajkhorshid E, Sligar SG, Rienstra CM. Tissue factor/factor VIIa complex: role of the membrane surface. *Thromb Res*. 2012;129 Suppl 2:S8-10.
14. Mertens K, Bertina RM. Pathways in the activation of human coagulation factor X. *Biochem J*. 1980;185(3):647-658.
15. Walker RK, Krishnaswamy S. The activation of prothrombin by the prothrombinase complex. The contribution of the substrate-membrane interaction to catalysis. *J Biol Chem*. 1994;269(44):27441-27450.

16. Wolberg AS, Campbell RA. Thrombin generation, fibrin clot formation and hemostasis. Transfusion and apheresis science : official journal of the World Apheresis Association : official journal of the European Society for Haemapheresis. Vol. 38; 2008:15-23.
17. Jesty J, Beltrami E. Positive feedbacks of coagulation: their role in threshold regulation. *Arterioscler Thromb Vasc Biol*. 2005;25(12):2463-2469.
18. Brandstetter H, Bauer M, Huber R, Lollar P, Bode W. X-ray structure of clotting factor IXa: active site and module structure related to Xase activity and hemophilia B. *Proc Natl Acad Sci U S A*. 1995;92(21):9796-9800.
19. Lammle B, Wuillemin WA, Huber I, et al. Thromboembolism and bleeding tendency in congenital factor XII deficiency--a study on 74 subjects from 14 Swiss families. *Thromb Haemost*. 1991;65(2):117-121.
20. Kubier A, O'Brien M. Endogenous anticoagulants. *Top Companion Anim Med*. 2012;27(2):81-87.
21. Oklu R. Thrombosis. *Cardiovasc Diagn Ther*. 2017;7(Suppl 3):S131-S133.
22. Previtali E, Bucciarelli P, Passamonti SM, Martinelli I. Risk factors for venous and arterial thrombosis. *Blood Transfus*. 2011;9(2):120-138.
23. Jackson SP. Arterial thrombosis--insidious, unpredictable and deadly. *Nat Med*. 2011;17(11):1423-1436.
24. Kume T, Akasaka T, Kawamoto T, et al. Assessment of coronary arterial thrombus by optical coherence tomography. *Am J Cardiol*. 2006;97(12):1713-1717.
25. Wendelboe AM, Raskob GE. Global Burden of Thrombosis: Epidemiologic Aspects. *Circ Res*. 2016;118(9):1340-1347.
26. Stone J, Hangge P, Albadawi H, et al. Deep vein thrombosis: pathogenesis, diagnosis, and medical management. *Cardiovasc Diagn Ther*. 2017;7(Suppl 3):S276-S284.
27. Beckman MG, Hooper WC, Critchley SE, Ortel TL. Venous thromboembolism: a public health concern. *Am J Prev Med*. 2010;38(4 Suppl):S495-501.
28. White RH. The epidemiology of venous thromboembolism. *Circulation*. 2003;107(23 Suppl 1):I4-8.
29. Lowe GD. Common risk factors for both arterial and venous thrombosis. *Br J Haematol*. 2008;140(5):488-495.
30. Khan S, Dickerman JD. Hereditary thrombophilia. *Thromb J*. 2006;4:15.
31. Kalafatis M, Bertina RM, Rand MD, Mann KG. Characterization of the molecular defect in factor VR506Q. *J Biol Chem*. 1995;270(8):4053-4057.
32. van der Meer FJ, Koster T, Vandenbroucke JP, Briet E, Rosendaal FR. The Leiden Thrombophilia Study (LETS). *Thromb Haemost*. 1997;78(1):631-635.
33. Ridker PM, Miletich JP, Hennekens CH, Buring JE. Ethnic distribution of factor V Leiden in 4047 men and women. Implications for venous thromboembolism screening. *JAMA*. 1997;277(16):1305-1307.
34. Press RD, Bauer KA, Kujovich JL, Heit JA. Clinical utility of factor V leiden (R506Q) testing for the diagnosis and management of thromboembolic disorders. *Arch Pathol Lab Med*. 2002;126(11):1304-1318.

35. Poort SR, Rosendaal FR, Reitsma PH, Bertina RM. A common genetic variation in the 3'-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. *Blood*. 1996;88(10):3698-3703.
36. Leroyer C, Mercier B, Oger E, et al. Prevalence of 20210 A allele of the prothrombin gene in venous thromboembolism patients. *Thromb Haemost*. 1998;80(1):49-51.
37. Bosler D, Mattson J, Crisan D. Phenotypic Heterogeneity in Patients with Homozygous Prothrombin 20210AA Genotype. A paper from the 2005 William Beaumont Hospital Symposium on Molecular Pathology. The Journal of molecular diagnostics : JMD. Vol. 8: American Society for Investigative Pathology; 2006:420-425.
38. Rosendaal FR. Venous thrombosis: a multicausal disease. *Lancet*. 1999;353(9159):1167-1173.
39. Heijboer H, Brandjes DP, Buller HR, Sturk A, ten Cate JW. Deficiencies of coagulation-inhibiting and fibrinolytic proteins in outpatients with deep-vein thrombosis. *N Engl J Med*. 1990;323(22):1512-1516.
40. Manco-Johnson MJ, Marlar RA, Jacobson LJ, Hays T, Warady BA. Severe protein C deficiency in newborn infants. *J Pediatr*. 1988;113(2):359-363.
41. Bovill EG, Bauer KA, Dickerman JD, Callas P, West B. The clinical spectrum of heterozygous protein C deficiency in a large New England kindred. *Blood*. 1989;73(3):712-717.
42. Gandrille S, Borgel D, Sala N, et al. Protein S deficiency: a database of mutations--summary of the first update. *Thromb Haemost*. 2000;84(5):918.
43. Dykes AC, Walker ID, McMahon AD, Islam SI, Tait RC. A study of Protein S antigen levels in 3788 healthy volunteers: influence of age, sex and hormone use, and estimate for prevalence of deficiency state. *Br J Haematol*. 2001;113(3):636-641.
44. Mateo J, Oliver A, Borrell M, Sala N, Fontcuberta J. Laboratory evaluation and clinical characteristics of 2,132 consecutive unselected patients with venous thromboembolism--results of the Spanish Multicentric Study on Thrombophilia (EMET-Study). *Thromb Haemost*. 1997;77(3):444-451.
45. Mahasandana C, Veerakul G, Tanphaichitr VS, Suvatte V, Opartkiattikul N, Hathaway WE. Homozygous protein S deficiency: 7-year follow-up. *Thromb Haemost*. 1996;76(6):1122.
46. Faioni EM, Valsecchi C, Palla A, Taioli E, Razzari C, Mannucci PM. Free protein S deficiency is a risk factor for venous thrombosis. *Thromb Haemost*. 1997;78(5):1343-1346.
47. Whisstock J, Lesk AM, Carrell R. Modeling of serpin-protease complexes: antithrombin-thrombin, alpha 1-antitrypsin (358Met-->Arg)-thrombin, alpha 1-antitrypsin (358Met-->Arg)-trypsin, and antitrypsin-elastase. *Proteins*. 1996;26(3):288-303.
48. Mateo J, Oliver A, Borrell M, Sala N, Fontcuberta J. Increased risk of venous thrombosis in carriers of natural anticoagulant deficiencies. Results of the family studies of the Spanish Multicenter Study on Thrombophilia (EMET study). *Blood Coagul Fibrinolysis*. 1998;9(1):71-78.

49. Hirsh J, Piovella F, Pini M. Congenital antithrombin III deficiency. Incidence and clinical features. *Am J Med*. 1989;87(3B):34S-38S.
50. Demers C, Ginsberg JS, Hirsh J, Henderson P, Blajchman MA. Thrombosis in antithrombin-III-deficient persons. Report of a large kindred and literature review. *Ann Intern Med*. 1992;116(9):754-761.
51. Sarper N, Orlando C, Demirsoy U, Gelen SA, Jochmans K. Homozygous antithrombin deficiency in adolescents presenting with lower extremity thrombosis and renal complications: two case reports from Turkey. *J Pediatr Hematol Oncol*. 2014;36(3):e190-192.
52. Watkins WM. Biochemistry and Genetics of the ABO, Lewis, and P blood group systems. *Adv Hum Genet*. 1980;10:1-136, 379-185.
53. Watkins WM. The ABO blood group system: historical background. *Transfus Med*. 2001;11(4):243-265.
54. Gill JC, Endres-Brooks J, Bauer PJ, Marks WJ, Jr., Montgomery RR. The effect of ABO blood group on the diagnosis of von Willebrand disease. *Blood*. 1987;69(6):1691-1695.
55. Orstavik KH, Magnus P, Reisner H, Berg K, Graham JB, Nance W. Factor VIII and factor IX in a twin population. Evidence for a major effect of ABO locus on factor VIII level. *Am J Hum Genet*. 1985;37(1):89-101.
56. Wiggins KL, Smith NL, Glazer NL, et al. ABO genotype and risk of thrombotic events and hemorrhagic stroke. *J Thromb Haemost*. 2009;7(2):263-269.
57. Wolpin BM, Kabrhel C, Varraso R, et al. Prospective study of ABO blood type and the risk of pulmonary embolism in two large cohort studies. *Thromb Haemost*. 2010;104(5):962-971.
58. Deykin D. Warfarin therapy. 1. *N Engl J Med*. 1970;283(13):691-694.
59. Damus PS, Hicks M, Rosenberg RD. Anticoagulant action of heparin. *Nature*. 1973;246(5432):355-357.
60. Bara L, Billaud E, Gramond G, Kher A, Samama M. Comparative pharmacokinetics of a low molecular weight heparin (PK 10 169) and unfractionated heparin after intravenous and subcutaneous administration. *Thromb Res*. 1985;39(5):631-636.
61. Frydman AM, Bara L, Le Roux Y, Woler M, Chauliac F, Samama MM. The antithrombotic activity and pharmacokinetics of enoxaparine, a low molecular weight heparin, in humans given single subcutaneous doses of 20 to 80 mg. *J Clin Pharmacol*. 1988;28(7):609-618.
62. Stangier J. Clinical pharmacokinetics and pharmacodynamics of the oral direct thrombin inhibitor dabigatran etexilate. *Clin Pharmacokinet*. 2008;47(5):285-295.
63. Raghavan N, Frost CE, Yu Z, et al. Apixaban metabolism and pharmacokinetics after oral administration to humans. *Drug Metab Dispos*. 2009;37(1):74-81.
64. Forslund T, Wettermark B, Andersen M, Hjemdahl P. Stroke and bleeding with non-vitamin K antagonist oral anticoagulant or warfarin treatment in patients with non-valvular atrial fibrillation: a population-based cohort study. *Europace*. 2018;20(3):420-428.

65. Nieuwenhuis HK, Albada J, Banga JD, Sixma JJ. Identification of risk factors for bleeding during treatment of acute venous thromboembolism with heparin or low molecular weight heparin. *Blood*. 1991;78(9):2337-2343.
66. Lip GY, Keshishian A, Kamble S, et al. Real-world comparison of major bleeding risk among non-valvular atrial fibrillation patients initiated on apixaban, dabigatran, rivaroxaban, or warfarin. A propensity score matched analysis. *Thromb Haemost*. 2016;116(5):975-986.
67. Stevens SM, Woller SC, Bauer KA, et al. Guidance for the evaluation and treatment of hereditary and acquired thrombophilia. *J Thromb Thrombolysis*. 2016;41(1):154-164.
68. Kuruvilla M, Gurk-Turner C. A review of warfarin dosing and monitoring. *Proc (Bayl Univ Med Cent)*. 2001;14(3):305-306.
69. Kretz CA, Weyand AC, Shavit JA. Modeling Disorders of Blood Coagulation in the Zebrafish. Current pathobiology reports. Vol. 3; 2015:155-161.
70. Gregory M, Hanumanthaiah R, Jagadeeswaran P. Genetic analysis of hemostasis and thrombosis using vascular occlusion. Blood cells, molecules & diseases; 2002.
71. Rallapalli PM, Orengo CA, Studer RA, Perkins SJ. Positive selection during the evolution of the blood coagulation factors in the context of their disease-causing mutations. *Mol Biol Evol*. 2014;31(11):3040-3056.
72. Grzegorski SJ, Hu Z, Liu Y, et al. Disruption of the kringle 1 domain of prothrombin leads to late onset mortality in zebrafish. *Sci Rep*. 2020;10(1):4049.
73. Hu Z, Lavik K, Liu Y, et al. Loss of fibrinogen in zebrafish results in an asymptomatic embryonic hemostatic defect and synthetic lethality with thrombocytopenia. Journal of Thrombosis and Haemostasis; 2019.
74. Hu Z, Liu Y, Huarng MC, et al. Genome editing of factor X in zebrafish reveals unexpected tolerance of severe defects in the common pathway. Blood. Vol. 130: American Society of Hematology; 2017:666-676.
75. Liu Y, Kretz CA, Maeder ML, et al. Targeted mutagenesis of zebrafish antithrombin III triggers disseminated intravascular coagulation and thrombosis, revealing insight into function. Blood. Vol. 124: American Society of Hematology; 2014:142-150.
76. Vo AH, Swaroop A, Liu Y, Norris ZG, Shavit JA. Loss of fibrinogen in zebrafish results in symptoms consistent with human hypofibrinogenemia. PloS one. Vol. 8: Public Library of Science; 2013:e74682.
77. Weyand AC, Grzegorski SJ, Rost MS, et al. Analysis of factor V in zebrafish demonstrates minimal levels needed for early hemostasis. Blood Advances; 2019.
78. Khandekar G, Kim S, Jagadeeswaran P. Zebrafish thrombocytes: functions and origins. *Adv Hematol*. 2012;2012:857058.
79. Lang MR, Gihr G, Gawaz MP, Müller II. Hemostasis in Danio rerio: is the zebrafish a useful model for platelet research? Journal of thrombosis and haemostasis : JTH. Vol. 8; 2010:1159-1169.

80. Thattaliyath B, Cykowski M, Jagadeeswaran P. Young thrombocytes initiate the formation of arterial thrombi in zebrafish. *Blood*. 2005;106(1):118-124.
81. Lin H-F, Traver D, Zhu H, et al. Analysis of thrombocyte development in CD41-GFP transgenic zebrafish. Blood. Vol. 106: American Society of Hematology; 2005:3803-3810.
82. Jagadeeswaran P, Sheehan JP, Craig FE, Troyer D. Identification and characterization of zebrafish thrombocytes. *Br J Haematol*. 1999;107(4):731-738.
83. Ponczek MB, Gailani D, Doolittle RF. Evolution of the contact phase of vertebrate blood coagulation. *J Thromb Haemost*. 2008;6(11):1876-1883.
84. Robinson AJ, Kropatkin M, Aggeler PM. Hageman factor (factor XII) deficiency in marine mammals. *Science*. 1969;166(3911):1420-1422.
85. Howe K, Clark MD, Torroja CF, et al. The zebrafish reference genome sequence and its relationship to the human genome. Nature; 2013.
86. Rogers RL, Cridland JM, Shao L, Hu TT, Andolfatto P, Thornton KR. Tandem Duplications and the Limits of Natural Selection in Drosophila yakuba and Drosophila simulans. *PLoS One*. 2015;10(7):e0132184.
87. Cunningham F, Achuthan P, Akanni W, et al. Ensembl 2019. *Nucleic Acids Res*. 2019;47(D1):D745-D751.
88. White RJ, Collins JE, Sealy IM, et al. A high-resolution mRNA expression time course of embryonic development in zebrafish. *Elife*. 2017;6.
89. Pasquier J, Cabau C, Nguyen T, et al. Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database. *BMC Genomics*. 2016;17:368.
90. Hunter S, Apweiler R, Attwood TK, et al. InterPro: the integrative protein signature database. *Nucleic Acids Res*. 2009;37(Database issue):D211-215.
91. Hanumanthaiah R, Day K, Jagadeeswaran P. Comprehensive Analysis of Blood Coagulation Pathways in Teleostei: Evolution of Coagulation Factor Genes and Identification of Zebrafish Factor VIIi. Blood Cells, Molecules, and Diseases. Vol. 29; 2002:57-68.
92. Glasauer SM, Neuhauss SC. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol Genet Genomics*. 2014;289(6):1045-1060.
93. Braasch I, Gehrke AR, Smith JJ, et al. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat Genet*. 2016;48(4):427-437.
94. Postlethwait JH, Woods IG, Ngo-Hazelett P, et al. Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res*. 2000;10(12):1890-1902.
95. Ravi V, Venkatesh B. The Divergent Genomes of Teleosts. *Annu Rev Anim Biosci*. 2018;6:47-68.
96. Zhou RF, Liu Y, Wang YX, Mo W, Yu M. Coagulation factor III (tissue factor) is required for vascularization in zebrafish embryos. *Genet Mol Res*. 2011;10(4):4147-4157.

97. Bardeen CR. Abnormal development of toad ova fertilized by spermatozoa exposed to the Roentgen rays. *Journal of Experimental Zoology*. 1907;4(1):1-44.
98. Beale G. The discovery of mustard gas mutagenesis by Auerbach and Robson in 1941. *Genetics*. 1993;134(2):393-399.
99. Moresco EM, Li X, Beutler B. Going forward with genetics: recent technological advances and forward genetics in mice. *Am J Pathol*. 2013;182(5):1462-1473.
100. Acevedo-Arozena A, Wells S, Potter P, Kelly M, Cox RD, Brown SD. ENU mutagenesis, a way forward to understand gene function. *Annu Rev Genomics Hum Genet*. 2008;9:49-69.
101. Stottmann R, Beier D. ENU Mutagenesis in the Mouse. *Curr Protoc Hum Genet*. 2014;82:15 14 11-10.
102. Weiss J, Hurley LA, Harris RM, et al. ENU mutagenesis in mice identifies candidate genes for hypogonadism. *Mamm Genome*. 2012;23(5-6):346-355.
103. Vitaterna MH, King DP, Chang AM, et al. Mutagenesis and mapping of a mouse gene, Clock, essential for circadian behavior. *Science*. 1994;264(5159):719-725.
104. Driever W, Solnica-Krezel L, Schier AF, et al. A genetic screen for mutations affecting embryogenesis in zebrafish. Development. Vol. 123; 1996.
105. Kimmel CB. Genetics and early development of zebrafish. *Trends Genet*. 1989;5(8):283-288.
106. Solnica-Krezel L, Schier AF, Driever W. Efficient recovery of ENU-induced mutations from the zebrafish germline. *Genetics*. 1994;136(4):1401-1420.
107. Nusslein-Volhard C. The zebrafish issue of Development. *Development*. 2012;139(22):4099-4103.
108. Geisler R, Rauch GJ, Geiger-Rudolph S, et al. Large-scale mapping of mutations affecting zebrafish development. *BMC Genomics*. 2007;8:11.
109. Gilchrist E, Haughn G. Reverse genetics techniques: engineering loss and gain of gene function in plants. *Brief Funct Genomics*. 2010;9(2):103-110.
110. Stainier DYR, Raz E, Lawson ND, et al. Guidelines for morpholino use in zebrafish. *PLoS Genet*. 2017;13(10):e1007000.
111. Bill BR, Petzold AM, Clark KJ, Schimmenti LA, Ekker SC. A primer for morpholino use in zebrafish. *Zebrafish*. 2009;6(1):69-77.
112. Kikuta H, Kawakami K. Transient and stable transgenesis using tol2 transposon vectors. Methods in molecular biology (Clifton, NJ). Vol. 546; 2009:69-84.
113. Suster ML, Kikuta H, Urasaki A, Asakawa K, Kawakami K. Transgenesis in zebrafish with the tol2 transposon system. *Methods Mol Biol*. 2009;561:41-63.
114. Foley JE, Maeder ML, Pearlberg J, Joung JK, Peterson RT, Yeh JR. Targeted mutagenesis in zebrafish using customized zinc-finger nucleases. *Nat Protoc*. 2009;4(12):1855-1867.
115. Amacher SL. Emerging gene knockout technology in zebrafish: zinc-finger nucleases. *Brief Funct Genomic Proteomic*. 2008;7(6):460-464.

116. Ma AC, Chen Y, Blackburn PR, Ekker SC. TALEN-Mediated Mutagenesis and Genome Editing. *Methods Mol Biol*. 2016;1451:17-30.
117. Zu Y, Tong X, Wang Z, et al. TALEN-mediated precise genome modification by homologous recombination in zebrafish. *Nat Methods*. 2013;10(4):329-331.
118. Ablain J, Durand Ellen M, Yang S, Zhou Y, Zon Leonard I. A CRISPR/Cas9 Vector System for Tissue-Specific Gene Disruption in Zebrafish. Developmental Cell. Vol. 32; 2015:756-764.
119. Irion U, Krauss J, Nüsslein-Volhard C. Precise and efficient genome editing in zebrafish using the CRISPR/Cas9 system. Development (Cambridge, England). Vol. 141; 2014:4827-4830.
120. Jao L-E, Wente SR, Chen W. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. Proceedings of the National Academy of Sciences of the United States of America. Vol. 110; 2013:13904-13909.
121. Bondy-Denomy J, Pawluk A, Maxwell KL, Davidson AR. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*. 2013;493(7432):429-432.
122. Wu RS, Lam II, Clay H, Duong D, Deo RC, Coughlin SR. A Rapid Method for Directed Gene Knockout for Screening in G0 Zebrafish. Developmental Cell. Vol. 46: Elsevier Inc.; 2018:112-125.e114.
123. Varshney GK, Pei W, LaFave MC, et al. High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9. Genome research. Vol. 25; 2015:1030-1042.
124. Kimura Y, Hisano Y, Kawahara A, Higashijima S-i. Efficient generation of knock-in transgenic zebrafish carrying reporter/driver genes by CRISPR/Cas9-mediated genome engineering. Scientific Reports. Vol. 4: Nature Publishing Group; 2014:6545.
125. Li W, Zhang Y, Han B, et al. One-step efficient generation of dual-function conditional knockout and geno-tagging alleles in zebrafish. *Elife*. 2019;8.
126. Wierson WA, Welker JM, Almeida MP, et al. GeneWeld: a method for efficient targeted integration directed by short homology. *bioRxiv*. 2019:431627.
127. Klarin D, Busenkell E, Judy R, et al. Genome-wide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease. *Nat Genet*. 2019;51(11):1574-1579.
128. Lindstrom S, Wang L, Smith EN, et al. Genomic and transcriptomic association studies identify 16 novel susceptibility loci for venous thromboembolism. *Blood*. 2019;134(19):1645-1657.
129. Reitsma PH. How to identify new genetic risk factors for VTE? *Thromb Res*. 2009;123 Suppl 4:S22-24.
130. Crous-Bou M, Harrington LB, Kabrhel C. Environmental and Genetic Risk Factors Associated with Venous Thromboembolism. *Semin Thromb Hemost*. 2016;42(8):808-820.

131. Link KG, Stobb MT, Sorrells MG, et al. A mathematical model of coagulation under flow identifies factor V as a modifier of thrombin generation in hemophilia A. *J Thromb Haemost*. 2020;18(2):306-317.
132. Davie EW, Kulman JD. An overview of the structure and function of thrombin. Seminars in thrombosis and hemostasis. Vol. 32 Suppl 1: Copyright © 2006 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA.; 2006:3-15.
133. Wolberg AS. Determinants of fibrin formation, structure, and function. Current Opinion in Hematology; 2012.
134. Coughlin SR. Protease-activated receptors in hemostasis, thrombosis and vascular biology. Journal of Thrombosis and Haemostasis; 2005.
135. Rosendaal FR, Doggen CJM, Zivelin A, et al. Geographic distribution of the 20210 G to A prothrombin variant. Thrombosis and Haemostasis; 1998.
136. Peyvandi F, Duga S, Akhavan S, Mannucci PM. Rare coagulation deficiencies. Haemophilia. Vol. 8: John Wiley & Sons, Ltd (10.1111); 2002:308-321.
137. Blanchard RA, Furie BC, Jorgensen M, Kruger SF, Furie B. Acquired vitamin K-dependent carboxylation deficiency in liver disease. The New England journal of medicine; 1981.
138. Bristol A, Ratcliffe JV, Roth DA, Jacobs MA, Furie BC, Furie B. Biosynthesis of prothrombin: intracellular localization of the vitamin K-dependent carboxylase and the sites of gamma-carboxylation. Blood; 1996.
139. Kotkow KJ, Deitcher SR, Furie B, Furie BC. The second kringle domain of prothrombin promotes factor Va-mediated prothrombin activation by prothrombinase. Journal of Biological Chemistry; 1995.
140. Deguchi H, Takeya H, Gabazza EC, Nishioka J, Suzuki K. Prothrombin kringle 1 domain interacts with factor Va during the assembly of prothrombinase complex. Biochem J; 1997.
141. Friedmann AP, Koutychenko A, Wu C, et al. Identification and characterization of a factor Va-binding site on human prothrombin fragment 2. Scientific Reports. Vol. 9: Springer US; 2019:2436.
142. Sun WY, Witte DP, Degen JL, et al. Prothrombin deficiency results in embryonic and neonatal lethality in mice. Medical Sciences. Vol. 95; 1998:7597-7602.
143. Xue J, Wu Q, Westfield LA, et al. Incomplete embryonic lethality and fatal neonatal hemorrhage caused by prothrombin deficiency in mice. Proceedings of the National Academy of Sciences; 1998.
144. Dewerchin M, Liang Z, Moons L, et al. Blood coagulation factor X deficiency causes partial embryonic lethality and fatal neonatal bleeding in mice. Thrombosis and Haemostasis. Vol. 83; 2000:185-190.
145. Cui J, O'Shea KS, Purkayastha A, Saunders TL, Ginsburg D. Fatal haemorrhage and incomplete block to embryogenesis in mice lacking coagulation factor V. Nature. Vol. 384; 1996:66-68.
146. Mullins ES, Kombrinck KW, Talmage KE, et al. Genetic elimination of prothrombin in adult mice is not compatible with survival and results in spontaneous hemorrhagic events in both heart and brain. Blood; 2009.

147. Ansell J, Hirsh J, Poller L, Bussey H, Jacobson A, Hijlek E. The pharmacology and management of the vitamin K antagonists: The Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. Chest; 2004.
148. Dulley K, Zon LI. Zebrafish: a model system for the study of human disease. Current Opinion in Genetics & Development. Vol. 10: Elsevier Current Trends; 2000:252-256.
149. Jagadeeswaran P, Liu YC. Developmental expression of thrombin in zebrafish embryos: A novel model to study hemostasis. Blood Cells, Molecules and Diseases; 1997.
150. Jagadeeswaran P, Paris R, Rao P. Laser-induced thrombosis in zebrafish larvae: a novel genetic screening method for thrombosis. Methods in molecular medicine. Vol. 129; 2006:187-195.
151. Fish RJ, Di Sanza C, Neerman-Arbez M. Targeted mutation of zebrafish fga models human congenital afibrinogenemia. Blood; 2014.
152. Monson CA, Sadler KC. Inbreeding Depression and Outbreeding Depression Are Evident in Wild-Type Zebrafish Lines. Zebrafish; 2010.
153. O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. Nucleic Acids Research; 2016.
154. Jenny NS, Lundblad RL, Mann KG. Thrombin. In: Colman RW, Marder VJ, Clowes AW, George JN, Goldhaber SJ eds. Hemostasis and Thrombosis: Basic Principles and Clinical Practice; 2008:193-214.
155. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research. Vol. 32; 2004:1792-1797.
156. Reyon D, Tsai SQ, Khgayter C, Foden JA, Sander JD, Joung JK. FLASH assembly of TALENs for high-throughput genome editing. Nature Biotechnology; 2012.
157. Thisse C, Thisse B. High-resolution in situ hybridization to whole-mount zebrafish embryos. Nature Protocols; 2008.
158. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative CTmethod. Nature Protocols; 2008.
159. Rost MS, Grzegorski SJ, Shavit JA. Quantitative methods for studying hemostasis in zebrafish larvae. Methods in Cell Biology; 2016.
160. Iuchi I, Yamamoto M. Erythropoiesis in the developing rainbow trout, Salmo gairdneri irideus: histochemical and immunochemical detection of erythropoietic organs. Journal of Experimental Zoology; 1983.
161. Paffett-Lugassy N, Zon L. Analysis of hematopoietic development in the zebrafish. Methods in molecular medicine; 2005:171-198.
162. Orcutt SJ, Krishnaswamy S. Binding of substrate in two conformations to human prothrombinase drives consecutive cleavage at two sites in prothrombin. Journal of Biological Chemistry; 2004.
163. Kim PY, Nesheim ME. Further evidence for two functional forms of prothrombinase each specific for either of the two prothrombin activation cleavages. Journal of Biological Chemistry; 2007.

164. Reilly PA, Lehr T, Haertter S, et al. The effect of dabigatran plasma concentrations and patient characteristics on the frequency of ischemic stroke and major bleeding in atrial fibrillation patients: The RE-LY trial (Randomized Evaluation of Long-Term Anticoagulation Therapy). Journal of the American College of Cardiology; 2014.
165. Gulilat M, Tang A, Gryn SE, et al. Interpatient Variation in Rivaroxaban and Apixaban Plasma Concentrations in Routine Care. Canadian Journal of Cardiology; 2017.
166. Monroe DM, Hoffman M. A mouse bleeding model to study oral anticoagulants. Thrombosis Research; 2014.
167. Lu G, Deguzman FR, Hollenbach SJ, et al. A specific antidote for reversal of anticoagulation by direct and indirect inhibitors of coagulation factor Xa. Nature Medicine; 2013.
168. Berghmans S, Butler P, Goldsmith P, et al. Zebrafish based assays for the assessment of cardiac, visual and gut function - potential safety screens for early drug discovery. Journal of Pharmacological and Toxicological Methods; 2008.
169. Koster R, Sassen WA. A molecular toolbox for genetic manipulation of zebrafish. Advances in Genomics and Genetics; 2015.
170. Rupp RAW, Snider L, Weintraub H. Xenopus embryos regulate the nuclear localization of XMyoD. Genes and Development; 1994.
171. Day K, Krishnegowda N, Jagadeeswaran P. Knockdown of prothrombin in zebrafish. Blood Cells, Molecules, and Diseases; 2004.
172. Hanumanthaiah R, Thankavel B, Day K, Gregory M, Jagadeeswaran P. Developmental Expression of Vitamin K-Dependent Gamma-Carboxylase Activity in Zebrafish Embryos: Effect of Warfarin. Blood Cells, Molecules, and Diseases. Vol. 27: Academic Press; 2001:992-999.
173. Day KR, Jagadeeswaran P. Microarray analysis of prothrombin knockdown in zebrafish. Blood Cells, Molecules, and Diseases; 2009.
174. Reininger AJ. Function of von Willebrand factor in haemostasis and thrombosis. Haemophilia; 2008.
175. Bryckaert M, Rosa JP, Denis CV, Lenting PJ. Of von Willebrand factor and platelets. Cellular and Molecular Life Sciences; 2015.
176. Ruggeri ZM. Structure of von Willebrand factor and its function in platelet adhesion and thrombus formation. Best Practice and Research: Clinical Haematology; 2001.
177. Rossi A, Kontarakis Z, Gerri C, et al. Genetic compensation induced by deleterious mutations but not gene knockdowns. Nature; 2015.
178. Wu W, Bancroft JD, Suttie JW. Structural features of the kringle domain determine the intracellular degradation of under- -carboxylated prothrombin: Studies of chimeric rat/human prothrombin. Proceedings of the National Academy of Sciences; 2002.
179. Doyle MF, Mann KG. Multiple active forms of thrombin. IV. Relative activities of meizothrombins. Journal of Biological Chemistry; 1990.
180. Dasgupta SK, Thiagarajan P. Inhibition of thrombin activity by prothrombin activation fragment 1.2. Journal of Thrombosis and Thrombolysis; 2007.

181. Haffter P, Granato M, Brand M, et al. The identification of genes with unique and essential functions in the development of the zebrafish, Danio rerio. *Development*. 1996;123:1-36.
182. Zhang J, Talbot WS, Schier AF. Positional cloning identifies zebrafish one-eyed pinhead as a permissive EGF-related ligand required during gastrulation. *Cell*. 1998;92(2):241-251.
183. Fishman MC. Zebrafish genetics: the enigma of arrival. *Proc Natl Acad Sci U S A*. 1999;96(19):10554-10556.
184. Brownlie A, Donovan A, Pratt SJ, et al. Positional cloning of the zebrafish sauternes gene: a model for congenital sideroblastic anaemia. *Nat Genet*. 1998;20(3):244-250.
185. Shimoda N, Knapik EW, Ziniti J, et al. Zebrafish genetic map with 2000 microsatellite markers. *Genomics*. 1999;58(3):219-232.
186. Leshchiner I, Alexa K, Kelsey P, et al. Mutation mapping and identification by whole-genome sequencing. *Genome Res*. 2012;22(8):1541-1548.
187. Voz ML, Coppieters W, Manfroid I, et al. Fast homozygosity mapping and identification of a zebrafish ENU-induced mutation by whole-genome sequencing. PloS one. Vol. 7: Public Library of Science; 2012:e34671.
188. Papadopoulos N. Introduction to positional cloning. *Clin Exp Allergy*. 1995;25 Suppl 2:116-118.
189. Irion U, Frohnhofer HG, Krauss J, et al. Gap junctions composed of connexins 41.8 and 39.4 are essential for colour pattern formation in zebrafish. *Elife*. 2014;3:e05125.
190. Jager M, Schubach M, Zemojtel T, Reinert K, Church DM, Robinson PN. Alternate-locus aware variant calling in whole genome sequencing. *Genome Med*. 2016;8(1):130.
191. Patowary A, Purkanti R, Singh M, et al. A sequence-based variation map of zebrafish. *Zebrafish*. 2013;10(1):15-20.
192. Guryev V, Koudijs MJ, Berezikov E, et al. Genetic variation in the zebrafish. *Genome Res*. 2006;16(4):491-497.
193. Butler MG, Iben JR, Marsden KC, Epstein JA, Granato M, Weinstein BM. SNPfisher: tools for probing genetic variation in laboratory-reared zebrafish. *Development*. 2015;142(8):1542-1552.
194. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research. Vol. 20; 2010:1297-1303.
195. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England). Vol. 25; 2009:1754-1760.
196. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England). Vol. 25; 2009:2078-2079.
197. Cingolani P, Patel VM, Coon M, et al. Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. Frontiers in genetics. Vol. 3; 2012:35.
198. Smit A, Hufley,R, Green, P. RepeatMasker Open-4.0. 2015.

199. Cingolani P, Platts A, Wang LL, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly. Vol. 6:80-92.
200. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4(7):1073-1081.
201. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17(1):122.
202. Suurvali J, Whiteley AR, Zheng Y, Gharbi K, Leptin M, Wiehe T. The laboratory domestication of zebrafish: from diverse populations to inbred substrains. *Mol Biol Evol*. 2019.
203. Bauer DC, McMorran BJ, Foote SJ, Burgio G. Genome-wide analysis of chemically induced mutations in mouse in phenotype-driven screens. *BMC Genomics*. 2015;16:866.
204. Matsumura S, Fujita Y, Yamane M, Morita O, Honda H. A genome-wide mutation analysis method enabling high-throughput identification of chemical mutagen signatures. *Sci Rep*. 2018;8(1):9583.
205. Mooney MR, Davis EE, Katsanis N. Analysis of Single Nucleotide Variants in CRISPR-Cas9 Edited Zebrafish Exomes Shows No Evidence of Off-Target Inflation. *Front Genet*. 2019;10:949.
206. Patnaik MM, Moll S. Inherited antithrombin deficiency: a review. Haemophilia : the official journal of the World Federation of Hemophilia. Vol. 14; 2008:1229-1239.
207. Brandt JT. Plasminogen and tissue-type plasminogen activator deficiency as risk factors for thromboembolic disease. *Arch Pathol Lab Med*. 2002;126(11):1376-1381.
208. Kuo WT. Endovascular therapy for acute pulmonary embolism. *J Vasc Interv Radiol*. 2012;23(2):167-179 e164; quiz 179.
209. Grzegorski SJ, Hu Z, Liu Y, et al. Disruption of the kringle 1 domain of prothrombin leads to late onset mortality in zebrafish. bioRxiv; 2019:576140.
210. Sander JD, Maeder ML, Reyon D, Voytas DF, Joung JK, Dobbs D. ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic Acids Res*. 2010;38(Web Server issue):W462-468.
211. Hwang WY, Fu Y, Reyon D, et al. Heritable and precise zebrafish genome editing using a CRISPR-Cas system. *PLoS One*. 2013;8(7):e68708.
212. Megy K, Downes K, Simeoni I, et al. Curated disease-causing genes for bleeding, thrombotic, and platelet disorders: Communication from the SSC of the ISTH. *J Thromb Haemost*. 2019;17(8):1253-1260.
213. Harter K, Levine M, Henderson SO. Anticoagulation drug therapy: a review. The western journal of emergency medicine. Vol. 16: California Chapter of the American Academy of Emergency Medicine (Cal/AAEM); 2015:11-17.
214. Shirai K, Chaudhary UB. Use of low molecular weight heparin and aminocaproic acid in chronic DIC associated with prostate cancer--a case report. *ScientificWorldJournal*. 2007;7:753-755.

215. Cooper DL, Sandler AB, Wilson LD, Duffy TP. Disseminated intravascular coagulation and excessive fibrinolysis in a patient with metastatic prostate cancer. Response to epsilon-aminocaproic acid. *Cancer*. 1992;70(3):656-658.
216. Kroh HK, Tans G, Nicolaes GA, Rosing J, Bock PE. Expression of allosteric linkage between the sodium ion binding site and exosite I of thrombin during prothrombin activation. *J Biol Chem*. 2007;282(22):16095-16104.
217. Gajsiewicz JM, Morrissey JH. Structure-Function Relationship of the Interaction between Tissue Factor and Factor VIIa. *Semin Thromb Hemost*. 2015;41(7):682-690.
218. Butenas S. Tissue factor structure and function. *Scientifica (Cairo)*. 2012;2012:964862.
219. Drake TA, Morrissey JH, Edgington TS. Selective cellular expression of tissue factor in human tissues. Implications for disorders of hemostasis and thrombosis. *Am J Pathol*. 1989;134(5):1087-1097.
220. Grover SP, Mackman N. Tissue Factor: An Essential Mediator of Hemostasis and Trigger of Thrombosis. *Arterioscler Thromb Vasc Biol*. 2018;38(4):709-725.
221. Ender F, Freund A, Quecke T, et al. Tissue factor activity on microvesicles from cancer patients. *J Cancer Res Clin Oncol*. 2020;146(2):467-475.
222. Aberg M, Siegbahn A. Tissue factor non-coagulant signaling - molecular mechanisms and biological consequences with a focus on cell migration and apoptosis. *J Thromb Haemost*. 2013;11(5):817-825.
223. Ruf W, Yokota N, Schaffner F. Tissue factor in cancer progression and angiogenesis. *Thromb Res*. 2010;125 Suppl 2:S36-38.
224. Belting M, Ahamed J, Ruf W. Signaling of the tissue factor coagulation pathway in angiogenesis and cancer. *Arterioscler Thromb Vasc Biol*. 2005;25(8):1545-1550.
225. Dorfleutner A, Hintermann E, Tarui T, Takada Y, Ruf W. Cross-talk of integrin alpha3beta1 and tissue factor in cell migration. *Mol Biol Cell*. 2004;15(10):4416-4425.
226. Kocaturk B, Versteeg HH. Tissue factor-integrin interactions in cancer and thrombosis: every Jack has his Jill. *J Thromb Haemost*. 2013;11 Suppl 1:285-293.
227. Bugge TH, Xiao Q, Kombrinck KW, et al. Fatal embryonic bleeding events in mice lacking tissue factor, the cell-associated initiator of blood coagulation. Proceedings of the National Academy of Sciences of the United States of America. Vol. 93; 1996:6258-6263.
228. Carmeliet P, Mackman N, Moons L, et al. Role of tissue factor in embryonic blood vessel development. Nature. Vol. 383; 1996:73-75.
229. Hogan KA, Weiler H, Lord ST. Mouse models in coagulation. *Thromb Haemost*. 2002;87(4):563-574.
230. Collen D, Rosen ED, Chan JCY, et al. Mice lacking factor VII develop normally but suffer fatal perinatal bleeding. Nature. Vol. 390: Nature Publishing Group; 1997:290-294.

231. Pawlinski R, Tencati M, Holscher T, et al. Role of cardiac myocyte tissue factor in heart hemostasis. *J Thromb Haemost*. 2007;5(8):1693-1700.
232. Pawlinski R, Wang JG, Owens AP, 3rd, et al. Hematopoietic and nonhematopoietic cell tissue factor activates the coagulation cascade in endotoxemic mice. *Blood*. 2010;116(5):806-814.
233. Wang S, Reeves B, Sparkenbaugh EM, et al. Protective and detrimental effects of neuroectodermal cell-derived tissue factor in mouse models of stroke. *JCI Insight*. 2016;1(11).
234. Parry GC, Erlich JH, Carmeliet P, Luther T, Mackman N. Low levels of tissue factor are compatible with development and hemostasis in mice. *J Clin Invest*. 1998;101(3):560-569.
235. Labun K, Montague TG, Gagnon JA, Thyme SB, Valen E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res*. 2016;44(W1):W272-276.
236. Traver D, Paw BH, Poss KD, Penberthy WT, Lin S, Zon LI. Transplantation and in vivo imaging of multilineage engraftment in zebrafish bloodless mutants. *Nat Immunol*. 2003;4(12):1238-1246.
237. Jin SW, Beis D, Mitchell T, Chen JN, Stainier DY. Cellular and molecular analyses of vascular tube and lumen formation in zebrafish. *Development*. 2005;132(23):5199-5209.
238. Schaffner F, Versteeg HH, Schillert A, et al. Cooperation of tissue factor cytoplasmic domain and PAR2 signaling in breast cancer development. *Blood*. 2010;116(26):6106-6113.
239. Belting M, Dorrell MI, Sandgren S, et al. Regulation of angiogenesis by tissue factor cytoplasmic domain signaling. *Nat Med*. 2004;10(5):502-509.
240. Xue Y, Liu D, Cui G, et al. A 3D Atlas of Hematopoietic Stem and Progenitor Cell Expansion by Multi-dimensional RNA-Seq Analysis. *Cell Rep*. 2019;27(5):1567-1578 e1565.
241. Farnsworth DR, Saunders LM, Miller AC. A single-cell transcriptome atlas for zebrafish development. *Dev Biol*. 2020;459(2):100-108.
242. Gailani D, Renné T. Intrinsic pathway of coagulation and arterial thrombosis. Arteriosclerosis, thrombosis, and vascular biology. Vol. 27; 2007:2507-2513.
243. Wu Y. Contact pathway of coagulation and inflammation. Thrombosis journal. Vol. 13; 2015:17.
244. Muluk SC, Vorp DA, Severyn DA, Gleixner S, Johnson PC, Webster MW. Enhancement of tissue factor expression by vein segments exposed to coronary arterial hemodynamics. *J Vasc Surg*. 1998;27(3):521-527.
245. El-Brolosy MA, Kontarakis Z, Rossi A, et al. Genetic compensation triggered by mutant mRNA degradation. *Nature*. 2019;568(7751):193-197.