Genomic Investigation of Inherited and Induced Hematological Disturbances in Thoroughbred Racehorses

By

ANNA DAHLGREN
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Integrative Genetics and Genomics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Carrie Finno, DVM PhD DACVIM, Chair

Fern Tablin, VMD, PhD

Anita Oberbauer, PhD

Committee in Charge

2021

Table of Contents

Table of Figures and Tables

Dedication

To my family and friends who have encouraged me, the horses who have kept me smiling and sane, the horses who have donated samples to my projects, and my lab mates and collaborators for teaching me and supporting me throughout my thesis work. Thank you to all of you.

Abstract

Epistaxis and sudden death due to cardiac events are substantial concerns in the Thoroughbred racing industry and often cannot be traced to a specific cause. Hematologic health in Thoroughbreds can be influenced by both inherited and induced conditions. To date, there have been five inherited bleeding disorders identified in the horse that all directly or indirectly affect coagulation protein concentration or function. Four also occur in other species, but one, Atypical Equine Thrombasthenia (AET), has only been diagnosed in Thoroughbred horses. AET is caused by platelets that cannot be properly activated by thrombin stimulation and also do not bind fibrinogen as efficiently, preventing affected horses from clotting normally after vascular injury. In racing Thoroughbreds, inducible causes for changes in red blood cell mass also lead to hematologic changes through illicit doping with recombinant erythropoietin (rHuEPO). This thesis will therefore use genomic tools to investigate both inherited and inducible hematologic disturbances that occur in the Thoroughbred racehorse.

This thesis provides an overview of what is currently known about inherited equine bleeding disorders, identifies and investigates putative variants for AET, determines the optimal RNA sequencing library preparation for long non-coding RNA (lncRNA) studies in horses, and identifies transcriptomic markers of rHuEPO micro-dosing. The first chapter provides an overview of the current body of knowledge of inherited equine bleeding disorders. The second chapter describes a whole-genome variant analysis study that was performed using AET affected and control Thoroughbreds. Three associated variants that were to be located in or near genes expressed in platelets were identified for further study: *SEL1L* c.1810A>G p.Ile604Val, *AL355838.1:g.26447375_26448962del*, and *VIPAS39:g.22685398_22685470del*. Since the lncRNA *AL355838.1* was identified, we sought to determine the best RNA sequencing library preparation method in the third chapter. Poly-A^+ selection was determined to be better

overall for identifying lncRNA as compared to rRNA-depletion.

AL355838.1:g.26447375_26448962del was subsequently excluded as it was not expressed in platelets. Based on molecular and functional studies, neither *SEL1L* c.1810A>G p.Ile604Val or *VIPAS39:g.22685398_22685470del* were conclusively identified as causative or excluded as putative variants and remain priorities for further investigation into the cause of AET. Lastly, chapter four identifies three transcriptomic markers (*C13H16orf54, PUM2 and CHTOP)* of rHuEPO micro-dosing using RNA sequencing (RNA-seq) of dosed and control Thoroughbreds that were on an exercise protocol comparable to the workload of a racehorse. However, reverse transcription quantitative PCR was not sensitive enough to replicate the differences observed with RNA-seq and thus, these are not suitable biomarkers to develop a commercial test to detect illicit rHuEPO micro-doping in racehorses.

Future studies should focus on further clarifying the functional effects of *SEL1L* c.1810A>G p.Ile604Val and *VIPAS39:g.22685398_22685470del* in platelets to determine the causative variant for AET. Additional work is also required to identify alternative methods of quantifying mRNA that are feasible as a commercial test or investigating the corresponding protein concentrations of these candidate genes to determine if identification of illicit drug dosing can be performed at a protein level.

This thesis also includes an addendum on the characterization of a novel heritable degenerative axonopathy identified in Quarter Horse foals. The disease was first identified by veterinarians who noted the close relation of the affected foals. This study, along with the studies on AET and rHuEPO biomarkers, emphasizes the important role that genomics can play in veterinary medicine to improve equine wellness while highlighting the value of collaborations between veterinarians and researchers.

Chapter 1: Genetics of Equine Bleeding Disorders

Authors: Anna R. Dahlgren[1], Fern Tablin[2], Carrie J. Finno[1]

[1] Department of Population Health and Reproduction, School of Veterinary Medicine, University of California Davis, Davis, CA 95616

[2] Department of Anatomy, Physiology and Cell Biology, School of Veterinary Medicine, University of California Davis, Davis, CA 95616

Keywords: Horse, hemophilia A, von Willebrand disease, prekallikrein deficiency, Glanzmann's Thrombasthenia, Atypical Equine Thrombasthenia

Reference: Dahlgren AR, Tablin F, Finno CJ. Genetics of Equine Bleeding Disorders. *Equine Veterinary Journal*. **2021**; 53(1):30-37. Published 2021 Jan. doi: 10.1111/evj.13290

Abstract

Genetic bleeding disorders can have a profound impact on a horse's health and athletic career. As such, it is important to understand the mechanisms of these diseases and how they are diagnosed. These diseases include hemophilia A, von Willebrand disease, prekallikrein deficiency, Glanzmann's Thrombasthenia, and Atypical Equine Thrombasthenia. Exercise-induced pulmonary hemorrhage also has a proposed genetic component. Genetic mutations have been identified for hemophilia A and Glanzmann's Thrombasthenia in the horse. Mutations are known for von Willebrand disease and prekallikrein deficiency in other species. In the absence of genetic tests, bleeding disorders are typically diagnosed by measuring platelet function, vWF, and other coagulation protein levels and activities. For autosomal recessive diseases, genetic testing can prevent the breeding of two carriers.

Background

Clotting is an essential biologic process. Several genetic bleeding disorders have been identified in horses, affecting the coagulation cascade and platelet function. These disorders are typically heritable and therefore passed down from parent to offspring. Genetic mechanisms have been identified for some equine bleeding disorders while others remain unknown. This review is focused on the genetics of equine diseases affecting the coagulation cascade and platelet function.

Hemostasis testing

Hemostasis tests are the cornerstone of diagnosing bleeding disorders across species. To appreciate how bleeding disorders in horses are diagnosed, it is necessary to understand how several of these tests work. Platelet count is important as it can indicate an ongoing disease or infection if the count is higher or lower than the normal range (94-232 x 10^3/uL at Cornell Clinical Pathology Laboratory). In addition to assessing overall platelet number on a complete blood count, a basic large animal coagulation panel, which includes activated partial thromboplastin time (aPTT), prothrombin time (PT), and fibrinogen, should be evaluated. The first two quantify the time for blood to clot following the addition of standard amounts of agonists to plasma. Blood is drawn into a tube that contains citrate which acts as an anticoagulant by binding all the extracellular calcium. Plasma is separated by centrifugation and removed. An excess of calcium is added to the plasma to allow it to clot. To determine aPTT, an activating substance for Factor XII (e.g. silica, celite, kaolin, ellagic acid) is added to simulate the intrinsic (also known as the contact) pathway [1]. To determine PT, tissue factor is added to activate the extrinsic pathway [1]. To measure fibrinogen activity, thrombin is added to plasma and the time it takes for a clot to form is optically measured and compared to a reference interval [1]. The

amount of fibrinogen is also compared to a species-specific standard curve to quantify how much fibrinogen is present in a volume of blood [1]. It is important to also keep in mind that increased fibrinogen can also be indicative of inflammation, not just a bleeding disorder [2]. Some laboratories also include the test for fibrin degradation products (FDPs) or D-dimers that are the result of the breakdown of a clot. FDPs and D-dimers differ in that D-dimers include the smallest crosslinked dimers of fibrin degradation products, while FDP measures all degradation products. This is an important distinction because the crosslinking only occurs when a clot is formed while fibrinogen breakdown can happen in the absence of clot formation. As such, D-dimer levels are more informative of clotting. The tests for FDPs and D-dimers are similar, consisting of using latex particles covered with antibodies against FDPs or D-dimers [1]. If FDPs or D-dimers are present, they will bind to the antibodies, causing the latex particles to bind together [1]. The extent to which the latex particles aggregate is quantified [1]. FDP and D-dimers levels generally are not informative for diagnosing genetic bleeding disorders.

Additional hemostasis tests important for diagnosing bleeding disorders include protein level and activity assays. The most well-known are probably factor assays where coagulation factor levels are measured. The primary method of diagnosing hemophilia A in horses is measuring factor VIII (FVIII) coagulant activity (FVIII:C). In this assay, horse plasma is mixed in various standard quantities with plasma that is deficient in FVIII. The amount of time it takes for each mixture to clot is measured, plotted, and compared to a standard curve to determine the levels of FVIII present [1]. If below the reference interval (50–200% at Cornell Comparative Coagulation Laboratory), the horse is considered deficient. Another important protein for diagnosing bleeding disorders is von Willebrand Factor (vWF). Two tests are typically used to measure vWF levels and function. The total plasma vWF antigen (vWF:Ag) quantifies vWF using an enzyme-linked immunosorbent assay (ELISA). Anti-vWF antibodies are used to specifically bind to vWF. These antibodies are also conjugated to an enzyme that reacts with a

subsequently added substrate to produce color which is quantified. vWF levels can then be measured against a standard curve. A second test measuring vWF activity uses ristocetin cofactor (vWF:RCo). Ristocetin is an antibiotic that induces binding of vWF to platelet glycoprotein Ib-IX-V, resulting in platelet-platelet aggregation which is then measured with an aggregometer. The rate of aggregation can be used to determine the levels of vWF. High molecular weight vWF multimers are visualized using agarose gel electrophoresis which is important when diagnosing the subtype of vWD. The severity of signs and total level of vWF present are used to split vWD into three different types (types 1, 2, and 3). Additionally, type 2 segregates into four subtypes (A, B, M, N), depending on quantity of high molecular weight multimers and abnormal binding affinity.

A method to specifically investigate platelet function in whole blood is thromboelastography. This method evaluates at the changes in the viscoelastic properties of whole blood during aggregation and fibrinolysis. Whole blood is placed in a reaction cup that oscillates along with an agonist (ex. kaolin) and calcium [3]. As the clot forms, tension is applied to a wire connected to a pin in the cup [3]. This tension is translated into an output trace. From this trace, multiple variables can be calculated, including the time it takes for clot to start forming, fibrinogen concentration, and strength of clot [3]. Lastly, a controversial method for testing platelet function *in vivo* is measuring template bleeding time (TBT). Briefly, this is done via a standard size shallow incision on the horses' forelimb and paper is used to absorb the blood [4]. The time from incision to when the bleeding stops is measured [4]. There is no standardized reference range for horses, and it has been published that TBT has wide variability between healthy horses [5]. However, a horse that consistently has a highly prolonged TBT likely has a bleeding disorder.

The results from a basic coagulation panel as well as subsequent tests can provide directions for the next steps in a diagnosis and treatment plan.

Inherited equine disorders affecting the coagulation cascade

The coagulation cascade describes the series of physiological events, including enzymatic activation of protein and recruitment to sites of injury, within the vasculature leading to hemostasis. Inherited equine disorders affecting this cascade include hemophilia A and prekallikrein deficiency.

Hemophilia A: Hemophilia A, caused by mutations in the *F8* gene that lead to FVIII deficiency, leads to recurrent bleeding. Hematomas are also often reported [6–8]. Under physiological conditions, in response to tissue injury, thrombin becomes activated through the clotting cascade which in turn cleaves FVIII into its active form (FVIIIa). Then, FVIIIa acts as a co-factor to factor IXa (FIXa), activating factor X (FX) leading to the cleavage of prothrombin into more thrombin which has positive feedback on several coagulation factors. In horses affected with hemophilia A, a severe deficiency in FVIII results in a decreased ability to maintain this cascade, so a clot does not properly form.

In horses with hemophilia A, aPTT is typically prolonged, PT is normal, and fibrinogen levels are decreased. However, as noted above, inflammation can increase fibrinogen activity, thus this should be taken into account during diagnosis [2]. The primary method of diagnosis is measuring FVIII:C. If the horse is deficient for FVIII, hemophilia A is the likely cause of abnormal bleeding. Horses that are carriers for hemophilia A typically have lower levels of FVIII, sometimes even below the reference range [7]. However, they do not have bleeding problems.

The mode of inheritance of hemophilia A is X-linked recessive since *F8* is on the X-chromosome. Affected females have two non-functional copies and males have one non-

functional copy of the F8 gene. As such, hemophilia A is more common in males, with all of the published records in male horses [6,9–14]. Hemophilia A has been reported in Thoroughbreds [7,10,11,15,16], Standardbreds [9,14], Quarter Horses [12,13], an Arabian [8], and a Tennessee Walker [6]. Other types of hemophilia (B and C) where other coagulation factors are deficient have been identified in humans [17,18], dogs [19–23], and cats [24]; however, hemophilia A is the only type that has been conclusively identified in horses. The hemophilic Arabian was reported to have deficiencies of other factors, though not to the extent that is usually seen in hemophilic individuals. While hemophilia A is the most common genetic bleeding disorder in horses, the genetic cause has only recently begun to be investigated. A Tennessee Walker affected with hemophilia A was found to have a four base pair (bp) deletion (EquCab3.0 chrX:127,502,317–127,502,314delAACA) and two linked single nucleotide polymorphisms (EquCab3.0 chrX:127,502,303G>C, chrX:127,502,320G>A) in intron 1 of *F8* [6]. This horse did not have detectable FVIII protein, though the authors discuss that the antibody used was not horse-specific and likely could not detect very low protein levels. However, they also determined that exons 1-2 could not be amplified from *FVIII* mRNA, and it was hypothesized that the intron 1 variants affected splicing [6]. In people, hemophilia A is very heterogeneous, with almost 3,000 causative variants identified [25]. Thus, there are likely other undiscovered genetic causes for hemophilia A in the horse.

Prekallikrein deficiency: Prekallikrein deficiency is a rare blood disorder in the horse. In healthy animals, factor XII (FXII) binds to the damaged endothelial surface and auto-activates to FXIIa. FXIIa then activates prekallikrein into plasma kallikrein, an additional activator of FXII, resulting in a positive feedback loop. FXIIa also plays a role in activating other coagulation factors leading to clot formation. Without functional prekallikrein, FXII is not activated as efficiently and can lead

to abnormal bleeding. However, the phenotype can be quite subtle, as this is not the only pathway of coagulation factor activation.

Prekallikrein deficiency causes prolonged aPTT with normal PT and fibrinogen levels [26,27]. Prekallikrein levels are measured similarly to FVIII:C where patient plasma is mixed with prekallikrein-deficient plasma and the time to clot is measured. The times are compared to a standard curve to quantify the level of prekallikrein present.

Prekallikrein deficiency has only been identified in two horse families, a Belgian family and a miniature horse family [26,27]. The initial Belgian horses bled excessively following castration and had normal levels of factors VIII, IX, XI, XII as well as a platelet count within the reference range [26]. However, prekallikrein deficiency does not always have obvious signs as in the case of the miniature horse [27]. The miniature horse was initially examined due to a metatarsophalangeal joint varus deformity. No other physical abnormalities were identified. The authors collected blood and observed its failure to clot after 30 minutes, prompting further investigation and the identification of prekallikrein deficiency [27]. There was no history of abnormal bleeding [27]. A full sister to the miniature horse also had low levels of prekallikrein and no abnormal bleeding was reported [27]. Similarly, the Belgian that bled excessively had two full siblings with low levels of prekallikrein but no obvious coagulopathy [26]. This demonstrates that, similar to what is found in people, prekallikrein deficiency does not typically lead to a severe bleeding phenotype compared to the other hemostatic disorders. It may be important to note that neither of these papers include information about platelet function (ex. thromboelastography) or vWF testing, possibly due to the resources available at the time of publication. Thus, there could be more going on than prekallikrein deficiency resulting in the abnormal bleeding. While a genetic cause for prekallikrein deficiency has not been elucidated in the horse, causative variants have been identified in the *kallikrein B1* gene which encodes prekallikrein in people and dogs [28–32]. The mode of inheritance for prekallikrein deficiency

has only been investigated in people, where it was shown to be inherited as an autosomal recessive trait [33]. Prekallikrein deficiency is also difficult to identify, likely due to normal clotting *in vivo*.

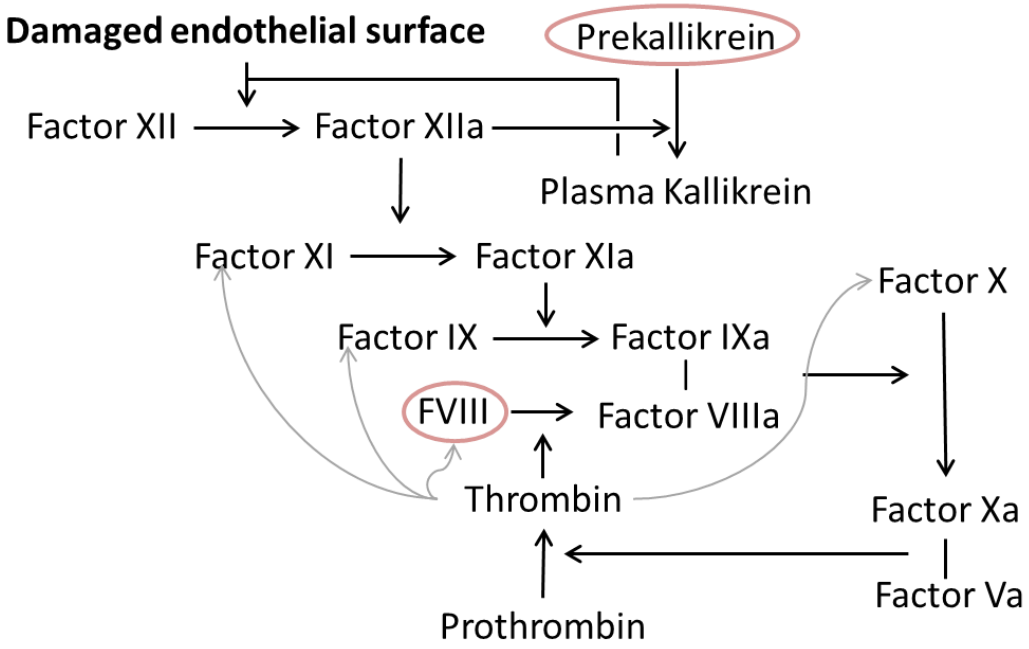


Figure 1.1: Schematic of coagulation factor signaling pathway affected by prekallikrein deficiency and hemophilia A. Red ovals indicate the step in the pathway where each disease affects coagulation signaling. Gray arrows indicate thrombin positive feedback.

Inherited equine disorders affecting platelet function

Platelets are essential to the clotting process as these anucleate cells activate, adhere to the subendothelium, bind fibrinogen, and aggregate to form a clot. Without normal platelet function, blood cannot clot properly. Platelet function disorders identified in the horse include von Willebrand Disease (vWD), Glanzmann's Thrombasthenia (GT), and Atypical Equine Thrombasthenia (AET).

Von Willebrand disease (vWD): vWD can result in epistaxis and abnormal bleeding after mild trauma or surgery. vWD is typically caused by mutations in the *von Willebrand Factor* gene, though acquired vWD has been reported rarely in humans and dogs [34–36]. The encoded protein (vWF) is a multimeric plasma glycoprotein that has several different roles. vWF binds to inactive FVIII to prevent degradation by activated protein C. vWF also binds to collagen that is exposed in damaged vascular sub-endothelium. Once vWF is bound to collagen, it can interact with platelets through the glycoprotein complex Ib-IX-V on the platelet membrane, which tethers the platelet to the site of injury. However, this interaction is not sufficient to make platelets firmly adhere to the site of injury. vWF binding to integrin $\alpha_{IIb}\beta_3$ provides a more stable platelet adhesion. Adenine diphosphate (ADP) and its receptors also play a necessary role in vWF-mediated adhesion. ADP is released from platelet dense granules as well as damaged red blood cells and can activate platelets by binding to receptors $P2Y_1$ and $P2Y_{12}$. Both receptors have been shown to be important in platelet aggregation after platelet adhesion to vWF bound to collagen [37]. ADP is also necessary for complete integrin $\alpha_{IIb}\beta_3$ activation and subsequent platelet adhesion [38]. Additionally, upregulated integrin $\alpha_{IIb}\beta_3$ binds to fibrinogen and engages in bidirectional signaling. While this signaling mediates platelet spreading, it also is regulated by glycoprotein VI (GPVI) and $\alpha_2\beta_1$. GPVI binds directly to collagen and is a major agonist for initial platelet activation. Integrin $\alpha_2\beta_1$ also is a major collagen receptor but requires inside-out activation. Another primary method of outside-in platelet activation is cleavage of a protease activated receptor on the platelet surface by thrombin which is quickly produced and activated by factor Xa, usually in a complex with factor V. Without sufficient levels or activity of vWF, platelets may not be able to protect FVIII from degradation, adhere to subendothelial collagen, or bind to platelet membrane proteins to tether platelets at the site of injury, leading to prolonged bleeding.

Horses affected with vWD may have mildly prolonged aPTT, but PT and fibrinogen are usually within normal ranges [39,40]. vWF:Ag and/or vWF:RCo will also be decreased depending on the type of vWD. When run on an agarose gel, the vWF may not be the correct size. vWD has been split into three types of which type 2 segregates into four subtypes to account for the variability of the disease. Type 1 describes a deficiency of vWF. In type 2 vWD, there is sufficient vWF, but it does not work properly. Subtype 2A indicates that the vWF multimers are not the correct size whereas subtype 2B indicates that vWF also is overly active. Subtype 2M describes vWF not being able to attach to the platelets. In subtype 2N, vWF does not bind properly to FVIII. Lastly, type 3 has very minimal levels of vWF.

vWD has been reported in two Quarter Horses [40,41] as well as a Thoroughbred mare and her colt [39]. The Quarter Horse filly [41] and the Thoroughbreds [39] were reported to have vWD type 2A characterized by decreased vWF activity and deficiency of high molecular weight vWF multimers. The Quarter Horse colt was hypothesized to have vWD type 1 due to signs of sufficient amounts of high molecular weight multimers [40]. However, formal subtyping was not performed [40]. No genetic mechanisms have been elucidated to date in the horse. There have been about 750 mutations identified in the human *vWF* gene associated with vWD [42] and at least four mutations associated with vWD in various dog breeds [43–46]. vWD has been observed to act in both a dominant and recessive mode of inheritance in humans depending on the specific mutation [42] and a recessive mode of inheritance in dogs [43–46]. Based on the few reports of vWD in horses, it does not appear to occur as frequently as it does in humans and dogs. To date, there are no studies on the mode of inheritance of vWD in the horse.

Glanzmann's Thrombasthenia (GT): Hallmarks of GT include epistaxis and prolonged bleeding. GT is caused by mutations leading to a loss of function or deficiency of integrin αIIbβ3. This integrin is composed of two subunits encoded by two different genes (*ITGA2B* and *ITGB3*). A

deleterious mutation in either gene leads to GT. Integrin αIIbβ3 is located on platelet membranes as well as on the platelet alpha granule membrane and binds fibrinogen, allowing activated platelets to aggregate and form a clot at the site of injury. With GT, there is either not enough integrin αIIbβ3 or αIIbβ3 does not properly bind to fibrinogen, inhibiting clot formation. In horses, there have only been reports of GT where there is an absence or very low expression of αIIbβ3.

The coagulation panel results in horses with GT remain within normal ranges. However, there is abnormal clot retraction and decreased platelet aggregation. Clot retraction is measured by drawing blood into a glass tube without anticoagulant and measuring the weight of the clot as well as the volume of remaining plasma upon complete contraction, compared with blood from control horses. It can also be assessed visually. Platelet aggregation, measured by an aggregometer, is determined by adding various agonists (e.g. adenine diphosphate (ADP), and collagen) to platelet-rich plasma. Thromboelastography has also been used as a diagnostic tool for GT. In one affected horse, clot formation time was increased and the clot was not as strong as a control horse [47]. While these tests can all suggest GT, they do not provide a conclusive diagnosis. Flow cytometry or western blot should be used to quantify αIIbβ3 protein levels with decreased protein expression indicating GT [47]. Ideally, a causative mutation identified in either *ITGA2B* and *ITGB3* would also confirm a GT diagnosis.

Two different mutations have been identified in the horse to cause GT. A missense mutation (Arg41Pro) in exon 2 of *ITGA2B* was identified in a Thoroughbred and an Oldenburg [48,49]. A 10 bp deletion in *ITGA2B* (EquCab3.0 chr11:19,247,983–19,247,992delCAGGTGAGGA) spanning the junction of exon 11 and intron 11 was identified as causative in a Peruvian Paso [50]. A GT-affected Quarter Horse was found to be a compound heterozygote for both of these variants [51]. Family studies have not been performed to look at the mode of inheritance of GT in horses, though GT has been observed to have a recessive

mode of inheritance in humans [52]. Based on the parental and sibling genotypes included in the literature, it appears that horses have the same mode of inheritance [48,51].

Atypical Equine Thrombasthenia (AET): AET is caused by abnormal platelet signaling leading to epistaxis and abnormal clotting after injury [53,54]. While pedigree analysis indicates that AET is heritable, the genetic cause has not yet been elucidated. The biochemical differences, however, have been thoroughly investigated. A number of proteins in the thrombin signaling pathway are decreased in quantity or activity in affected horses [53]. The biochemical hallmark of AET is that thrombin stimulated platelets from affected horses do not activate normally and bind fibrinogen less efficiently [55]. However, they do respond normally to other agonists (ADP) [55]. This is in contrast to GT where platelet aggregation is absent in response to all agonists.

An AET-affected horse was reported to have normal aPTT and slightly decreased PT [54]. Some factor activity was below the reference range, but this is more likely a result of long-term bleeding than indicative of factor deficiency. Additionally, the TBT was increased [54]. Currently, there is no readily available test that can be used to diagnose AET. It must be diagnosed via a fibrinogen-binding assay that uses thrombin as the stimulant after platelets have been confirmed to respond normally to ADP [55]. However, this test has high inter- and intra-individual variability, indicating the need for a more precise diagnostic tool (unpublished, Tablin lab).

The first horse diagnosed with AET was a mare and identified when she could not clot after pin firing. Subsequently, one of her offspring was diagnosed with the disease and her other offspring had inconclusive results with the fibrinogen-binding assay [55]. At this time, AET has only been identified in Thoroughbreds. Within a single breeding and training farm, AET had a prevalence of every one in 150 Thoroughbreds [56].

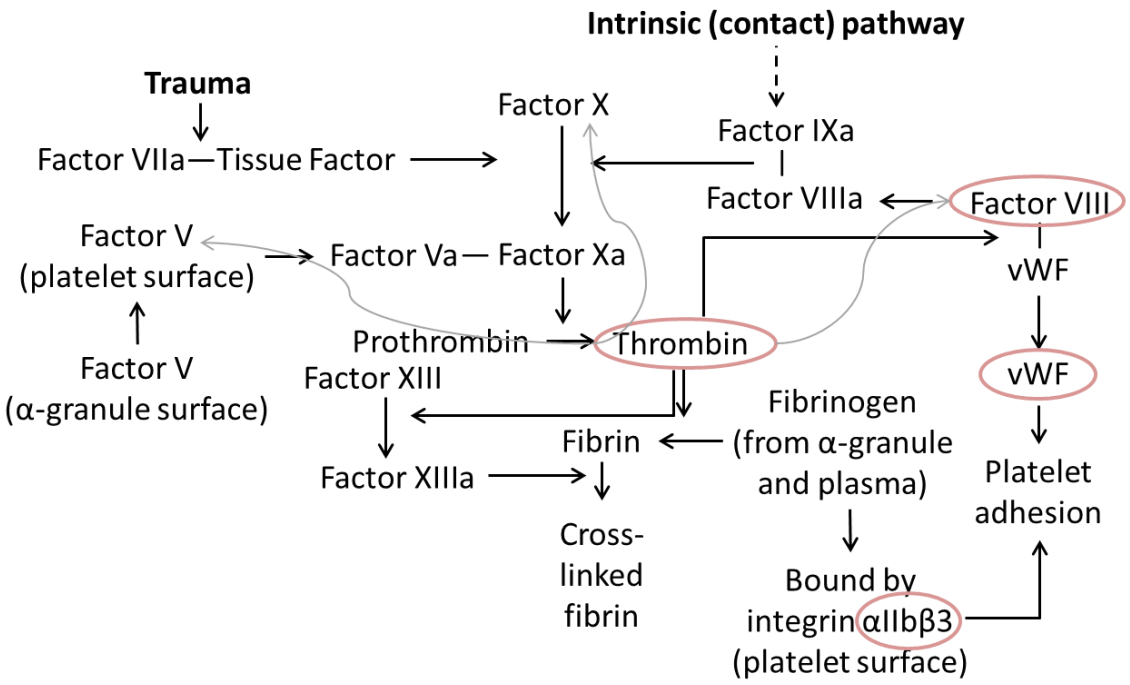


Figure 1.2: Schematic of coagulation factor and platelet signaling pathways affected by vWD, GT, and AET. Red ovals indicate the step in the pathway where each disease affects coagulation signaling. Gray arrows indicate thrombin positive feedback.

Other bleeding disorders

Exercise-induced pulmonary hemorrhage (EIPH): EIPH describes any hemorrhage in the lungs, typically resulting from intense physical exercise. It is generally thought to be caused by a cyclic pattern of increased pulmonary vascular pressures leading to thickening of pulmonary vein walls and vice versa. Together, this leads to pulmonary capillary failure and blood leaking into the interstitial and alveolar spaces in the lung. However, while it is not always consistent [57], some have shown in Thoroughbreds and Standardbreds that platelets in EIPH-affected horses may not be as responsive to ADP stimulation as compared to control horses [58,59]. Thus, there may be an unknown mechanism involving decreased platelet function in affected horses. Horses that already have decreased coagulation signaling or platelet dysfunction from an

inherited disorder are likely at a higher risk for hemorrhagic disorders, including EIPH,. EIPH has a high frequency in Thoroughbreds (44[60] –75%[61]), Standardbreds (87%[62]), and Quarter horses (62.3%[63]). The heritability of EIPH has been investigated in Thoroughbreds and found to be between 0.23 and 0.5, depending on the population and model used [64,65]. In these retrospective studies, EIPH was defined as presence of blood in nostrils. EIPH is also diagnosed via endoscopy and bronchoalveolar lavage as hemorrhaging in the lungs is not always severe enough to cause epistaxis. A small (*n*=6) study in Standardbreds indicated that PT and aPTT are all within the reference range [66]. One complication of EIPH diagnoses is that other potential causes for hemorrhaging should be excluded. Hemophilia A, vWD, GT, and AET have all been reported to cause epistaxis [6,8,39,48,50,54] and potentially contribute to the previously determined heritability of EIPH. While most of these bleeding disorders would likely prevent a horse from becoming a successful racehorse, AET has not been reported to have as severe of a phenotype. Thus, if truly at a high frequency [56], AET could potentially influence the calculated EIPH heritability. However, additional research is necessary to further elucidate the role genetics and platelet function play in EIPH.

Conclusion

Abnormal bleeding can arise from coagulation factor deficiency, decreased platelet function, strenuous exercise, and primary medical conditions. Genetic variants can lead to factor deficiencies, abnormal platelet function, and potentially increase risk of bleeding during exercise. To date, causative genetic mutations have only been identified for hemophilia A and GT in the horse. However, the genetic mechanisms for vWD and prekallikrein deficiency are known in humans and dogs. Similar future discoveries will therefore likely be made in horses. The genetic mechanism for AET remains unknown. These inherited bleeding diseases may

contribute to the risk for EIPH, and EIPH has been shown to potentially have a heritable component. However, additional studies are required to understand the interaction between genetics and EIPH.

Coagulation, platelet function, and molecular tests are currently used to diagnose bleeding disorders either through direct tests of the coagulation factors or by process of elimination. Identified carriers of autosomal recessive traits should not be bred to other carriers. Ideally, the mutation causing the bleeding disease in the affected offspring should be identified and a genetic test developed to assist in making breeding decisions.

Table 1.1: Summary of bleeding disorders with a genetic component

Disease	**Clinical signs**	**Diagnostic findings**
Hemophilia A	Recurrent bleeding, hematomas	Prolonged aPTT, highly deficient FVIII:C, decreased fibrinogen, normal PT, normal FDP/d-dimers
Prekallikrein deficiency	Abnormal bleeding, often asymptomatic	Prolonged aPTT, deficient prekallikrein levels, normal PT, normal FDP/d-dimers
von Willebrand disease	Epistaxis, abnormal bleeding after trauma or surgery	Decreased vWF:Ag, decreased vWF:RC, smaller multimers, mildly prolonged aPTT, normal PT, normal FDP/d-dimers
Glanzmann's Thrombasthenia	Epistaxis, prolonged bleeding	Abnormal clot retraction, decreased platelet aggregation to all agonists, increased clot formation time, decreased clot strength, decreased protein expression of αIIbβ3, normal PT, normal aPTT, normal FDP/d-dimers
Atypical Equine Thrombasthenia	Epistaxis, abnormal bleeding after injury	Increased TBT, abnormal platelet response to thrombin stimulation, slightly decreased PT, normal aPTT, normal FDP/d-dimers

Exercise-induced pulmonary hemorrhage	Epistaxis, poor performance	Visual inspection, endoscopy, bronchoalveolar lavage, normal PT, normal aPTT, normal FDP/d-dimers

References

[1] Van Cott EM, Laposata M. Coagulation. In: Jacobs DS, Oxley DK, DeMott WR, editors. Lab. Test Handb. 5th ed., Cleveland: Lexi-Comp; 2001, p. 327–58.

[2] Davalos D, Akassoglou K. Fibrinogen as a key regulator of inflammation in disease. Semin Immunopathol 2012;34:43–62. doi:10.1007/s00281-011-0290-8.

[3] Mendez-Angulo JL, Mudge MC, Couto CG. Thromboelastography in equine medicine: Technique and use in clinical research. Equine Vet Educ 2012;24:639–49. doi:10.1111/j.2042-3292.2011.00338.x.

[4] Alcott C, Wong D, Brockus C, Sponseller B. Hemostasis. Compend Contin Educ Pract Vet (Equine Ed) 2009;4:78–89.

[5] Segura D, Monreal L. Poor Reproducibility of Template Bleeding Time in Horses. J Vet Intern Med 2008;22:238–41. doi:10.1177/016555157900100507.

[6] Norton EM, Wooldridge AA, Stewart AJ, Cusimano L, Schwartz DD, Johnson CM, et al. Abnormal coagulation factor VIII transcript in a Tennessee Walking Horse colt with hemophilia A. Vet Clin Pathol 2016;45:96–102. doi:10.1111/vcp.12315.

[7] Littlewood JD, Bevan SA, Corke MJ. Haemophilia A (classic haemophilia, factor VIII deficiency) in a Thoroughbred colt foal. Equine Vet J 1991;23:70–2. doi:10.1111/j.2042-3306.1991.tb02719.x.

[8] Hinton M, Jones DRE, Lewis I., Thomson PE. A Clotting Defect in an Arab Colt Foal.

Equine Vet J 1977;9:1–3.

[9] Hutchins DR, Lepherd EE, Crook IG. A case of equine haemophilia. Aust Vet J 1967;43:83–7.

[10] Archer R., Allen BV. True haemophilia in horses. Vet Rec 1972;91:655–6.

[11] Nossel HL, Archer RK, MacFarlane RG. Equine haemophilia: report of a case and its response to multiple infusions in heterospecific AHG. Br J Haematol 1962;8:335–42.

[12] Henninger RW. Hemophilia A in two related quarter horse colts. J Am Vet Med Assoc 1988;193:91–4.

[13] Feldman BF, Giacopuzzi RL. Hemophilia A (factor Vlll deficiency) in a colt. Equine Pract 1982;4:24–30.

[14] Sanger VL, Mairs RE, Trapp AL. Hemophilia in a foal. J Am Med Assoc 1964;144:259–64.

[15] Archer RK. True haemophilia (haemophilia A) in a Thoroughbred foal. Vet Rec 1961;73:338–40.

[16] Mills JN, Bolton JR. Haemophilia A in a 3-year-old thoroughbred horse. Aust Vet J 1983;60.

[17] Aggeler PM, White SG, Glendening MB, Page EW, Leake TB, Bates G. Plasma Throinboplastin Component (PTC) Deficiency: A New Disease Resembling Hemophilia. Proc Soc Exp Biol Med 1952;79:692–4.

[18] Rosenthal RL, Herman Dreskin O, Rosenthal N. New Hemophilia-like Disease Caused by Deficiency of a Third Plasma Thromboplastin Factor. Proc Soc Exp Biol Med 1953;82:171–4.

[19] Evans JP, Brinkhous KM, Brayer GD, Reisner HM, High KA. Canine hemophilia B resulting from a point mutation with unusual consequences. Proc Natl Acad Sci U S A 1989;86:10095–9. doi:10.1073/pnas.86.24.10095.

[20] Mauser AE, Whitlark J, Whitney KM, Lothrop CDJ. A deletion mutation causes hemophilia B in Lhasa Apso dogs. Blood 1996;88:3451–5.

[21] Brooks MB, Gu W, Ray K. Complete deletion of factor IX gene and inhibition of factor IX activity in a labrador retriever with hemophilia B. J Am Vet Med Assoc 1997;211:1418–21.

[22] Gu W, Brooks MB, Catalfamo J, Ray J, Ray K. Two Distinct Mutations Cause Severe Hemophilia B in Two Unrelated Canine Pedigrees. Thromb Haemost 1999;82:1270–5.

[23] Brooks MB, Gu W, Barnas JL, Ray J, Ray K. A Line 1 insertion in the Factor IX gene segregates with mild hemophilia B in dogs. Mamm Genome 2004;14:788–95. doi:10.1007/s00335-003-2290-z.

[24] Goree M, Catalfamo JL, Aber S, Boudreaux MK. Characterization of the mutations causing hemophilia B in 2 domestic cats. J Vet Intern Med 2005;19:200–4. doi:10.1892/0891-6640(2005)19<200:COTMCH>2.0.CO;2.

[25] Centers for Disease Control and Prevention. CDC Hemophilia Mutation Project 2018.

[26] Geor RJ, Jackson ML, Lewis KD, Fretz PB. Prekallikrein deficiency in a family of Belgian horses. J Am Vet Assoc 1990;197:741–5.

[27] Turrentine MA, Sculley PW, Green EM, Johnson GS. Prekallikrein deficiency in a family of miniature horses. Am J Vet Res 1986;47:2464–7.

[28] François D, Trigui N, Leterreux G, Flaujac C, Horellou M-H, Mazaux L, et al. Severe prekallikrein deficiencies due to homozygous C529Y mutations. Blood Coagul Fibrinolysis

2007;18:283–6.

[29] Shigekiyo T, Fujino O, Kanagawa Y, Matsumoto T. Prekallikrein (PK) Tokushima : PK deficiency caused by a Gly401->Glu mutation. Int Soc Thromb Haemost 2003;1:1314–6.

[30] Lombardi AM, Sartori MT, Cabrio L, Fadin M, Zanon E. Severe prekallikrein (Fletcher factor) deficiency due to a compound heterozygosis (383Trp stop codon and Cys529Tyr). Thromb Haemost 2003;90:1040–5. doi:10.1160/TH03-05-0275.

[31] Jones DW, Russell G, Allford SL, Burdon K, Hawkins GA, Donald W, et al. Severe prekallikrein deficiency associated with homozygosity for an Arg94Stop nonsense mutation. Br J Haematol 2004;127:220–3. doi:10.1111/j.1365-2141.2004.05180.x.

[32] Okawa T, Yanase T, Miyama TS, Hiraoka H, Baba K. Prekallikrein Deficiency in a Dog. J Vet Med Sci 2011;73:107–11.

[33] Wuepper KD, T DRMH, Lacombe MJ. Flaujeac Trait Deficiency of Human Plasma Kininogen. J Clin Invest 1975;56:1663–72.

[34] Budde U, Bergmann F, Michiels JJ. Acquired von Willebrand Syndrome: Experience from 2 Years in a Single Laboratory Compared with Data from the Literature and an International Registry. Semin Thromb Hemost 2002;28:227–38.

[35] Rand JH, Budde U, Van Genderen PJJ, Mohri H, Meyer D, Rodeghiero F, et al. Acquired von Willebrand syndrome: Data from an international registry. Thromb Haemost 2000;84:345–9. doi:10.1055/s-0037-1614018.

[36] Dodds WJ. Von Willebrand’s Disease in Dogs. Mod Vet Pract 1984;65:681–6.

[37] Turner NA, Moake JL, Mclntire L V. Blockade of adenosine diphosphate receptors P2Y 12 and P2Y 1 is required to inhibit platelet aggregation in whole blood under flow. Blood 2001;98:3340–5. doi:10.1182/blood.V98.12.3340.

[38] Mazzucato M, Cozzi MR, Pradella P, Ruggeri ZM, De Marco L. Distinct roles of ADP receptors in von Willebrand factor-mediated platelet signaling and activation under high flow. Blood 2004;104:3221–7. doi:10.1182/blood-2004-03-1145.

[39] Rathgeber RA, Brooks MB, Bain FT, Byars DT. Von Willebrand Disease in a Thoroughbred Mare and Foal. J Vet Intern Med 2001;15:63–6. doi:10.1093/rheumatology/ket281.

[40] Laan TTJM, Goehring LS, Sloet van Oldruitenborgh- Oosterbaan MM. Von Willebrand's disease in an eight-day-old quarter horse foal. Vet Rec 2005;157:322–4. doi:10.1136/vr.157.11.322.

[41] Brooks M, Leith G, Allen A, Woods P, Benson R, Dodds W. Bleeding disorder (von Willebrand disease) in a quarter horse. J Am Vet Med Assoc 1991;198:114–6.

[42] Jong A De, Eikenboom J. Von Willebrand disease mutation spectrum and associated mutation mechanisms. Thromb Res 2017;159:65–75. doi:10.1016/j.thromres.2017.09.025.

[43] Venta PJ, Li J, Yuzbasiyan-gurkan V, Brewer GJ, Schall WD. Mutation Causing von Willebrand's Disease in Scottish Terriers. J Vet Intern Med 2000;14:10–9.

[44] Vos-Loohuis M, van Oost BA, Dangel C, Langbein-Detsch I, Leegwater PA. A novel VWF variant associated with type 2 von Willebrand disease in German Wirehaired Pointers and German Shorthaired Pointers. Animal 2017;48:493–6. doi:10.1111/age.12544.

[45] Rieger M, Schwarz HP, Turecek PL, Dorner F, Mourik JA Van, Mannhalter C. Identification of Mutations in the Canine von Willebrand Factor Gene Associated with Type III von Willebrand Disease. Thromb Haemost 1998;80:332–7.

[46] Kramer JW, Venta PJ, Klein SR, Cao Y, Schall WD, Yuzbasiyan-Gurkan V. A von

Willebrand ' s Factor Genomic Nucleotide Variant and Polymerase Chain Reaction Diagnostic Test Associated with Inheritable Type-2 von Willebrand ' s Disease in a Line of German Shorthaired Pointer Dogs. Vet Pathol 2004;41:221–8.

[47] Macieira S, Rivard G-E, Champagne J, Lavoie J, Be C. Glanzmann thrombasthenia in an Oldenbourg filly. Vet Clin Pathol 2007;36:204–8.

[48] Christopherson PW, Insalaco TA, Van Santen VL, Livesey L, Bourne C, Boudreaux MK. Characterization of the cDNA encoding alpha IIb and beta 3 in normal horses and two horses with Glanzmann thrombasthenia. Vet Pathol 2006;43:78–82. doi:10.1354/vp.43-1-78.

[49] Macieira S, Lussier J, Bédard C. Characterization of the cDNA and genomic DNA sequence encoding for the platelet integrin alpha IIB and beta III in a horse with Glanzmann thrombasthenia. Can J Vet Res 2011;75:222–7.

[50] Sanz MG, Wills TB, Christopherson P, Hines MT. Glanzmann thrombasthenia in a 17-year-old Peruvian Paso mare. Vet Clin Pathol 2011;40:48–51. doi:10.1111/j.1939-165X.2011.00289.x.

[51] Christopherson PW, Santen VL, Livesey L, Boudreaux MK. A 10-Base-Pair Deletion in the Gene Encoding Platelet Glycoprotein IIb Associated with Glanzmann Thrombasthenia in a Horse. J Vet Intern Med 2007;21:196–198. doi:10.1111/j.1939-1676.2007.tb02947.x.

[52] Reichert N, Seligsohn U, Ramot B. Clinical and Genetic Aspects of Glanzmann's Thrombasthenia in Israel. Thromb Haemost 1975;34:806–20.

[53] Norris JW, Pombo M, Shirley E, Blevins G, Tablin F. Association of Factor V Secretion with Protein Kinase B Signaling in Platelets from Horses with Atypical Equine Thrombasthenia. J Vet Intern Med 2015;29:1387–94. doi:10.1111/jvim.13595.

[54] Fry MM, Walker NJ, Blevins GM, Magdesian K, Tablin F. Platelet Function Defect in Thoroughbred Filly. J Vet Intern Med 2005:359–62. doi:10.1007/s11999-009-0725-x.

[55] Norris JW, Pratt SM, Auh J, Wilson SJ, Clutter D, Magdesian KG, et al. Investigation of a novel, heritable bleeding diathesis of Thoroughbred horses and development of a screening assay. J Vet Intern Med 2006;20:1450–6. doi:10.1111/j.1939-1676.2006.tb00765.x.

[56] Norris JW, Pratt SM, Hunter JF, Gardner IA, Tablin F. Prevalence of reduced fibronogen binding to platelets in a population of Thoroughbreds. Am J Vet Res 2007;68:716–21. doi:10.2460/ajvr.68.7.716.

[57] Weiss DJ, Mcclay CB, Smith CM, Rao GHR, White JG. Platelet Function in the Racing Thoroughbred: Implication for Exercise-Induced Pulmonary Hemorrhage. Vet Clin Pathol 1990;19:35–9.

[58] Johnstone IB, Viel L, Crane S, Whiting T. Hemostatic studies in racing standardbred horses with exercise-induced pulmonary hemorrhage. Hemostatic parameters at rest and after moderate exercise. Can J Vet Res 1991;55:101–6.

[59] Bayly WM, Meyers KM, Keck MT, Huston LJ, Grant BD. Effects of Exercise on the Hemostatic System of Thoroughbred Horses Displaying Post-Exercise Epistaxis. Equine Vet Sci 1983;3:191–3.

[60] Pascoe J, Ferraro G, Cannon J, Arthur R, Wheat J. Exercise-induced pulmonary hemorrhage in racing thoroughbreds: a preliminary study. Am J Vet Res 1981;42:703–7.

[61] Raphel C, Soma L. Exercise-induced pulmonary hemorrhage in Thoroughbreds after racing and breezing. Am J Vet Res 1982;43:1123–7.

[62] Lapointe J, Vrins A, Mccarvlll E. A survey of exercise-induced pulmonary haemorrhage in

Quebec Standardbred racehorses. Equine Vet Educ 1994;26:482–5.

[63] Hillidge C, Lane T, Johnson E, Asquith R. Preliminary investrigations of exercised-induced pulmonary hemorrhage in racing quarter horses. Equine Vet Sci 1984;4:21–3.

[64] Weideman H, Schoeman SJ, Jordaan GF. A genetic analysis of epistaxis as associated with EIPH in the Southern African Thoroughbred. South African J Anim Sci 2004;34:265–73.

[65] Velie BD, Raadsma HW, Wade CM, Knight PK, Hamilton NA. Heritability of epistaxis in the Australian Thoroughbred racehorse population. Vet J 2014;202:274–8. doi:10.1016/j.tvjl.2014.06.010.

[66] Perez-Moreno CI, Couëtil LL, Pratt SM, Ochoa-Acuña HG, Raskin RE, Russell MA. Effect of furosemide and furosemide-carbazochrome combination on exercise-induced pulmonary hemorrhage in standardbred racehorses. Can Vet J 2009;50:821–7.

Chapter 2: Identification of Putative Variants for Atypical Equine Thrombasthenia in Thoroughbred Horses

Authors: Anna R. Dahlgren[1], Fern Tablin[2], Carrie J. Finno[1]

[1] Department of Population Health and Reproduction, School of Veterinary Medicine, University of California Davis, Davis, CA 95616

[2] Department of Anatomy, Physiology and Cell Biology, School of Veterinary Medicine, University of California Davis, Davis, CA 95616

Key words: SEL1L, VIPAS39, platelets, clotting

Abstract

Atypical Equine Thrombasthenia (AET) is a heritable platelet disorder that, to date, has only been identified in the Thoroughbred breed. Aberrant cell signaling after thrombin stimulation prevents platelets from efficiently binding to fibrinogen, leading to abnormal clotting. Affected Thoroughbreds can experience repeated hematoma formation and prolonged bleeding after a vascular injury. Despite the negative effect on horse health and performance, the underlying etiology of AET is unknown, though pedigrees of affected horses indicate that AET is heritable. A whole-genome association study using five affected, one equivocal, and eleven control Thoroughbreds identified an associated haploblock on chromosome 24. Three putative variants in this block were determined to be located in or near genes known to be expressed in platelets: *SEL1L* c.1810A>G p.Ile604Val, *AL355838.1*:g.26447375_26448962del, and *VIPAS39*:g.22685398_22685470del. RT-PCR excluded the variant in *AL355838.1*, and no significant differential expression of *SEL1L* or *VIPAS39* was identified at the transcript level. However, there was a significant decrease in SEL1L protein expression in AET-affected platelets. It was determined that SEL1L is located intracellularly, but not in the α-granule, and

moves to the surface of the platelet upon activation with thrombin. Functional investigations using a collagen spreading assay yielded significant differences for horses that were homozygous for the *SEL1L* variant, but template bleeding times did not yield significant differences. Additional studies are required to definitively identify the resulting functional effects of the prioritized *SEL1L* c.1810A>G p.Ile604Val and *VIPAS39*:g.22685398_22685470del variants in conjunction with the AET phenotype.

Introduction

Atypical Equine Thrombasthenia (AET) is a heritable platelet disorder that, to date, has only been identified in the Thoroughbred breed [1]. One prevalence study has been performed and demonstrated that on a single breeding farm, one in every 150 Thoroughbreds (TBs) was affected [1]; however, the frequency of the disease in the general TB population is not known. AET inhibits platelets response to the physiological agonist thrombin, leading to abnormal bleeding after vascular injury [2,3]. Furthermore, this inhibition can result in prolonged bleeding and, in one reported case, epistaxis during racing [2]. A diagnostic method has been developed using flow cytometry to quantify the amount of fibrinogen bound by platelets after thrombin activation [3]. However, this diagnostic test is time-intensive and requires specialized equipment [3], which prevents affected horses from being readily and conclusively diagnosed on the racetrack or in the clinic. The potentially severe effects on horse health and racing career as well as the difficulty in diagnosing the disease, demonstrate the need to further understand the etiology of AET and develop a genetic test to identify affected horses.

Typically, when a horse is injured, clotting factors are stimulated in the damaged vessels, leading to the conversion of prothrombin to thrombin, the most potent activator of platelets. Other primary physiological agonists include adenosine diphosphate (ADP) and collagen. Once

platelets are activated, they bind fibrinogen, allowing them to aggregate and adhere to the site of injury. Thrombin also cleaves fibrinogen into fibrin, an insoluble protein that forms a net-like structure gathering platelets and other blood cells forming a stable blood clot [4].

Detailed biochemical assays have determined steps in this pathway that are aberrant in AET-affected horses' platelets. Coagulation factors were found to be within normal limits, demonstrating that AET is caused by changes in the platelets [2]. Platelet aggregometry studies, using physiological agonists, demonstrate that AET platelets respond normally to ADP, only respond to collagen with supplementary calcium, and have a minimal response to the most potent agonist, thrombin [2,3]. Further research has identified biochemical differences in the AET platelet response to thrombin. Subsequent to activation, the membrane phosphoinositol pathway is upregulated resulting in the eventual phosphorylation of protein kinase B (AKT) and release of platelet alpha granule contents. AET platelets have disruptions in the phosphoinositol activation pathway, as evidenced by decreased levels of phosphatidylinositol-3,4,5-trisphosphate 5-phosphatase 1 (SHIP1) in the platelet membrane, as well as the increased association of PIK3C2B, a phosphoinositide 3-kinase (PI3K) subunit with the membrane. These two changes likely impact the phosphorylation of AKT and downstream α-granule release [5]. The α-granule protein Factor V is critical for the acceleration of thrombin activation. AET platelet α-granules contain similar levels of Factor V to control platelets; however, the amount of Factor V released from AET platelet α-granules is decreased (**Fig. 2.1**) [5]. This suggests a failing in the exocytosis of α-granules by activated platelets.

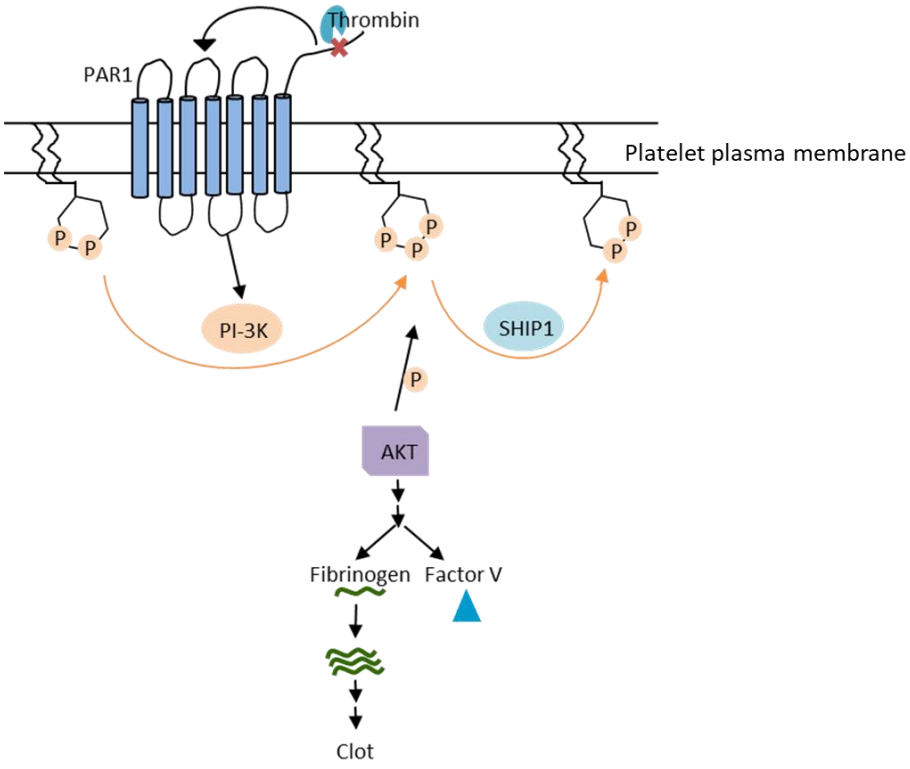


Figure 2.1. Summary of thrombin signaling pathway.

The goal of this study was to identify the causative variant for AET to guide the development of a genetic test. Due to the known biochemical changes, we initially hypothesized that a genetic variant within a gene coding for one of these proteins in the thrombin signaling pathway is responsible for AET. However, after a candidate gene approach using whole-genome sequencing data did not yield any variants that segregated with disease phenotype, we broadened our hypothesis to genetic variants in a gene that encodes a protein involved in the thrombin signaling pathway.

Methods

Horses and whole genome sequencing

Seventeen TBs were identified as AET-affected (*n*=5), control (*n*=11), or equivocal (*n*=1) in a previously published study [1]. Briefly, platelets were activated with thrombin and the amount of fibrinogen the platelets bound was quantified via flow cytometry and compared to the average of three unaffected horses [3]. The assay was performed using at least two technical replicates per horse. A strict cut-off of 35% of fibrinogen bound by platelets or less compared to control horses was used to definitively diagnose affected horses. Whole-genome sequencing was performed on these 17 horses using an Illumina HiSeq 2500 at a targeted 50x coverage with paired end 150 base-pair (bp) reads.

Variant analysis

Raw reads were trimmed with Trimmomatic [6], mapped to the EquCab3.0 reference genome [7] using Burrows-Wheeler Aligner (BWA-MEM) [8], and sorted by coordinate with samtools [9]. PCR duplicates were removed with Picard tools (http://broadinstitute.github.io/picard/). Integrative Genome Viewer (IGV) [10], was used to genotype all horses for the two variants shown to cause Glanzmann's Thrombasthenia [11–14]. Freebayes [15] was used to call single nucleotide variants (SNVs) and small insertions/deletions. Delly was used to identify larger structural variants [16]. The functional effects of the variants were predicted using SnpEff [17], and the variants were filtered by segregation using Fischer's Exact Test with SnpSift [17]. PlateletWeb (http://plateletweb.bioapps.biozentrum.uni-wuerzburg.de/plateletweb.php) was used to identify variants in and near genes known to be expressed in human platelets. KaryoploteR [18] was used to identify haploblocks among the variants that were $P<0.0001$.

Candidate gene approach

A list of candidate genes was developed from the previous biochemical work [5] on AET-affected platelets mentioned above (**Fig. 2.1**). Genes that encode known aberrant proteins (*AKT1, AKT2, F5, SHIP1, PIK3C2B*) were prioritized as candidate genes. Additionally, since

aberrant protein activity has only been identified within the thrombin signaling pathway, prothrombin (*F2*) and the thrombin receptor (*F2R*), which encodes protease-activated receptor 1 (*PAR1*) were included as candidate genes. Other subunits and classes of PI3K (*PIK3C2A, PIK3C2G, PIK3CA, PIK3CB, PIK3CG, PIK3R1, PIK3R4, PIK3R5, PIK3R6*) were also added. One kilobase of sequence before and after the genes was included to account for the 5' and 3' untranslated regions that could harbor regulatory variants. The functional effects of the variants were predicted using SnpEff [17], and the variants were filtered by segregation using Fisher's Exact Test with SnpSift [19].

RNA isolation and reverse transcription PCR

All identified variants across the whole genome were filtered by *P*-value (*P*<0.0001) with SnpSift [17], predicted effect on protein function (Freebayes variants only; HIGH and MODERATE), and expression in platelets which yielded three candidate variants – *SEL1L* c.1810A>G p.Ile604Val, *AL355838.1*:g.26447375_26448962del, and *VIPAS39:*g.22685398_22685470del. For all three candidate variants, reverse transcription PCR (RT-PCR) was performed to determine gene expression in equine platelets. Blood was drawn into acid citrate dextrose tubes from one horse that was wild type (WT) for all three variants. Platelets were separated from the whole blood by density centrifugation, treated with prostaglandin E1, and leukocyte depleted using an anti-equine F6 antibody generously provided by Dr. Jeff Stott. RNA was extracted from the platelets using a phenol-chloroform phase separation protocol, as previously described [20]. Turbo DNase (Thermo Fisher Scientific, Waltham, MA) was used to ensure there is no DNA contamination. Superscript III (Thermo Fisher Scientific, Waltham, MA) was then used to reverse transcribe 500 ng of RNA into cDNA. Primers that yield products spanning two exons were designed for *SEL1L* and the upstream lncRNA. To ensure that the primers (**Supplemental Table 2.S1**) amplified the targeted sequence, end-point PCR was performed on cDNA from testicular tissue, where the transcripts were known to be expressed [21].

Genotyping additional horses

A larger population of 114 TBs, including some of the WGS horses that had been phenotyped with the flow cytometry diagnostic method [1,3] were genotyped for the two remaining variants (*SEL1L* c.1810A>G p.Ile604Val and *VIPAS39:g.22685398_22685470del*). Polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) was used to genotype the *SEL1L* missense variant (**Supplemental Table 2.S1**). The restriction enzyme HpyCH4III cuts at the alternate allele allowing for distinction between the reference and alternate alleles. The *VIPAS39* deletion was large enough to visualize on a gel (**Supplemental Table 2.S1**). An ANOVA was performed to identify significant differences between genotypes with significance was set at $P<0.05$.

Table 2.S1. Primer Sequences			
Primer Target	**Coordinates**	**Forward**	**Reverse**
SEL1L transcript	chr24:26,074,490-26,076,942	CCTAGCACTCTGCAAATTAGGC	CAAGTTGGGTGAAGACATCTCG
VIPAS39 transcript	chr24:22,684,454-22,686,432	CAGGACCCTGAGAAACGAAA	CCAGGAGCGTGTAATGGTCT
lncRNA transcript	chr24:26,327,936-26,328,024	CTGCTGAGCTCTCTGGGTGT	CAGCTGGAGTGACTGAGCATT
SEL1L genotyping	chr24:26,080,981-26,082,166	CAACGTCAGTGCCAAATCC	ACATTGCCTTTCCAGCAGTC
VIPAS39 genotyping	chr24:22,685,271-22,685,559	TGAGCCACAGCCACTTGTTA	TTCCAGACCCGTCTAGCATC
ACTB transcript	chr13:4,382,919-4,383,011	AAGGAGAAGCTCTGCTATGTCG	GGGCAGCTCGTAGCTCTTC

Reverse transcription quantitative PCR (RT-qPCR) for SEL1L (*c.1810A>G p.Ile604Val) *and VIPAS39 (*g.22685398_22685470del) *variants

Reverse transcription quantitative PCR (RT-qPCR) was performed to determine if there was differential expression of *SEL1L or VIPAS39* since RT-PCR had excluded the *AL355838.1:g.26447375_26448962del* variant. Platelets and RNA were isolated as described for RT-PCR from horses genotyped for both *SEL1L* c.1810A>G p.Ile604Val (*n*=12 WT, 6 heterozygous, 2 homozygous) and *VIPAS39:g.22685398_22685470del* (*n*=16 WT, 3 heterozygous, 1 homozygous). Complementary DNA was run in triplicate using SYBR green on an AriaMx Real-Time PCR System (Agilent, Santa Clara, CA). The housekeeping gene *ACTB* was used for normalization (**Supplemental Table 2.S1**). The ΔΔCt values were calculated to determine differential transcript expression of *SEL1L* between healthy and AET-affected TBs. Due to the small sample size, a Kruskal-Wallis test was used to determine differential expression of *SEL1L* or *VIPAS39* between genotypes with significance set at $P<0.05$.

Flow cytometry

Since *SEL1L* c.1810A>G p.Ile604Val was a missense variant and predicted to have a moderate effect on protein function, it was prioritized for functional studies. Blood was drawn into acid citrate dextrose tubes from *n*=7 WT TBs, 5 TBs heterozygous, and 2 TBs homozygous for *SEL1L* c.1810A>G p.Ile604Val. Platelets were separated from whole blood by density centrifugation. Washed equine platelets were fixed in 1% paraformaldehyde, permeabilized in 0.1% NP-40 and labeled with a monoclonal SEL1L antibody conjugated to AlexaFluor488 (Santa Cruz Biotechnology, Dallas TX) and a monoclonal P-Selectin antibody conjugated to APC (Invitrogen, Waltham, MA). Samples were analyzed on a Beckkman Coulter FC500 flow cytometer and data analyzed with FlowJo. A Kruskal-Wallis test with multiple comparisons was

used to determine if any differential protein expression of SEL1L was significant, with a false discovery rate (FDR) correction for multiple testing. Significance was set at $P_{FDR}<0.05$.

Immunofluorescence

Washed platelets were obtained from three homozygous reference and three homozygous alternate TBs. Resting platelets (10 x 10^6 plt/mL) were incubated on poly L-lysine coverslips (37^{o}C/1hr). Platelets from the same horses (10 x 10^6 plt/mL) were activated with thrombin (bovine alpha-thrombin, Haematologic technologies, Essex Junction, VT) and then incubated on poly L-lysine coverslips (37^{o}C/1hr). The platelets were then fixed with 1% paraformaldehyde, permeabilized with 0.1% NP40, and blocked with 10% bovine serum albumin. Platelets were incubated with anti-P-selectin (BD Pharmaceuticals, diluted 1:10 in 5% goat serum), followed by a secondary antibody conjugated to an AF488 fluorophore (Thermofisher, Waltham, Massachusetts; diluted 1:20 in 5% goat serum). Both incubations were performed for 2 hours rocking at room temperature. Then, fixed platelets were incubated with a cocktail of anti-SEL1L and anti-CD42b (Abcam, Waltham, MA; diluted 1:200 and 1:10 respectively in 5% goat serum) rocking overnight at 4^{o}C. Lastly, the platelets were incubated with a cocktail of secondary antibodies conjugated to AF555 and AF405 (Thermofisher, diluted 1:300 and 1:10 respectively in 5% goat serum) for 2 hours rocking at room temperature. The stained platelets were imaged on a Leica TCS SP8 STED 3X (Leica Microsystems, Wetzlar, Germany).

Collagen spreading assay

Samples for the collagen spreading assay included *n*=11 WT, 5 heterozygous and 3 homozygous horses for the *SEL1L* c.1810A>G p.Ile604Val variant. These samples were used to evaluate the platelets' ability to adhere to and spread on collagen. Washed platelets (9 x 10^9 plt/L) were allowed to adhere to type 1 collagen coated coverslips (Electron Microscopy Sciences, Hatfield PA; 37^{o}C/2hrs), washed in buffer, and fixed in 1% paraformaldehyde (room

temperature/1hr). The mean platelet volume of the platelet rich plasma was determined. Coverslips were mounted, and platelets were imaged using an Olympus BX62 microscope. Platelet area was determined in ImageJ 1.51K (http://imagej.nih.gov/ij). The interpreter of the platelet area was blinded to the phenotypes and genotypes of the samples. A Kruskal-Wallis test with multiple comparisons was used to compare differences in platelet spreading, with an FDR correction for multiple testing. Significance was set at $P_{FDR}<0.05$.

Template bleeding times

Template bleeding times (TBTs) were preliminarily measured on 32 horses (18 WT, 11 heterozygous, 3 homozygous alternate) to validate the *SEL1L* missense variant with prolonged bleeding. Template bleeding times were conducted as previously described [22]. Briefly, a template system (MedexSupply, Passaic NJ) was used on a clipped area just distal to the accessory carpal bone on the caudomedial and caudolateral aspects of the forelimbs. A sphygmomanometer cuff was placed proximal to the carpus and inflated to 40 mmHg for 60 sec to achieve cutaneous venostasis. Blood flow from the incision was collected on a No. 40 absorbent filter paper at 30 sec intervals until no capillary flow can be absorbed. Bleeding times were performed at a standardized time each morning on both forelimbs over three consecutive days and mean bleeding times determined. Blood samples for a complete blood count and a basic coagulation panel (prothrombin time, activated partial thromboplastin time, and fibrinogen) were obtained prior to each template bleeding time determination to ensure that the bleeding times were not influenced by changes in platelet count or clotting defects. Due to horse availability, the 32 horses were split into three trial groups. A Kruskal-Wallis test identified a significant effect of group, so a follow-up study of 13 horses (10 WT, 3 homozygous alternate) was conducted over the same three-day period to confirm results.

Student's t-tests were used to determine differences in TBTs, platelet count, or basic coagulation variables, with significance set at $P<0.05$.

Results

Horses

As previously described [1], the five TBs identified as AET affected had 16.6% - 32.4% fibrinogen bound by platelets as compared to controls (**Table 2.1**). All unaffected TBs had >90% fibrinogen was bound by platelets compared to controls, except for horse #6, whose platelets bound an average of 74.3% fibrinogen. Horse #6 was included since her dam (horse 3) and full sibling (horse 4) were AET affected. A large degree of intra- and inter-individual variability was identified.

Table 2.1. Sample Phenotypes

Horse#	AET Phenotype	% fibrinogen bound compared to controls				Average % fibrinogen	Number of replicates (n)	Standard Deviation (for n > 2 trial only)
		Trial 1	*Trial 2*	*Trial 3*	*Trial 4*			
1	Affected	25.6	17.3	16.4	7.3	16.6	4	7.5
2		36.4	40.6	20.3		32.4	3	10.7
3		18.8	22.2			20.5	2	
4		13.8	28.2			21.0	2	
5		20.1	17.5			18.8	2	
6	Unaffected	49.9	83.4	89.6		74.3	3	21.3
7		93.1	117.7	82.0		97.6	3	18.3
8		107.9	90.4			99.2	2	
9		91.4	115.1	111.3		105.9	3	12.7
10		147.2	105.0			126.1	2	
11		142.8	159.0			150.9	2	
12		93.0	97.7			95.4	2	
13		115.9	113.1			114.5	2	
14		100.8	100.9			100.9	2	
15		133.4	138.3			135.8	2	
16		91.4	101.7	110.9		101.3	3	9.8
17		74.8	107.1	96.0		92.6	3	16.4

Candidate gene investigation

Since several proteins in the thrombin signaling pathway were expressed or functioned abnormally [5], the genes that encoded these 16 proteins were prioritized. Across these 16 genomic regions, 8,364 variants were identified (**Supplemental Table 2.S2**), but were not significantly associated with AET (P>0.05).

Table 2.S2. Number of variants identified in candidate genes	
Gene	Total variants
AKT1	251
AKT2	181
F2	166
F2R	64
F5	448
FGA, FGB	152
PIK3C2A	329
PIK3C2B	421
PIK3C2G	2598
PIK3CA	97
PIK3CB	688
PIK3CG	128
PIK3R1	374
PIK3R4	1144
PIK3R5, PIK3R6	602
SHIP1	721

Similarly, there were 133 variants across these 16 candidate genomic regions predicted to have a “MODERATE” effect, but not significantly associated with the phenotype (P>0.05). We next sought to identify any significantly associated variants using a 10% FDR, since a Bonferroni corrected P-value ($P_{bonferroni}=5.98 \times 10^{-6}$) is highly conservative [23]. However, even with a more generous cut-off, no variants were found to be associated with AET. We subsequently pursued a whole-genome association study.

Whole-genome association approach

Freebayes identified 18,367,554 variants and Delly identified 180,965 variants across the whole genome. The variants were filtered based on *P*-value and predicted effect on protein function. In order to minimize true negative associations, a cutoff of $P<0.0001$ was selected as this allows for half of the control horses to be heterozygous. This cutoff yielded 3,769 Freebayes variants (**Fig 2.2**) and 10 Delly variants. Approximately 70% (2,626 variants) of these Freebayes variants were in a ~6Mb region on chromosome 24 (21Mb - 27Mb; **Fig 2.2**). All 10 of the Delly variants were in this same region on chromosome 24.

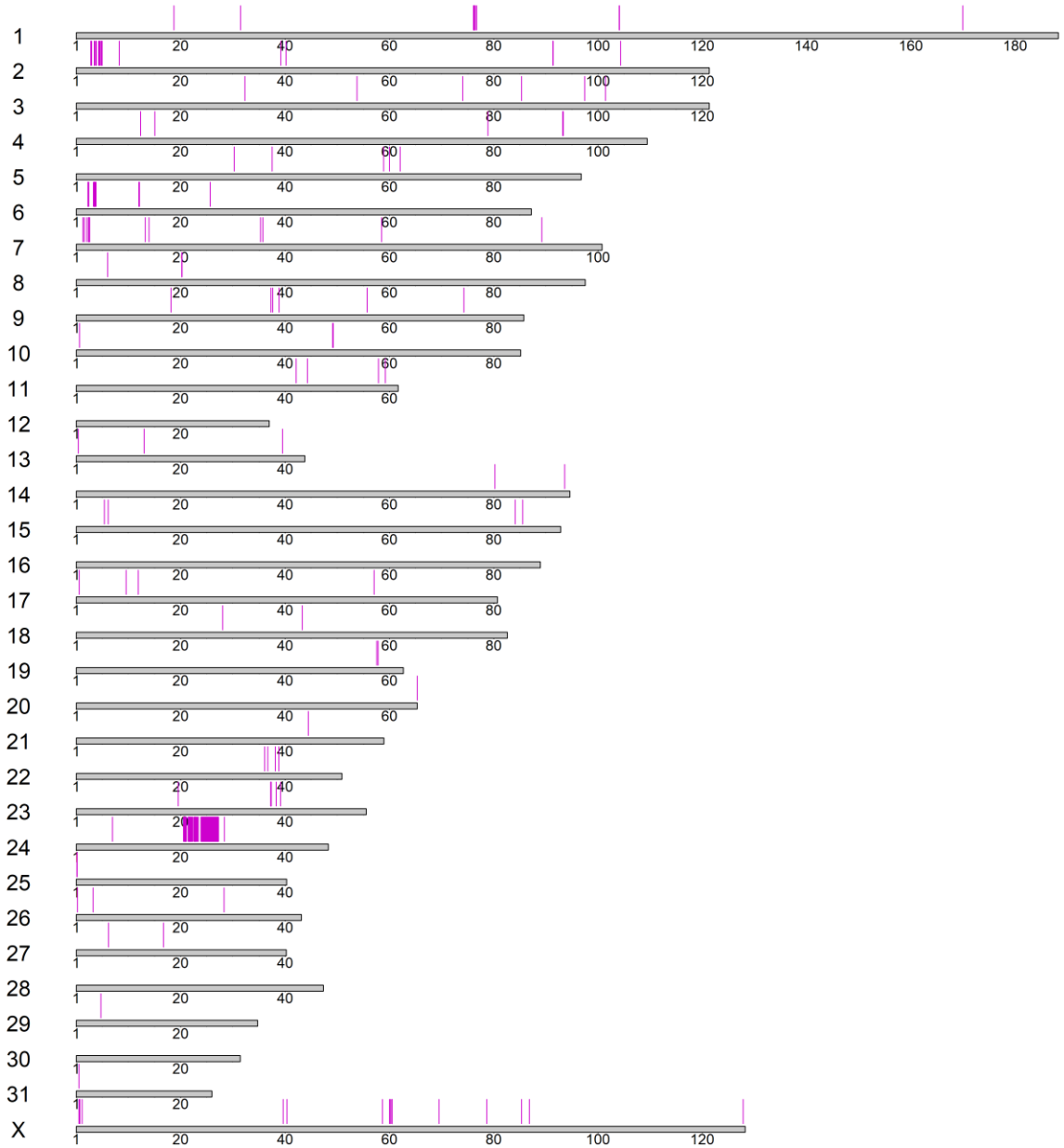


Figure 2.2. Representation of 3,769 Freebayes variants across genome. Each chromosome is represented by a grey bar and the magenta lines indicate variant position along chromosome. A large ~6Mb haploblock was identified on chromosome 24 (21Mb – 27Mb).

There were no variants predicted to have a HIGH effect, but five Freebayes variants were predicted to have a MODERATE effect on protein function (**Table 2.2**). Only one of these

variants was in a gene known to be expressed in platelets and interact with other platelet proteins (https://plateletweb.bioapps.biozentrum.uni-wuerzburg.de/plateletweb.php). The variant was a missense variant in exon 17 of *suppressor/enhancer of lin-12-like (SEL1L* c.1810A>G p.Ile604Val*; P*=7.63x10^{-6}*)*. The Delly variants were filtered solely based on *P*-value. There were 10 variants with *P*<0.0001, all of which were on chromosome 24. Two of them were in or near a gene expressed in platelets. One was a deletion in an uncharacterized lncRNA upsteam of *SEL1L* according TransMap (genome.ucsc.edu; *AL355838.1:g.26447375_26448962del*; *P*=7.63x10^{-6}) and the other was an intronic deletion in *VIPAS39* (g.22685398-22685470del ; *P*=3.74x10^{-5}). All three of these variants are in the associated region on chromosome 24, thereby prioritizing these variants for further investigation.

Table 2.2. Associated Variants with *P*<0.0001

Position	***P*-values**	**Nearest Gene**	**Expression and Interaction in Platelets** (PlateletWeb)
chr24:22162998	3.74E-05	*ANGEL1*	No
chr24:22642450	3.74E-05	*SAMD15*	No
chr24:22642470	3.74E-05	*SAMD15*	No
chr24:22888518	3.74E-05	*ALKBH1*	No
chr24:26081856	7.63E-06	*SEL1L*	Yes
chr24:21867050	3.74E-05	*ESRRB*	No
chr24:21999935	3.74E-05	*ESRRB*	No
chr24:22685398	3.74E-05	*VIPAS39*	Yes
chr24:24095086	9.73E-06	*NRXN3*	No
chr24:24460139	9.73E-06	*NRXN3*	No

chr24:24525130	9.73E-06	*NRXN3*	No
chr24:24709312	2.29E-05	*NRXN3*	No
chr24:25392066	3.74E-05	*CEP128*	No
chr24:25429259	3.74E-05	*CEP128*	No
chr24:26447375	7.63E-06	*SEL1L (AL355838.1)*	Yes

RT-PCR

To confirm that *SEL1L*, the upstream lncRNA, and *VIPAS39* were expressed in equine platelets, RNA was isolated from leukocyte-depleted platelets, converted to cDNA, and end-point PCR was performed, with cDNA from testicular tissue used as a positive control. The primers targeting the lncRNA upstream of *SEL1L* did not amplify a product from cDNA of equine platelets, but a product was amplified from cDNA of testicular tissue, indicating that the lncRNA is not expressed in platelets and thus is unlikely to play a role in the etiology of AET. Primers targeting *SEL1L* and *VIPAS39* amplified the expected products in cDNA from both platelets and testicular tissue, confirming that these genes are expressed in equine platelets and may play a role in the etiology of AET (**Supplemental Fig. 2.S1**).

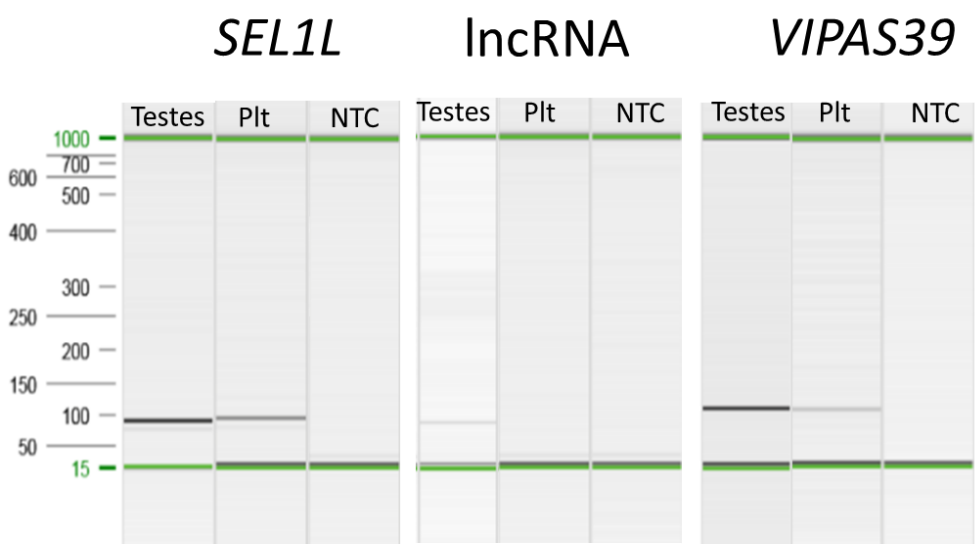


Figure 2.S1. Reverse transcription PCR showing transcript presence in testes (positive control) and platelets (Plt). *SEL1L* and *VIPAS39* transcripts are shown to be present in platelets while the lncRNA transcript is not. Expected size of *SEL1L* PCR product was 83bp, lncRNA PCR product was 89bp, and *VIPAS39* PCR product was 100bp.

Genotyping additional horses

A larger population of 114 TBs, including some of the WGS horses that had been phenotyped with the flow cytometry diagnostic method [1,3] were then genotyped for *SEL1L* c.1810A>G p.Ile604Val and *VIPAS39:g.22685398_22685470del* variants. All horses whose platelets bound <35% fibrinogen as compared to control horses were homozygous for *SEL1L* c.1810A>G p.Ile604Val (i.e. alternate allele (Alt/Alt). Horses that were heterozygous (Ref/Alt) and homozygous WT (Ref/Ref) for *SEL1L* c.1810A>G p.Ile604Val had platelets that bound similar amounts of fibrinogen. There was a significant difference in percent fibrinogen bound between the Alt/Alt horses and the other two genotypes (**Fig. 2.3A**; P=2.0x10^{-9}). For the *VIPAS39:g.22685398_22685470del* variant, not all horses that bound <35% fibrinogen as compared to control horses were homozygous (Alt/Alt), but there was a significant difference between the Alt/Alt horses and the other two genotypes (**Fig. 2.3B**; P=2.0x10^{-9}).

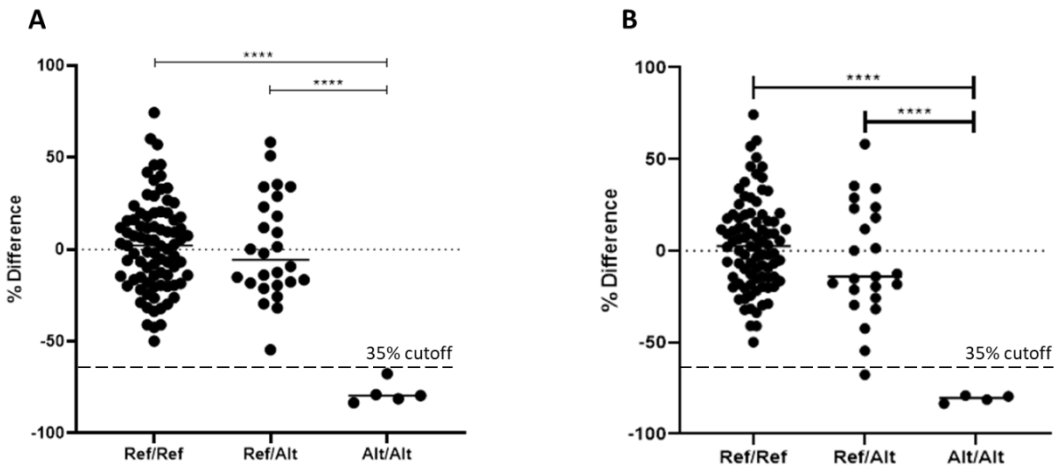


Figure 2.3. Genotypes of larger population of phenotyped Thoroughbreds for (A) *SEL1L* c.1810A>G p.Ile604Val and (B) *VIPAS39:g.22685398_22685470del*. Y-axis is % difference in fibrinogen binding as compared to control horses. The 35% cutoff line characterizing affected horses is indicated by the dashed line. ****P<0.0001

RT-qPCR

Because the *SEL1L* c.1810A>G p.Ile604Val variant was classified as a coding missense variant, platelet gene expression was assessed in platelets from horses that were WT (Ref/Ref; n=6), heterozygotes (Ref/Alt; n=3), and homozygotes (Alt/Alt; n=2) for this variant. There was no significant difference in gene expression between any of the three genotypes (P=0.74; **Fig. 2.4A**).

Additionally, since *VIPAS39:g.22685398_22685470del* is a large deletion in a region potentially important for splicing, platelet gene expression was also assessed from horses that were WT (Ref/Ref; n=16), heterozygotes (Ref/Alt; 3), and homozygote (Alt/Alt; 1). Primers were designed to target the exons that surround the intron where the deletion is located. Similar to *SEL1L*, there was no significant difference in gene expression between any of the three genotypes (P=0.64; **Fig. 2.4B**).

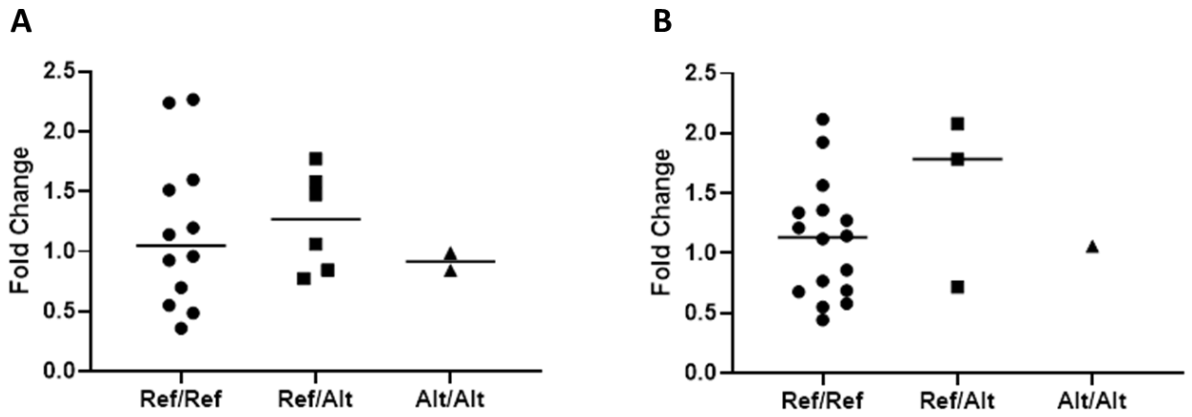


Figure 2.4. Quantitative reverse transcription PCR for (A) *SEL1L* and (B) *VIPAS39* by genotype. There were no significant differences in transcript expression between genotypes.

Flow Cytometry

Since point mutations can affect protein function without affecting mRNA expression [24], we proceeded with investigating SEL1L protein expression using flow cytometry. Platelets from WT (Ref/Ref; *n*=7), heterozygotes (Ref/Alt; 5), and homozygotes (Alt/Alt; 2) were assayed. There were significantly fewer platelets with intracellular SEL1L in horses that were homozygous for the *SEL1L* variant as compared to WT horses (*P*=0.022; **Fig. 2.5A**). There also was a decrease in platelets with intracellular SEL1L between Ref/Ref and Ref/Alt horses but it was not significant (*P*=0.08).

Since the function of SEL1L in platelets has not yet been evaluated, we next questioned whether SEL1L might play a role in platelet interactions following activation. Equine platelets were activated with thrombin, which has been shown to be an aberrant pathway in AET-affected horses [3,5]. While SEL1L relocated to the platelet surface upon activation (**Fig. 2.5B**) and there tended to be fewer platelets expressing surface SEL1L in AET-affected horses (*P*=0.09), a significant difference was not observed.

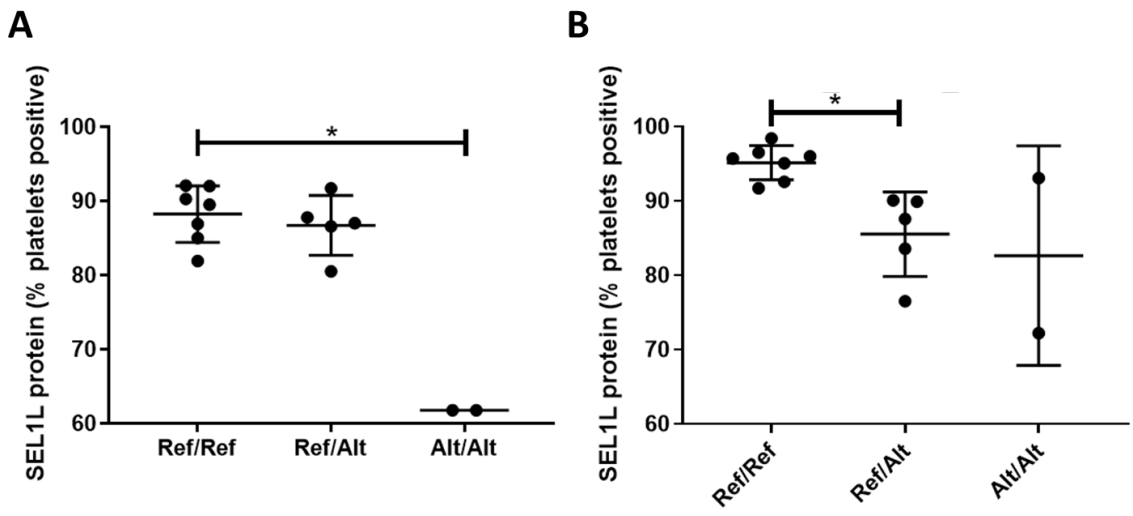


Figure 2.5. SEL1L protein expression from flow cytometry (A) in resting permeabilized platelets and (B) on thrombin-activated platelets. *P<0.05

Immunofluorescence

Since flow cytometry indicated that SEL1L moves to the surface upon thrombin activation, it was hypothesized that SEL1L would be located in the α-granule membrane. These granules exocytose their contents after activation, thus protein in the membrane subsequently becomes incorporated in the platelet membrane [4]. To investigate this hypothesis, permeabilized platelets from Ref/Ref TBs were fluorescently stained for p-selectin, a marker of α-granules, and SEL1L. The fluorescent markers of these two proteins did not overlap consistently, indicating that SEL1L does not localize to the α-granule (**Fig. 2.6A**). Stained Alt/Alt platelets looked similar to Ref/Ref platelets (**Supplemental Fig. 2.S2A**).

Since the flow cytometry indicated that SEL1L moves to the platelet membrane after activation with thrombin but immunofluorescence demonstrated that SEL1L does not localize to the α-granule, we used immunofluorescence to confirm that SEL1L does, in fact, relocate to the membrane. We stained activated platelets from Ref/Ref and Alt/Alt TBs for p-selectin and

SEL1L. SEL1L did localize to the platelet membrane of thrombin-activated platelets from all horses (**Fig. 2.6B and Supplemental Fig. 2.S2**), confirming the flow cytometry results.

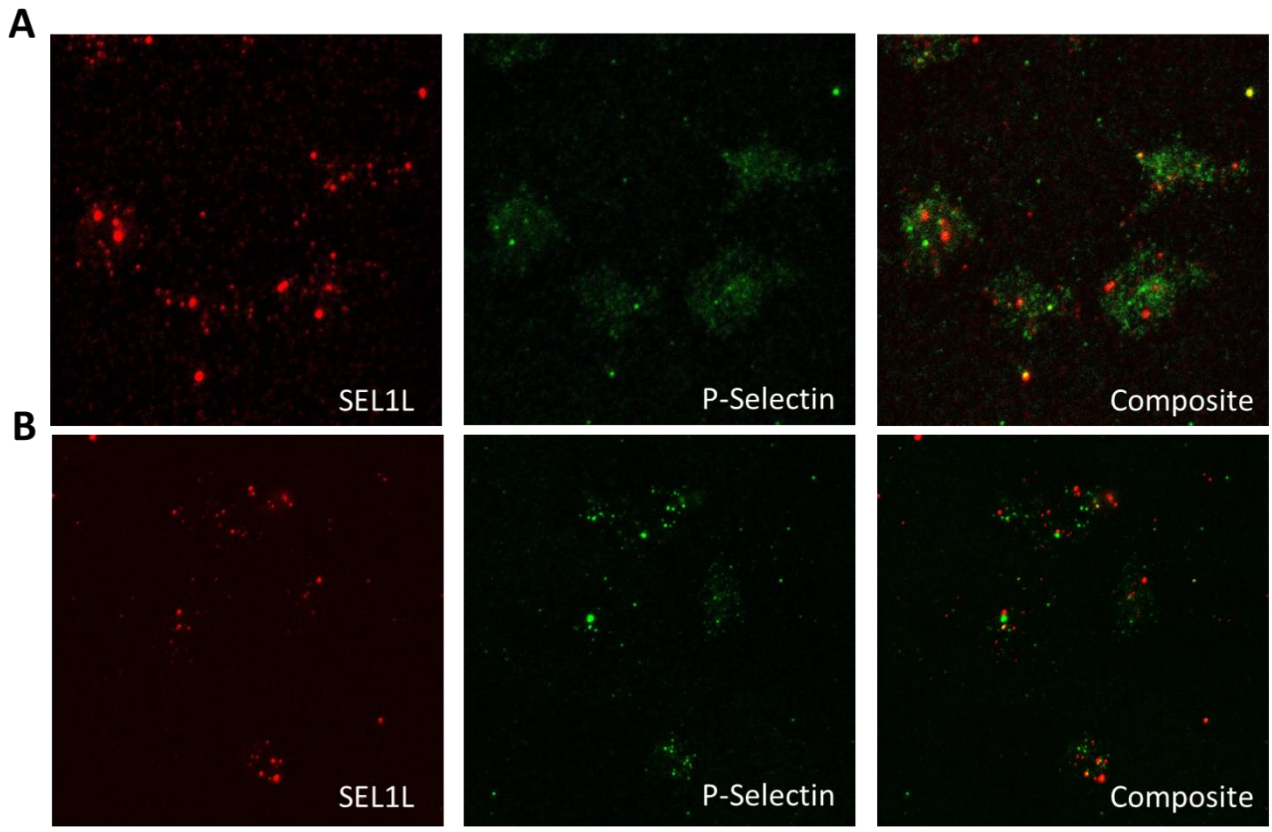


Figure 2.6. SEL1L (red) localization in relation to P-selectin (green) in (A) resting permeabilized platelets and (B) thrombin activated platelet surface of horses that are Ref/Ref for *SEL1L* c.1810A>G p.Ile604Val. SEL1L did not localize to the surface in resting permeabilized platelets (A) but did localize to the surface in thrombin-activated platelets (B).

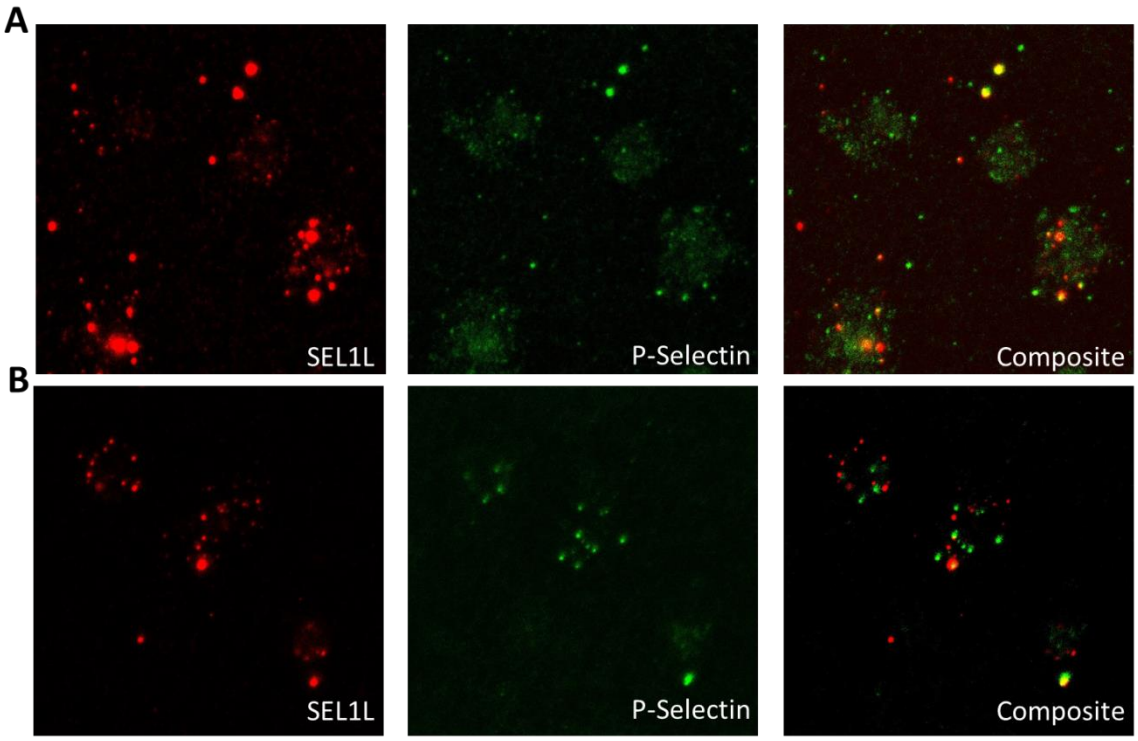


Figure 2.S2. SEL1L (red) localization in relation to P-selectin (green) in (A) resting permeabilized platelets and (B) thrombin activated platelet surface of horses that are Alt/Alt for *SEL1L* c.1810A>G p.Ile604Val. SEL1L did not localize to the surface in resting permeabilized platelets (A) but did localize to the surface in thrombin-activated platelets (B), similar to Ref/Ref horses.

Collagen Spreading Assay

To further elucidate the direct effect of *SEL1L* c.1810A>G p.Ile604Val on platelet function, a collagen spreading assay was performed. Platelets from WT (Ref/Ref; *n*=11), heterozygous (Ref/Alt; 5), and homozygous (Alt/Alt; 3) TBs were used. The mean platelet volume of the platelet rich plasma was determined to ensure there was no significant difference between the groups (*P*=0.45; **Supplemental Fig. 2.S3**).

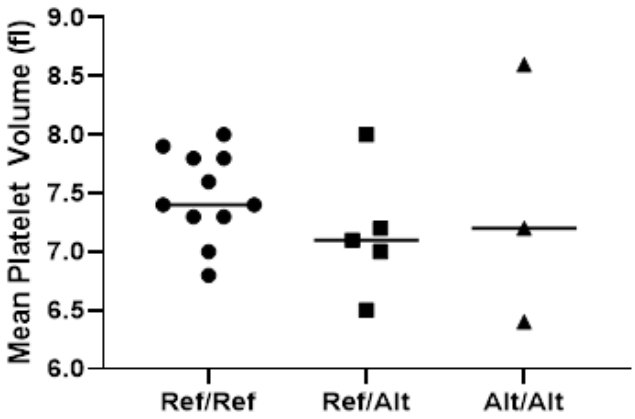


Figure 2.S3. Mean platelet volumes of platelet rich plasma for each individual horse separated by genotype for *SEL1L* c.1810A>G p.Ile604Val. There is no significant difference between genotypes.

The average area of the activated platelets was then determined for each horse. While the platelets from two of the horses that were homozygous for *SEL1L* c.1810A>G p.Ile604Val spread substantially less than the heterozygous or WT horses, the platelets from a third homozygous horse spread similarly to the other genotypes (*P*=0.45; **Supplemental Fig. 2.S4**). After investigations into this difference, we discovered that the AET affected outlier was ovariectomized (OVX), which has been shown to lead to increased platelet aggregation in response to both ADP and collagen [39]. The Ref/Ref horse with the highest median platelet area was also OVX, so these two horses were excluded from analysis. Comparison between the remaining horses showed that the decrease in median platelet area was significant between Ref/Ref and Alt/Alt horses (*P*=0.023; **Fig. 2.7**).

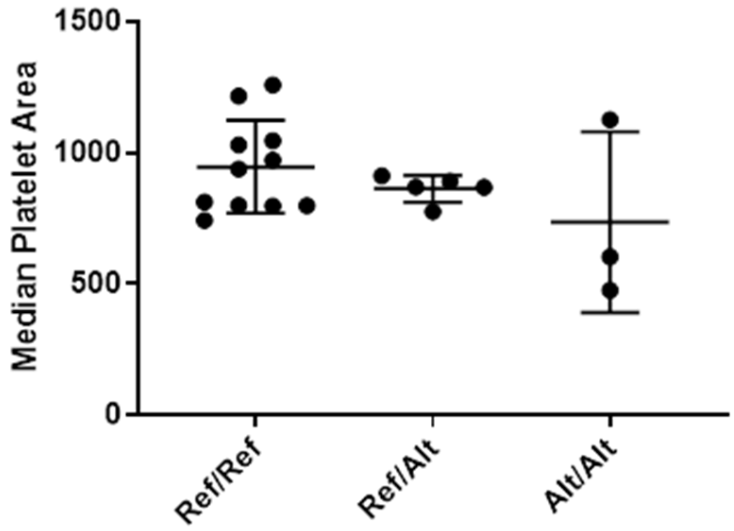


Figure 2.S4. Collagen spreading assay comparing median platelet area to genotype. There was no significant difference between homozygous Ref and Alt genotypes.

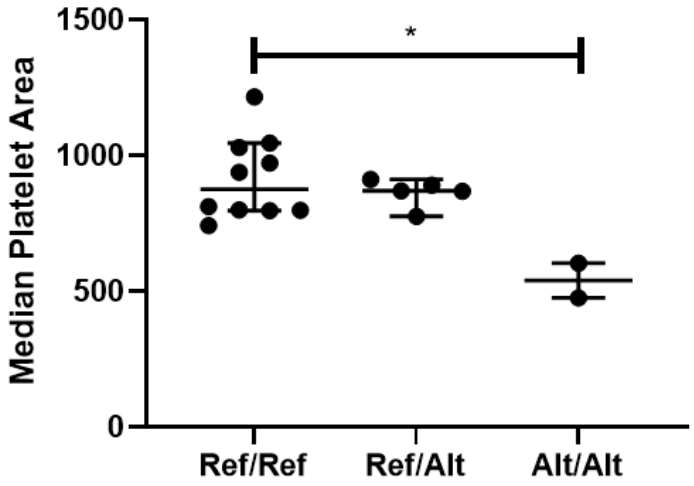


Figure 2.7. Collagen spreading assay comparing median platelet area to genotype, excluding OVX mares. There was a significant difference between homozygous Ref and alternate genotypes. *P<0.05

Template Bleeding Time

Next, we investigated if horses that were homozygous (Alt/Alt) for *SEL1L* c.1810A>G p.Ile604Val would have a longer template bleeding time (TBT) than heterozygous (Ref/Alt) or

WT (Ref/Ref) horses. A previous study highlighted the high degree of inter- and intra- horse variability in the TBT assay [25]. However, in the initial paper describing AET, the index case had a TBT of >120 min [2]. Thus, we hypothesized that horses that were Alt/Alt for *SEL1L* c.1810A>G p.Ile604Val would have a substantially longer TBT than Ref/Ref or Ref/Alt horses. To fully evaluate each horse's coagulation parameters prior to the study onset, platelet count, prothrombin time, activated partial thromboplastin time, and fibrinogen were measured. There were no differences in clotting times or platelet and fibrinogen count between groups (**Supplemental Fig. 2.S5A**). Due to horse availability, TBT trials were conducted in three batches. A significant effect of batch, or group, (*P* =0.0004; **Supplemental Fig. 2.S5B**) was identified, with the second group having the highest average TBT. This group consisted of TBs that were on an exercise protocol while the horses in the other two groups were not regularly exercised.

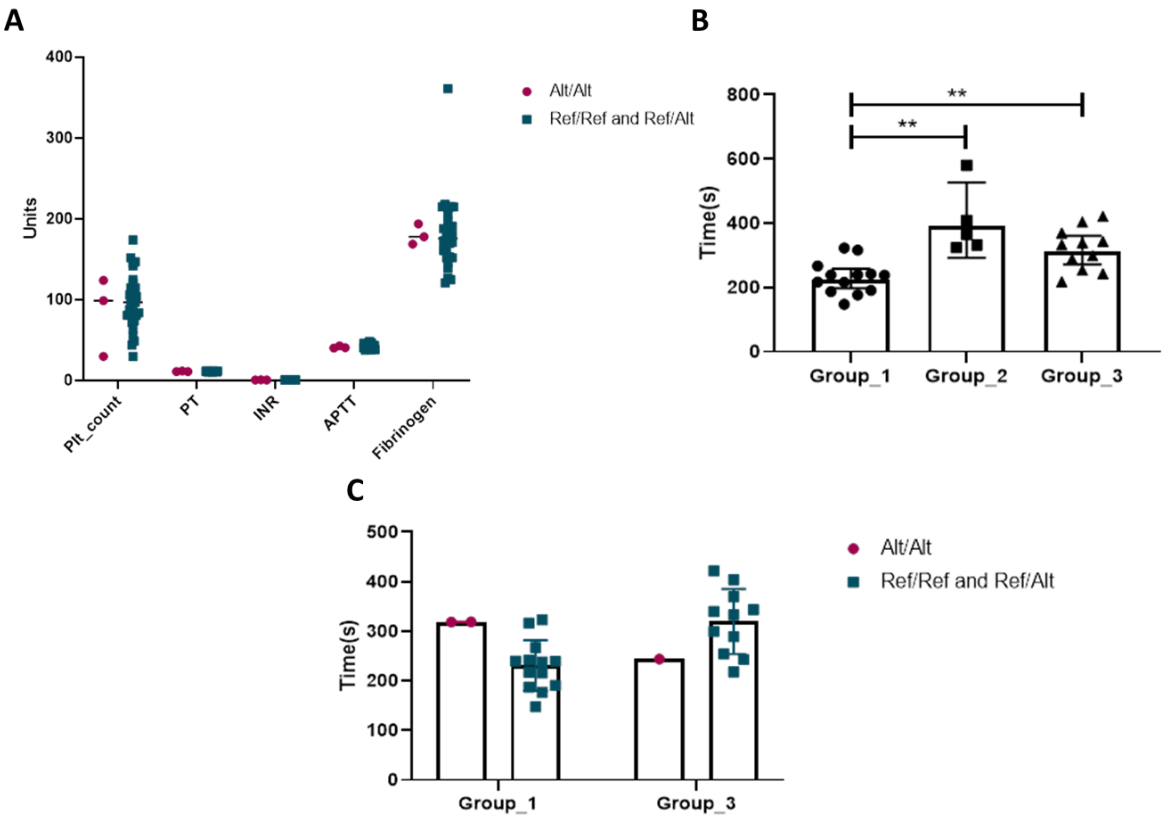


Figure 2.S4. (A) Platelet count and coagulation panels *SEL1L* c.1810A>G p.Ile604Val genotype, demonstrating no significant differences in any assayed parameters. (B) Average template bleeding times (TBTs; in seconds) of Ref/Ref and Ref/Alt control horses by group, showing substantial effect of group; (C) Average TBTs (in seconds) of all horses, in groups 1 and 3 only. Circle data points indicate horses that were Alt/Alt for *SEL1L* missense variant and square data points indicate horses that were Ref/Alt or Ref/Ref. ***P*<0.01

Due to the "group" effect of conducting the experiment in three different batches, a follow-up TBT study was performed using Ref/Ref (*n*=10) and Alt/Alt (3) TBs. This study confirmed that there was no significant difference in TBT between the two genotypes (*P*=0.94; **Fig. 2.8**).

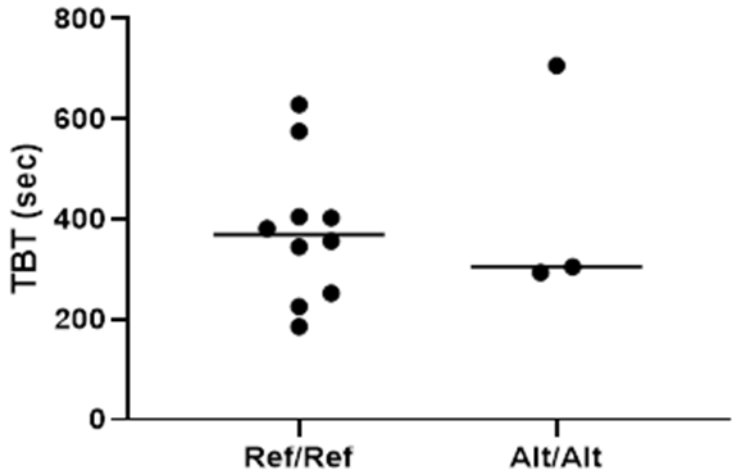


Figure 2.8. TBTs from follow up study using only horses that were Ref/Ref or Alt/Alt for the *SEL1L* c.1810A>G p.Ile604Val variant. There is no significant difference between genotypes.

Discussion

This study investigated the genetic etiology of AET, an inherited platelet disorder that prevents affected horses from clotting efficiently following vascular injury. Whole-genome sequencing

from five affected, eleven controls, and one equivocal horse was utilized to investigate candidate genes, and subsequently, associated genetic variants across the entire genome. The equivocal horse was included in the analysis since her dam and full sibling were affected. A total of 16 candidate genes (*AKT1, AKT2, F5, SHIP1, PIK3C2B, F2, F2R, PIK3C2A, PIK3C2G, PIK3CA, PIK3CB, PIK3CG, PIK3R1, PIK3R4, PIK3R5,* and *PIK3R6*) were identified from a previous study [5] that had examined differences in protein expression and function in AET affected horses. Since there were no associated genetic variants in any of these 16 genes, whole-genome sequencing analysis was performed, identifying a region on chromosome 24 harboring 70% of the significantly associated variants (P<0.001). Within this region, three putative variants (*SEL1L* c.1810A>G p.Ile604Val, *AL355838.1:g.26447375_26448962del*, and *VIPAS39:g.22685398_22685470del*) were identified as associated with AET and in or near genes known to be expressed in platelets. The *AL355838.1:g.26447375_26448962del* variant was subsequently excluded as RT-PCR demonstrated that the lncRNA is not expressed in equine platelets. While there was no significant change in transcript expression of *SEL1L* or *VIPAS39*, there was evidence of changes at the protein level of SEL1L. The lack of change in expression of *SEL1L* is also further evidence that the upstream *AL355838.1* is not affecting the *SEL1L* transcript even in megakaryocytes, the cell that produces platelets [4].

Further investigation into *SEL1L* identified changes in protein expression, with Alt/Alt horses having significantly less platelets with SEL1L. We subsequently demonstrated that, while SEL1L is an intracellular protein, it does not localize to the α-granule in resting platelets, but it does relocate to the plasma membrane upon thrombin stimulation. This indicates that SEL1L likely localized to a granule or organelle within the platelet that is involved in exocytosis post-activation. One of our hypotheses for the mechanism of action of SEL1L is that, since the AET clinical phenotype is only seen after vascular injury, thrombin activation and SEL1L relocation is necessary for SEL1L to function as a cell signaling protein to further propagate the clotting

process. *SEL1L* knockdown in zebrafish leads to cardiac and brain edema, decreased blood circulation, and numerous blood stases in the tail [26], which further suggests that SEL1L does play a role in circulation and in particular normal circulation.

While *SEL1L* in platelets has not yet been studied, in other cell types, the primary known function of SEL1L is in endoplasmic reticulum (ER) associated degradation (ERAD) [27,28]. Knocking out *SEL1L* is embryonic lethal in mice [27,28]. SEL1L forms a complex with Hrd1 [29] and is necessary for ER homeostasis and proper protein degradation and secretion [27,28]. Even though platelets do not have a complete ER, platelets contain ER fragments [30]. Thus, SEL1L could play a similar role in protein degradation in platelets. Recent studies have identified novel mechanisms of SEL1L regulating cell-signaling pathways by regulating the ubiquitination of a transcription factor and by controlling the degradation of a specific receptor [31,32]. SEL1L may have a similar mechanism in AET where it regulates a protein involved in platelet function, likely in the thrombin signaling pathway since that is where differences were previously identified in AET affected horses [3,5]. However, platelets do not have a complete ER [30], and no studies on the whether ERAD occurs in platelets have been published.

While SEL1L has an established role in ERAD, a recent study validated two additional isoforms of SEL1L that do not solely localize to the ER in human lymphocytes [33]. One of these isoforms (SEL1L-B) in particularly appears to associate with membranes and participate in multi-protein complexes [33]. The variant *SEL1L* c.1810A>G p.Ile604Val is not in the published versions of either of these shorter isoforms that contain exons 1-9 (SEL1L-B) or exons 1-8 (SEL1L-C) [34]. However, it has been noted that work in at least one pancreatic neoplastic cell line does not confirm published splicing of SEL1L-B [34], so it is likely that there are additional isoforms in other cell types. Moreover, there are two other computationally predicted isoforms that have not yet been experimentally validated, and one of them does contain the exon where *SEL1L*

c.1810A>G p.Ile604Val is located [34]. We hypothesize that, since platelets do not have a complete ER and we demonstrate that SEL1L relocates to the plasma membrane, an alternative isoform is likely present in platelets that plays a role in cell signaling between platelets post-thrombin activation.

While three candidate genetic variants were identified based on our whole-genome association study (*SEL1L* c.1810A>G p.Ile604Val, *AL355838.1:g.26447375_26448962del*, and *VIPAS39:g.22685398_22685470del*), we prioritized *SEL1L* c.1810A>G p.Ile604Val because; *(1)* this is a coding variant, *(2)* the whole-genome sequenced AET affected horses were homozygous, and *(3)* flow cytometry demonstrated decreased protein expression of SEL1L in Alt/Alt horses. However, the genotyping of the larger 114 horse population for the two putative variants expressed in equine platelets confirmed that the *VIPAS39:g.22685398_22685470del* variant is also still associated with AET. *VIPAS39* has been shown to be necessary for α-granule biogenesis in megakaryocytes and platelets [35]. Genetic mutations naturally occur in *VIPAS39* in humans, causing Arthrogryposis, Renal Dysfunction and Cholestasis (ARC) Syndrome, which is a multisystem genetic disorder [36]. Signs of the abnormal bleeding due to ARC include hemorrhaging, petechiae, and epistaxis [37,38]. While ARC is substantially more severe than AET, there are similarities in the signs and in that α-granule exocytosis may be affected in AET horses, as demonstrated by the decreased amount of Factor V released from AET platelet α-granules [5]. These similarities highlight that we cannot exclude *VIPAS39:g.22685398_22685470del* as a putative variant for AET, and it should be further studied to determine if it plays a role in the etiology of the disease.

A limitation throughout this study was the number of horses available for study. We only identified *n*=3 TBs that were Alt/Alt for the *SEL1L* c.1810A>G p.Ile604Val variant and that were readily accessible, since fresh platelets were required for most of the experiments. Additionally,

the difficulty in sufficiently phenotyping the horses posed a problem in identifying the causative variant. But it also provided an opportunity to seek out other methods to functionally evaluate the disease, such as the collagen spreading assay and template bleeding times.

In examining functional effects of *SEL1L* c.1810A>G p.Ile604Val, two of the homozygous horses had decreased platelet spreading on collagen. However, an additional third homozygous horse was identified later in the study and the platelets from this horse, who was an OVX mare, spread similarly to Ref/Ref horses. It was also noted that the Ref/Ref horse that spread the most was the only other OVX mare in the population. While OVX mares are rare in the general horse population, these two mares are teaching and research horses at the University of California Davis and used as jump mares. In OVX pigs removal of the ovaries led to increased platelet aggregation in response to both ADP and collagen [39]. Thus, it is likely that even with the *SEL1L* c.1810A>G p.Ile604Val variant, the OVX mare's platelets had a strong enough reaction to collagen to stimulate the platelets to spread similarly to Ref/Ref horses. Therefore, the OVX mares were excluded and a significant decrease in platelet area was identified between Ref/Ref and Alt/Alt horses. This suggests that SEL1L likely plays a role in platelet activation and the *SEL1L* c.1810A>G p.Ile604Val variant prevents normal function. Further study is needed to better understand the mechanism of SEL1L in platelet activation.

The TBT study included all three homozygous horses, but the high degree of both intra- and inter- horse variability [25] likely led to the non-significant results. Thus, a larger cohort of Alt/Alt *SEL1L* c.1810A>G p.Ile604Val horses is required to further identify any functional differences in Alt/Alt horses. It is noteworthy though that horses in a regular exercise schedule did have substantially higher TBTs. This has not been published previously in equine athletes, but there have been studies done in humans with varying results [40,41]. One study found that it depends on the type of exercise, with shorter (15 minute) exercise times causing an increase in TBT and

longer (several hours) exercise times causing a decrease in TBTs [41]. This supports our findings with the horses since their exercise time was less than 30 min each day.

In summary, these results suggest *SEL1L* c.1810A>G p.Ile604Val may play a role in the etiology of AET. Additionally, our work did not exclude *VIPAS39:g.22685398_22685470del* as a candidate and it should be further evaluated to determine functional impact on platelet function. Identifying the definitive causative variant of AET will allow for affected TBs to be easily identified through genetic testing as well as identify carriers to inform breeding decisions. Additionally, since neither candidate gene has a known role in clotting, elucidating the mechanism of AET may define novel pathways involved in coagulation.

References

1. Norris, J. W.; Pratt, S. M.; Hunter, J. F.; Gardner, I. A.; Tablin, F. Prevalence of Reduced Fibronogen Binding to Platelets in a Population of Thoroughbreds. *Am. J. Vet. Res.* **2007**, *68* (7), 716–721. https://doi.org/10.2460/ajvr.68.7.716.
2. Fry, M. M.; Walker, N. J.; Blevins, G. M.; Magdesian, K.; Tablin, F. Platelet Function Defect in Thoroughbred Filly. *J. Vet. Intern. Med.* **2005**, No. 19, 359–362. https://doi.org/10.1007/s11999-009-0725-x.
3. Norris, J. W.; Pratt, S. M.; Auh, J.; Wilson, S. J.; Clutter, D.; Magdesian, K. G.; Ferraro, G. L.; Tablin, F. Investigation of a Novel, Heritable Bleeding Diathesis of Thoroughbred Horses and Development of a Screening Assay. *J. Vet. Intern. Med.* **2006**, *20* (6), 1450–1456. https://doi.org/10.1111/j.1939-1676.2006.tb00765.x.
4. Beutler, E.; Lichtman, M. A.; Coller, B. S.; Kipps, T. J.; Seligsohn, U. *Williams Hematology*, 6th ed.; McGraw-Hill: New York, 2001.
5. Norris, J. W.; Pombo, M.; Shirley, E.; Blevins, G.; Tablin, F. Association of Factor V

Secretion with Protein Kinase B Signaling in Platelets from Horses with Atypical Equine Thrombasthenia. *J. Vet. Intern. Med.* **2015**, *29* (5), 1387–1394. https://doi.org/10.1111/jvim.13595.

6. Bolger, A. M.; Lohse, M.; Usadel, B. Genome Analysis Trimmomatic : A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30* (15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
7. Kalbfleisch, T. S.; Rice, E. S.; DePriest, M. S.; Walenz, B. P.; Hestand, M. S.; Vermeesch, J. R.; O'Connell, B. L.; Fiddes, I. T.; Vershinina, A. O.; Saremi, N. F.; Petersen, J. L.; Finno, C. J.; Bellone, R. R.; McCue, M. E.; Brooks, S. A.; Bailey, E.; Orlando, L.; Green, R. E.; Miller, D. C.; Antczak, D. F.; MacLeod, J. N. Improved Reference Genome for the Domestic Horse Increases Assembly Contiguity and Composition. *Commun. Biol.* **2018**, *1* (197), 1–8. https://doi.org/10.1038/s42003-018-0199-z.
8. Li, H.; Durbin, R. Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* **2009**, *25* (14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
9. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25* (16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
10. Thorvaldsdóttir, H.; Robinson, J. T.; Mesirov, J. P. Integrative Genomics Viewer (IGV): High-Performance Genomics Data Visualization and Exploration. *Brief. Bioinform.* **2013**, *14* (2), 178–192. https://doi.org/10.1093/bib/bbs017.
11. Christopherson, P. W.; Insalaco, T. A.; Van Santen, V. L.; Livesey, L.; Bourne, C.; Boudreaux, M. K. Characterization of the CDNA Encoding Alpha IIb and Beta 3 in Normal Horses and Two Horses with Glanzmann Thrombasthenia. *Vet. Pathol.* **2006**, *43* (1), 78–82. https://doi.org/10.1354/vp.43-1-78.

12. Macieira, S.; Lussier, J.; Bédard, C. Characterization of the CDNA and Genomic DNA Sequence Encoding for the Platelet Integrin Alpha IIB and Beta III in a Horse with Glanzmann Thrombasthenia. *Can. J. Vet. Res.* **2011**, *75*, 222–227.

13. Sanz, M. G.; Wills, T. B.; Christopherson, P.; Hines, M. T. Glanzmann Thrombasthenia in a 17-Year-Old Peruvian Paso Mare. *Vet. Clin. Pathol.* **2011**, *40* (1), 48–51. https://doi.org/10.1111/j.1939-165X.2011.00289.x.

14. Christopherson, P. W.; Santen, V. L.; Livesey, L.; Boudreaux, M. K. A 10-Base-Pair Deletion in the Gene Encoding Platelet Glycoprotein IIb Associated with Glanzmann Thrombasthenia in a Horse. *J. Vet. Intern. Med.* **2007**, *21* (1), 196–198. https://doi.org/10.1111/j.1939-1676.2007.tb02947.x.

15. Garrison, E.; Marth, G. Haplotype-Based Variant Detection from Short-Read Sequencing. *arXiv Prepr. arXiv1207.3907 [q-bio.GN]* **2012**, 1–9. https://doi.org/arXiv:1207.3907 [q-bio.GN].

16. Rausch, T.; Zichner, T.; Schlattl, A.; Stütz, A. M.; Benes, V.; Korbel, J. O. DELLY: Structural Variant Discovery by Integrated Paired-End and Split-Read Analysis. *Bioinformatics* **2012**, *28* (18), 333–339. https://doi.org/10.1093/bioinformatics/bts378.

17. Cingolani, P.; Platts, A.; Wang, L. L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S. J.; Lu, X.; Ruden, D. M. A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff: SNPs in the Genome of Drosophila Melanogaster Strain w 1118; Iso-2; Iso-3. *Fly (Austin).* **2012**, *6* (2), 80–92. https://doi.org/10.4161/fly.19695.

18. Gel, B.; Serra, E. KaryoploteR: An R/Bioconductor Package to Plot Customizable Genomes Displaying Arbitrary Data. *Bioinformatics* **2017**, *33* (19), 3088–3090. https://doi.org/10.1093/bioinformatics/btx346.

19. Cingolani, P.; Patel, V. M.; Coon, M.; Nguyen, T.; Land, S. J.; Ruden, D. M.; Lu, X. Using Drosophila Melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Front. Genet.* **2012**, *3* (MAR), 1–9.

https://doi.org/10.3389/fgene.2012.00035.

20. Amisten, S.; Braun, O. Ö.; Bengtsson, A.; Erlinge, D. Gene Expression Profiling for the Identification of G-Protein Coupled Receptors in Human Platelets. *Thromb. Res.* **2008**, *122*, 47–57. https://doi.org/10.1016/j.thromres.2007.08.014.

21. Lonsdale, J.; Thomas, J.; Salvatore, M.; Phillips, R.; Lo, E.; Shad, S.; Hasz, R.; Walters, G.; Garcia, F.; Young, N.; Foster, B.; Moser, M.; Karasik, E.; Gillard, B.; Ramsey, K.; Sullivan, S.; Bridge, J.; Magazine, H.; Syron, J.; Fleming, J.; Siminoff, L.; Traino, H.; Mosavel, M.; Barker, L.; Jewell, S.; Rohrer, D.; Maxim, D.; Filkins, D.; Harbach, P.; Cortadillo, E.; Berghuis, B.; Turner, L.; Hudson, E.; Feenstra, K.; Sobin, L.; Robb, J.; Branton, P.; Korzeniewski, G.; Shive, C.; Tabor, D.; Qi, L.; Groch, K.; Nampally, S.; Buia, S.; Zimmerman, A.; Smith, A.; Burges, R.; Robinson, K.; Valentino, K.; Bradbury, D.; Cosentino, M.; Diaz-Mayoral, N.; Kennedy, M.; Engel, T.; Williams, P.; Erickson, K.; Ardlie, K.; Winckler, W.; Getz, G.; DeLuca, D.; Daniel MacArthur; Kellis, M.; Thomson, A.; Young, T.; Gelfand, E.; Donovan, M.; Meng, Y.; Grant, G.; Mash, D.; Marcus, Y.; Basile, M.; Liu, J.; Zhu, J.; Tu, Z.; Cox, N. J.; Nicolae, D. L.; Gamazon, E. R.; Im, H. K.; Konkashbaev, A.; Pritchard, J.; Stevens, M.; Flutre, T.; Wen, X.; Dermitzakis, E. T.; Lappalainen, T.; Guigo, R.; Monlong, J.; Sammeth, M.; Koller, D.; Battle, A.; Mostafavi, S.; McCarthy, M.; Rivas, M.; Maller, J.; Rusyn, I.; Nobel, A.; Wright, F.; Shabalin, A.; Feolo, M.; Sharopova, N.; Sturcke, A.; Paschal, J.; Anderson, J. M.; Wilder, E. L.; Derr, L. K.; Green, E. D.; Struewing, J. P.; Temple, G.; Volpi, S.; Boyer, J. T.; Thomson, E. J.; Guyer, M. S.; Ng, C.; Abdallah, A.; Colantuoni, D.; Insel, T. R.; Koester, S. E.; A Roger Little; Bender, P. K.; Lehner, T.; Yao, Y.; Compton, C. C.; Vaught, J. B.; Sawyer, S.; Lockhart, N. C.; Demchok, J.; Moore, H. F. The Genotype-Tissue Expression (GTEx) Project. *Nat. Genet.* **2013**, *45* (6), 580–585. https://doi.org/10.1038/ng.2653.

22. Kopp, K. J.; Moore, J. N.; Byars, T. D.; Brooks, P. Template Bleeding Time and Thromboxane Generation in the Horse: Effects of Three Non-steroidal Anti-inflammatory

Drugs. *Equine Vet. J.* **1985**, *17* (4), 322–324. https://doi.org/10.1111/j.2042-3306.1985.tb02509.x.

23. Salkind, N. J. Bonferroni Procedure. In *Encylopedia of Research Design, Volume 1*; SAGE, 2010; pp 98–101.
24. Vogel, C.; Marcotte, E. M. Insights into Regulation of Protein Abundance from Proteomics and Transcriptomis Analyses. *Nat. Rev. Genet.* **2013**, *13* (4), 227–232. https://doi.org/10.1038/nrg3185.Insights.
25. Segura, D.; Monreal, L. Poor Reproducibility of Template Bleeding Time in Horses. *J. Vet. Intern. Med.* **2008**, *22*, 238–241. https://doi.org/10.1177/016555157900100507.
26. Barbieri, A.; Carra, S.; De Blasio, P.; Cotelli, F.; Biunno, I. Sel1l Knockdown Negatively Influences Zebrafish Embryos Endothelium. *J. Cell. Physiol.* **2018**, *233* (7), 5396–5404. https://doi.org/10.1002/jcp.26366.
27. Sun, S.; Shi, G.; Han, X.; Francisco, A. B.; Ji, Y.; Mendonca, N.; Liu, X.; Locasale, J. W.; Simpson, K. W.; Duhamel, G. E.; Kersten, S.; Yates, J. R.; Long, Q.; Qi, L. Sel1L Is Indispensable for Mammalian Endoplasmic Reticulum-Associated Degredation, Endoplasmic Reticulum Homeostasis, and Survival. *Proc. Natl. Acad. Sci.* **2014**, *111*, 582–591. https://doi.org/10.1073/pnas.1308963111.
28. Francisco, A. B.; Singh, R.; Li, S.; Vani, A. K.; Yang, L.; Munroe, R. J.; Diaferia, G.; Cardano, M.; Biunno, I.; Qi, L.; Schimenti, J. C.; Long, Q. Deficiency of Suppressor Enhancer Lin12 1 like (SEL1L) in Mice Leads to Systemic Endoplasmic Reticulum Stress and Embryonic Lethality. *J. Biol. Chem.* **2010**, *285* (18), 13694–13703. https://doi.org/10.1074/jbc.M109.085340.
29. Lilley, B. N.; Ploegh, H. L. Multiprotein Complexes That Link Dislocation, Ubiquitination, and Extraction of Misfolded Proteins from the Endoplasmic Reticulum Membrane. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (40), 14296–14301. https://doi.org/10.1073/pnas.0505014102.

30. Daimon, T.; Gotoh, Y. Cytochemical Evidence of the Origin of the Dense Tubular System in the Mouse Platelet. *Histochemstry* **1982**, *76*, 189–196.

31. Bhattacharya, A.; Sun, S.; Wang, H.; Liu, M.; Long, Q.; Yin, L.; Kersten, S.; Zhang, K.; Qi, L. Hepatic Sel1L-Hrd1 ER-associated Degradation (ERAD) Manages FGF21 Levels and Systemic Metabolism via CREBH. *EMBO J.* **2018**, *37* (22), 1–18. https://doi.org/10.15252/embj.201899277.

32. Shrestha, N.; Liu, T.; Ji, Y.; Reinert, R. B.; Torres, M.; Li, X.; Zhang, M.; Tang, C. H. A.; Hu, C. C. A.; Liu, C.; Naji, A.; Liu, M.; Lin, J. D.; Kersten, S.; Arvan, P.; Qi, L. Sel1L-Hrd1 ER-Associated Degradation Maintains β Cell Identity via TGF-β Signaling. *J. Clin. Invest.* **2020**, *130* (7), 3499–3510. https://doi.org/10.1172/JCI134874.

33. Cattaneo, M.; Lotti, L. V.; Martino, S.; Cardano, M.; Orlandi, R.; Mariani-Costantini, R.; Biunno, I. Functional Characterization of Two Secreted SEL1L Isoforms Capable of Exporting Unassembled Substrate. *J. Biol. Chem.* **2009**, *284* (17), 11405–11415. https://doi.org/10.1074/jbc.M805408200.

34. Biunno, I.; Cattaneo, M.; Orlandi, R.; Canton, C.; Biagiotti, L.; Ferrero, S.; Barberis, M.; Pupa, S. M.; Scarpa, A.; Mé Nard, S. SEL1L a Multifaceted Protein Playing a Role in Tumor Progression. *J. Cell. Physiol.* **2006**, *208*, 23–38. https://doi.org/10.1002/jcp.20574.

35. Urban, D.; Li, L.; Christensen, H.; Pluthero, F. G.; Chen, S. Z.; Puhacz, M.; Garg, P. M.; Lanka, K. K.; Cummings, J. J.; Kramer, H.; Wasmuth, J. D.; Parkinson, J.; Kahr, W. H. A. The VPS33B-Binding Protein VPS16B Is Required in Megakaryocyte and Platelet a -Granule Biogenesis. *Blood* **2012**, *120* (25), 5032–5040. https://doi.org/10.1182/blood-2012-05-431205.There.

36. Cullinane, A. R.; Straatman-Iwanowska, A.; Zaucker, A.; Wakabayashi, Y.; Bruce, C. K.; Luo, G.; Rahman, F.; Gürakan, F.; Utine, E.; Özkan, T. B.; Denecke, J.; Vukovic, J.; Rocco, M. Di; Mandel, H.; Cangul, H.; Matthews, R. P.; Thomas, S. G.; Rappoport, J. Z.; Arias, I. M.; Wolburg, H.; Knisely, A. S.; Kelly, D. A.; Müller, F.; Maher, E. R.; Gissen, P.

Mutations in VIPAR Cause an Arthrogryposis, Renal Dysfunction and Cholestasis Syndrome Phenotype with Defects in Epithelial Polarization. *Nat. Genet.* **2010**, *42* (4), 303–312. https://doi.org/10.1038/ng.538.Mutations.

37. Lo, B.; Li, L.; Gissen, P.; Christensen, H.; Mckiernan, P. J.; Ye, C.; Abdelhaleem, M.; Hayes, J. A.; Williams, M. D.; Chitayat, D.; Kahr, W. H. A. Requirement of VPS33B , a Member of the Sec1 / Munc18 Protein Family , in Megakaryocyte and Platelet a-Granule Biogenesis. *Blood* **2005**, *106* (13), 4159–4167. https://doi.org/10.1182/blood-2005-04-1356.B.L.

38. Abbott, L. S.; Li, L.; Christensen, H.; Pluthero, F.; Urban, D.; Dror, Y.; Kahr, W. H. Platelet Functional and Morphological Abnormalities In a Case Of Arthrogryposis, Renal Dysfunction and Cholestasis (ARC) Syndrome With a Novel Genetic Profile. *Blood* **2013**, *122* (21).

39. Jayachandran, M.; Owen, W. G.; Miller, V. M. Effects of Ovariectomy on Aggregation, Secretion, and Metalloproteinases in Porcine Platelets. *Am J Physiol Hear. Circ Physiol* **2003**, *284*, 1679–1685. https://doi.org/10.1152/ajpheart.00958.2002.

40. Chen, H.; Tang, Y.; Wu, H.-J.; Jen, C. J. Effects of Acute Exercise on Bleeding Time, Bleeding Amount and Blood Cell Counts: A Comparative Study. *Thromb. Res.* **1989**, *55*, 503–510.

41. Carter, J. W.; Ready, A. E.; Singhroy, S.; Duta, E.; Gerrard, J. M. The Effect of Exercise on Bleeding Time and Local Production of Prostacyclin and Thromboxane. *Eur. J. Appl. Physiol. Occup. Physiol.* **1989**, *59*, 355–359.

Chapter 3: Comparison of Poly-A+ Selection and rRNA Depletion in Detection of lncRNA in Two Equine Tissues Using RNA-seq

Authors: Anna R. Dahlgren[1], Erica Y. Scott[2], Tamer Mansour[1], Erin N. Hales[1], Pablo J. Ross[2], Theodore S. Kalbfleisch[3], James N. MacLeod[3], Jessica L. Petersen[4], Rebecca R. Bellone[1,5], Carrie J. Finno[1]

[1] Department of Population Health and Reproduction, School of Veterinary Medicine, University of California Davis, Davis, CA 95616

[2] Department of Animal Science, College of Agricultural and Environmental Sciences, University of California Davis, Davis, CA 95616

[3] Gluck Equine Research Center, Department of Veterinary Science, University of Kentucky, Lexington, KY 40546

[4] Department of Animal Science, University of Nebraska Lincoln, Lincoln, NE 68583

[5] Veterinary Genetics Laboratory, School of Veterinary Medicine, University of California Davis, CA 95616

Keywords: Annotation, transcriptome, regulatory, horse

Reference: Dahlgren AR, Scott E, Mansour T, Hales EN, Ross P, Kalbfleisch T, MacLeod J, Petersen J, Bellone R, Finno CJ. Comparison of poly-A+ selection and rRNA depletion in detection of lncRNA in two equine tissues using RNA-seq. *Non-coding RNA* 2020 6(3); pii: E32. Published 2020 Aug 21. doi: 10.3390/ncrna6030032.

Abstract

Long non-coding RNAs (lncRNAs) are untranslated regulatory transcripts longer than 200 nucleotides that can play a role in transcriptional, post-translational, and epigenetic regulation. Traditionally, RNA-sequencing (RNA-seq) libraries have been created by isolating transcriptomic RNA via poly-A^+ selection. In the past 10 years, methods to perform ribosomal

RNA (rRNA) depletion of total RNA have been developed as an alternative, aiming for better coverage of whole transcriptomic RNA, both polyadenylated and non-polyadenylated transcripts. The purpose of this study was to determine which library preparation method is optimal for lncRNA investigations in the horse. Using liver and cerebral parietal lobe tissues from two healthy Thoroughbred mares, RNA-seq libraries were prepared using standard poly-A^+ selection and rRNA-depletion methods. Averaging the two biologic replicates, poly-A^+ selection yielded 327 and 773 more unique lncRNA transcripts for liver and parietal lobe, respectively. More lncRNA were found to be unique to poly-A^+ selected libraries, and rRNA-depletion identified small nucleolar RNA (snoRNA) to have a higher relative expression than in the poly-A^+ selected libraries. Overall, poly-A^+ selection provides a more thorough identification of total lncRNA in equine tissues while rRNA-depletion may allow for easier detection of snoRNAs.

Introduction

Long non-coding RNAs (lncRNAs) are untranslated transcripts longer than 200 nucleotides (nt). They have been shown to have a wide range of functions in the regulation of transcription, translation, epigenetics, differentiation, and the cell cycle [1–7]. In recent years, lncRNAs have been increasingly shown to play important roles in diseases, particularly cancer [8–10] and neurodegeneration [11,12]. However, many functional roles of lncRNAs in cell biology, development, and disease pathogenesis remain unknown, especially in the horse. As sequence conservation of lncRNA among species is low [13], lncRNAs identified in the human and mouse often are not expected to have a similar genomic sequence in horses.

Previously, an equine lncRNA pipeline and database (https://github.com/eyscott/lncRNA) was developed using RNA sequencing (RNA-seq) data from various laboratories and across disease phenotypes [13]. While filling a necessary gap in knowledge in equine genetics, the

database had some limitations. One of the largest limitations was that many of the horses had one of several diseases. This prevents the public database from being used as a baseline from which to identify aberrant lncRNA expression or splicing in disease-affected horses. Additionally, some RNA samples were prepared using ribosomal RNA (rRNA)-depletion whereas other samples were prepared with poly-A^+ selection. However, both methods were not used on any single tissue, so there was no way to assess and quantify different transcription profiles as a function of the library preparation methods used. Previous human studies have demonstrated that the transcripts that are sequenced may differ in quantity and identity between the two methods, with poly-A^+ selection limited to transcripts with a polyadenylated tail and rRNA-depleted libraries having the additional challenge of often including intronic and intergenic regions [14]. In the horse, biologic validation of putative lncRNAs is lacking; therefore, it is difficult to distinguish between novel lncRNA and true intronic and intergenic reads.

The objective of this study was to determine which RNA-seq library preparation method would most reliably capture lncRNA in the equine genome. Long ncRNA was the focus of this study since substantial work is already in progress to annotate equine protein-coding genes, lncRNA are more likely than protein-coding genes to differ from other species [13,15–17], and there are potentially fewer lncRNA with poly-A tails. Using liver and cerebral parietal lobe tissues collected from two healthy Thoroughbred mares as part of the Functional Annotation of Animal Genomes (FAANG) initiative, direct comparisons between library preparations was performed. These two tissues were chosen to be representative of homogenous and heterogeneous tissue, respectively. As non-polyadenylated lncRNA have been identified in other species [18,19], our hypothesis was that rRNA-depleted libraries would be preferable for annotating lncRNA as this method is not dependent on the transcripts being polyadenylated. Determining the RNA-seq library preparation method best for identifying lncRNA is an essential step toward annotation of the horse genome to identify genetic regions and variants associated with diseases.

Methods

Samples and Sequencing

Liver and parietal lobe of the cerebrum tissues from two healthy Thoroughbred mares was obtained from the functional annotation of the animal genome biobank [20] to investigate lncRNA expression in both a homogeneous and a complex tissue. RNA was isolated using a phenol-chloroform method with a column clean up. RNA quality was measured using an Agilent Bioanalyzer (RIN=8.7). Two RNA-seq libraries were prepared from each tissue sample, one based on poly-A^+ selection and the other using rRNA-depletion. Poly-A^+ selected libraries were made using a strand-specific poly-A^+ capture protocol (TruSeq Stranded mRNA, Illumina, San Diego, CA, USA). A bioanalyzer was used to ensure all poly-A^+ selected libraries had adequate size distributions. The rRNA was depleted (Ribo-Zero, Illumina, San Diego, CA, USA) and prepared as strand-specific (TruSeq Stranded Total RNA Library pre kit, Illumina, San Diego, CA, USA). The libraries were size selected for 140 bp ± 10% fragments and sequenced on a HiSeq 4000 to an average depth of 30 M mapped reads (ERX2600970, ERX2600971). The reads are paired end and 125 bp long.

lncRNA Identification

The reads were trimmed with Sickle [32], mapped with STAR (2-pass) [33], map quality checked with samtools flagstat (>99% mapped and properly paired) [32], down-sampled to similar read counts across samples with samtools view [34], and annotated with Stringtie [35]. The EquCab3 reference genome and corresponding annotation was obtained from NCBI [36]. A lncRNA pipeline slightly modified from the one published by Scott and Mansour [13] was used to isolate the lncRNA and compare the two library preparation methods. Known protein coding transcripts were removed using a combination of filtering out any transcript GffCompare [37]

identified as an exact match with protein coding transcripts from the reference and anti_join [38] against known protein coding transcript names. A histogram of TPMs was used to identify the cut-off to filter out the single exon transcripts with low TPMs that were categorized as false positives (TPM < 2) and to determine the TPM cut-off for the remaining transcripts (TPM < 0.4). As lncRNA are defined as longer than 200 bp, the transcripts were also filtered based on length. Remaining protein coding regions were computationally identified by combining predicted ORFs, protein domain models (Pfam) [39,40], and protein sequence database (hmmsearch; http://hmmer.org/). Additionally, the reads were BLAST'd (NCBI) against known human protein coding cDNA and protein peptide sequences. These findings were merged and filtered out. Known human lncRNA from the Ensembl ncRNA database (ftp.ensembl.org/pub/release-86/fasta/homo_sapiens/ncrna/Homo_sapiens.GRCh38.ncrna.fa.gz) were then compared to the filtered out coding regions and any matches were returned to the final lncRNA file. The lncRNA analysis and final bed files are detailed at https://github.com/ADahlgren/PolyA_ribozero.

Analysis

BEDtools intersect (version 2.29.2) [41] was used to identify lncRNA that do not overlap (by 50%) with the opposite library to identify the number lncRNA that were unique to each library preparation for each tissue. BEDtools intersect was also used to identify transcripts in one library that overlap with transcripts from another library (by 90%) and to identify transcripts that are likely RefSeq lncRNAs.

Correlation plots were made in Rstudio. The spearman rho (r) value and *p*-values were also calculated in Rstudio. EdgeR [42] was used to create MDS scaling plots based on log fold change to determine the role library preparation plays in the lncRNA transcriptome. It was also used to identify the most differentially expressed transcripts between liver and parietal lobe for each library preparation method, with *p* values corrected by a false discovery rate of <0.05. Principal component analysis was done using the dataset from EdgeR in Rstudio.

qRT-PCR

RNA from AH1 and AH2 was reverse transcribed into cDNA using SuperScript III (ThermoFisher; Waltham, MA, USA) according to the manufacturer's instructions. Primers were designed using Primer 3 Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) to span two exons (**Supplemental Table 3.S1**). Endpoint PCR showed specific amplification of the correct product. qRT-PCR was performed on an AriaMx Real-time PCR System (Agilent; Santa Clara, CA, USA) using Brilliant III Ultra-Fast SYBR qPCR Master Mix (Agilent; Santa Clara, CA, USA). cDNA was pooled and serially diluted to ensure efficient amplification and the optimal dilution. All samples were run in triplicate and delta Cqs were calculated with *ACTB* as the reference gene.

Table 3.S1. qRT-PCR Primer Sequences

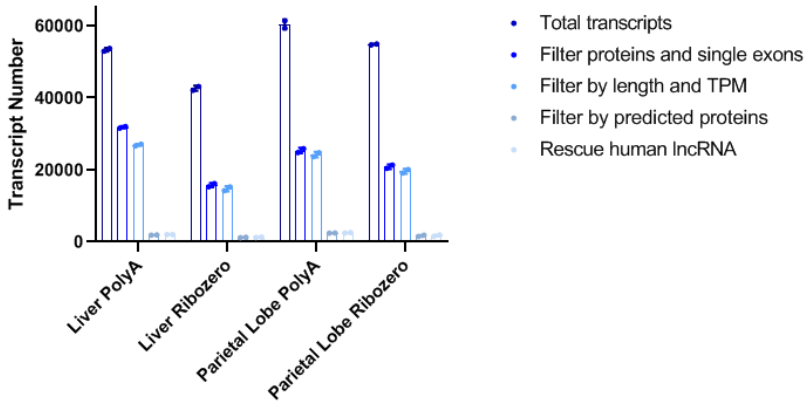
Transcript	Forward Primer	Reverse Primer
H19	GAAGAAGTCCGGGTTCCAA	CCCAAGAACCCTCAAGATGA
LINC02586	TCTGTCACAGTCACCGAAGC	ACCATCCAGATGGACAAGGA
MIR124-2HG	AGAGCCGGTCTAGCTCAGAGA	GAGACGCGGATTACCCAAAC
LINC01351	TATCCCGATACCGCACAAAC	TTCCACTCTACCTTTCCCTCCT
ACTB	AAGGAGAAGCTCTGCTATGTCG	GGGCAGCTCGTAGCTCTTC

Results

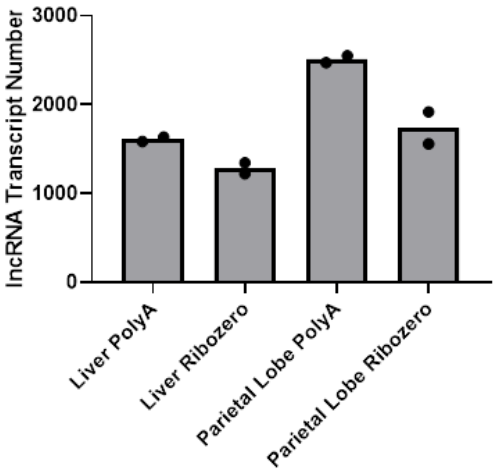
Liver and parietal lobe of the cerebrum were collected from two healthy Thoroughbred mares (adult horse 1: AH1 and adult horse 2: AH2) [20]. RNA was isolated and prepared for

sequencing with poly-A$^+$ selection and rRNA-depletion. Four filters were applied to the resulting RNA-seq datasets as previously described [13] to isolate lncRNA transcripts. First, single exon transcripts with low transcripts per million (TPMs) were filtered out as done previously [13] to remove likely uninformative reads and polymerase mistakes. Then, known protein-coding transcripts were filtered out. Next, the remaining transcripts were filtered based on the definition of lncRNA (>200 nt) and by TPM. To ensure no protein-coding transcripts remain, protein-coding transcripts were computationally predicted and removed. Lastly, previous work has shown that this pipeline removes some true lncRNA, so a rescue step is required [13]. This was done by comparing the removed transcripts to known human lncRNA.

Filtering out known protein-coding and single exon transcripts expressed at low levels resulted in the greatest removal of transcripts for all the samples (**Fig. 3.1A**). Libraries prepared with rRNA-depletion had more protein-coding transcripts removed across tissues and biologic replicates (**Supplemental Fig. 3.S1**), and poly-A$^+$ selection yielded more unique lncRNA (**Fig. 3.1B**). Additionally, as expected, the more complex parietal lobe samples had more unique lncRNA transcripts than the liver samples, which have a more homogenous cellular composition (**Fig. 3.1B**).



(**A**)



(**B**)

Figure 3.1. (**A**) Bar graph showing the average number of total transcripts between the two horses after each filtering step. TPM = transcripts per million. (**B**) Number of unique long non-coding RNAs (lncRNA) in each tissue library preparation combination (same as the last bar in A). Each data point represents one horse. PolyA indicates poly-A^+ selection.

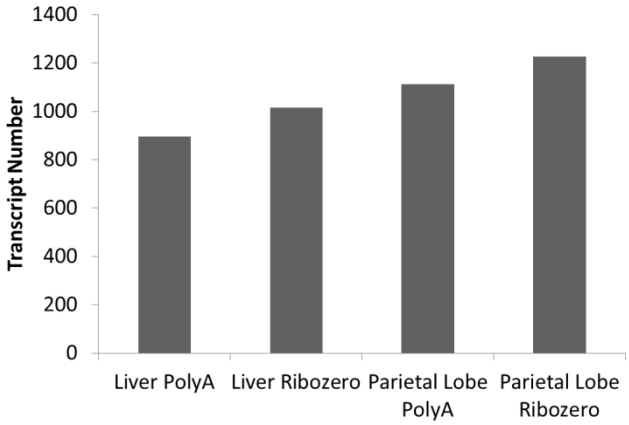


Figure 3.S1. Average number of protein-coding transcripts.

To investigate if the lncRNA expression was similar between the two biologic replicates, we plotted the TPM values for each horse against each other for each tissue and library preparation method in a correlation plot. Analogous unannotated transcripts between biologic replicates were identified via bedtools intersect (**Supplemental Table 3.S2**). Correlation was significant

across both tissue types (**Fig. 3.2**). In each dataset, there were unique transcripts that were outliers (**Supplemental Table 3.S3**) with high TPM values. However, even when the outliers were removed, correlations of biologic replicates between library preparations remained significant (Spearman $r_{(liver\_polyA)}$=0.45, P=8.98×10^{-97}; Spearman $r_{(liver\_ribo)}$=0.524, P=3.24×10^{-86}; Spearman $r_{(parietal\_polyA)}$=0.53, P=6.08×10^{-172}; Spearman $r_{(parietal\_ribo)}$=0.588, P=4.47 × 10^{-127}). Taken together, these findings demonstrate strong correlation of biologic replicates within library preparations and tissue types.

Table 3.S2. Number of lncRNA Used in Correlation Analysis by Tissue and Library Preparation

Tissue and Library Preparation	Number of lncRNA Used in Correlation
Liver PolyA	1934
Liver Ribozero	1208
Parietal Lobe PolyA	2371
Parietal Lobe Ribozero	1364

Table 3.S3. lncRNA Removed in Biologic Replicate Correlation

Liver		Parietal Lobe	
PolyA	**Ribozero**	**PolyA**	**Ribozero**
chr7:25389286-25390378	chr31:8722887-8723106	chr1:112301900-112302558	chr1:21949607-21950018
chr27:34137241-34138781		chr10:76657675-76658212	chr1:112301900-112302558
		chr31:8722887-8723106	chr1:158746022-158746364
		chr31:16560679-16562097	chr27:5219606-5221622
			chr31:8722887-8723106

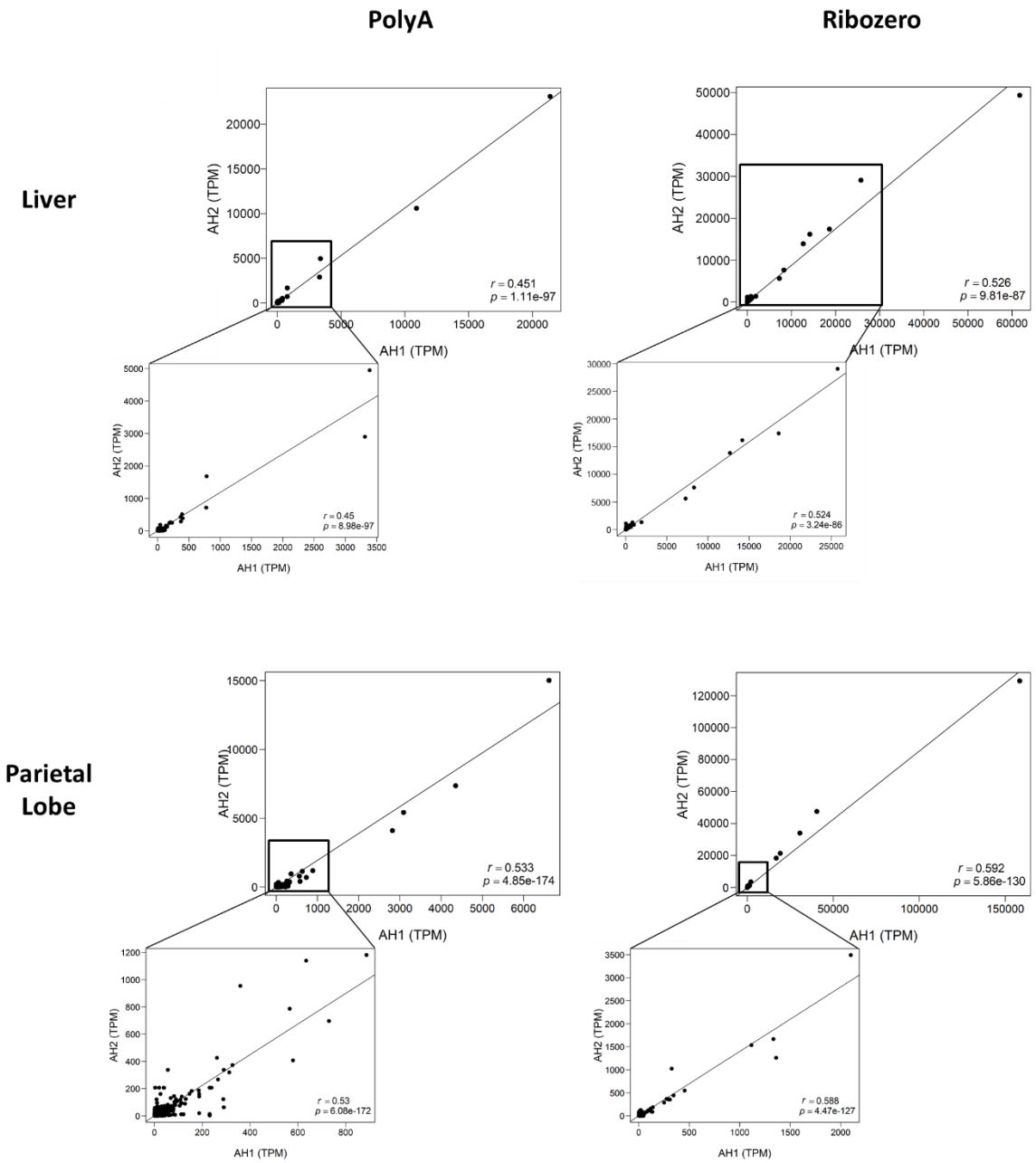


Figure 3.2. Correlation plots for lncRNA expression between biologic replicates. The poly-A$^+$ selected libraries are on the left, the rRNA depleted libraries on the right. The top two graphs are from the liver while the bottom two are from the parietal lobe of the cerebrum. The x-axis indicates expression of the individual lncRNA in adult horse 1 (AH1). The y-axis shows expression of the same lncRNAs in adult horse 2 (AH2). Spearman correlation (r) and *p*-value in bottom right corner.

Differentially expressed (DE) lncRNAs between liver and parietal lobe samples for each library preparation were determined. While most of the DE lncRNA were unannotated, *H19* was identified by both poly-A^+ selection and rRNA-depletion as being expressed higher in the liver than the parietal lobe, similar to findings in humans [21]. In poly-A^+ selected libraries, lncRNAs that appear similar to *LINC00643*, *LINC02586*, and *LOC100128494* in humans have similar expression patterns in liver and parietal lobe [21]. For example, *LINC00643* is highly expressed in the brain in humans and only minimally in the liver [21], which parallels what we see in our RNA-seq data. In rRNA-depleted libraries, lncRNAs that have similarities in sequence or genomic position to *MIR124-2HG*, *LINC00643*, *RP4-785G19.5*, and *LINC01351* in humans have parallel expression patterns in liver and parietal lobe in the horse [21]. *H19, LINC02586, MIR124-2HG,* and *LINC01351* expression in the parietal lobe and liver was confirmed with quantitative reverse transcription PCR (qRT-PCR) in the same horses (**Supplemental Fig. 3.S2**). This suggests that both library preparations are accurately demonstrating relative expression between tissue types.

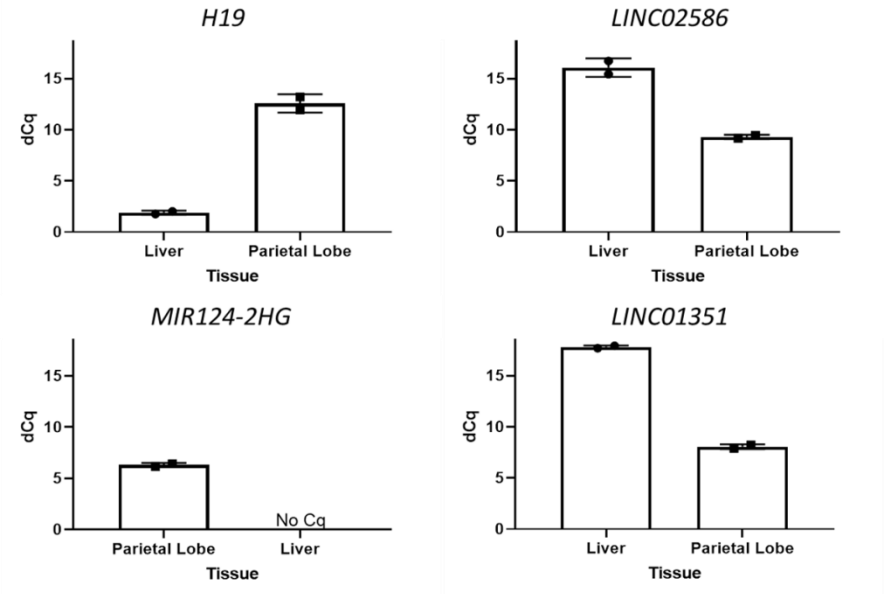


Figure 3.S2. Bar graphs showing delta Cq in parietal lobe and liver for four lncRNA transcripts

Comparing the DE lncRNAs that are annotated by National Center for Biotechnology Information (NCBI) showed that the top two lncRNA are the same for poly-A^+ selection and rRNA-depletion. One in an unknown lncRNA (rna69770) and the other is *H19* (rna41570). Additionally, within the top 10 DE lncRNA, there are two other lncRNA that show up in both library preparations. There is another unknown lncRNA (rna12060) and the other is similar to human *LINC00643* (rna64504). So, only four of the top 10 annotated DE lncRNA are the same between library preparations, indicating that there is a substantial difference in the quantity of lncRNA that is detected by each library preparation.

To further address the impact that library preparation plays in defining the lncRNA transcriptome, a multi-dimensional scaling (MDS) plot was evaluated. While tissue type caused the largest difference between samples (dimension 1; x-axis), library preparation caused the second largest difference (dimension 2; y-axis**; Fig. 3.3**). Principal component analysis (PCA) also showed an even greater difference between tissues when rRNA-depletion was used (PC1=70.9%, PC2=12%; **Supplemental Fig. 3.S3**).

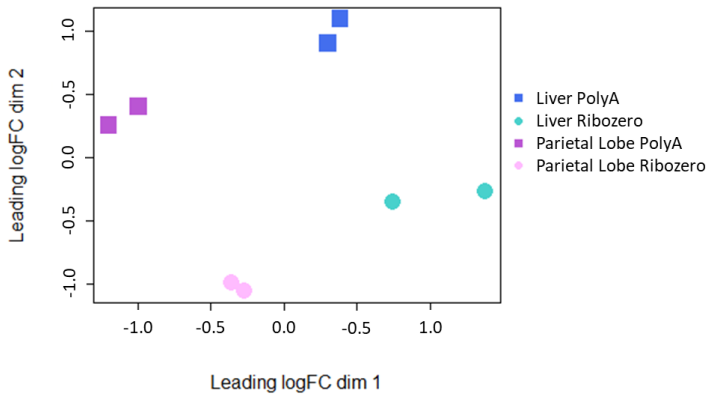


Figure 3.3. Multidimensional scaling plot of the lncRNA expression from each tissue/library preparation for each horse.

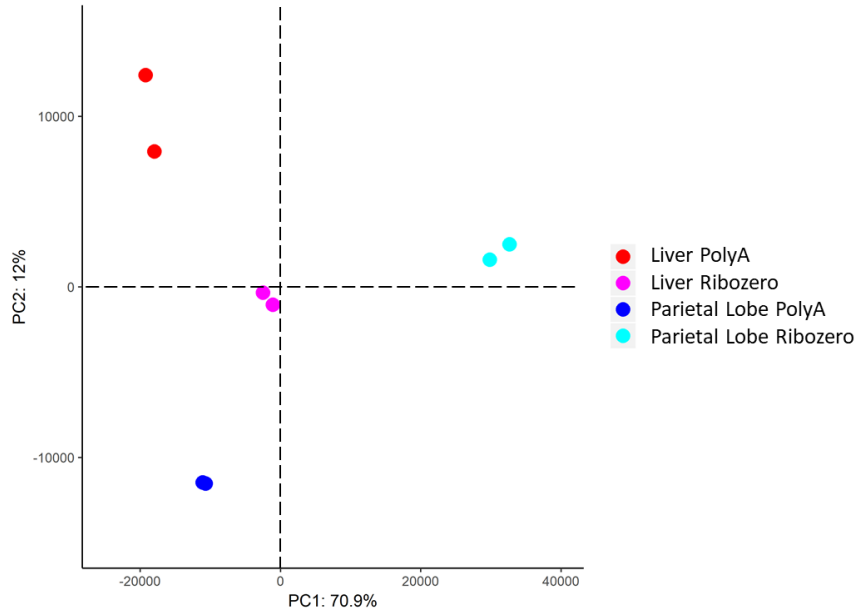


Figure 3.S3. Principle component plot of the lncRNA expression.

We also investigated how many lncRNAs only appear in one library preparation. Within poly-A^+ selected libraries for the liver and parietal lobe, 1276 and 2602 unique lncRNA were identified, respectively. Fewer lncRNA were unique to the rRNA-depleted libraries, with 977 and 1467 lncRNAs identified for the liver and parietal lobe, respectively. This suggests poly-A^+ selection captures more lncRNA than rRNA-depletion.

To continue investigating the differences between library preparations, correlation plots were constructed between library preparation methods for each horse and each tissue (**Fig. 3.4 and Supplemental Fig. 3.S4**). There was only a moderate degree of correlation between library preparations (Spearman $r_{(liver,subset)}$=0.476, P=3.62×10^{-86}; Spearmen $r_{(parietal,subset)}$=0.473, P=9.25×10^{-89}).

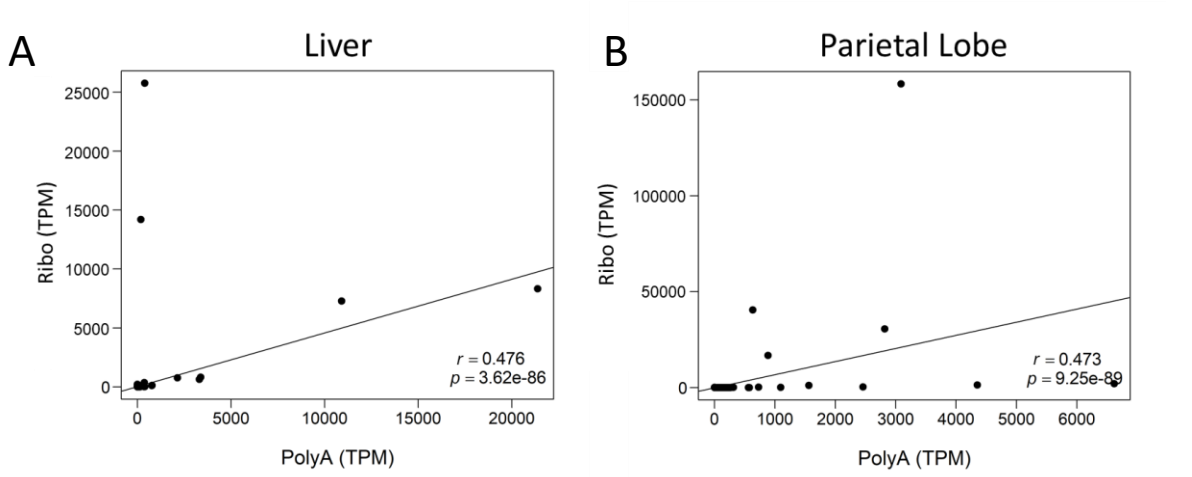


Figure 3.4. Correlation plots for lncRNA expression in AH1 with annotated transcripts. The x-axis indicates expression of the individual lncRNA in the poly-A$^+$ selected library. The y-axis shows expression of the same lncRNAs in the rRNA-depleted library. (**A**) Liver (**B**) Parietal lobe of the cerebrum. Spearman correlation (r) and *p*-value in bottom right corner.

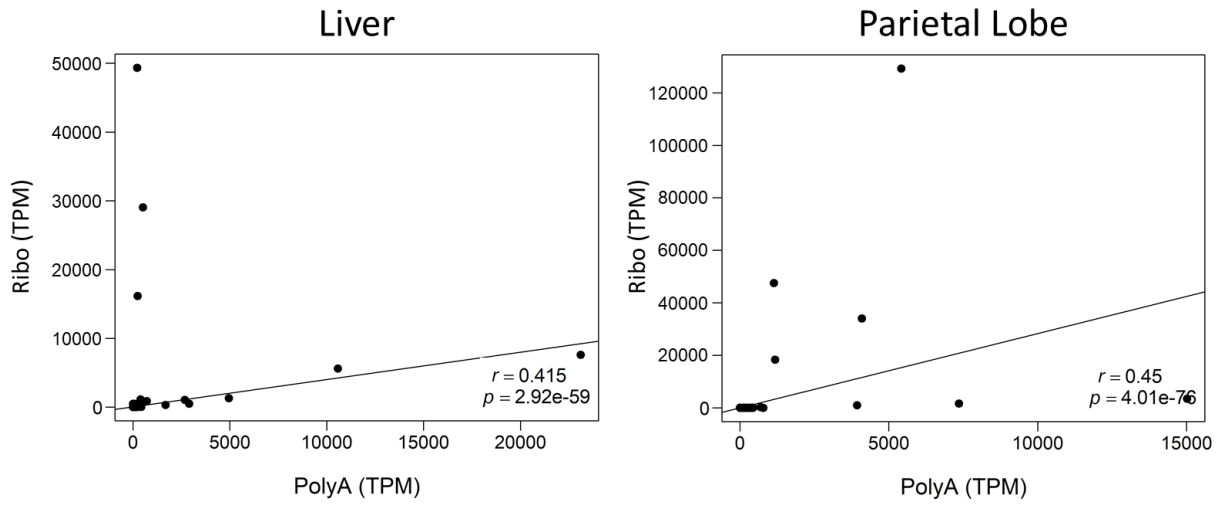


Figure 3.S4. Correlation plots for lncRNA expression in AH2 with annotated transcripts.

When evaluating the annotated lncRNA that had substantially higher TPMs in one library preparation as compared to the other, small nucleolar RNAs (snoRNAs) were consistently higher in rRNA-depleted libraries. There is minimal difference in the number of snoRNAs identified between each library preparation (**Supplemental Table 3.S4**); however

there appears to be a large difference in the relative expression of snoRNAs in rRNA-depleted libraries.

Table 3.S4. lncRNA annotated as snoRNA for each horse, tissue, and library preparation method

Horse	Tissue	Preparation method	lncRNA annotated as snoRNA
AH1	Liver	PolyA	178
		Ribozero	180
	Parietal Lobe	PolyA	201
		Ribozero	178
AH2	Liver	PolyA	213
		Ribozero	181
	Parietal Lobe	PolyA	237
		Ribozero	158

Limiting the correlation analysis to the EquCab3.0 RefSeq annotated lncRNA (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/863/925/GCF_002863925.1_EquCab3.0/), which excludes snoRNAs, removed all transcripts that had high expression (>60 TPM) in the rRNA-depleted libraries and substantially disproportionate low expression in the poly-A^+ selected libraries. However, the correlation between library preparations slightly decreased (Spearman $r_{(liver,subset)}$=0.425, P=2.47×10^{-19}; Spearman $r_{(parietal)l}$=0.416, P=1.84x10^{-19}; **Fig. 3.5 and Supplemental Fig. 3.S5**), but this is likely due to the decrease in number of transcripts used in the analysis.

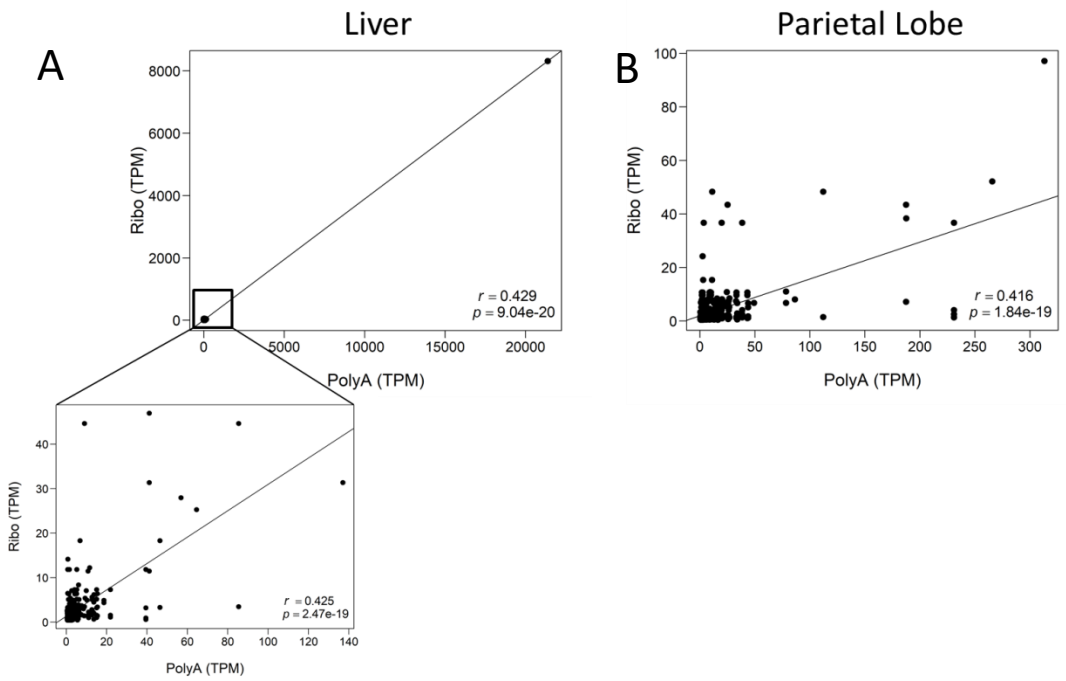


Figure 3.5. Correlation plots for lncRNA expression in AH1 with RefSeq annotated lncRNA. The x-axis indicates expression of the individual lncRNA in the poly-A$^+$ selected library. The y-axis shows expression of the same lncRNAs in the rRNA-depleted library. (**A**) Liver (**B**) Parietal lobe of the cerebrum. Spearman correlation (r) and *p*-value in bottom right corner.

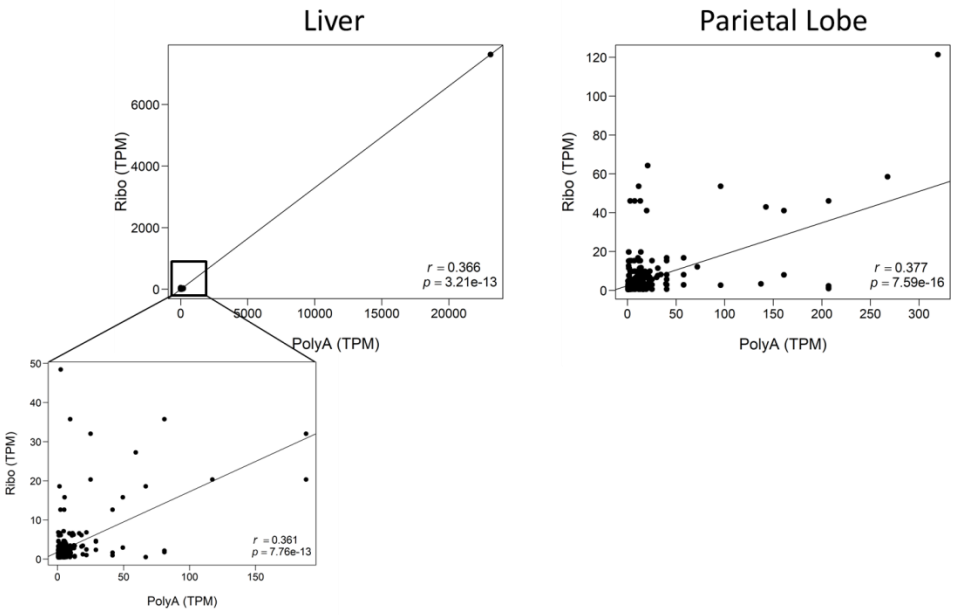


Figure 3.S5. Correlation plots for lncRNA expression in AH2 with RefSeq annotated lncRNA

Discussion

Long ncRNAs play a role in many cellular functions [1–7] and have been implicated in cancer [8–10] and neurodegeneration [11,12]. Additionally, lncRNAs may play a large role in athletic performance in the horse [22]. Thus, the goal of this study was to determine which RNA-seq library preparation method most reliably identifies lncRNA in horses. We hypothesized that rRNA-depleted libraries would be preferable as selection is not dependent on a poly-A tail. However, with an average coverage of 30 M reads, we identified more total lncRNA in the poly-A^+ selected libraries, suggesting that poly-A^+ selection may be more efficient at capturing lncRNAs than rRNA-depletion. It has been shown that some non-polyadenlyated lncRNAs are stabilized by other secondary structure features [23]. These features could have prevented the isolation of the associated transcripts during library preparation, leading to fewer lncRNAs being detected in the rRNA-depleted libraries. Additionally, the high TPM values obtained for snoRNAs suggest that these transcripts are overrepresented in rRNA-depleted libraries, preventing other ncRNAs from being identified. Previous studies on RNA-seq from human cell lines and blood have detected more or equal numbers of lncRNA in rRNA-depleted libraries [14,24–26], though one study explicitly reported that rRNA-depleted libraries yielded fewer usable reads than poly-A^+ selection [14]. The study that found this included RNA from the colon, however this RNA sample was obtained directly from a commercial source and was not isolated by the researchers. Our study is unique in that all RNA was isolated by a single researcher from flash-frozen tissue collected from healthy, well-phenotyped individuals before proceeding to library preparation. This minimizes the potential variation from multiple individuals performing RNA isolations as well as any pathological variation. As a result, we did identify a strong correlation between biologic replicates.

While this study highlights many of the drawbacks of rRNA-depletion library preparation methods, poly-A^+ selection has its own disadvantages. As seen here, some transcriptomic information is lost if it does not have a polyadenylated tail. Additionally, it is well known that poly-A^+ selected libraries have a 3′-bias. In an effort to identify a library preparation method that avoids some of the biggest disadvantages of poly-A^+ selection and rRNA-depletion, the rRNA-depletion protocol could be further optimized. Alternatively, a NuGEN Ovation v2 protocol, which utilized both random and oligo(dT) primers to remove rRNAs, performed well in lncRNA identification in one comparison study [27]. This method addressed the poly-A tail disadvantage; however, it performed poorly when looking at protein-coding transcripts and had a substantial 3′-bias [27]. In short, further research is needed to develop an improved RNA-seq library preparation protocol.

Investigating DE lncRNA between tissues for poly-A^+ selection and rRNA-depletion as a proof of concept identified several lncRNA that are similar to lncRNA found in humans with similar expression differences between brain and liver tissues [20]. This suggests that both methods may be used to annotate lncRNA that is already known in other species. However, more lncRNA transcripts were unique to poly-A^+ selected libraries than to rRNA-depleted libraries, indicating that, for a specific sequencing depth, poly-A^+ selection may yield more informative lncRNA data.

When comparing poly-A^+ selection and rRNA-depletion methods in humans, it is common to limit the comparison to already annotated lncRNA [14,24–26]. While this may be sufficient when using human data, there are not enough lncRNAs annotated in the equine reference genome to identify a substantial correlation between library preparation methods [14,24]. However, we can still observe a moderate correlation between library preparations. rRNA-depletion is often recommended for poor-quality RNA where a full transcript is likely not attached to a polyA-tail

[28]. The RNA used in this study was of high quality and therefore our comparative results only apply to high-quality RNA library preparations.

As suggested in Scott et al. [13], library preparation does play a large role in the lncRNA that are detected. As such, rRNA-depleted datasets should not be considered equivalent to poly-A^+ selected datasets. While this might not raise problems in the annotation of the equine genome, differential transcript expression studies between two equine populations would require additional biologic replicates to overcome the variation between library preparations. A primary advantage of using rRNA-depletion appears to be enhanced identification of snoRNAs. Data from human studies supports this. In a previous report using the HEK293 cell line, a snoRNA was one of the top three highly expressed lncRNA [25]. That particular study raised a valid concern that these highly expressed snoRNAs and similar transcripts lowered the sequencing depth for other RNAs [25]. Similarly, when using pooled blood RNA and a single colon RNA sample, a large portion of rRNA-depleted libraries consisted of a small number of lncRNAs and small RNAs (smRNAs) [14]. Therefore, for annotation of these RNAs, rRNA-depletion would likely be the most thorough. However, poly-A^+ selection can identify these RNAs and may simply require deeper sequencing to better detect these transcripts. Previous study of HEK293 cells support this finding of highly expressed lncRNA from rRNA-depleted libraries also being present in poly-A^+ selected libraries [25].

As only two tissues were used in this study, the results obtained here do not provide a thorough annotation of the equine genome. By evaluating eight different tissues, Scott et al. identified 20,800 putative lncRNA [13]. This number of putative lncRNA far exceeds what was identified in our study; however, tissues used in the Scott et al. study included both nervous and embryonic tissues, which likely have a substantially different lncRNA transcriptional profile as compared to adult horses [29]. In humans, rRNA-depletion has been reported to include more intergenic and

intronic reads than poly-A^+ selection [14]. Unfortunately, as many non-coding RNAs and untranslated regions are not identified in the horse, an accurate measure of the non-exonic reads in our dataset cannot be obtained. Similarly, we do not know the true distribution of lncRNA in horses. However, the identification of similar lncRNA that are differentially expressed between the liver and parietal lobe in both humans and horses suggest a potential method for future annotation.

Potential limitations of this study include the use of only two tissues from two biological replicates. Since the liver and parietal lobe are quite different in terms of cellular make-up complexity, they were considered to be good representative tissues. However, there could be some factors concerning RNA-seq library preparations we are not observing with the limited sample number. A limitation of the pipeline used is a combination of the strict filtering of predicted protein-coding sequences and an ineffective rescue of known lncRNA from human data. All transcripts with an open reading frame (ORF) were filtered out which likely excluded some lncRNA as there are reports of lncRNA with short ORFs in mice [30,31]. As noted from the increase in lncRNA after rescuing filtered out known lncRNA, a substantial number of lncRNA were incorrectly filtered out. An alternative pipeline may remove this part of the filter, though there is then the possibility of retaining unannotated or truncated protein-coding transcripts. Additionally, due to the low sequence conservation of lncRNA between horse and human [13], there are likely lncRNA that are not rescued as this step uses nucleotide BLAST (BLASTN). An improved rescue might utilize lncRNA known to be expressed in a specific tissue in a more thoroughly annotated species, such as human, and identify lncRNA in syntenic regions of the organism of interest, such as horse.

In summary, poly-A^+ selection allowed for the identification of more lncRNA and missed fewer lncRNAs compared to rRNA-depletion. While changes to the pipeline could improve annotation, using poly-A^+ selection in equine samples provides thorough identification of lncRNA.

References

1. Guttman, M.; Amit, I.; Garber, M.; French, C.; Lin, M.F.; Feldser, D.; Huarte, M.; Zuk, O.; Carey, B.W.; Cassady, J.P.; et al. Chromatin Signature Reveals over a Thousand Highly Conserved Large Non-Coding RNAs in Mammals. *Nature* **2009**, *458*, 223–227, doi:10.1038/nature07672.
2. Martens, J.A.; Laprade, L.; Winston, F. Intergenic Transcription Is Required to Repress the Saccharomyces Cerevisiae SER3 Gene. *Nature* **2004**, *429*, 571–574, doi:10.1038/nature02538.
3. Blume, S.W.; Meng, Z.; Shrestha, K.; Snyder, R.C.; Emanuel, P.D. The 5′-Untranslated RNA of the Human Dhfr Minor Transcript Alters Transcription Pre-Initiation Complex Assembly at the Major (Core) Promoter. *J. Cell. Biochem.* **2003**, *88*, 165–180, doi:10.1002/jcb.10326.
4. Martignetti, J.A.; Brosius, J. BC200 RNA: A Neural RNA Polymerase III Product Encoded by a Monomeric Alu Element. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 11563–11567, doi:10.1073/pnas.90.24.11563.
5. Redrup, L.; Branco, M.R.; Perdeaux, E.R.; Krueger, C.; Lewis, A.; Santos, F.; Nagano, T.; Cobb, B.S.; Fraser, P.; Reik, W. The Long Noncoding RNA Kcnq1ot1 Organises a Lineage-Specific Nuclear Domain for Epigenetic Gene Silencing. *Development* **2009**, *136*, 525–530, doi:10.1242/dev.031328.
6. Dinger, M.; Amaral, P.; Mercer, T. Long Noncoding RNAs in Mouse Embryonic Stem Cell Pluripotency and Differentiation. *Genome Res.* **2008**, 1433–1445, doi:10.1101/gr.078378.108.7.

7. Khalil, A.M.; Faghihi, M.A.; Modarresi, F.; Brothers, S.P.; Wahlestedt, C. A Novel RNA Transcript with Antiapoptotic Function Is Silenced in Fragile X Syndrome. *PLoS ONE* **2008**, *3*, doi:10.1371/journal.pone.0001486.
8. Liang, W.-C.; Fu, W.-M.; Wong, C.-W.; Wang, Y.; Wang, W.-M.; Hu, G.-X.; Zhang, L.; Xiao, L.-J.; Wan, D.C.-C.; Zhang, J.-F.; et al. The LncRNA H19 Promotes Epithelial to Mesenchymal Transition by Functioning as MiRNA Sponges in Colorectal Cancer. *Oncotarget* **2015**, *6*, doi:10.18632/oncotarget.4154.
9. Li, H.; Yu, B.; Li, J.; Su, L.; Yan, M.; Zhu, Z.; Liu, B. Overexpression of LncRNA H19 Enhances Carcinogenesis and Metastasis of Gastric Cancer. *Oncotarget* **2014**, *5*, 2318–2329, doi:10.18632/oncotarget.1913.
10. Zhang, Y.; Pitchiaya, S.; Cieślik, M.; Niknafs, Y.S.; Tien, J.C.Y.; Hosono, Y.; Iyer, M.K.; Yazdani, S.; Subramaniam, S.; Shukla, S.K.; et al. Analysis of the Androgen Receptor-Regulated LncRNA Landscape Identifies a Role for ARLNC1 in Prostate Cancer Progression. *Nat. Genet.* **2018**, *50*, 814–824, doi:10.1038/s41588-018-0120-1.
11. Johnson, R. Long Non-Coding RNAs in Huntington's Disease Neurodegeneration. *Neurobiol. Dis.* **2012**, *46*, 245–254, doi:10.1016/j.nbd.2011.12.006.
12. Spreafico, M.; Grillo, B.; Rusconi, F.; Battaglioli, E.; Venturin, M. Multiple Layers of CDK5R1 Regulation in Alzheimer's Disease Implicate Long Non-Coding RNAs. *Int. J. Mol. Sci.* **2018**, *19*, 1–14, doi:10.3390/ijms19072022.
13. Scott, E.Y.; Mansour, T.; Bellone, R.R.; Brown, C.T.; Mienaltowski, M.J.; Penedo, M.C.; Ross, P.J.; Valberg, S.J.; Murray, J.D.; Finno, C.J. Identification of Long Non-Coding RNA in the Horse Transcriptome. *BMC Genomics* **2017**, *18*, 1–11, doi:10.1186/s12864-017-3884-2.
14. Zhao, S.; Zhang, Y.; Gamini, R.; Zhang, B.; Von Schack, D. Evaluation of Two Main RNA-Seq Approaches for Gene Quantification in Clinical RNA Sequencing: PolyA+ Selection versus RRNA Depletion. *Sci. Rep.* **2018**, *8*, 1–12, doi:10.1038/s41598-018-23226-4.

15. Ulitsky, I.; Shkumatava, A.; Jan, C.H.; Sive, H.; Bartel, D.P. Conserved Function of LincRNAs in Vertebrate Embryonic Development Despite Rapid Sequence Evolution. *Cell* **2011**, *147*, 1537–1550, doi:10.1016/j.cell.2011.11.055.Conserved.
16. Hezroni, H.; Koppstein, D.; Schwartz, M.G.; Avrutin, A.; Bartel, D.P.; Ulitsky, I. Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. *Cell Rep.* **2015**, *11*, 1110–1122, doi:10.1016/j.celrep.2015.04.023.
17. Muret, K.; Désert, C.; Lagoutte, L.; Boutin, M.; Gondret, F.; Zerjal, T.; Lagarrigue, S. Long Noncoding RNAs in Lipid Metabolism : Literature Review and Conservation Analysis across Species. *BMC Genomics* **2019**, *20*, 1–18.
18. Wilusz, J.E.; Freier, S.M.; Spector, D.L. 3′ End Processing of Long Nuclear-Retained Non-Coding RNA Yields a TRNA-like Cytoplasmic RNA. *Cell* **2009**, *135*, 919–932, doi:10.1016/j.cell.2008.10.012.3.
19. Cheng, J.; Kapranov, P.; Drenkow, J.; Dike, S.; Brubaker, S.; Patel, S.; Long, J.; Stern, D.; Tammana, H.; Helt, G.; et al. Transcriptional Maps of 10 Human Chromosomes at 5-Nucleotide Resolution. *Science* **2005**, *308*, 1149–1154, doi:10.1126/science.1108625.
20. Burns, E.N.; Bordbari, M.H.; Mienaltowski, M.J.; Affolter, V.K.; Barro, M.V.; Gianino, F.; Gianino, G.; Giulotto, E.; Kalbfleisch, T.S.; Katzman, S.A.; et al. Generation of an Equine Biobank to Be Used for Functional Annotation of Animal Genomes Project. *Anim. Genet.* **2018**, *49*, 564–570, doi:10.1111/age.12717.
21. Lonsdale, J.; Thomas, J.; Salvatore, M.; Phillips, R.; Lo, E.; Shad, S.; Hasz, R.; Walters, G.; Garcia, F.; Young, N.; et al. The Genotype-Tissue Expression (GTEx) Project. *Nat. Genet.* **2013**, *45*, 580–585, doi:10.1038/ng.2653.
22. Capomaccio, S.; Vitulo, N.; Verini-Supplizi, A.; Barcaccia, G.; Albiero, A.; D'Angelo, M.; Campagna, D.; Valle, G.; Felicetti, M.; Silvestrelli, M.; et al. RNA Sequencing of the Exercise Transcriptome in Equine Athletes. *PLoS ONE* **2013**, *8*, doi:10.1371/journal.pone.0083504.

23. Wilusz, J.E.; JnBaptiste, C.K.; Lu, L.Y.; Kuhn, C.D.; Joshua-Tor, L.; Sharp, P.A. A Triple Helix Stabilizes the 3′ Ends of Long Noncoding RNAs That Lack Poly(A) Tails. *Genes Dev.* **2012**, *26*, 2392–2407, doi:10.1101/gad.204438.112.
24. Guo, Y.; Zhao, S.; Sheng, Q.; Guo, M.; Lehmann, B.; Pietenpol, J.; Samuels, D.C.; Shyr, Y. RNAseq by Total RNA Library Identifies Additional RNAs Compared to Poly(A) RNA Library. *Biomed Res. Int.* **2015**, 1–9, doi:10.1155/2015/862130.
25. Sultan, M.; Amstislavskiy, V.; Risch, T.; Schuette, M.; Dökel, S.; Ralser, M.; Balzereit, D.; Lehrach, H.; Yaspo, M.L. Influence of RNA Extraction Methods and Library Selection Schemes on RNA-Seq Data. *BMC Genom.* **2014**, *15*, 1–13, doi:10.1186/1471-2164-15-675.
26. Cui, P.; Lin, Q.; Ding, F.; Xin, C.; Gong, W.; Zhang, L.; Geng, J.; Zhang, B.; Yu, X.; Yang, J.; et al. A Comparison between Ribo-Minus RNA-Sequencing and PolyA-Selected RNA-Sequencing. *Genomics* **2010**, *96*, 259–265, doi:10.1016/j.ygeno.2010.07.010.
27. Chao, H.P.; Chen, Y.; Takata, Y.; Tomida, M.W.; Lin, K.; Kirk, J.S.; Simper, M.S.; Mikulec, C.D.; Rundhaug, J.E.; Fischer, S.M.; et al. Systematic Evaluation of RNA-Seq Preparation Protocol Performance. *BMC Genom.* **2019**, *20*, 1–20, doi:10.1186/s12864-019-5953-1.
28. Schuierer, S.; Carbone, W.; Knehr, J.; Petitjean, V.; Fernandez, A.; Sultan, M.; Roma, G. A Comprehensive Assessment of RNA-Seq Protocols for Degraded and Low-Quantity Samples. *BMC Genom.* **2017**, *18*, 1–13, doi:10.1186/s12864-017-3827-y.
29. Yan, L.; Yang, M.; Guo, H.; Yang, L.; Wu, J.; Li, R.; Liu, P.; Lian, Y.; Zheng, X.; Yan, J.; et al. Single-Cell RNA-Seq Profiling of Human Preimplantation Embryos and Embryonic Stem Cells. *Nat. Struct. Mol. Biol.* **2013**, *20*, 1131–1139, doi:10.1038/nsmb.2660.
30. Ingolia, N.T.; Lareau, L.F.; Weissman, J.S. Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity of Mammalian Proteomes. *Cell* **2011**, *147*, 789–802, doi:10.1038/jid.2014.371.
31. Nelson, B.R.; Makarewich, C.A.; Anderson, D.M.; Winders, B.R.; Troupes, C.D.; Wu, F.; Reese, A.L.; McAnally, J.R.; Chen, X.; Kavalali, E.T.; et al. A Peptide Encoded by a

Transcript Annotated as Long Noncoding RNA Enhances SERCA Activity in Muscle. *Science* **2016**, *351*, 271–275, doi:10.1038/nbt.3301.Mammalian.

32. Joshi, N.; Fass, J. *Sickle: A Sliding-Window, Adaptive, Quality-Based Trimming Tool for FastQ Files,* Version 1.33 [Software]; 2011. Available at https://github.com/najoshi/sickle.
33. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* **2013**, *29*, 15–21, doi:10.1093/bioinformatics/bts635.
34. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079, doi:10.1093/bioinformatics/btp352.
35. Pertea, M.; Pertea, G.M.; Antonescu, C.M.; Chang, T.C.; Mendell, J.T.; Salzberg, S.L. StringTie Enables Improved Reconstruction of a Transcriptome from RNA-Seq Reads. *Nat. Biotechnol.* **2015**, *33*, 290–295, doi:10.1038/nbt.3122.
36. Kalbfleisch, T.S.; Rice, E.S.; DePriest, M.S.; Walenz, B.P.; Hestand, M.S.; Vermeesch, J.R.; O′Connell, B.L.; Fiddes, I.T.; Vershinina, A.O.; Saremi, N.F.; et al. Improved Reference Genome for the Domestic Horse Increases Assembly Contiguity and Composition. *Commun. Biol.* **2018**, *1*, 1–8, doi:10.1038/s42003-018-0199-z.
37. Pertea, G. GffCompare. Available online: http://ccb.jhu.edu/software/stringtie/gffcompare.shtml (accessed on April 6, 2020).
38. Wickham, H.; François, R.; Henry, L.; Müller, K. *Dplyr*; 2018. Available at https://cran.r-project.org/web/packages/dplyr/index.html
39. Sonnhammer, E.L.; Eddy, S.R.; Durbin, R. Pfam: A Comprehensive Database of Protein Domain Families Based on Seed Alignments. *Proteins Struct. Funct. Bioinforma.* **1997**, *28*, 405–420.
40. Finn, R.D.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Mistry, J.; Mitchell, A.L.; Potter, S.C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; et al. The Pfam Protein Families Database:

Towards a More Sustainable Future. *Nucleic Acids Res.* **2016**, *44*, 279–285, doi:10.1093/nar/gkv1344.

41. Quinlan, A.R.; Hall, I.M. BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features. *Bioinformatics* **2010**, *26*, 841–842, doi:10.1093/bioinformatics/btq033.
42. Robinson, M.; McCarthy, D.; Smyth, G. EdgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* **2010**, *26*, 139–140, doi:10.1093/bioinformatics/btp616.

Chapter 4: Transcriptomic Markers of Recombinant Human Erythropoietin Micro-Dosing in Thoroughbred Horses

Authors: Anna R. Dahlgren[1], Heather K. Knych[2], Rick Arthur[3], Blythe Durbin-Johnson[4], Carrie J. Finno[1]

[1] Department of Population Health and Reproduction, School of Veterinary Medicine, University of California Davis, Davis, CA 95616

[2] K.L. Maddy Equine Analytical Pharmacology Lab and Department of Molecular Biosciences, School of Veterinary Medicine, University of California Davis, Davis, CA 95616

[3] School of Veterinary Medicine, University of California Davis, Davis, CA 95616

[4] Bioinformatics Core Facility, Genome Center, University of California, Davis, CA 95616

Key words: RNA sequencing, biomarkers, rHuEPO, doping

Abstract

Recombinant human erythropoietin (rHuEPO) is a well-known performance enhancing drug in human athletes, and there is anecdotal evidence of it being used in horse racing for the same purpose. rHuEPO, like endogenous EPO, increases arterial oxygen content and thus aerobic power. Micro-doping, or injecting smaller doses over a longer period of time, has become an important concern in both human and equine athletics since it is more difficult to detect. Horses offer an additional challenge of a contractile spleen, thus large changes in the red blood cell mass occur naturally. To address the challenge of detecting rHuEPO doping in horse racing, we determined the transcriptomics effects of rHuEPO micro-dosing over seven weeks in exercised Thoroughbreds. RNA-sequencing of peripheral blood mononuclear cells isolated at several time points throughout the study identified three transcripts (*C13H16orf54, PUM2* and *CHTOP*) that

were significantly (P_{FDR}<0.05) different between the treatment groups across two or three time point comparisons. *PUM2* and *CHTOP* play a role in erythropoiesis while not much is known about *C13H16orf54*, but it is primarily expressed in whole blood. However, gene expression differences were not large enough to detect via RT-qPCR, thereby precluding their utility as biomarkers of micro-doping.

Introduction

Erythropoietin (EPO) is a protein predominantly secreted from the kidney to stimulate red blood cell (RBC) production in the bone marrow. As such, it increases hemoglobin mass, arterial oxygen content, and thus aerobic power [1]. Recombinant human erythropoietin (rHuEPO) is well known as a performance enhancing agent in human athletes [2–4]. Similar to training at higher elevations to naturally stimulate EPO [5], injecting rHuEPO has the same end result of enhancing an individual's aerobic power [2]. This has led to its abuse in human sports and placement on the World Anti-Doping Agency's list of prohibited substances [6]. rHuEPO has also been used clinically to replace endogenous EPO in human patients suffering from various conditions associated with a decline in red blood cell counts [7].

In human sports, various methods of detecting rHuEPO administration have been employed. These include setting upper limits for hemoglobin and hemacrit levels [8], using isoelectric focusing (IEF) to detect differences in the charge profiling of endogenously versus exogenously produced EPO in urine [9], and using sarcosyl polyacrylamide gel electrophoresis (SAR-PAGE) [10] to detect rHuEPO in serum and urine. For the latter two methods, the detection window is quite short, ranging from 24-85 hours from time of administration [9,10]. Recently, there have been reports of micro-dosing, where small amounts of rHuEPO are administered over a long period of time to avoid detection [11]. Currently, the Athlete Biological Passport (ABP) regulates

rHuEPO abuse and other blood doping agents in human athletes indirectly by monitoring for changes in hematological values such as hemoglobin, reticulocytes, and red blood cell count in an individual overtime [11]. However, this screening method fails to detect micro-doses of rHuEPO [11]. An alternative that has recently been investigated is using transcriptomic biomarkers to detect micro-dosing and supplement the ABP [12].

In Thoroughbred racehorses, there is anecdotal evidence of rHuEPO administration to enhance performance. A single small study has shown that rHuEPO administration does increase aerobic power and performance in horses, but these horses were splenectomized [13], which prevents horses from sustaining athletic performance [14]. There are no other studies to date that have evaluated the efficacy of rHuEPO administration, but methods of detection of rHuEPO administration have been evaluated in horses [15–18]. Routine testing of samples from equine athletes for detection of rHuEPO doping has been performed with liquid chromatography tandem mass spectrometry (LC-MS/MS) based assays [19,20], which are time consuming and cumbersome. Additionally, the detection window is short and micro-dosing is unlikely to be detected [19,20] since rHuEPO has a half-life of about 12.9 hours in horses [21]. Monitoring hematological values over time, similar to the screening performed in human athletes, is not a viable option for detecting rHuEPO doping in horses due to their high RBC reserve and contractile spleen [14]. Several methods of detection have been investigated in the horse, including serial analysis of gene expression libraries [15], digital droplet qPCR [16], microfluidic quantitative PCR [17], and whole-genome resequencing [18]. However, none of the studies addressed micro-dosing detection and the bulk of the work optimizing testing methods were primarily *in vitro*. The studies that did administer rHuEPO to horses to test their methods did not use horses that were at a similar age or fitness level as racehorses [15–17].

To investigate the effects of rHuEPO micro-dosing in exercising Thoroughbred horses, we performed RNA-sequencing (RNA-seq) of peripheral blood mononuclear cells (PBMCs) in an experimental study. All horses were on an exercise schedule that simulated the workload of Thoroughbred racehorses. Our goal was to identify transcriptomic markers of rHuEPO micro-dosing to identify illicit doping of racehorses. We hypothesized that genes involved in erythropoiesis would be differentially expressed in micro-dosed horses, thereby providing a mechanism for testing of micro-dosing in exercising Thoroughbreds.

Methods

Horses

Five Thoroughbred mares, aged 4-7 years (median=6), and 5 Thoroughbred geldings, aged 5-6 years (median=5), were examined by a veterinarian and determined to be healthy. These ten horses were put on a consistent exercise protocol prior to and throughout the study. This protocol consisted of three days a week on an Equicizer (5 minutes walk, 20 minutes trot, 5 minutes walk; Centaur Horse Walkers Inc, Mira Loma, CA) and two days per week on a high speed treadmill (5 min walk at 1.6 m/s, then 5 min trot at 4m/s, 5 min canter at 7m/s and a cool out of 5 min walk at 1.6 m/s, then 5 min walk at 1.6 m/s, incline to 4% and 10 min trot at 4 m/s and then 5 min walk at 1.6 m/s; Mustang 2200, Graber AG, Switzerland). Before beginning the study, the ten horses were determined to be healthy by physical examination by a veterinarian, complete blood count, and serum biochemistry panel. The horses did not receive any other medication for at least two weeks prior to beginning the study.

Six of the horses (3 mares, 3 geldings) were randomly assigned to the treatment group and the remaining four horses (2 mares, 2 geldings) to the saline control group. The horses in the treatment group received a subcutaneous injection of 20 IU/kg of rHuEPO (EPOGEN, Amgen,

Thousand Oaks, CA) into the loose skin on the lower chest two times a week, alternating sides, **(Fig. 4.1)** for a total of seven weeks. This dose and frequency was chosen based on a previous study of micro-doping done in humans [11]. The route of administration was chosen based on previous studies that investigated markers of rHuEPO doping [15]. Horses in the control group received a comparable volume of 0.9% saline subcutaneously at the same times. All animal procedures were approved by the University of California-Davis Institutional Animal Care and Use Committees (IACUC #20319).

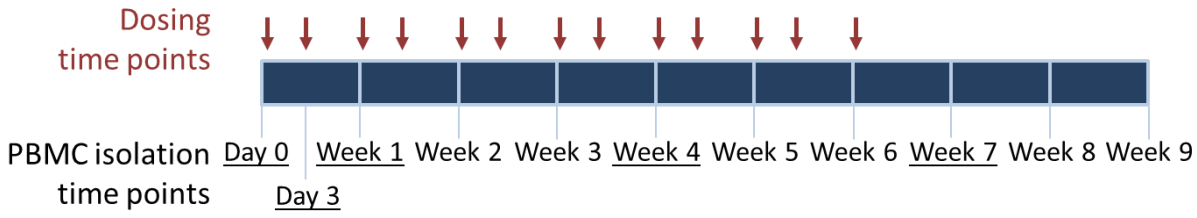


Figure 4.1. Schematic outlining time points of rHuEPO or saline dosing and blood draws for peripheral blood mononuclear cell (PBMC) isolations. The underlined time points indicate the time points that were submitted for RNA sequencing.

Sample Collection

Twenty mL of blood per horse were drawn into EDTA vacutainers immediately prior to initial treatment, three days after initial treatment, once a week for the remainder of seven weeks, and once a week for three weeks following the final rHuEPO treatment.

PBMC and RNA isolations

Peripheral blood mononuclear cells were isolated from the EDTA blood using Histopaques 1119 and 1077 (Sigmal Aldrich, St. Louis, MO) following the company's protocol, with the exception that 5 mL of blood was layered onto the histopaques discontinuous gradient and only the

mononuclear layer was collected. RNA was isolated immediately after PBMC isolation using a trizol-chloroform phase separation followed by a column clean up (Zymo, Irvine, CA).

RNA sequencing and analysis

RNA samples from all horses from day 0, day 3, week 1, week 4, and week 7 were chosen for sequencing. All RNA had an RNA integrity number >7. Strand-specific libraries (Universal Plus mRNA) were created and sequenced on an Illumina NovaSeq to a targeted depth of 30 million reads/sample.

Raw reads were trimmed with trimmomatic [22] to remove low quality reads and adapter sequences. The read were mapped to EquCab3.0 using STAR aligner [23] and the RefSeq annotation (GCF_002863925.1_EquCab3.0). Count data was normalized with transcript per million mapped read normalization in EdgeR (version 3.34.0) [24] and limma-voom (3.48.1) [25] was used to analyze differential expression. Transcripts that were significantly different between treatment groups across two time points were identified using a model including factors for treatment group, time point (as a categorical variable), the interaction between treatment and time point, and sex. These transcripts were then prioritized to those that were significantly different across at least two time point comparisons. Reported *P*-values were adjusted via the Benjamini-Hochberg false discovery rate (FDR), with $P_{FDR}<0.05$ considered significant.

In order to compare RNA-seq results to these obtained from the RT-qPCR study, counts per million (CPMs) for the top three differentially expressed transcripts were transformed by taking the $\log_2$ of the CPM+1 and then graphed. A mixed-effects model, where horse was a random effect and time point, treatment, and their interaction were fixed, was used to confirm significant differences between treatment groups. GraphPad Prism (GraphPad Software, San Diego, California USA, www.graphpad.com) was used to create the model and perform post-hoc

multiple comparison testing. Stringtie [26] was used to determine if there were multiple isoforms of the three transcripts of interest present in the RNA-seq.

RT-q PCR validation

To validate the top transcripts across all time points, primers were designed to span two exons of the target transcripts (**Supplemental Table 4.S1)**. The RNA-seq data was used to design the primers and ensure reads were present at the primer locations. The same aliquot of RNA (1200 ng) that was used for RNA-seq from each horse at every time point was reverse transcribed into cDNA with Superscript III (Thermo Fisher Scientific, Waltham, MA). Reverse transcriptase quantitative PCR (RT-qPCR) was performed in triplicate using SYBR green on an AriaMx Real-Time PCR System (Agilent, Santa Clara, CA). The housekeeping gene *B2M* was used for normalization. The ΔΔCt values and fold changes were calculated to determine differential transcript expression between treatment groups and time points. A 2-way repeated measures ANOVA was used to identify if there was any significant differences between saline and rHuEPO treated groups or time points.

Supplemental Table 4.S1: Primer sequences

Transcript	**Coordinates**	**Forward Sequence**	**Reverse Sequence**
C13H16orf54	chr13:20,454,043-20,454,707	CAGCATGATGGGGATACAGG	AGAGAGGAGCCCAGGACTGA
PUM2	chr15:76,186,319-76,188,177	CTGCTGAGCGACAAATGGTA	TGACCACGAATACGAGTAGCC
CHTOP	chr5:40,507,416-	GCATCGGCTGTTTGTTC	AGTCAGCGCCGAAAGT

	40,508,838	TTC	TG
B2M	chr1:145,961,343-145,964,675	TCGGGCTACTCTCCCTGACT	CGTGAGTAAACCTGAACCTTCG

Results

Horses

All horses successfully completed both the exercise and micro-dosing protocol, except for one horse (#790) in the treatment group that colicked after rHuEPO administration ended and temporarily stopped the exercise protocol. There is no evidence that the colic was related to rHuEPO administration. Administration of subcutaneous rHuEPO did result in apparent discomfort in 5/6 horses after each injection, as indicated by pawing after the injection, leg swelling in one horse, and hives and swelling or edema at the injection site in four horses the day after injection.

RNA sequencing and differential transcript analysis

Our targeted depth of 30 million reads/sample was achieved, with an average of 37.2M raw reads and 36.8M trimmed reads. Multidimensional scale analysis indicated that there was no obvious grouping of samples by time point, treatment, or individual **(Supplemental Fig. 4.S1)**.

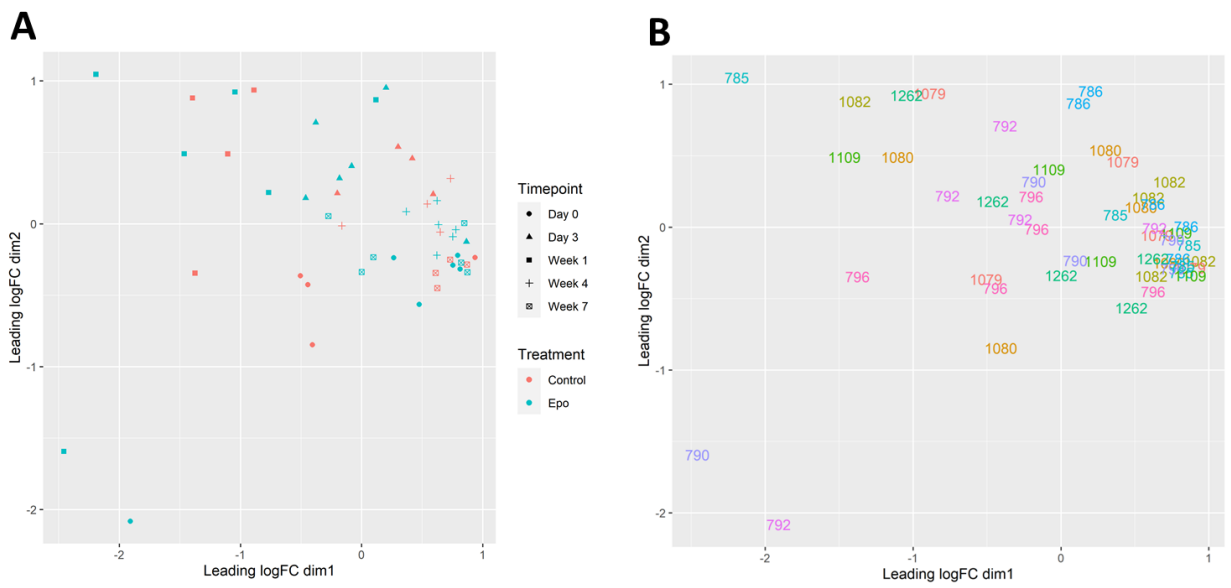


Figure 4.S1. Multidimensional scaling (MDS) plots of read counts normalized by EdgeR. (A) MDS plot indicating time point by shape and treatment group by color. (B) MDS plot showing each individual horse (denoted by each individuals' numbers).

Twenty-one transcripts were determined to be differentially expressed at P_{FDR}<0.05 **(Supplemental Table 4.S2)**. Of the transcripts that were only significant at one time point comparison, 13 were downregulated with EPO administration and 4 were upregulated. Four transcripts (*C13H16orf54* upregulated, *LOC106783439* upregulated then downregulated, *PUM2* upregulated, *CHTOP* downregulated) were significantly dysregulated across two or more time comparisons **(Supplemental Table 4.S2, bolded)** and prioritized for RT-qPCR.

Table 4.S2: Transcripts significantly differentially expressed between treatment groups (EPO vs Control) across two time points. D = day, logFC = log of fold change, AveExpr = average expression across all samples, *P*.Value = raw *P*-value, adj.*P*.Val = FDR corrected *P*-value. Bolded transcripts are those that are significantly differentially expressed at more than one time comparison.

Time point	Gene.name	logFC	AveExpr	*P*.Value	adj.*P*.Val
D7 v D0	BCORL1	-6.91	3.58	1.04E-06	7.90E-03
	C13H16orf54	6.68	3.98	1.48E-06	7.90E-03
	LOC100630352	6.32	5.13	2.25E-06	8.04E-03
	SLC16A13	-6.61	4.45	7.80E-06	1.53E-02
	ARHGEF3	-7.18	3.39	1.06E-05	1.53E-02
	LOC111776172	-4.96	5.93	1.11E-05	1.53E-02
	LOC106783439	4.70	5.05	1.14E-05	1.53E-02
	LOC111767722	-6.38	2.94	1.15E-05	1.53E-02
	PUM2	4.40	5.19	1.28E-05	1.53E-02
	CCDC6	-6.17	2.92	1.43E-05	1.53E-02
	INTS10	-6.56	2.85	1.61E-05	1.57E-02
	SAV1	-6.48	3.80	2.74E-05	2.45E-02
	RBBP7	-5.70	4.16	3.41E-05	2.80E-02
	ACVR1B	-6.41	3.24	3.66E-05	2.80E-02
	CHTOP	-4.03	4.68	4.42E-05	3.15E-02
	NSG1	5.34	3.92	7.67E-05	4.98E-02
	ZBTB14	-6.47	1.41	7.91E-05	4.98E-02
D7 v D3	**LOC106783439**	5.31	5.05	2.76E-07	2.42E-03
	CHTOP	-5.00	4.68	4.53E-07	2.42E-03
	PUM2	4.11	5.19	1.21E-05	4.32E-02
	C13H16orf54	5.55	3.98	2.08E-05	4.64E-02
	SLC20A1	-5.43	3.42	2.17E-05	4.64E-02

D28 v D0	SNX32	-0.39	13.85	9.48E-07	1.01E-02
D28 v D7	**LOC106783439**	-5.08	5.05	2.90E-07	3.10E-03
	LOC100066849	7.49	2.30	1.32E-06	7.08E-03
	WBP11	5.08	5.08	3.27E-06	1.17E-02
	CHTOP	4.24	4.68	1.25E-05	3.35E-02
D49 v D7	**LOC106783439**	-5.90	5.05	7.55E-09	8.08E-05

RT-qPCR

Of the four prioritized transcripts (**Supplemental Table 4.S2, bolded)**, one transcript (*LOC106783439*) did not have any exons that aligned between individual horses in our data set or with the reference exons, indicating an error with the reference. This transcript was thus excluded from further investigation. Primers were successfully designed for the remaining three transcripts (*C13H16orf54, PUM2* and *CHTOP*; **Supplementary Table 4.S1**) and reference transcript (*B2M*; **Supplementary Table 4.S1**). There were no significant differences identified between treatment groups for any of the transcripts of interest via RT-qPCR (**Fig. 4.2**).

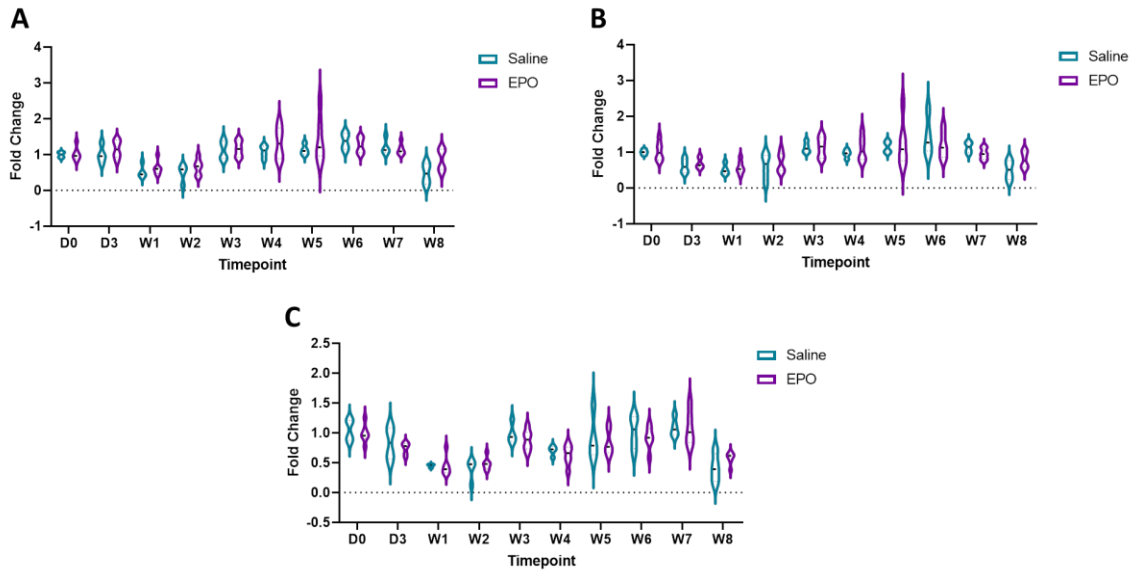


Figure 4.2. Fold change of (A) *C13H16orf54,* (B) *PUM2,* and (C) *CHTOP* mRNA as determined by RT-qPCR at each time point. Expression is in relation to *B2M*. D stands for day and W indicates week. Each data point is an individual horse, saline treatment as teal and EPO-dosed horses as purple. There was no significant difference of expression between treatment groups.

To confirm that the results of our model were present in the raw data, we graphed the CPMs (**Supplemental Fig. 4.S2**) of the three transcripts of interest at each time point. After transforming the CPMs, *PUM2* was not significant between treatments, but *C13H16orf54* and *CHTOP* were significantly different at 7 days (P=2.9x10^{-5}, P=0.016, respectively).

We hypothesized that there may be multiple isoforms of the transcripts of interest and that differential expression of the transcripts is what is leading to the conflicting results between the RNA-seq and RT-qPCR. We determined that, in our data set, there were nine *PUM2* isoforms and five isoforms of *CHTOP*. Seventeen of the possible 22 exons in *PUM2* are conserved across all transcripts present in the data set, and in *CHTOP*, there are 4 exons conserved across all transcripts with 7 total possible exons. The primers for RT-qPCR were designed to amplify exons that were conserved across the majority of isoforms. Primers to amplify *PUM2* spanned exons 6-7, and primers for *CHTOP* spanned exons 2-3. However, this prevented the detection of different isoforms.

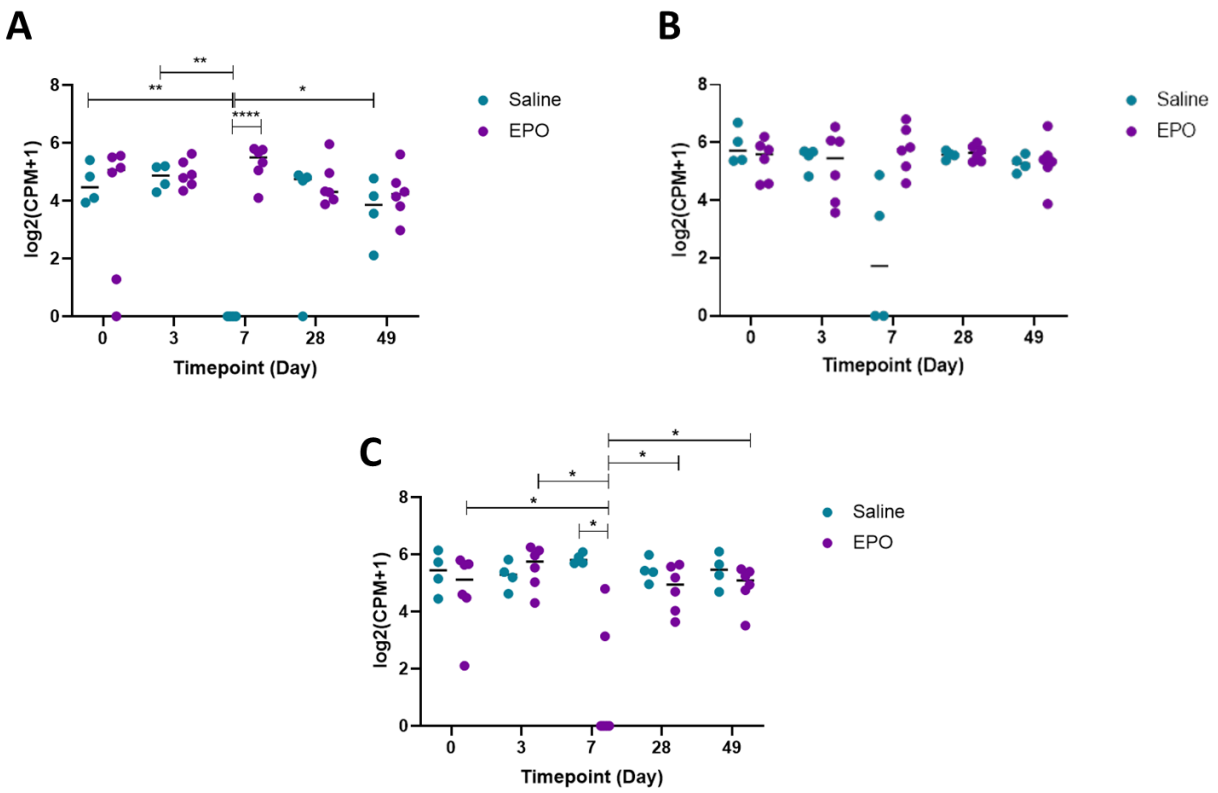


Figure 4.S2. CPMs transformed by $\log_2$(CPM+1) of (A) *C13H16orf54,* (B) *PUM2,* and (C) *CHTOP* across five time points (in days) from RNA-seq. Each data point is an individual horse with saline treatment as teal and EPO-dosed horses as purple. There is a significant difference between treatment groups at 7 days for *C13H16orf54* and *CHTOP. * P_{FDR}<0.05, ** P_{FDR}<0.01, **** P_{FDR}<0.0001*

Discussion

This study set out to identify transcriptomic markers of rHuEPO micro-doping in Thoroughbred racehorses. Like endogenous EPO, rHuEPO binds to the EPO receptor on erythroid progenitor cells, initiating a signaling cascade that leads to the binding of key transcription factors that induce production of more red blood cells [27]. Increasing the number of red blood cells increases the total oxygen available and aerobic power [1]. Thus, we hypothesized that we would identify significantly differential transcripts that were involved in erythropoiesis.

RNA-seq from PBMCs isolated throughout the experiment identified three transcripts that changed significantly over time between treatment groups. *C13H16orf54* and *PUM2* were upregulated and *CHTOP* was downregulated with EPO administration. The function of *C13H16orf54* is unknown, but in humans, it is primarily expressed in whole blood [28]. Slightly more is known about the involvement of *PUM2* and *CHTOP* in erythropoiesis. *PUM2,* an RNA-binding protein, has been shown to have a role in hematopoietic stem cell survival and proliferation [29]. Hematopoietic stem cells are progenitor cells to erythroid stem cells that develop into red blood cells [30], so it follows that *PUM2* changes with rHuEPO dosing. *CHTOP* is an important regulator of γ-globin gene expression, which is a fetal globin gene [31]. A previous small study administered rHuEPO to patients with sickle cell anemia and β-thalassemia and identified an increase in their γ-globin levels [32]. rHuEPO has been further investigated in the treatment of β-thalassemia patients in combination with the fetal globin gene inducer butyrate [33]. A pilot study demonstrated that the rHuEPO was required in some patients to respond to butyrate [33], indicating that rHuEPO and γ-globin, can work synergistically to increase total hemoglobin concentrations. Our study further supports the relationship between rHuEPO dosing and γ-globin expression.

With the significant gene expression differences not validated using RT-qPCR, the development of a diagnostic test to detect rHuEPO doping in racehorses is hindered. Although the same RNA was used for both RNA-seq and RT-qPCR, RT-qPCR is subject to more variability [34]. Complementary cDNA synthesis and use of a reference gene in particular add to the variability of the technology [34]. Additionally, since both *PUM2* and *CHTOP* have several known transcript isoforms, specific isoforms may be differentially expressed with rHuEPO administration. In our data set, there were nine *PUM2* isoforms and five isoforms of *CHTOP*. The RT-qPCR primers were designed to target the exons that were most conserved throughout

the isoforms. However, this might have masked significant isoforms that were detected by the more sensitive RNA-seq. This poses a point of consideration for future studies evaluating transcriptomic markers of doping.

A limitation to this study is the small sample size of ten horses, with six administered rHuEPO and four saline. Six horses have been shown to be a sufficient population size to identify significantly differential gene expression in horses [35]. In an effort to identify differentially expressed transcripts with a limit of ten horses, the horses were separated into unbalanced groups. However, our results require validation in another larger population of exercising Thoroughbreds. Furthermore, the dosing regimen selected for this study was designed based on previous studies done in horses and humans [11,15], but it was not evaluated to determine effect on performance.

Recent, highly-publicized positive drug results in horse racing have highlighted the need for reliable tests. While these were not identified as rHuEPO, preventing any illicit drug doping protects the health and safety of both horses and jockeys, and it protects the integrity of a sport that has a substantial betting culture. A case study was published reporting on two Thoroughbreds that developed anti-rHuEPO antibodies which ended up cross-reacting with endogenous EPO, decreasing erythropoiesis and causing anemia [36]. While the risk of this happening more broadly is not known, it is further motivation to deter the use of rHuEPO as a performance enhancing drug.

In summary, we identified two transcripts that were significantly upregulated and one transcript that was significantly downregulated in horses administered micro-doses of rHuEPO. The changes in *C13H16orf54, PUM2* and *CHTOP* provide insight into the effects of rHuEPO dosing. Since these differences were not detectable via RT-qPCR, these transcripts are not suitable

biomarkers of rHuEPO micro-doping. Further work is required to validate these findings and determine the optimal way to use transcriptomic data to inform micro-doping detection tests.

References

1. Jacob, J.; John, Mj.; Jaison, V.; Jain, K.; Kakkar, N. Erythropoietin Use and Abuse. *Indian J. Endocrinol. Metab.* **2012**, *16* (2), 220–227. https://doi.org/10.4103/2230-8210.93739.
2. Ekblom, B.; Berglund, B. Effect of Erythropoietin Administration on Mammal Aerobic Power. *Scand. J. Med. Sci. Sports* **1991**, *1* (2), 88–93. https://doi.org/10.1111/j.1600-0838.1991.tb00276.x.
3. Thomsen, J. J.; Rentsch, R. L.; Robach, P.; Calbet, J. A. L.; Boushel, R.; Rasmussen, P.; Juel, C.; Lundby, C. Prolonged Administration of Recombinant Human Erythropoietin Increases Submaximal Performance More than Maximal Aerobic Capacity. *Eur. J. Appl. Physiol.* **2007**, *101* (4), 481–486. https://doi.org/10.1007/s00421-007-0522-8.
4. Durussel, J.; Daskalaki, E.; Anderson, M.; Chatterji, T.; Wondimu, D. H.; Padmanabhan, N.; Patel, R. K.; McClure, J. D.; Pitsiladis, Y. P. Haemoglobin Mass and Running Time Trial Performance after Recombinant Human Erythropoietin Administration in Trained Men. *PLoS One* **2013**, *8* (2), 1–8. https://doi.org/10.1371/journal.pone.0056151.
5. Siri, W. E.; Van Dyke, D. C.; Winchell, H. S.; Pollycove, M.; Parker, H. G.; Cleveland, A. S. Blood , and Physiological Responses To Severe Hypoxia in Man. *J. Appl. Physiol.* **1966**, *21*, 73–80.
6. Wold-Anti Doping Agency. The 2021 WADA Prohibited List. *World Anti-Doping* **2021**, 1–24.
7. Ng, T.; Marx, G.; Littlewood, T.; Macdougall, I. Recombinant Erythropoietin in Clinical Practice. *Postgrad. Med. J.* **2003**, *79* (933), 367–376. https://doi.org/10.1136/pmj.79.933.367.

8. Parisotto, R.; Gore, C. J.; Emslie, K. R.; Ashenden, M. J.; Brugnara, C.; Howe, C.; Martin, D. T.; Trout, G. J.; Hahn, A. G. A Novel Method Utilizing Markers of Altered Erythropoiesis for the Detection of Recombinant Human Erythropoietin Abuse in Athletes. *Haematologica* **2000**, *85* (6), 564–572.

9. Lundby, C.; Achman-Andersen, N. J.; Thomsen, J. J.; Norgaard, A. M.; Robach, P. Testing for Recombinant Human Erythropoietin in Urine: Problems Associated with Current Anti-Doping Testing. *J. Appl. Physiol.* **2008**, *105* (2), 417–419. https://doi.org/10.1152/japplphysiol.90529.2008.

10. Schwenke, D. Improved Detection of EPO in Blood and Urine Based on Novel Velum SAR Precast Horizontal Gels Optimized for Routine Analysis. **2015**, 1–6.

11. Ashenden, M.; Gough, C. E.; Garnham, A.; Gore, C. J.; Sharpe, K. Current Markers of the Athlete Blood Passport Do Not Flag Microdose EPO Doping. *Eur. J. Appl. Physiol.* **2011**, *111* (9), 2307–2314. https://doi.org/10.1007/s00421-011-1867-6.

12. Wang, G.; Durussel, J.; Shurlock, J.; Mooses, M.; Fuku, N.; Bruinvels, G.; Pedlar, C.; Burden, R.; Murray, A.; Yee, B.; Keenan, A.; McClure, J. D.; Sottas, P. E.; Pitsiladis, Y. P. Validation of Whole-Blood Transcriptome Signature during Microdose Recombinant Human Erythropoietin (RHuEpo) Administration. *BMC Genomics* **2017**, *18* (Suppl 8). https://doi.org/10.1186/s12864-017-4191-7.

13. McKeever, K. H.; McNally, B. A.; Hinchcliff, K. W.; Lehnhard, R. A.; Poole, D. C. Effects of Erythropoietin on Systemic Hematocrit and Oxygen Transport in the Splenectomized Horse. *Respir. Physiol. Neurobiol.* **2016**, *225*, 38–47. https://doi.org/10.1016/j.resp.2016.02.001.

14. Persson, S. G. B.; Ekman, L.; Lydin, G.; Tufvesson, G. Circulatory Effects of Splenectomy in the Horse: III. Effect on Pulse-work Relationship. *Zentralblatt für Veterinärmedizin R. A* **1973**, *20* (7), 521–530. https://doi.org/10.1111/j.1439-0442.1973.tb01066.x.

15. Bailly-Chouriberry, L.; Noguier, F.; Manchon, L.; Piquemal, D.; Garcia, P.; Popot, M. A.; Bonnaire, Y. Blood Cells RNA Biomarkers as a First Long-Term Detection Strategy for EPO Abuse in Horseracing. *Drug Test. Anal.* **2010**, *2* (7), 339–345. https://doi.org/10.1002/dta.146.

16. Tozaki, T.; Ohnuma, A.; Takasu, M.; Kikuchi, M.; Kakoi, H.; Hirota, K. I.; Kusano, K.; Nagata, S. I. Droplet Digital Pcr Detection of the Erythropoietin Transgene from Horse Plasma and Urine for Gene-Doping Control. *Genes (Basel).* **2019**, *10* (3). https://doi.org/10.3390/genes10030243.

17. Tozaki, T.; Ohnuma, A.; Kikuchi, M.; Ishige, T.; Kakoi, H.; Hirota, K. I.; Kusano, K.; Nagata, S. I. Microfluidic Quantitative PCR Detection of 12 Transgenes from Horse Plasma for Gene Doping Control. *Genes (Basel).* **2020**, *11* (4). https://doi.org/10.3390/genes11040457.

18. Tozaki, T.; Ohnuma, A.; Takasu, M.; Nakamura, K.; Kikuchi, M.; Ishige, T.; Kakoi, H.; Hirora, K. ichi; Tamura, N.; Kusano, K.; Nagata, S. ichi. Detection of Non-Targeted Transgenes by Whole-Genome Resequencing for Gene-Doping Control. *Gene Ther.* **2020**. https://doi.org/10.1038/s41434-020-00185-y.

19. Fragkaki, A. G.; Kioukia-Fougia, N.; Kiousi, P.; Kioussi, M.; Tsivou, M. Challenges in Detecting Substances for Equine Anti-Doping. *Drug Test. Anal.* **2017**, *9* (9), 1291–1303. https://doi.org/10.1002/dta.2162.

20. Guan, F.; Uboh, C. E.; Soma, L. R.; Birks, E.; Chen, J.; Mitchell, J.; You, W.; Rudy, J.; Xu, F.; Li, X.; Mbuy, G. LC-MS/MS Method for Confirmation of Recombinant Human Erythropoietin and Darbepoetin α in Equine Plasma. *Anal. Chem.* **2007**, *79* (12), 4627–4635. https://doi.org/10.1021/ac070135o.

21. Souillard, A.; Audran, M.; Bressolle, F.; Jaussaud, P.; Gareau, R. Pharmacokinetics and Haematological Parameters of Recombinant Human Erythropoietin after Subcutaneous Administrations in Horses. *Biopharm. Drug Dispos.* **1996**, *17* (9), 805–815.

https://doi.org/10.1002/(SICI)1099-081X(199612)17:9<805::AID-BDD995>3.0.CO;2-H.

22. Bolger, A. M.; Lohse, M.; Usadel, B. Genome Analysis Trimmomatic : A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30* (15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

23. Dobin, A.; Davis, C. A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T. R. STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* **2013**, *29* (1), 15–21. https://doi.org/10.1093/bioinformatics/bts635.

24. Robinson, M. D.; McCarthy, D. J.; Smyth, G. K. EdgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* **2010**, *26* (1), 139–140. https://doi.org/10.1093/bioinformatics/btp616.

25. Law, C. W.; Chen, Y.; Shi, W.; Smyth, G. K. Voom: Precision Weights Unlock Linear Model Analysis Tools for RNA-Seq Read Counts. *Genome Biol.* **2014**, *15*. https://doi.org/10.1186/gb-2014-15-2-r29.

26. Pertea, M.; Pertea, G. M.; Antonescu, C. M.; Chang, T. C.; Mendell, J. T.; Salzberg, S. L. StringTie Enables Improved Reconstruction of a Transcriptome from RNA-Seq Reads. *Nat. Biotechnol.* **2015**, *33* (3), 290–295. https://doi.org/10.1038/nbt.3122.

27. He, R.; Oliveira, J. L.; Hoyer, J. D.; Viswanatha, D. S. 24 - Molecular Hematopathology. In *Foundations in Diagnostic Pathology*; Hsi, E. D. B. T.-H. (Third E., Ed.; Elsevier: Philadelphia, 2018; pp 712-760.e18. https://doi.org/https://doi.org/10.1016/B978-0-323-47913-4.00024-0.

28. Lonsdale, J.; Thomas, J.; Salvatore, M.; Phillips, R.; Lo, E.; Shad, S.; Hasz, R.; Walters, G.; Garcia, F.; Young, N.; Foster, B.; Moser, M.; Karasik, E.; Gillard, B.; Ramsey, K.; Sullivan, S.; Bridge, J.; Magazine, H.; Syron, J.; Fleming, J.; Siminoff, L.; Traino, H.; Mosavel, M.; Barker, L.; Jewell, S.; Rohrer, D.; Maxim, D.; Filkins, D.; Harbach, P.; Cortadillo, E.; Berghuis, B.; Turner, L.; Hudson, E.; Feenstra, K.; Sobin, L.; Robb, J.; Branton, P.; Korzeniewski, G.; Shive, C.; Tabor, D.; Qi, L.; Groch, K.; Nampally, S.; Buia,

S.; Zimmerman, A.; Smith, A.; Burges, R.; Robinson, K.; Valentino, K.; Bradbury, D.; Cosentino, M.; Diaz-Mayoral, N.; Kennedy, M.; Engel, T.; Williams, P.; Erickson, K.; Ardlie, K.; Winckler, W.; Getz, G.; DeLuca, D.; Daniel MacArthur; Kellis, M.; Thomson, A.; Young, T.; Gelfand, E.; Donovan, M.; Meng, Y.; Grant, G.; Mash, D.; Marcus, Y.; Basile, M.; Liu, J.; Zhu, J.; Tu, Z.; Cox, N. J.; Nicolae, D. L.; Gamazon, E. R.; Im, H. K.; Konkashbaev, A.; Pritchard, J.; Stevens, M.; Flutre, T.; Wen, X.; Dermitzakis, E. T.; Lappalainen, T.; Guigo, R.; Monlong, J.; Sammeth, M.; Koller, D.; Battle, A.; Mostafavi, S.; McCarthy, M.; Rivas, M.; Maller, J.; Rusyn, I.; Nobel, A.; Wright, F.; Shabalin, A.; Feolo, M.; Sharopova, N.; Sturcke, A.; Paschal, J.; Anderson, J. M.; Wilder, E. L.; Derr, L. K.; Green, E. D.; Struewing, J. P.; Temple, G.; Volpi, S.; Boyer, J. T.; Thomson, E. J.; Guyer, M. S.; Ng, C.; Abdallah, A.; Colantuoni, D.; Insel, T. R.; Koester, S. E.; A Roger Little; Bender, P. K.; Lehner, T.; Yao, Y.; Compton, C. C.; Vaught, J. B.; Sawyer, S.; Lockhart, N. C.; Demchok, J.; Moore, H. F. The Genotype-Tissue Expression (GTEx) Project. *Nat. Genet.* **2013**, *45* (6), 580–585. https://doi.org/10.1038/ng.2653.

29. Naudin, C.; Hattabi, A.; Michelet, F.; Miri-Nezhad, A.; Benyoucef, A.; Pflumio, F.; Guillonneau, F.; Fichelson, S.; Vigon, I.; Dusanter-Fourt, I.; Lauret, E. PUMILIO/FOXP1 Signaling Drives Expansion of Hematopoietic Stem/Progenitor and Leukemia Cells. *Blood* **2017**, *129* (18), 2493–2506. https://doi.org/10.1182/blood-2016-10-747436.

30. Beutler, E.; Lichtman, M. A.; Coller, B. S.; Kipps, T. J.; Seligsohn, U. *Williams Hematology*, 6th ed.; McGraw-Hill: New York, 2001.

31. Van Dijk, T. B.; Gillemans, N.; Pourfarzad, F.; Van Lom, K.; Von Lindern, M.; Grosveld, F.; Philipsen, S. Fetal Globin Expression Is Regulated by Friend of Prmt1. *Blood* **2010**, *116* (20), 4349–4352. https://doi.org/10.1182/blood-2010-03-274399.

32. Bourantas, K. L.; Georgiou, I.; Seferiadis, K. Fetal Globin Stimulation during a Short Term Trial of Erythropoietin in HbS/Beta-Thalassemia Patients. *Acta Haematol.* **1994**, *92* (2), 79–82. https://doi.org/10.1159/000204179.

33. Perrine, S. P.; Castaneda, S. A.; Boosalis, M. S.; White, G. L.; Jones, B. M.; Bohacek, R. Induction of Fetal Globin in β-Thalassemia: Cellular Obstacles and Molecular Progress Susan. *Ann. N. Y. Acad. Sci.* **2005**, *1054*, 257–265. https://doi.org/10.1196/annals.1345.033.Induction.

34. Bustin, S. A.; Nolan, T. Pitfalls of Quantitative Real-Time Reverse-Transcription Polymerase Chain Reaction. *J. Biomol. Tech.* **2004**, *15*, 155–166.

35. Finno, C. J.; Bordbari, M. H.; Valberg, S. J.; Lee, D.; Herron, J.; Hines, K.; Monsour, T.; Scott, E.; Bannasch, D. L.; Mickelson, J.; Xu, L. Transcriptome Profiling of Equine Vitamin E Deficient Neuroaxonal Dystrophy Identifies Upregulation of Liver X Receptor Target Genes. *Free Radic. Biol. Med.* **2016**, *101*, 261–271. https://doi.org/10.1016/j.freeradbiomed.2016.10.009.Transcriptome.

36. Piercy, R.; Swardson, C.; Hinchcliff, K. Erythroid Hypoplasia and Anemia Following Administration of Recombinant Human Erythropoietin to Two Horses. *J. Am. Vet. Med. Assoc.* **1998**, *212* (2), 244–247.

Concluding Discussion

The investigations into the genetic mechanism of Atypical Equine Thrombasthenia (AET), RNA sequencing (RNA-seq) library preparation methods for long non-coding RNA (lncRNA) detection, and transcriptomic markers of recombinant human erythropoietin (rHuEPO) micro-dosing presented in this thesis have advanced the current knowledge of the genomics of inherited and induced hematological disturbances in Thoroughbred racehorses. A whole-genome variant analysis study and subsequent molecular and functional studies identified two putative variants (*SEL1L* c.1810A>G p.Ile604Val and *VIPAS39:g.22685398_22685470del*), which may play a novel role in platelet function. Further work is necessary to elucidate the definitive causative variant, but identifying the mechanism of AET will deepen our understanding of platelet function during injury. Additionally, determining the causative variant for AET will allow for the identification of affected horses to ensure any injuries are managed closely.

After identifying a lncRNA associated with AET, an investigation into the optimal RNA-seq library preparation method for lncRNA detection was performed that can inform future equine lncRNA study design. A comparison between two tissues from the same two horses was used in the first systematic comparison of RNA-seq library preparations in the horse. Poly-A^+ selection was determined to be better overall for identifying lncRNA as compared to rRNA depletion. However, rRNA-depletion identified small nucleolar RNA (snoRNA) at higher relative levels than with poly-A^+ selection, which may make rRNA-depletion better for snoRNA studies.

In addition to studying an inherited bleeding disorder in Thoroughbred racehorses, we also investigated an induced hematological disturbance by identifying transcriptomic makers of rHuEPO micro-dosing. This was the first study in equine athletes to focus on micro-dosing and the only rHuEPO study to use Thoroughbreds that were of similar age and fitness to American

Thoroughbred racehorses. We identified three transcriptomic markers (*C13H16orf54, PUM2 and CHTOP)* of rHuEPO micro-dosing, but RT-qPCR was not sensitive enough to replicate the differences observed with RNA-seq. While these are therefore not suitable biomarkers to develop a commercial test to detect illicit rHuEPO micro-doping in racehorses, identification of these dysregulated transcripts support altered hematologic properties in these horses following micro-dosing. Further study is needed to determine a more accurate way to commercially quantify these biomarkers or to investigate additional biomarkers of rHuEPO micro-dosing that are feasible to use as a diagnostic test for doping.

A broader understanding of the genomics of inherited and induced hematological disturbances in horses may provide methods to better support the overall health of the Thoroughbred breed. Identification of inherited bleeding disorders can inform breeding and management decisions, and accurate detection of illicit doping can deter the use of the drugs which, in turn, can prevent unintended side effects in the horses. These changes may lead to lower frequencies of epistaxis and sudden death due to cardiac events, both of which can impact the economic viability and overall integrity of the racehorse industry.

Addendum: Equine Juvenile Degenerative Axonopathy in Quarter Horse Foals – Clinical, Histologic and Genetic Characterization

Authors: Anna R. Dahlgren[1], Kevin Woolard [2], Callum G. Donnelly[1], Ana Pacheco[3], Katherine Watson[4], David Ricks[5], Christine Boeckh[5], Emily Berryhill[6], Sarah Humphreys[7], Andrew Willis[7], Berkley Chesen[8], John Ragsdale[9], James E. Tompkins[10], Carrie J. Finno[1]

[1] Department of Population Health and Reproduction, School of Veterinary Medicine, University of California Davis, Davis, CA 95616

[2] Department of Pathology, Microbiology and Immunology, School of Veterinary Medicine, University of California, Davis, CA 95616

[3] Carlson College of Veterinary Medicine, Oregon State University, Corvallis, OR 97331

[4] California Animal Health and Food Safety Lab, School of Veterinary Medicine, University of California Davis, Davis, CA 95616

[5] Brazos Valley Equine Hospital, Stephenville, TX 76401

[6] Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California Davis, Davis, CA 95616

[7] Veterinary Medical Teaching Hospital, School of Veterinary Medicine, University of California Davis, Davis, CA 95616

[8] Equine Comprehensive Wellness, Santa Fe, NM 87111

[9] New Mexico Department of Agriculture Veterinary Diagnostic Services (Ragsdale), Albuquerque, NM 87102

[10] Mountain Veterinary Clinic (Tompkins), Tucumcari, NM 88401

Key Words: inherited, neurologic, spinal cord, Wallerian

Abstract

Background: Genetic mutations have been identified for inherited degenerative axonopathies in other species with none identified in horses to date. In 2020, six related Quarter horse (QH) foals were identified that had developed a novel acute onset of ataxia at 1-4 weeks of age.

Hypothesis/Objectives: We aimed to define the clinicopathologic and histologic findings in these QH foals, to determine the mode of inheritance, and to identify an associated genomic region.

Animals: Six affected foals (*n*=3 females and *n*=3 males) underwent clinical and postmortem evaluations. Whole-genome sequencing (WGS) was performed on *n*=5 affected, *n*=8 unaffected related (*n*=4 parents and *n*=4 half or full siblings) and *n*=11 unaffected unrelated QHs.

Methods: Routine blood work, cerebrospinal fluid cytology, imaging, serum α-tocopherol assessment and necropsy were performed. Pedigrees were evaluated and homozygosity mapping performed on WGS to identify a genomic region of association with the phenotype.

Results: Hyperglycemia and increased serum GGT were common findings in the affected QH foals. At necropsy, Wallerian degeneration was most prominent in the dorsal spinocerebellar tract but also within the cuneate and ventromedial tracts of the spinal cord, with no lesions identified in the brainstem or cerebellum. The degenerative axonopathy was inherited as an autosomal recessive trait. WGS identified a 2.28 Mb region of association on chromosome 11.

Conclusions and Clinical Importance: An autosomal recessive degenerative axonopathy exists in QH foals with the likely putative functional variant in a haplotype region on chromosome 11. Elucidating the genetic cause for this disease will allow for genetic testing to prevent future cases.

Introduction

Degenerative axonopathies have been characterized by progressive pelvic limb ataxia and paraparesis [1–6] in other species, including cattle and sheep. This ultimately leads to permanent recumbency in a responsive animal. Progressive limb ataxia, mainly affecting the pelvic limbs, and paraparesis primarily indicate lesions in the spinal cord between the second thoracic (T2) and third lumbar (L3) segments; however, peripheral nervous system, brainstem, cerebellum, or even cerebrum may also be involved [7]. In cattle, progressive limb ataxia and paraparesis can result from trauma [8], vertebral/extradural abscessation [9], degenerative myelopathies due to organophosphate toxicity [10] or inherited defects [11,12]. In horses, trauma [13] and infection [14] have been reported to result in similar clinical signs; however, there are no reports to date of inherited defects leading to an early-onset progressive limb ataxia and paraparesis. Thus, the aims of this prospective study were to characterize the clinical and pathologic findings of a novel inherited degenerative axonopathy in the American Quarter Horse (QH).

Methods

Animals

Ethics statement. All procedures were approved by the University of California-Davis Institutional Animal Care and Use Committee (protocol #20751) and carried out in accordance with guidelines and regulations. Written owner consent was obtained for all sample collections.

2020 affected foals. Six QH foals diagnosed with equine juvenile degenerative axonopathy (EJDA) were used for this study. In March of 2020, a 3-week old Quarter horse (QH) filly (*Case #1*) presented to Oregon State University Carlson College of Veterinary Medicine with an acute

onset of ataxia (grade 4/5 [15], pelvic >> thoracic), which progressed within five days, to pelvic limb paresis. Mentation and cranial nerve evaluations were normal, and there was no history of trauma as the foal was maintained under video surveillance. Routine blood work revealed leukocytosis, hyperglycemia and increased serum GGT and SDH concentrations (**Table A.S1**). Cerebrospinal fluid (CSF) was collected, found to be cytologically normal and tested negative for equine protozoal myeloencephalitis (EPM) using the Indirect Fluorescent Antibody Test. Computed tomography (CT) with contrast and a CT myelogram were performed with no abnormalities identified. Whole blood selenium concentrations were normal and serum α-tocopherol concentrations were high in the foal (**Table A.S1**) and normal in the dam (2.64 µg/mL). Despite intensive care, the filly became unable to stand within five days of presentation and was subsequently euthanized. A full necropsy evaluation at Oregon State University did not identify any gross lesions. Histologic evaluation was performed at University of California, Davis (KW). Within the spinal cord, dilated myelin sheaths with digestion chambers were identified, most prominent in the dorsal spinocerebellar tract, but also within the cuneate and ventromedial tracts, consistent with a degenerative axonopathy. There were no lesions identified in the brainstem or cerebellum.

One week after this case, two additional cases (*Cases #2* and *3*) were identified in Texas with similar clinical presentations and diagnostic findings (**Table A.S1**). These foals were subsequently euthanized and samples sent to the California Animal Health and Food Safety Laboratory for DNA collection and histologic review. Lesions at necropsy followed a similar distribution to *Case #1*. In April of 2020, an 11-day old QH filly (*Case #4*) was presented to UC Davis with similar clinical signs (**Table A.S1**). Neurologic deficits progressed to recumbency within 3 days, necessitating euthanasia. Necropsy findings were consistent with previous cases. In May of 2020, another two related QH foals (*Cases #5* and *6*) in New Mexico were identified with similar clinical presentations (**Table A.S1**). Samples were collected and shipped to UC

Davis for necropsy evaluation, with similar findings as the previous cases. There were no lesions within the brainstem and both serum and hepatic vitamin E concentrations were high in these foals (foals had all been supplemented at birth), eliminating equine neuroaxonal dystrophy/ degenerative myeloencephalopathy (eNAD/EDM) as a potential cause.

Table A.S1: Summary of clinical presentation and diagnostic findings for six QH foals with SCA. CBC=complete blood count, Chem=serum biochemistry, CSF=cerebrospinal fluid, CT=computed tomography, EPM=equine protozoal myeloencephalitis, N/A=not available, SCA=spinocerebellar ataxia, TP=total protein, WNL=within normal limits; d=day, w=week, month. Values in red are elevated relative to reference ranges for foals

Case	Signal-ment	Age of onset	CBC / Chem Results				Other Dx Results			Vitamin E in serum unless otherwise noted(µg/mL)	Necropsy Findings
			WBC (/µL)	Glucose (mg/dL)	GGT (IU/L)	Bilirub (mg/dL)	CSF	EPM	Imaging		
1	3 w female	2 w	13,890	265	156	2.6	WNL	Neg	CT myelogram - WNL	21.8	SCA
2	2 w female	2 w	9,200	163	74	1.6	Elevated TP (99 mg/dL)	N/A	MRI – WNL	N/A	SCA
3	4 w male	1 mo	15,200	168	274	3.7	N/A	N/A	Radiographs- suspect L5-6 fracture	N/A	SCA – no fracture
4	1.5 w female	7 d	10,860	162	39	1.7	WNL	Neg	N/A	33	SCA
5	2 w female	1 w	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	SCA
6	5 w male	1 mo	N/A	N/A	N/A	N/A	N/A	N/A	N/A	200 hepatic	SCA

Unaffected related control horses. Blood samples were collected from all dams, one sire, one unaffected full sibling, and three unaffected half-siblings for DNA extraction.

Unaffected unrelated control horses. Eleven healthy adult unrelated QHs (5 geldings, 6 mares, age range 3-15 years) sequenced as part of the Pioneer 100 Horse Health Project at the UC Davis Center for Equine Health were used as additional control samples. These horses had been clinically phenotyped by two of the investigators (CGD and CJF) in 2020 and 2021, with no neurologic abnormalities identified.

DNA extraction

Genomic DNA was isolated from whole blood samples according to the WIZARD Blood DNA Extraction Kit protocol (Promega, Madison, WI). If no blood sample was available, genomic DNA was isolated from available tissue (3 horses) using the Gentra Puregene Tissue kit (Qiagen, Germantown MD).

Pedigree Analysis

Six-generation pedigrees were available for all affected QH foals and constructed using Pedigraph [16].

Whole-genome sequencing

Genomic DNA from *n*=5 affected, *n*=8 unaffected related and *n*=11 unaffected unrelated QHs was sequenced on the Illumina HiSeq2500 at approximately 30X coverage. The reads were trimmed with fastp [17], mapped to EquCab3.0 with Burrows-Wheeler Aligner (BWA-MEM) [18,19] and sorted by coordinate with samtools [20]. PCR duplicates were removed with

Sambamba markdup [21], and Freebayes [22] and Delly [23] were used to call variants. The functional effects of the variants were predicted using SnpEff [24], and the variants were filtered by segregation using Fisher's Exact Test with SnpSift [25].

Homozygosity Mapping

SNPSift [24,25] was first used to filter the resulting .vcf file by quality, using a variant Phred threshold of 30 (Q≥30) and then pruned for strong local linkage disequilibrium using PLINK (--indep-pairwise 50 50 0.2) [26], leaving 1,873,121 variants for homozygosity mapping. Homozygosity mapping was performed in plink [26] using a 50 kb sliding window and allowing for *n*=8 heterozygotes in each window (i.e. 8 related unaffected horses). Overlapping segments were identified using –homozyg-group and further filtered by regions of shared homozygous alleles in only the five affected foals.

Positional Candidate Gene Evaluation

The variants in the associated haploblocks identified by homozygosity mapping were evaluated for conservation across 100 vertebrate species by using University of California Santa Cruz (UCSC) Genome Browsers' (genome.ucsc.edu) LiftOver and Table Browser tools. Variants that had a phastCons score of over 0.6 were prioritized and the presence of the variants was evaluated in a cohort of 159 horses that were whole genome sequenced for other projects. Variants that were homozygous in healthy horses were excluded.

Results

Pedigree analysis

The pedigrees of the six affected foals were obtained and a pedigree analysis was conducted using Pedigraph [16]. Four of the affected foals shared a sire and all foals were related on the sires' side within four generations. The foals were also related on the dams' side within four generations. The pedigree analysis indicates an autosomal recessive mode of inheritance for the disease (**Fig. A.1**).

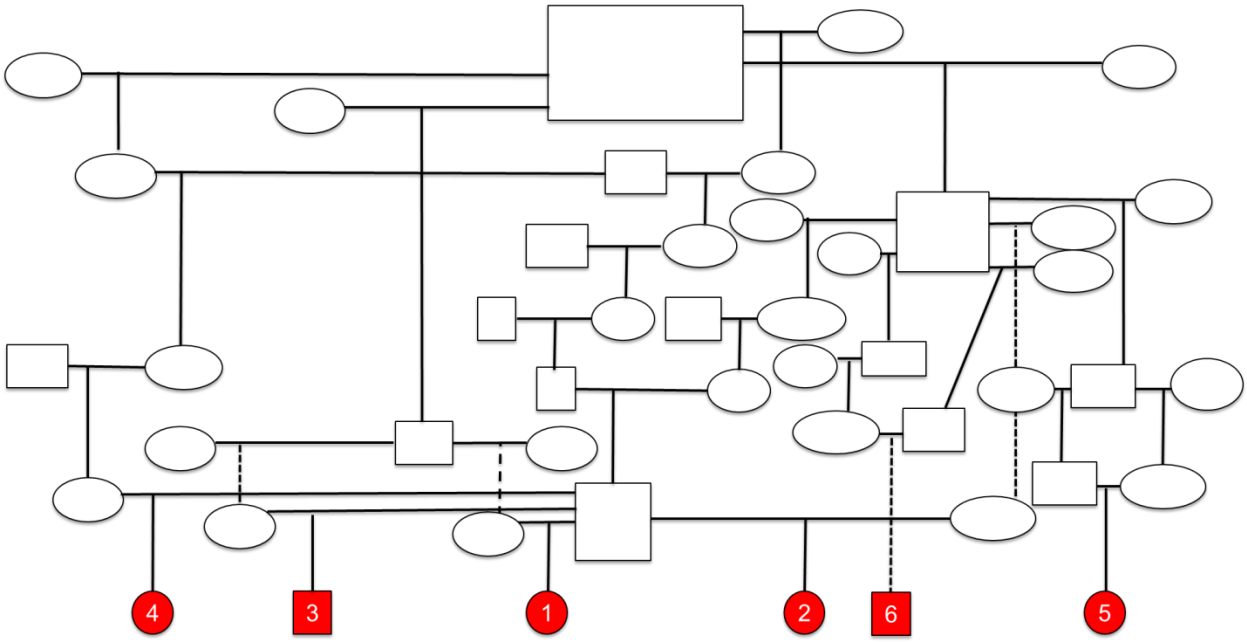


Figure A.1. Pedigrees of affected horses as indicated by red shapes. Number correspond to case number. Squares indicate males and circles are females.

Necropsy results

Necropsies were performed on all affected foals. Examined tissues included the gluteal muscle, semitendinosus muscle, epaxial lumbar, epaxial thoracic, kidney, intestine, spleen, liver, lung, heart, spinal column, brainstem, cerebellum and cerebrum. There were no gross lesions consistently observed across all the cases. Case #2 did have hemorrhage in the epaxial muscles as well as epidural hemorrhage, but this was not seen in any other cases. In all the cases, myelin sheaths with digestion chambers were identified, most prominent in the dorsal

spinocerebellar tract (**Fig. A.2**) but also within the cuneate and ventromedial tracts of the spinal cord, consistent with degenerative axonopathy. Case #2 also had axonal spheroids and gliosis. No lesions were identified in the brainstem or cerebellum.

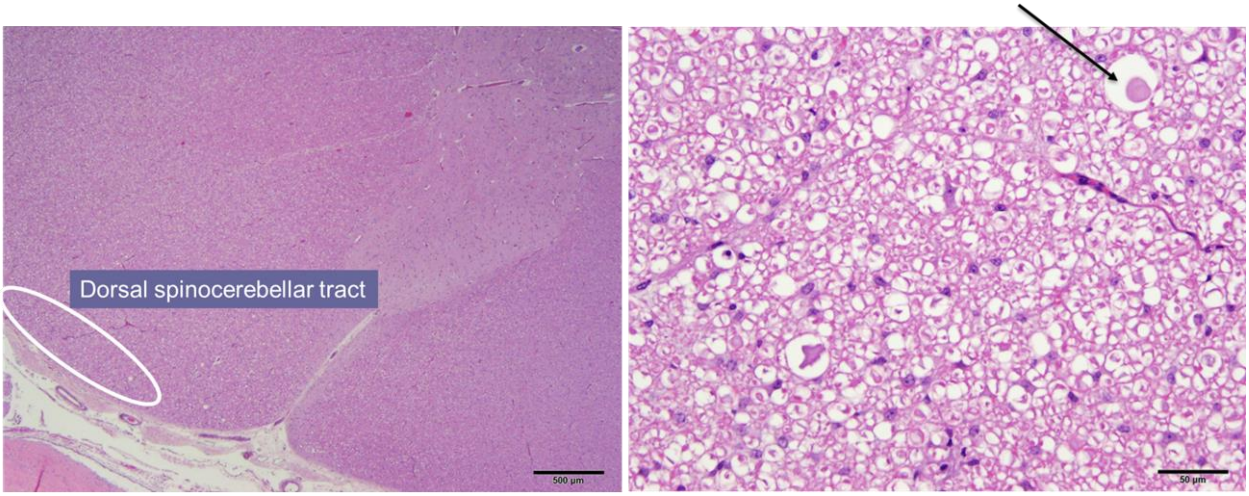


Figure A.2. Digestion chambers (arrow) in dorsal spinocerebellar tract of representative affected foal.

Homozygosity mapping

Whole genome sequencing was performed on 5 cases, 3 of the dams, 1 of the sires, 4 siblings/half-siblings, and 11 unrelated controls QHs. Variant analysis identified an associated region on chromosome 11 (**Fig. A.3**).

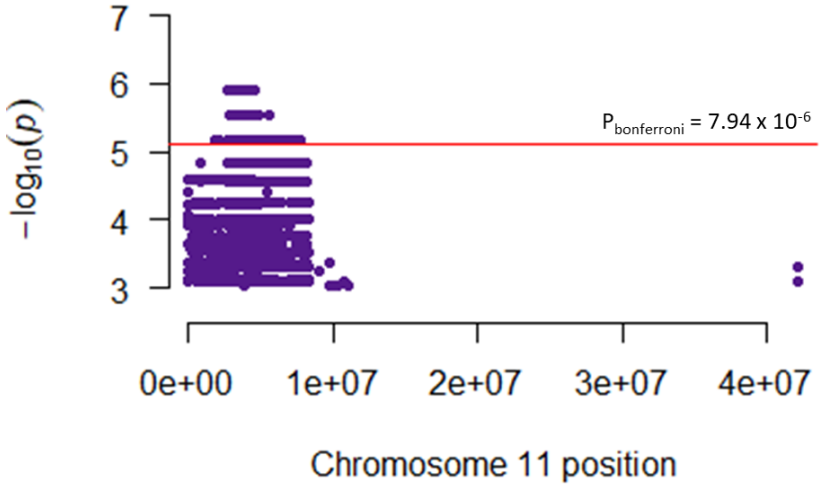


Figure A.3. Manhattan plot of chromosome 11 and corresponding QQ plot. All variants that have p<0.001 are included. Red line indicates Bonferroni correction.

Homozygosity mapping identified two unique runs of homozygosity present in all affected foals and none of the related or unrelated controls horses. Both blocks of homozygosity were on chromosome 11 (chr11:2783556-5064747, chr11:5073050-7955023). Within these blocks, there were 854 variants. Using the LiftOver and Table Brower functions on UCSC Genome Browser (genome.ucsc.edu), phastCons conservation scores were determined for 604 of the variants. The remaining 250 variants could not be successfully lifted over to the analogous position in the human genome. Of these 604 variants, 12 variants had a phastCons score >0.6. One of these variants was subsequently identified to be homozygous in a healthy QH in a data set of 159 whole genome sequenced horses of varying phenotypes. The variant was excluded, leaving 11 prioritized variants for further investigation (**Table A.1**).

Table A.1: Prioritized Variants		
Coordinates	**Ref allele**	**Alt allele**
chr11:4105363	C	T
chr11:4127390	A	G
chr11:4131093	T	C
chr11:4404747	T	C
chr11:4394325	G	A
chr11:2998975	C	A
chr11:2987471	C	G
chr11:3282492	G	A
chr11:4077222	G	C
chr11:4078820	G	C
chr11:3705792	T	C

Discussion

The purpose of this study was to describe a novel degenerative axonopathy recently identified in QH foals, to determine the mode of inheritance and to identify an associated genomic region. The affected QH foals consistently presented with acute onset of ataxia within the first month of life that progressed to pelvic limb paresis and death (if the foal was not euthanized). At necropsy, there were no consistent gross lesions identified across all samples; however, histologic assessment identified myelin sheaths with digestion chambers prominently in the dorsal spinocerebellar tract but also within the cuneate and ventromedial tracts of the spinal cord. No lesions were identified in the brainstem or cerebellum. This histologic diagnosis is consistent with a degenerative axonopathy.

Pedigree analysis of the affected horses supports a recessive mode of inheritance and thus a genomic investigation was pursued. Without strong candidate genes for EJDA, a whole genome-association study was performed and an associated region on chromosome 11 was identified. Eleven putative variants were identified within this region. Genotyping additional

unaffected siblings and affected cases will be necessary to investigate functional effects of the putative variants.

One limitation of this study is the small number of affected horses (*n*=5). However, a previous study of a recessive genetic disease in horses identified the causative variant by whole genome sequencing only two affected horses [27]. While the two horses in that study were Thoroughbreds which are more inbred than QHs [28], increasing the affected population to five horses and including more related unaffected horses (*n*=8) should be sufficient to identify the causative variant in a more genetically diverse breed [28].

A recent retrospective study of the causes of spinal ataxia of horses in California from 2005-2017 found that of the 2,861 horses necropsied during the time period, 316 of them were ataxic [29]. The most common causes of ataxia were cervical vertebral compressive myelopathy, equine neuroaxonal dystrophy/equine degenerative myeloencephalopathy (eNAD/EDM) and trauma [29]. In QHs specifically, eNAD/EDM was the most common diagnosis [29]. Thus, eNAD/EDM was a strong differential diagnosis for this early-onset ataxia in QH foals. However, the defining lesions for eNAD – spheroids in the lateral accessory cuneate, medial cuneate and gracile nuclei of the brainstem and nucleus thoracicus of the spinal cord [30] – were not identified in these foals. Although degeneration of the dorsal spinocerebellar tracts can be identified in horses with EDM [30], the absence of spheroids in the grey matter tracts precludes a diagnosis of eNAD/EDM. Additionally, vitamin E (alpha-tocopherol) concentrations were normal or high in all QH foals that had concentrations measured. Therefore, eNAD/EDM was excluded as a diagnosis in these foals.

To date, no degenerative axonopathies have previously been identified in horses, but similar axonopathies have been described in cattle and sheep [1–6]. Of the early-onset degenerative

axonopathies, only degenerative axonopathy (Demetz syndrome) in Tyrolean Grey cattle and bovine spinal dysmyelination in Brown Swiss have lesions confined to the brainstem and spinal cord [1,31]. In Tyrolean Grey cattle, degenerative axonopathy is clinically characterized by progressive ataxia and paraparesis most prominent in the pelvic limbs starting at one to one and a half months of age [1]. Affected calves eventually become recumbent [1]. This is similar to EJDA in the QH foals, where the pelvic limbs are more severely affected, and it progresses to recumbency. However, EJDA occurs slightly earlier in life (within first month) than in cattle with degenerative axonopathy. Bovine spinal dysmyelination affected Brown Swiss cattle are unable to stand from birth [31]. In contrast, EJDA foals typically appear healthy at first then progress to being unable to stand. For these two comparative inherited degenerative axonopathies of cattle, putative variants have been identified and validated. A SNP in *mitofusin 2* (*MFN2*) causing a splicing defect was found to be associated with degenerative axonopathy in Tyrolean Grey cattle, and a missense mutation in *spastin* (*SPAST*) was determined to be associated with bovine spinal dysmyelination. Neither *MFN2* nor *SPAST* are located in the associated region for EJDA on equine chromosome 11, so these genes are unlikely to be implicated in EJDA.

Other inherited axonpathies (central and peripheral axonopathy of Rouge-des-Pres cattle, progressive ataxia of Charolais and neuroaxonal dystrophy of sheep) have lesions extending into the cerebrum and/or cerebellum [32–34]. Additionally, canine hereditary ataxia in Old English Sheepdogs and Gordon Setters is classified by autophagosome accumulation in cerebellar Purkinje cells [35]. Central and peripheral axonopathy of Rouge-des-Pres cattle presents as a progressive ataxia most prominently affecting the pelvic limbs, and affected calves consistently have a swaying motion of the hindquarters that leads to loss of balance [4]. Signs begin between five and sixteen weeks and end in recumbency within one to three weeks following onset of signs [4]. QHs affected with EJDA show signs a little sooner and progress quicker. A SNP in *Solute Carrier Family 25 Member 46* (*SLC25A46*) was determined to be

significantly associated with the phenotype. Progressive ataxia in Charolais cattle is clinically characterized by progressive ataxia beginning at six months to five years that gradually increases in severity over several months, finally leading to recumbency. EJDA is an earlier onset disease with quicker progression. A SNP in *kinesin family member 1C* (*KIF1C*) leading to a functional knock out was found to be associated with the phenotype. Neuroaxonal dystrophy (NAD) in sheep has an age of onset of around six weeks [34], later than EJDA affected QH foals. The NAD affected lambs were ataxic with a stiff gait that increased in severity over several days to recumbency [34]. Two mutations in p*hospholipase A2 group VI* (*PLA2G6*) were identified as associated with NAD in Swaledale sheep and occur frequently together in affected sheep so that the sheep are compound heterozygotes [34]. Lastly, canine hereditary ataxia in Old English Sheepdogs and Gordon Setters is clinically characterized by hypermetria, truncal sway, and tremors which slowly progresses to more severe gait abnormalities [35–37]. But, unlike EJDA, it does not progress to recumbency [35–37]. Histologically, this disease is classified by autophagosome accumulation in cerebellar Purkinje cells [35]. An associated SNP in *RAB24 member RAS oncogene family* (*RAB24*) was identified as associated with canine hereditary ataxia. With the differences between these four axonopathies and EJDA, it is unsurprising that none of the above four genes are located in the chromosome 11 associated haploblock and thus are unlikely to be involved in the mechanism of EJDA.

In summary, the novel fatal neurodegenerative disease identified in QHs in the first month of life was determined to be a degenerative axonopathy based on the lesions identified in the spinal cord. Additionally, EJDA has an autosomal recessive mode of inheritance, and an associated region on chromosome 11 was identified for further study. Once the causative variant for EJDA is validated, carriers can be identified through genetic testing to inform breeding decisions.

References

1. Drögemüller, C.; Reichart, U.; Seuberlich, T.; Oevermann, A.; Baumgartner, M.; Boghenbor, K. K.; Stoffel, M. H.; Syring, C.; Meylan, M.; Müller, S.; Müller, M.; Gredler, B.; Sölkner, J.; Leeb, T. An Unusual Splice Defect in the Mitofusin 2 Gene (Mfn2) Is Associated with Degenerative Axonopathy in Tyrolean Grey Cattle. *PLoS One* **2011**, *6* (4), 1–9. https://doi.org/10.1371/journal.pone.0018931.
2. Harper, P. A. W.; Duncan, D. W.; Plant, J. W.; Smeal, M. G. Cerebellar Abiotrophy and Segmental Axonopathy: Two Syndromes of Progressive Ataxia of Merino Sheep. *Aust. Vet. J.* **1986**, *63* (1), 18–21. https://doi.org/10.1111/j.1751-0813.1986.tb02865.x.
3. Harper, P. A.; Healy, P. J. Neurological Disease Associated with Degenerative Axonopathy of Neonatal Holstein-Friesian Calves. *Aust. Vet. J.* **1989**, *66* (5), 143–146. https://doi.org/10.1111/j.1751-0813.1989.tb09781.x.
4. Timsit, E.; Albaric, O.; Colle, M. A.; Costiou, P.; Cesbron, N.; Bareille, N.; Assie, S. Clinical and Histopathologic Characterization of a Central and Peripheral Axonopathy in Rouge-Des-Pre´s (Maine Anjou) Calves. *J. Chem. Inf. Model.* **2011**, *25*, 386–392.
5. Harper, P. A.; Plant, J. W.; Walker, K. H.; Timmins, K. G. Progressive Ataxia Associated with Degenerative Thoracic Myelopathy in Merino Sheep. *Aust. Vet. J.* **1991**, *68* (11), 357–358. https://doi.org/10.1111/j.1751-0813.1991.tb00735.x.
6. Hartley, W. J.; Loomis, L. N. Murrurundi Disease: An Encephalopathy of Sheep. *Aust. Vet. J.* **1981**, *57*, 399–400.
7. Lorenz, M.; Kornegay, J. Localization of Lesions in the Nervous System. In *Handbook of Veterinary Neurology*; Saunders: St. Louis, MO, 2004; pp 45–74.
8. Watson, A. G.; Wilson, J. H.; Cooley, A. J.; Donovan, G. A.; Spencer, C. P. Occipito-Atlanto-Axial Malformation with Atlanto-Axial Subluxation in an Ataxic Calf. *J. Am. Vet. Med. Assoc.* **1985**, *187* (7), 740–742.

9. Braun, U.; Schweizer, G.; Gerspach, C.; Feige, K. Clinical Findings in 11 Cattle with Abscesses in the Thoracic Vertebrae. *Vet Rec.* **2003**, *152*, 782–784. https://doi.org/10.1136/vr.152.25.782.

10. Perdrizet, J. A.; Cummings, J. F.; deLahunta, A. Presumptive Organophosphate-Induced Delayed Neurotoxicity in a Paralyzed Bull. *Cornell Vet.* **1985**, *75* (3), 401–410.

11. Baird, J. D.; Sarmiento, U. M.; Basrur, P. K. Bovine Progressive Degenerative Myeloencephalopathy (Weaver Syndrome) in Brown Swiss Cattle in Canada: A Literature Review and Case Report. *Can. Vet. J.* **1988**, *29* (4), 370–377.

12. Edwards, J.; Richards, R.; Carrick, M. Progressive Spinal Myelinopathy in Murray Grey Cattle. *Aust. Vet. J. Vet. J.* **1988**, *65* (4), 108–109. https://doi.org/10.1080/00480169.1987.35437.

13. Feige, K.; Furst, A.; Kaser-Hotz, B.; Ossent, P. Traumatic Injury to the Central Nervous System in Horses: Occurrence , Diagnosis and Outcome. *Equine Vet. Educ.* **2000**, *12* (4), 220–224.

14. Khatibzadeh, S. M.; Gold, C. B.; Keggan, A. E.; Perkins, G. A.; Glaser, A. L.; Dubovi, E. J.; Wagner, B. West Nile Virus–Specific Immunoglobulin Isotype Responses in Vaccinated and Infected Horses. *Am. J. Vet. Res.* **2015**, *76* (1), 92–100. https://doi.org/10.2460/ajvr.76.1.92.

15. Lunn, D. P.; Mayhew, I. G. The Neurological Evaluation of Horses. *Equine Vet. Educ.* **1989**, *1* (2), 94–101. https://doi.org/10.1111/j.2042-3292.1989.tb01355.x.

16. Garbe, J. R.; Da, Y. Pedigraph: A Software Tool for the Graphing and Analysis of Large Complex Pedigree. University of Minnesota 2008.

17. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor. *Bioinformatics* **2018**, *34* (17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

18. Li, H.; Durbin, R. Fast and Accurate Short Read Alignment with Burrows-Wheeler

Transform. *Bioinformatics* **2009**, *25* (14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

19. Li, H. Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. *arXiv:1303.3997v2 [q-bio.GN]* **2013**, 1–3. https://doi.org/arXiv:1303.3997 [q-bio.GN].
20. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25* (16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
21. Tarasov, A.; Vilella, A. J.; Cuppen, E.; Nijman, I. J.; Prins, P. Sambamba: Fast Processing of NGS Alignment Formats. *Bioinformatics* **2015**, *31* (12), 2032–2034. https://doi.org/10.1093/bioinformatics/btv098.
22. Garrison, E.; Marth, G. Haplotype-Based Variant Detection from Short-Read Sequencing. *arXiv Prepr. arXiv1207.3907 [q-bio.GN]* **2012**, 1–9. https://doi.org/arXiv:1207.3907 [q-bio.GN].
23. Rausch, T.; Zichner, T.; Schlattl, A.; Stütz, A. M.; Benes, V.; Korbel, J. O. DELLY: Structural Variant Discovery by Integrated Paired-End and Split-Read Analysis. *Bioinformatics* **2012**, *28* (18), 333–339. https://doi.org/10.1093/bioinformatics/bts378.
24. Cingolani, P.; Platts, A.; Wang, L. L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S. J.; Lu, X.; Ruden, D. M. A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff: SNPs in the Genome of Drosophila Melanogaster Strain w 1118; Iso-2; Iso-3. *Fly (Austin).* **2012**, *6* (2), 80–92. https://doi.org/10.4161/fly.19695.
25. Cingolani, P.; Patel, V. M.; Coon, M.; Nguyen, T.; Land, S. J.; Ruden, D. M.; Lu, X. Using Drosophila Melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Front. Genet.* **2012**, *3* (MAR), 1–9. https://doi.org/10.3389/fgene.2012.00035.
26. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M. A. R.; Bender, D.; Maller,

J.; Sklar, P.; de Bakker, P. I. W.; Daly, M. J.; Sham, P. C. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **2007**, *81* (3), 559–575. https://doi.org/10.1086/519795.

27. Rivas, V. N.; Gary Magdesian, K.; Fagan, S.; Slovis, N. M.; Luethy, D.; Javsicas, L. H.; Caserto, B. G.; Miller, A. D.; Dahlgren, A. R.; Peterson, J.; Hales, E. N.; Peng, S.; Watson, K. D.; Khokha, M. K.; Finno, C. J. A Nonsense Variant in Rap Guanine Nucleotide Exchange Factor 5 (RAPGEF5) Is Associated with Equine Familial Isolated Hypoparathyroidism in Thoroughbred Foals. *PLoS Genet.* **2020**, *16* (9). https://doi.org/10.1371/journal.pgen.1009028.
28. Petersen, J. L.; Mickelson, J. R.; Cothran, E. G.; Andersson, L. S.; Axelsson, J.; Bailey, E.; Bannasch, D.; Binns, M. M.; Borges, A. S.; Brama, P.; Leeb, T.; Lindgren, G.; Lohi, H.; Lopes, M. S.; Mcgivney, B. A. Genetic Diversity in the Modern Horse Illustrated from Genome-Wide SNP Data. *PLoS One* **2013**, *8* (1), 1–15. https://doi.org/10.1371/journal.pone.0054997.
29. Hales, E. N.; Aleman, M.; Marquardt, S. A.; Katzman, S. A.; Woolard, K. D.; Miller, A. D.; Finno, C. J. Postmortem Diagnoses of Spinal Ataxia in 316 Horses in California. *J. Am. Vet. Med. Assoc.* **2021**, *258* (12), 1386–1393. https://doi.org/10.2460/javma.258.12.1386.
30. Aleman, M.; Finno, C. J.; Higgins, R. J.; Puschner, B.; Gericota, B.; Gohil, K.; Lecouteur, R. a; Madigan, J. E. Evaluation of Epidemiological, Clinical, and Pathological Features of Neuroaxonal Dystrophy in Quarter Horses. *J. Am. Vet. Med. Assoc.* **2011**, *239* (6), 823–833. https://doi.org/10.2460/javma.239.6.823.
31. Thomsen, B.; Nissen, P. H.; Agerholm, J. S.; Bendixen, C. Congenital Bovine Spinal Dysmyelination Is Caused by a Missense Mutation in the SPAST Gene. *Neurogenetics* **2010**, *11*, 175–183. https://doi.org/10.1007/s10048-009-0214-0.
32. Duchesne, A.; Vaiman, A.; Castille, J.; Beauvallet, C.; Gaignard, P.; Floriot, S.; Rodriguez, S.; Vilotte, M.; Boulanger, L.; Passet, B.; Albaric, O.; Guillaume, F.; Boukadiri,

A.; Richard, L.; Bertaud, M.; Timsit, E.; Guatteo, R.; Jaffrézic, F.; Calvel, P.; Helary, L.; Mahla, R.; Esquerré, D.; Péchoux, C.; Liuu, S.; Vallat, J.-M.; Boichard, D.; Slama, A.; Vilotte, J.-L. Bovine and Murine Models Highlight Novel Roles for SLC25A46 in Mitochondrial Dynamics and Metabolism, with Implications for Human and Animal Health. *PLoS Genet.* **2017**, *13* (4), 1–30. https://doi.org/10.1371/journal.pgen.1006597.

33. Duchesne, A.; Vaiman, A.; Frah, M.; Floriot, S.; Legoueix-Rodriguez, S.; Desmazières, A.; Fritz, S.; Beauvallet, C.; Albaric, O.; Venot, E.; Bertaud, M.; Saintilan, R.; Guatteo, R.; Esquerré, D.; Branchu, J.; Fleming, A.; Brice, A.; Darios, F.; Vilotte, J.-L.; Stevanin, G.; Boichard, D.; El Hachimi, K. H. Progressive Ataxia of Charolais Cattle Highlights a Role of KIF1C in Sustainable Myelination. *PLoS Genet.* **2018**, *14* (8), e1007550. https://doi.org/10.1371/journal.pgen.1007550.

34. Letko, A.; Strugnell, B.; Häfliger, I. M.; Paris, J. M.; Waine, K.; Cord Drögemüller; Scholes, S. Compound Heterozygous PLA2G6 Loss-of-Function Variants in Swaledale Sheep with Neuroaxonal Dystrophy. *Mol. Genet. Genomics* **2021**, *296*, 235–242. https://doi.org/10.1007/s00438-020-01742-1.

35. Agler, C.; Nielsen, D. M.; Urkasemsin, G.; Singleton, A.; Tonomura, N.; Sigurdsson, S.; Tang, R.; Linder, K.; Arepalli, S.; Hernandez, D.; Lindblad-Toh, K.; Van De Leemput, J.; Motsinger-Reif, A.; O'brien, D. P.; Bell, J.; Harris, T.; Steinberg, S.; Olby, N. J. Canine Hereditary Ataxia in Old English Sheepdogs and Gordon Setters Is Associated with a Defect in the Autophagy Gene Encoding RAB24. *PLoS Genet.* **2014**, *10* (2), 1–13. https://doi.org/10.1371/journal.pgen.1003991.

36. Steinberg, H. S.; Thomas, D. ; Winkle, V.; Bell, J. S.; Alexander De Lahunta, D. ; *Cerebellar Degeneration in Old English Sheepdogs*; 2000; Vol. 217.

37. Steinberg, H. S.; Troncoso, J. C.; Cork, L. C.; Price, D. L. Clinical Features of Inherited Cerebellar Degeneration in Gordon Setters. *J. Am. Vet. Med. Assoc.* **1981**, *179* (9), 886–890.