Resolving the major dyslipidemia phenotypes and genetic risk factors for familial hyperlipidemia in Miniature Schnauzers

A DISSERTATION

SUBMITTED TO THE FACULTY OF

THE UNIVERSITY OF MINNESOTA

BY

Nicole M. Tate

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Eva Furrow, Advisor

September 2022

© Nicole M. Tate 2022

Acknowledgements

Thank you to everyone who has helped me get to this point. I have appreciated every scientific discussion, the endless amount of advice, and moral support.

Dr. Eva Furrow: I thank you most for your relentless belief in me. I cannot express enough gratitude for taking me on as a mentee, sharing your endless knowledge of and enthusiasm for Miniature Schnauzers, writing and research skills, and for supporting countless opportunities for me to develop professionally.

Dr. Molly McCue: Thank you for giving me my start. You have been a source of unending encouragement and have believed in me even when I didn't. I look forward to continuing to learn and grow under your guidance.

Dr. James Mickelson: Thank you for being a constant source of guidance, knowledge, and support over my years in the EGGL and CGL.

Dr. Steven Friedenberg: Thank you for sharing your expertise and ideas throughout this project. Your support and encouragement have meant a lot to me over the years.

Dr. Joerg Steiner: Thank you for your collaboration on this project, your kindness and support, and for sharing your knowledge of lipoprotein profiling and Miniature Schnauzers. Hopefully we can meet in person someday!

Dr. Panagiotis Xenoulis: Thank you for your collaboration, valuable input on lipoprotein profiling, and sharing your enthusiasm for this project. I hope one day we can meet in person!

Members of Canine and Equine Genetics and Genomics Laboratory

Thank you to my fellow CGL and EGGL members, both past and present, for your friendship, wisdom, moral support, and constant celebration of my successes: Dr. Annette McCoy, Dr. Raffaella Teixeira, Dr. Elaine Norton, Dr. Sian Durward-Akhurst, Dr. Rob Schaefer, Dr. Felipe Avila, Dr. Lauren Hughes, Jonah Cullen, Kevin Murray, and Amy Treeful.

An extra special thanks goes to Katie Minor, who has not only shared her "secrets of the trade" but has also honored me with her unending friendship. I cannot express enough gratitude for all the support, guidance, kindness, the shoulder to cry on, allowing me to vent about all things, the many laughs and memories, our silly phrases/inside jokes (i.e., "holy cats on Tuesday" and "Love you, bye!"), and our weekly Zoom "meetings" during the pandemic, which made things feel a whole lot more normal.

I have been immensely grateful for the research, tuition, and stipend support provided in part by the American Kennel Club, the Joanne Schmidt O'Brien/Bee Hanlon Research Fellowship, and the Dr. Carl Osborne Scholarship.

Lastly, an immense amount of appreciation needs to go to the dogs, their owners, trainers, breeders, and referring veterinarians who participated in this research. This project would not be possible without you and your dedication to the health of Miniature Schnauzers.

Dedication

To those I have chosen as family, Gene and Linda Miller. The unending support, constant belief in me, and the unconditional love you have both given me over the years has meant more than you will ever know. I would not have gotten here without you both. Gene, I will forever be grateful for that day you asked me if I wanted a job. I thank you for the years of memories, all our cereal and coffee breaks, sharing your expertise and experience with Standardbreds, and for always making sure I was able to leave by noon to go work in the laboratory. I love you both so much.

To my number one supporter and husband, Matthew James. You jumped headfirst into this journey with me. Your support has been unwavering. I have appreciated the many snack runs, all-nighters, sleeping on my office floor to keep me company, your immense patience, taking care of our many animals when I was busy, the endless emotional and mental support, and making sure I am fed and taking care of myself. Most of all, I thank you for being willing to do whatever it took to make this dream of mine possible. I could not have gotten here without you.

Abstract

Hyperlipidemia is common in the Miniature Schnauzer breed, especially those over the age of 10. Hyperlipidemia is defined as an increased concentration of lipids (i.e., triglycerides and/or cholesterol) in the blood. Hyperlipidemia predisposes Miniature Schnauzers to severe consequences such as pancreatitis, gallbladder mucoceles, and glomerular proteinuria. However, the underlying molecular derangements and cause remains unresolved. It is suspected that hyperlipidemia in Miniature Schnauzers is due to an underlying genetic risk factor(s). Additionally, the varied responses to management strategies in the breed, suggests a potential for multiple subtypes. Thus, the goal of this thesis was to identify the spectrum of dyslipidemia subtypes and ascertain the metabolic and genetic risk factors underlying hyperlipidemia in Miniature Schnauzers.

The possibility of multiple dyslipidemia subtypes within the breed was evaluated using lipoprotein profile data and unsupervised hierarchical cluster analysis. The results support the hypothesis that multiple dyslipidemia subtypes exist in Miniature Schnauzers and that the major distinguishing factor between the subtypes may be differences in low-density lipoproteins. Additional studies are warranted to confirm the range and number of distinct lipoprotein profiles within this breed.

The lipidome and metabolome of Miniature Schnauzers with moderate-to-severe primary hyperlipidemia were compared to those from Miniature Schnauzers with normal serum triglyceride concentrations to elucidate the underlying pathophysiological processes of hyperlipidemia in the breed. Differences in the lipidome and metabolome were identified between the two groups. The differentiating lipid and metabolite species suggest involvement and/or disruption of the pathways and products of glycerolipid,

glycerophospholipid, glycosphingolipid, and fatty acid metabolism. The results of this study provide insights into the underlying pathways. However, it is still unknown whether these pathways are causal of hyperlipidemia or if the disturbances are in response to elevated triglyceride (TG) concentrations.

This thesis also used unsupervised hierarchical cluster analysis to compare the lipidome and metabolome of Miniature Schnauzers dogs with normal serum triglyceride concentrations, mild triglyceride elevations, moderate-to-severe triglyceride elevations, and triglyceride elevations due to endocrinopathies (i.e., secondary hyperlipidemia). The most notable finding being that Miniature Schnauzers with mild HTG cannot be definitively classified as having primary HTG, as their lipid disturbances do not reliably differ from dogs with NTG.

Whole genome sequencing (WGS) of eight Miniature Schnauzers with primary hyperlipidemia was screened for risk variants in six HTG candidate genes: *LPL, APOC2, APOA5, GPIHBP1, LMF1,* and *APOE*. A monogenic cause for primary hyperlipidemia in the breed was not identified in the evaluated candidate genes. Two variants passed the filtering criteria, a deletion in the TATA box of *APOE* and a missense variant in *GPIHBP1*. While the two variants did not have sufficient evidence to support a strong impact, neither can be ruled out as contributors to the disease. These findings, and the growing data on dyslipidemia subtypes in Miniature Schnauzers, suggest that hyperlipidemia in the breed is likely a polygenic or complex trait.

Finally, a key challenge in genetic studies is the prioritization of identified variants. Many *in silico* tools have been developed to use features of amino acids and proteins to determine if a variant is likely pathogenic. However, these methods are

typically trained using human variants and have not been validated for use in non-human species. Thus, this thesis evaluates the performance of eight tools for pathogenicity prediction of missense variants (MutPred2, PANTHER, PhD-SNP, PolyPhen2-HumDiv, PolyPhen2-HumVar, Provean, SIFT, and SNPs&GO) for use in the dog and horse. The findings of this study suggest that these methods can be effectively used in veterinary species.

Table of Contents

List of Tables ix

List of Figures xi

Chapter 1 – Introduction 1

Chapter 2 - Clustering analysis of lipoprotein profiles to identify subtypes of primary hyperlipidemia in Miniature Schnauzers

- Title Page 6
- Abstract/Summary 7
- Introduction 8
- Materials and Methods 9
- Results 11
- Discussion 14

Chapter 3 - Serum metabolomic and lipidomic analysis of Miniature Schnauzers with hypertriglyceridemia

- Title Page 29
- Abstract/Summary 30
- Introduction 31
- Materials and Methods 32
- Results 37
- Discussion 43

Chapter 4 - Sequence analysis of six candidate genes in Miniature Schnauzers with primary hypertriglyceridemia

- Title Page 63
- Abstract/Summary 64
- Introduction 65
- Materials and Methods 66

Results 70
Discussion 72

Chapter 5 - Performance of Variant Pathogenicity Prediction Methods on Canine and Equine Missense Variants
Title Page 86
Abstract/Summary 87
Introduction 88
Materials and Methods 89
Results 94
Discussion 98

Chapter 6 - Conclusions and Future Directions 112

References 123

List of Tables

Chapter 2

Table 1. Eleven density lipoprotein fractions identified in dogs…………….22

Table 2. Clinical characteristics of Miniature Schnauzers with lipoprotein profiles……………………………………………………………………..23

Table 3. Individual contribution of lipoprotein fractions to variance in principal component analysis…….……………………………………………..24

Table 4. Comparison of clinical variables across six clusters……………….25

Chapter 3

Table 1. Clinical characteristics of Miniature Schnauzers with lipidomic and metabolomic data……………………………………………………………….49

Table 2. Comparison of clinical variables across six clusters using lipidomic data…………………………………………………………………………50

Table 3. Comparison of clinical variables across four clusters using metabolomic data……………………………………………………………….51

Chapter 4

Table 1. Six candidate genes for hypertriglyceridemia………………………76

Table 2. Variants of interest from whole genome sequence analysis of candidate genes……………………………………………………………… 77

Table 3. Regression results for the effect of age, sex, and *APOE* genotype on triglyceride concentrations…………………………………………………….. 78

Table 4. Breed counts from a private whole genome sequencing database….79

Table 5. Conservation of the *APOE* deletion in 67 vertebrate species……….82

Chapter 5

Table 1. Summary of eight variant pathogenicity prediction tools…………105

Table 2. Performance of eight variant pathogenicity prediction tools………107

Table 3. Number of variant pathogenicity prediction methods in concordance………………………………………………………………110

Table 4. Comparison of method accuracy between trait and disease variants…...…………………………………………………………………111

List of Figures

Chapter 1

Figure 1. Schematic overview of the project…………………………………….5

Chapter 2

Figure 1. Principal component analysis of lipoprotein profile data…………26

Figure 2. Score and loading plots for the first and third and second and third principal components……………………………………………………………..27

Figure 3. Hierarchical cluster analysis and heatmap of lipoprotein profiling data………………………………………………………………………………..28

Chapter 3

Figure 1. Distribution of lipid classes represented by the differentiating lipid species……………………………………………………………………….52

Figure 2. Heatmaps of the top 10 lipid species differentiating hypertriglyceridemia clinical classifications of Miniature Schnauzers……..53

Figure 3. Pathway analysis of the differentiating lipid species……………...54

Figure 4. Random forest model classifications and variable importance plot using lipidomic data…………………………………………………………….55

Figure 5. Distribution of metabolite classes represented by the differentiating metabolite species…………………………………………………………….56

Figure 6. Heatmaps of the top 9 metabolite species differentiating hypertriglyceridemia clinical classifications of Miniature Schnauzers……. 57

Figure 7. Pathway analysis of the differentiating metabolite species……… 58

Figure 8. Random forest model classifications and variable importance plot using metabolomic data…………………………………………………..59

Figure 9. Principal component analysis of lipidomic and metabolomic data………………………………………………………………………..60

Figure 10. Hierarchical cluster analysis of lipidomic data…………………….61

Figure 11. Hierarchical cluster analysis of metabolomic data………………62

Chapter 4

Figure 1. Range of triglyceride concentrations for each *APOE* genotype. 85

CHAPTER 1

Introduction

Introduction

Hyperlipidemia is defined as increased concentrations of lipids (i.e., triglycerides and/or cholesterol) in the blood. Hyperlipidemia can occur postprandially, secondary to certain diseases (e.g., diabetes mellitus, hypothyroidism, hyperadrenocorticism, pancreatitis, obesity) or drug administration (e.g., glucocorticoids, phenobarbital), or as a primary form without an identifiable underlying cause.[1] Hyperlipidemia is particularly common in Miniature Schnauzers and is characterized by hypertriglyceridemia, with or without hypercholesterolemia, due to increases in triglyceride-rich lipoproteins (very-low density lipoproteins [VLDL] or a combination of chylomicrons and VLDL).[2-4] Primary hypertriglyceridemia is subclinical in Miniature Schnauzers, unless they develop secondary consequences.[5] These secondary consequences include pancreatitis, gallbladder mucoceles, hepatic disease, ocular disease, glomerular proteinuria, or even neurologic disease.[4,6-10] Hypertriglyceridemia is an age-related condition in the breed, with both severity and prevalence increasing with age.[4] By the time Miniature Schnauzers reach 6 years of age or greater, more than 80% have moderate-to-severe serum triglyceride concentrations (>400 mg/dL). Additionally, by the age of 10 more than 75% of the breed is affected.

The cause of hypertriglyceridemia in Miniature Schnauzers remains unclear. In humans, normal serum triglyceride concentrations are maintained in the blood through an equilibrium between the rates of intake, production, and the rate of clearance.[11] A disturbance to this equilibrium leads to hypertriglyceridemia. Given the high prevalence of hypertriglyceridemia in Miniature Schnauzers, it is theorized that a genetic risk factor

is responsible. However, to the author's knowledge, no genetic risk variants have been reported to date.

Hypotheses and Objectives

The overarching objective of this project is to characterize hyperlipidemia subtypes metabolically and genetically within the Miniature Schnauzer breed. This will be accomplished through analysis of lipoprotein, lipid, and metabolite profiles to detect metabolic subtypes and their pathways and analysis of whole genome sequencing data to detect genetic risk variants (Figure 1).

Hypothesis 1 (Chapter 2) - More than one dyslipidemia phenotype exists within the Miniature Schnauzer breed.

Objective 1 - Identify potential subtypes of dyslipidemia in Miniature Schnauzers using hierarchical cluster analysis of lipoprotein profiles.

Hypothesis 2 (Chapter 3) - The lipidome and metabolome differ between Miniature Schnauzers with primary hyperlipidemia compared to those with normal serum triglyceride concentrations, and these differences will reveal pathways contributing to or responding to hyperlipidemia.

Objective 1 - Ascertain serum lipid species and metabolites that distinguish Miniature Schnauzers with primary hypertriglyceridemia from those with normal serum triglyceride concentrations.

Objective 2 - Determine how the lipidome and metabolome profile of Miniature Schnauzers with presumed secondary hypertriglyceridemia (affected by an underlying endocrinopathy) differs from those with primary hypertriglyceridemia.

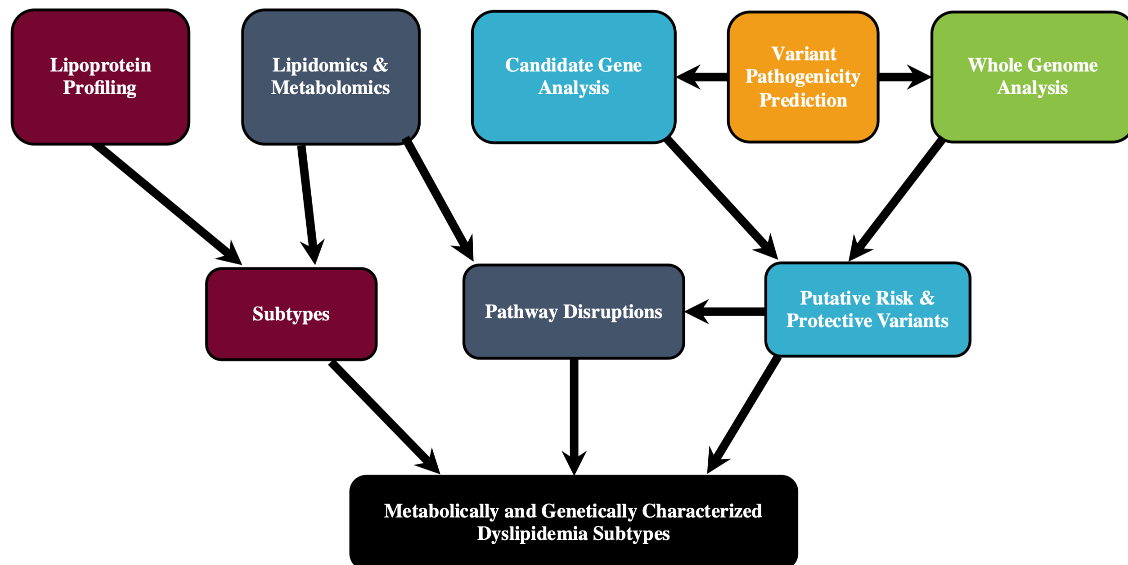
Hypothesis 3 (Chapter 4) - A putative causal variant associated with primary hypertriglyceridemia in Miniature Schnauzers can be identified in one of six major candidate genes.

Objective 1 - Utilize whole genome sequencing to discover putative risk variants for primary hypertriglyceridemia in Miniature Schnauzers in six major candidate genes: *LPL*, *APOC2*, *APOA5*, *GPIHBP1*, *LMF1*, and *APOE*.

Hypothesis 4 (Chapter 5) - One or more variant pathogenicity prediction programs, alone or in combination, will provide a reliable method for variant prioritization in dogs and horses.

Objective 1 - Evaluate the performance of eight variant pathogenicity prediction programs using missense variants in the dog and horse.

Figure 1. Schematic overview of the project.



CHAPTER 2

Clustering analysis of lipoprotein profiles to identify subtypes of primary hyperlipidemia in Miniature Schnauzers

Abstract/Summary

Primary hyperlipidemia is prevalent in Miniature Schnauzers, predisposing them to life-threatening diseases. Varied responses to management strategies suggest the possibility of multiple subtypes. The objective of this study is to identify potential subtypes of hyperlipidemia in Miniature Schnauzers through cluster analysis of lipoprotein profiles. We hypothesize that multiple hyperlipidemia phenotypes exist. Twenty Miniature Schnauzers with normal serum triglyceride concentrations (NTG), 25 with primary hypertriglyceridemia (HTG), and 5 with secondary HTG. Lipoprotein profiles were generated using a continuous lipoprotein density profiling method and clustered with hierarchical cluster analysis. Clinical data (age, sex, body condition score, and dietary fat content) was compared between clusters. Six clusters were identified. Three clusters comprised predominantly HTG dogs. One showed the highest intensities for triglyceride-rich lipoprotein (TRL) and LDL fractions. The second showed moderately increased TRL fraction intensities with intermediate intensities across other fractions. The third showed the lowest LDL fraction intensities and intermediate TRL fraction intensities; HTG cases in this cluster were mild. Two clusters comprised only NTG dogs with lower TRL intensities and low-to-intermediate LDL intensities. The remaining cluster included a mix of NTG and mild HTG dogs with increased LDL but variable TRL fraction intensities. The clinical data was not a significant source of differences between clusters. The results support a spectrum of lipoprotein phenotypes within Miniature Schnauzers that cannot be predicted by triglyceride concentration alone. Lipoprotein profiling may be a useful tool to determine if subtypes have different origins, clinical consequences, and response to treatment.

Introduction

By 10 years of age, >75% of Miniature Schnauzers develop primary hyperlipidemia, predisposing them to pancreatitis, gallbladder mucoceles, glomerular proteinuria, and other disease. [4,6-10] Primary hyperlipidemia in Miniature Schnauzers is characterized by hypertriglyceridemia (HTG) with or without hypercholesterolemia without an identifiable underlying cause.[4] Primary hyperlipidemia is believed to have a genetic origin, though the underlying cause is unresolved.[12-13]

Historically, primary hyperlipidemia in Miniatures Schnauzers was thought to be a single condition characterized by increased triglyceride-rich lipoproteins (TRL) and a decrease in low density lipoproteins (LDL).[3] The increase in TRL is even notable in Miniature Schnauzers without overt HTG, leading to a distinct lipoprotein profile that defines the breed.[3] However, Miniature Schnauzers have varied responses to management strategies, such as feeding a low-fat diet, that may indicate the presence of multiple subtypes.[14-15] The identification and characterization of hyperlipidemia subtypes in Miniature Schnauzers has potential implications for the clinical management, research on hyperlipidemia complications, and genetic approaches to discover risk variants.

Cluster analysis can identify phenotypic subtypes of disease.[16-21] The objective of this study was to identify potential subtypes of hyperlipidemia in Miniature Schnauzers through clustering analysis of lipoprotein profiles. We hypothesized that more than one hyperlipidemia phenotype exists within the Miniature Schnauzer breed.

Materials and Methods

Samples

Samples were selected from 95 Miniature Schnauzer dogs with serum biobanked (-80˚C) at the University of Minnesota Canine Genetics Laboratory from past and ongoing research projects.[9,22,23] Samples were selected if a serum triglyceride concentration (measured by either a Roche/Hitachi Modular Analytics D2400 Module, Roche Diagnostics, Indianapolis, Indiana or Beckman Coulter AU480 Chemistry Analyzer, Beckman Coulter, Brea, California) was available (collected after asking owners to withhold food for 12-18 hours) and if the dog met the below criteria for primary HTG, secondary HTG, or normal serum triglyceride concentration (NTG). Samples were excluded if the dog was receiving glucocorticoids, fibrates, or statins at the time of serum collection.

Clinical Categorization of Hypertriglyceridemia

Data extracted from medical records included sex, age, body condition score (BCS, 1-9 scale), fat content for the primary diet fed (g/100 kcal), medications, diagnoses, and results of endocrine testing. Primary HTG was defined as a fasting serum triglyceride concentration greater than 108 mg/dL (1.2 mmol/L) and no diagnosis or clinical suspicion of an underlying condition that can cause HTG (e.g., diabetes mellitus, hypothyroidism, hyperadrenocorticism, nephrotic syndrome) at the time of sampling and up to 6 months thereafter. Dogs with proteinuria and HTG were included in the primary HTG category if they had no other evidence of renal dysfunction (non-azotemic and normoalbuminemic).[23] Secondary HTG was defined as a fasting serum triglyceride

concentration greater than 108 mg/dL (1.2 mmol/L) in a dog with a diagnosis of an endocrinopathy known to cause hyperlipidemia (e.g., diabetes mellitus, hypothyroidism, hyperadrenocorticism). The severity of HTG was characterized as mild for dogs with triglyceride concentrations of 109-400 mg/dL (1.2-4.5 mmol/L) and moderate-to-severe for concentrations >400 mg/dL (>4.5 mmol/L).[4] Dogs with a fasting serum triglyceride concentration less than or equal to 108 mg/dL (1.2 mmol/L) at 8 years of age or older were categorized as having NTG.

Lipoprotein Profile Analysis

Lipoprotein profiles were generated as previously described using a continuous lipoprotein density profiling method that uses bismuth sodium ethylenediaminetetraacetic acid (NaBiEDTA) as a self-generating density gradient solution.[24] Lipoprotein fractions were imaged as previously described using a custom fluorescence imaging system consisting of a digital camera and a metal halide continuous light source.[24] This method identifies 11 distinct density lipoprotein fractions in dogs based solely on density characteristics (Table 1).[3]

Data Analysis

All statistical analyses were performed using R statistical software (R, version 4.1.2, www.r-project.org).[25] Data normality was evaluated using the Shapiro-Wilks test and quantile-quantile plot graphs (QQ-plot). Clinical data, including age, sex, dietary fat, and BCS were compared between dogs with primary HTG, secondary HTG, or NTG using one-way analysis of variance (ANOVA) for parametric variables (dietary fat), the

Kruskal-Wallis test for nonparametric variables (age, BCS), and the Fisher's exact test for count data (sex).

The 11 fractions of the lipoprotein profile data (described in Table 1) were log transformed and Pareto scaled prior to analysis. Principal component analysis (PCA) was performed to visually assess lipoprotein profile data clusters. The clustering tendency of the data was also assessed using the Hopkins statistic. Variables were clustered using unsupervised, agglomerative, hierarchical clustering. The Ward linkage method was applied on Spearman correlation distances between variables. Clinical data, including age, sex, dietary fat, and BCS, were compared between each cluster. Clusters were compared using a one-way ANOVA for parametric variables (dietary fat), the Kruskal-Wallis test for nonparametric variables (age, BCS), and the Fisher's exact test for count data (sex, hypertriglyceridemia classification).

For all analyses, a $P < 0.05$ was considered statistically significant. The median and range of triglyceride concentrations were also determined for each cluster; however, no statistical comparison was performed since triglycerides contribute to the lipoprotein profiles. Similarly, the proportion of dogs classified as NTG, primary HTG, and secondary HTG were determined for each cluster, but no statistical comparison was performed.

Results

Samples

A total of 50 samples from Miniature Schnauzers were included in the study. Of these, 20 were categorized as NTG, 25 as primary HTG (19 mild and 6 moderate-to-

severe), and 5 as secondary HTG (2 mild and 3 moderate-to-severe). Signalment, BCS, triglyceride concentrations, and dietary fat content are summarized in Table 2; none of these variables had statistically significant differences between dogs with primary HTG, secondary HTG, or NTG. Serum cholesterol concentrations were not measured as part of this study but were available for 8/20 NTG dogs (all within the laboratory reference interval), 14/25 primary HTG dogs (1 above reference interval), and 5/5 secondary HTG dogs (3/5 above reference interval). As per the inclusion criteria, none of the dogs with primary HTG had clinical suspicion for an endocrinopathy. However, the extent of diagnostic screening varied between dogs. All 25 primary HTG dogs had fasting blood or serum glucose concentrations measured (median 100 mg/dL, range 67 - 129 mg/dL [1.1 mmol/L, range 0.76-1.46 mmol/L]). Thirteen primary HTG dogs were screened for hyperadrenocorticism (11 urine cortisol:creatinine ratios and 2 low dose dexamethasone suppression tests), and 12 were screened for hypothyroidism (total thyroxine concentration), with results within laboratory reference intervals. The secondary HTG group included 2 dogs with untreated hypothyroidism and 3 dogs with diabetes mellitus (1 diagnosed the day the serum sample was obtained, 1 diagnosed 9 days prior and unregulated, and 1 diagnosed 2 years prior and described as well-regulated in the medical records). Two of the dogs in the secondary HTG group had no TG concentration available prior to the diagnosis of the endocrinopathy. The other two had mild elevations at 172 and 261 mg/dL (1.9 and 3.0 mmol/L) noted prior to the diagnosis of diabetes mellitus and hypothyroidism, respectively; the dog with hypothyroidism was not screened for thyroid function at the time the HTG was first noted.

Data Analysis

Cluster tendency was evaluated visually with PCA and the Hopkins statistic. The PCA score and loading plot for the first and second principal component (PC) are shown in Figure 1. Clustering of NTG, primary HTG, and secondary HTG dogs can be seen. The first PC was influenced by the TRL and LDL fractions. The second PC was influenced by the TRL, LDL fractions 4 and 5, and some of the high-density lipoprotein (HDL) fractions ($HDL_{2b,\ 3b,\ and\ 3c}$). The third PC was strongly influenced by the HDL fractions. Score and loading plots for the first and third and second and third PCs are included in Figure 2. A list of the individual contributions of each lipoprotein fraction to PC 1, PC 2, and PC 3 is provided in Table 3.

Hierarchical cluster analysis identified six clusters. These clusters and a heatmap corresponding to the distribution of 11 lipoprotein fractions in the clusters are shown in Figure 3. Clinical data (i.e., age, sex, dietary fat, and BCS) were compared between clusters, and none were found to differ significantly by cluster (Table 4).

The most dissimilar nodes (separated by the first branch of the dendrogram and referred to as A and B), predominantly differed in the LDL 4 and LDL 5 fractions, with dogs in the B clusters having a higher intensity across these two fractions compared to those in A clusters. The second major separation occurred between clusters B1 and B2. The lipoprotein profiles for dogs in cluster B1 were characterized by lower intensities over the HDL_{2a} fraction. The dogs in cluster B1 also had low to intermediate intensities across the TRL fraction and increases across LDL fractions, mainly fractions 4 and 5. Cluster B1 comprised 5 NTG and 4 primary HTG dogs (all mild). Cluster B2 was characterized by the highest intensities across the LDL 1-3 and TRL fractions. Cluster B2

included 1 NTG and 11 HTG dogs. Of those with HTG, 7 had primary and 4 secondary (2 with untreated hypothyroidism, 1 with recently diagnosed, unregulated diabetes mellitus, and 1 with well-regulated diabetes mellitus). This cluster comprised 37% (11/30) of the dogs with any degree of HTG and 78% (7/9) of the dogs with moderate-to-severe HTG. This cluster also included 80% (4/5) of the secondary HTG dogs. Due to the relatively high number of dogs with secondary HTG in this cluster, endocrine testing results were reviewed for the 7 dogs with primary HTG to determine the comprehensiveness of their evaluations: 4 were screened for both hypothyroidism and hyperadrenocorticism, 2 were only screened for hyperadrenocorticism, and 1 was only screened for hypothyroidism.

Within the A node, clusters were less distinct. Clusters A2 and A3 had the lowest intensities for the TRL fraction but differed in intensity over the LDL fractions. Dogs in cluster A2 had lower intensities over these fractions. These two clusters contained only NTG dogs, altogether including 55% (11/20) of the NTG dogs in the study. Overall, dogs in cluster A1 had low intensities in the TRL fraction and the lowest intensities across the LDL fractions. This cluster included 2 NTG and 8 primary HTG (all mild) dogs. Most dogs in cluster A4 had mildly higher TRL intensities; the intensities of other fractions were relatively intermediate. Six of 7 dogs in cluster A4 had HTG, including 4 mild HTG (all primary) and 2 moderate-to-severe (1 primary HTG and 1 secondary HTG, diagnosed with diabetes mellitus on the day the serum was collected]).

Discussion

In this study, we used hierarchical cluster analysis of lipoprotein profiles in Miniature Schnauzers with NTG, primary HTG, or secondary HTG and identified 6

clusters that might represent different dyslipidemia phenotypes in this breed. Three clusters were composed almost entirely of HTG dogs (11/12, 6/7, and 8/10 dogs per cluster). Each of these three clusters had different lipoprotein profile characteristics. One was characterized by the highest intensities in the TRL and all LDL fractions, one by a moderately increased intensity in the TRL fraction with intermediate or variable intensity across other fractions, and the third by the lowest LDL fraction intensities and intermediate TRL fraction intensities (all HTG cases in this third cluster were mild). Two clusters comprised only NTG dogs; both were characterized by lower intensities of the TRL fraction compared to the other clusters and low to intermediate LDL fraction intensities. The remaining cluster included a 50:50 mix of dogs with NTG and mild HTG. These clusters appeared to be driven more by differences in intensities across LDL and HDL fractions rather than TRL. The clinical data evaluated (i.e., age, sex, BCS, and dietary fat) were not identified as a source of differences between clusters. The clustering results support a spectrum of lipoprotein phenotypes within the breed that cannot be predicted by triglyceride concentration alone.

The most dissimilar clusters (A and B clusters) predominantly differed in the LDL 4 and LDL 5 fractions, with dogs in the B clusters having a higher intensity across these two fractions. Increases of the fractions corresponding to LDLs, mainly LDL 2 to LDL 4, occur in dogs with pancreatitis.[26] However, dogs with pancreatitis differ from those in cluster B of this study in other fractions. Specifically, dogs with pancreatitis have decreases in the fractions corresponding to TRL, HDL_{2a}, and HDL_{3c}. In contrast, most dogs in cluster B had increased TRL fraction intensities, and none had decreased intensities across the TRL, HDL_{2a}, and HDL_{3c} fractions.

The largest cluster, B2, comprised 37% (11/30) of the dogs with any degree of HTG and 78% (7/9) of those with moderate-to-severe HTG. This cluster also included 80% (4/5) of the secondary HTG dogs. The inclusion of both primary and secondary dogs in the same cluster could indicate that the number of dogs with endocrinopathies was too small for the clustering analysis to capture unique phenotypes of these disorders and segregate them into their own cluster. In support of this, 4 of the 5 dogs with secondary HTG (2 with hypothyroidism and 2 with diabetes mellitus) clustered together, meaning that they were more like each other than most other dogs within the cluster. However, the other dog with secondary HTG (diabetes mellitus) had a different lipoprotein profile pattern and was not even within the B node. It is possible that the endocrinopathy was not the source of HTG in that dog and that genetic risk factors were instead the major underlying cause (i.e., the dog had primary HTG, despite the concurrent endocrinopathy). Even in the dogs with secondary HTG that clustered together, it is likely that the HTG is not solely from the endocrinopathy but rather the sum of multiple risk factors that affect lipid metabolism. In support of this theory, two of the dogs with secondary HTG had mild HTG documented prior to their diagnosis of an endocrinopathy. Triglyceride concentrations are inconsistently increased in dogs with diabetes mellitus, hypothyroidism or hyperadrenocorticism, suggesting that the development of HTG is not an assured outcome of those disorders.[27-29] In humans, HTG is viewed as a continuum, with various degrees of genetic and environmental factors contributing to disease.[11] Another possible explanation for the clustering of primary and secondary HTG cases in the B2 cluster is that some of the dogs categorized as having primary HTG had an undiagnosed subclinical endocrinopathy; two primary HTG dogs in this cluster were not

screened for hypothyroidism and one was not screened for hyperadrenocorticism, although there was no clinical suspicion of an endocrinopathy in any dog classified as having primary HTG. The lack of comprehensive endocrine screening for all HTG dogs is a limitation of the current study.

Two other clusters, A1 and A4, were also primarily HTG dogs (8/10 and 6/7 dogs, respectively). The dogs in cluster A1 had small peaks in the TRL fraction and low intensities across LDL fractions; all 8 dogs with HTG in this cluster were mild cases. The A4 cluster included 5 primary HTG (4 mild and 1 moderate-to-severe) and 1 secondary HTG (diabetes mellitus, moderate-to-severe). The A4 cluster had low to intermediate intensities in LDL fractions. The pattern observed in the dogs in the A1 and A4 clusters are most similar to what has previously been described in Miniature Schnauzers with NTG and HTG, respectively.[3] Using a similar method, a 2013 study determined that Miniature Schnauzers with HTG typically have increased TRL and decreased LDL fractions relative to dogs of other breeds with NTG. The authors of this study also determined that lipoprotein profiles in Miniature Schnauzers with NTG, have a similar, albeit less pronounced, changes to their lipoprotein profiles. The conclusion of these findings was that serum TG concentrations alone are not enough to detect differences in lipoprotein metabolism in dogs and that it is possible that the majority of Miniature Schnauzers differ in their basic lipoprotein metabolism from dogs of other breeds.[3] The results of this study also demonstrate that TG concentrations are an incomplete method to assess lipoprotein metabolism in a patient with HTG. Furthermore, the results suggest that the presence of more than one lipoprotein profile pattern in Miniature Schnauzers with primary HTG. Dogs in the B2 versus A1 and A4 clusters might have distinct

mechanisms contributing to HTG development that determine whether there is an increase only in VLDLs or chylomicrons or whether LDL increases also occur.

The mechanisms that contribute to different lipoprotein profile patterns could be genetic, environmental, or both. For example, lipoprotein lipase deficiency, which could be due to a loss of function mutation in the lipoprotein lipase gene, is associated with decreased levels of LDL and HDL.[30] In contrast, hepatic lipase deficiency, which could result from a loss of function mutation of the hepatic lipase C (*LIPC*) gene, is characterized by increased levels of TG, LDL, and HDL.[30] In terms of environmental contributors to dyslipidemia, none of the clinical variables tested (i.e., age, sex, BCS, or dietary fat) differed by cluster. However, other contributing factors to dyslipidemia that were not evaluated for the purpose of this study include insulin resistance, central obesity (versus overall BCS), and exercise.[11] Another important environmental factor could be different types of dietary fat. Different amounts of fat types could theoretically affect lipoprotein profiles, especially in genetically predisposed individuals.[31,32] Also, as mentioned above, four of the dogs in the B2 cluster had endocrinopathies. Alterations in lipoprotein profiles occur with hypothyroidism, diabetes mellitus, and hyperadrenocorticism.[27,28,33] Lipoprotein profiles of dogs with hypothyroidism and diabetes mellitus are generally characterized by increases across all fractions (TRL, LDL, and HDL), while hyperadrenocorticism primarily increases LDL fractions. It is important to note that the methods for the above referenced studies differed from those used in this study, which can alter the subfractions. Thus, a direct comparison to the lipoprotein profiles identified in this study is not possible. Also, in dogs, the composition of lipoprotein density subfractions and their functional characteristics are currently

unknown and assignment to traditional functional classes, such as LDL and HDL, can only be done nominally.

Two of the clusters, A2 and A3, comprised only NTG dogs, containing about half (11/20) of all NTG dogs included in the study. Both clusters overall had lower intensities of the TRL fraction but differed in intensity of the LDL fractions, with most dogs in cluster A2 having lower intensities of these fractions. Low intensities across LDL fractions are a feature of the Miniature Schnauzer breed pattern, described by Xenoulis et al. in their 2013 study.[3] We did not include non-Schnauzer breeds in this study, which makes it difficult to know whether the A2 and A3 clusters are normal variations of lipoprotein profiles in healthy dogs or whether cluster A2 represents a breed-specific dyslipidemia.

The remaining cluster, B1, comprised an equal proportion of dogs with NTG and mild HTG. The dogs in this cluster had variable intensities across the TRL fraction and increases across LDL fractions 4 and 5 and, less consistently, 1 to 3. One possible explanation is that the dogs in cluster B1 have a mild/early form of the dyslipidemia phenotype present in cluster B2. It is also possible that the mild HTG in dogs in this cluster is not due to genetic risk factors but rather an effect of another patient or dietary factor that was not measured in this study.

In this population of Miniature Schnauzers, dietary fat did not differ by cluster. Feeding a low-fat diet has been shown to resolve HTG in approximately half of Miniature Schnauzers and can alter lipoprotein profiles.[14,15] There were 10 dogs included in this study on diets with fat contents <3.0 g/100 kcal, including 1 NTG, 6 primary HTG, and 3 secondary HTG. It is unknown whether these dogs would have clustered in different

groups if fed a higher fat diet, but they appeared to cluster with other dogs with similar TG concentrations. We also were not able to identify an effect of age, sex, or BCS as a source of differences between clusters. Mild elevations in triglyceride concentrations occur in dogs in association with aging[34,35], and aging is associated with mild increases in the TRL and decreases in the LDL and HDL fractions.[35] Multiple mechanisms have been observed in humans, one of which is an age-related decrease in lipoprotein lipase activity.[36] Lipoprotein lipase activity is also lower in older Miniature Schnauzers with HTG compared to young Miniature Schnauzers with NTG, but data is not available comparing young to old dogs with NTG.[9] An overweight or obese condition is also reported to affect lipoprotein profiles in dogs.[38] When compared to dogs with ideal body condition, obese and overweight dogs have higher TRL and HDL.[38] In humans, a high body mass index is associated with increased levels of LDL and decreased levels of HDL, while gender had no influence on LDL or HDL.[37] The absence of differences in dietary fat, age, and BCS between clusters might be related to the size of the study and relatively low numbers of dogs representing extreme ends of these variables (e.g., few dogs were on low-fat diets, all dogs were between 8-12 years of age, and most dogs in this study had BCS between 5-6).

Our study is limited by a small sample size, especially for dogs with moderate-to-severe HTG. Analysis of additional dogs could reveal further separations and better capture the true range of HTG subtypes. Furthermore, we are limited in our ability to interpret what is driving the clusters. Correlation of genetic variants with specific clusters might reveal genetic drivers of dyslipidemia subtypes, but genomic data was not within the scope of this study. Inclusion of more dogs with secondary HTG of both the

Miniature Schnauzer breed and other breeds not reported to have primary HTG could help separate disturbances that are directly associated with an endocrinopathy from those due to underlying genetic risk factors. Finally, it is also important to note, that while all the serum samples in this study were from dogs after withholding food for 12-18 hours, these lipoprotein profiles only represent a “snapshot” from a single moment in a dog’s life. Fasting triglyceride concentrations can vary considerably in Miniature Schnauzers with HTG and range from mild to severe on different sample dates, without discernible changes in environment between dates.[14]

In conclusion, we identified six clusters of lipoprotein profiles in Miniature Schnauzer dogs, three of which primarily included dogs with HTG. The data supports the hypothesis that more than one HTG phenotype exists in Miniature Schnauzers, and differences in LDL fractions might be the major distinguishing factor. Further investigation is warranted to confirm the range and number of distinct lipoprotein profiles within this breed. Lipoprotein profiling may be a useful tool for future research to determine if subtypes of HTG in Miniature Schnauzers have different origins or clinical consequences.

Table 1. Eleven density lipoprotein fractions identified in dogs using a continuous lipoprotein density profiling method that uses bismuth sodium ethylenediaminetetraacetic acid (NaBiEDTA) as a self-generating density gradient solution.

Fraction	Density (g/mL)	Potential Classification*
R1	<1.019	TRL
R2	1.019-1.023	LDL_1
R3	1.023-1.029	LDL_2
R4	1.029-1.039	LDL_3
R5	1.039-1.050	LDL_4
R6	1.050-1.063	LDL_5
R7	1.063-1.091	HDL_{2b}
R8	1.091-1.110	HDL_{2a}
R9	1.110-1.133	HDL_{3a}
R10	1.133-1.156	HDL_{3b}
R11	1.156-1.179	HDL_{3c}

TRL- Triglyceride-rich lipoproteins (including chylomicrons and very-low density lipoproteins; LDL-low-density lipoproteins; HDL-high-density lipoproteins

*The functional characteristics and composition of most lipoprotein density subfractions in dogs are currently unknown. Thus, all density subfractions can only be nominally assigned to traditional functional classes.

Table 2. Clinical characteristics of 50 Miniature Schnauzers with normal serum triglyceride concentration (NTG), primary hypertriglyceridemia (HTG), or secondary HTG. *P* values are reported for statistical comparisons between all three classifications.

Group Characteristics	NTG (<108 mg/dL)	Primary HTG (>108 mg/dL)	Secondary HTG	*P* value
Total number, n	20	25	5	-
Age yrs, median (range)	9 (8-13)	10 (6-14)	12 (8-14)	0.39
Sex[a], male, female	15, 5	15, 10	3, 1	0.35
BCS[b], median (range)	6 (3-8)	6 (3-7)	5(5)	0.4
Serum TG mg/dL, median (range)	58 (35-102)	266 (110-2821)	772 (399-1848)	-
Dietary fat[c], g/100 kcal, median (range), # of dogs fed a diet with <3.0 g/100 kcal	4.0 (2.8-4.8), 1	3.8 (2.1-5.3), 6	2.8 (2.8-3.4), 3	0.22

BCS, body condition score; HTG, hypertriglyceridemia; NTG, normal serum triglyceride concentration

[a]All dogs were spayed or neutered except for one intact male with secondary HTG.

[b]BCS unknown for 1 primary and 2 secondary HTG dogs.

[c]Dietary fat content unknown for 2 NTG and 4 primary HTG dogs.

Table 3. Individual contribution of lipoprotein fractions to variance in principal component analysis in Miniature Schnauzers with primary hypertriglyceridemia, secondary hypertriglyceridemia, or normal serum triglyceride concentrations.

	Individual contribution to variance (%)		
	PC 1	PC 2	PC 3
TRL	14	13	2
LDL_1	20	5	0
LDL_2	20	1	0
LDL_3	18	1	0
LDL_4	15	11	0
LDL_5	10	14	0
HDL_{2b}	1	14	26
HDL_{2a}	2	1	11
HDL_{3a}	1	1	32
HDL_{3b}	0	23	16
HDL_{3c}	0	16	13

Data are given for the first 3 principal components which explain >85% of the variance.

TRL-triglyceride-rich lipoproteins; LDL - low-density lipoproteins; HDL - high-density lipoproteins

Table 4. Comparison of clinical variables across six clusters of lipoprotein profiles from 50 Miniature Schnauzers, including 20 with a normal serum triglyceride concentration (NTG), 25 with primary hypertriglyceridemia (1 HTG), and 5 with secondary hypertriglyceridemia (2 HTG). *P* values are for comparisons to determine if the clinical variable differs by cluster. Mild HTG is defined as a fasting serum triglyceride concentration of 109 – 400 mg/dL, and moderate-to-severe (mod/severe) HTG is defined as a concentration >400 mg/dL.

Cluster	A1	A2	A3	A4	B1	B2	
Variables	n=10	n=8	n=3	n=7	n=10	n=12	*P* value
Age (median, range)	10 (8-14)	10 (9-13)	9 (9-10)	10 (8-14)	9 (8-11)	11 (6-14)	0.5
Sex[a] (m,f)	4,6	5,3	3,0	6,1	7,3	9,3	0.38
BCS[b] (median, range)	6 (4-7)	5 (3-6)	6 (6-7)	5 (3-7)	6 (5-8)	6 (3-7)	0.14
Dietary fat[c] (median, range), # of dogs fed a diet with <3.0 g/100 kcal	3.4 (2.1-4.3), 3	4.0 (3.6-4.8), 1	3.9 (3.5-4.0), 0	3.3 (2.8-5.3), 3	4.0 (2.5-4.3), 1	4.0 (2.8-4.8), 2	0.36
TG (mg/dL) (median, range)	179 (35-392)	42 (35-91)	61 (54-65)	334 (85-720)	136 (42-338)	643 (41-2821)	-
# of NTG	2	8	3	1	5	1	-
# of 1 HTG	8	0	0	5	5	7	-
mild	8	-	-	4	5	2	
mod/severe	0	-	-	1	0	5	
# of 2 HTG	0	0	0	1	0	4	-
mild	-	-	-	0	-	2	
mod/severe	-	-	-	1	-	2	

[a]All dogs were spayed or neutered except for one intact male with secondary HTG.

[b]BCS unknown for 1 primary and 2 secondary HTG dogs.

[c]Dietary fat content unknown for 2 NTG and 4 primary HTG dogs.

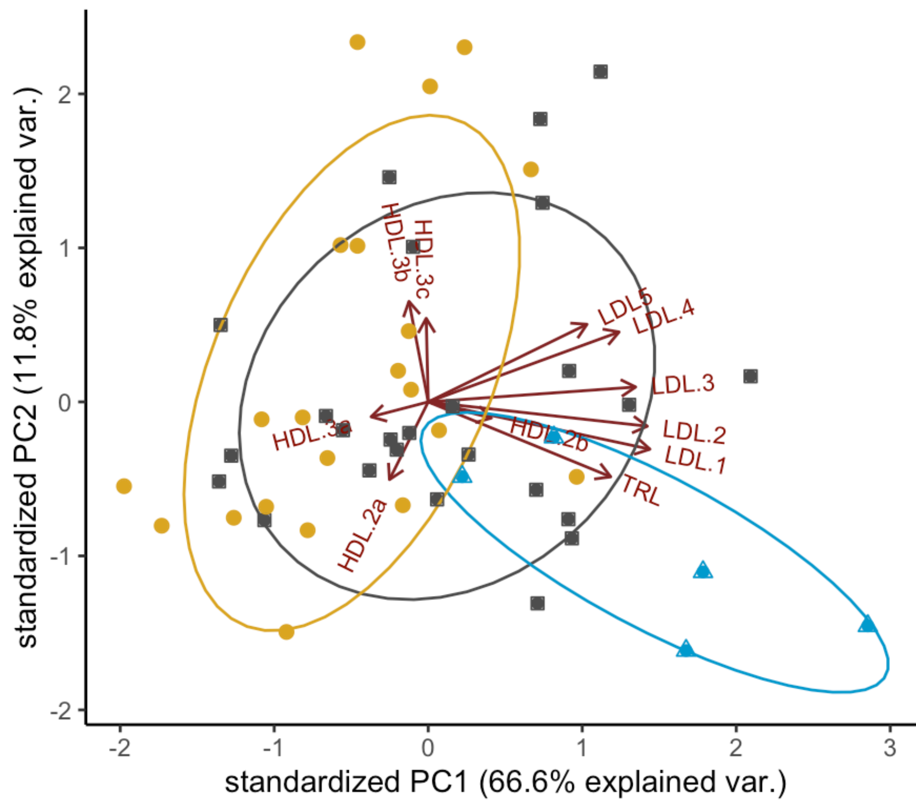
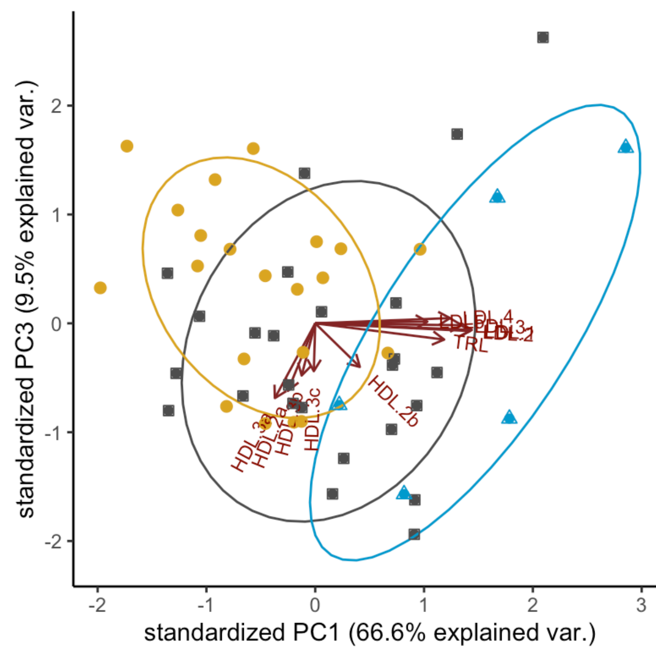


Figure 1. Principal component analysis biplot to visualize clusters in lipoprotein profile data from 20 Miniature Schnauzers with normal serum TG concentrations (NTG), 25 with primary hypertriglyceridemia (HTG), and 5 with secondary HTG. The first two principal components are plotted with loading vectors and ellipses drawn at 95% confidence intervals around the mean data points within NTG (circles), primary HTG (squares), and secondary HTG (triangles) groups.



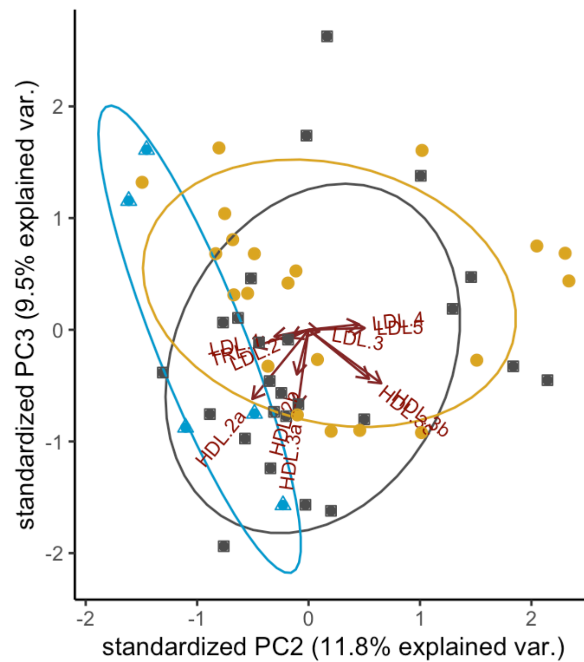


Figure 2. Principal component analysis biplots to visualize clusters in lipoprotein profile data from 20 Miniature Schnauzers with normal serum TG concentrations (NTG), 25 with primary hypertriglyceridemia (HTG), and 5 with secondary HTG. A) First and third principal components and B) the second and third principal components are shown. Ellipses drawn at 95% confidence intervals around the mean data points within NTG (circles), primary HTG (squares), and secondary HTG (triangles).

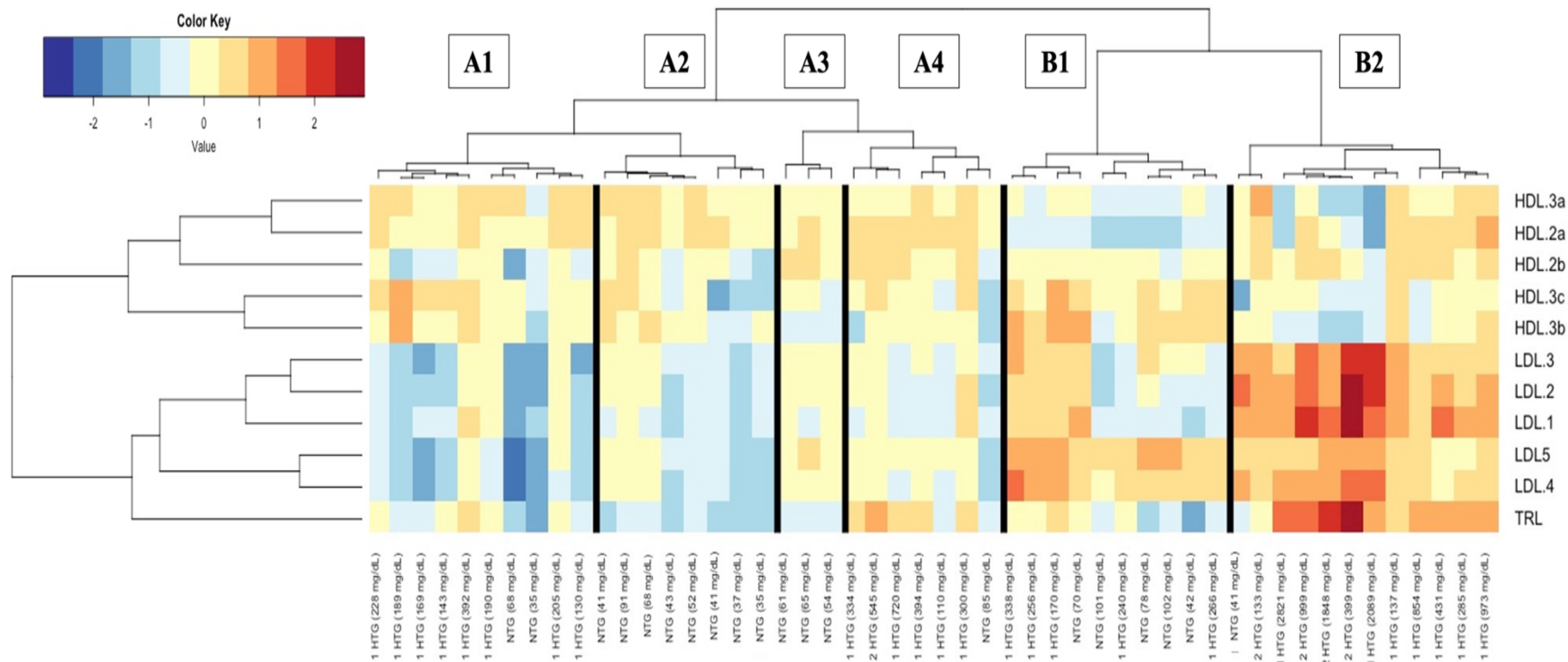


Figure 3. Hierarchical cluster analysis and heatmap of lipoprotein profiling data from 20 Miniature Schnauzer dogs with normal serum triglyceride concentrations (NTG), 25 with primary hypertriglyceridemia (1 HTG), and 5 with secondary HTG (2 HTG). The heatmap corresponds to the intensity of the 11 lipoprotein fractions across dogs. For each dog the clinical classification and triglyceride concentration (in parentheses) are included along the bottom. Clusters are separated by black lines. Cluster analysis was performed with Ward's method using the Spearman correlation distance for samples and Euclidean distance for lipoprotein fractions.

TRL – triglyceride-rich lipoproteins; LDL – low-density lipoproteins; HDL – high-density lipoproteins.

CHAPTER 3

Serum metabolomic and lipidomic analysis of Miniature Schnauzers with hypertriglyceridemia

Abstract/Summary

Analysis of lipidomic and metabolomic data has the potential to reveal pathophysiological processes and specific derangements that contribute to disease. The aim of this study was to ascertain lipid species and metabolites that distinguish Miniature Schnauzers with primary hypertriglyceridemia (HTG) from those with normal serum triglyceride concentrations (NTG). Serum samples from 30 Miniature Schnauzers with primary HTG (11 mild [109-250 mg/dL] and 19 moderate-to-severe [>250 mg/dL]), 24 with NTG [<109 mg/dL], and 5 with endocrinopathies and presumed secondary HTG were analyzed using lipidomic and metabolomic platforms. Comparison of dogs with moderate-to-severe primary HTG to those with NTG identified 881 lipid species and 9 metabolites that significantly differed (adjusted $p<0.05$). The differentiating lipid and metabolite species suggest involvement or disruption of the pathways and products of glycerolipid, glycerophospholipid, glycosphingolipid, and fatty acid metabolism. Hierarchical cluster analysis revealed that 7 of 11 Miniature Schnauzers with mild primary HTG clustered with NTG dogs rather than with the moderate-to-severe primary HTG dogs. This suggests that mild elevations in triglyceride concentrations might not always indicate a primary lipid disturbance. Dogs with presumed secondary HTG did not form a distinct cluster in the lipidomic analysis, but they did cluster together in the metabolomic analysis. These findings enable the further classification of Miniature Schnauzers with HTG for future studies and offer insights into potential underlying pathways and biomarkers for primary HTG.

Introduction

Primary hyperlipidemia is common in Miniature Schnauzers, with >75% affected by 10 years of age.[4] Within the breed, primary hyperlipidemia is characterized by hypertriglyceridemia (HTG) due to increases in triglyceride-rich lipoproteins (very-low density lipoproteins [VLDL] or a combination of chylomicrons and VLDL).[2,3] Hypertriglyceridemia is associated with an increased risk for pancreatitis, gallbladder mucoceles, glomerular proteinuria, and several other complications.[4,6-10]

The pathogenesis of HTG in Miniature Schnauzers is thought to be complex, with environmental and genetic contributors (see chapters 2 and 4).[4,12] However, the precise molecular mechanisms underlying HTG in Miniature Schnauzers are unknown. Distinctive alterations in the serum lipidome and serum or plasma metabolome occur with various diseases in dogs, including obesity, hyperadrenocorticism, hypothyroidism, diabetes mellitus, and gallbladder mucoceles.[39-46] Analysis of serum lipidomic and metabolomic data from Miniature Schnauzers with HTG could similarly elucidate the underlying pathogenic mechanisms.

The aim of this study was to ascertain serum lipid species and metabolites that distinguish Miniature Schnauzers with primary HTG from those with normal serum triglyceride concentrations (NTG). An additional aim was to determine how the lipidomic and metabolomic profiles of Miniature Schnauzers with presumed secondary HTG (underlying endocrinopathy) compare to those with primary HTG.

Materials and Methods

Samples

Serum samples were selected from Miniature Schnauzers with serum biobanked (-80˚C) at the University of Minnesota Canine Genetics Laboratory from past and ongoing research projects approved by the University of Minnesota Institutional Animal Care and Use Committee (protocols 1207A-17243, 1509-33019A, and 1807-36213A).[9,22,23] Samples with a fasting serum triglyceride (TG) concentration available (collected after asking owners to withhold food for 12-18 hours) were considered for inclusion. TG concentrations were measured by either a Roche/Hitachi Modular Analytics D2400 Module (Roche Diagnostics, Indianapolis, Indiana) or a Beckman Coulter AU480 Chemistry Analyzer (Beckman Coulter, Brea, California).

A sample qualified for inclusion in the study if obtained from a dog that met the criteria described below for having primary HTG, secondary HTG, or NTG. Samples were excluded if the dog was receiving glucocorticoids, fibrates, or statins at the time of serum collection. A minimum volume of 300 μL was required per sample to be run on both the complex lipid panel and the global metabolomics platform. If a sample only had enough serum available for one platform (<150 μL), the complex lipid panel was prioritized. As some samples had undergone previous freeze-thaw cycles to remove aliquots for other research, the total number of freeze-thaw cycles per sample was recorded.

Clinical Classification of Dogs

Information acquired from medical record data included age, sex, TG concentration, body condition score (BCS, 1-9 scale), fat content of the primary diet fed (g/100 kcal), current medications, diagnoses, and results of any endocrine testing performed. Dogs were classified as having primary HTG if their fasting serum TG concentration was >108 mg/dL and they had no previous diagnosis or clinical suspicion of an underlying condition reported to cause HTG (e.g., diabetes mellitus, hypothyroidism, hyperadrenocorticism, nephrotic syndrome) at the time of sampling and up to six months after. Dogs with proteinuria were permitted in the primary HTG classification if they were non-azotemic and normoalbuminemic.[23]

The severity of HTG was classified as mild for dogs with TG concentrations of 109-250 mg/dL and moderate-to-severe for concentrations >250 mg/dL. These classifications differ from previous studies that used 400 mg/dL as the cut off for mild versus moderate-to-severe.[4,9,23,47] The lower cut off was selected for two reasons. First, elevations in TG concentrations up to 150-250 mg/dL are commonly observed with aging in dogs.[34,35,38] Second, Miniature Schnauzers with TG concentrations ≤250 mg/dL often cluster with those with NTG based on analysis of lipoprotein profiles (see chapter 2). Secondary HTG was defined as a fasting serum TG concentration >108 mg/dL in a dog with concurrent endocrinopathy known to cause hyperlipidemia (e.g., diabetes mellitus, hypothyroidism, hyperadrenocorticism). To be classified as NTG, dogs had to be 8 years

of age or older at the time of sampling and have a fasting serum TG concentration ≤108 mg/dL.

Serum Metabolomic and Lipidomic Analysis

Serum samples were shipped overnight on dry ice to a commercial laboratory, Metabolon, Inc. (Durham, NC), where they were analyzed using Metabolon's Complex Lipid Panel (detects up to 1,100 lipid species per sample) and Precision Metabolomics global metabolomics platform (detects up to 5,200 metabolites per sample). Metabolite profiling was performed using Metabolon's standard protocols and software as previously described.[48] Briefly, serum samples were extracted and analyzed on gas chromatography-mass spectrometry and ultra-high performance liquid chromatography-tandem mass spectrometry platforms. Known lipid species and other metabolites were identified by matching ion chromatographic retention index and mass spectral fragmentation signatures with entries in the reference library.

Statistical Analyses

Data processing and all statistical analyses were performed using MetaboAnalyst5.0 (www.metaboanalyst.ca) and R statistical software (R, version 4.1.2, www.r-project.org).[25,49] Data distribution was evaluated with the Shapiro-Wilks test and quantile-quantile plot graphs. Clinical data were compared between moderate-to-severe primary HTG and NTG groups using a Student's t-test for age, the Wilcoxon rank sum test for dietary fat and BCS, and the Fisher's exact test for count data (sex). The level of statistical significance for these analyses was set at $p<0.05$. The median and range of TG

concentrations were determined for the moderate-to-severe primary HTG and NTG groups but were not statistically compared, as TG concentration was used for these group classifications.

Data filtering was not performed on the metabolomic and lipidomic datasets because both contained <5,000 features. Zero or missing values were replaced by the minimum positive value for each variable. Data were normalized to the sum, log transformed, and Pareto scaled prior to analysis. For both the lipid species and metabolites, linear models were used to identify those that differentiated moderate-to-severe primary HTG and NTG groups, with covariate adjustment for age, sex, dietary fat, and number of freeze thaw cycles (ranging from 1-3). As TG concentration was used to define the classification of dogs, it was not included in the linear model as a covariate. Due to the inability of MetaboAnalyst5.0 to analyze ordinal data and missing values for some dogs, BCS was not included as a covariate. Raw *p*-values were corrected for multiple testing using Benjamini and Hochberg's method, and the level of significance was set at an adjusted $p<0.05$.

Heatmaps of statistically significant features were generated from each linear model to identify clustering lipid species and metabolites. Lipid species and metabolites were further analyzed using the pathway analysis module, which integrates two methods, pathway enrichment and pathway impact analysis, from MetaboAnalyst5.0. Significant differences in pathways were evaluated using a Fisher's exact test, and pathway topology was analyzed based on the relative-betweenness centrality.

Random Forest analysis was performed for feature selection and classification of the metabolomic and lipidomic data sets using MetaboAnalyst5.0

(http:www.metaboanalyst.ca). All metabolites as well as age, sex, dietary fat, and freeze thaw cycles were tested as predictors of classification as NTG versus moderate-to-severe primary HTG.

The lipidomic and metabolomic datasets from all dogs, including those with NTG, any degree of primary HTG (mild or moderate-to-severe), and secondary HTG, were further analyzed using hierarchical cluster analysis to investigate if the lipidome and metabolome cluster according to these four classifications. First, principal component analysis (PCA) was performed to visually assess cluster tendency of the lipidomic and metabolomic data across all dogs. The first two principal components, which capture the most data variation, were examined for their ability to separate groups. The clustering tendency of the data was further assessed using the Hopkins statistic. A Hopkins statistic >0.5 indicates that data is non-random. Prior to cluster analysis, very highly correlated variables (r>0.9) were removed. The optimal number of clusters was determined using the 'fviz_nbclust' R package. Lipid species and metabolites were clustered using unsupervised, agglomerative, hierarchical clustering. The Ward linkage method was applied to Spearman correlation distances between variables. The median and range for clinical data, including age, sex, dietary fat, TG concentration (metabolomics data only), and BCS were determined for each cluster. Clusters were compared using a Kruskal-Wallis test for age, BCS, dietary fat, and TG concentration, and the Fisher's exact test for count data (i.e., sex, counts of HTG classification). The level of statistical significance for these analyses was set at $p<0.05$. TG concentrations were not statistically compared between clusters created from the lipidomic dataset because multiple TG species are included in the complex lipid panel.

Results

Samples

Biobanked serum was available from 95 Miniature Schnauzers. Of these, 78 dogs had a fasting serum TG concentration available. Nineteen dogs were excluded due to not meeting age criteria (n=1, NTG but only 5 years old), subsequent diagnosis of an endocrinopathy within a six month time frame (n=1, diabetes mellitus diagnosed three months after sample collection), chronic kidney disease (n=3), inconsistent HTG classification (n=4, all with NTG at the time of sample collection but one with HTG one year later and three with previous HTG that resolved on a low-fat diet with or without fibrate therapy), or suspected secondary cause of HTG (n=10; 1 with suspected but unconfirmed hypothyroidism, 3 with cholestasis, and 6 with suspected but unconfirmed hyperadrenocorticism).

The 59 remaining samples were included in the study and analyzed using Metabolon's Complex Lipid Panel. The breakdown by clinical classification was: 24 NTG, 30 primary HTG (11 mild and 19 moderate-to-severe), and 5 secondary HTG. Eight dogs did not have enough serum available to run both platforms. Thus, only 51 dogs were analyzed using Metabolon's Precision Metabolomics platform (22 NTG, 25 primary HTG [11 mild and 14 moderate-to-severe], and 4 secondary HTG). Forty-three of these dogs were previously included in an analysis of lipoprotein profiles to identify potential subtypes of hyperlipidemia in Miniature Schnauzers (Chapter 2). Age, sex, BCS, TG concentrations, and dietary fat are summarized in Table 1. No statistically significant differences were found in these variables between dogs with NTG versus moderate-to-severe primary HTG.

The extent of diagnostic testing for underlying endocrinopathies varied for the dogs with primary HTG. All 30 had fasting blood or serum glucose concentrations available (median 101 mg/dL, range 67 - 129 mg/dL). Fourteen were screened for hyperadrenocorticism (12 urine cortisol:creatinine ratios and 2 low dose dexamethasone suppression tests), and 14 were screened for hypothyroidism (total thyroxine concentration); results for these endocrine function tests were within laboratory reference intervals. The secondary HTG group comprised 3 dogs with untreated hypothyroidism and 2 dogs with diabetes mellitus (1 diagnosed the day the serum sample was obtained and 1 diagnosed 9 days prior and unregulated).

Comparison of Lipidomic Profiles Between Dogs With Moderate-to-Severe Primary HTG and NTG

Linear Model with Covariate Adjustment

After adjustment for multiple comparisons, 881 lipid species (out of 988 detected) were identified that differentiated NTG from moderate-to-severe primary HTG dogs (Supplementary Table 1). A pie chart representing the distribution of the lipid classes for the differentiating lipid species is shown in Figure 1. The majority (56%) of the differentiating lipid species were TG, with all 496 differentiating TG species increased in dogs with moderate-to-severe primary HTG. Also increased in moderate-to-severe primary HTG dogs were 38 (of 40) diglyceride (DG) species and 2 (of 12) of the monoglyceride (MG) species that differentiated between classifications. In contrast, all other differentiating lipid species (cholesterol esters (CE), ceramides (CER), dihydroceramides (DCER), hexosylceramides (HCER), lactosylceramides (LCER),

lysophosphatidylcholines (LPC), lysophosphatidylethanolamines (LPE), phosphatidylcholines (PC), phosphatidylethanolamines (PE), phosphatdiylinositols (PI), and sphingomyelins (SM)) were decreased in the moderate-to-severe primary HTG dogs. The highest proportion of these were PC (12%) and PE (9%).

A heatmap including the top 10 differentiating lipid species is shown in Figure 2. Total TG and seven TG species were included in the top 10 differentiating species and were increased in the moderate-to-severe primary HTG dogs relative to NTG dogs. Two LCER species, LCER(24:0) and LCER(24:1) were also in the top 10 differentiating lipid species and were decreased in moderate-to-severe primary HTG dogs.

MetaboAnalyst5.0 Pathway Enrichment and Impact Analysis

Of the 881 lipid species that significantly differentiated dogs with NTG from those with moderate-to-severe HTG, MetaboAnalyst's Pathway Analysis recognized the compound names for 598 (Supplementary Table 1); this subset was used in both the pathway enrichment and impact analyses. The pathway enrichment analysis determined that glycerophospholipid metabolism (p=0.0016) and linoleic acid metabolism (p=0.035) were enriched in the differentiating lipid species (Figure 3). Pathway impact analysis found the differentiating lipid species impacted glycerophospholipid metabolism, ether lipid metabolism, glycerolipid metabolism, and glycosylphosphatidylinositol (GPI)-anchor biosynthesis.

Random Forest Classification

Random forest classification with the lipidomic data showed perfect prediction (100%) for dogs with NTG and moderate-to-severe primary HTG with an out of bag error rate of 0% (Figure 4A). The random forest variable importance plot identified 15 lipid species ranked by their contribution to classification accuracy (Figure 4B). Of these, the majority were species of PC (n=3), PE (n=5), and TG (n=4).

Metabolomic Comparison of Moderate-to-Severe Primary HTG versus NTG

Linear Model with Covariate Adjustment

After adjustment for multiple comparisons, 9 metabolites (of 803 detected) were identified that differentiated dogs with NTG from those with moderate-to-severe primary HTG (Supplementary Table 2). A pie chart representing the distribution of the metabolite class for the differentiating metabolites is shown in Figure 5. The majority (n=4) of the differentiating metabolites were part of lipid metabolism. A heatmap of the differentiating metabolites is shown in Figure 6. Metabolites involved in long chain fatty acid (saturated and unsaturated) metabolism were increased in dogs with moderate-to-severe primary HTG. These included arachidoylcarnitine (C20), dihomolinolenate (20:3n3 or 3n6), and behenoylcarnitine. Two metabolites were decreased in moderate-to-severe primary HTG dogs, including one of Metabolon's proprietary metabolites and branched chain 14:0 dicarboxylic acid (involved in dicarboxylate fatty acid metabolism).

MetaboAnalyst5.0 Pathway Enrichment and Impact Analysis

Of the nine metabolites that significantly differentiated dogs with NTG from dogs with moderate-to-severe HTG, four were recognized by MetaboAnalyst's Pathway Analysis (Supplementary Table 2). Pathway enrichment analysis and pathway impact analysis determined biotin metabolism to be impacted by the differentiating metabolites (p=0.019) (Figure 7).

Random Forest Classification

The metabolomic data showed excellent prediction of classification (92%), with all 22 dogs with NTG and 11/14 dogs with moderate-to-severe primary HTG classified correctly and an out of bag error rate of 0.08% (Figure 8A). The random forest variable importance plot identified 15 metabolites ranked by their contribution to classification accuracy (Figure 8B). Of these, 10 were part of lipid metabolism, including the following pathways: endocannabinoid, fatty acid metabolism (acyl carnitine, long chain saturated), fatty acid metabolism (acyl choline), fatty acid metabolism (dicarboxylate), long chain polyunsaturated fatty acid (n3 and n6), long chain saturated fatty acid, and lysophospholipid.

Hierarchical Cluster Analysis of All Dogs

Principal Component Analysis

Cluster tendency was evaluated with the Hopkins statistic and visually with PCA. The Hopkins statistic for the lipidomic and metabolomic datasets were 0.70 and 0.61, respectively. The PCA of the lipidomic data indicated clear separations of all four classifications, with the first principal component driven by severity of HTG (Figure 9A).

The first two principal components accounted for 53% and 11% of the total data variation. The PCA of the metabolomic data revealed little to no separation between classifications (Figure 9B). The first two principal components accounted for 18% and 9% of the total data variation.

Hierarchical Cluster Analysis of Lipidomic Data

After removal of highly correlated variables, 303 lipid species were included in the hierarchical cluster analysis. The optimal number of clusters was determined to be six. These clusters are shown in Figure 10. Triglyceride concentrations differed significantly by cluster (Table 2). The other clinical data (i.e., age, sex, dietary fat, and BCS) did not differ significantly by cluster (Table 2).

The most dissimilar nodes, separated by the first branch of the dendrogram, were referred to as A and B. Node A comprised all the dogs with moderate-to-severe primary HTG (n=19) and secondary HTG (n=5), and 4/11 dogs with mild primary HTG. Node B comprised all the dogs with NTG and 7/11 with mild primary HTG. Dogs in node A had higher abundance of TG and DG and a lower abundance of all other species (CE, CER, MG, DCER, HCER, LCER, LPC, LPE, PC, PE, PI, and SM). The inverse was true for dogs in node B.

Dogs in cluster A1 (10 with moderate-to-severe primary HTG and 2 with secondary HTG) had the highest abundance of TG and DG species and lowest abundance of all other species. The most dissimilar cluster from A1 was cluster B3. Dogs in cluster

B3 (15 NTG and 1 mild primary HTG) had the lowest abundances of TG and DG species and high abundances of all other species.

Hierarchical Cluster Analysis of Metabolomic Data

After removal of highly correlated variables, 732 metabolites were included in the hierarchical cluster analysis, using four as the optimal number of clusters. The clusters are shown in Figure 11. Clinical data (i.e., age, sex, TG concentration, dietary fat, and BCS) for each cluster is summarized in Table 3; serum TG concentration was the only variable that differed significantly between clusters.

Cluster B3 comprised all four secondary HTG dogs included in this study, as well as three dogs with moderate-to-severe primary HTG. The remaining three clusters were not as clearly defined with NTG, mild primary HTG, and moderate-to-severe primary HTG dogs present in each.

Discussion

The serum lipidome and, to a lesser degree, metabolome differ between Miniature Schnauzers with moderate-to-severe primary HTG (defined in this study as fasting serum TG concentrations >250 mg/dL) and those with NTG (<109 mg/dL). Eighty-nine percent of the lipid species measured differed between these two groups with the majority being TG species, which were increased in Miniature Schnauzers with moderate-to-severe primary HTG. The differentiating lipid species suggest possible involvement of the pathways and products of glycerolipid, glycerophospholipid, and glycosphingolipid metabolism. In contrast to the lipidome, only 1% of the metabolites

significantly differed between groups. Most differentiating metabolites are involved in fatty acid metabolism pathways. Cluster analysis of lipidomic data resulted in perfect separation of moderate-to-severe HTG (primary or secondary) from NTG. In contrast, dogs with mild HTG (109-250 mg/dL) were unpredictable with many clustering with NTG dogs. Cluster analysis of the metabolome resulted in one cluster comprising all four dogs with secondary HTG and three with moderate-to-severe primary HTG. The remaining three clusters contained a mix of NTG, mild primary HTG, and moderate-to-severe primary HTG dogs.

In the analysis of the lipidomic data, 56% of the lipid species that differentiated moderate-to-severe HTG from NTG were TG species. All 496 differentiating TG species were increased in Miniature Schnauzers with moderate-to-severe primary HTG. Increases in 38 of the 40 species of DG were also seen in dogs with moderate-to-severe HTG. These two lipid classes are involved in the glycerolipid metabolism pathway. In this pathway, phosphatidic acid phosphatase converts phosphatidic acid to DG.[50] The DG can then be converted to TG through the addition of an acyl group by DG acyltransferases. TG can be stored in the liver or secreted in the form of VLDL.[51]

All other differentiating lipid species, including CE, CER, DCER, HCER, LCER, LPC, LPE, PE, PC, PI, SM, and all but two MG species, were decreased in dogs with moderate-to-severe primary HTG. Of these, a PC and HCER species were most important for classification in the random forest analysis. However, removal of any individual lipid species had a relatively small effect on the mean accuracy, consistent with the discovery that most lipid species tested differentiated the groups. The two LCER species that were among the top 10 differentiating lipid species, LCER 24:0 and 24:1, are also decreased in

rhesus monkeys with metabolic syndrome.[52] While HCER, LPE, PC, and PE species were lower in the Miniature Schnauzers with moderate-to-severe primary HTG in this study, the abundance of these species is greater in the lipidome of dogs with hypothyroidism and hyperadrenocorticism compared to healthy dogs.[40,42] Dogs with hypothyroidism also have increases in CE, CER, LPE, LPC, and PI species.[40] Thus, these findings suggest that lipidome disturbances in Miniature Schnauzers with primary HTG differ from those driven by endocrinopathies.

Alterations in lipid species abundance might reflect specific pathways contributing to HTG. The phospholipids PC and PE are synthesized through the glycerophospholipid metabolism pathway. Mice deficient in one of the transferase enzymes that convert DG to PC or PE (CTP:phosphoethanolamine cytidylyltransferase) develop HTG due to redirection of DG to TG.[53] The decrease in PC and PE in Miniature Schnauzers with primary HTG might indicate inefficiencies in glycerophospholipid metabolism.

In contrast to the lipidomic analysis, where the bulk of lipid species were differentiating, only nine metabolites differed between Miniature Schnauzers with moderate-to-severe primary HTG and those with NTG in the global metabolomics data. Seven of these were increased in dogs with moderate-to-severe primary HTG, including two long-chain acylcarnitines (arachidoylcarnitine (C20) and behenoylcarnitine (C22)). Increases in long-chain and very-long-chain acylcarnitines are linked to obesity, insulin resistance, and type 2 diabetes.[54-57] The dogs with moderate-to-severe primary HTG also had increases in dihomolinolenate (20:3n3 or 3n6). Similar to acylcarnitines, increased

dihomolinolenate (20:3n3 or 3n6) is associated with obesity, insulin resistance, and risk of type 2 diabetes.[58-60]

Lipidomic and metabolomic data from two additional groups of Miniature Schnauzers, those with mild primary HTG (108-249 mg/dL) and secondary HTG, were also used in the hierarchical cluster analysis to determine how all clinical classifications of HTG compare with clustering of lipid and metabolite profiles. Six clusters were identified in the hierarchical cluster analysis of lipidomic data from all four classifications. Cluster analysis was able to distinctly separate the moderate-to-severe primary HTG dogs (A clusters) from those with NTG (B clusters), consistent with the significantly different lipidomic profiles of these two classifications. However, cluster analysis of the lipidomic data did not result in a separation of primary from secondary HTG dogs. This could suggest that lipidomics is insufficient to distinguish lipidomic profiles of Miniature Schnauzers with secondary HTG from those with moderate-to-severe primary HTG. However, only five secondary dogs were included in the study, two with diabetes mellitus and three with hypothyroidism. Thus, it is also possible that the number of dogs with secondary HTG, especially for each disease, was insufficient to capture the characteristic lipidome features of these disorders and segregate them into their own cluster. Cluster analysis was also unable to segregate dogs with mild primary HTG into their own cluster. The majority (7/11) of mild primary HTG dogs were clustered in those containing NTG dogs (B clusters) rather than in the clusters containing moderate-to-severe primary HTG dogs (A clusters). This suggests that dogs with mild fasting serum TG concentrations (109-249 mg/dL) cannot be definitively classified as

having primary HTG. It is possible that for some dogs, the mild TG elevations could be a result of age or other environmental factors, such as diet.[4,77,78]

Four clusters were identified in the hierarchical cluster analysis of metabolomic data from all four classifications. One cluster contained all the secondary HTG dogs and three moderate-to-severe primary HTG dogs. This might be due to metabolic differences in dogs with endocrinopathies that are independent of the hyperlipidemia. While our study was underpowered to test this hypothesis, previous studies have identified major alterations of the metabolome in dogs with hyperadrenocorticism, hypothyroidism, and diabetes mellitus.[40,42-45] These studies suggest the involvement of glycolysis/gluconeogenesis, tryptophan metabolism, pentose phosphate pathway, aminoacyl-tRNA biosynthesis, and pyrimidine metabolism. The combination of moderate-to-severe primary HTG dogs with the secondary HTG dogs in the same cluster could indicate similarities in the metabolic response to increased TG concentrations in these dogs. Alternatively, it is possible that the three moderate-to-severe primary HTG dogs in this cluster had an undiagnosed subclinical endocrinopathy. The remaining three clusters were not clearly defined with NTG, mild primary HTG, and moderate-to-severe primary HTG dogs scattered throughout those three clusters.

In conclusion, this study identified differences of the serum lipidome and metabolome between Miniature Schnauzers with moderate-to-severe primary HTG and those with NTG. These changes suggest possible involvement or disruption of the pathways and products of glycerolipid, glycerophospholipid, glycosphingolipid, and fatty acid metabolism. Of note, the direction of change for many of the non-TG lipid species was the opposite of what is found in dogs with hyperadrenocorticism or hypothyroidism,

supporting that Miniature Schnauzers with primary HTG have a different type of dyslipidemia. Based on the results of the hierarchical cluster analysis, Miniature Schnauzers with mild HTG cannot be definitively classified as having primary HTG, as their lipid disturbances do not reliably differentiate them from dogs with NTG. These findings inform the definition and classification of primary HTG in Miniature Schnauzers and offer insights into potential underlying pathways and biomarkers for the disorder.

Table 1. Clinical characteristics of Miniature Schnauzers with normal serum fasting triglyceride concentrations (NTG), mild primary hypertriglyceridemia (HTG), moderate-to-severe primary HTG, or secondary HTG with serum analyzed using Metabolon's Complex Lipid Panel and Precision Metabolomics platforms. *P* values are reported for statistical comparisons between the NTG and moderate-to-severe primary HTG groups.

Group Characteristics	NTG (<108 mg/dL)	Moderate-to-severe primary HTG (>250mg/dL)	*P* value for NTG vs moderate-to-severe primary HTG	Mild primary HTG (108-249 mg/dL)	Secondary HTG
Complex Lipid Panel					
Total number, n	24	19	-	11	5
Age yrs, median (range)	10 (7-13)	11 (6-15)	0.28	10 (8-13)	13 (8-14)
Sex# male,female	17,7	11,8	0.52	6,5	3,2
Body condition score (BCS), median (range)	5 (3-8)*	6 (3-7)*	0.23	6 (4-7)*	5 (5-5)*
Serum TG mg/dL, median (range)	53 (14-102)	467 (254-2821)	-	170 (130-246)	999 (399-1848)
Dietary fat, g/100 kcal, median (range)	4.0 (2.5-4.8)	4.0 (2.5-5.9)	0.19	3.5 (2.7-4.6)^	2.8 (2.6-3.4)
Global Metabolomics					
Total number, n	22	14	-	11	4
Age yrs, median (range)	10 (7-13)	10 (6-14)	0.70	10 (8-13)	13 (8-14)
Sex# male,female	15,7	9,5	1	6,5	3,1
BCS, median, range	5 (3-8)*	5 (5-7)*	0.51	6 (4-7)*	5 (5-5)*
Serum TG mg/dL, median (range)	53 (14-102)	385 (254-2821)	-	170 (130-246)	772 (399-1848)
Dietary fat, g/100 kcal, median (range)	4.0 (2.5-4.8)	4.0 (2.5-5.9)	0.13	3.5 (2.7-4.6)^	2.9 (2.8-3.4)

*BCS unknown for 2 NTG, 1 ModSev primary HTG, 1 mild primary HTG, and 2 secondary HTG.

^Dietary fat content unknown for 2 mild HTG

#All dogs were spayed or neutered except for two intact males, one with mild HTG and one with secondary HTG.

Table 2. Comparison of clinical variables across six clusters of lipidomic data from 59 Miniature Schnauzers, including 24 with normal serum triglyceride concentrations (NTG), 11 with mild primary hypertriglyceridemia, 19 with moderate-to-severe primary HTG and 5 with secondary hypertriglyceridemia. *P* values are for comparisons to determine if the clinical variable differed by cluster. Mild primary HTG is defined as a fasting serum triglyceride concentration of 109 – 249 mg/dL, and moderate-to-severe primary HTG is defined as a concentration >250 mg/dL.

Cluster	A1 n=12	A2 n=5	A3 n=11	B1 n=11	B2 n=4	B3 n=16	*P* value
Variables							
Age yrs, median (range)	12 (8-15)	11 (6-14)	11 (6-14)	9 (8-12)	11 (9-13)	10 (7-13)	0.51
Sex# male,female	7,5	5,0	6,5	6,5	3,1	11,5	0.55
BCS, median, range	6 (5-7)	6 (5-7)*	5 (3-7)*	5 (3-6)*	6 (5-8)	5 (4-6)*	0.41
Dietary fat, g/100 kcal, median (range)	4 (2.5-5.9)	3.6 (2.8-4.0)^	3.6 (2.8-4.4)^	3.9 (2.5-4.6)	3.8 (3.1-4.3)	4 (2.7-4.8)	0.60
Serum TG mg/dL, median (range)	914 (266-2821)	285 (190-399)	300 (68-1848)	68 (14-170)	145 (85-246)	52 (24-130)	<0.001
Classifications							
# of NTG	0	0	0	7	2	15	<0.001
# of Mild primary HTG	0	2	2	4	2	1	
# of Moderate-to-severe primary HTG	10	2	7	0	0	0	
# of secondary HTG	2	1	2	0	0	0	

*indicates unknown BCS for 2 NTG, 1 ModSev HTG, 1 mild HTG, and 2 secondary HTG.

^indicates unknown dietary fat content for 2 mild HTG

#All dogs were spayed or neutered except for two intact males, one with mild HTG and one with secondary HTG.

Table 3. Comparison of clinical variables across four clusters of metabolomic data from 51 Miniature Schnauzers including 22 with normal serum fasting triglyceride concentrations (NTG), 11 with mild primary hypertriglyceridemia (Mild pHTG), 14 with moderate-to-severe primary HTG (ModSev pHTG), and 5 with secondary hypertriglyceridemia (2HTG). P values are for comparisons to determine if the clinical variable differs by cluster. Mild primary HTG is defined as a fasting serum triglyceride concentration of 109 – 249 mg/dL, and moderate-to-severe primary HTG is defined as a concentration >250 mg/dL.

Cluster	A1 n=19	B1 n=19	B2 n=6	B3 n=7	*P* value
Variables					
Age yrs, median (range)	10 (7-14)	10 (6-13)	11 (9-13)	13 (8-14)	0.11
Sex# male,female	10,9	14,5	3,3	6,1	0.32
Body Condition Score (BCS), median (range)	6 (4-8)	5 (3-7)	5 (3-6)	5 (5-7)	0.64
Dietary fat, g/100 kcal, median (range)	4.0 (2.5-4.3)	3.7 (2.7-4.8)	4.3 (2.8-5.9)	2.9 (2.8-5.1)	0.09
Serum TG mg/dL, median (range)	101 (14-2821)	137 (35-467)	121 (35-720)	854 (334-1848)	**0.004**
Classifications					
# of NTG	11	8	3	0	
# of Mild primary HTG	2	8	1	0	**0.001**
# of Moderate-to-severe primary HTG	6	3	2	3	
# of secondary HTG	0	0	0	4	

*BCS known for 2 NTG, 1 ModSev HTG, 1 mild HTG, and 2 secondary HTG.

^Dietary fat content unknown for 2 mild HTG

#All dogs were spayed or neutered except for two intact males, one with mild HTG and one with secondary HTG.

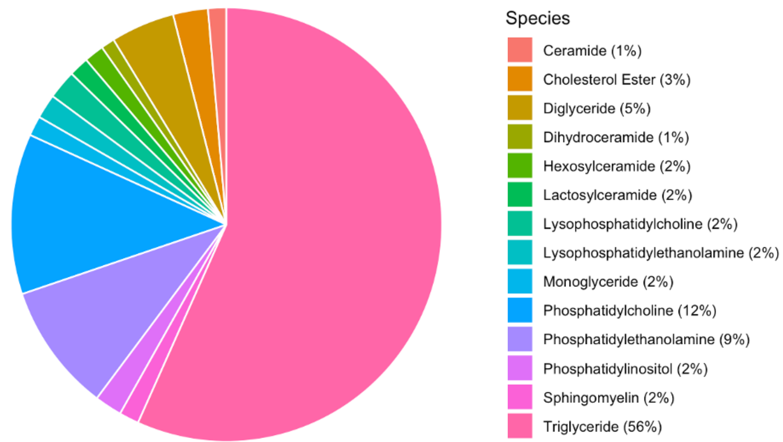


Figure 1. Pie chart displaying the distribution of lipid classes represented by 881 lipid species that differentiated serum from 19 Miniature Schnauzers with moderate-to-severe primary hypertriglyceridemia from 24 with normal serum triglyceride concentrations.

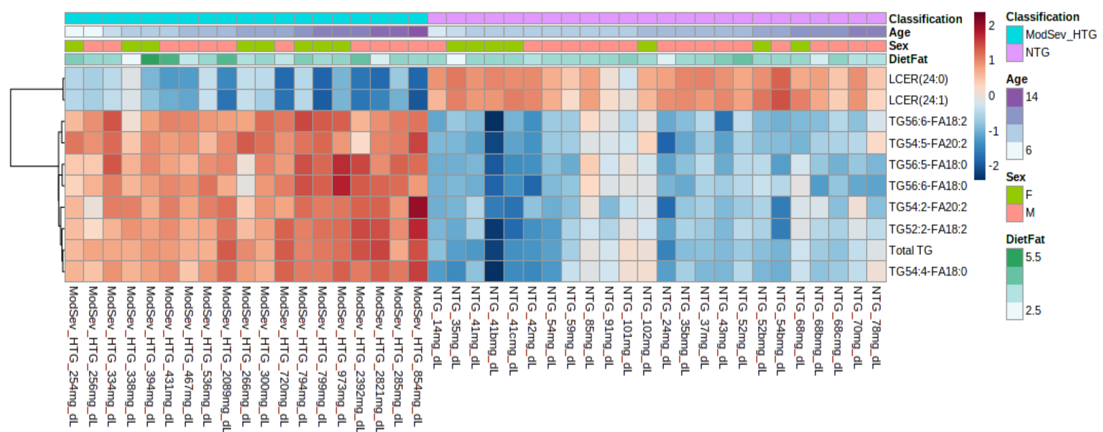


Figure 2. Heatmaps representing the top 10 lipid species that differentiate (*p*<0.05) between serum from Miniature Schnauzers with moderate-to-severe primary hypertriglyceridemia (ModSev_HTG; n=19) versus those with normal serum triglyceride concentrations (NTG; n=24). Age (years), sex, and dietary fat (DietFat; g/100 kcal) are also indicated for each dog (rows two through four). The heatmap corresponds to the intensity of the lipid species for each dog (column). The clinical classification and triglyceride concentration for each dog are included along the bottom. The lipid species name are included to the right (rows).

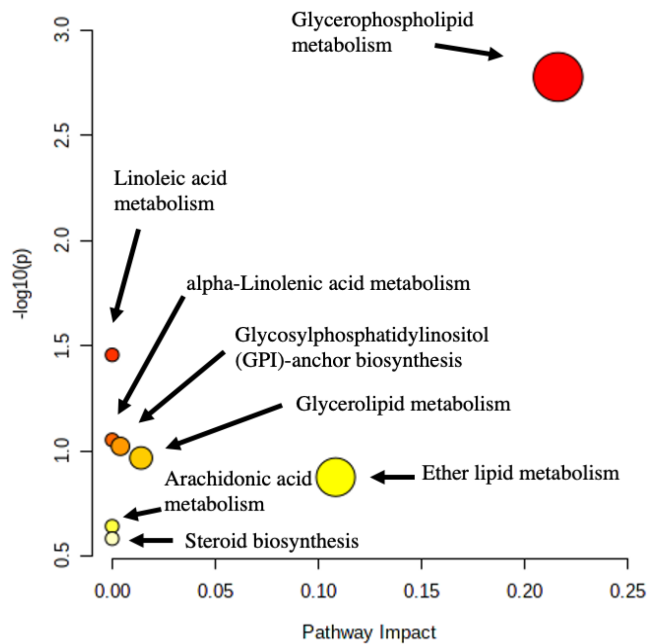


Figure 3. Pathway analysis of the lipid species that significantly differ between serum from 19 Miniature Schnauzers with moderate-to-severe primary hypertriglyceridemia and 24 with normal serum triglyceride concentrations. Only those that were found to be significant with pathway enrichment analysis (y axis) and/or those found to have an impact with pathway impact analysis are labeled. The node radius is based on pathway impact values and color is based on *p*-value.

A

		Groups assigned by random forest		
Classification	**Dogs, n**	**ModSev_HTG**	**NTG**	**Classification Error**
ModSev_HTG	19	19	0	0
NTG	24	0	24	0

B

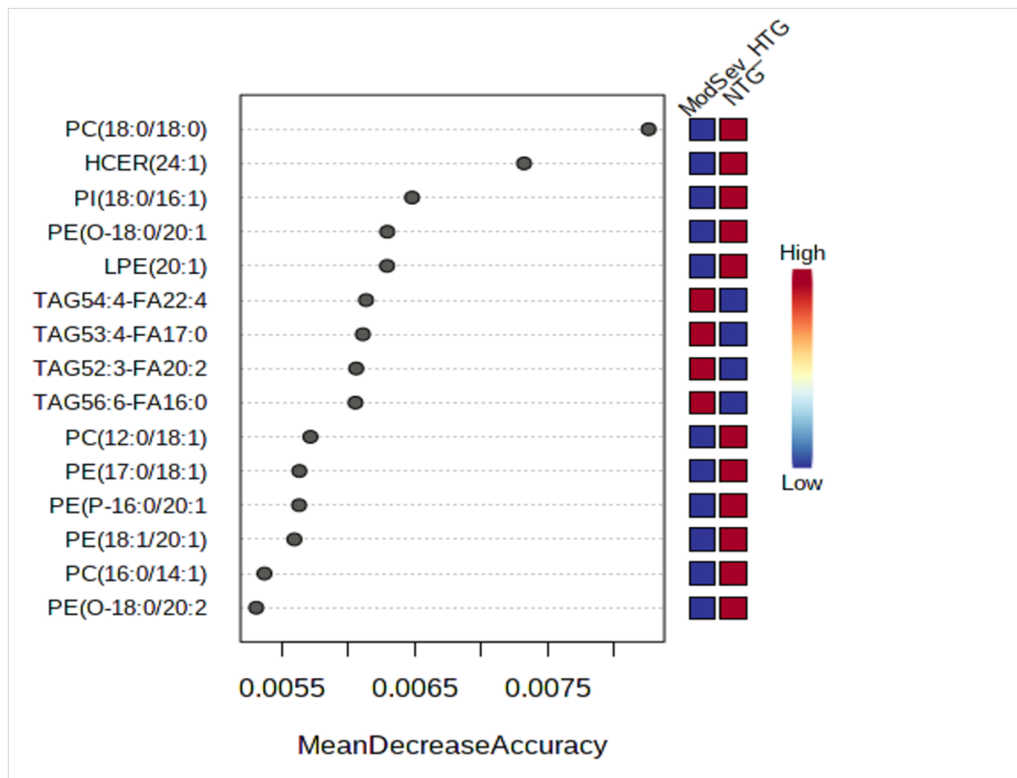


Figure 4. Random forest model to classify Miniature Schnauzers with moderate-to-severe primary hypertriglyceridemia (ModSev_HTG) versus those with normal serum triglyceride concentrations (NTG) using serum lipidomic data. A) The lipidomic random forest model correctly classified 100% of dogs. B) Random forest variable importance plots from lipidomic data analysis. Mean decrease accuracy is a measure of the performance of the model without each metabolite. Higher values indicate the importance of a lipid species in predicting classification (NTG or moderate-to-severe primary HTG) and its removal would result in a loss of accuracy of the model.

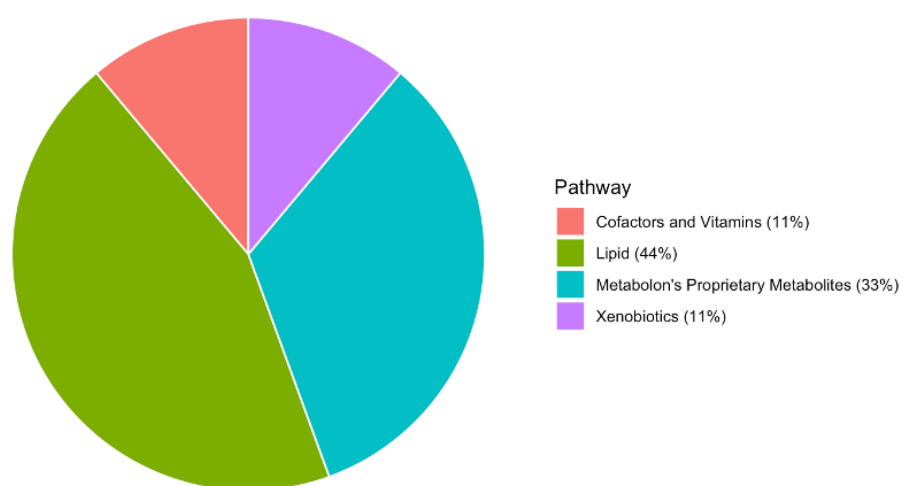


Figure 5. Pie chart displaying the distribution of metabolite classes represented by nine metabolites that differentiated serum from 19 Miniature Schnauzers with moderate-to-severe primary hypertriglyceridemia from 24 with normal serum triglyceride concentrations.

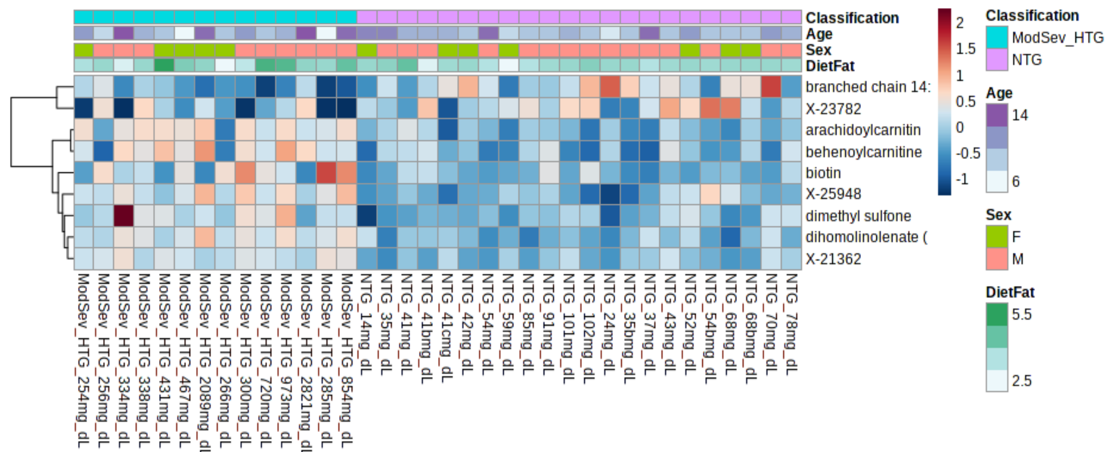


Figure 6. Heatmap of nine metabolites that differentiate ($p<0.05$) between serum from Miniature Schnauzers with moderate-to-severe primary hypertriglyceridemia (ModSev_HTG; n=14) versus those with normal serum triglyceride concentrations (NTG; n=22). Age (years), sex, and dietary fat (DietFat; g/100 kcal) are also indicated for each dog (rows two through four). The heatmap corresponds to the intensity of the metabolites for each dog (column). The clinical classification and triglyceride concentration for each dog are included along the bottom. The metabolite names are included to the right (rows).

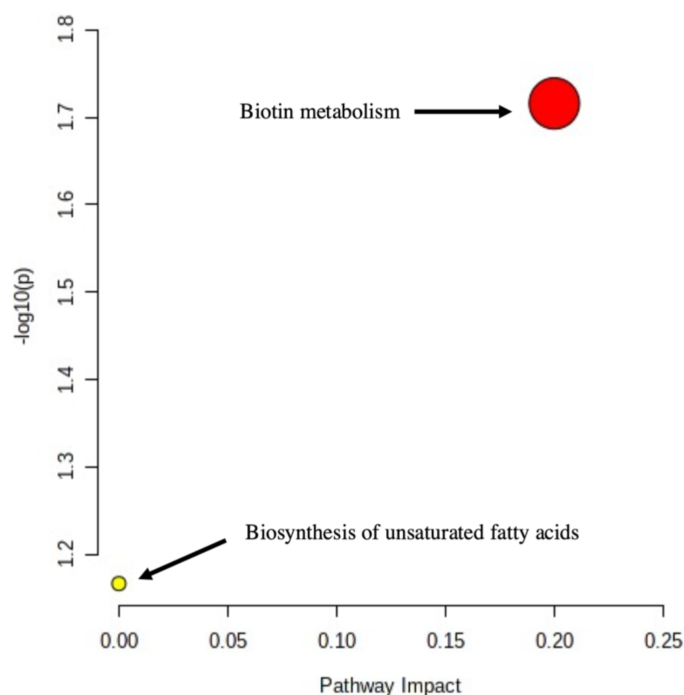


Figure 7. Pathway analysis of the metabolites which significantly differ between serum from 14 Miniature Schnauzers with moderate-to-severe primary hypertriglyceridemia and 22 with normal serum triglyceride concentrations. Only those which were found to be significant with pathway enrichment analysis (y axis) and/or those found to have an impact with pathway impact analysis are labeled. The node radius is based on pathway impact values and color is based on *p*-value.

A

		Groups assigned by random forest		
Classification	**Dogs, n**	**ModSev_HTG**	**NTG**	**Classification Error**
ModSev_HTG	14	11	3	0.21
NTG	22	0	22	0

B

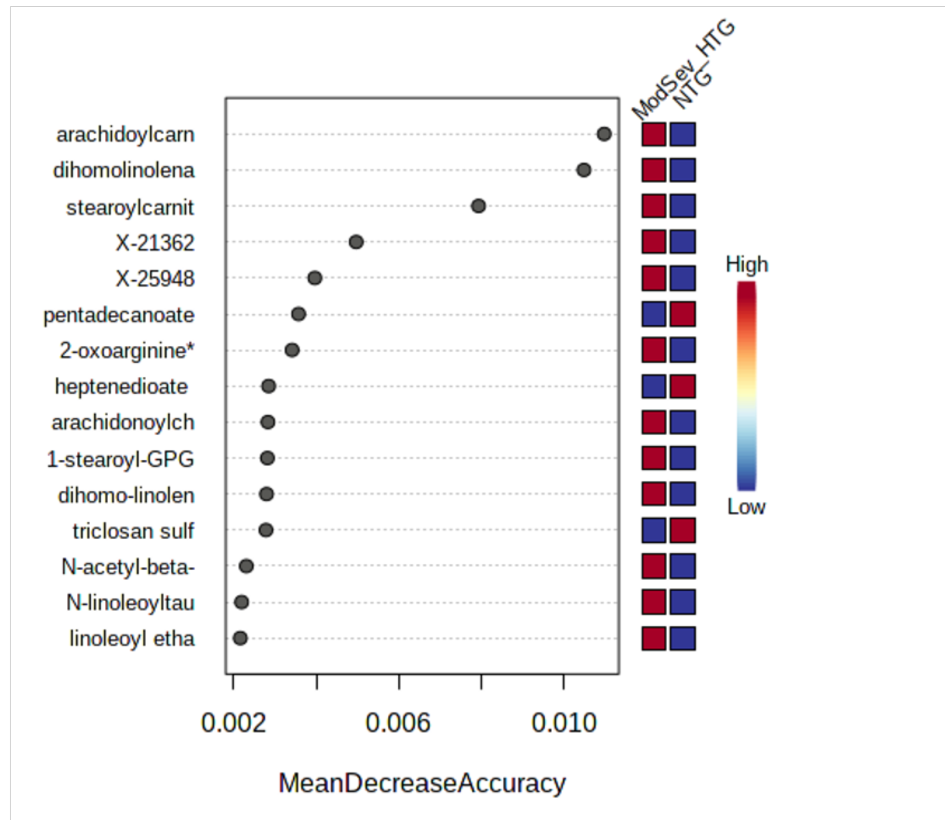


Figure 8. Random forest model to classify Miniature Schnauzers with moderate-to-severe primary hypertriglyceridemia (ModSev_HTG) versus those with normal serum triglyceride concentrations (NTG) using serum metabolomic data. A) The metabolomic random forest model correctly classified 92% of dogs. B) Random forest variable importance plots from metabolomic data analysis. Mean decrease accuracy is a measure of the performance of the model without each metabolite. Higher values indicate the importance of a lipid species in predicting classification (NTG or moderate-to-severe primary HTG) and its removal would result in a loss of accuracy of the model.

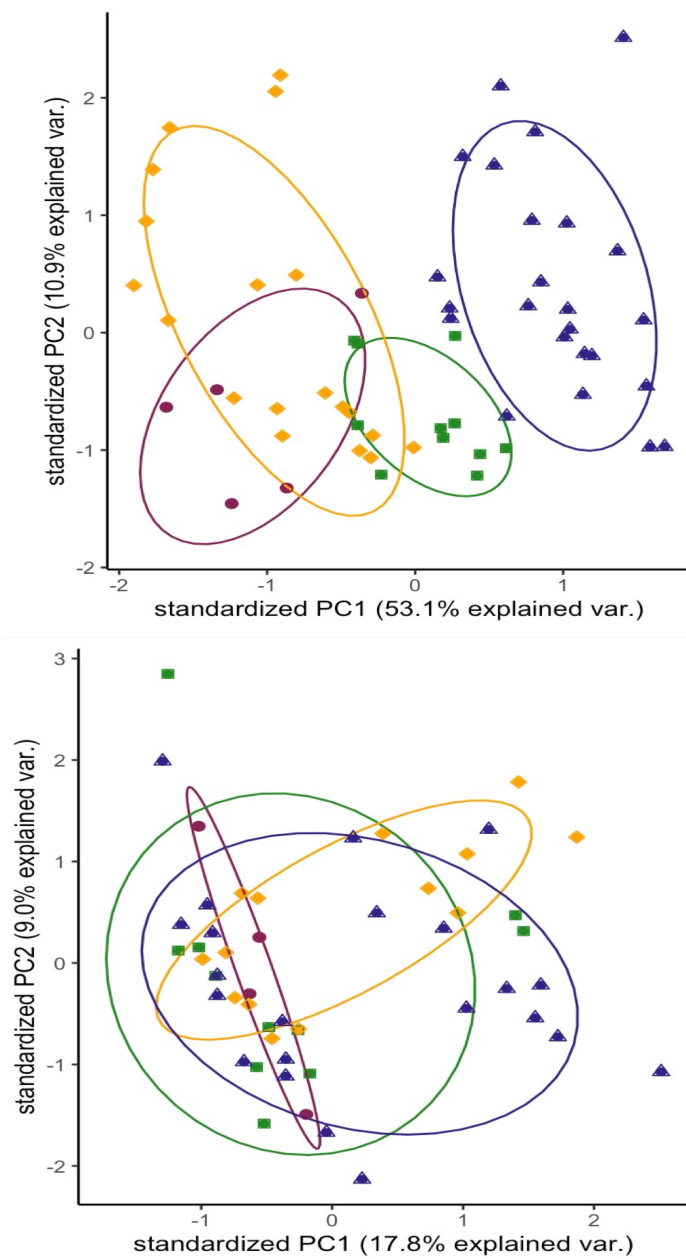


Figure 9. Principal component analysis of the first and second principal component of lipidomic (top) and metabolomic data (bottom) from Miniature Schnauzers with normal serum triglyceride concentrations (triangles), mild primary hypertriglyceridemia (squares), moderate-to-severe primary hypertriglyceridemia (diamonds), and secondary hypertriglyceridemia (circles). For each, the first two principal components are plotted with ellipses drawn at 95% intervals around the mean.

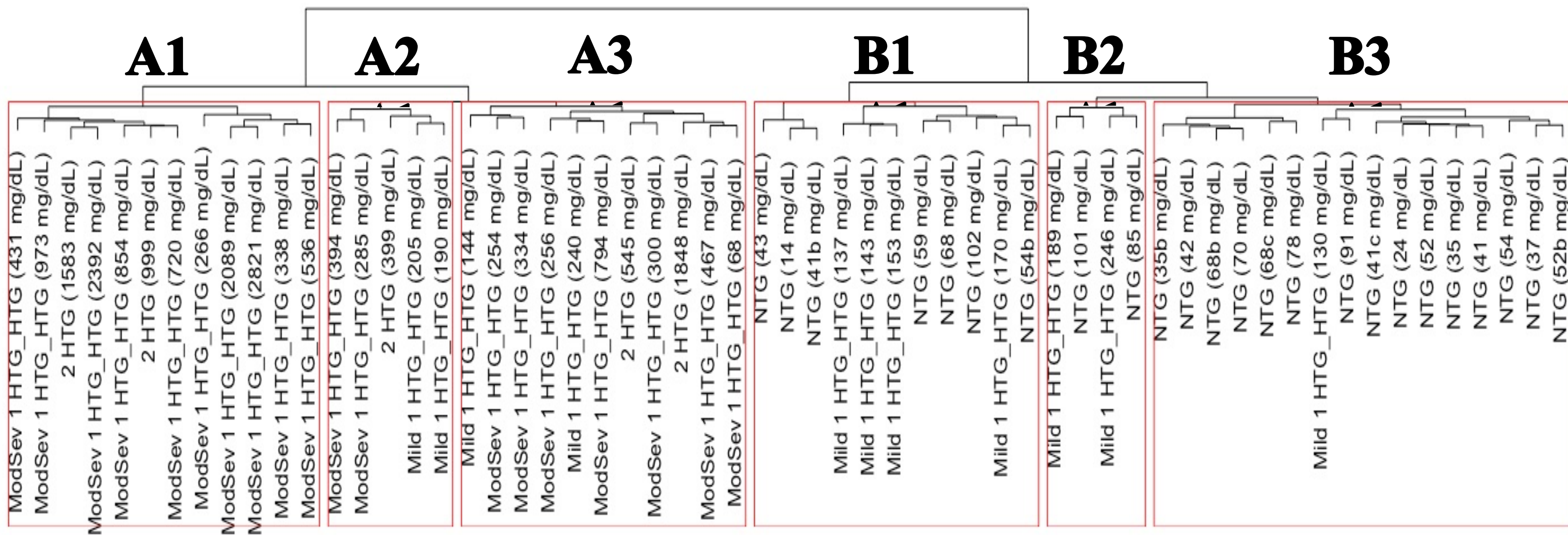


Figure 10. Hierarchical cluster analysis of serum lipidomic data from 24 Miniature Schnauzer dogs with normal serum fasting triglyceride concentrations, 11 with mild primary hypertriglyceridemia (Mild 1 HTG), 19 with moderate-to-severe primary HTG (ModSev 1 HTG), and 5 with endocrinopathies and presumed secondary HTG (2 HTG). For each dog the clinical classification and triglyceride concentration (in parentheses) are included along the bottom. Clusters are indicated with boxes. Cluster analysis was performed with Ward's method using the Spearman correlation distance for samples and Euclidean distance for lipid species.

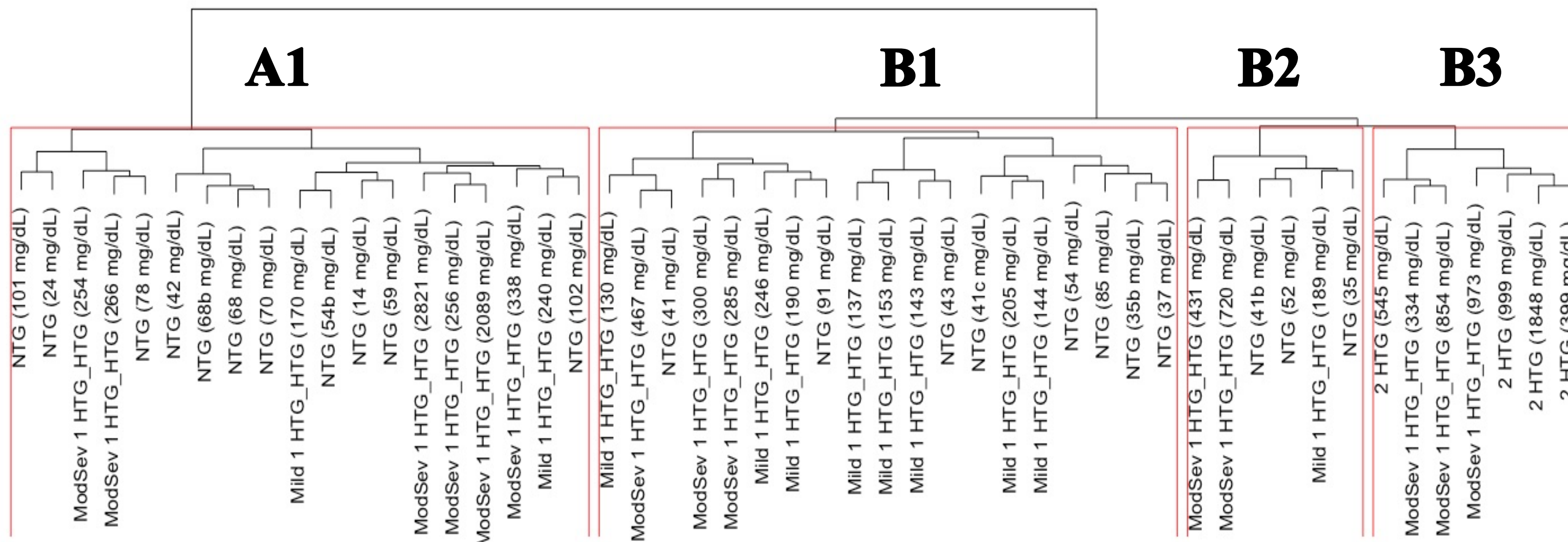


Figure 11. Hierarchical cluster analysis of serum metabolomic data from 22 Miniature Schnauzer dogs with normal serum triglyceride concentrations (NTG), 11 with mild primary hypertriglyceridemia (Mild 1 HTG), 14 with moderate-to-severe primary HTG (ModSev 1 HTG), and 4 with secondary HTG (2 HTG). For each dog the clinical classification and triglyceride concentration (in parentheses) are included along the bottom. Clusters are separated by boxes. Cluster analysis was performed with Ward's method using the Spearman correlation distance for samples and Euclidean distance for lipid species.

CHAPTER 4

Sequence analysis of six candidate genes in Miniature Schnauzers with primary hypertriglyceridemia

Abstract

Miniature Schnauzers are predisposed to primary hypertriglyceridemia (HTG), but the underlying genetic determinants are unknown. In this study, we performed whole genome sequencing of eight Miniature Schnauzers with primary HTG and screened for risk variants in six HTG candidate genes: *LPL, APOC2, APOA5, GPIHBP1, LMF1,* and *APOE*. Variants were filtered to identify those present in two or more Miniature Schnauzers with primary HTG and uncommon (<10% allele frequency) in a genomic variant database, including 613 dogs from 61 other breeds. Two variants passed filtering: a three bp deletion in the TATA box of *APOE* and a missense variant in Ly-6 domain of *GPIHBP1*. The *APOE* variant was exclusive to Miniature Schnauzers and disturbed a conserved base. It was genotyped in a cohort of 159 Miniature Schnauzers, including 91 with primary HTG and 68 control dogs without HTG. A multivariable regression, including age and sex, did not identify an effect of *APOE* genotype on triglyceride concentration (estimate = -0.03, std. error = 0.13; $P < 0.82$). The *GPIHBP1* missense variant resided at a position with low base-wise conservation and was predicted to have a neutral impact on the protein; follow-up genotyping was not performed. In conclusion, we did not identify a monogenic cause for primary HTG in Miniature Schnauzers in the six genes evaluated. However, if HTG in Miniature Schnauzers is a complex disease resulting from the cumulative effects of multiple variants and environment, one or both variants identified might be contributing factors. Evaluation of other genes is also warranted.

Introduction

Miniature Schnauzers have a predisposition for primary hypertriglyceridemia (HTG), defined as an elevated serum or plasma triglyceride concentration in the absence of an identifiable underlying cause.[4] Complications associated with HTG include pancreatitis, gallbladder mucoceles, glomerular proteinuria, and others.[4,6-10] An underlying genetic risk factor is suspected to be responsible for HTG in the breed. If HTG risk variants are identified, screening would allow for early identification of susceptible dogs and might inform clinical care to prevent the development of HTG and its complications.

In humans, 42-54% of patients with severe primary HTG have rare variants in one of five major lipid metabolism genes: *lipoprotein lipase (LPL), apolipoprotein C-II (APOC2), apolipoprotein A-V (APOA5), glycosylphosphatidylinositol-anchored HDL-binding protein 1 (GPIHBP1)*, and *lipase maturation factor 1* (*LMF1*).[85,86] Previous sequencing of *LPL* and *APOC2* in Miniature Schnauzers with primary HTG did not reveal any variants.[12,13] However, the dog genome was not well annotated at the time of the *LPL* sequencing and not all exons were captured. Also, non-coding regions were not thoroughly evaluated for either gene; non-coding variants, such as core promoter regions, which can influence gene expression.[61] Therefore, re-evaluation of *LPL* and *APOC2* and analysis of additional major lipid metabolism genes are warranted in Miniature Schnauzers with primary HTG. Along with the five major susceptibility genes for HTG in humans, *apolipoprotein E (APOE)* is another potential candidate gene for HTG in Miniature Schnauzers. Rare variants in *APOE* cause HTG and lipoprotein

glomerulopathy in humans, and similar glomerular lesions occur in Miniature Schnauzers with HTG.[9,10]

Whole genome sequencing (WGS) offers a rapid and comprehensive alternative to conventional sequencing of multiple candidate genes.[62,63] The objective of this study was to use WGS to discover putative risk variants in the aforementioned six candidate genes for HTG in Miniature Schnauzers. We hypothesized that we would identify a putative causal variant associated with HTG in Miniature Schnauzers.

Materials and Methods

Miniature Schnauzers WGS Cohort

Eight Miniature Schnauzers with primary HTG were recruited for WGS. This sample size was chosen to capture a greater than 80% probability of detecting a variant present in at least 20% of the population based on basic probability of events $(1- (0.8)^n)$. This calculation underestimates the true probability of detecting a variant at this population frequency, as dogs are diploid. Primary HTG cases were dogs with fasting serum triglyceride concentrations >250 mg/dL and no clinical suspicion or previous diagnosis of a condition that can cause HTG (e.g., hypothyroidism, diabetes mellitus, hyperadrenocorticism). All primary HTG dogs were managed by veterinary board-certified internists and had blood samples submitted to the University of Minnesota Canine Genetics Laboratory for participation in hyperlipidemia research. Three dogs were recruited through an outside veterinary clinic after diagnosis with glomerular lipid thromboemboli through the International Veterinary Renal Pathology Service.[9] The other five dogs were recruited through the patient population at the University of Minnesota

Veterinary Medical Center. Informed owner consent was obtained, and the University of Minnesota Institutional Animal Care and Use Committee (Protocol #1509-33019A) approved the study. Four of the primary HTG dogs were previously included in an analysis of *APOC2* for coding variants.[12]

DNA Isolation and WGS

Ethylenediaminetetraacetic acid (EDTA) blood samples (2-4 mL) were obtained for genomic DNA isolation. Genomic DNA was isolated using a commercial kit (Gentra Puregene Blood Kit, Qiagen Sciences). WGS was performed using 150 base pair (bp) paired-end reads on an Illumina HiSeq 2500 with an average coverage of 18x. Quality control, mapping, and variant calling were performed using a previously described standardized pipeline[64] and the UU_Cfam_GSD_1.0/canFam4.0 dog reference assembly. Variants were annotated using Ensembl's Variant Effect Predictor (VEP).[65]

Candidate Gene Analysis

Variants present in two or more cases and located in the exons (including the 5' and 3' untranslated regions), intron-exon boundaries (within 20 bp of an exon), or within 150 bp upstream of the six candidate genes were extracted (Table 1).[66,67] Variant allele frequencies were determined for 613 dogs of 61 non-Miniature Schnauzer breeds and 30 Miniature Schnauzers with unknown HTG phenotypes in a private database of WGS variant calls (Table 2). Variant allele frequencies were calculated separately for the 8 Miniature Schnauzer cases with primary HTG, the 30 database Miniature Schnauzers with unknown phenotypes, and the 613 database dogs from 61 non-Miniature Schnauzer

breeds. Variants were filtered to identify those with a <0.10 allele frequency in 613 dogs of non-Miniature Schnauzer breeds; the Miniature Schnauzers with unknown phenotypes were not used for this step, given the high prevalence of HTG in the breed.

Variant locations were assessed for base-wise conservation using the "100 Vertebrates Basewise Conservation by phyloP (phyloP100way)" track on the UCSC Genome browser.[68] The absolute values of the scores from phyloP100way are the -log10(p-value) for rejecting the null hypothesis of neutral evolution. Conservation is indicated by positive scores, while negative scores indicate acceleration. The "Vertebrate Multiz Alignment & Conservation (100 Species)" track was used to determine the number of species the base position was conserved across.[69] Three variant pathogenicity prediction methods, SNPs&GO, PolyPhen2 HumDiv, and PolyPhen2 HumVar, were used to assess the pathogenicity of missense variants (see Chapter 5).[70,71] For both programs, scores >0.5 are considered pathogenic predictions. InterProScan was used to determine if variants resided in protein domains or other important sites.[72]

APOE Variant Follow-Up Genotyping

A variant of interest in *APOE* was selected for follow-up genotyping to determine if it was associated with primary HTG in the Miniature Schnauzer breed. Samples for the genotyping cohort were selected from Miniature Schnauzers from past and ongoing research projects with DNA biobanked (-80˚C) at the University of Minnesota Canine Genetics Laboratory.[9,22,23] Samples were genotyped for the *APOE* variant if they had a fasting serum triglyceride concentration available, regardless of age or sex. Dogs were excluded from the genotyping population if they had suspected or confirmed causes of

secondary HTG (e.g., diabetes mellitus, hypothyroidism, hyperadrenocorticism, or corticosteroid therapy).

NEBcutter V2.0 was used to identify differences in restriction enzyme sites for commercially available enzymes.[73] One enzyme, PsiI, was found to cut the reference sequence, but not the *APOE* variant. Primer3 was used to design primers to amplify a 710 base pair (bp) product encompassing the *APOE* variant: forward primer 5'-AGATGTCACCTCCCTTCGTG-3' and reverse primer 5'-GGAAGAGGATGCACGCAG-3'.[74] Standard polymerase chain reaction (PCR) amplification was performed with 35 cycles and a 60°C annealing temperature on a MJ Research PTC-100 thermal cycler. The PCR product was incubated overnight at 37°C with 1 unit of the PsiI enzyme. The PCR-RFLP (restriction fragment length polymorphism) assay products were resolved using gel electrophoresis. Dogs homozygous for the *APOE* variant had a single 383 bp product, dogs homozygous for the reference had a 201 bp product and a 181 bp product, and dogs heterozygous for the variant had all three products (383, 201, and 181 bp). Samples from one dog of each genotype were used as controls for the genotyping assay.

R statistical software was used for all statistical analyses. (R, version 4.1.2, www.r-project.org).[25] Data normality was evaluated using the Shapiro-Wilks test and quantile-quantile plot graphs (QQ-plot). A multivariable regression model was fit on the genotyping cohort with log-transformed triglyceride concentration as the dependent variable. In this model, the independent variable was the genotype (coded additively). Sex and age were included as covariates.

Results

Miniature Schnauzer WGS Cohort

The median fasting serum triglyceride concentration for the eight primary HTG dogs was 773 mg/dL (range 380-2,089 mg/dL). The median age was 11 years (range 6-13 years). There were three females and five males; all were spayed or neutered. The median fasting serum cholesterol concentration was 314 mg/dL (range 162-491 mg/dL). The three dogs with glomerular lipid thromboemboli were diagnosed by evaluation of renal biopsy specimens[9]; none were azotemic or hypoalbuminemic, but all were proteinuric with urine protein-to-creatinine ratios (UPC) of 3.0, 6.0, and 9.3. Two additional dogs without renal biopsies performed had proteinuria with UPCs of 2.0 and 2.2; these dogs were also neither azotemic nor hypoalbuminemic. The other three dogs did not have UPCs measured.

Candidate Gene Analysis

Twenty variants were recorded within the target regions of the six candidate genes in two or more Miniature Schnauzers with primary HTG. Eighteen variants were excluded due to presence at ≥0.10 allele frequency in the database of non-Miniature Schnauzer dogs. This left two variants: one was in *APOE* and the other in *GPIHBP1*. Variant details, including genotype and allele frequencies, are provided in Table 2.

The *APOE* variant was a three bp deletion in the 5' untranslated region. The *APOE* deletion resulted in a loss of the "TAT" of the TATA box, an important promoter sequence, previously described in mouse and human *APOE*.[75,76] The base-wise conservation score for one of these three nucleotides indicated significant conservation

(Table 2). The three deleted nucleotides are conserved in 57/67 vertebrate species (Table 5); the 10 species where the deleted nucleotides were not conserved included 3 bird, 3 reptile, 3 mammal, and 1 amphibian species. The *APOE* promoter deletion was found in a heterozygous state in two cases and was absent from the non-Miniature Schnauzer dogs in the WGS database. In the 30 Miniature Schnauzers with unknown phenotypes from the WGS database, the variant allele frequency was 0.30.

The *GPIHBP1* missense variant (p.S91L) was present in the Ly-6 domain, as determined by InterProScan (IPR016054, amino acids 87-162). The variant was predicted to have a neutral impact on the protein by SNPs&GO, PolyPhen2-HumDiv, and PolyPhen2-HumVar (0.14, 0.28, and 0.08, respectively) and was not conserved across vertebrate species. It was found in a heterozygous state in two Miniature Schnauzers with primary HTG. It was also present in a heterozygous state in one non-Miniature Schnauzer in the WGS database (a Dachshund), but it was not present in any of the 30 Miniature Schnauzers with unknown phenotypes.

APOE Variant Follow-Up Genotyping

One-hundred and fifty-nine samples were available from Miniature Schnauzers with fasting triglyceride concentrations, including the 8 dogs with primary HTG used for WGS. These dogs were genotyped for the *APOE* promoter deletion (g.111237170_111237172del). The median fasting serum triglyceride concentration for the genotyping cohort was 153 mg/dL (range 14-3975 mg/dL) with 91 primary HTG cases (triglyceride concentration >108 mg/dL) and 68 controls (≤108 mg/dL). The median age was 10 years (range 5-15 years). There were 73 females and 86 males.

Within the genotyping population, the *APOE* variant was present at an allele frequency of 0.52; 51 dogs were homozygous for the variant, 63 dogs were heterozygous, and 45 dogs were homozygous for the reference alleles. Results of the multivariable regression are shown in Table 3. There was no statistically significant effect of *APOE* genotype on triglyceride concentration in the multivariable regression analysis (Figure 1). Age was a predictor of triglyceride concentration, but sex was not.

Discussion

In this study, we used WGS variant call data to evaluate Miniature Schnauzers with primary HTG for putative risk variants in six candidate genes: *LPL, APOC2, APOA5, GPIHBP1, LMF1,* and *APOE*. Two variants met the filtering criteria: a 5'UTR deletion in *APOE* and a missense variant in *GPIHBP1*. The *APOE* deletion was considered the variant of greatest interest because it was unique to the Miniature Schnauzer breed, conserved in vertebrate species, and altered an important promoter sequence. However, no association between the *APOE* variant genotype and triglyceride concentration could be identified in a follow-up cohort of 159 Miniature Schnauzers with known HTG phenotypes. The *GPIHBP1* missense variant altered a poorly conserved base and was predicted to have a neutral impact on the protein; it was not evaluated further. The results of this study do not support a monogenic cause for HTG in the six candidate genes evaluated in Miniature Schnauzers.

Apolipoprotein E contributes to the metabolism of triglyceride-rich lipoproteins (very-low density lipoproteins and chylomicrons) by influencing lipoprotein lipolysis and controlling receptor-mediated clearance.[77] The *APOE* promoter variant was a deletion of

the “TAT” in the TATA box. The TATA box is a sequence in the core promoter region and is involved in the initiation of transcription in TATA-containing genes.[66] Genes containing TATA box promoter motifs are enriched in metabolism pathways, such as carbohydrate, amino acid, and lipid metabolism.[78] Variants in the TATA box generally regulate gene expression but can lead to a range of effects including alteration of the transcriptional start site or splicing defects.[79] Previously, a variant in the TATA box of apolipoprotein A-1 (*APOA1*) was associated with APOA1 deficiency and a resultant low concentration of high-density lipoproteins in a human.[80] It is possible that the *APOE* TATA box deletion impacts the metabolism of triglyceride-rich lipoproteins through one of these mechanisms. However, the variant was not associated with triglyceride concentration in follow-up genotyping of a large Miniature Schnauzer cohort. Thus, the *APOE* TATA box deletion does not appear to be solely responsible for primary HTG in Miniature Schnauzers. The apparent exclusivity of the variant to the Miniature Schnauzer breed is still of interest. It is possible the *APOE* variant is a modifier or contributor to the condition instead of a monogenic cause. Further experiments to confirm transcript levels, protein levels, or both could verify the impact of the TATA box deletion on *APOE* transcription.

Glycosylphosphatidylinositol-anchored HDL-binding protein 1 promotes the processing of triglyceride-rich lipoproteins and aids in the transport of LPL to the capillary lumen.[81] The *GPIHBP1* variant discovered in this study, a missense variant in the Ly-6 protein domain, altered a poorly conserved nucleotide and was predicted to have a neutral impact on the protein. For these reasons, follow-up genotyping of the *GPIHBP1* variant was not performed. However, while the variant pathogenicity prediction methods

used in this study have been shown to be accurate in classifying missense variants in animal species, classification accuracy is not perfect (Chapter 5). Therefore, we cannot rule out that the *GPIHBP1* variant contributes to primary HTG in some Miniature Schnauzers, and testing for an association with HTG might still be warranted.

There are a few possible explanations for why a monogenic cause for HTG was not discovered in these candidate genes. First, our analysis did not include an evaluation for structural variations (e.g., copy number variations, inversions, and translocations). Copy number variations in *LPL* and *GPIHBP1* have been identified in individuals with primary HTG and might contribute to disease severity.[82,83] A second possibility is that a putative causal variant for primary HTG in Miniature Schnauzers resides in a gene not included in this study. In humans, variants in the examined HTG candidate genes explain 42-54% of severe primary HTG cases.[85,86] Thus, approximately half of primary HTG cases are due to other susceptibility genes. Other genes for consideration include several of the angiopoietin-like proteins (*ANGPTL3, ANGPTL4, and ANGPTL8*), which inhibit LPL activity, apolipoprotein C-III (*APOC3*), which inhibits LPL activity by displacing LPL from triglyceride-rich lipoproteins, lipase C, and hepatic type (*LIPC*), which aids in the conversion of very-low density lipoproteins and intermediate-density lipoproteins to low-density lipoproteins.[11] Another, and perhaps the most likely, possibility is that primary HTG in Miniature Schnauzers is a polygenic or complex trait. In humans, HTG often develops from the cumulative effects of common and rare variation in multiple genes under the influence of the environment.[11,87,88] In dogs, triglyceride concentrations are influenced by age and diet, suggesting an environmental contribution to the

disease.[14,15,34,35] Furthermore, analysis of lipoprotein profiles from Miniature Schnauzers with primary HTG supports the possibility of multiple subtypes (Chapter 2).

In conclusion, a monogenic cause for primary HTG in Miniature Schnauzers could not be identified in the six candidate genes evaluated in this study. Though two variants passed filtering criteria, a deletion in the TATA box of *APOE* and a missense variant in *GPIHBP1*, neither had sufficient evidence to support a strong effect on primary HTG in Miniature Schnauzers due to a lack of association with serum triglyceride concentrations and a predicted neutral effect, respectively. However, neither can be ruled out as contributors to the disease. Given these findings and growing data on hyperlipidemia subtypes in Miniature Schnauzers, it is possible that HTG is a polygenic or complex trait in the breed.

Table 1. Six candidate genes previously associated with primary hypertriglyceridemia in people.

Gene Symbol	Gene Name	Function[77]
LPL	Lipoprotein lipase	hydrolyzes circulating triglycerides
APOC2	Apolipoprotein C-II	essential cofactor for LPL activity
APOA5	Apolipoprotein A-V	stimulates LPL activity
LMF1	Lipase maturation factor 1	essential for LPL enzymatic function
GPIHBP1	Glycosylphosphatidylinositol-anchored HDL binding protein 1	facilitates LPL transport to cell surface and promotes processing of triglyceride-rich lipoproteins
APOE	Apolipoprotein E	regulates clearance and lipolysis of triglyceride-rich lipoproteins

Table 2. Variants of interest identified by whole genome sequencing in eight Miniature Schnauzers with primary hypertriglyceridemia. Genomic locations are based on the UU-Cfam_GSD_1.0/canFam4 assembly.

Gene	**Description**	**Variant Type**	**PhyloP Score**	**Population**	***ref/ref***	***ref/var***	***var/var***	**AF**
APOE	g.111237170_111237172del	5' UTR	1.95, 0.83, 0.43	MS HTG	6	2	0	0.13
				Other MS#	14	14	2	0.3
				Non-MS Breeds^	613	0	0	0
GPIHBP1	g.37746233C>T	missense	-3.66	MS HTG	6	2	0	0.13
				Other MS#	30	0	0	0
				Non-MS Breeds^	599	1	0	0

^ Some dogs in population were not genotyped at this variant.

#Variant calls from 30 Miniature Schnauzers with unknown phenotypes in the WGS database.

Table 3. Multivariable regression model for the effects of age, sex, and *APOE* promoter deletion genotype on log transformed triglyceride concentrations in Miniature Schnauzers.

Variable	**Estimate of the coefficient**	**Std. Error**	***P* value**
APOE genotype	-0.03	0.13	0.81
Age (years)	0.13	0.05	**0.01**
Sex (male)	-0.07	0.21	0.73

P values in bold denote significance (<0.05).

Table 4. Breed counts from the private WGS database.

Breed	#
Akita	1
American Staffordshire Terrier	2
American Foxhound	1
Australian Cattle Dog	2
Australian Shepherd	3
Bichon Frise	3
Border Collie	8
Border Terrier	1
Boston Terrier	5
Bouvier des Flandres	10
Boxer	37
Boykin Spaniel	3
Brittany Spaniel	3
Bulldog	28
Bullmastiff	19
Cairn Terrier	7
Cavalier King Charles Spaniels	23
Collie	5
Coonhound	1
Corgi	8

Dachshund	15
Doberman Pinscher	8
English Bulldog	6
English Cocker Spaniel	1
English Mastiff	1
French Bulldog	21
German Shepherd	27
Golden Retriever	46
Goldendoodle	2
Great Dane	25
Great Pyrenees	3
Havanese	2
Irish Setter	3
Irish Wolfhound	20
Labradoodle	2
Labrador Retriever	22
Lhasa Apso	3
Miniature Poodle	10
Miniature Schnauzer	30
Mixed Breed	9
Newfoundland	15
Pomeranian	13
Portuguese Water Dog	11

Pug	4
Rhodesian Ridgeback	4
Rottweiler	17
Scottish Deerhound	10
Scottish Terrier	6
Sheltie	9
Shih Tzu	1
Shiloh Shepherd	9
Siberian Huskey	16
Spanish Greyhound	1
Spinoni Italiano	1
Standard Poodle	28
Toy Poodle	4
Wachtelhund	2
Welsh Springer Spaniel	4
Welsh Terrier	1
Whippet	15
West Highland White Terrier	4
Yorkshire Terrier	42

Table 2. Conservation of the 3 bp deletion (underlined) of the *APOE* TATA box sequence in 67 vertebrate species.

Human	TATA
Chimp	TATA
Gorilla	TATA
Gibbon	TATA
Rhesus	TATA
Crab-eating macaque	TATA
Baboon	TATA
Green monkey	TATA
Marmoset	TATA
Squirrel monkey	TATA
Bushbaby	TATA
Chinese tree shrew	TATA
Squirrel	TATA
Less Egyptian Jerboa	TATA
Prairie Vole	TATA
Chinese Hamster	TATA
Golden Hamster	TATA
Mouse	TATA
Rat	TATA
Naked Mole-rat	TATA

Guinea Pig	TATA
Chinchilla	TATA
Rabbit	TATA
Pika	TATA
Pig	TATA
Alpaca	TATA
Bactrian Camel	TATA
Dolphin	TATA
Killer Whale	TATA
Tibetan Antelope	TATA
Cow	TATA
Sheep	TATA
Domestic Goat	TATA
Cat	TATA
Dog	TATA
Ferret	TATA
Panda	TATA
Pacific Walrus	TATA
Weddell Seal	TATA
Black Flying-fox	TATA
Megabat	TATA
Big Brown Bat	TATA
David's Myotis Bat	TATA

Little Brown Bat	TATA
Hedgehog	TATA
Shrew	TATA
Star-nosed Mole	TATA
Elephant	TATA
Cape Elephant Shrew	TATA
Manatee	TATA
Cape Golden Mole	TATA
Tenrec	TATA
Aardvark	TATA
Armadillo	TATA
Tasmanian Devil	TATA
Wallaby	TATA
Brush-tailed rat	TATA
Zebra Finch	CAGA
Green Sea Turtle	CAGG
Chinese Softshell Turtle	CAGG
Scarlet Macaw	GAG-
Opossum	GGGA
American Alligator	GGGT
X. tropicalis	TCTA
Collared Flycatcher	TGGG
Horse	TTTA

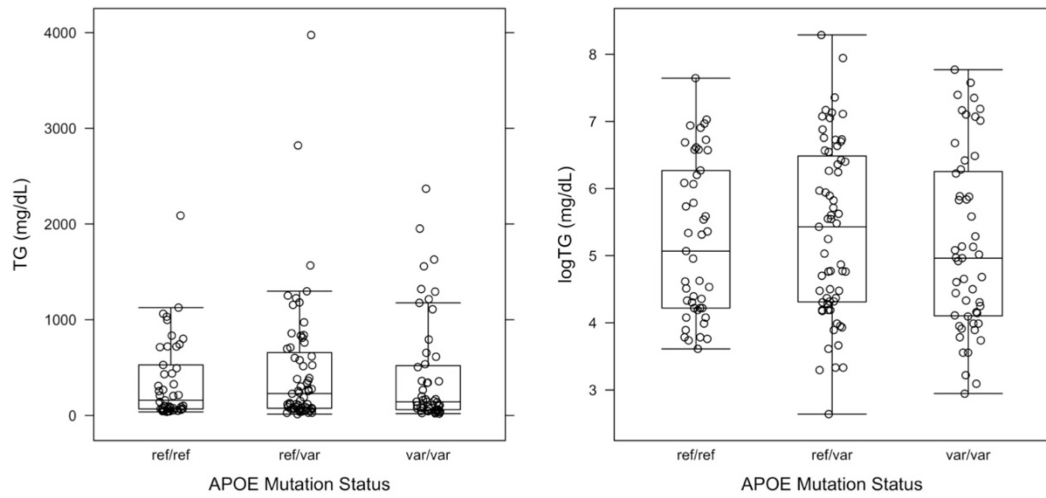


Figure 1. Range of A) triglyceride (TG) and B) log transformed triglyceride (logTG) concentrations for each *APOE* genotype. Ref/ref = homozygous for the reference sequence, ref/var = heterozygous for the *APOE* promoter variant, and var/var = homozygous for the *APOE* promoter variant.

CHAPTER 5

Performance of Variant Pathogenicity Prediction Methods Utilizing Canine and Equine Missense Variants

Abstract

Whole genome and exome sequencing has become a routine method for discovery of causal variants for hereditary disease. A key challenge in sequencing studies is variant effect prediction. Many *in silico* tools exist to predict variant pathogenicity but are not validated for analysis of veterinary datasets. This study evaluates the performance of eight pathogenicity prediction methods (MutPred2, PANTHER, PhD-SNP, PolyPhen2-HumDiv, PolyPhen2-HumVar, Provean, SIFT, and SNPs&GO) for missense variants in the dog and horse. Performance was evaluated using a collection of 169 trait/disorder variants (dog=127, horse=42) and 373 benign variants (dog=216, horse=157). For each prediction program, the prediction rate, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, F-score, and Matthews correlation coefficient (MCC) were calculated. The tool with the poorest performance within and across species was PANTHER, with the lowest specificity, accuracy, and MCC, and the highest number of unclassified variants (lowest prediction rate). Overall, the remaining tools performed comparably within and across both species, with sensitivities >74%, specificities >79%, and MCCs >0.5. However, consensus across all programs was uncommon. Forty percent of trait/disorder variants and 38% of benign variants were incorrectly classified by at least two methods. These results further inform the usage of *in silico* variant pathogenicity prediction tools in veterinary species. While several programs

perform well for predicting the pathogenicity of missense variants in dogs and horses, it is common to have disagreement between programs.

Introduction

Increased availability and decreased cost have made whole genome and exome sequencing a routine method for discovery of causal variants for hereditary disease in humans and other species.[63,89,90] This high-throughput technology allows for the detection of thousands of variants per individual.[89] However, filtering thousands of variants to identify those most likely to cause disease is challenging. First pass variant prioritization is often based on the effect of the variant on the coding sequence.[91] Yet, the impact of missense variants, the most common type of non-synonymous coding variant, is highly variable.[91,92] Some amino acid substitutions, such as those at critical sites in the protein sequence, affect protein conformation, stability, and protein-protein interactions, while others have little to no effect on protein function.[92]

Functional studies can provide strong evidence of variant effect and pathogenicity but are not always available or practical. An alternative, albeit weaker, source of evidence comes from computational pathogenicity prediction by *in silico* tools.[91,93] Broadly, these programs use features such as amino acid side chain properties, amino acid conservation, protein structure, protein domains, and splice sites to determine if a variant is likely

pathogenic.[70,71,94-98] The available tools differ in prediction algorithm(s), methods used to develop the algorithm, and datasets used for training and validation.

Variant pathogenicity prediction programs are typically trained using human variants, and many are unable to analyze variants from non-human genomes. Some tools can be adapted to incorporate information from other species.[70,99-102] However, performance between species can differ[103], and most programs have not been validated for non-human species.

Two veterinary species, the dog and horse, have gained recognition as spontaneous models for hereditary human disorders. The relatively low within-breed genetic diversity in these species is advantageous for discovery of disease-causing variants.[63,104,105] The aim of this study was to evaluate the performance of eight tools for pathogenicity prediction of missense variants in the dog and horse. We hypothesized that one or more of these programs, alone or in combination, would provide a reliable method for variant prioritization in these species.

Materials and Methods

Variant Datasets

For each species, a trait/disorder dataset of missense variants (Supplementary Tables 3 and 4) was built from the Online Mendelian Inheritance in Animals (OMIA,

omia.org) database by searching for entries with "*Canis lupus familiaris"* or "*Equus caballus*" as the species scientific name and "missense" as the molecular genetics term. Additional variants were identified through a PubMed search for the species common name ("dog" or "horse") with "missense mutation" and "AND missense" (example: "dog missense mutation", "horse AND missense"). An initial search was completed on September 14, 2018; the search was repeated on October 7, 2021 with the goal of identifying newly discovered variants. All variants described as causal for a trait or disorder were included, even if qualifier terms, such as "likely" and "probably," were used. Variants were excluded if the relationship with the disease was described as "unknown" or "undetermined." A benign dataset (Supplementary Tables 5 and 6) of missense variants was assembled for each species using variant calls from an in-house database of 250 dogs from 44 breeds and a publicly available database of 534 horses from 44 breeds.[90] Variant calls were pruned for missense variants in protein coding genes with an allele frequency between 0.49 and 0.51. This range was selected because variants in approximately 50% of the population are unlikely to cause disease and allowed us to capture up to 4x the number of variants in the disease dataset. Variants were excluded from the benign dataset if they did not have a homologous human protein-coding gene; all variants within the disease dataset had a human homolog.

Pathogenicity Prediction Methods

We evaluated eight variant pathogenicity prediction methods: MutPred2, PANTHER, PhD-SNP, PolyPhen2-HumDiv, PolyPhen2-HumVar, Provean, SIFT, and SNPs&GO.[70,71,94-98] Methodology, development, and additional attributes for each method are summarized in Table 1. These methods were chosen for their 1) ability to analyze data from any species, 2) performance with human data, and 3) public availability.[106-109]

Input data was submitted and analyzed through the on-line platforms. Where applicable, default parameters were applied for each variant in the dataset. PANTHER and PhD-SNP were run through SNPs&GO using the "all methods" option. For genes where GO terms were unavailable, SNPs&GO was run without. SIFT scores were obtained through Ensembl's Variant Effect Predictor (VEP).[65] The VEP allows for the analysis of data from any species with an assembled genome and provides SIFT scores for 14 species, including dog and horse (additional species can be found here: https://www.ensembl.org/info/genome/variation/prediction/protein_function.html#sift). If SIFT scores were not returned by the VEP, SIFT scores were obtained through the online SIFT platform (https://sift.bii.a-star.edu.sg). During this study, new canine genomes were released, and VEP no longer accepted genomic positions from canFam3.1. Thus,

genomic positions from Dog10K_Boxer_Tasha/canFam6 were used to obtain VEP results after December 2021.

Data Gathering

Amino acid sequences were obtained from Ensembl through the Bioconductor interface, biomaRt.[110,111] In the case of multiple transcripts, the longest transcript was chosen for the analysis.[112,113] If the amino acid was not present at the noted position in the longest transcript, we identified the transcript that correlated to the position by inputting the genomic position into VEP. This allowed us to review protein positions for all transcripts and select the appropriate one. Amino acid sequences that were not available through Ensembl were obtained from the National Center for Biotechnology Institute database (NCBI, https://www.ncbi.nlm.nih.gov).[114]

One tool evaluated in this study, SNPs&GO, allows for the input of Gene Ontology (GO) terms. The Bioconductor interface, biomaRt, was used to obtain GO terms. A subset of 50 canine missense variants (n=25 trait/disorder, n=25 benign) was used to determine if the choice of GO database had a significant effect on the performance of SNPs&GO (Supplementary Table 3). Differences in performance using no GO terms, GO terms pulled from the canine database, and GO terms pulled from the human database, were compared using a Fisher's exact test. Statistical analysis was performed using R statistical software (R core team 2021, https://www.r-project.org).[25]

Classification

The pathogenicity prediction result was considered a true positive (TP) if a trait/disorder missense variant was classified as pathogenic by the program. A true negative (TN) was a benign missense variant classified by the program as benign. Alternatively, a false positive (FP) was classification of a benign variant as pathogenic, and a false negative (FN) was classification of a trait/disorder variant as benign.

Performance Evaluation

The performance for each method was evaluated using eight parameters: classification rate, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, F-score, and Matthews Correlation Coefficient (MCC) (equations below). For all parameters other than F-score and MCC, a Wilson's 95% confidence interval (CI) was calculated. The accuracy of each variant pathogenicity prediction method was compared between missense variants that result in a trait and those that result in disease using a Fisher's exact test.

$$Classification\ Rate = \frac{\#\ variants - \#\ unclassified}{\#\ of\ variants}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{FP + TN}$$

$$Positive\ Predictive\ Value\ (PPV) = \frac{TP}{TP + FP}$$

$$Negative\ Predictive\ Value\ (NPV) = \frac{TN}{TN + FN}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$F - Score = 2\frac{PPV\ x\ Sensitivity}{PPV + Sensitivity}$$

$$Matthews\ Correlation\ Coefficient\ (MCC)$$

$$= \frac{TP\ x\ TN - FP\ x\ FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Results

Variant Datasets

179 trait/disorder variants (131 canine and 48 equine) were obtained from the Online Mendelian Inheritance in Animals (OMIA, omia.org database and PubMed searches, and 375 benign variants (216 canine and 159 equine) were extracted from the variant call database searches. Twelve variants (10 trait/disorder and two benign) were excluded from the datasets (Supplementary Table 3-6). Three variants in *TTN* (one dog trait/disorder and two horse benign) were excluded because of missing results from some tools due to the large size of the protein sequences. One variant in *RYR1* (canine trait/disorder) was excluded because of missing results from one of the tools (SIFT).

Three variants (one equine and two canine trait/disorder) were excluded because they had been removed from OMIA between 2018 and 2021 due to updated evidence that challenged their pathogenicity.[115-117] Two variants (equine trait/disorder) were excluded because they resided in an incompletely annotated protein (SLC45A2). The final three variants were excluded because they were in a non-target species (trait/disease in the donkey, *Equus asinus*).

The final list of variants included in the study are provided in Supplementary Tables 3-6. The canine trait/disorder dataset included 127 missense variants that resulted in disease (n=115) and non-disease traits (i.e., coat color, coat length, fecundity; n=12) (Supplementary Table 3). The canine benign dataset consisted of 216 missense variants (Supplementary Table 5). The equine trait/disorder dataset included 42 missense variants that resulted in disease (n=17) and non-disease traits (i.e., coat color, eye color, environmental adaptations; n=25) (Supplementary Table 4). The equine benign dataset comprised 157 missense variants (Supplementary Table 6).

Use of GO Terms

SNPs&GO algorithm uses gene ontology (GO) terms to improve accuracy. Therefore, we compared sources for GO terms before running this analysis. The canine GO term database was more likely to have missing GO terms for genes than the human GO term database ($p = 2.4 \times 10^{-15}$). For the 274 genes in our canine dataset, 77 had no GO terms using the canine GO database, compared to 9 with no GO terms using the human database. To determine if the use of GO terms impacted SNPs&GO prediction outcomes, a subset of 50 canine variants were evaluated (25 benign and 25 trait/disorder). Using

human GO terms had a greater sensitivity than not using GO terms (p = 0.0096, OR = 5.4, 95% CI 1.4-23.3). There was no significant difference in specificity between using human GO terms and not using GO terms (p = 0.24, OR = 0.3, 95% CI 0.03-2.3). There was also no difference in sensitivity (p = 0.35, OR = 0.5, 95% CI 0.1-2) or specificity (p = 0.7, OR = 0.7, 95% CI 0.1-4) between using human versus canine GO terms. Therefore, human GO terms were used for SNPs&GO predictions in both canine and equine datasets.

Performance of Pathogenicity Prediction Methods

Results for eight performance parameters (prediction rate, sensitivity, specificity, PPV, NPV, accuracy, F-score, and MCC) for each pathogenicity prediction method are presented in Table 2.

In the combined dataset (canine and equine), sensitivity ranged from 74-91%, with PANTHER being the lowest and PolyPhen2-HumDiv being the highest. Specificity ranged from 72-86%, with PANTHER being the lowest and SNPs&GO the highest. The PPV ranged from 60-74%, with PANTHER being the lowest and SNPs&GO the highest. The NPV ranged from 83-95%, with PANTHER being the lowest and PolyPhen2-HumDiv the highest. Accuracy ranged from 73-86%, with PANTHER being the lowest and SNPs&GO and PolyPhen2-HumVar being the highest. The F-score ranged from 0.67-0.80, with PANTHER being the lowest and PolyPhen2-HumVar being the highest. Lastly, the MCC ranged from 0.45-0.70, with PANTHER being the lowest and SNPs&GO and PolyPhen2-HumVar being the highest.

When evaluating the canine and equine datasets separately, all methods had lower PPV for equine variants than canine variants. The PPV ranged from 65-78% for canine variants and 52-68% for equine variants. However, the 95% CI for the PPV overlapped between the two species for all methods with the exception of MutPred2. In dogs, the PPV for MutPred2 was 77% (95% CI 69% - 84%), and in horses it was 52% (95% CI 40% - 63%). Additionally, dogs had a lower NPV (75%, 95% CI 67% - 82%) for PANTHER than horses (93%, 95% CI 86% - 97%). For all programs, the F-score was lower for equine variants, and MCC was lower for 5 of 8 programs.

Five of the eight methods, including MutPred2, SIFT, SNPs&GO, PhD-SNP, and Provean, classified 100% of the variants. Both versions of PolyPhen2 provided classifications for 94% of the variants. In contrast, PANTHER only provided classifications for 69% of the variants with the others deemed “unclassified.”

In the combined dataset, complete concordance across all methods was uncommon (Table 3). Forty percent of the trait/disorder variants and 38% of the benign variants were incorrectly classified by at least two methods. Nine trait/disorder variants (seven canine and two equine) were classified as neutral by at least six methods; this included four coat color and five disease variants. (Supplementary Table 3 and 4). Twenty-six variants in the benign dataset (15 canine and 11 equine) were classified as likely pathogenic by six or more methods (Supplementary Table 5 and 6). For each method, there was no difference in accuracy for missense variants causal for a trait versus those that result in a disorder (Table 4).

Discussion

In this study, we evaluated the performance of eight methods to predict the pathogenicity of missense variants in the dog and horse. The programs evaluated in this study were selected based on their adaptability for non-human data. Seven of the methods tested performed well within and across both species with accuracies of 80% or greater and MCCs greater than 0.5. The exception was PANTHER, which had the lowest accuracy and MCC, and a high proportion of unclassified variants. Complete consensus across all programs was uncommon, with 40% of trait/disorder variants and 38% of benign variants incorrectly classified by at least two methods.

Seven of the methods (SNPs&GO, MutPred2, Provean, PhD-SNP, SIFT, PolyPhen2-HumDiv, and PolyPhen2-HumVar) performed similarly with sensitivities ranging from 76% to 91% and specificities from 79% to 86% in the combined dataset. SNPs&GO and PolyPhen2-HumVar had the highest MCC (0.7) and accuracy (86%) in the combined dataset. SNPs&GO was the most specific program (86%), while PolyPhen2-HumDiv was the most sensitive (91%). The accuracy of all 8 methods did not differ between variants impacting traits, such as performance or coat color, compared with those that result in disease. The high performance of these programs is similar to what has been reported for human variant data.[106-109] For comparison, in humans, SNPs&GO is reported to have specificities between 92% and 95%, and PolyPhen2-HumVar is reported to have sensitivities between 83% and 87%.[106-109]

The performance of SNPs&GO was evaluated using human GO terms for classification. This decision was based on analyses in a subset of canine missense variants, where we compared performance without GO terms, with GO terms obtained

from the canine database, and with GO terms obtained from the human database. Human GO terms had a greater sensitivity than using no GO terms. This is consistent with previous findings that inclusion of GO terms improves the accuracy and MCC of the SNPs&GO method.[71] While there was no difference detected in performance between human and canine GO terms, the human database was selected because it was less likely to be missing GO terms for genes. Also, the horse does not currently have a species-specific GO term data resource, requiring equine variants to be analyzed using the human GO database as well.

Panther was the overall lowest performing program with the lowest specificity, accuracy, and MCC and the highest number of unclassified variants in all datasets. PANTHER uses evolutionarily related protein sequences to estimate the probability of an amino acid to occur at a specified position and thus, the likelihood of a variant to cause a functional effect.[95] PANTHER has variable performance using human datasets. Some studies report a high specificity, accuracy, and MCC[106,107,109]; whereas others, report poorer performance compared to other variant effect prediction programs.[118,119] The reasons for PANTHER's low performance in our datasets and its variable performance with human datasets is unknown.

PANTHER also had the lowest prediction rate in this study with only 69% of variants classified by this method compared to prediction rates of 94-100% for the other methods. This low prediction rate is similar to that reported by PANTHER's developers (76%) for variants from the Human Gene Mutation Database.[95] A low prediction rate (52-68%) was also reported for PANTHER in performance evaluations using human variant data from the Human Gene Mutation Database, recently published variants in *Nature*

*Genetic*s, and the Comparative High-Throughput Analyses of Resistance Gene Evolution database.[108,118] Frousios et al. attribute the low prediction rate to variants residing in protein positions that are not covered by the multiple sequence alignment in the PANTHER library.[108] It is possible that this also explains the large number of unclassified variants in our dataset.

Although none of the programs tested were designed specifically for dogs or horses, our findings suggest that they can be effectively used in these species. We did not identify a single method that scored the highest across all parameters. Therefore, we recommend investigators using these tools carefully consider what criteria are most valuable to the goal of their specific study. For example, if the aim is to maximize detection of all possibly pathogenic variants, the two PolyPhen2 methods are recommended due to their high sensitivity within and across both species. In contrast, if the priority is to minimize false positive calls, SNPs&GO had the highest specificity. SNPs&GO and PolyPhen2-HumVar have high MCC and are suggested for an overall accurate method for variant classification; MCC is a measure of classification accuracy for binary classifiers and is less susceptible to imbalance issues within a dataset.[120] Since our dataset was limited to canine and equine variants, we cannot infer how these programs would perform analyzing data from other species. However, SNPs&GO and PolyPhen2-HumVar, similarly have high MCC scores in humans.[106,107,109] Therefore, we suggest these for species in which variant pathogenicity prediction methods have not yet been evaluated.

The methods in this study were publicly available and did not require adaptation for non-human species. Some variant pathogenicity prediction methods, such as

PolyPhen2 and CADD, can be modified or adapted to use species-specific information.[70,99-102] PolyPhen2 has species-specific information built-in for orangutan, mouse, rat, dog, zebrafish, and fruitfly and can be adapted for other species through the use of the species genome, annotation, and species-specific protein sequences.[70] CADD uses the species genome, annotation, species-specific protein sequences, and uses species-specific gene ontology terms.[99-101] Previously, CADD models have been adapted and trained for use in mice, pigs, and chickens.[99-101] CADD was not used in this study as, unless adapted for a specific species, it requires the input of the homologous position in the human genome for dog or horse. It is possible that methods adapted to incorporate species-specific information would perform better than the methods evaluated in this study. Fido-SNP, a canine-specific method similar to PhD-SNP, was not included in this study because the training set used in the development of Fido-SNP included variants present in our trait/disorder dataset.[121] Thus, using variants that were part of the Fido-SNP training set would have created a circularity problem and biased results towards this tool. Future research is needed to determine if species-specific methods are more effective in classifying/prioritizing variants in dogs or horses.

Forty percent of the trait/disorder variants and 38% of the benign variants in this study were incorrectly classified by at least two programs. Trait/disorder variants were obtained from the OMIA database but were not rigorously assessed to determine if they met AMCG Standards and Guidelines for pathogenicity.[93] It is possible that some variants in our trait/disorder dataset are not the true causal variant. In fact, three variants were removed from OMIA between the 2018 and 2021 searches because new data challenged their pathogenicity.[115-117] However, the low concordance across all eight

programs argues against the ACMG Standards and Guidelines recommendation that *in silico* programs should not be used as supporting evidence unless all programs agree.[93] This recommendation has also been challenged by other authors, indicating the requirement may be too strict and could result in decreased overall performance of the guidelines.[122-126] Regardless, it is important to interpret classifications from variant pathogenicity prediction methods carefully and not use these classifications as the sole source of evidence.[127]

The benign datasets were assembled from in-house databases by pruning for variants in protein coding genes with an allele frequency between 0.49 and 0.51. While expected to be infrequent, it is possible that a variant present in roughly 50% of the population could be functional and confer a desired phenotype or disorder. For example, variants that influence coat length, curl, and furnishings are common in dogs and might be present at the 50% cut off if half of the sample population has these traits.[128]

In this study, performance was evaluated using variants reported to contribute to Mendelian traits and disorders. Pathogenicity prediction programs might not perform as well when applied to the analysis of variants contributing to complex traits. For example, variants associated with complex disease might occur at less conserved amino acid residues[129] and are often in non-coding regions of the genome.[130] PolyPhen2 recommends HumDiv, one of its two outputs, for the evaluation of rare alleles at loci potentially involved in complex traits [10]. Other methods, like Oligogenic Resource for Variant AnaLysis (utilizes VarCoPP and Digenic Effect Predictor), assess the pathogenicity of combinations of variants in two genes.[70] Additionally, methods such as FATHMM, FunSeq2, LINSIGHT, and ORION, have been developed to assess the pathogenicity of

non-coding variants.[131-134] The optimal approach for identifying coding or non-coding small effect-size variants in dogs and horses remains to be determined.

Overall, we evaluated 542 variants from 394 genes. Previous studies have shown variability in the performance of pathogenicity prediction methods between genes.[124,125] We did not balance the number of variants per gene across the datasets, and *KIT* variants were 29% (12/42) of the equine trait/disorder dataset. Another imbalance was the presence of 21 variants in olfactory receptor genes in the canine benign dataset and 30 in the equine benign dataset, but none in either species trait/disorder dataset. The olfactory receptor gene family is the largest gene family found in mammalian genomes.[135] The composition of the genes included in our datasets might have impacted performance results.

In conclusion, the results of this study inform the usage of *in silico* variant pathogenicity prediction tools in dogs and horses by showing high accuracies of multiple methods developed for human datasets. The lower performance and higher rate of unclassified variants by PANTHER suggest that this method is not optimal for canine or equine data. Of the other methods, none was rated highest across all parameters, though SNPs&GO and PolyPhen2-HumVar had the highest MCC and accuracy. Ultimately, users should consider what aspects are most valuable to their study. For example, to maximize sensitivity, one might utilize more than one tool with high sensitivity and investigate any variant with predicted pathogenicity from either program. In contrast, if specificity is desired, one might utilize more than one tool with high specificity and only investigate variants with predicted pathogenicity in both programs. Regardless, it is

important to interpret classifications from variant pathogenicity prediction methods carefully and not use these classifications as the sole source of evidence.

Table 1. Summary of the eight variant pathogenicity prediction tools evaluated.

Tool	Method	Training Data	Conservation Analysis	Structural Attributes	Annotations	Score Threshold for Pathogenicity	Website
MutPred2[94]	Random Forest	Human Gene Mutation Database	SIFT, Pfam, PSI-BLAST	Yes	No	> 0.5	http://mutpred2.mutdb.org
PANTHER[95]	Alignment Scores	-*	PANTHER Library, Hidden Markov Model	No	No	> 0.5	http://pantherdb.org/tools/csnpScoreForm.jsp?
PhD-SNP[96]	Support Vector Machine	Swiss-Prot	Sequence Environment, Sequence Profiles	No	No	> 0.5	https://snps.biofold.org/phd-snp/phd-snp.html
PolyPhen2-HumDiv[70]	Bayesian classification	UniProtKB	PSIC profiles	Yes	No	> 0.5	http://genetics.bwh.harvard.edu/pph2/
PolyPhen2-HumVar[70]	Bayesian classification	UniProtKB	PSIC profiles	Yes	No	> 0.5	http://genetics.bwh.harvard.edu/pph2/
Provean[97]	Alignment Scores	UniProt	Delta Alignment Score	No	No	≤ -2.5	http://provean.jcvi.org/index.php

SIFT[98]	Alignment scores	-	Multiple Sequence Alignment	No	No	≤ 0.05	https://www.ensembl.org/Tools/VEP https://sift.bii.a-star.edu.sg/www/SIFT_seq_submit2.html
SNPs&GO[71]	Support Vector Machine	Swiss-Prot	Sequence Environment, Sequence Profiles, PANTHER library	No	Gene Ontology	> 0.5	https://snps.biofold.org/snps-and-go//snps-and-go.html

Modified from Thusberg et al., 2011.[106]

*Training data was not reported, however PANTHER was tested using data from the Human Gene Mutation Database and dbSNP.[95]

Table 2. Performance of 8 variant pathogenicity prediction tools for classification of canine and equine missense variants.

	MutPred2	SIFT	SNPs&GO	Panther	PhD-SNP	Provean	PolyPhen2-HumDiv	PolyPhen2-HumVar
Performance of variant pathogenicity tools (combined)								
tp	128	142	145	102	131	134	150	144
fn	41	27	24	35	38	35	14	20
tn	313	299	322	170	312	318	273	292
fp	60	74	51	67	61	55	72	53
unclassified	0	0	0	168	0	0	33	33
Classification Rate	1.00	1.00	1.00	0.69	1.00	1.00	0.94	0.94
Sensitivity (95% CI)	0.76 (0.69, 0.82)	0.84 (0.78, 0.89)	0.86 (0.80, 0.90)	0.74 (0.67, 0.81)	0.78 (0.71, 0.83)	0.79 (0.73, 0.85)	0.91 (0.86, 0.95)	0.88 (0.82, 0.92)
Specificity (95% CI)	0.84 (0.80, 0.87)	0.80 (0.76, 0.84)	0.86 (0.82, 0.89)	0.72 (0.66, 0.77)	0.84 (0.80, 0.87)	0.85 (0.81, 0.88)	0.79 (0.75, 0.83)	0.85 (0.80, 0.88)
PPV (95% CI)	0.68 (0.61, 0.74)	0.66 (0.59, 0.72)	0.74 (0.67, 0.80)	0.60 (0.53, 0.67)	0.68 (0.61, 0.74)	0.71 (0.64, 0.77)	0.68 (0.61, 0.73)	0.73 (0.67, 0.79)
NPV (95% CI)	0.88 (0.85, 0.91)	0.92 (0.88, 0.94)	0.93 (0.90, 0.95)	0.83 (0.77, 0.87)	0.89 (0.85, 0.92)	0.9 (0.87, 0.93)	0.95 (0.92, 0.97)	0.94 (0.90, 0.96)
Accuracy (95% CI)	0.81 (0.78, 0.84)	0.81 (0.78, 0.84)	0.86 (0.83, 0.89)	0.73 (0.68, 0.77)	0.82 (0.78, 0.85)	0.83 (0.80, 0.86)	0.83 (0.80, 0.86)	0.86 (0.82, 0.88)
F-score	0.72	0.74	0.79	0.67	0.73	0.75	0.78	0.80
MCC	0.58	0.61	0.70	0.45	0.59	0.63	0.67	0.70

Table 2. Performance of 8 variant pathogenicity prediction tools for classification of canine and equine missense variants.

Performance of variant pathogenicity tools (dog)								
tp	94	105	110	71	98	96	112	106
fn	33	22	17	30	29	31	11	17
tn	188	174	185	90	182	186	154	161
fp	28	42	31	39	34	30	41	34
unclassified	0	0	0	113	0	0	25	25
Classification Rate	1.00	1.00	1.00	0.67	1.00	1.00	0.93	0.93
Sensitivity (95% CI)	0.74 (0.66, 0.81)	0.83 (0.75, 0.88)	0.87 (0.80, 0.91)	0.70 (0.61, 0.78)	0.77 (0.69, 0.84)	0.76 (0.67, 0.82)	0.91 (0.85, 0.95)	0.86 (0.79, 0.91)
Specificity (95% CI)	0.87 (0.82, 0.91)	0.81 (0.75, 0.85)	0.86 (0.80, 0.90)	0.70 (0.61, 0.77)	0.84 (0.79, 0.89)	0.86 (0.81, 0.90)	0.79 (0.73, 0.84)	0.83 (0.77, 0.87)
PPV (95% CI)	0.77 (0.69, 0.84)	0.71 (0.64, 0.78)	0.78 (0.71, 0.84)	0.65 (0.55, 0.73)	0.74 (0.66, 0.81)	0.76 (0.68, 0.83)	0.73 (0.66, 0.80)	0.76 (0.68, 0.82)
NPV (95% CI)	0.85 (0.80, 0.89)	0.89 (0.84, 0.92)	0.92 (0.87, 0.95)	0.75 (0.67, 0.82)	0.86 (0.81, 0.90)	0.86 (0.80, 0.90)	0.93 (0.88, 0.96)	0.90 (0.85, 0.94)
Accuracy (95% CI)	0.82 (0.78, 0.86)	0.81 (0.77, 0.85)	0.86 (0.82, 0.89)	0.70 (0.64, 0.76)	0.81 (0.77, 0.85)	0.82 (0.78, 0.86)	0.84 (0.79, 0.87)	0.84 (0.80, 0.88)
F-score	0.76	0.77	0.82	0.67	0.76	0.76	0.81	0.81
MCC	0.62	0.62	0.71	0.40	0.61	0.62	0.68	0.67

Table 2. Performance of 8 variant pathogenicity prediction tools for classification of canine and equine missense variants.

Performance of variant pathogenicity tools (horse)								
tp	34	37	35	31	33	38	38	38
fn	8	5	7	5	9	4	3	3
tn	125	125	137	80	130	132	119	131
fp	32	32	20	28	27	25	31	19
unclassified	0	0	0	55	0	0	8	8
Classification Rate	1.00	1.00	1.00	0.72	1.00	1.00	0.96	0.96
Sensitivity (95% CI)	0.79 (0.65, 0.89)	0.86 (0.73, 0.93)	0.81 (0.67, 0.90)	0.84 (0.69, 0.92)	0.77 (0.62, 0.87)	0.88 (0.75, 0.95)	0.90 (0.78, 0.96)	0.90 (0.78, 0.96)
Specificity (95% CI)	0.80 (0.73, 0.85)	0.80 (0.73, 0.85)	0.87 (0.81, 0.92)	0.74 (0.65, 0.81)	0.83 (0.76, 0.88)	0.84 (0.78, 0.89)	0.79 (0.72, 0.85)	0.87 (0.81, 0.92)
PPV (95% CI)	0.52 (0.40, 0.63)	0.54 (0.42, 0.65)	0.64 (0.50, 0.75)	0.53 (0.40, 0.65)	0.55 (0.42, 0.67)	0.60 (0.47, 0.71)	0.55 (0.43, 0.66)	0.67 (0.54, 0.78)
NPV (95% CI)	0.93 (0.88, 0.96)	0.95 (0.90, 0.98)	0.94 (0.89, 0.97)	0.93 (0.86, 0.97)	0.93 (0.87, 0.96)	0.96 (0.92, 0.98)	0.97 (0.92, 0.99)	0.97 (0.93, 0.99)
Accuracy (95% CI)	0.80 (0.73, 0.85)	0.81 (0.75, 0.86)	0.86 (0.81, 0.90)	0.77 (0.69, 0.83)	0.81 (0.76, 0.86)	0.85 (0.79, 0.89)	0.82 (0.76, 0.87)	0.88 (0.83, 0.92)
F-score	0.62	0.66	0.71	0.65	0.64	0.71	0.68	0.77
MCC	0.51	0.57	0.63	0.51	0.53	0.64	0.60	0.70

tp = true positive; fn = false negative; tn = true negative; fp = false positive; MCC = Matthews correlation coefficient

Table 3. Number of variant pathogenicity prediction methods in concordance for classifying variants correctly.

# of Programs:	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Trait/Disorder Variants									
Dog, n = 127	4 (3%)	1 (<1%)	4 (3%)	8 (6%)	8 (6%)	12 (9%)	13 (10%)	27 (21%)	50 (39%)
Horse, n = 42	2 (5%)	1 (2%)	1 (2%)	1 (2%)	2 (5%)	4 (10%)	6 (14%)	25 (60%)	0 (0%)
Combined, n = 169	6 (4%)	2 (1%)	5 (3%)	9 (5%)	10 (6%)	16 (9%)	19 (11%)	52 (31%)	50 (29%)
Benign Variants									
Dog, n = 216	6 (3%)	6 (3%)	4 (2%)	10 (5%)	8 (4%)	24 (11%)	32 (15%)	76 (35%)	50 (23%)
Horse, n = 157	9 (6%)	0 (0%)	3 (2%)	6 (4%)	9 (6%)	16 (10%)	9 (6%)	55 (35%)	50 (32%)
Combined, n = 373	15 (4%)	6 (2%)	7 (2%)	16 (4%)	17 (5%)	40 (11%)	41 (11%)	131 (35%)	100 (27%)

Table 4. Comparison of method accuracy between missense variants that result in a trait versus those which result in disease.

	MutPred2	SIFT	SNPs&GO	PANTHER	PhD-SNP	Provean	PolyPhen2-HumDiv	PolyPhen2 HumVar
Disease								
TP	97	112	113	75	102	104	120	115
Total	132	132	132	103	132	132	130	130
Accuracy	73%	85%	86%	73%	77%	79%	92%	88%
Trait								
TP	31	30	32	27	29	30	30	29
Total	37	37	37	34	37	37	34	34
Accuracy	84%	81%	86%	79%	78%	81%	88%	85%
p-value	0.68	0.89	1	0.77	1	1	0.89	1

CHAPTER 6

Conclusions and Future Directions

Disease and Overall Objective

Hyperlipidemia affects more than 30% of Miniature Schnauzers, with prevalence and severity increasing with age.[4] In this breed, hyperlipidemia is characterized by hypertriglyceridemia (HTG), with or without hypercholesterolemia. The hypertriglyceridemia is characterized by increases in triglyceride-rich lipoproteins (very-low density lipoproteins (VLDL) or a combination of chylomicrons and VLDL).[2,3] Hypertriglyceridemia predisposes Miniature Schnauzers to severe consequences, such as pancreatitis, gallbladder mucoceles, glomerular proteinuria, and other conditions.[4,6-10] However, the precise molecular mechanisms and cause of HTG in Miniature Schnauzers remains unclear. It is suspected that a genetic risk factor(s) is responsible. Additionally, there is a potential for multiple dyslipidemia subtypes to exist within the breed, as Miniature Schnauzers with HTG have varied responses to management strategies.

The objective of this thesis is to identify the spectrum of dyslipidemia subtypes within the Miniature Schnauzer breed, metabolically characterize dyslipidemia in Miniature Schnauzers, and detect genetic risk variants associated with primary HTG in Miniature Schnauzers. Accomplishing these three objectives will lay the groundwork for the identification of dogs at risk, allowing for earlier intervention and informing treatment strategies.

Research aims, Findings, and Future Directions

The presence of varied responses to management strategies, such as feeding a low-fat diet[14,15], indicates a potential for multiple HTG subtypes within the breed. Thus, one objective of this research was to ascertain the range of dyslipidemia subtypes in the

Miniature Schnauzer breed. In **Chapter 2**, hierarchical cluster analysis of lipoprotein profiles from Miniature Schnauzers with primary HTG and those with normal serum triglyceride concentrations (NTG) was used to identify potential subtypes in the breed. The hierarchical cluster analysis identified six clusters, three of which comprised predominantly HTG dogs and two comprised only NTG dogs. The final cluster comprised a mix of NTG and mild HTG dogs. Generally, TRL fraction intensities were increased in clusters containing HTG dogs, and the major differences between clusters occurred in the low-density lipoprotein (LDL) fractions. The clinical data evaluated in this study (i.e., age, sex, BCS, and dietary fat) was not a significant source of differences between clusters. These results support the hypothesis that multiple dyslipidemia subtypes exist in the Miniature Schnauzer breed and that differences in LDL fractions may be a major distinguishing factor between the subtypes.

In future studies, inclusion of additional dogs, both Miniature Schnauzers and dogs of non-Miniature Schnauzer breeds, could reveal further separations which capture the true range of dyslipidemia subtypes in the breed and determine whether any are breed specific. For example, as most dogs included in this study reside within the Midwest, it is possible that including Miniature Schnauzers from other geographic locations would capture additional subtypes within the breed. The inclusion of non-Miniature Schnauzer breeds could clarify which differences between the clusters are normal variations of lipoprotein profiles and whether any features are breed specific. Furthermore, including additional dogs with secondary HTG, of both Miniature Schnauzer and non-Miniature Schnauzer breeds, could help separate differences in lipoprotein profiles that are directly associated with an endocrinopathy (secondary) from those due to underlying genetic risk

factors (primary). Finally, other methods of lipoprotein profiling do exist, such as LipoPrint, which allows for the further separation of lipoprotein subfractions.[136] Specifically, LipoPrint allows for further separation of the high-density lipoprotein (HDL) subfractions. Cluster analysis with this method may identify further distinguishing factors in the lipoprotein profiles between subtypes. We have generated LipoPrint profiles for 36 Miniature Schnauzers with known HTG phenotypes (21 HTG and 15 NTG). Of these, 11 also have lipoprotein profiles generated from the continuous lipoprotein density profiling method described in Chapter 2. In the future, hierarchical cluster analysis will be applied to the LipoPrint profile data in a similar approach.

The precise molecular mechanisms underlying HTG in Miniature Schnauzers are unknown. The analysis of lipidomic and metabolomic data has the potential to reveal pathophysiological processes and specific derangements that contribute to HTG in Miniature Schnauzers. Thus, the analysis of this data was used to ascertain the lipid species and metabolites that distinguish Miniature Schnauzers with primary HTG from those with NTG (**Chapter 3**). Nearly all (89%) of the lipid species differed between these two groups, with the majority being triglyceride species that were increased in those with moderate-to-severe primary HTG. The differentiating lipid species suggest possible involvement or disruption of the pathways and products of glycerolipid, glycerophospholipid, and glycosphingolipid metabolism. In contrast to the lipidome, only 1% of the metabolites significantly differed between groups. Many of the differentiating metabolites are involved in fatty acid metabolism pathways. These findings offer insights into the underlying pathways and potential biomarkers for moderate-to-severe primary HTG in Miniature Schnauzers. Further, there are major differences in the alterations that

occur in dogs with endocrinopathies from what we found in Miniature Schnauzers with primary HTG, suggesting a distinct type of dyslipidemia.

The results of this study provide insights into the underlying pathways. It is still unknown which pathways are contributors to HTG development versus which are secondary responses to HTG. Additional studies are warranted to determine the role of these pathways in HTG in Miniature Schnauzers. Our approach specifically looked for lipid species and metabolites that distinguish Miniature Schnauzers with moderate-to-severe primary HTG from those with NTG. However, without non-Miniature Schnauzer breeds, it is unknown if the HTG changes are breed-specific or found in any breed with primary HTG and if NTG Miniature Schnauzers are in fact "normal."

A second objective of **Chapter 3** was to use hierarchical cluster analysis to determine how the lipidomic and metabolomic profiles of Miniature Schnauzers with presumed secondary HTG compare to those with primary HTG and NTG. Hierarchical cluster analysis of the lipidomic data resulted in the perfect separation of moderate-to-severe HTG (primary or secondary) from NTG. In contrast, many dogs with mild HTG clustered with NTG dogs, suggesting that dogs with mild elevations in triglyceride concentrations may not always represent a significant dyslipidemia. Instead, mild TG elevations in these dogs could be the result of age or other environmental factors such as diet.[4,14,34,38] A distinct cluster did not form for dogs with presumed secondary HTG with the lipidomic data, but they did cluster together with the metabolomic data. Specifically, one cluster comprised all four dogs with secondary HTG and three with moderate-to-severe primary HTG. This could indicate similarities in the metabolic response to increased triglyceride concentrations in both classifications. Alternatively, it is possible

that the three dogs with moderate-to-severe primary HTG in this cluster had an undiagnosed subclinical endocrinopathy. It is possible that including additional dogs with secondary HTG, and specifically for each endocrinopathy, would result in a distinct cluster(s) for secondary HTG. The remaining three clusters contained a mix of NTG, mild primary HTG, and moderate-to-severe primary HTG. As stated above, the inclusion of non-Miniature Schnauzer breeds with and without primary HTG, could help resolve if any changes are breed-specific and provide lipidomic and metabolomic profiles of dogs that are "normal" for comparison.

Evaluation of a larger number of Miniature Schnauzers with secondary HTG could identify lipid species and/or metabolites that distinguish them from Miniature Schnauzers with primary HTG. This would not only be useful for proper classification in future studies, but also has an advantage for clinical management. For example, the ability to determine whether elevated triglyceride concentrations in a Miniature Schnauzer are due to primary causes or secondary to an endocrinopathy could inform the clinical assessment of whether the underlying endocrinopathy is well managed or would benefit from adjustment in therapy. Additionally, these studies could validate the findings of previous studies in dogs with endocrinopathies or identify new biomarkers.[40,42-45] Earlier detection of HTG in Miniature Schnauzers is also important, as it would allow for intervention and treatment before the development of severe consequences. The analysis of the lipidome and metabolome before and after the onset of HTG could identify biomarkers that present before serum TG concentrations elevate.

An underlying genetic risk factor is suspected to be responsible for HTG in the breed. The identification of HTG risk variants would allow for early detection of susceptible dogs and could inform clinical care to prevent the development of HTG and its severe consequences. In **Chapter 4**, we used whole genome sequencing of eight affected dogs to discover putative risk variants in six candidate genes for HTG (i.e., *lipoprotein lipase (LPL), apolipoprotein C-II (APOC2), apolipoprotein A-V (APOA5), glycosylphosphatidylinositol-anchored HDL-binding protein 1 (GPIHBP1)*, and *lipase maturation factor 1* (*LMF1*), and *apolipoprotein E (APOE))*. Two variants passed the filtering criteria. The first was a three bp deletion in the TATA box of *APOE*. This variant was of particular interest as it was exclusive to Miniature Schnauzers and disturbed a conserved base. However, no association was identified between the *APOE* genotype and serum TG concentration in a follow-up cohort of 159 Miniature Schnauzers. The second was a *GPIHBP1* missense variant in the Ly-6 protein domain. While this variant resided in a protein domain, the specific base was poorly conserved, and the variant was predicted to have a neutral impact on the protein. Follow-up genotyping was not performed for the *GPIHBP1* variant and thus, the association between the variant genotype and TG concentration is unknown. The results of this study do not support a monogenic cause for HTG in the six candidate genes evaluated. Given these findings and growing data on hyperlipidemia subtypes in Miniature Schnauzers, it is possible that HTG is a polygenic or complex trait. Future genetic studies should consider the possibility of the cumulative effects of variants in multiple genes and the contribution of the environment.

A future step is to genotype the *GPIHBP1* variant in the follow-up cohort of Miniature Schnauzers. This will allow us to determine if the variant genotype is associated with triglyceride concentration. Additionally, as structural variants (e.g., copy number variations, inversions, and translocations) associated with primary HTG have been identified in *LPL* and *GPIHBP1,* all six candidate genes will be evaluated for structural variants.[82,83] Further, the whole genome sequences of the dogs from this study have been merged with those from a cohort of Miniature Schnauzers from the United Kingdom. The merged dataset will be analyzed through a pipeline developed by our collaborators in the United Kingdom. Briefly, after variant mapping and discovery, the pipeline prioritizes variants by frequency to identify both risk and protective variants, determines the impact of the variant on the protein, and draws from outside resources and information, such as the pathways and metabolites identified in **Chapter 3**, results from the analysis of RNASeq data, Gene Ontology, UniProt, and the GWAS catalog. Variants of interest identified by this analysis will be genotyped in a cohort of 200 Miniature Schnauzers using a genotyping array that allows multiple variants to be genotyped in many samples at once. The association between the genotypes and TG concentration will be tested in a multivariable regression.

A key challenge in genetic studies is the prioritization of identified variants. Missense variants have the potential to affect protein conformation, stability, and protein-protein interactions.[92] Many *in silico* tools have been developed to use features of amino acids and proteins to determine if a variant is likely pathogenic.[70,71,94-98] However, variant pathogenicity prediction programs are typically trained using human variants and most programs have not been validated for use in non-human species. In **Chapter 5**, the

performance of eight tools for pathogenicity prediction of missense variants (MutPred2, PANTHER, PhD-SNP, PolyPhen2-HumDiv, PolyPhen2-HumVar, Provean, SIFT, and SNPs&GO) were evaluated for use in the dog and horse. Seven of the variant pathogenicity prediction methods performed well within and across both species, though it was common to have two or more methods in disagreement for any given variant. Unlike the other methods, PANTHER had low accuracy and failed to provide a classification for a high proportion of variants. These findings suggest that the variant pathogenicity prediction methods evaluated in this study can be effectively used in dogs and horses. We recommend that investigators carefully consider what criteria are most valuable to the goal of their specific study (i.e., sensitivity or specificity) when choosing which method(s) to use. Additionally, it is important to interpret classifications from variant pathogenicity prediction methods carefully and not use these classifications as the sole source of evidence.[127]

Since our dataset was limited to dog and horse variants, we cannot infer how these programs would perform analyzing data from other species. Thus, performing similar evaluations of these programs in additional species is warranted. Additionally, the methods evaluated in this study are not the only methods available that can analyze missense variants from non-human species. Evaluating other variant pathogenicity methods could identify additional methods effective at predicting the impact of missense variants in non-human species. However, complete agreement across all eight programs was uncommon. It would be interesting to determine if the trait/disorder variants predicted as neutral by most of the methods are truly trait/disorder variants or if it is possible that they are not the causal variant. While functional studies would provide the

strongest evidence of variant effect, they may not be practical. Instead, evaluating these variants using the standardized criteria for classifying pathogenic variants could determine if enough evidence exists to classify the variants as pathogenic.[93] Finally, our study evaluated the performance of the methods using variants reported to contribute to Mendelian traits and disorders. The performance of these methods, when applied to the analysis of variants contributing to complex traits, is unknown and future studies are warranted to determine if performance is impacted.

Final Conclusions

This research investigated the metabolic and genetic risk factors of primary HTG in Miniature Schnauzers. Hierarchical cluster analysis of lipoprotein profiles provided evidence of multiple dyslipidemia subtypes in the Miniature Schnauzer breed. The major distinguishing factor between subtypes may be differences in LDL fractions. Analysis of the serum lipidome and metabolome of Miniature Schnauzers identified ways in which those with moderate-to-severe primary HTG differ from those with NTG. The alterations of the lipidome and metabolome in Miniature Schnauzers with moderate-to-severe primary HTG suggest the possible involvement or disruption of the pathways and products of glycerolipid, glycerophospholipid, glycosphingolipid, and fatty acid metabolism. Hierarchical cluster analysis of the lipidomic data determined that Miniature Schnauzers with mild HTG cannot be definitively classified as having primary HTG, as their lipid disturbances do not reliably differ from dogs with NTG. More research is needed to determine if the inclusion of additional secondary dogs could identify potential biomarkers that allow for the differentiation of primary from secondary HTG. The

candidate gene analysis did not support a monogenic cause for HTG in the candidate genes screened. The two identified variants were deemed unlikely to be causal, but neither could be ruled out as contributing factors. These findings, as well as the increasing evidence of dyslipidemia subtypes in the Miniature Schnauzer breed, suggest that HTG is a polygenic or complex trait.

References

1. Xenoulis PG, Steiner JM. Lipid metabolism and hyperlipidemia in dogs. *Vet J*. 2010;183(1):12-21.
2. Whitney MS, Boon GD, Rebar AH, Story JA, Bottoms GD. Ultracentrifugal and electrophoretic characteristics of lipoproteins of miniature schnauzer dogs with idiopathic hyperlipoproteinemia. *J Vet Intern Med*. 1993;7(4):253-260.
3. Xenoulis PG, Cammarata PJ, Walzem RL, Macfarlane RD, Suchodolski JS, and Steiner JM. Novel lipoprotein density profiling in healthy dogs of various breeds, healthy Miniature Schnauzers, and Miniature Schnauzers with hyperlipidemia. *BMC Vet Res*. 2013;9:47.
4. Xenoulis PG, Suchodolski JS, Levinski MD, Steiner JM. Investigation of hypertriglyceridemia in healthy Miniature Schnauzers. *J Vet Intern Med*. 2007;21:1224-1230.
5. Xenoulis PG, Steiner JM. Canine hyperlipidaemia. *J Small Anim Pract*. 2015;56(10):595-605.
6. Xenoulis PG, Suchodolski JS, Ruaux CG, Steiner JM. Association between serum triglyceride and canine pancreatic lipase immunoreactivity concentrations in miniature schnauzers. *J Am Anim Hosp Assoc*. 2010;46:229-34.
7. Xenoulis PG, Levinski MD, Suchodolski JS, Steiner JM. Serum triglyceride concentrations in Miniature Schnauzers with and without a history of probable pancreatitis. *J Vet Intern Med*. 2011;25:20-25.

8. Kutsunai M, Kanemoto H, Fukushima K, Fujino Y, Ohno K, Tsujimoto H. The association between gallbladder mucoceles and hyperlipidaemia in dogs: a retrospective case control study. *Vet J*. 2014;199(1):76-9.
9. Furrow E, Jaeger JQ, Parker VJ, Hinchcliff KW, Johnson SE, Murdoch SJ, de Boer IH, Sherding RG, and Brunzell JD. Proteinuria and lipoprotein lipase activity in Miniature Schnauzer dogs with and without hypertriglyceridemia. *Vet J*. 2016;212:83-89.
10. Furrow E, Lees GE, Brown CA, and Cianciolo RE. Glomerular lesions in proteinuric Miniature Schnauzer dogs. *Vet Pathol*. 2017;54(3):484-489.
11. Lewis GF, Xiao C, Hegele RA. Hypertriglyceridemia in the genomic era: a new paradigm. *Endocr Rev*. 2015;36(1):131-47.
12. Xenoulis PG, Tate NM, Bishop MA, Steiner JM, Suchodolski J, Furrow E. Sequence analysis of the coding regions of the apolipoprotein C2 (*APOC2*) gene in Miniature Schnauzers with idiopathic hypertriglyceridemia. *Vet J*. 2020;265:105559.
13. Schickel R. Identification of the nucleotide sequence of the lipoprotein lipase gene as well as its role in the development of hyperlipidemia and pancreatitis in the Miniature Schnauzer. 2005. Ludwig-Maximilians University (Dr.med.vet. thesis)
14. Xenoulis PG, Cammarata PJ, Walzem RL, Suchodolski JS, Steiner JM. Effect of a low-fat diet on serum triglyceride and cholesterol concentrations and lipoprotein profiles in Miniature Schnauzers with hypertriglyceridemia. *J Vet Intern Med*. 2020;34:2605-2616.

15. de Albuquerque P, De Marco V, Vendramini THA, Amaral AR, Catanozi S, Santana KG, Nunes VS, Nakandakare ER, Brunetto MA. Supplementation of omega-3 and dietary factors can influence the cholesterolemia and triglyceridemia in hyperlipidemic Schnauzer dogs: A preliminary report. *PLoS One*. 2021;16:e0258058.

16. van Rooden SM, Heiser WJ, Kok JN, Verbaan D, van Hilten JJ, Marinus J. The identification of Parkinson's disease subtypes using cluster analysis: a systemic review. *Mov Disord*. 2010;25(8):969-978.

17. Zinchuk A, Yaggi HK. Phenotypic subtypes of OSA: a challenge and opportunity for precision medicine. *Chest*. 2020;157(2):403-420.

18. García-Solano B, Gallegos-Cabriales EC, Gómez-Meza MV, García-Madrid G, Flores-Merlo M, García-Solano M. Hierarchical clusters in families with type 2 diabetes. *SAGE Open Med*. 2015;3:2050312115622957.

19. Youroukova VM, Dimitrova DG, Valerieva AD, Lesichkova SS, Velikova TV, Ivanova-Todorova EI, Tumangelova-Yuzeir KD. Phenotypes determined by cluster analysis in moderate to severe bronchial asthma. *Folia Med (Plovdiv)*. 2017;59(2):165-173.

20. Pikoula M, Quint JK, Nissen F, Hemingway H, Smeeth L, Denaxas S. Identifying clinically important COPD sub-types using data-driven approaches in primary care population based electronic health records. *BMC Med Inform Decis Mak*. 2019; 19(1):86.

21. Burgel PR, Paillasseur JL, Peene B, Dusser D, Roche N, Coolen J, Troosters T, Decramer M, Janssens W. Two distinct chronic obstructive pulmonary disease

(COPD) phenotypes are associated with high risk of mortality. *PLoS One*. 2012;7(12):e51048.

22. Furrow E, Patterson EE, Armstrong PJ, Osborne CA, Lulich JP. Fasting urinary calcium-to-creatinine and oxalate-to-creatinine ratios in dogs with calcium oxalate urolithiasis and breed-matched controls. *J Vet Intern Med*. 2015;29(1):113-119.
23. Smith RE, Granick JL, Stauthammer CD, Polzin DJ, Heinrich DA, Furrow E. Clinical consequences of hypertriglyceridemia associated proteinuria in Miniature Schnauzers. *J Vet Intern Med*. 2017;31(6):1740-1748.
24. Minamoto T, Parambeth JC, Walzem RL, Payne HR, Lidbury JA, Suchodolski JS, Steiner JM. Evaluation of density gradient ultracentrifugation serum lipoprotein profiles in healthy dogs and dogs with exocrine pancreatic insufficiency. *J Vet Diagn Invest*. 2018;30(6):878-886.
25. R Core Team (2021). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org.
26. Xenoulis PG, Cammarata PJ, Walzem RL, Suchodolski JS, Steiner JM. Serum triglyceride and cholesterol concentrations and lipoprotein profiles in dogs with naturally occurring pancreatitis and healthy control dogs. *J Vet Intern Med*. 2020;34(2):644-652.
27. Rogers WA, Donovan EF, Kociba GJ. Lipids and lipoproteins in normal dogs and in dogs with secondary hyperlipoproteinemia. *J Am Vet Med Assoc*. 1975;166:1092-1100.

28. Seage EC, Drobatz KJ, Hess RS. Spectrophotometry and ultracentrifugation for measurement of plasma lipids in dogs with diabetes mellitus. *J Vet Intern Med*. 2018;32:93-98.

29. Kim K, Han S, Jeon K, Kim H, Li Q, Ryu M, Song W, Park S, Youn H. Clinical relationship between cholestatic disease and pituitary-dependent hyperadrenocorticism in dogs: a retrospective case series. *J Vet Intern Med*. 2017;31:335-342.

30. Tani M, Horvath KV, Lamarche B, Couture P, Burnett JR, Schaefer EJ, Asztalos BF. High-denisty lipoprotein subpopulation profiles in lipoprotein lipase and hepatic lipase deficiency. *Atherosclerosis*. 2016;253:7-14.

31. DiNicolantonio JJ, O'Keefe JH. Effects of dietary fats on blood lipids: a review of direct comparison trials. *Open Heart* 2018;5:e000871.

32. Hannon BA, Edwards CG, Thompson SV, Burke SK, Burd NA, Holscher HD, Teran-Garcia M, Khan NA. Genetic Variants in Lipid Metabolism Pathways Interact with Diet to Influence Blood Lipid Concentrations in Adults with Overweight and Obesity. *Lifestyle Genom*. 2020;13(6):155-163.

33. Barrie J, Watson TDG, Stear MJ, Nash AS. Plasma cholesterol and lipoprotein concentrations in the dog: the effects of age, breed, gender, and endocrine disease. *J Small Anim Pract*. 1993;10:507-512.

34. Kawasumi K, Kashiwado N, Okada Y, Sawamura M, Sasaki Y, Iwazaki E, Mori N, Yamamoto I, Arai T. Age effects on plasma cholesterol and triglyceride profiles and metabolite concentrations in dogs. *BMC Vet Res*. 2014;10:57.

35. Usui, S, Mizoguchi Y, Yasuda H, Arai N, Koketsu Y. Dog age and breeds associated with high plasma cholesterol and triglyceride concentrations. *J Vet Med Sci*. 2014; 76(2):269-272.
36. Spitler KM, Davies BSJ. Aging and plasma triglyceride metabolism. *J Lipid Res*. 2020;61(8):1161-1167.
37. Milyani AA, Al-Agha AE. The effect of body mass index and gender on lipid profile in children and adolescents in Saudi Arabia. *Ann Afr Med*. 2019;18(1):42-46.
38. Usui S, Yasuda H, Koketsu Y. Lipoprotein cholesterol and triglyceride concentrations associated with dog body condition score; effect of recommended fasting duration on sample concentrations in Japanese private clinics. *J Vet Med Sci*. 2015;77(9):1063-1069.
39. Söder J, Höglund K, Dicksved J, Hagman R, Röhnisch HE, Moazzami AA, Wernersson S. Plasma metabolomics reveals lower carnitine concentrations in overweight Labrador Retriever dogs. *Acta Vet Scand*. 61(1):10.
40. Sieber-Ruckstuhl NS, Tham WK, Baumgartner F, Selva JJ, Wenk MR, Burla B, Boretti FS. Serum lipidome signatures of dogs with different endocrinopathies associated with hyperlipidemia. *Metabolites*. 2022;12(4):306.
41. Vendramini THA, Macedo HT, Zafalon RVA, Macegoza MV, Pedrinelli V, Risolia LW, Ocampos FMM, Jeremias JT, Pontieri CFF, Ferriolli E, Colnago LA, Brunetto MA. Serum metabolomics analysis reveals that weight loss in obese dogs results in a similar metabolic profile to dogs in ideal body condition. *Metabolomics*. 2021;17(3):27.

42. Muñoz-Prieto A, Rubić I, Horvatić A, Rafaj RB, Cerón JJ, Tvarijonaviciute A, Mrljak V. Metabolic profiling of serum from dogs with pituitary-dependent hyperadrenocorticism. *Res Vet Sci*. 2021;138:161-166.
43. Muñoz-Prieto A, González-Arostegui LG, Rubić I, Cerón JJ, Tvarijonaviciute A, Horvatić A, Mrljak V. Untargeted metabolomic profiling of serum in dogs with hypothyroidism. *Res Vet Sci*. 2021;136:6-10.
44. O'Kell AL, Garrett TJ, Wasserfall C, Atkinson MA. Untargeted metabolomic analysis in naturally occurring canine diabetes mellitus identifies similarities to human type I diabetes. *Sci Rep*. 2017;7(1):9467.
45. O'Kell AL, Garrett TJ, Wasserfall C, Atkinson MA. Untargeted metabolomic analysis in non-fasted diabetic dogs by UHPLC-HRMS. *Metabolomics*. 2019;15(2):15.
46. Gookin JL, Mathews KG, Cullen J, Seiler G. Qualitative metabolomics profiling of serum and bile from dogs with gallbladder mucocele formation. *PLoS One*. 2018;13(1):e0191076.
47. Xenoulis PG, Suchodolski JS, Levinski MD, Steiner JM. Serum liver enzyme activities in healthy Miniature Schnauzers with and without hypertriglyceridemia. *J Am Vet Med Assoc*. 2008;232(1):63-67.
48. Li Q, Laflamme DP, Bauer JE. Serum untargeted metabolomic changes in response to diet intervention in dogs with preclinical myxomatous mitral valve disease. *PLoS One*. 2020;15(6):e0234404.

49. Pang Z, Chong J, Zhou G, Morais DAL, Chang L, Barrette M, Gauthier C, Jacques PE, Li S, Xia J. MetaboAnalyst5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res*. 2021;49(W1):W388-W396.
50. Coleman RA, Mashek DG. Mammalian triacylglycerol metabolism: synthesis, lipolysis, and signaling. *Chem Rev*. 2011;111(10):6359-6386.
51. Alves-Bezzerra M, Cohen DE. Triglcyeride metabolism in the liver. *Compr Physiol*. 2017;8(1):1-8.
52. Smith AB, Schill JP, Gordillo R, Gustafson GE, Rhoads RW, Burhans MS, Broman AT, Colman RJ, Sherer PE, Anderson RM. Ceramides are early responders in metabolic syndrome development in rhesus monkeys. *Sci Rep*. 2022;12(1):9960.
53. Singh RK, Fullerton MD, Vine D, Bakovic M. Mechanism of hypertriglyceridemia in CTP:phosphoethanolamine cytidylyltransferase-deficient mice. *J Lipid Res*. 2012;53(9):1811-1822.
54. Mihalik SJ, Goodpaster BH, Kelley DE, Chace DH, Vockley J, Toledo FG, DeLany JP. Increased levels of plasma acylcarnitines in obesity and type 2 diabetes and identification of a marker of glucolipotoxicity. *Obesity (Silver Spring)*. 2010;18(9):1695-700.
55. Mai M, Tönjes A, Kovacs P, Stumvoll M, Fiedler GM, Leichtle AB. Serum levels of acylcarnitines are altered in prediabetic conditions. *PLoS One*. 2013;8(12):e82459.
56. Zhang X, Zhang C, Chen L, Han X, Ji L. Human serum acylcarnitine profiles in different glucose tolerance states. *Diabetes Res Clin Pract*. 2014;104(3):376-82.

57. Schooneman MG, Vaz FM, Houten SM, Soeters MR. Acylcarnitines: reflecting or inflicting insulin resistance? *Diabetes*. 2013;62(1):1-8.

58. Hua S, Qi Q, Kizer JR, Williams-Nguyen J, Strickler HD, Thyagarajan B, Daviglus M, Talavera GA, Schneiderman N, Cotler SJ, Cai J, Kaplan R, Isasi CR. Association of liver enzymes with incident diabetes in US Hispanic/Latino adults. *Diabet Med*. 2021;38(8):e14522.

59. Tsurutani Y, Inoue K, Sugisawa C, Saito J, Omura M, Nishikawa T. Increased serum dihomo-γ-linolenic acid levels are associated with obesity, body fat accumulation, and insulin resistance in Japanese patients with type 2 diabetes. *Intern Med*. 2018;57(20):2929-2935.

60. Lankinen MA, Stančáková A, Uusitupa M, Ågren J, Pihlajamäki J, Kuusisto J, Schwab U, Laakso M. Plasma fatty acids as predictors of glycaemia and type 2 diabetes. *Diabetologia*. 2015;58(11):2533-44.

61. Zhang F, Lupski JR. Non-coding genetic variants in human disease. *Hum Mol Genet*. 2015;24(R1):R102-R110.

62. Turro E, Astle WJ, Megy K, Graf S, Greene D, Shamardina O, et al. Whole-genome sequencing of patients with rare diseases in a national health system. *Nature*. 2020;583(7814):96-102.

63. van Steenbeek FG, Hytönen MK, Leegwater PAJ, Lohi H. The canine era: the rise of a biomedical model. *Anim Genet*. 2016;47(5):519-527.

64. Friedenberg SG, Meurs KM. Genotype imputation in the domestic dog. *Mamm Genome*. 2016;27:485-494.

65. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17:122.

66. Haberle V, Stark A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol*. 2018;19(10):621-637.

67. Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res*. 1987;15(17):7155-7174.

68. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res*. 2010;20(1):110–121.

69. Blanchette M, Kent WJ, Reimer C, Elnitski L, Smit AF, Roskin KM, Baertsch R, Rosenbloom K, Clawson H, Green ED, Haussler D, Miller W. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res*. 2004;14(4):708–715.

70. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010; 7(4):248-249.

71. Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum Mutat*. 2009;30(8):1237-1244.

72. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan5:genome-scale protein function classification. *Bioinformatics*. 2014;30(9):1236-1240.

73. Vincze, T., Posfai, J. and Roberts, R.J. NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Res*. 2003;31: 3688-3691.

74. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M and Rozen SG. Primer3--new capabilities and interfaces. *Nucleic Acids Res*. 2012;40(15):e115.

75. Horuchi K, Tajima S, Menju M, Yamamoto A. Structure and expression of mouse apolipoprotein E gene. *J Biochem*. 1989;106(1):98-103.

76. Maloney B, Ge YW, Alley GM, Lahiri DK. Important differences between human and mouse APOE gene promoters: limitation of mouse APOE model in studying Alzheimer's disease. *J Neurochem*. 2007;103(3):1237-1257.

77. Wu SA, Kersten S, Qi L. Lipoprotein Lipase and Its Regulators: An Unfolding Story. *Trends Endocrinol Metab*. 2021;32(1):48-61.

78. Bae SH, Han HW, Moon J. Functional analysis of the molecular interactions of TATA box-containing genes and essential genes. *PLoS One*. 2015;10(3):e0120848.

79. Zukunft J, Lang T, Richter T, et al. A natural CYP2B6 TATA box polymorphism (82T-> C) leading to enhanced transcription and relocation of the transcriptional start site. *Mol Pharmacol*. 2005;67:1772-1782.

80. Matsunaga A, Sasaki J, Han H, Huang W, Kugi M, Koga T, Ichiki S, Shinkawa T, Arakawa K. Compound heterozygosity for an apolipoprotein A1 gene promoter mutation and a structural nonsense mutation with apolipoprotein A1 deficiency. *Arterioscler Thromb Vasc Biol*. 1999;19(2):348-55.

81. Young SG, Fong LG, Beigneux AP, Allan CM, He C, Jiang H, Nakajima K, Meiyappan M, Birrane G, Ploug M. GPIHBP1 and Lipoprotein Lipase, Partners in Plasma Triglyceride Metabolism. *Cell Metab*. 2019;30(1):51-65.

82. Dron JS, Wang J, McIntyre AD, Cao H, Robinson JF, Duell PB, Manjoo P, Feng J, Movsesyan I, Malloy MJ, Pullinger CR, Kane JP, Hegele RA. Partial LPL deletions: rare copy-number variants contributing towards severe hypertriglyceridemia. *J Lipid Res*. 2019;60(11):1593-1958.

83. Patni N, Brothers J, Xing C, Garg A. Type 1 hyperlipoproteinemia in a child with large homozygous deletion encompassing *GPIHBP1*. *J Clin Lipidol*. 2016;10(4):1035-1039.

84. Wang J, Cao H, Ban MR, Kennedy BA, Zhu S, Anand S, Yusuf S, Pollex RL, Hegele RA. Resequencing genomic DNA of patients with severe hypertriglyceridemia (MIM 144650). *Arterioscler Thromb Vasc Biol*. 2007;27:2450-2455.

85. Wang J, Cao H, Ban MR, Kennedy BA, Zhu S, Anand S, Yusuf S, Pollex RL, Hegele RA. Resequencing genomic DNA of patients with severe hypertriglyceridemia (MIM 144650). *Arterioscler Thromb Vasc Biol*. 2007;27:2450-2455.

86. Surendran RP, Visser ME, Heemelaar S, Wang J, Peter J, Defesche JC, Kuivenhoven JA, Hosseini M, Peterfy M, Kastelein JJP, Johansen CT, Hegele RA, Stroes ESG, Dallinga-Thie GM. Mutations in LPL, APOC2, APOA5, GPIHBP1, and LMF1 in patients with severe hypertriglyceridemia. *J Intern Med*. 2012;272(2):185-196.

87. Dron JS, Wang J, Cao H, McIntyre AD, Iacocca MA, Menard JR, Movsesyan I, Malloy MJ, Pullinger CR, Kane JP, Hegele RA. Severe hypertriglyceridemia is primarily polygenic. *J Clin Lipidol*. 2019;13(1):80-88.

88. Dron JS, Wang J, McIntyre AD, Cao H, Hegele RA. The polygenic nature of mild-to-moderate hypertriglyceridemia. *J Clin Lipidol*. 2020;14(1):28-34.e2.

89. Rehder C, Bean LJH, Bick D, Chao E, Chung W, Das S, O'Daniel J, Rehm H, Shashi V, Vincent LM, ACMG Laboratory Quality Assurance Committee. Next-generation sequencing for constitutional variants in the clinical laboratory, 2021 revision: a technical standard of the American College of Medical Genetics and Genomics (ACMG). *Genet Med*. 2021;23(8):1399-1415.

90. Durward-Akhurst SA, Schaefer RJ, Grantham B, Carey WK, Mickelson JR, McCue ME. Genetic variation and the distribution of variant types in the horse. *Front Genet*. 2021;12:758366.

91. Eilbeck K, Quinlan A, Yandell M. Settling the score: variant prioritization and Mendelian disease. *Nat Rev Genet*. 2017;18(10):599-612.

92. Thusberg J, Vihinen M. Pathogenic or not? And if so, then how? Studying the effects of missense mutations using bioinformatics methods. *Hum Mutat*. 2009;30(5):703-714.

93. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical

Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-424.

94. Pejaver V, Urresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam H, Mort M, Cooper DN, Sebat J, Iakoucheva LM, Mooney SD, Radivojac P. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat Commun*. 2020;11(1):5918.
95. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res*. 2003;13(9):2129-2141.
96. Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics*. 2006;22(22):2729-2734.
97. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *Plos One*. 2012;7(10):e46688.
98. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. *Genome Res*. 2001; 11(5):863-874.
99. Groß C, de Ridder D, Reinders M. Predicting the deleteriousness in non-human species: applying the CADD approach in mouse. *BMC Bioinformatics*. 2018;19(1):373.
100. Groß C, Derks M, Megens H, Bosse M, Groenen MAM, Reinders M, de Ridder D. pCADD: SNV prioritization in Sus scrofa. *Genet Sel Evol*. 2020a;52(1):4.

101. Groß C, Bortoluzzi C, de Ridder D, Megens H, Groenen MAM, Reinders M, Bosse M. Prioritizing sequence variants in conserved non-coding elements in the chicken genome using chCADD. *PLoS Genet*. 2020b;16(9):e1009027.

102. Hunt SE, McLaren W, Gil L, Thormann A, Schuilenburg H, Sheppard D, Parton A, Armean IM, Trvanion SJ, Flicek P, Cunningham F. Ensembl variation resources. *Database (Oxford)*. 2018.

103. Miosge LA, Field MA, Sontani Y, Cho V, Johnson S, Palkova A, Balakishnan B, Liang R, Zhang Y, Lyon S, Beutler B, Whittle B, Bertram EM, Enders A, Goodnow CC, Andrews TD. Comparison of predicted and actual consequences of missense mutations. *Proc Natl Acad Sci USA*. 2015;112(37):E5189-5198.

104. Shearin AL, Ostrander EA. Leading the way: canine models of genomics and disease. *Dis Model Mech*. 2010;3(1-2):27-34.

105. Raudsepp T, Finno CJ, Bellone RR, Petersen JL. Ten years of the horse reference genome: insights into equine biology, domestication and population dynamics in the post-genome era. *Anim Genet*. 2019;50(6):569-597.

106. Thusberg J, Olatubosun A, Vihinen M. Performance of mutation pathogenicity prediction methods on missense variants. *Hum Mutat*. 2011;32(4):358-368.

107. Suybeng V, Koeppel F, Harlé A, Rouleau E. Comparison of pathogenicity prediction tools on somatic variants. *J Mol Diagn*. 2020;22(12):1383-1392.

108. Frousios K, Iliopoulos CS, Schlitt T, Simpson MA. Predicting the functional consequences of non-synonymous DNA sequence variants – evaluation of bioinformatic tools and development of a consensus strategy. *Genomics*. 2013;102(4):223-228.

109. de la Campa EA, Padilla N, de la Cruz X. Development of pathogenicity predictors specific for variants that do not comply with clinical guidelines for the use of computational evidence. *BMC Genomics*. 2017; 18(5):569.

110. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics. 2005;*21(16):3439–3440.

111. Durinck S, Spellman P, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc.* 2009;4(8):1184–1191.

112. Niroula A, Urolagin S, Vihinen M. PON-P2: prediction method for fast and reliable identification of harmful variants. *PLoS One*. 2015;10(2):e0117380.

113. Knecht C, Mort M, Junge O, Cooper DN, Krawczak M, Caliebe A. IMHOTEP – a composite score integrating popular tools for predicting the functional consequences of non-synonymous sequence variants. *Nucleic Acids Res*. 2017;45(3):e13.

114. National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [cited 2017 Apr 06]. Available from: https://www.ncbi.nlm.nih.gov/

115. Leeb T. Concern regarding the publication by Posbergh et al. "A nonsynonymous change in adhesion G protein-coupled receptor L3 associated with risk for Equine Degenerative Myeloencephalopathy in the Caspian horse. *J Equine Vet Sci*. 2019;72:124.

116. Marquardt SA, Wilcox CV, Burns EN, Peterson JA, Finno CJ. Previously identified genetic variants in ADGRL3 are not associated with risk for equine degenerative myeloencephalopathy across breeds. *Genes – Basel*. 2019;10(9):681.

117. Bannasch DL, Kaelin CB, Letko A, Loechel R, Hug P, Jagannathan V, Henkel J, Roosje P, Hytönen MK, Lohi H, Arumilli M, DoGA consortium, Minor KM, Mickelson JR, Drögemüller C, Barsh GS, Leeb T. Dog colour patterns explained by modular promoters of ancient canid origin. *Nat Ecol Evol*. 2021;5(10):1415-1423.

118. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, Liu X. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet*. 2015;24(8):2125-2137.

119. Hassan MS, Shaalan AA, Dessouky MI, Abdelnaiem AE, ElHefnawi M. Evaluation of computational techniques for predicting non-synonymous single nucleotide variants pathogenicity. *Genomics*. 2019;111(4):869-882.

120. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*. 2020;21(1):6.

121. Capriotti E, Montanucci L, Profiti G, Rossi I, Giannuzzi D, Aresu L, Fariselli P. Fido-SNP: the first webserver for scoring the impact of single nucleotide variants in the dog genome. *Nucleic Acids Res*. 2019;47(W1):W136-W141.

122. Maxwell KN, Hart SN, Vijai J, Schrader KA, Slavin TP, Thomas T, Wubbenhorst B, Ravichandran V, Moore RM, Hu C, Guidugli L, Wenz B,

Domchek SM, Robson ME, Szabo C, Neuhausen SL, Weitzel JN, Offit K, Couch FJ, Nathanson KL. Evaluation of ACMG-guideline-based variant classification of cancer susceptibility and non-cancer-associated genes in families affected by breast cancer. *Am J Hum Genet*. 2016;98(5):801-817.

123. Ghosh R, Oak N, Plon SE. Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. *Genome Biol*. 2017;18(1):225.

124. Cubuk C, Garret A, Choi S, King L, Loveday C, Torr B, Burghel GJ, Durkie M, Callaway A, Robinson R, Drummond J, Berry I, Wallace A, Eccles D, Tischkowitz M, Whiffin N, Ware JS, Hanson H, Turnbull C, CanVIG-UK. Clinical likelihood ratios and balanced accuracy for 44 in silico tools against multiple large-scale functional assays of cancer susceptibility genes. *Genet Med*. 2021;23:2096-2104.

125. Hart SN, Hoskin T, Shimelis H, Moore RM, Feng B, Thomas A, Lindor NM, Polley EC, Goldgar DE, Iversen E, Monteiro ANA, Suman VJ, Couch FJ. Comprehensive annotation of BRCA1 and BRCA2 missense variants by functionally validated sequence-based computational prediction models. *Genet Med*. 2019;21(1):71-80.

126. Gunning AC, Fryer V, Fasham J, Crosby AH, Ellard S, Baple EL, Wright CF. Assessing the performance of pathogenicity predictors using clinically relevant variant datasets. *J Med Genet*. 2021;58(8):547-555.

127. Vihinen M. Problems in variation interpretation guidelines and in their implementation in computational tools. *Mol Genet Genomic Med*. 2020;8(9):e1206.

128. Cadieu E, Neff MW, Quignon P, Walsh K, Chase K, Parker HG, Vonholdt BM, Rhue A, Boyko A, Byers A, Wong A, Mosher DS, Elkahloun AG, Spady TC, André C, Lark KG, Cargill M, Bustamante CD, Wayne RK, Ostrander EA. Coat variation in the domestic dog is governed by variants in three genes. *Science*. 2009;326(5949):150-153.

129. Thomas PD, Kejariwal A. Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. *Proc Natl Acad Sci USA*. 2004;101(43):15398-15403.

130. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA*. 2009;106(23):9362-7.

131. Shihab HA, Rogers MF, Gough J, Mort M, Cooper DN, Day IN, Gaunt TR, Campbell C. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics*. 2015;31(10):1536-43.

132. Fu Y, Liu Z, Lou S, Bedford J, Mu XJ, Yip KY, Khurana E, Gerstein M. FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer. *Genome Biol*. 2014;15(10):480.

133. Gronau I, Arbiza L, Mohammed J, Siepel A. Inference of natural selection from interspersed genomic elements based on polymorphism and divergence. *Mol Biol Evol*. 2013;30(5):1159-71.

134. Gussow AB, Copeland BR, Dhindsa RS, Wang Q, Petrovski S, Majoros WH, Allen AS, Goldstein DB. Orion: Detecting regions of the human non-coding genome that are intolerant to variation using population genetics. *PLoS One*. 2017;12(8):e0181604.

135. Glusman G, Yanai I, Rubin I, Lancet D. The complete human olfactory subgenome. *Genome Res*. 2001;11(5):685-702.

136. Oravec S, Dostal E, Dukát A, Gavorník P, Kucera M, Gruber K. HDL subfractions analysis: a new laboratory diagnostic assay for patients with cardiovascular diseases and dyslipoproteinemia. *Neuro Endocrinol Lett*. 2011;32(4):502-9.