Impact of Warfarin Pharmacogene Variation on Drug Metabolism and Pharmacological Response in Alaska Native and American Indian Populations

Lindsay Maritza Henderson

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Kenneth Thummel, Chair

Allan Rettie

Edward Kelly

Program Authorized to Offer Degree:

Pharmaceutics

University of Washington

Abstract

Impact of Warfarin Pharmacogene Variation on Drug Metabolism and Pharmacological Response in Alaska Native and American Indian Populations

Lindsay Maritza Henderson

Chair of the Supervisory Committee:
Dr. Kenneth Thummel
Department of Pharmaceutics

The oral vitamin K antagonist warfarin (Coumadin®) is used to prevent stroke in patients with atrial fibrillation and for secondary prevention of venous thromboembolism. Despite newer treatment options such as the direct oral anticoagulants, warfarin remains a mainstay in anticoagulation therapy. Warfarin therapy requires intensive monitoring and dose titration due to its narrow therapeutic index and wide inter-individual response, due in part to genetic variation, as well as clinical, demographic, and environmental factors. While variation in *VKORC1, CYP2C9, CYP4F2* and *GGCX* genes have been associated with the warfarin dose required to achieve a therapeutic anticoagulation response, these findings are based largely on variant alleles and their frequencies found in populations largely of European descent, and may not be

generalizable to other, less studied populations such as Alaska Native and American Indian (AN/AI) people. It is important to characterize the unique genetic variation that exists in the AN/AI population because there are clinical implications of having uncharacterized genetic variation (eg, phenotype misclassification), particularly with drug metabolizing enzymes and drug target proteins. The projects described in this dissertation proposal inform on personalized warfarin therapy for AN/AI patients living in remote communities of Alaska and investigate the missing heritability in warfarin dose variance.

Chapter 2 and 3 describe the *in vitro* and *in vivo* characterization of novel variation in *CYP2C9* found in the indigenous population of Alaska. In Chapter 4, the impact of *CYP2C9, VKORC1, CYP4F2, CYP4F11*, and *GGCX* gene variation on stable warfarin dose was determined in an AN/AI population living in Southcentral Alaska. Chapter 5 seeks to better understand the variability in warfarin dose through a pathway-based analysis identifying potential regulators of *VKORC1* that influence VKOR expression, and therefore warfarin dose requirement.

This dissertation research has improved our understanding of the clinical relevance of *VKORC1* as a determinant of therapeutic warfarin dose requirement in AN/AI people as well as the pharmacokinetic impact of novel *CYP2C9* variation on warfarin metabolism and pharmacological response, in addition to other narrow therapeutic index CYP2C9 substrates. Overall, this data can be used to implement population-specific genetic variation in clinical decisions associated with personalized or precision medicine, and to assess whether pharmacogenetic testing provides unique advantages in rural AN/AI communities.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my advisor, Dr. Kenneth Thummel for his guidance and support throughout this process. I appreciate all of his contribution of time and ideas to make my graduate school experience productive and stimulating. The enthusiasm that Ken has for research is inspiring and I will be forever grateful for the training he provided to help shape me into an independent researcher.

I would also like to thank my doctoral supervisory committee – Dr. Allan Rettie, Dr. Edward Kelly, Dr. Timothy Thornton, and Dr. Steve White – for their thoughtful discussions and direction during the course of my research.

To the members of the Thummel Lab, I am thankful for your sound scientific advice, unwavering support, and for your friendship. I would also like to acknowledge the positive impact that the supportive environment, cultivated by the faculty, staff, and students of the Pharmaceutics, Medicinal Chemistry, and Pharmacy Departments has had on me as a graduate student.

Thank you to everyone who is a part of the Northwest-Alaska Pharmacogenetic Research Network – your collaboration and contributions made this dissertation research possible. I thank the Yup'ik people of the Yukon Kuskokwim Delta and the customer-owners at Southcentral Foundation for their partnership and participation in the research studies described in this dissertation.

Finally, I would like to thank my friends and family for their love and encouragement during this journey. To my husband, Andres, thank you for always encouraging me to follow my dreams and for being my biggest supporter and my best friend. I cannot imagine going through the past five years without you.

Chapter 1. INTRODUCTION

(Part of this chapter was published as a review article "*P450* Pharmacogenetics in Indigenous North American Populations." (2018) Journal of Personalized Medicine 8(1): 9)

1.1 BACKGROUND

1.1.1 *Introduction*

Alaska Native and American Indian (AN/AI) people are under-represented in health research, particularly genomic research, but are likely to have unique genetic variation profiles that may critically impact their response to drug therapy. Importantly, AN/AI people have a higher incidence of cardiovascular disease, particularly stroke, compared to other world populations, necessitating a need to optimize in this population disease prevention therapies. Warfarin is an oral anticoagulant approved for treatment of thromboembolic disorders including stroke prophylaxis. While very effective, warfarin therapy requires extensive monitoring due to its narrow therapeutic index and wide inter-individual response to treatment. Warfarin pharmacogenes encode enzymes that are involved in the drug's disposition through metabolism (CYP2C9), its pharmacological target (VKOR), activation by γ-glutamyl carboxylation of vitamin K clotting factors (GGCX), and catabolism of the GGCX cofactor vitamin K (CYP4F2 and CYP4F11). While many studies have investigated the functional consequences of warfarin pharmacogene variation and the potential value of prospective genetic testing in Caucasian, Asian, and African American populations, little is known about their contribution to variability in warfarin response in AN/AI people. In order to optimize clinical warfarin therapy and minimize adverse drug events in underrepresented AN/AI populations, these groups should be included in pharmacogenetic studies to inform on the clinical dosing of this drug and advance the goal of precision medicine.

1.1.2 *Precision Medicine*

Pharmacogenetics, a form of genomic and precision medicine, aims to establish how genetic variation can affect an individual's response to drugs, guiding the selection of the best drug and dose for a patient in order to improve healthcare quality [1]. The field of pharmacogenetics has the potential to improve health outcomes and reduce the cost of care by maximizing therapeutic success and minimizing the risk of adverse drug reactions or therapeutic failure at the population and potentially the individual level. However, a major issue in the translation of pharmacogenetic research into clinical practice is that existing databases, which feed warfarin dose titration algorithms (standard of care), have been populated from studies that lack significant ethnic and racial diversity.

Although diversity in genomic research, including pharmacogenetics, has increased in recent years, oversampling of populations of European ancestry continues to be a problem in the field. Indeed, the latest analysis of genome-wide association studies found that 81% of samples were from individuals of European ancestry [2]. The non-European portion was comprised of mostly Asian ancestry, leaving just 5% for the rest of the world's populations [2]. Failure to include diverse populations in genomic studies leads to a biased understanding of the health implications of genetic variation and resulting medical findings may be preferentially beneficial to patients of European ancestry. Increased attention to genetic data from diverse populations is required to give "everyone the best chance at good health" [3].

While much work has been conducted investigating the clinical importance of the cytochrome P450 (P450) gene variants, there is relatively little data specifically addressing P450 variation and its consequences in Indigenous populations [4]. Geographical isolation, unbalanced resource allocation, failure of researchers to include Indigenous communities in study design and

reluctance to participate in studies due to historical and recent research misconduct all contribute to under-representation of Indigenous populations in biomedical (including genetic) research studies [2,5-9]. However, it is important to characterize the unique genetic variation that exists in these historically isolated populations, because there are clinical implications of having uncharacterized genetic variation (phenotype misclassification), particularly with drug metabolizing enzymes.

1.1.3 *P450 Pharmacogenetics*

The *P450* genes encode a group of highly polymorphic enzymes that play a critical role in drug metabolism [10]. There are 57 *P450* genes in humans, with members of the CYP1, CYP2, and CYP3 families being responsible for most of the metabolic clearance of the approximately 75% of all drugs that are eliminated from blood by this process [11-13]. Variation in these *P450* genes can result in proteins with altered catalytic activity or abundance (referred to collectively hereafter as 'enzyme activity'), leading to high inter-individual variability in systemic drug concentration and pharmacological response without dose adjustment [14]. Gene sequence changes (single nucleotide variation and structural variation, referred to collectively as alleles) that lead to altered P450 enzyme activity can be classified into four phenotypic groups: poor metabolizer (PM), intermediate metabolizer (IM), extensive metabolizer (EM), and ultra-rapid metabolizer (UM). PMs are generally homozygous for a variant allele that causes a complete loss of enzyme activity (null allele), IMs can be heterozygous for a reference allele and a null allele or a combination of reduced function alleles, EMs have two reference activity alleles, and UMs have multiple copies of the reference allele or an active variant that increases total enzyme activity, relative to the reference state. Enzyme activity has proportional effects on the intrinsic parent drug clearance and

thus is inversely related to systemic parent drug exposure, which drives most pharmacological responses.

Indigenous populations can have unique variants that confer altered enzyme activity as well as distinct variant allele frequencies, both of which are related to historic geographical isolation and arise due to genetic drift, selective pressures, and the founder effect. The population-level differences in *P450* allele frequencies in Indigenous peoples require consideration to avoid or reduce negative clinical outcomes that include phenotypic misclassification and inappropriate drug utilization, further contributing to health care disparities.

1.1.4 *Inclusion of AN/AI People in Pharmacogenetic Research*

This dissertation focuses on unique variation found in the AN/AI population and its potential impact on drug metabolism and response. The 2010 United States Census reported that 5.2 million AN/AI people live in the United States, with the AN/AI population having grown 39% in the preceding decade [15]. For discussion purposes (and statistical power in experimental design presented later), I have intentionally clustered these peoples by their shared heritage and geographical proximity, though it is acknowledged that each tribe has their own unique histories, languages, and cultural traditions.

The Northwest-Alaska Pharmacogenomic Research Network (NWA-PGRN) was created to address the dearth of genetic data from AN/AI people with the goal of identifying and characterizing novel variation that contributes to drug disposition and pharmacological response in this population. NWA-PGRN is focused on applying both novel and generalizable approaches to predicting gene-environment-drug interactions and is particularly interested in assessing whether pharmacogenetic testing provides unique advantages in AN/AI communities, who might

have less therapeutic monitoring due to predominantly rural home settings. The current clinical focus of pharmacogenetic research by NWA-PGRN is on cardiovascular disease.

Individuals of AN/AI descent are more likely to be affected by cardiovascular disease and die from stroke at an earlier age, compared to other world populations [16]. Socioeconomic status and gene-environment interactions are potential causes for these observed epidemiological differences, contributing to disease states and health outcomes following medical intervention [17,18]. To better understand the biological and environmental factors affecting cardiovascular disease and its treatment, academic investigators at the University of Alaska Fairbanks (UAF), most now at Oregon Health Sciences Center (OHSU), established genetic and environmental research partnerships with rural communities in the Yukon-Kuskokwim River Delta (YK Delta), that is home to a relatively large (~23,000) Yup'ik AN population. Similar health research is being spearheaded by the Southcentral Foundation (SCF), a tribally owned and operated regional health corporation that provides healthcare services to ~65,000 AN/AI customer-owners. Partnerships between the University of Washington, UAF/OHSU, and SCF were developed to investigate the impact of genetic variation in AN/AI populations of Alaska on drug responses in patients with cardiovascular disease that are known to be influenced by genetic variation; one such example is warfarin anticoagulation therapy used to prevent thromboembolic events in at-risk patients. The investigation described in this dissertation proposal capitalized on these existing research collaborations.

1.1.5 *Warfarin Pharmacogenetics*

The oral vitamin K antagonist warfarin (Coumadin®) is used chronically in patients with atrial fibrillation to prevent venous thromboembolic events and stroke or for acute prevention of deep venous thrombosis and pulmonary embolism secondary to a transient risk factor [19]. While

very effective, warfarin therapy requires intensive monitoring and extensive dose titration due to its narrow therapeutic index and wide inter-individual (up to 30-fold), and intra-individual response to treatment, due in part to genetic variation [20]. To ensure adequate anticoagulation and avoid bleeding events, warfarin therapy is monitored measuring prothrombin time and using an international normalized ratio (INR) of prothrombin time. An INR of 2-3 or 2.5-3.5 is recommended for most clinical indications, whereas an INR above 4 is associated with increased risk of major, sometimes fatal, bleeding events [21]. While clinical and demographic factors contribute to the variability of warfarin's therapeutic response [22], the dose required to achieve therapeutic anticoagulation also differs across ethnicities [23], leading to the identification of genetic factors that could explain some of this variability. The pharmacogenes encoding enzymes that have been associated with variable warfarin disposition and pharmacological response are shown in **Figure 1.1**. The mechanism of action of warfarin involves depletion of endogenous vitamin K-dependent clotting factors (e.g., Factor II) through inhibition of the Vitamin K Epoxide Reductase (VKOR) enzyme. This lowers the steady-state level of reduced hepatic vitamin K, the essential cofactor for GGCX, and slows the rate of clotting factor post-translational modification, which generates " biologically active" proteins [24].

The activity of CYP2C9, VKORC1, CYP4F2, CYP4F11, and GGCX (in the liver) in AN/AI populations may differ from that found in other US and world populations. Indeed, the average daily dose of warfarin for long-term care (over 22 weeks of treatment) targeting for an INR between 2-3, for the AN/AI people at SCF was found to be lower (4.34 mg/day) than the average at other North American and European sites (5.19 mg/day) [25] using the same DAWN AC Anticoagulation Management Software (4S Information Systems Ltd., Cumbria, England). Approximately 60% of the observed variance in warfarin maintenance dose can be explained, with

clinical factors (patient age, body surface area, gender, and concurrent medications) responsible for ~20% [22], and genetic variation in *VKORC1, CYP2C9, CYP4F2,* and *GGCX* accounting for 25%, 10%, 2%, and 2%, respectively [26-31], in populations of European origin. However, the percent of dose variability explained by genotype differs across racial groups due to different allele frequencies in these populations, as has been demonstrated with both *VKORC1 [32,33]* and *CYP2C9 [33]*. For instance, while *VKORC1 -1639C>T* (rs9923231) decreases stable warfarin dose requirement across African American, AN/AI, Asian, and European populations, there is a stronger association in populations with higher frequencies of the variant allele [32,34]. Despite the associations with *VKORC1* genotype and stable warfarin dose across different populations, knowledge about the regulation of *VKORC1* and its interactions with other hepatocellular proteins remains incomplete. The ~40% unexplained variability in a stable therapeutic warfarin dose may be due to uncharacterized novel variation in known warfarin pharmacogenes [35] or variants in genes that regulate these warfarin pharmacogenes [36], as well as non-genetic factors such as dietary or environment interactions (intake of ω3 polyunsaturated fatty acids) [37].

Recently, full gene resequencing of *CYP2C9, VKORC1, CYP4F2, CYP4F11,* and *GGCX* in the AN/AI population identified novel coding-region single nucleotide variants (SNVs) and established population frequencies in the Yup'ik AN as well as an aggregate of AN/AI from SCF [38]. **Table 1.1** shows the minor allele frequencies (MAFs) in the Yukon-Kuskokwim delta Yup'ik AN population and the AN/AI population at SCF, compared to major global populations from 1000 Genomes [39,40]. A preponderance of the H1 haplotype (-1639 AA diplotype) predicts low warfarin maintenance dose in Yup'ik AN individuals [26,38], similar to the lower dose observed in Asian populations [41]. In addition, two novel coding-region *CYP2C9* SNVs in the Yup'ik and SCF populations, *CYP2C9M1L (M1L)* and *CYP2C9N218I (N218I),* and a relatively new and

poorly characterized variant, *CYP2C9P279T (P279T)*, were identified [38]. These three coding SNVs are present at relatively high frequencies in the Yup'ik AN population, where the MAFs of *M1L*, *N218I*, and *P279T* are 6.3%, 4.4%, and 1.9%, respectively (Error! Reference source not found.).

Figure 1.2 depicts the human *CYP2C9* gene and the commonly studied SNVs, *CYP2C9*2* and **3*, as well as the novel coding-region variants, *M1L*, *N218I,* and *P279T.* The switch from a methionine to leucine at the start codon for *M1L* is predicted to prevent protein synthesis, conferring a CYP2C9 poor metabolizer phenotype *in vivo*. *N218I* was also found to be common in some SCF AN/AI subgroups (i.e., Yup'ik and Athabascan) and is predicted to impair the catalytic function of the enzyme and reduce *in vivo* metabolic clearance of CYP2C9 substrates. If each of the new alleles impact catalytic activity, in aggregate, a sizeable proportion of the Yup'ik and SCF populations are expected to exhibit warfarin sensitivity from these new *CYP2C9* alleles alone. Finally, the *CYP4F2*3* variant, significantly associated with higher average daily warfarin dose [42], was found at a higher frequency (51%) in the Yup'ik population that that seen in almost all world populations. Together, these markedly different warfarin pharmacogene allele frequencies, compared to European ancestry populations, and predicted phenotypes of coding variation, support the need to investigate the functional activity of these variants to fully characterize their potential to affect the pharmacokinetics and pharmacodynamics of warfarin in the AN/AI population. This lack of foundational genotype-phenotype data pertaining to the AN/AI population could affect the clinical utility of pharmacogenetic screening.

The projects described in this dissertation proposal investigate the missing heritability in warfarin dose variance and inform on personalized warfarin therapy for AN/AI patients living in remote communities. Work aimed at characterizing novel variation in *CYP2C9* in the AN/AI

population is described Chapters 2 and 3, a project evaluating *CYP2C9, VKOR, CYP4F2, CYP4F11*, and *GGCX* gene variation and warfarin dose requirement in an AN/AI population is presented in Chapter 4, and a pathway-based analysis identifying potential regulators of *VKORC1* that influence VKOR synthesis, and therefore warfarin dose requirement, is explored in Chapter 5.

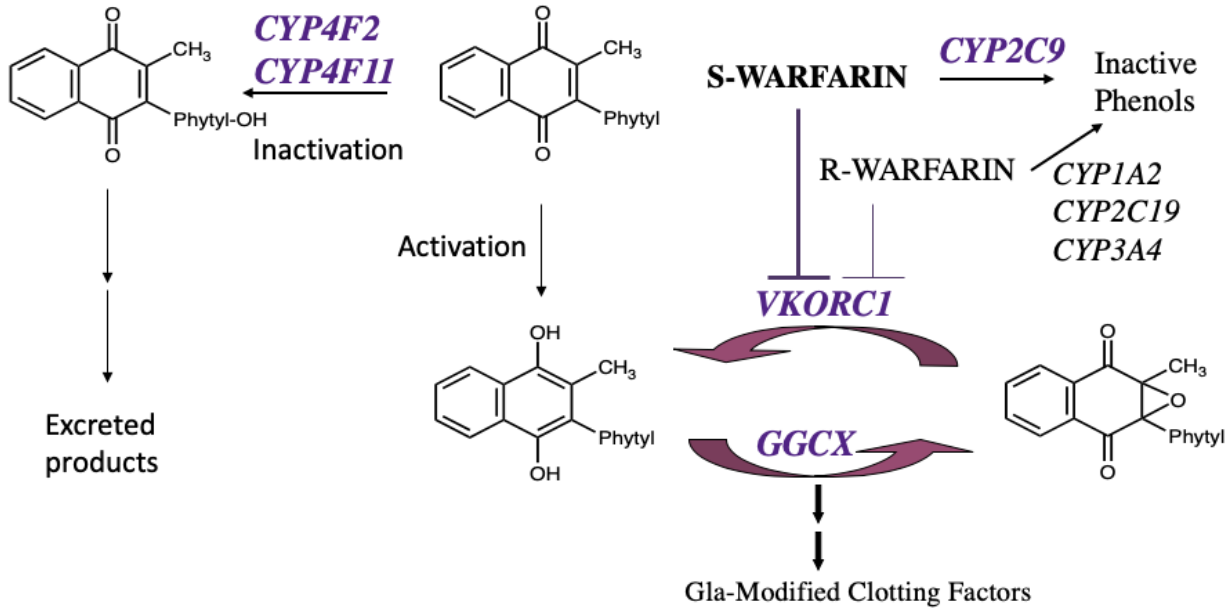

Figure 1.1. Vitamin K cycle-associated genes that impact warfarin pharmacokinetics and pharmacodynamics.

Table 1.1. Yup'ik AN and AN/AI population from SCF and world populations from 1000 Genomes[#] [39].

SNV	Yup'ik MAF (%)	AN/AI (SCF) MAF (%)	Asian[#] MAF (%)	European[#] MAF (%)	African American[#] MAF (%)
VKORC1 -1639G>A	78*	60*	92*	40	7
VKORC1 1173C>T	78*	60*	92*	40	7
CYP2C9 M1L	6	1	-	-	-
*CYP2C9*2*	<1	5	<1	12	2
CYP2C9 N218I	4	1	-	-	-
*CYP2C9*29*	2	0	-	-	-
*CYP2C9*3*	2	3	4	6	<1
*CYP4F2*3*	51*	32	21	27	9
*CYP4F2*2*	4	11	7	16	24
CYP4F11 R276C	<1	9	7	27	29

The reported allele frequency is for the variant allele, an alternative allele to the reference allele, which is defined by the global population. *The variant allele frequency in the population is the major allele for the racial/ethnic group.

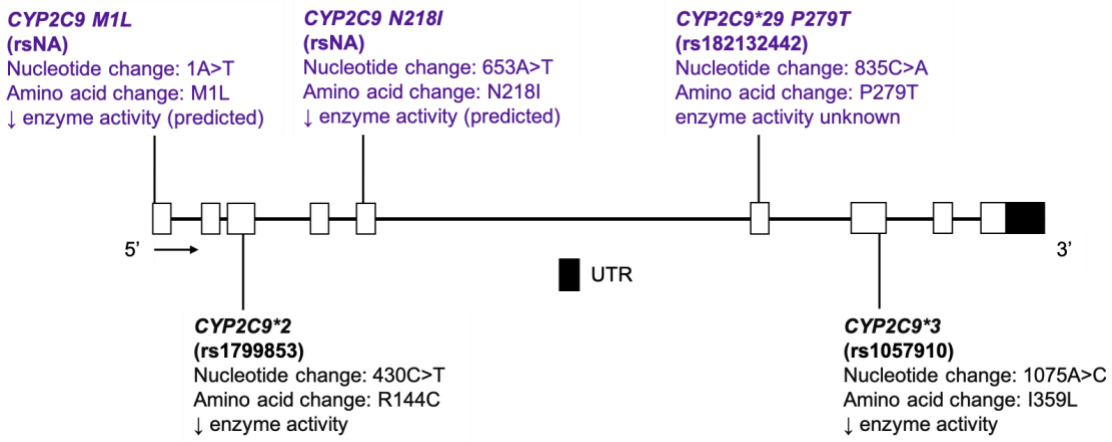

Figure 1.2. The human *CYP2C9* gene and commonly studied (*CYP2C9*2,*3*) as well as novel (*M1L, N218I, P279T*) variants identified in an Alaska Native population.
Open boxes represent exons, lines represent introns, and the shaded box represents untranslated region (UTR). Adapted from Henderson et al. 2018 [32].

1.2 Hypothesis and Specific Aims

Novel variants or allele frequencies in the vitamin K gene pathway contribute to the safety and efficacy of therapeutic inhibitors of the clotting cascade in AN/AI populations.

The specific aims of this dissertation research are:

Specific Aim 1: Obtain *in vitro* functional data on novel *CYP2C9* variants identified in AN people to inform on their potential to affect probe substrate pharmacokinetic parameters.

Specific Aim 2: Characterize the catalytic efficiency of CYP2C9 M1L *in vivo* to determine if disruption of the translation start codon confers a poor metabolizer phenotype.

Specific Aim 3: Determine whether inheritance of *CYP2C9, VKORC1, CYP4F2, CYP4F11,* and *GGCX* gene variants in AN/AI populations affects the dose of warfarin needed to

achieve a therapeutic INR in order to inform on the utility of personalized warfarin therapy for AN/AI communities.

Specific Aim 4: Determine associations between *VKORC1* mRNA levels and genome-wide transcript and microRNA variation.

1.3 REFERENCES

1. Relling, M.V.; Evans, W.E. Pharmacogenomics in the clinic. *Nature* **2015**, *526*, 343-350, doi:10.1038/nature15817.
2. Popejoy, A.B.; Fullerton, S.M. Genomics is failing on diversity. *Nature* **2016**, *538*, 161-164, doi:10.1038/538161a.
3. Collins, F.S.; Varmus, H. A new initiative on precision medicine. *N Engl J Med* **2015**, *372*, 793-795, doi:10.1056/NEJMp1500523.
4. Jaja, C.; Burke, W.; Thummel, K.; Edwards, K.; Veenstra, D.L. Cytochrome p450 enzyme polymorphism frequency in indigenous and native american populations: a systematic review. *Community Genet* **2008**, *11*, 141-149, doi:10.1159/000113876.
5. Boyer, B.B.; Dillard, D.; Woodahl, E.L.; Whitener, R.; Thummel, K.E.; Burke, W. Ethical issues in developing pharmacogenetic research partnerships with American Indigenous communities. *Clin Pharmacol Ther* **2011**, *89*, 343-345, doi:10.1038/clpt.2010.303.
6. Drabiak-Syed, K. Lessons from Havasupai Tribe v. Arizona State University Board of Regents: Recognizing Group, Cultural, and Dignitary Harms as Legitimate Risks Warranting Integration into Research Practice. *Journal of Health & Biomedical Law* **2010**, *6*, 175-226.
7. Harmon, A. Indian Tribe Wins Fight to Limit Research of Its DNA The New York Times: 2010.
8. Foulks, E.F. Misalliances in the Barrow Alcohol Study. *Am Indian Alsk Native Ment Health Res* **1989**, *2*, 7-17.
9. Wiwchar, D. Nuu-chah-nulth blood returns to west coast. Ha-Shilth-Sa: 2004.
10. Zanger, U.M.; Turpeinen, M.; Klein, K.; Schwab, M. Functional pharmacogenetics/genomics of human cytochromes P450 involved in drug biotransformation. *Anal Bioanal Chem* **2008**, *392*, 1093-1108, doi:10.1007/s00216-008-2291-6.
11. Guengerich, F.P. Cytochrome p450 and chemical toxicology. *Chem Res Toxicol* **2008**, *21*, 70-83, doi:10.1021/tx700079z.
12. Zanger, U.M.; Schwab, M. Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol Ther* **2013**, *138*, 103-141, doi:10.1016/j.pharmthera.2012.12.007.
13. Wienkers, L.C.; Heath, T.G. Predicting in vivo drug interactions from in vitro drug discovery data. *Nat Rev Drug Discov* **2005**, *4*, 825-833, doi:10.1038/nrd1851.
14. Zhou, S.F.; Liu, J.P.; Chowbay, B. Polymorphism of human cytochrome P450 enzymes and its clinical impact. *Drug Metab Rev* **2009**, *41*, 89-295, doi:10.1080/03602530902843483.
15. Norris, T.; Vines, P.L.; Hoeffel, E.M. The American Indian and Alaska Native Population: 2010. In *2010 Census Briefs*, U.S. Department of Commerce Economics and Statistics Administration U.S. CENSUS BUREAU: 2012.
16. Jolly, S.E.; Howard, B.V.; Umans, J.G. Cardiovascular Disease Among Alaska Native Peoples. *Curr Cardiovasc Risk Rep* **2013**, *7*, doi:10.1007/s12170-013-0362-5.
17. Horner, R.D.; Day, G.M.; Lanier, A.P.; Provost, E.M.; Hamel, R.D.; Trimble, B.A. Stroke mortality among Alaska Native people. *Am J Public Health* **2009**, *99*, 1996-2000, doi:10.2105/AJPH.2008.148221.

18. Baye, T.M.; Abebe, T.; Wilke, R.A. Genotype-environment interactions and their translational implications. *Per Med* **2011**, *8*, 59-70, doi:10.2217/pme.10.75.
19. Coumadin® (warfarin sodium) [package insert]. Princeton, NJ: Bristol-Myers Squibb. 2017.
20. Rettie, A.E.; Tai, G. The pharmocogenomics of warfarin: closing in on personalized medicine. *Mol Interv* **2006**, *6*, 223-227, doi:10.1124/mi.6.4.8.
21. Hirsh, J.; Fuster, V.; Ansell, J.; Halperin, J.L.; Association, A.H.; Foundation, A.C.o.C. American Heart Association/American College of Cardiology Foundation guide to warfarin therapy. *Circulation* **2003**, *107*, 1692-1711, doi:10.1161/01.CIR.0000063575.17904.4E.
22. Gage, B.F.; Eby, C.; Milligan, P.E.; Banet, G.A.; Duncan, J.R.; McLeod, H.L. Use of pharmacogenetics and clinical factors to predict the maintenance dose of warfarin. *Thromb Haemost* **2004**, *91*, 87-94, doi:10.1160/TH03-06-0379.
23. Dang, M.T.; Hambleton, J.; Kayser, S.R. The influence of ethnicity on warfarin dosage requirement. *Ann Pharmacother* **2005**, *39*, 1008-1012, doi:10.1345/aph.1E566.
24. Suttie, J.W. The biochemical basis of warfarin therapy. *Adv Exp Med Biol* **1987**, *214*, 3-16.
25. Schilling, B. Anticoagulation care for Alaska Native customer-owners within the Nuka model of care, 7th ed. Dawn AC Anticoagulation Management Software North American User Group Meeting: La Jolla, CA, 2013.
26. Rieder, M.J.; Reiner, A.P.; Gage, B.F.; Nickerson, D.A.; Eby, C.S.; McLeod, H.L.; Blough, D.K.; Thummel, K.E.; Veenstra, D.L.; Rettie, A.E. Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose. *N Engl J Med* **2005**, *352*, 2285-2293, doi:10.1056/NEJMoa044503.
27. Schwarz, U.I.; Stein, C.M. Genetic determinants of dose and clinical outcomes in patients receiving oral anticoagulants. *Clin Pharmacol Ther* **2006**, *80*, 7-12, doi:10.1016/j.clpt.2006.04.008.
28. Caldwell, M.D.; Awad, T.; Johnson, J.A.; Gage, B.F.; Falkowski, M.; Gardina, P.; Hubbard, J.; Turpaz, Y.; Langaee, T.Y.; Eby, C., et al. CYP4F2 genetic variant alters required warfarin dose. *Blood* **2008**, *111*, 4106-4112, doi:10.1182/blood-2007-11-122010.
29. Rieder, M.J.; Reiner, A.P.; Rettie, A.E. Gamma-glutamyl carboxylase (GGCX) tagSNPs have limited utility for predicting warfarin maintenance dose. *J Thromb Haemost* **2007**, *5*, 2227-2234, doi:10.1111/j.1538-7836.2007.02744.x.
30. Au, N.; Rettie, A.E. Pharmacogenomics of 4-hydroxycoumarin anticoagulants. *Drug Metab Rev* **2008**, *40*, 355-375, doi:10.1080/03602530801952187.
31. Johnson, J.A.; Caudle, K.E.; Gong, L.; Whirl-Carrillo, M.; Stein, C.M.; Scott, S.A.; Lee, M.T.; Gage, B.F.; Kimmel, S.E.; Perera, M.A., et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Pharmacogenetics-Guided Warfarin Dosing: 2017 Update. *Clin Pharmacol Ther* **2017**, *102*, 397-404, doi:10.1002/cpt.668.
32. Henderson, L.M.; Claw, K.G.; Woodahl, E.L.; Robinson, R.F.; Boyer, B.B.; Burke, W.; Thummel, K.E. P450 Pharmacogenetics in Indigenous North American Populations. *J Pers Med* **2018**, *8*, doi:10.3390/jpm8010009.
33. Limdi, N.A.; Brown, T.M.; Yan, Q.; Thigpen, J.L.; Shendre, A.; Liu, N.; Hill, C.E.; Arnett, D.K.; Beasley, T.M. Race influences warfarin dose changes associated with genetic factors. *Blood* **2015**, *126*, 539-545, doi:10.1182/blood-2015-02-627042.

34. Limdi, N.A.; Wadelius, M.; Cavallari, L.; Eriksson, N.; Crawford, D.C.; Lee, M.T.; Chen, C.H.; Motsinger-Reif, A.; Sagreiya, H.; Liu, N., et al. Warfarin pharmacogenetics: a single VKORC1 polymorphism is predictive of dose across 3 racial groups. *Blood* **2010**, *115*, 3827-3834, doi:10.1182/blood-2009-12-255992.
35. Manolio, T.A.; Collins, F.S.; Cox, N.J.; Goldstein, D.B.; Hindorff, L.A.; Hunter, D.J.; McCarthy, M.I.; Ramos, E.M.; Cardon, L.R.; Chakravarti, A., et al. Finding the missing heritability of complex diseases. *Nature* **2009**, *461*, 747-753, doi:10.1038/nature08494.
36. Allyn-Feuer, A.; Ade, A.; Luzum, J.A.; Higgins, G.A.; Athey, B.D. The pharmacoepigenomics informatics pipeline defines a pathway of novel and known warfarin pharmacogenomics variants. *Pharmacogenomics* **2018**, *19*, 413-434, doi:10.2217/pgs-2017-0186.
37. Au, N.T.; Ryman, T.; Rettie, A.E.; Hopkins, S.E.; Boyer, B.B.; Black, J.; Philip, J.; Yracheta, J.; Fohner, A.E.; Reyes, M., et al. Dietary Vitamin K and Association with Hepatic Vitamin K Status in a Yup'ik Study Population from Southwestern Alaska. *Mol Nutr Food Res* **2017**, doi:10.1002/mnfr.201700746.
38. Fohner, A.E.; Robinson, R.; Yracheta, J.; Dillard, D.A.; Schilling, B.; Khan, B.; Hopkins, S.; Boyer, B.; Black, J.; Wiener, H., et al. Variation in genes controlling warfarin disposition and response in American Indian and Alaska Native people: CYP2C9, VKORC1, CYP4F2, CYP4F11, GGCX. *Pharmacogenet Genomics* **2015**, *25*, 343-353, doi:10.1097/FPC.0000000000000143.
39. Abecasis, G.R.; Auton, A.; Brooks, L.D.; DePristo, M.A.; Durbin, R.M.; Handsaker, R.E.; Kang, H.M.; Marth, G.T.; McVean, G.A.; Consortium, G.P. An integrated map of genetic variation from 1,092 human genomes. *Nature* **2012**, *491*, 56-65, doi:10.1038/nature11632.
40. Consortium, G.P.; Auton, A.; Brooks, L.D.; Durbin, R.M.; Garrison, E.P.; Kang, H.M.; Korbel, J.O.; Marchini, J.L.; McCarthy, S.; McVean, G.A., et al. A global reference for human genetic variation. *Nature* **2015**, *526*, 68-74, doi:10.1038/nature15393.
41. Takahashi, H.; Wilkinson, G.R.; Nutescu, E.A.; Morita, T.; Ritchie, M.D.; Scordo, M.G.; Pengo, V.; Barban, M.; Padrini, R.; Ieiri, I., et al. Different contributions of polymorphisms in VKORC1 and CYP2C9 to intra- and inter-population differences in maintenance dose of warfarin in Japanese, Caucasians and African-Americans. *Pharmacogenet Genomics* **2006**, *16*, 101-110.
42. McDonald, M.G.; Rieder, M.J.; Nakano, M.; Hsia, C.K.; Rettie, A.E. CYP4F2 is a vitamin K1 oxidase: An explanation for altered warfarin dose in carriers of the V433M variant. *Mol Pharmacol* **2009**, *75*, 1337-1346, doi:10.1124/mol.109.054833.

Chapter 2. *IN VITRO* FUNCTIONAL CHARACTERIZATION OF NOVEL *CYP2C9* VARIANTS IDENTIFIED IN AN ALASKA NATIVE POPULATION

(A version of this chapter is currently a draft manuscript "Heterologous Expression and Functional Characterization of Novel CYP2C9 Variants in an Alaska Native Population." (2019) with planned submission to Drug Metabolism and Disposition.)

2.1 INTRODUCTION

Although gene duplications/deletion for *CYP2C9* have not been reported, an array of single nucleotide coding polymorphisms that confer substantial functional deficits to the enzyme that alter drug elimination and response are known [1]. The most widely studied of these are *CYP2C9*2* and **3*, which are present at high allele frequencies in people of European ancestry but are much less prevalent in other ethnic populations [2]. *CYP2C9*2* is essentially absent in East Asians and functionally defective *CYP2C9*5, *6, *8,* and **11* alleles are the predominant reduced or loss of function alleles in people of African ancestry. The clinical impact of the latter alleles is often overlooked. For example, they were not included in the analysis for the recent USA-wide COAG randomized clinical trial [3], that was intended to test whether pharmacogenetic information could improve clinical outcomes for patients receiving warfarin anticoagulation therapy. A later study that did account for these alleles demonstrated improved outcomes for African-Americans being treated with this narrow therapeutic index drug [4]. These findings demonstrate the need to consider population-selective alleles in pharmacogenetic studies to avoid healthcare disparities in under-represented populations [5].

In the Alaska Native (AN) population, ultra-rare or novel *CYP2C9* variants, *M1L, N218I,* and *P279T,* are expressed with higher frequencies than the well-characterized *CYP2C9*2* and *CYP2C9*3* alleles [6]. The altered initiation start site allele, *Met1Leu (M1L),* was identified in multiple AN tribal groups, with the highest minor allele frequency (6.3%) in Yup'ik residents of the Yukon-Kuskokwim delta (N=350). In the same AN sub-population, *Asn218Ile (N218I)* and *Pro279Thr* (*P279T*) variants were present with allele frequencies of 3.8% and 2.1%, respectively, and the *Asn218Ile* variant was found at a minor allele frequency of 5.4% in Athabascan AN sub-populations living in central Alaska [6]. The *Pro279Thr* variant (*CYP2C9*29*) was first detected in Han Chinese with a low allele frequency of 0.19 [7] and studied *in vitro* using enzyme variant expressing insect cells. It exhibited catalytic efficiencies for flurbiprofen and phenytoin metabolism that were maintained within a factor of two, relative to wild-type CYP2C9 [8,9]. However, no functional data exist for the *N218I* and *M1L* variants.

This chapter reports on *in vitro* functional characterization of the CYP2C9 P279T, N218I, and M1L enzymes. HepG2 cells were used to assess RNA and protein expression in mammalian cells, while *E.coli* was used for high level expression suitable for purification and detailed catalytic studies with CYP2C9 substrates after reconstitution with required coenzymes and lipid. Enzyme expression, transcript and protein analysis, as well as some of the functional assessment studies were conducted by personnel in the lab of Dr. Allan Rettie (Dr. Matthew McDonald, Dr. Sutapa Ray, John Kowalski, Amanda Johnson, and Dr. Catherine Yeung), Department of Medicinal Chemistry, University of Washington, but is presented in this chapter for completeness and clarity.

2.2 METHODS

2.2.1 *General Reagents*

The 6-hydroxywarfarin-d5, 7-hydroxywarfarin-d5 and 5-(p-hydroxyphenyl)-5-phenylhydantoin-d5 (p-HPPH-d5) chemical standards (as well as their unlabeled analogs) were previously synthesized, as were the unlabeled o-HPPH and m-HPPH phenytoin metabolite standards. Unlabeled flurbiprofen, 4'-OH-flurbiprofen, *(S)*-naproxen, and racemic O-desmethyl naproxen-d_3 were purchased from Toronto Research Chemicals (Ontario, Canada). DLPC was procured from Avanti Polar Lipids, Inc. (Alabaster, AL), and unilamellar vesicles (i.e. liposomes) were freshly prepared, immediately prior to their use in incubations with purified CYP2C9 protein, by repeatedly passing a 1 mM DLPC solution (in 100 mM potassium phosphate buffer, pH 7.4) through a 0.1 μm nucleopore membrane using a mini-extruder apparatus according to the manufacturer's protocol (Avanti). Cytochrome P450 oxido-reductase (CPR) and cytochrome b5 utilized in purified CYP2C9 variant protein incubations were expressed and purified by the Rettie Lab. Organic solvents and buffer salts (potassium phosphates, NaCl, etc.) were obtained from Fisher Scientific (Fair Lawn, NJ) and all other chemicals (including racemic warfarin, phenytoin, O-desmethyl naproxen, β-NADPH, etc.) were purchased from Sigma Aldrich (St. Louis, MO).

2.2.2 *Expression of CYP2C9 Variants in HepG2 Cells and E. coli*

HepG2 (liver-derived) cells that stably expressed CYP2C9 WT, M1L, N218I, P279T, and I359L (*3), generated using lentivirus mediated heterologous expression, were obtained from the Rettie Lab. CYP2C9 reference (WT) and variant genes were cloned into lentiviral vector, transfected into HEK293T cells, and the viral supernatant was used to transduce desired eukaryotic cells. Gene expression was determined by RT-PCR. The mRNA expression of each sample was

normalized to its own relative fold change, determined by comparing expression to a representative pool of seven human liver tissue samples (all CYP2C9 WT). A 30 μg protein aliquot of HepG2 cell lysate (CYP2C9 WT, M1L, N218I, P279T, I359L, and EGFP lysate, a mock-transduced negative control) and 0.25 pmol rCYP2C9 Supersomes (positive control) were evaluated by Western blot using an anti-CYP2C9 polyclonal antibody. Ms. Amanda Johnson performed this analysis.

Recombinant plasmids were used to transform *E. coli* DH5αF′IQ to express the mutant proteins, CYP2C9 M1L, N218I, and P279T, which were obtained from the Rettie Lab. Proteins were expressed as previously described [10] without the use of stabilizing ligands. Purification was performed via nickel-nitrilotriacetic acid Superflow (Qiagen, Valencia, CA) with imidazole elution. The protein was dialyzed (3 x 1 L) with a buffer containing 100 mM KPi, 20% glycerol, 1 mM EDTA (pH 7.4) and concentrated in a centrifugal concentrator (Amicon Ultra 30K NMWL, Millipore Sigma, Burlington, MA). Dr. Cathy Yeung oversaw this procedure. Dr. Matthew McDonald performed tryptic digest analysis of recombinant protein using high resolution LC/MS to confirm the amino acid sequence of the variant proteins.

2.2.3 *Quantitative CO Binding Assay*

Carbon monoxide binding spectra by reduced CYP2C9 protein variants were recorded on an Olis modernized Aminco DW-2 spectrophotometer (Olis, Bogart, GA). Purified protein was diluted 10-fold in 100 mM KPi buffer, pH 7.4. Sodium dithionite (excess, powder) was added and the reduced protein mix was then split between sample and reference quartz cuvettes before taking a baseline scan from 400 to 500 nm, at 25°C. Carbon monoxide gas was bubbled to saturation through the sample cuvette, and difference scans were taken of the CO-bound enzyme. Holo-CYP2C9 variant concentrations were calculated according to Beer's law taking the absorbance

difference between the peak maximum (at ~450 nm) and the baseline (at 490 nm) and using an extinction coefficient of 91 $mM^{-1}cm^{-1}$.

2.2.4 *Enzyme Reconstitution and Assay Incubation Conditions*

Master protein mixes were prepared from concentrated stock solutions of purified CYP2C9 variant protein (1 molar equivalent), CPR (2 molar equivalents) and DLPC (160 molar equivalents - added from a freshly extruded 1mM aqueous stock solution). Mixtures were incubated on ice for 30 minutes and then diluted with reaction buffer (100 mM potassium phosphate, pH 7.4). One molar equivalent of cytochrome b5 was added and the mixtures were incubated on ice for an additional 5 minutes. The master mixes were then aliquoted for individual incubation reactions and substrate was added. A standard incubation reaction contained 10 pmol of CYP2C9 variant, 20 pmol CPR, 10 pmol cytochrome b5 and substrate at a variable concentration (added from 100x concentrated stock solutions made up in 1:1 methanol/water) in a 250 μL final incubation volume. For kinetic experiments, a substrate concentration range of 100 nM to 100 μM was used for warfarin, 1 to 250 μM for phenytoin, 1 to 300 μM for flurbiprofen, and 5 to 1800 μM for naproxen. Enzyme and substrate were pre-incubated at 37° C/70 rpm in a shaking water bath for 3 minutes prior to initiation of the reaction with addition of NADPH (1 mM final concentration). Twenty minutes later, warfarin and phenytoin incubations were quenched with addition of 10 μL of ice cold 70% $HClO_4$ and internal standards were added (10 pmol each of 6- and 7-hydroxywarfarin-d_5 were added for warfarin metabolic experiments, while 10 pmol of p-hydroxyphenyl-5-phenylhydantoin-d_5 (p-HPPH-d_5) were added to incubations where phenytoin was used as the substrate). Flurbiprofen and naproxen incubations were quenched with the addition of 1 mL of ice cold methanol, containing 2% formic acid after 20 minutes of incubation, then internal standards were added (200 ng of O-desmethyl naproxen were added for flurbiprofen metabolic experiments,

while 200 ng of O-desmethyl naproxen-d_3 was added to incubations where naproxen was used as substrate). The reaction mixtures were centrifuged to remove protein and supernatants were analyzed either by LC-MS/MS for warfarin and phenytoin (performed by Dr. Matthew McDonald), or LC/MS for flurbiprofen and naproxen.

Calibration curves for the quantitation of warfarin and phenytoin metabolites were prepared by spiking variable amounts of either the unlabeled 6- and 7-hydroxywarfarins or the unlabeled hydroxy phenytoin (i.e. p-HPPH, m-HPPH and o-HPPH), respectively, into 250 μL volumes of potassium phosphate buffer in order to generate standard mixtures with final metabolite concentrations of between 1 and 1000 nM. Calibration curves for flurbiprofen and naproxen metabolites were prepared by spiking variable amounts of, respectively, either the unlabeled 4-hydroxy flurbiprofen or the unlabeled O-desmethyl naproxen, into 250 μL volumes of potassium phosphate buffer in order to generate standard mixtures with final concentrations of between 0.1 and 30 μM. These standard solutions, prepared in duplicate for each metabolite concentration, were immediately worked-up and analyzed in an identical fashion to that described for the enzyme incubation samples above.

2.2.5 *Analysis of Metabolites by LC-MS*

Flurbiprofen and naproxen metabolite concentrations were assessed by LC/MS on an Agilent 1956B single quadrupole mass spectrometer coupled with an Agilent 1200 series (Santa Clara, CA) liquid chromatography system. Chromatographic separation was achieved on a Luna C18 (2 mm x 50 mm x 5μm) column (Torrence, CA) with a mobile phase flow rate of 0.3 mL/min. The mobile phase consisted of 10 mM ammonium formate (A, pH 3.5) and methanol (B), and linear gradients were applied with B% increasing from 40% to 80% between 3 and 8 minutes and decreasing to 40% at 9 minutes. Column temperature was maintained at 35°C. The electrospray

ionization source of the mass spectrometer was operated in positive ion mode. Quantitation was achieved by selected ion monitoring of the following ion channels: *m/z* = 278.1 for 4-hydroxy flurbiprofen, *m/z* = 234.1 for *(S)*-6-O-desmethyl naproxen and *m/z* = 237.1 for racemic O-desmethyl naproxen-d_3. Data acquisition and analysis were performed using the Agilent MassHunter software. Calibration curves were constructed by plotting the peak area ratio of each compound to the respective internal standard against a range of concentrations.

LC-MS/MS analyses for *(S)*-warfarin and phenytoin were conducted by Dr. Matthew McDonald on a Waters Xevo TQ-S Tandem Quadrupole Mass Spectrometer (Waters Co., Milford, MA) coupled to an ACQUITY Ultra Performance LC™ (UPLC™) System with integral autoinjector (Waters). The Xevo was operated in ESI^+-MS/MS (SRM) mode at a source temperature of 150°C and a desolvation temperature of 350°C. The following mass transitions were monitored in separate ion channels for the various oxidative warfarin metabolites/standards: *m/z* 325 > 179 (6- and 7-hydroxywarfarins-d_0) and *m/z* 330 > 179 (6- and 7-hydroxywarfarins-d_5); and phenytoin metabolites/standards: *m/z* 269 > 198 (meta- and p-HPPH-d_0), *m/z* 269 > 209 (o-HPPH) and *m/z* 274 > 203 (p-HPPH-d_5). Optimized cone voltages and collision energies were set to 25 V and 15 eV for all metabolites and standards of warfarin, while the cone voltage was set to 35 V with collision energies of 15, 15 or 20 eV for the phenytoin metabolites p-HPPH (both d_0 and d_5-labeled), m-HPPH and o-HPPH, respectively. Metabolic products from the warfarin incubations were separated on an Acquity BEH Phenyl, 1.7 μ, 2.1 x 150 mm UPLC column (Waters, Corp) using an isocratic gradient of 45% solvent A (0.1% aqueous formic acid) and 55% solvent B (methanol), with a constant flow rate of 0.35 mL/min. Phenytoin metabolites were separated using this same BEH Phenyl UPLC column with a solvent gradient of water (solvent A) and acetonitrile (solvent B), both of which contained 0.1% formic acid, running at a flow rate of 0.3 mL/min.

Initially, solvent B was set to 28%, where it was maintained for 4.5 minutes, then increased linearly to 95% over 0.5 minutes where it was left for an additional 1.5 minutes. Metabolites were quantified through comparison of their peak area ratios (relative to either the 6- and 7-hydroxywarfarin-d_5 or p-HPPH-d_5 internal standard peak areas) to calibration curves using linear regression analysis.

2.2.6 *Kinetic Data Analysis*

Kinetic experiments for the assessment of CYP2C9 probe substrate metabolism were performed in triplicate, with duplicate data points for each substrate concentration. GraphPad Prism v.6 was used to estimate K_m and V_{max} parameters. Data presented are means ± S.D. of the three separate determinations unless otherwise stated.

2.3 RESULTS

2.3.1 *Expression of CYP2C9 Variants in HepG2 Cells*

Robust transcription of *CYP2C9* was observed across all cell lines except the *EGFP* negative control, as expected (**Figure 2.1**). While mRNA expression of the *CYP2C9 M1L, N218I,* and *P279T* gene variants and the reference (wildtype) protein in HepG2 cells were similar, the M1L variant protein was undetectable (**Figure 2.2**). This is expected, due to disruption of the *CYP2C9* start codon.

2.3.2 *Expression and Purification of CYP2C9 Variants in E. coli*

Concentrations for the CYP2C9 protein variants were determined by UV-vis spectroscopy using a quantitative CO binding assay. Upon binding carbon monoxide, all three enzymes showed an identical λ_{max} at 449 nm. Enzyme quantitation was performed in triplicate for each protein and

the mean specific content was found to be 19.4 ± 2.2, 7.0 ± 0.3 and 2.5 ± 0.6 μM, respectively, for the CYP2C9 WT, P279T and N218I protein variants.

2.3.3 *Comparison of Probe Substrate Metabolism by CYP2C9 Variant Enzymes*

Substrate saturation kinetic experiments were performed, in triplicate, for the enzyme variants using a reconstituted system that consisted of purified CYP2C9 protein, P450 oxidoreductase, cytochrome b5 and DLPC lipid (**Figure 2.3**). K_m and V_{max} values were determined for each variant using a selection of CYP2C9-specific probe substrates with both high (*S*-naproxen, flurbiprofen) and low (*S*-warfarin, phenytoin) turnover rates (**Table 2.1**). The same ordering of specific activities for the three enzymes was maintained across all four probe substrates, with the P279T variant showing a low to moderate decrease in catalytic efficiency (60 – 90% of WT activity, depending on the substrate), and the N218I variant showing a much more significant decrease (14 – 38% of WT activity). Interestingly, although the differences in overall catalytic efficiencies for the variants were similar across all four substrate probes, the individual components of K_m and V_{max} varied. P279T metabolized *S*-warfarin (to both 6- and 7-hydroxywarfarin) with a K_m value that was essentially identical to the WT protein, but showed both an increase in K_m for phenytoin and flurbiprofen hydroxylation, and a decrease in K_m for *S*-naproxen O-desmethylation, in comparison to the WT protein. Likewise, although the V_{max} values for *S*-warfarin, phenytoin and flurbiprofen metabolism by P279T were all very close, though slightly lower, than the WT rate, the turnover rate for naproxen dealkylation by P279T was approximately half that of WT enzyme. In contrast, the N218I variant showed a similar binding effect across all four substrates, with K_m increasing by 1.6 to 2.3-fold compared to WT. However, while the Vmax values for warfarin, phenytoin and naproxen metabolism by N218I were all

similarly low (i.e. less than 50%) there was only a 15% reduction in the flurbiprofen 4'-hydroxylation rate, compared to WT CYP2C9.

2.4 DISCUSSION

Coding-region variants of *CYP2C9* can dramatically influence the pharmacokinetics and pharmacodynamic response of CYP2C9 narrow therapeutic substrates, necessitating dose changes to avoid adverse drug responses when drug clearance is reduced. The rare coding *CYP2C9* variants identified in the AN/AI population described in Chapter 1 include *M1L, N218I* and *P279T*. The most prevalent of these novel variants, *M1L*, occurs at the translation start site and was predicted to disrupt translation and confer a CYP2C9 poor metabolizer phenotype. The *N218I* variant was also predicted to decrease CYP2C9 enzyme activity [6], with the neighboring variant, *Q214L* (*CYP2C9*28*), displaying a significant decline in *S*-warfarin 7-hydroxylation activity in CYP2C9.28 expressing COS-7 cells [11]. The *P279T* allele was originally identified in a Japanese population with a MAF of 0.002 [12], and is now described as *CYP2C9*29* in the star nomenclature. The *CYP2C9*29* variant was predicted to result in reduced enzyme activity, based on several *in vitro* studies that have investigated the catalytic capabilities of CYP2C9.29 towards diverse substrates [11,13-15]. The P279T enzyme expressed well in COS-7 cells and exhibited 54% of the catalytic efficiency of the wild-type enzyme for formation of (S)-7-hydroxywarfarin due entirely to a decrease in Vmax [11]. The same investigators reported that CYP2C9.29 exhibited 72% of the activity of wild-type towards tolbutamide when reactions were conducted at the high substrate concentrations of 1 mM.

Functional characterization of the *Met1Leu* (1A>T) and *N218I* variants has never been reported. Expression of *M1L* in HepG2 cells confirmed the predicted lack of transcription of the mRNA that was abundantly present. Consequently, *M1L* should be categorized as a no function

allele, similar to the previously reported base frameshift and premature stop codon variants: *CYP2C9*6, CYP2C9*15,* and *CYP2C9*25*. The absence of mRNA translation from a different altered start codon allele, *CYP2C9 M1V* (*CYP2C9*36*, rs114071557), is reported in the ExAc database with a low frequency of 0.0004% across all studied populations [16]. Notably, PharmVar describes the start site allele, *CYP2C9*36,* as a variant of unknown function, and while no experimental data have been provided as yet, the foregoing considerations suggest that *CYP2C9*36* is also null. *CYP2C19*4*, (ATG→TTG) was one of the first null variants to be discovered that conferred poor metabolizer status for metabolism of CYP2C19 substrates because of loss of the translation start site [17].

The *in vitro* results presented in this chapter support the hypothesis that the *M1L* variant is null for CYP2C9 enzyme activity. Despite the high mRNA signal observed for the *M1L* variant, no CYP2C9 protein accumulated in HepG2 cells. To test if the *M1L* variant confers CYP2C9 poor metabolizer status, an *in vivo* pharmacogenetic-pharmacokinetic study in genotyped Yup'ik participants was conducted and is described in Chapter 3. This *in vivo* study guided the selection of the probe substrates selected for *in vitro* analysis in this chapter. The established CYP2C9 probes flurbiprofen, a non-steroidal anti-inflammatory drug, as well as warfarin and phenytoin, which are narrow therapeutic index drugs, all require a prescription in the USA. Flurbiprofen was originally proposed as the CYP2C9 substrate for the *in vivo* study, but ultimately the over the counter medication *(S)*-naproxen was selected, further discussed in Chapter 3. For this reason, *(S)*-naproxen was included in the *in vitro* studies with purified CYP2C9 protein variants in this chapter in order to assess its pharmacokinetic behavior compared to the other well-established CYP2C9 probe substrates.

In contrast to the *M1L* variant, His-tagged wild-type protein and the N218I and P279T variants proteins expressed well in E. coli and were highly purified after affinity chromatography. Upon reconstitution with cytochrome P450 reductase and cytochrome b5, the N218I and P279T variants metabolized *S*-warfarin, phenytoin, flurbiprofen and *S*-naproxen to the expected mono-hydroxylated or O-desmethylated metabolites. Kinetic analysis revealed that the relative catalytic efficiency towards probe substrate metabolism were 14 – 38% for N218I and 60-90% for P279T, compared to purified WT enzyme. These deficits were the result of a combination of changes in K_m and V_{max}, indicating changes in both substrate binding and $k_{cat,}$ depending on the substrate. The latter change could be the result of altered function of the CYP2C9 catalytic heme intermediate or in electron transfer from P450 reductase or cytochrome b_5 to CYP2C9.

With regard to *(S)*-naproxen O-demethylation by N218I, the K_m was increased by 1.7-fold and the V_{max} was decreased by 2.3-fold, compared to the WT protein, resulting in an overall decrease in catalytic efficiency to 25% of that of WT. The *N218I* variant is located between the F and G helices, which encapsulate the active site [18], and, thus, a decrease in the k_{cat} and an increase in the K_m due to amino acid substitution in a region of the enzyme known to be important for catalytic activity [19] was expected. Moreover, the reduction in catalytic activity of the N218I protein was of similar magnitude as the CYP2C9*3 enzyme and, therefore, is also predicted to alter the metabolic clearance of *(S)*-warfarin and other CYP2C9 substrates *in vivo.*

P279T is located between the H and I helices of the enzyme and resulted in a decrease in catalytic efficiency to approximately 60% of that of the WT protein, for *(S)*-naproxen O-demethylation. This decrease in catalytic efficiency was mainly attributed to a decrease in V_{max} by 2.4-fold, compared to WT enzyme. Helix I borders part of the CYP2C9 active site cavity [20] and perturbances in helix I due to amino acid substitutions have been hypothesized to disturb the active

site topography and reduce enzyme activity [21]. Therefore, the reduction in O-desmethylation of *(S)*-naproxen observed with the P279T variant was not surprising, however, it is important to note that this variant displays substrate selectivity and had similar activity (~90% of that of WT enzyme) for *(S)*-warfarin metabolism.

Novel variants identified in AN/AI populations require characterization, as functionally defective variants such as *M1L* can have significant clinical implications, especially considering the percentage of Yup'ik people with genetically-determined impairment in CYP2C9 activity would be 15-16%. This high percentage is a consequence of identifying novel alleles in the Yup'ik population at a combined frequency of 13%, compared to the <3% frequency for the more commonly studied *CYP2C9*2* and **3* variants. The potential of *CYP2C9 M1L*, *N218I*, and perhaps even *P279T*, to alter the pharmacokinetics of drugs metabolized by CYP2C9 may put carriers at risk of exacerbated therapeutic effects from drugs that rely predominately on CYP2C9 for their metabolic clearance. With the emergence of precision medicine and pre-emptive pharmacogenetic testing, clinical implications for the AN people include phenotypic misclassification from currently available tests, which typically screen only for the common *CYP2C9*2* and **3* variants. Studies to further characterize the impact of variant alleles on drug disposition and response in AN people are needed to avoid additional healthcare disparities.

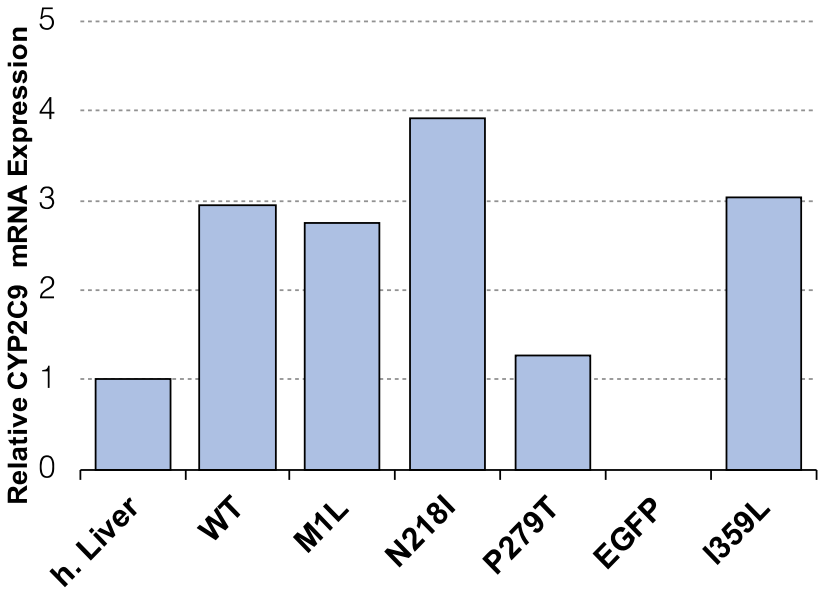

Figure 2.1. *CYP2C9* mRNA expression in HepG2 cells.

Each sample was normalized to its own relative fold change determined by comparing expression to h.Liver (set to 1.0), a representative pool of seven human liver samples, all *CYP2C9 WT*).

Figure 2.2. CYP2C9 protein expression in HepG2 cells.

The upper band of the Western blot is 30 μg protein from HepG2 cell lysate and the lower band is actin loading control. r2C9 is 0.25 pmol of CYP2C9 Supersomes.

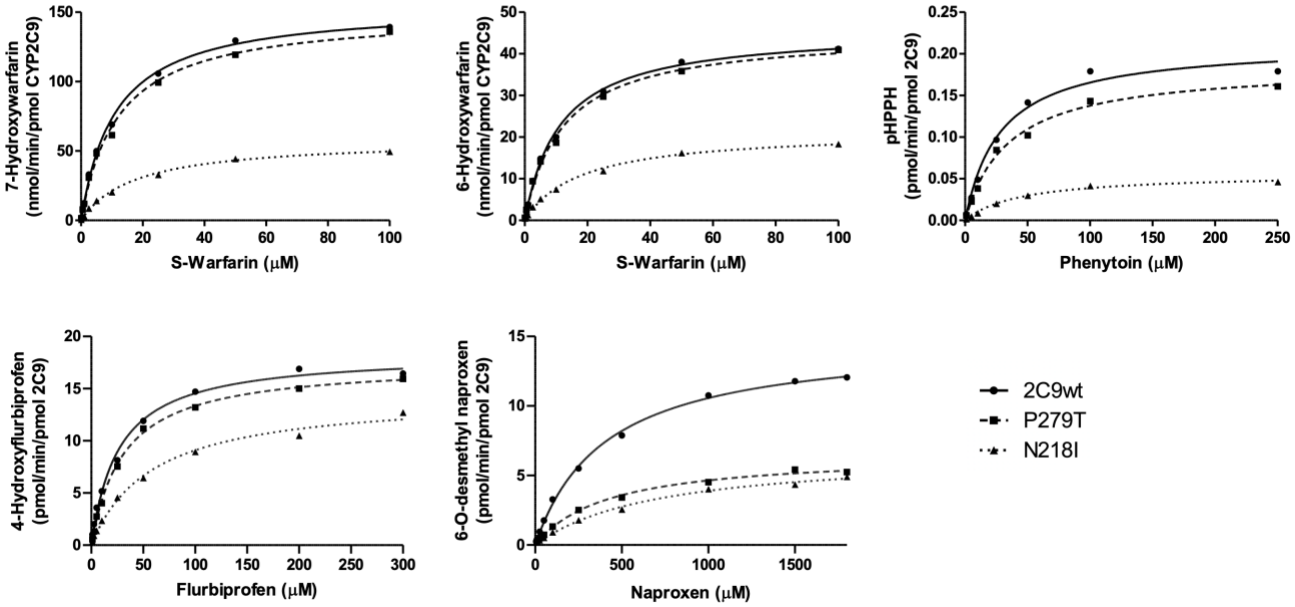

Figure 2.3. Kinetics of *CYP2C9* probe substrates by purified CYP2C9 variants and WT enzyme.

Table 2.1. Kinetics of probe substrate metabolism by purified CYP2C9 variant enzymes.

Metabolite	Variant	K_m (μM)	V_{max} (pmol/min/pmol)	CL_{int} (mL/min/pmol)	Catalytic Efficiency (%)
6-OH-Warf	WT	11.1 ± 0.3	0.046 ± 0.006	4.1 ± 0.5	100
	P279T	11.0 ± 1.1	0.041 ± 0.004	3.7 ± 0.1	90
	N218I	18.0 ± 1.5	0.022 ± 0.001	1.2 ± 0.1	30
7-OH-Warf	WT	10.7 ± 0.4	0.153 ± 0.021	14.3 ± 1.8	100
	P279T	10.7 ± 1.2	0.133 ± 0.015	12.4 ± 0.2	87
	N218I	17.1 ± 0.9	0.060 ± 0.002	3.5 ± 0.3	24
p-HPPH	WT	27.4 ± 4.7	0.204 ± 0.008	7.6 ± 1.2	100
	P279T	34.4 ± 1.5	0.171 ± 0.013	5.0 ± 0.5	66
	N218I	47.9 ± 3.2	0.050 ± 0.006	1.1 ± 0.2	14
4'-OH-Flurb	WT	26.0 ± 0.40	17.7 ± 0.53	680 ± 10	100
	P279T	39.0 ± 2.4	16.8 ± 2.4	430 ± 39	63
	N218I	59.8 ± 0.25	15.3 ± 0.28	260 ± 3.7	38
DM-Nap	WT	419 ± 59	15.3 ± 0.72	37 ± 3.5	100
	P279T	310 ± 80	6.3 ± 0.25	22 ± 4.2	59
	N218I	725 ± 106	6.6 ± 1.1	9.1 ± 0.2	25

Kinetic parameters are mean values from 3 replicate experiments.

2.5 REFERENCES

1. Daly, A.K.; Rettie, A.E.; Fowler, D.M.; Miners, J.O. Pharmacogenomics of CYP2C9: Functional and Clinical Considerations. *J Pers Med* **2017**, *8*, doi:10.3390/jpm8010001.
2. Zhou, Y.; Ingelman-Sundberg, M.; Lauschke, V.M. Worldwide Distribution of Cytochrome P450 Alleles: A Meta-analysis of Population-scale Sequencing Projects. *Clin Pharmacol Ther* **2017**, *102*, 688-700, doi:10.1002/cpt.690.
3. Kimmel, S.E.; French, B.; Kasner, S.E.; Johnson, J.A.; Anderson, J.L.; Gage, B.F.; Rosenberg, Y.D.; Eby, C.S.; Madigan, R.A.; McBane, R.B., et al. A pharmacogenetic versus a clinical algorithm for warfarin dosing. *N Engl J Med* **2013**, *369*, 2283-2293, doi:10.1056/NEJMoa1310669.
4. Drozda, K.; Wong, S.; Patel, S.R.; Bress, A.P.; Nutescu, E.A.; Kittles, R.A.; Cavallari, L.H. Poor warfarin dose prediction with pharmacogenetic algorithms that exclude genotypes important for African Americans. *Pharmacogenet Genomics* **2015**, *25*, 73-81, doi:10.1097/FPC.0000000000000108.
5. Henderson, L.M.; Claw, K.G.; Woodahl, E.L.; Robinson, R.F.; Boyer, B.B.; Burke, W.; Thummel, K.E. P450 Pharmacogenetics in Indigenous North American Populations. *J Pers Med* **2018**, *8*, doi:10.3390/jpm8010009.
6. Fohner, A.E.; Robinson, R.; Yracheta, J.; Dillard, D.A.; Schilling, B.; Khan, B.; Hopkins, S.; Boyer, B.; Black, J.; Wiener, H., et al. Variation in genes controlling warfarin disposition and response in American Indian and Alaska Native people: CYP2C9, VKORC1, CYP4F2, CYP4F11, GGCX. *Pharmacogenet Genomics* **2015**, *25*, 343-353, doi:10.1097/FPC.0000000000000143.
7. Dai, D.P.; Xu, R.A.; Hu, L.M.; Wang, S.H.; Geng, P.W.; Yang, J.F.; Yang, L.P.; Qian, J.C.; Wang, Z.S.; Zhu, G.H., et al. CYP2C9 polymorphism analysis in Han Chinese populations: building the largest allele frequency database. *Pharmacogenomics J* **2014**, *14*, 85-92, doi:10.1038/tpj.2013.2.
8. Wang, L.; Bao, S.H.; Pan, P.P.; Xia, M.M.; Chen, M.C.; Liang, B.Q.; Dai, D.P.; Cai, J.P.; Hu, G.X. Effect of CYP2C9 genetic polymorphism on the metabolism of flurbiprofen in vitro. *Drug Dev Ind Pharm* **2015**, *41*, 1363-1367, doi:10.3109/03639045.2014.950274.
9. Chen, L.G.; Wang, Z.; Zhu, Y.; Xiong, J.H.; Sun, L.R.; Dai, D.P.; Cai, J.P.; Hu, G.X. In vitro metabolism of phenytoin in 36 CYP2C9 variants found in the Chinese population. *Chem Biol Interact* **2016**, *253*, 93-99, doi:10.1016/j.cbi.2016.04.040.
10. Cheesman, M.J.; Baer, B.R.; Zheng, Y.M.; Gillam, E.M.; Rettie, A.E. Rabbit CYP4B1 engineered for high-level expression in Escherichia coli: ligand stabilization and processing of the N-terminus and heme prosthetic group. *Arch Biochem Biophys* **2003**, *416*, 17-24, doi:10.1016/s0003-9861(03)00278-9.
11. Niinuma, Y.; Saito, T.; Takahashi, M.; Tsukada, C.; Ito, M.; Hirasawa, N.; Hiratsuka, M. Functional characterization of 32 CYP2C9 allelic variants. *Pharmacogenomics J* **2014**, *14*, 107-114, doi:10.1038/tpj.2013.22.
12. Maekawa, K.; Fukushima-Uesaka, H.; Tohkin, M.; Hasegawa, R.; Kajio, H.; Kuzuya, N.; Yasuda, K.; Kawamoto, M.; Kamatani, N.; Suzuki, K., et al. Four novel defective alleles and comprehensive haplotype analysis of CYP2C9 in Japanese. *Pharmacogenet Genomics* **2006**, *16*, 497-514, doi:10.1097/01.fpc.0000215069.14095.c6.

13. Dai, D.P.; Wang, Y.H.; Wang, S.H.; Geng, P.W.; Hu, L.M.; Hu, G.X.; Cai, J.P. In vitro functional characterization of 37 CYP2C9 allelic isoforms found in Chinese Han population. *Acta Pharmacol Sin* **2013**, *34*, 1449-1456, doi:10.1038/aps.2013.123.
14. Dai, D.P.; Wang, S.H.; Geng, P.W.; Hu, G.X.; Cai, J.P. In vitro assessment of 36 CYP2C9 allelic isoforms found in the Chinese population on the metabolism of glimepiride. *Basic Clin Pharmacol Toxicol* **2014**, *114*, 305-310, doi:10.1111/bcpt.12159.
15. Wang, Y.H.; Pan, P.P.; Dai, D.P.; Wang, S.H.; Geng, P.W.; Cai, J.P.; Hu, G.X. Effect of 36 CYP2C9 variants found in the Chinese population on losartan metabolism in vitro. *Xenobiotica* **2014**, *44*, 270-275, doi:10.3109/00498254.2013.820007.
16. Lek, M.; Karczewski, K.J.; Minikel, E.V.; Samocha, K.E.; Banks, E.; Fennell, T.; O'Donnell-Luria, A.H.; Ware, J.S.; Hill, A.J.; Cummings, B.B., et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **2016**, *536*, 285-291, doi:10.1038/nature19057.
17. Ferguson, R.J.; De Morais, S.M.; Benhamou, S.; Bouchardy, C.; Blaisdell, J.; Ibeanu, G.; Wilkinson, G.R.; Sarich, T.C.; Wright, J.M.; Dayer, P., et al. A new genetic defect in human CYP2C19: mutation of the initiation codon is responsible for poor metabolism of S-mephenytoin. *J Pharmacol Exp Ther* **1998**, *284*, 356-361.
18. Williams, P.A.; Cosme, J.; Ward, A.; Angove, H.C.; Matak Vinković, D.; Jhoti, H. Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature* **2003**, *424*, 464-468, doi:10.1038/nature01862.
19. Wester, M.R.; Yano, J.K.; Schoch, G.A.; Yang, C.; Griffin, K.J.; Stout, C.D.; Johnson, E.F. The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0-A resolution. *J Biol Chem* **2004**, *279*, 35630-35637, doi:10.1074/jbc.M405427200.
20. Mosher, C.M.; Hummel, M.A.; Tracy, T.S.; Rettie, A.E. Functional analysis of phenylalanine residues in the active site of cytochrome P450 2C9. *Biochemistry* **2008**, *47*, 11725-11734, doi:10.1021/bi801231m.
21. Maekawa, K.; Adachi, M.; Matsuzawa, Y.; Zhang, Q.; Kuroki, R.; Saito, Y.; Shah, M.B. Structural Basis of Single-Nucleotide Polymorphisms in Cytochrome P450 2C9. *Biochemistry* **2017**, *56*, 5476-5480, doi:10.1021/acs.biochem.7b00795.

Chapter 3. CHARACTERIZATION OF THE *IN VIVO* CATALYTIC EFFICIENCY OF *CYP2C9 M1L*, A NOVEL AND COMMON VARIANT IN THE YUP'IK ALASKA NATIVE POPULATION

(A version of this chapter is currently a draft manuscript "Validation of *(S)*-Naproxen as a CYP2C9 Activity Probe and Characterization of the *In Vivo* Catalytic Efficiency of *CYP2C9 M1L*, a Novel and Common Variant in the Yup'ik Alaska Native Population.")

3.1 INTRODUCTION

The cytochrome P450 family 2 subfamily C member 9 (CYP2C9) enzyme is responsible for the metabolic clearance of approximately 15% of all medications cleared through a P450-mediated pathway [1,2]. CYP2C9 has a broad range of clinical substrates including anticoagulants, anticonvulsants, angiotensin II blockers, hypoglycemic agents, and nonsteroidal anti-inflammatory drugs. The *CYP2C9* gene is highly polymorphic, with coding-region variation (*CYP2C9*2* and **3*) that confer a poor metabolizer phenotype, dramatically influencing the pharmacokinetics and drug response of commonly used narrow therapeutic index medications (e.g., *(S)*-warfarin, phenytoin) [3-5].

Recently, our group identified a novel *CYP2C9 Met1Leu* (*M1L*) variant in a Yup'ik Alaska Native (AN) population [6]. The substitution of leucine for methionine at the first amino acid position is predicted to slow or stop RNA translation. Indeed, *in vitro* studies with *M1L* gene transfected HepG2 cells demonstrated that the CYP2C9 *Leu1* variant protein does not accumulate in this human liver-derived cell line (see Chapter 3, McDonald, unpublished 2019). In the Yup'ik

population, the *M1L* variant is expressed at a higher frequency (6.3%) than the well-characterized *CYP2C9*2* (0.3%) and *CYP2C9*3* (2.1%) alleles [6]. The Yup'ik people live in the relatively remote Yukon-Kuskokwim (YK) Delta of Southwestern Alaska. There are 58 communities in the YK Delta in a 75,000 square mile area, and all communities are off the road system. Communities have health clinics staffed by community health aids and primary care is offered through sub-regional health clinics or the regional hub hospital in Bethel, Alaska. This geographic isolation of communities away from primary care providers creates challenges to medical service that may not be experienced in urban areas and could contribute to health disparities in the region. Moreover, pharmacotherapy with narrow therapeutic index drugs can be more challenging, because of the difficulty in monitoring drug responses. With specific regard to CYP2C9 substrates, such as warfarin, phenytoin and tolbutamide, variation in the *CY2C9* gene contributes to inter-individual differences in dose requirement [3-5,7]. Genetic testing, as a form of precision medicine, may enhance clinical utility for managing these and other drug therapies in the Yup'ik population. To advance the goals of precision medicine, it is necessary to fully understand the frequency and function of variation in important pharmacogenes such as CYP2C9. Moreover, it is critical to investigate previously unknown variants, such as *M1L,* that are common in the Yup'ik population [6] and are expected to impair CYP2C9 activity.

Characterization of enzyme function *in vivo* is commonly accomplished with a pharmacokinetic study that involves administration of a probe drug selectively metabolized by the enzyme of interest. Established CYP2C9 probes include the same narrow therapeutic index drugs warfarin, phenytoin and tolbutamide, as well as the non-steroidal anti-inflammatory drugs celecoxib and flurbiprofen. However, for a study in the Yup'ik population, selection of a commonly used drug known to be safe and recognizable to participants (over the counter) is

considered just as important as selectivity for CYP2C9 activity. Thus, we elected to validate and use *(S)*-naproxen as the *in vivo* enzyme probe. *(S)*-naproxen undergoes O-dealkylation primarily by CYP2C9, with minor involvement from other P450 enzymes [8,9]. *(S)*-naproxen is well absorbed [10,11], highly bound to albumin [11], and almost completely eliminated in the urine as naproxen glucuronide (60% of the dose), unchanged naproxen (1%), and secondary glucuronide and sulfate metabolites of O-desmethyl naproxen (20%) [12-15]. Although a minor pathway of *(S)*-naproxen elimination, a change in the total urinary *(S)*-O-desmethyl naproxen/*(S)*-naproxen concentration ratio is indicative of a change in CYP2C9 intrinsic formation clearance. This study's objective was to verify the selectivity of the *(S)*-naproxen O-dealkylation reaction for CYP2C9 and then determine the catalytic efficiency of the novel *M1L* variant *in vivo* to inform on its potential to affect the drug disposition and pharmacological response of medications metabolized by CYP2C9.

Pharmacogenetic testing offers an opportunity to improve the tailoring of medication therapy for agents in which genetic variation, at least in part, controls response, such as is the case for warfarin anticoagulation therapy. Prevalence and frequency of genetic variation is diverse across racial and ethnic populations, and many groups such as the AN community, do not currently benefit from personalized drug therapy due to a lack of foundational genotype-phenotype data. Our research collaborations with the Yukon-Kuskokwim Health Corporation, which serves about 23,000 Yup'ik people, have the goal of gathering pharmacogenetic information to improve drug response and minimize adverse drug events in this under-studied AN population.

3.2 METHODS

3.2.1 *Setting*

Study recruitment was conducted at 10 communities based in the YK Delta of Alaska. Approximately two-thirds of the AN population in Alaska live in rural communities with populations of 50 – 1000 people, many only accessible by air or water [16]. Dr. Bert Boyer and Ms. Scarlett Hopkins, now based at Oregon Health & Science University, have ongoing genetic research partnerships with 11 out of the 58 rural communities in the YK-Delta.

3.2.2 *Study Participants*

Study participants were selected from a cross-sectional population of Yup'ik men and women over 18 years old, for whom *CYP2C9 M1L* genotype was previously determined, and who consented to be contacted for future research investigations. Participants were in good health and not taking non-steroidal anti-inflammatory agents or other drugs known or suspected of altering CYP2C9 function.

3.2.3 *Study Design*

The University of Alaska Fairbanks and Oregon Health and Science University Institutional Review Boards (IRB) and the YK Health Corporation Human Studies Committee and Executive Board approved this study. The University of Washington (UW) IRB approved the overall research project, as UW was the academic home of the grant funding this research (NIH P01 GM116691) and its principal investigators.

Following informed consent, participants were asked to fast overnight to start the pharmacokinetic study and provided a baseline urine sample. A single 220 mg naproxen sodium

caplet (200 mg *(S)*-naproxen), was administered with a glass of water. Urine was then collected for the next 24 hours. Due to the instability of naproxen acyl glucuronides in alkaline media, urine pH was stabilized by adding 13.6 g monobasic potassium phosphate to each urine collection container before use. At the end of the collection interval, study participants returned the urine collection container to the study site, where the urine volume was measured and recorded. The urine was well mixed and two 5 mL aliquots were taken from the collection container and stored initially at -15°C in a portable freezer, and then at -80°C until analysis.

3.2.4 *Genotyping*

To identify *Met1/*Leu1 heterozygotes and *Leu1/Leu1* homozygotes from the Yup'ik population, the Fluidigm platform was used to perform genotype analysis of DNA extracted from white blood cells, targeting the CYP2C9 exome, as previously described in Fohner et al., 2015 [6]. Based on prior gene sequencing work, the following *CYP2C9* variants were tested: *Met1Leu* (1A>T), **29, *2* (rs1799853), **3* (rs1057910), **8* (rs7900194), **11* (rs28371685), **13* (rs72558187), **14* (rs72558189), Iso218Asp T>A, and **29* (Thre279Pro A>C). A total of 1112 individuals from the Yup'ik population were genotyped.

3.2.5 *Validation of (S)-Naproxen as a CYP2C9 Probe Substrate*

Comprehensive *in vitro* experiments were performed to validate the selectivity and sensitivity of naproxen as a probe for CYP2C9 activity. Unlabeled *(S)*-naproxen, and racemic O-desmethyl naproxen-d_3 were purchased from Toronto Research Chemicals (Ontario, Canada). Unlabeled *(S)*-O-desmethyl naproxen, furafylline, sulfaphenazole, and β-NADPH were purchased from Sigma Aldrich (St. Louis, MO). Pooled human liver microsomes (HLMs) were purchased

from XenoTech (Kansas City, KS). Individual HLMs were isolated from the University of Washington School of Pharmacy Human Liver bank, as previously reported [17]. Individual recombinantly-expressed cytochrome P450 Supersome preparations were obtained from Corning Life Sciences (Woburn, MA). All other chemicals were analytical grade or better and obtained from various commercial vendors.

(S)-naproxen was incubated with pooled HLMs (0.5 mg/mL final concentration) in the presence of NADPH (1 mM final concentration) in 50 mM KH_2PO_4 with 1.27 mM EDTA, pH 7.4, in a total volume of 200 µl. In experiments using selective inhibitors, sulfaphenazole (prepared in methanol, with final concentration below 0.2%) and furafylline (prepared in DMSO, with final concentration below 0.1%), the final inhibitor concentration was 10 µM. Microsomal incubations with furafylline underwent a 20 minute preincubation with the CYP1A2 inhibitor prior to *(S)*-naproxen reaction initiation. Reactions were carried out for 20 minutes at 37°C, over a *(S)*-naproxen concentration range of 5 – 1800 µM. The microsomal incubation reaction was quenched with the addition of 1 mL ice cold methanol containing 2% formic acid. To the quenched samples, 80 ng of O-desmethyl naproxen-d_3, internal standard, were added. The samples were then centrifuged at 3000 *g* for 10 minutes, decanted into glass culture tubes, and dried with nitrogen gas, and resuspended in 50 µl mobile phase. A volume of 20 µl was injected onto the LC/MS.

A CYP Supersome screen was performed by evaluating CYP1A1, CYP1A2, CYP2A6, CYP2B6, CYP2C8, CYP2C9*1, CYP2C9*2, CYP2C9*3, CYP2C19, CYP2D6, CYP2E1, CYP3A4, CYP3A5, and CYP3A7 metabolic activity towards *(S)*-naproxen. *(S)*-naproxen was incubated with 10 pmol of each Supersome preparation in 50 mM KH_2PO_4 with 1.27 mM EDTA buffer (except CYP2A6, for which 50 mM Tris buffer was used) in a total volume of 200 µl. Reactions were initiated with the addition of NADPH (1 mM final concentration) and incubated

for 20 minutes at 37°C at the *(S)*-naproxen concentrations of 25 μM (sub-saturating concentration) and 1000 μM (saturating concentration). The incubation reaction was quenched with 1 mL ice cold methanol containing 2% formic acid and 200 ng of O-desmethyl naproxen-d_3, internal standard, was added. The samples were centrifuge at 3000 *g* for 10 minutes, decanted into glass culture tubes, and dried with N2 gas, and resuspended in 50 μl mobile phase. A volume of 5 μl was injected onto the LC/MS. For CYP enzymes that catalyzed *(S)*-O-desmethyl naproxen formation, additional reactions were carried out for 20 minutes at 37°C, over the *(S)*-naproxen concentration range of 5 – 1800 μM, to determine Michaelis-Menten kinetic parameters.

To evaluate the effect of varying CYP1A2 content on naproxen metabolism, single donor HLMs from the UW Human Liver Bank were selected based on CYP1A2 protein content (determined by LC-MS/MS analysis). Five high CYP1A2 expressors and five low CYP1A2 expressors, with comparable CYP2C9 protein expression (Error! Reference source not found.), w ere incubated with 20 μM *(S)*-naproxen, NADPH (1 mM final concentration), in the presence or absence of 10 μM sulfaphenazole and furafylline, in 50 mM KH_2PO_4 with 1.27 mM EDTA, pH 7.4, in a total volume of 200 μl. Reactions were carried out as described above with pooled HLM experiments.

Kinetic experiments, for the assessment of CYP2C9 probe substrate metabolism, were performed in triplicate on different days, with duplicate data points for each substrate concentration. Calibration curves for *(S)*-naproxen metabolites were prepared by spiking variable amounts of unlabeled *(S)*-O-desmethyl naproxen, into 200 μl of potassium phosphate buffer in order to generate standard mixtures with final concentrations of 0.2 – 10 μM for HLM incubations, 0.1 – 5 μM for HLM inhibition experiments, 0.1 – 30 μM for Supersome experiments. Standard

solutions, prepared in duplicate for each concentration, were immediately worked-up and analyzed in an identical fashion to that described for the incubation samples above.

3.2.6 *Urine Sample Preparation*

For *(S)*-naproxen detection, urine samples (50 µl) were prepared by adding 100 µl of HPLC grade water and 100 µl of 1 nmol racemic naproxen-d_3 (internal standard for *(S)*-naproxen). For naproxen acyl glucuronide detection, urine samples were diluted 1:20 in blank urine then 50 µl of the diluted sample was combined with 100 µl of HPLC grade water and 100 µl of 1 nmol racemic flurbiprofen acyl glucuronide (internal standard for naproxen acyl glucuronide). For total *(S)*-O-desmethyl naproxen detection, urine was diluted 1:4 in blank urine, then the diluted urine sample (50 µl) was combined with 80 µl HPLC grade water, 20 µl 6M HCL, 100 µl of internal standard (1 nmol racemic O-desmethyl naproxen-d_3) followed by vortexing and incubating at 90°C for 60 minutes to facilitate glucuronide and sulfate cleavage via acid hydrolysis. All samples were vortexed, centrifuged at 14,000 *g* for 5 minutes, and then 50 µl of sample transferred to autosampler vials and 2 µl were injected.

3.2.7 *Urine Analysis*

To evaluate the effect of *M1L* variation on CYP2C9 function, the urinary *(S)*-O-desmethyl naproxen to unchanged naproxen metabolite to parent ratio was determined from 24-hour urine collection. Naproxen and metabolite concentrations were accessed by LC/MS using an Agilent 1956B single quadrupole mass spectrometer coupled with an Agilent 1200 series (Santa Clara, CA) liquid chromatography system. Chromatographic separation was achieved on a Luna C18 (2 mm x 50 mm x 5µm) column (Torrence, CA) with a mobile phase flow rate of 0.3 mL/min. The mobile phase consisted of 10 mM ammonium formate (A, pH 3.5) and methanol (B), and linear

gradients were applied with B% increasing from 40% to 80% between 3 and 8 minutes and decreasing to 40% at 9 minutes. Quantitation was achieved by selected ion monitoring centered on m/z values of 248.1 for *(S)*-naproxen, 251.1 for racemic naproxen-d3, 234.1 for *(S)*-O-desmethyl naproxen, 237.1 for racemic O-desmethyl naproxen-d_3, 424.1 for naproxen acyl glucuronide, and 438.1 for racemic flurbiprofen acyl glucuronide. Data acquisition and analysis were performed using the Agilent MassHunter software. Calibration curves were constructed by plotting the peak area ratio of each compound to the respective internal standard against a range of concentrations. We measured the urinary concentration of the major naproxen metabolite, naproxen acyl glucuronide, in order to ensure comparable dose recovery and urine collection compliance. The intra-day variation for did not exceed 2% for the low concentration quality control (QC) for *(S)*-O-desmethyl naproxen and did not exceed 6% for the high concentration QC. The relative errors of the two QC concentrations tested in three independent experiments were within 5% and 8% for the low and high concentration QCs, respectively.

3.2.8 *Statistical Analysis*

The primary statistical analysis plan involved testing by regression analysis for an additive-gene dose effect (0, 1, or 2 functional alleles), where 10 participants per genotype group were needed to achieve a power of 0.8 and a significance level of 0.05. Considering possible difficulties in recruiting 10 Leu1 homozygotes, we proposed *a priori* an alternative statistical analysis plan where the homozygous variant Leu1/Leu1 group would be combined with the heterozygous variant Met1/Leu1 group. Under this plan, a sample size of 15 reference and 15 Leu1 carriers and homozygotes would achieve the same power at the same level of significance.

3.3 RESULTS

3.3.1 *Study Enrollment Based on M1L Genotype*

A total of 1112 individuals were genotyped for *CYP2C9* variants. After removing duplicate records from repeat visits (*n=193)* and genotypes with no calls (*n=6)*, a total of 913 genomes were considered for the pharmacokinetic study **(Figure 3.1).** Individuals carrying *CYP2C9*2*, **3, *8, *11, *13, *14, *29,* or *N218I* alleles were excluded due to their confounding effects on CYP2C9 activity. A total of 8 *Leu1/Leu1* homozygotes were identified from 5 different communities, with an average age of 36 years. These individuals were all unrelated at the parent-child and sibling level. A total of 85 *Met1/Leu1* heterozygotes, with an average age of 37 years, and 629 in the *Met1/Met1* reference (wildtype), with an average age of 36 year, were also identified.

3.3.2 *Selectivity and Sensitivity of Naproxen as a Probe for CYP2C9 Enzyme Activity*

Comprehensive *in vitro* studies determined that CYP2C9 is the predominant enzyme metabolizing *(S)*-naproxen to *(S)*-O-desmethyl naproxen. A representative Michaelis-Menten plot of *(S)*-O-desmethyl naproxen formation in pooled HLMs is shown in **Figure 3.2**, with a mean K_m of 420 μM and V_{max} of 0.92 nmol/min/mg microsomal protein from three repeated experiments (**Table 3.1**). The CYP Supersome screen showed that *(S)*-O-desmethyl naproxen was formed from 25 μM naproxen by only three P450 enzymes – CYP2C9, CYP2C8, CYP1A2 (**Figure 3.3**). This is a more clinically relevant concentration, considering the *(S)*-naproxen C_{max} (~400 μM) (FDA label) and extensive protein binding (>99%). The activity of CYP2C9 was much higher than that of CYP1A2, and CYP2C8. However, with 1000 μM of *(S)*-naproxen, the contribution of CYP1A2

increased by over 30-fold and small amounts of product were detected in incubations with several other P450 Supersomes (**Table 3.2**).

Full kinetic experiments were conducted to assess the *(S)*-O-desmethyl naproxen intrinsic formation clearances of CYP2C9, CYP1A2, and CYP2C8 Supersomes (**Figure 3.4**). The mean V_{max} values for CYP2C9 and CYP1A2 were similar, 31.7 and 41.7 pmol/min/pmol P450, respectively, while their K_m values were markedly different, 280 μM for CYP2C9 and 1000 uM for CYP1A2 (**Table 3.3).** Given liver abundances of 73, 52, and 24 pmol P450/mg protein [18] for CYP2C9, CYP1A2, and CYP2C8, respectively, the average contribution from each of these enzymes to *(S)*-O-desmethyl naproxen formation was predicted to be 78% for CYP2C9, 20% for CYP1A2, and 2% for CYP2C8.

We also estimated the fraction of *(S)*-naproxen metabolized to *(S)*-O-desmethyl naproxen in HLMs by CYP2C9 and CYP1A2 from selective enzyme inhibitor experiments conducted with 20 μM naproxen, a substrate concentration 5-fold below the K_m determined in pooled HLMs. *(S)*-O-desmethyl naproxen formation was reduced by 76.9 ± 1.5% with 10 μM sulfaphenazole, a selective CYP2C9 inhibitor, by 21.5 ± 1.6% with 10 μM furafylline, a selective CYP1A2 inhibitor, and by 95.8 ± 2.1% with both sulfaphenazole and furafylline (**Figure 3.5**). The solvents for the inhibitors had negligible effects on the percent inhibition (**Figure 3.5**). The effect of CYP2C9 and CYP1A2 inhibition on *(S)*-O-desmethyl naproxen formation by HLMs was also assessed in two groups of single donor HLMs, high CYP1A2 expressors (*n = 5)* with an average CYP1A2 content of 31.2 ± 10.8 pmol/mg microsomal protein, and low CYP1A2 expressors (*n = 5)* with an average of 2.8 ± 2.3 pmol/mg microsomal protein. CYP2C9 content was 53.2 ± 13.3 and 36.6 ± 6.3 pmol/mg microsomal protein in the high and low CYP1A2 groups, respectively. As predicted, the percent inhibited by 10 μM furafylline was greater in the high CYP1A2 group (39.7 ± 7.0 pmol/mg

microsomal protein), compared to the low CYP1A2 group (23.6 ± 7.6 pmol/mg microsomal protein), while the percent inhibited by 10 μM sulfaphenazole was greater in the low CYP1A2 group (85.2 ± 11.8 pmol/mg microsomal protein), compared to the high CYP1A2 group (65.5 ± 4.1 pmol/mg microsomal protein) (**Table 3.4**). Although the CYP1A2 content was 11-fold greater in the high CYP1A2 group, compared to the low CYP1A2 group, the percent inhibited by furafylline was only 1.7-fold greater. Thus, CYP1A2 contribution to *(S)*-O-desmethyl naproxen formation in HLMs was always minor, in comparison to the CYP2C9 contribution.

3.3.3 *Impact of M1L on Urinary Metabolite to Parent Ratio*

In a preliminary analysis of approximately two-thirds of the final study population, the mean ratio of *(S)*-O-desmethyl naproxen to naproxen was greater for the homozygous reference group (17.3 ± 7.8, n = 10), compared to the *M1L* variant group (10.1 ± 6.0) that includes 8 *Met1*/Leu1 heterozygotes and 3 *Leu1/Leu1* homozygotes (**Figure 3.6**). Pairwise comparison (assuming unequal variance between groups) was significant (p = 0.024), suggesting reduced activity for the *Leu1* variant. There was no evidence of metabolic shifting towards the parent glucuronide elimination pathway in *Leu1* carriers, as the urinary metabolite to parent ratio carriers of the *Leu1* allele (62.5) was similar to that of the reference group (70.5). We plan to continue recruitment for the *in vivo* urine analysis study to reach a total of 15 homozygote variants and heterozygotes (combined), and 15 *CYP2C9* reference individuals.

3.4 DISCUSSION

Yup'ik Alaska Native people are under-represented in genetic research, but have unique pharmacogene variation that may critically impact their response to drug therapy. This is the first

study to characterize the *in vivo* functional effect of the novel, relatively common, *CYP2C9 M1L* single nucleotide polymorphism identified in the Yup'ik AN people. Given the mean contributions of CYP2C9 (80%) and CYP1A2 (20%) to *(S)*-O-desmethyl naproxen formation in HLMs, it was predicted that a *Leu1* variant group (comprised of three *Leu1/Leu1* homozygotes and eight heterozygotes) would have a 51% reduction in urinary ratio of *(S)*-O-desmethyl naproxen to unchanged naproxen, compared to the reference group. The observed 42% reduction in the *Leu1* variant group is in good agreement with this prediction. Reduced activity of the *Leu1* variant has clinical implications, including phenotypic misclassification from currently available pharmacogenetic tests, in the Alaska Native population.

While the *M1L* variant is a novel *CYP2C9* impaired function variant found in the AN population (and at a lower frequency in American Indians) [6], it is not the only example of loss of the translation start codon conferring poor metabolizer status in the P450 2C subfamily. *CYP2C19*4* (rs28399504) is a loss-of-function allele, resulting from a substitution of methionine to valine at the first amino acid position [19]. However, based on 1000 genomes data, the *CYP2C19*4* variant is only found at low frequencies across world populations: 0.8% in a Mexican population (California, USA), 0.5% in a Han Chinese population (Beijing, China), and the allele was not detected in Europeans (Utah residents with Northern and Western European ancestry) or in African Americans (Southwestern USA) [20]. By contrast, *M1L* is present at a relatively high minor allele frequency of 6.3% in the Yup'ik population and, thus, can contribute to the risk of an adverse drug response from narrow therapeutic index CYP2C9 substrates such as (S)-warfarin, phenytoin, and tolbutamide.

In order to characterize the catalytic efficiency of the *M1L* variant, this study first had to establish the use of *(S)*-naproxen as an over the counter probe substrate to assess CYP2C9 enzyme

activity. Earlier studies characterizing the *in vitro* metabolism of *(S)*-naproxen downplayed its utility as a probe substrate due to involvement of CYP1A2 [8,21], and because an *in vivo* study in a Korean population did not observe a difference in the mean plasma concentration-time profile of *(S)*-naproxen in *CYP2C9*1/*3,* compared to *CYP2C9* reference individuals [22]. However, lack of change in *(S)*-naproxen concentration alone, does not provide evidence for the absence of a pharmacogenetic-pharmacokinetic relationship between *CYP2C9* genotype and naproxen metabolism because *(S)*-naproxen is primarily glucuronidated (60% of the dose) [14]. Approximately 20% of the dose is eliminated as *(S)*-O-desmethyl naproxen and its secondary glucuronide and sulfate metabolites [12-15]. Therefore, to detect the effect of *CYP2C9* variation on *(S)*-naproxen, it is necessary to consider both the unchanged *(S)*-naproxen as well as its metabolites that are cleared through a CYP2C9-mediated pathway, as was done in the current study. Furthermore, the *in vitro* experiments conducted here demonstrate that at physiologically relevant concentrations, CYP2C9 is the major enzyme responsible for naproxen O-dealkylation and that CYP1A2 only plays a minor role. Moreover, the results of inhibitor experiments conducted in single donor HLMs demonstrate that the overall contribution of CYP1A2 to *(S)*-O-desmethyl naproxen formation does not increase substantially with increasing CYP1A2 protein abundance (**Table 3.4**). Thus, elevated CYP1A2 expression and activity, due to genotype [23] or xenobiotic exposure [24,25], is not expected to significantly impact CYP2C9's predominant role in the O-desmethylation of *(S)*-naproxen *in vivo*.

The identification of a novel *CYP2C9* variant that impairs enzyme function, and is unique to a population under-represented in biomedical, and especially genetic research [24], illustrates the importance of population-specific pharmacogenetic studies to guide medication therapy. A pharmacogenetic algorithm that is based on polymorphisms from a specific sub-set of the global

population may not be as clinically beneficial for populations in which the frequency of variant alleles is markedly different or if enzyme activity is determined by uncharacterized genetic variation. This was demonstrated by the conflicting results published by two randomized clinical trials, the European Pharmacogenetics of Anticoagulant Therapy (EU-PACT) [25], and the Clarification of Optimal Anticoagulation through Genetics (COAG) [26] trials. The EU-PACT trial showed a benefit for genotype-guided warfarin dosing over standard clinical care, but the COAG trial did not find a significant difference between the two groups [25,26]. Variation in the ethnicities and genetics of the sample populations likely contributed to the different results [27]. While the EU-PACT participants were primarily European, the COAG study population was composed of 27% African-Americans, who have lower frequencies of *CYP2C9*2* and **3* (the only *CYP2C9* variant alleles considered in the pharmacogenetic algorithm), but higher frequencies of other reduced function *CYP2C9* variants (e.g., **5, *6, *8,* and **11*) [28]. Similarly, current pharmacogenetic warfarin dose algorithms would likely not optimize warfarin dosing for the Yup'ik population, where the *CYP2C9*2* and **3* frequencies are low and novel reduced or loss of function variants such as *M1L* are present. The Yup'ik population may benefit from the consideration and inclusion of population-specific genetic variation in clinical decisions surrounding personalized medication therapy. Clearly, an understanding of genetic variation in under-represented minority populations is essential if pharmacogenetic testing is to reach its optimal clinical utility in patients of all ethnicities.

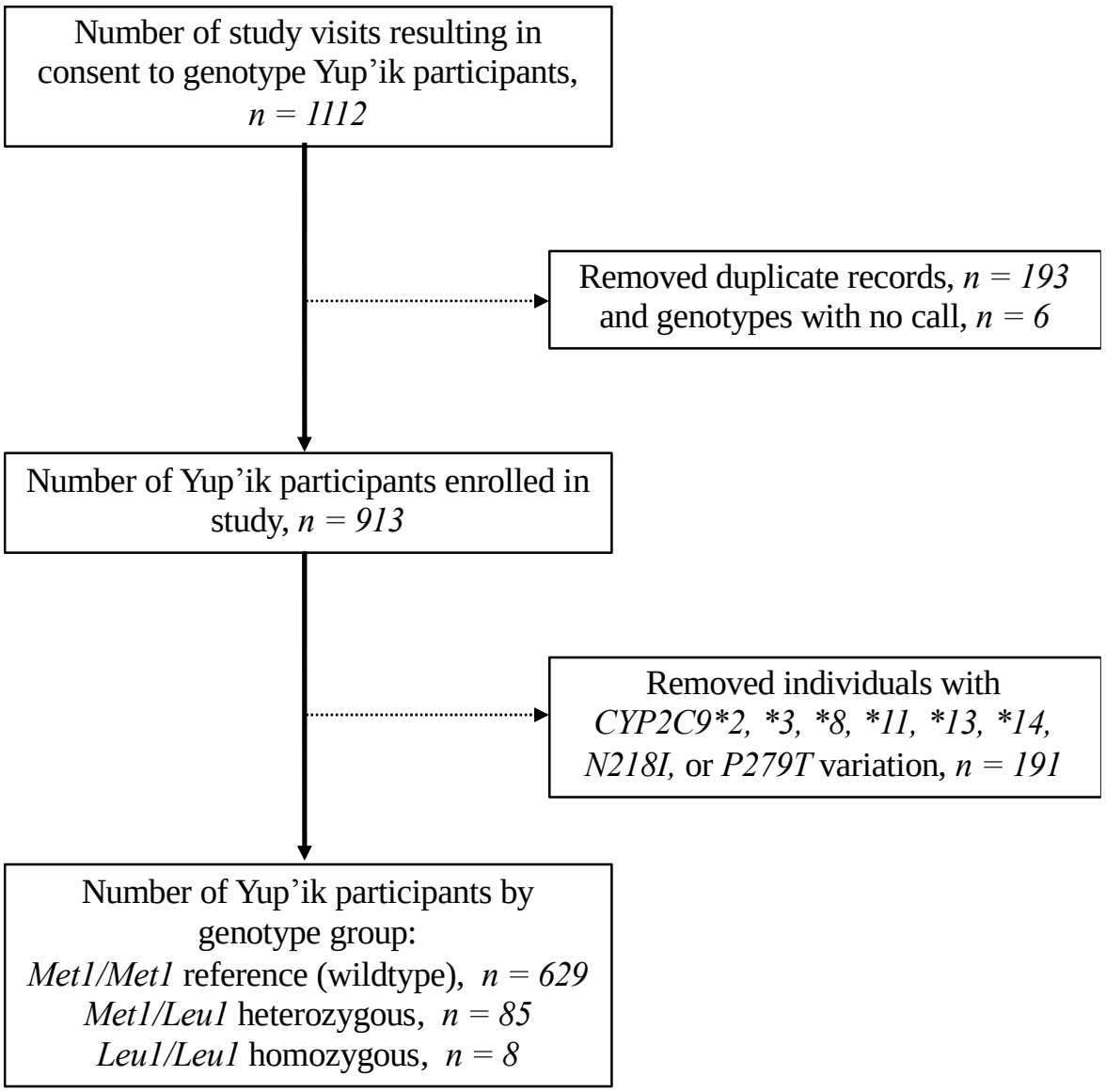

Figure 3.1. Study inclusion based on *CYP2C9 M1L* screening results.

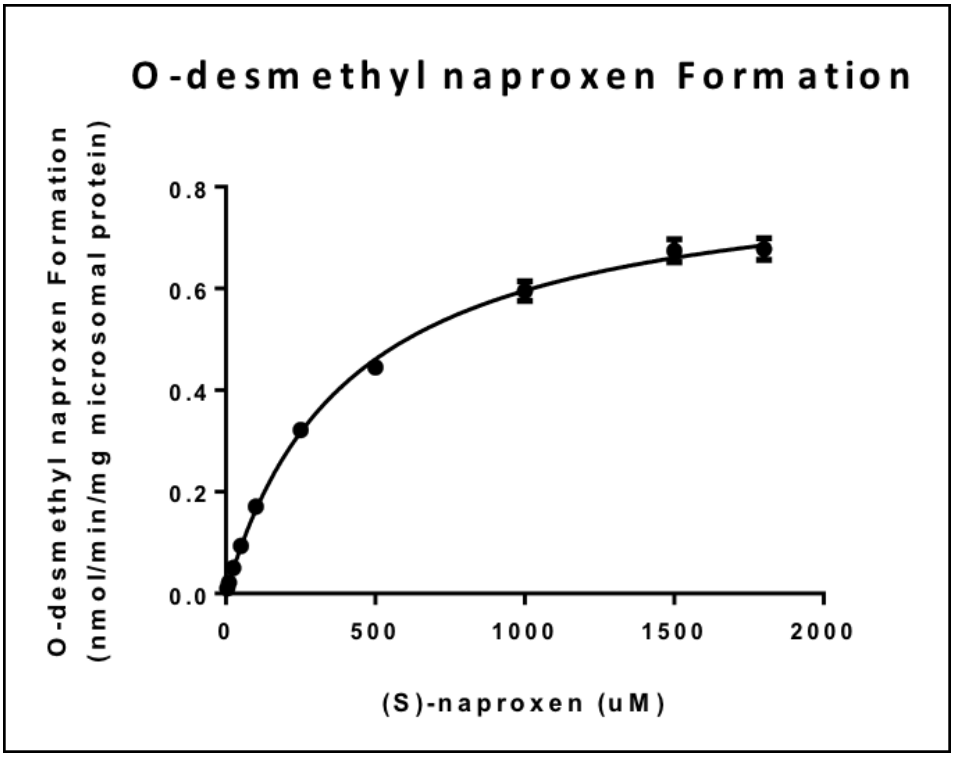

Figure 3.2. Representative Michaelis-Menten plot from a single experiment.

Table 3.1. Kinetic parameters for *(S)*-naproxen metabolism in pooled HLMs.

Kinetic Parameter	**Mean ± stdev**
K_m (μM)	420 ± 2
V_{max} (nmol/min/mg microsomal protein)	0.92 ± 0.06
CL_{int} (μL/hr/mg microsomal protein)	130 ± 8

Kinetic parameters reported are the mean values from three repeated experiments.

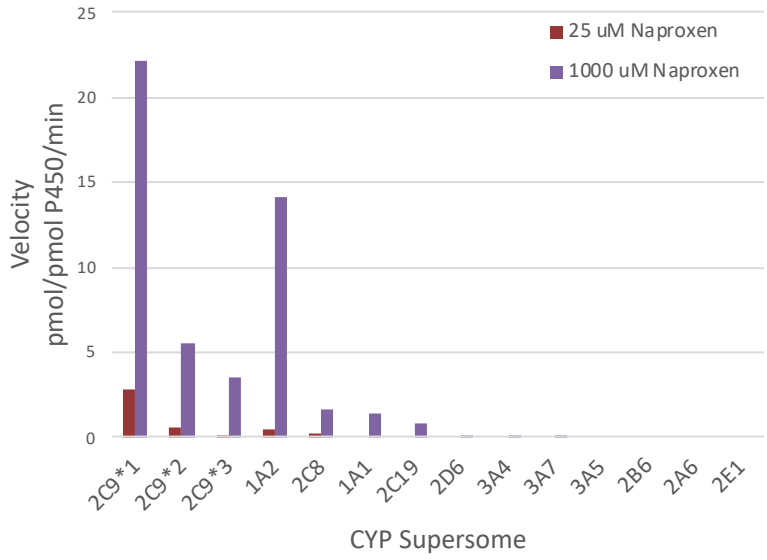

Figure 3.3. CYP Supersome screen at a clinically relevant *(S)*-naproxen concentration (25 μM) and a saturating concentration (1000 μM).

Table 3.2. *(S)*-O-desmethyl naproxen formation by CYP Supersomes at a clinically relevant concentration (25 μM) and a saturating concentration (1000 μM).

	Velocity (pmol/pmol P450/min)	
P450	**25 uM**	**1000 uM**
2C9*1	2.85	22.16
2C9*2	0.51	5.47
2C9*3	0.14	3.48
1A2	0.43	14.13
2C8	0.27	1.65
1A1	-	1.40
2C19	-	0.75
2D6	-	0.11
3A4	-	0.03
3A7	-	0.03
3A5	-	-
2B6	-	-
2A6	-	-
2E1	-	-

Reported velocity is the mean value from 2 repeated experiments, each with technical triplicates.

Dashes indicate that no product formation was detected.

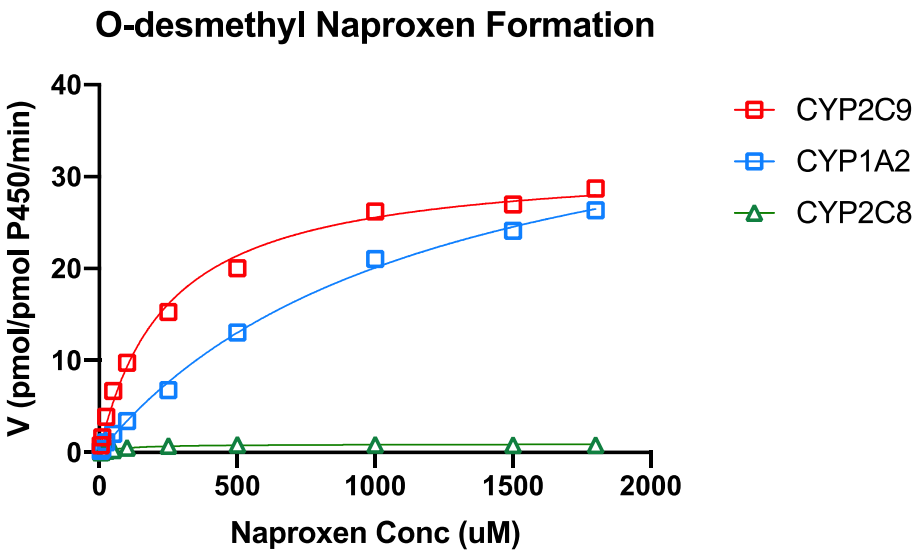

Figure 3.4. Representative Michaelis-Menten plot of *(S)*-O-desmethyl naproxen formation by CYP2C9, CYP1A2, and CYP2C8 Supersomes.

Table 3.3. Kinetic parameters for *(S)*-O-desmethyl naproxen formation by CYP2C9, CYP1A2, and CYP2C8 Supersomes.

P450	K_m (μM)	V_{max} (pmol/min/pmol P450)	CL_{int} (uL/min/pmol P450)
CYP2C9	280 ± 8.9	31.7 ± 2.5	0.11 ± 0.10
CYP1A2	1000 ± 97	41.7 ± 2.4	0.04 ± 0.003
CYP2C8	90	0.90	0.01

Reported parameters are the mean values across three repeated experiments for CYP2C9 and CYP1A2 Supersomes, (duplicate experiments for CYP2C8), each with technical duplicates.

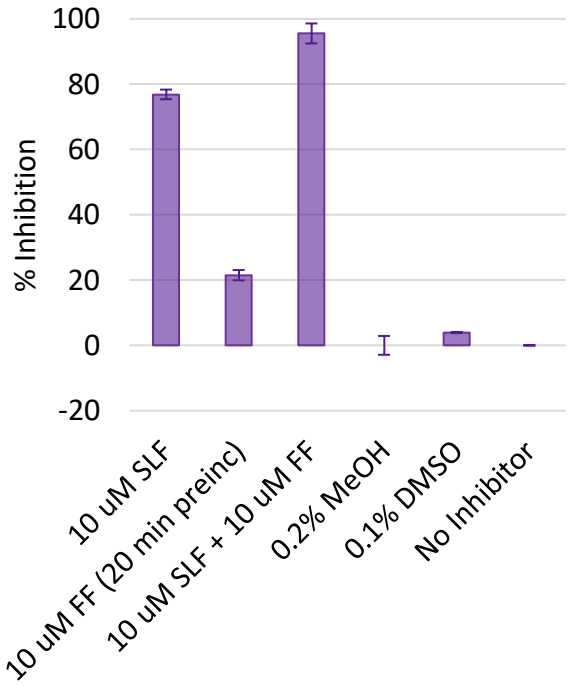

Figure 3.5. Effects of selective CYP2C9 and CYP1A2 inhibitors on *(S)*-O-desmethyl naproxen formation in pooled HLMs.

Sulfaphenazole (SLF) was dissolved in methanol (MeOH) <0.2% final concentration, and furafylline (FF) was dissolved in DMSO, <0.1% final concentration.

Table 3.4. Inhibition of *(S)*-O-desmethyl naproxen formation in single donor HLMs by sulfaphenazole and furafylline.

CYP1A2 Extreme	**High** *n = 5*	**Low** *n = 5*
Avg 1A2 (pmol/mg microsomal protein)	31.2 ± 10.8	2.8 ± 2.3
Avg 2C9 (pmol/mg microsomal protein)	53.2 ± 13.3	36.6 ± 6.3
Furafylline Inhibition (%)	39.7 ± 7.0	23.6 ± 7.6
Sulfaphenazole Inhibition (%)	65.5 ± 4.1	85.2 ± 11.8

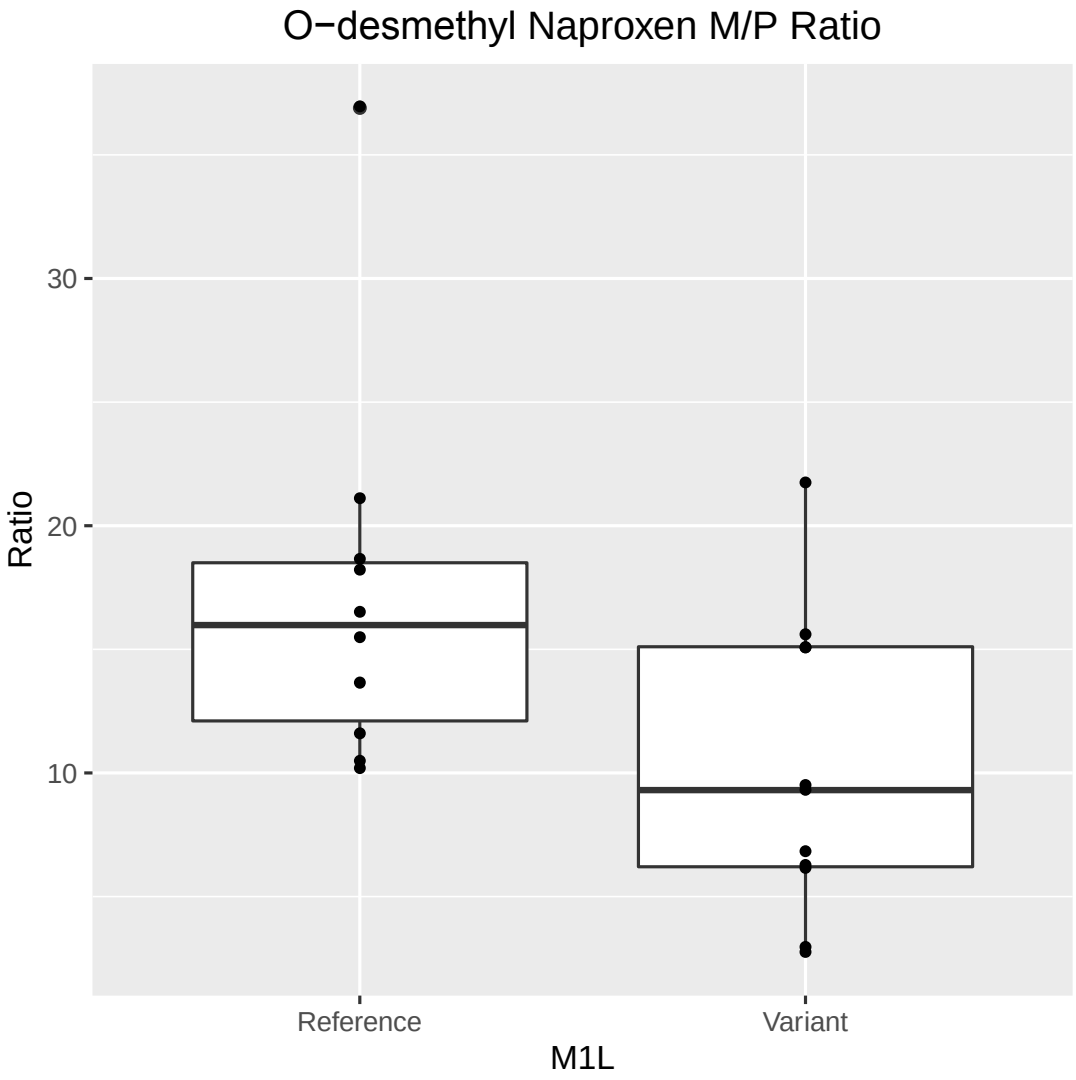

Figure 3.6. Urinary metabolite to parent ratio of *(S)*-O-desmethyl naproxen to unchanged *(S)*-naproxen by *M1L* genotype.

3.5 REFERENCES

1. Zanger, U.M.; Turpeinen, M.; Klein, K.; Schwab, M. Functional pharmacogenetics/genomics of human cytochromes P450 involved in drug biotransformation. *Anal Bioanal Chem* **2008**, *392*, 1093-1108, doi:10.1007/s00216-008-2291-6.
2. Van Booven, D.; Marsh, S.; McLeod, H.; Carrillo, M.W.; Sangkuhl, K.; Klein, T.E.; Altman, R.B. Cytochrome P450 2C9-CYP2C9. *Pharmacogenet Genomics* **2010**, *20*, 277-281, doi:10.1097/FPC.0b013e3283349e84.
3. Flora, D.R.; Rettie, A.E.; Brundage, R.C.; Tracy, T.S. CYP2C9 Genotype-Dependent Warfarin Pharmacokinetics: Impact of CYP2C9 Genotype on R- and S-Warfarin and Their Oxidative Metabolites. *J Clin Pharmacol* **2017**, *57*, 382-393, doi:10.1002/jcph.813.
4. Caudle, K.E.; Rettie, A.E.; Whirl-Carrillo, M.; Smith, L.H.; Mintzer, S.; Lee, M.T.; Klein, T.E.; Callaghan, J.T.; Consortium, C.P.I. Clinical pharmacogenetics implementation consortium guidelines for CYP2C9 and HLA-B genotypes and phenytoin dosing. *Clin Pharmacol Ther* **2014**, *96*, 542-548, doi:10.1038/clpt.2014.159.
5. Johnson, J.A.; Caudle, K.E.; Gong, L.; Whirl-Carrillo, M.; Stein, C.M.; Scott, S.A.; Lee, M.T.; Gage, B.F.; Kimmel, S.E.; Perera, M.A., et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Pharmacogenetics-Guided Warfarin Dosing: 2017 Update. *Clin Pharmacol Ther* **2017**, *102*, 397-404, doi:10.1002/cpt.668.
6. Fohner, A.E.; Robinson, R.; Yracheta, J.; Dillard, D.A.; Schilling, B.; Khan, B.; Hopkins, S.; Boyer, B.; Black, J.; Wiener, H., et al. Variation in genes controlling warfarin disposition and response in American Indian and Alaska Native people: CYP2C9, VKORC1, CYP4F2, CYP4F11, GGCX. *Pharmacogenet Genomics* **2015**, *25*, 343-353, doi:10.1097/FPC.0000000000000143.
7. Becker, M.L.; Visser, L.E.; Trienekens, P.H.; Hofman, A.; van Schaik, R.H.; Stricker, B.H. Cytochrome P450 2C9 *2 and *3 polymorphisms and the dose and effect of sulfonylurea in type II diabetes mellitus. *Clin Pharmacol Ther* **2008**, *83*, 288-292, doi:10.1038/sj.clpt.6100273.
8. Miners, J.O.; Coulter, S.; Tukey, R.H.; Veronese, M.E.; Birkett, D.J. Cytochromes P450, 1A2, and 2C9 are responsible for the human hepatic O-demethylation of R- and S-naproxen. *Biochem Pharmacol* **1996**, *51*, 1003-1008.
9. Tracy, T.S.; Marra, C.; Wrighton, S.A.; Gonzalez, F.J.; Korzekwa, K.R. Involvement of multiple cytochrome P450 isoforms in naproxen O-demethylation. *Eur J Clin Pharmacol* **1997**, *52*, 293-298.
10. Runkel, R.A.; Kraft, K.S.; Boost, G.; Sevelius, H.; Forchielli, E.; Hill, R.; Magoun, R.; Szakacs, J.B.; Segre, E. Naproxen oral absorption characteristics. *Chem Pharm Bull (Tokyo)* **1972**, *20*, 1457-1466.
11. Aleve® (naproxen sodium) [product monograph]. Morristown, NJ: Bayer.
12. Davies, N.M.; Anderson, K.E. Clinical pharmacokinetics of naproxen. *Clin Pharmacokinet* **1997**, *32*, 268-293, doi:10.2165/00003088-199732040-00002.
13. Kiang, C.H.; Lee, C.; Kushinsky, S. Isolation and identification of 6-desmethylnaproxen sulfate as a new metabolite of naproxen in human plasma. *Drug Metab Dispos* **1989**, *17*, 43-48.

14. Vree, T.B.; van den Biggelaar-Martea, M.; Verwey-van Wissen, C.P.; Vree, J.B.; Guelen, P.J. Pharmacokinetics of naproxen, its metabolite O-desmethylnaproxen, and their acyl glucuronides in humans. *Biopharm Drug Dispos* **1993**, *14*, 491-502.
15. Sugawara, Y.; Fujihara, M.; Miura, Y.; Hayashida, K.; Takahashi, T. Studies on the fate of naproxen. II. Metabolic fate in various animals and man. *Chem Pharm Bull (Tokyo)* **1978**, *26*, 3312-3321.
16. Norris, T.; Vines, P.L.; Hoeffel, E.M. The American Indian and Alaska Native Population: 2010. In *2010 Census Briefs*, U.S. Department of Commerce Economics and Statistics Administration U.S. CENSUS BUREAU: 2012.
17. Shirasaka, Y.; Chaudhry, A.S.; McDonald, M.; Prasad, B.; Wong, T.; Calamia, J.C.; Fohner, A.; Thornton, T.A.; Isoherranen, N.; Unadkat, J.D., et al. Interindividual variability of CYP2C19-catalyzed drug metabolism due to differences in gene diplotypes and cytochrome P450 oxidoreductase content. *Pharmacogenomics J* **2016**, *16*, 375-387, doi:10.1038/tpj.2015.58.
18. Rowland-Yeo, K.; Rostami-Hodjegan, A.; Tucker, G. Abundance of cytochromes P450 in human liver: a
meta-analysis. British Journal of Clinical Pharmacology, 2004; Vol. 57, pp 687–688.
19. Ferguson, R.J.; De Morais, S.M.; Benhamou, S.; Bouchardy, C.; Blaisdell, J.; Ibeanu, G.; Wilkinson, G.R.; Sarich, T.C.; Wright, J.M.; Dayer, P., et al. A new genetic defect in human CYP2C19: mutation of the initiation codon is responsible for poor metabolism of S-mephenytoin. *J Pharmacol Exp Ther* **1998**, *284*, 356-361.
20. Consortium, G.P.; Auton, A.; Brooks, L.D.; Durbin, R.M.; Garrison, E.P.; Kang, H.M.; Korbel, J.O.; Marchini, J.L.; McCarthy, S.; McVean, G.A., et al. A global reference for human genetic variation. *Nature* **2015**, *526*, 68-74, doi:10.1038/nature15393.
21. Rodrigues, A.D. Impact of CYP2C9 genotype on pharmacokinetics: are all cyclooxygenase inhibitors the same? *Drug Metab Dispos* **2005**, *33*, 1567-1575, doi:10.1124/dmd.105.006452.
22. Bae, J.W.; Kim, J.H.; Choi, C.I.; Kim, M.J.; Kim, H.J.; Byun, S.A.; Chang, Y.S.; Jang, C.G.; Park, Y.S.; Lee, S.Y. Effect of CYP2C9*3 allele on the pharmacokinetics of naproxen in Korean subjects. *Arch Pharm Res* **2009**, *32*, 269-273, doi:10.1007/s12272-009-1232-z.
23. Thorn, C.F.; Aklillu, E.; Klein, T.E.; Altman, R.B. PharmGKB summary: very important pharmacogene information for CYP1A2. *Pharmacogenet Genomics* **2012**, *22*, 73-77, doi:10.1097/FPC.0b013e32834c6efd.
24. Zevin, S.; Benowitz, N.L. Drug interactions with tobacco smoking. An update. *Clin Pharmacokinet* **1999**, *36*, 425-438, doi:10.2165/00003088-199936060-00004.
25. Dobrinas, M.; Cornuz, J.; Oneda, B.; Kohler Serra, M.; Puhl, M.; Eap, C.B. Impact of smoking, smoking cessation, and genetic polymorphisms on CYP1A2 activity and inducibility. *Clin Pharmacol Ther* **2011**, *90*, 117-125, doi:10.1038/clpt.2011.70.
26. Popejoy, A.B.; Fullerton, S.M. Genomics is failing on diversity. *Nature* **2016**, *538*, 161-164, doi:10.1038/538161a.
27. Pirmohamed, M.; Burnside, G.; Eriksson, N.; Jorgensen, A.L.; Toh, C.H.; Nicholson, T.; Kesteven, P.; Christersson, C.; Wahlström, B.; Stafberg, C., et al. A randomized trial of genotype-guided dosing of warfarin. *N Engl J Med* **2013**, *369*, 2294-2303, doi:10.1056/NEJMoa1311386.

28. Kimmel, S.E.; French, B.; Kasner, S.E.; Johnson, J.A.; Anderson, J.L.; Gage, B.F.; Rosenberg, Y.D.; Eby, C.S.; Madigan, R.A.; McBane, R.B., et al. A pharmacogenetic versus a clinical algorithm for warfarin dosing. *N Engl J Med* **2013**, *369*, 2283-2293, doi:10.1056/NEJMoa1310669.
29. Scott, S.A.; Lubitz, S.A. Warfarin pharmacogenetic trials: is there a future for pharmacogenetic-guided dosing? *Pharmacogenomics* **2014**, *15*, 719-722, doi:10.2217/pgs.14.18.
30. Limdi, N.A.; Brown, T.M.; Yan, Q.; Thigpen, J.L.; Shendre, A.; Liu, N.; Hill, C.E.; Arnett, D.K.; Beasley, T.M. Race influences warfarin dose changes associated with genetic factors. *Blood* **2015**, *126*, 539-545, doi:10.1182/blood-2015-02-627042.

Chapter 4. IMPACT OF *CYP2C9, VKORC1, CYP4F2, CYP4F11,* AND *GGCX* GENE VARIANTS ON STABLE WARFARIN DOSE IN THE AN/AI POPULATION

(A version of this chapter was published as "*VKORC1* and Novel *CYP2C9* Variation Predict Warfarin Response in Alaska Native and American Indian People." (2019) Clinical Translational Science 12(3):312-320)

4.1 INTRODUCTION

While variation in *VKORC1, CYP2C9, CYP4F2* and *GGCX* genes have all been associated with the warfarin dose needed to achieve a common degree of anticoagulation response [1-5], these findings are based largely on alleles and associated frequencies discovered in populations with limited ethnic/racial heterogeneity (i.e. Caucasian, Asian, and African American), and may not be generalizable to other, less studied populations such as the indigenous people of North America [6,7]. In that regard, we recently reported the discovery within the indigenous population of Alaska (Alaska Native and American Indian, AN/AI) of common, novel coding variation in the *CYP2C9* gene (*M1L* and *N218I*) predicted to decrease CYP2C9 enzyme function [8]. We also observed relatively high frequencies of known function-altering *VKORC1 (-1639G>A*) and *CYP4F2 (*3*, V433M) alleles in some, but not all AN/AI subgroups. These data warrant further investigation to understand their clinical impact on warfarin therapy in AN/AI people. Variation in *CYP4F11* was included for testing, as the gene product can also catalyze vitamin K catabolism [9]. Thus, the goal of this investigation was to determine whether inheritance of *CYP2C9, VKORC1, CYP4F2, CYP4F11,* and *GGCX* gene variants, particularly novel variants in an AN/AI population, affect the dose of warfarin required to achieve a therapeutic INR in order to better understand the value of

genetic testing to guide warfarin therapy for the AN/AI population, and potentially other indigenous peoples of North America.

4.2 METHODS

4.2.1 *Setting*

Southcentral Foundation (SCF), a tribally owned and operated regional health corporation provides pre-paid healthcare services to 65,000 AN/AI customer-owners. The Anchorage Service Unit (ASU) and Cook Inlet Region Villages served by SCF are comprised of both urban and rural areas, including Anchorage, the Matanuska-Susitna Borough, and 76 outlying villages (most with fewer than 500 residents). It provides primary care services to 46% of the AN population in the ASU at six SCF primary care clinics on the Alaska Native Medical Center (ANMC) campus where participant recruitment took place.

4.2.2 *Study Participants*

Between 2011 and 2013 a representative convenience sample of 118 AN/AI customer-owners, ≥18 years of age, receiving warfarin therapy at SCF, were recruited, and consent obtained by research staff members at SCF's primary care clinics. Study participants completed a short demographic questionnaire (self-reported gender, date of birth, and self-reported heritage). Consented customer-owners were then provided two small, sterile swabs to collect epithelial cheek cells for DNA analysis of *CYP2C9*, *VKORC1*, *CYP4F2*, *CYP4F11*, and *GGCX* gene variation. Swabs were then placed in sterile tubes and were stored at −80°C until genotyping analysis. were then placed in sterile tubes and were stored at −80°C until genotyping analysis.

4.2.3 *Study Design*

The Alaska Area Institutional Review Board (IRB), and the SCF and Alaska Native Tribal Health Consortium tribal review boards approved work conducted at SCF on the ANMC campus. The University of Washington (UW) IRB approved the overall research project, as UW is the academic home of the grant funding this research (Pharmacogenetics in Rural and Underserved Populations) and its principal investigators. The National Institute of General Medical Sciences and the Indian Health Service granted a Certificate of Confidentiality for protection of customer-owner information, and the respective Alaska Area IRB approved forms for written consent prior to initiating research. Community based participatory research at SCF and the Center for Alaska Native Health Research was used to develop research questions.

This retrospective cohort study was conducted at one anticoagulation clinic based in Anchorage, Alaska. Customer-owner care for this study was managed by a credentialed anticoagulation pharmacist with physician oversight. A standardized approach aided by commercial anticoagulation software was employed and follow-up averaged a little more than two weeks. All customer-owners below 65 years of age received the same initial dose, 5 mg/day, with subsequent dose adjustments made based on INR results.

Customer-owners' medical records were retroactively queried by SCF Data Services Department staff for specific data elements (i.e., ICD-9 codes and diagnoses, purpose, and explanation of the data queried). For elements extracted prior to October 1, 2011, the Legacy electronic medical record system Resource Patient Management System was used, and for dates on or later than October 1, 2011, the Cerner electronic medical record system was used. Data abstracted included: dates of warfarin initiation/stabilization and indication for treatment (deep vein thrombosis, pulmonary embolism, pulmonary hypertension, atrial fibrillation/flutter,

cardiomyopathy, left ventricular dilation, stroke, post-orthopedic variables, valve replacement, bleeding events, and thromboembolic events). Comorbidities recorded included chronic liver disease, heart disease, diabetes, valve replacement, stroke, anemia, kidney dysfunction, ulcers, hyperlipidemia, and cancer, as well as the use of medications known to interact with warfarin. The following demographic/background variables that may impact dose of warfarin or likelihood of bleeding events were also collected: gender, age, heritage, height, and weight. Password protected electronic databases were used to store customer-owner questionnaire and medical record information.

To account for population substructure within the SCF cohort, customer-owners were asked for self-reported tribal affiliation. Customer-owners classified as AN included the following tribes: Inupiaq, Athabascan subgroups, Tlingit, Tsimshian, Haida, Eyak, Aleut/Unangan, Central Yup'ik, Cup'ik, Sugpiaq/Alutiiq. Customer-owners also were given the option of choosing affiliation with multiple tribes in the lower 48 states of the US and were classified as AI. Due to the small sample size, we intentionally grouped these customer-owners by their shared heritage and geographic proximity for data analysis, though we acknowledge that each tribe has its own unique culture and history.

4.2.4 *Genotyping Methods and Linkage Disequilibrium Calculations*

Genomic DNA from buccal swabs was extracted using a QIAamp DNA Blood Midi/Maxi kit (Qiagen, Valencia, California, USA). Quality and concentration of DNA were determined using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Wilmington, Delaware, USA). Population relevant SNVs and small insertions/ deletions were identified previously through gene resequencing in 188 AN/AI customer-owners in partnership with SCF in Anchorage, AK [8]. DNA samples from the current study participants were genotyped for the novel and known variants of

interest using the Fluidigm Platform (South San Francisco, CA), as previously reported [8]. This included 8 SNVs in *CYP2C9,* 4 in *CYP4F2*, 1 in *CYP4F11*, 2 in *VKORC1*, and 2 in *GGCX*. Samples with overall call rates below 95% were removed from further analysis. Of all DNA samples genotyped, 9 were excluded from further analysis due to call rates below 95%. The no call rate was 0.4%. Allele frequencies and pairwise linkage disequilibrium (LD) were calculated using Haploview 4.2 software [10]. All SNVs identified were tested for deviations from Hardy–Weinberg equilibrium using a χ^2-test. Allele frequency results were compared to previously reported AN/AI sequencing results from SCF customer-owners [8].

4.2.5 *Outcomes*

The outcome variable for this study was stable therapeutic warfarin dose (mg/day), which was ascertained from available clinical data and prescription records from 1/2000 – 6/2017. Stable warfarin dose was defined in two ways: (INR-based) the dose of warfarin required to achieve an INR within the target range, at least six months after starting warfarin therapy, with two matching doses at least two weeks apart; or (consecutive, non-INR-based) two matching consecutive doses at least six months after starting warfarin therapy and at least two weeks apart. The INR was considered in therapeutic range if it was measured to be within the target range for a given indication. For example, if the target INR was 2.0 to 3.0, then the first INR result between 2.0 to 3.0, and at least six months after initiating therapy, was defined as within range. Once the customer-owner was determined to reach their stable warfarin dose, his/her maintenance dose was recorded, allowing for comparison by various genotypes in warfarin pharmacogenes.

4.2.6 *Statistical Analysis*

Customer-owners were included in two groups based on available clinical data: INR-based stable warfarin dose and consecutive (non-INR based) dose. All customer owners included in the INR-based cohort were also included in the consecutive cohort. For each cohort, associations between pharmacogenetics and pharmacodynamics were evaluated with a univariate and multivariate linear regression models using RStudio version 1.0.143 (RStudio, Inc., Boston, MA). To account for heteroscedasticity, robust standard errors were calculated using the lmtest package in R. Covariates included in the International Warfarin Pharmacogenetics Consortium (IWPC) pharmacogenetic dosing algorithm were considered; however, no customer-owners in this analysis were using CYP2C9 inducers concurrently and the races in the IWPC algorithm (Asian, Black, or Caucasian) were not used [11], instead we used heritage (AN or AI). Height and weight were collected at the time of genotyping; however, this may not correspond to the same height and weight of a customer-owner at the time of reaching stable warfarin dose, so body mass index was not included in the multivariate regression model.

We elected to include age, gender, and heritage in the multivariate regression model *a priori*, as these demographic variables have been shown to influence warfarin dose across many world populations [11]. The potential confounding effects of the concurrent use of CYP2C9 inhibitors (amiodarone, fluconazole, and disulfiram), statins (simvastatin), or antibiotics on warfarin dose were also evaluated. Concurrent medication classes were added to the model one at a time to assess potential confounding effects on warfarin dose and were included in the multivariate regression model if the medication was used by at least 10 customer-owners in the study population and if the covariate significantly impacted warfarin dose. The primary analysis assessed associations between single SNVs and INR-based stable warfarin dose adjusting for age,

gender, AN or AI heritage, and concurrent statin use in a multivariate regression model. For SNVs that included reference, heterozygote, and homozygous variant observations, an additive model was used for genotype.

Secondary analysis of the data was conducted, testing for associations with a combination of impaired function *CYP2C9* variants as well as adjusting for SNVs found to be significantly associated with stable warfarin dose in the primary analysis model, allowing for detection of stronger, independent signals from other variants potentially associated with dose. The primary and secondary analysis was repeated in the consecutive (non-INR based) stable warfarin dose cohort. Bonferroni adjusted p-values were used to correct for multiple testing of five independent tests of the five candidate loci of interest. The SCF team independently validated the statistical analysis by reviewing and running all R code. Any discrepancies in interpretation were discussed among the team and consensus was reached.

4.3 RESULTS

A total of 118 AN/AI customer-owners at SCF receiving warfarin therapy were enrolled in this study. **Figure 4.1** depicts the number of customer-owners included in the final analysis and **Table 4.1** describes study participant characteristics and stable warfarin dose. The primary INR-based stable warfarin dose definition (dose required to achieve an INR within the target range, at least six months after starting warfarin therapy, with two matching doses, at least two weeks apart) included 50 customer-owners in the genotype-phenotype analysis, with a breakdown of 43 AN and 7 AI individuals. The secondary non-INR-based warfarin dose definition (two matching consecutive doses six months after starting warfarin therapy, and at least two weeks apart) included 78 customer-owners, with 68 AN and 10 AI individuals. The average stable warfarin doses for the

INR-based dose definition and the consecutive (non-INR-based) dose definition cohorts were similar (5.0 vs. 4.9 mg/day) and range (1.7 – 13.0 vs. 1.5 – 12.5 mg/day) (**Table 4.1**).

The multivariate regression analysis model controlled for age, gender, self-reported heritage, and concurrent statin use. We elected to include age, gender, and heritage in the model *a priori*, although only age and heritage were found to be significantly associated with stable warfarin dose by univariate regression analysis (**Table 4.2**). Statins were the only medication class included in the multivariate regression model, as they were associated with a 1.7 mg/day decrease in warfarin dose ($p < 0.01$) and more than 10 customer-owners in the cohort were using simvastatin concomitantly. There were not enough observations of CYP2C9 inhibitors or antibiotics to include these medications in the regression model.

The results of the analysis adjusting for age, heritage, gender, and concurrent statin use are shown in **Table 4.3**. Among the INR-based stable warfarin dose cohort of 50 customer-owners, *VKORC1 -1639G>A* and *VKORC1 1173C>T* were significantly associated with stable dose, decreasing the dose required to achieve therapeutic INR by 1.7 mg/day per allele (t-test of coefficients, unadjusted p = 1.4e-05, Bonferroni adjusted p = 7.0e-05). This relationship remained significant when only AN customer-owners were included in the analysis (Bonferroni adjusted p = 3.3e-04). *VKORC1 -1639G>A* and *1173C>T* were in full linkage disequilibrium (**Figure 4.2**), and thus their effect size was the same. *VKORC1* genotype explained 34% of warfarin dose variability (R^2) in the AN/AI population at SCF. The mean stable dose by *VKORC1 -1639G>A* reference (wildtype), heterozygote, and homozygous variant genotype was 7.5, 5.2 and 3.2 mg/day, respectively, by univariate regression analysis ($p < 0.05$ for each genotype comparison) (**Figure 4.3A**).

The mean stable dose for *CYP2C9 N218I* heterozygotes was significantly lower than reference individuals (3.7 vs. 5.1 mg/day, p = 0.005) by univariate regression analysis (**Figure 4.3B**). In the multivariate regression analysis, the stable warfarin dose for customer-owners carrying one copy of the *N218I* allele was 1.1 mg/day lower compared to individuals carrying two reference alleles, although the difference missed statistical significance (t-test of coefficients, p = 0.077) (**Table 4.3**). Other variants identified, *GGCX-intron14*, *CYP4F2 g72220026G, CYP4F2*3, CYP4F11 R276C*, and *CYP2C9*2* and **3*, were not significantly associated with warfarin dose requirement in the INR-based cohort. *CYP2C9*8, CYP2C9*11, CYP2C9*14,* CYP2C9 *P279T,* and *GGCXG421A* variant alleles were not present in this cohort, and *CYP2C9 M1L, CYP4F2 SpliceCG,* and *CYP4F2 G185V*, which had been detected previously in the AN/AI population of Alaska [8], had too few observations of variant alleles in the study population to accurately assess their effect on warfarin dose. While the combination of *CYP2C9 M1L* or *N218I* heterozygotes compared to reference individuals was not significantly associated with stable warfarin dose, the average effect trended in the expected direction by univariate analysis (-0.99 mg/day, p = 0.070) as well as by multivariate regression analysis (-0.89 mg/day, p = 0.13). No additional SNVs were found to be significantly associated with dose in a secondary analysis that adjusted for the low-dose associated *VKORC1* gene variants in the multivariate regression model. The combination of *VKORC1 -1639G>A* and *CYP2C9 N218I*, as well as age, gender, and statin use in the multivariate regression model explained 44% of the variation in the therapeutic warfarin dose in the INR-based dose definition cohort (adjusted $R^2 = 0.38$).

Interestingly, AN customer-owners required a significantly lower daily warfarin dose to achieve therapeutic INR (4.6 mg/day), compared to those of AI heritage (7.0 mg/day) (t-test of coefficients, p = 0.049) (**Figure 4.4**). This was well explained by a substantial difference in the

diagnostic *VKORC1* allele frequencies in these heritage groups, with heritage losing statistical significance in the multivariate regression model when evaluating the gene-dose effect from *VKORC1*. The MAF of the low-dose associated *VKORC1 -1639G>A* variant was 61.6% the AN population, compared to 28.6% in AI customer-owners in this study cohort.

Analysis of the consecutive (non-INR-based) warfarin dose definition in 78 customer-owners showed similar statistical trends. *VKORC1 -1639G>A* and *VKORC1 1173C>T* remained significantly associated with a lower warfarin dose requirement ($p < 0.05$), as was *CYP2C9 N218I* ($p < 0.05$) (**Table 4.4**). While not found to be associated with warfarin dose in the INR-based cohort, *GGCX-intron14* was suggestively associated ($p = 0.087$) with lower warfarin dose, compared to reference in the consecutive dose definition cohort. Again, *CYP2C9*8, CYP2C9*14,* CYP2C9 *P279T* variant alleles were not present, while *CYP2C9*11, CYP2C9 M1L, GGCX G421A,* and *CYP4F2 SpliceCG* had too few observations of variant alleles to accurately assess their effect on dose. *CYP4F2 g72220026G, CYP4F2*3, CYP4F2 G185V, CYP4F11 R276C*, and *CYP2C9*2* and **3* did not appear to be associated with warfarin dose requirement. No additional SNVs were found to be significantly associated with dose in a secondary analysis that adjusted for the low-dose associated *VKORC1* gene variants in the multivariate regression model.

4.4 DISCUSSION

Warfarin has been the focus of many pharmacogenetic studies, however there are few data specifically addressing the clinical implications of genetic variation in the AN/AI population [6,7]. To our knowledge, this is the first study to evaluate genotype-phenotype associations for warfarin pharmacogenes with stable warfarin dose in an AN/AI population. Results of this retrospective study suggest that [1] *VKORC1* genotype is strongly associated with stable warfarin dose in the AN/AI population, explaining 34% of the variability in a therapeutic warfarin dose (R^2), somewhat

higher than that seen in other populations worldwide; [2] AN people require a lower warfarin dose than AI people on average, explained by a difference in the diagnostic *VKORC1* allele frequencies; [3] the novel *CYP2C9* variant, *N218I,* lowers warfarin dose requirement to a clinically meaningful degree. The combination of *VKORC1 -1639G>A* and *CYP2C9 N218I*, as well as age, gender, and statin use explained 44% of the variation in the therapeutic warfarin dose, establishing the clinical validity of pharmacogenetic associations in this population. Other factors for consideration of clinical utility of pharmacogenetic testing of customer-owners at SCF starting warfarin therapy include cost, access to genetic testing, and potential limitations of the current empirical approach to warfarin dosing when many of the customer-owners live in remote communities throughout the state.

The significantly lower warfarin dose in AN, compared to AI customer owners, was well explained by a higher frequency of the low dose associated *VKORC1* alleles (*-1639G>A* and *1173C>T*) in the AN heritage group. While heritage was associated with stable warfarin dose in a univariate regression model, it was no longer a statistically significant adjustment in the multivariate regression model when evaluating the additive gene-dose effect from *VKORC1*. Body mass index (BMI) between the two heritage groups did not greatly differ, with the mean BMI for the AN and AI customer-owners being 30.7 and 30.5, respectively. Racial/ethnic differences in drug response can arise from multiple non-genetic factors (e.g., diet, environment), but in the case of warfarin it would appear that *VKORC1* variation is the predominant determinant.

The percent of warfarin dose variability explained by *VKORC1* genotype has previously been shown to differ by race, due to differences in *VKORC1* MAFs across racial groups. The low-dose *VKORC1* alleles were found in the highest frequency in East Asian populations and associated with a lower average dose requirement than populations of European or African origin

[3,12]. The *VKORC1* alleles were also significantly associated with a decrease in dose for both European Americans and African Americans, but with a greater effect in Europeans, who had a higher frequency of the *VKORC1* variant allele, compared to African Americans [13].

The relatively small sample size in this study likely reduced our ability to find significant associations with the well-established *CYP2C9* variants (i.e., **2* and **3*). We did not find a significant association with the *CYP2C9*3* allele, but the mean warfarin dose requirement was lower in carriers of the variant, compared to the reference group; 0.14 mg/day less for INR-based stable dose and 0.46 mg/day less for the consecutive (non-INR-based) stable dose definition. In the absence of more definitive data, we hypothesize that those alleles would have the expected effect (reduced dose requirement) on average in the SCF population.

Novel variants unique to the AN/AI population, such as *CYP2C9 M1L* or *N218I,* may also be important to consider. While only three AN customer-owners were heterozygous for *CYP2C9 N218I,* the change is predicted to be disruptive to CYP2C9 enzyme function, as it resulted in a lower warfarin dose in both the INR-based and consecutive stable warfarin dose cohorts. Chapter 2 includes a discussion of a causal relationship between *N218I* variation and altered warfarin metabolism, with the recombinant *N218I* variant enzyme purified from *E. coli* exhibiting ~70% reduction in intrinsic clearance to the major 7-hydroxy and 6-hydroxy warfarin metabolites, compared to wildtype CYP2C9 protein.

The *CYP4F2*3* allele has been associated with reduced hepatic concentrations of the CYP4F2 enzyme, decreasing vitamin K catabolism [14], and was therefore expected to confer an increase in stable warfarin dose requirement in individuals carrying the variant allele. This was not the case. Possible reasons for deviation from the expected result include uncontrolled dietary effects (e.g.,

consumption of vitamin K rich foods) or possibly a difference in the penetrance of the *CYP4F2*3* allele because of other uncontrolled genetic and regulatory factors [15].

With regard to the significant covariates, statins have been shown to decrease the stable warfarin dose requirement, although it is unclear how they may be involved in the causal pathway of warfarin response. Possible explanations are that they inhibit OAT2 mediated uptake of warfarin into hepatocytes [16], or that they affect the composition of lipoproteins that facilitate the trafficking of vitamin K into and out of hepatocytes [17]. As expected, increasing age was significantly associated with a reduction in warfarin dose requirement, reflecting a variety of anatomical and biochemical changes that occur with aging and that can affect drug disposition and response [18]. The lack of a significant association with gender may be attributed to its small effect size and the relatively small sample size of the study, as a causal relationship is expected based on the bulk of the published literature.

The main limitation of this study is the relatively small number of customer-owners included in the primary analysis, however a small sample size is inherent to the study population. There are approximately 113,000 AN/AI people living in Alaska [19], with only a fraction of those receiving warfarin anticoagulation therapy. Nonetheless, it is important to conduct studies such as ours, despite inherent limitations, to expand the knowledge base, guide future research, and support the introduction of precision medicine, where appropriate, for this historically underserved population. The primary INR-based stable dose definition was selected to allow for a more rigorous dose tied to therapeutic INR, increasing precision, but it also reduced the sample size appreciably, decreasing the study's power. By excluding customer-owners who were not receiving warfarin chronically (at least 6 months), we likely excluded individuals being treated for deep venous thrombosis and pulmonary embolism, as these indications typically require three months of

warfarin therapy, whereas atrial fibrillation or mechanical valves require life-long prophylaxis treatment. This allowed for the analysis of customer-owners receiving chronic warfarin therapy instead of the initial dosing period, where warfarin dose is often unstable [20], and during which individuals carrying *CYP2C9*2* or **3* variant alleles require additional dose adjustments and take longer to achieve a stable, therapeutic warfarin dose [1].

Another consequence of the small study sample size is that there were only a few AI customer-owners included in the study. To address this limitation, we included the consecutive (non-INR-based) stable dose definition in an effort to include more customer-owners, and AI individuals, in the final study analysis. Interestingly, the observation of a higher mean therapeutic warfarin dose in customer-owners of AI heritage, compared to AN, in both the INR-based and consecutive dose cohorts is consistent with years of clinical experience by the SCF anticoagulation team. Future work should aim to analyze the effect of genetic variation across the different tribes in the lower 48 states, rather than grouping them together under AI heritage.

In conclusion, in an AN/AI population, *VKORC1 -1639G>A* was significantly associated with a lower warfarin dose requirement, as was a novel *CYP2C9* gene variant (*N218I*). These findings could potentially be incorporated into pharmacogenetic screening to help guide dose selection to improve warfarin safety and efficacy for this rural population. These observations and marked differences in allele frequencies found in AN/AI subgroups [8] compels further investigation in order to better understand the role of genetic variation in warfarin therapy outcomes and potentially optimize personalized medicine for this underserved population. The AN/AI population must be adequately represented in pharmacogenetic studies that assess genotype-phenotype associations in order to fully inform on the dosing of medications with

established clinical associations with pharmacogene variation, especially narrow therapeutic index substrates such as warfarin.

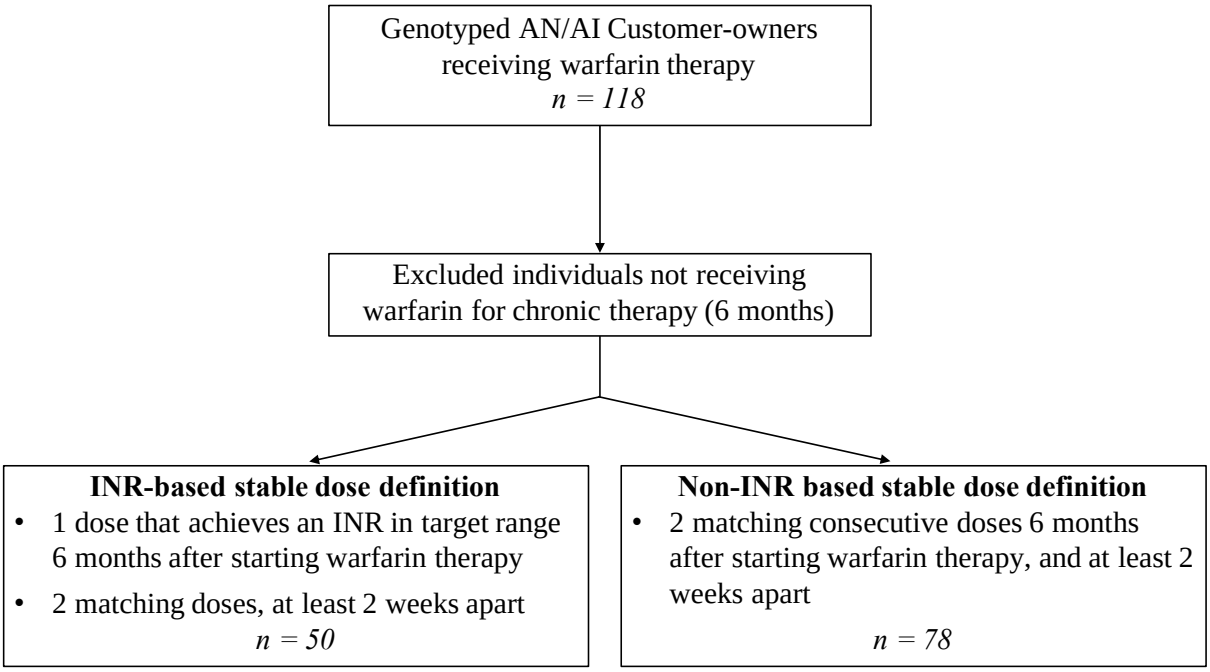

Figure 4.1. Comparison of customer-owner characteristics and warfarin dose in the INR-based and consecutive (non-INR-based) stable warfarin dose definition cohorts.

Table 4.1. Comparison of customer-owner characteristics and warfarin dose in the INR-based and consecutive (non-INR-based) stable warfarin dose definition cohorts.

Variable	INR-based Stable Dose Cohort (*n = 50*)	Consecutive Stable Dose Cohort (*n = 78*)
Demographics		
Alaska Native, No. (%)	43 (86.0)	68 (87.2)
American Indian, No. (%)	7 (14.0)	10 (12.8)
Men, No. (%)	29 (58.0)	45 (57.7)
Age, mean (SD), y	57.9 (14.9)	59.6 (15.0)
BMI, mean (SD)	30.6 (7.6)	30.7 (6.9)
Indication for warfarin, No. (%)		
Cardiomyopathy / Left Ventricular Dilation	1 (2.0)	4 (5.1)
Deep Vein Thrombosis	8 (16.0)	10 (12.8)
Pulmonary Hypertension	22 (44.0)	35 (44.9)
Stroke	1 (2.0)	3 (3.8)
Concomitant Medications, No. (%)		
CYP2C9 Inhibitor	2 (4.0)	4 (5.1)
CYP2C9 Inducer	0 (0.0)	0 (0.0)
Statin (Simvastatin)	13 (26.0)	15 (19.2)
Antibiotic	4 (8.0)	12 (15.4)
Stable Warfarin Dose, (mg/day)		
Mean	5.0	4.9
Median	4.3	4.3
Range	1.7 – 13.0	1.5 – 12.5

Table 4.2. Effects of clinical and demographic factors on stable warfarin dose by univariate linear regression analysis.

Covariate	p-value	Trend in univariate linear regression
Heritage	$p < 0.049$	-2.4 mg/day for AN
Age	$p < 0.008$	-0.05 mg/day for increase of 1 year in age
Gender	$p < 0.30$	+0.72 mg/day for males
Statin	$p < 0.013$	-1.67 mg/day with concurrent statin use

Heritage, age, gender, and statin use were all included in the multivariate regression model to assess genotype-dose associations.

Table 4.3. Effects of genotype on stable warfarin dose in the INR-based cohort using multivariate regression analysis (adjusting for age, heritage, gender, and concurrent statin use) as well as univariate regression analysis for comparison.

Variant (rs ID)	Trend with Stable Warfarin Dose (Significance)		# Reference	# Heterozygotes	# Homozygote Variants	Allele Frequency in AN/AI at SCF (%)
	Multivariate Analysis	Univariate Analysis				
VKORC1 -1639G>A (rs9923231)	-1.7 mg/day (p = 1.4e-05‡)	-2.1 mg/day (p = 1.3e-06‡)	9	25	16	59.7*
VKORC1 1173C>T (rs9934438)	-1.7 mg/day (p = 1.4e-05‡)	-2.1 mg/day (p = 1.3e-06‡)	9	25	16	59.7*
*CYP2C9*2* (rs1799853)	+0.43 mg/day (p = 0.58)	+0.78 mg/day (p = 0.50)	42	7	0	5.2
*CYP2C9*3* (rs1057910)	-0.14 mg/day (p = 0.91)	+1.3 mg/day (p = 0.072)	44	4	0	3.4
CYP2C9 M1L (rsNA)	-0.16 mg/day (p = 0.78)	+0.16 mg/day (p = 0.66)	49	1	0	1.0
CYP2C9 N218I (rsNA)	-1.1 mg/day (p = 0.077)	-1.4 mg/day (p = 0.005‡)	47	3	0	1.4
GGCX-intron14 (rs11676382)	-0.06 mg/day (p = 0.93)	-0.68 mg/day (p = 0.43)	44	6	0	3.8
CYP4F2 g72220026G (rs2189784)	-0.02 mg/day (p = 0.95)	-0.02 mg/day (p = 0.95)	22	21	7	31.0
*CYP4F2*3* (rs2108622)	-0.44 mg/day (p = 0.31)	-0.65 mg/day (p = 0.16)	23	22	5	31.5
CYP4F2 SpliceCG (rsNA)	-3.5 mg/day (p = 3.0e-08‡)	-3.0 mg/day (p = 3.0e-09‡)	48	2	0	1.4

CYP4F2 G185V (rs3093153)	+1.7 mg/day (p = 0.22)	+1.4 mg/day (p = 0.37)	47	2	1	2.2
CYP4F11 R276C (rs8104361)	+1.1 mg/day (p = 0.22)	+1.0 mg/day (p = 0.23)	42	7	1	9.1

‡ p-value remains significant after adjusting for multiple testing using Bonferroni correction. Variant allele frequency from an AN/AI population at SCF is included for comparison (Fohner 2015). *CYP2C9*8, CYP2C9*11, CYP2C9*14,* CYP2C9 *P279T,* or *GGCXG421A* variant alleles were not present in the INR-based dose cohort. The reported allele frequency is for the variant allele, an alternative allele to the reference allele, which is defined by the global population. *The variant allele frequency in the AN/AI population is the major allele for the *VKORC1 -1639G>A* and *1173C>T*.

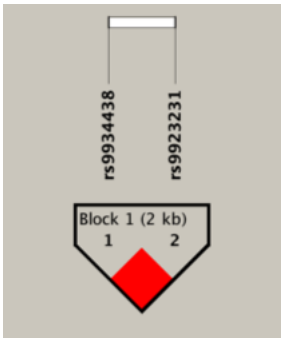

Figure 4.2. Linkage Disequilibrium (LD) in the *VKORC1* locus for rs9923231 (–1639G>A) and rs9934438 (1173C>T) SNVs in all 118 genotyped customer-owners from SCF.

R^2 was 1.0 for this variant pair.

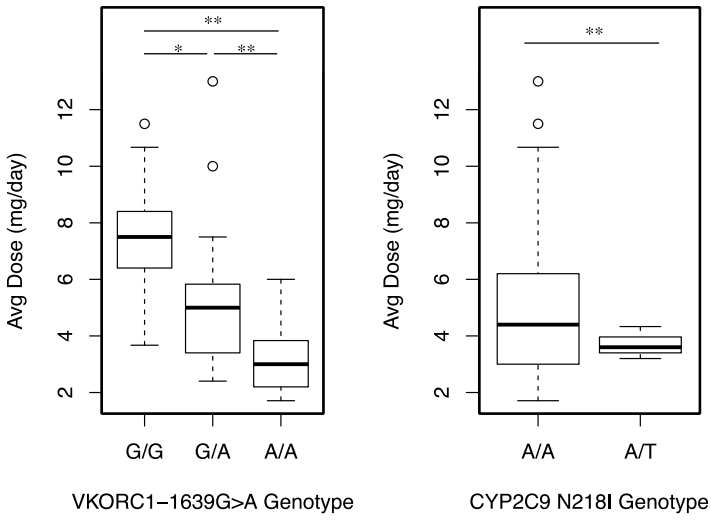

Figure 4.3. Effect of (A) *VKORC1* and (B) *CYP2C9 N218I* genotype on warfarin dose, assessed by univariate regression analysis in the INR-based cohort.

* denotes $p < 0.05$ and ** $p < 0.01$ by t-test of coefficients.

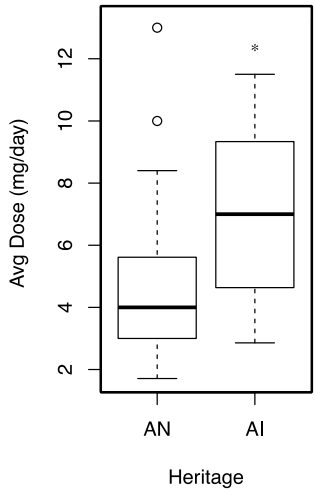

Figure 4.4. Effect of self-reported heritage on warfarin dose, assessed by univariate regression analysis in the INR-based cohort.

The Alaska Native (AN) subgroup includes the following tribes: Inupiaq, Athabascan subgroups, Tlingit, Tsimshian, Haida, Eyak, Aleut/Unangan, Central Yup'ik, Cup'ik, Sugpiaq/Alutiiq. The American Indian (AI) subgroup includes participants affiliated with tribes in the lower 48 states.

* denotes $p < 0.05$ by t-test of coefficients, compared to the AN subgroup.

Table 4.4. Effects of genotype on stable warfarin dose in the consecutive (non-INR-based) cohort using multivariate regression analysis (adjusting for age, heritage, gender, and concurrent statin use) as well as univariate regression analysis for comparison.

Variant (rs ID)	Trend with Stable Warfarin Dose (Significance)		# Reference	# Heterozygotes	# Homozygote Variants	Allele Frequency in AN/AI at SCF (%)
	Multivariate Analysis	Univariate Analysis				
VKORC1 -1639G>A (rs9923231)	-1.8 mg/day (p = 1.6e-08‡)	-1.8 mg/day (p = 8.1e-08‡)	14	43	21	59.7*
VKORC1 1173C>T (rs9934438)	-1.8 mg/day (p = 1.6e-08‡)	-1.8 mg/day (p = 8.1e-08‡)	14	43	21	59.7*
*CYP2C9*2* (rs1799853)	+0.16 mg/day (p = 0.80)	+0.15 mg/day (p = 0.83)	68	9	0	5.2
*CYP2C9*3* (rs1057910)	-0.46 mg/day (p = 0.61)	+0.03 mg/day (p = 0.97)	69	7	0	3.4
*CYP2C9*11* (rs28371685)	-1.0 mg/day (p = 0.29)	+0.07 mg/day (p = 0.79)	77	1	0	0.0
CYP2C9 M1L (rsNA)	+0.37 mg/day (p = 0.47)	+0.50 mg/day (p = 0.06)	77	1	0	1.0
CYP2C9 N218I (rsNA)	-1.1 mg/day (p = 0.014)	-1.3 mg/day (p = 0.001‡)	75	3	0	1.4
GGCX-intron14 (rs11676382)	-1.2 mg/day (p = 0.087)	-1.3 mg/day (p = 0.039)	69	9	0	3.8
GGCX G421A (rsNA)	+0.25 mg/day (p = 0.61)	+0.05 mg/day (p = 0.84)	76	1	0	0.6
CYP4F2 g72220026G (rs2189784)	+0.27 mg/day (p = 0.40)	+0.22 mg/day (p = 0.53)	35	32	11	31.0
*CYP4F2*3* (rs2108622)	-0.25 mg/day (p = 0.48)	-0.35 mg/day (p = 0.34)	35	36	6	31.5

CYP4F2 SpliceCG (rsNA)	-2.3 mg/day (p = 1.2e-04‡)	-2.2 mg/day (p = 3.7e-06‡)	74	3	0	1.4
CYP4F2 G185V (rs3093153)	+0.72 mg/day (p = 0.49)	+0.63 mg/day (p = 0.57)	73	4	1	2.2
CYP4F11 R276C (rs8104361)	+0.57 mg/day (p = 0.35)	+0.58 mg/day (p = 0.34)	63	14	1	9.1

‡ p-value remains significant after adjusting for multiple testing using Bonferroni correction.

Variant allele frequency from an AN/AI population at SCF is included for comparison (Fohner 2015). *CYP2C9*8, CYP2C9*14,* or CYP2C9 *P279T* variant alleles were not present in the consecutive (non-INR-based) dose cohort. The reported allele frequency is for the variant allele, an alternative allele to the reference allele, which is defined by the global population. *The variant allele frequency in the AN/AI population is the major allele for the *VKORC1 -1639G>A* and *1173C>T*.

4.5 REFERENCES

1. Higashi, M.K.; Veenstra, D.L.; Kondo, L.M.; Wittkowsky, A.K.; Srinouanprachanh, S.L.; Farin, F.M.; Rettie, A.E. Association between CYP2C9 genetic variants and anticoagulation-related outcomes during warfarin therapy. *JAMA* **2002**, *287*, 1690-1698.
2. Limdi, N.A.; McGwin, G.; Goldstein, J.A.; Beasley, T.M.; Arnett, D.K.; Adler, B.K.; Baird, M.F.; Acton, R.T. Influence of CYP2C9 and VKORC1 1173C/T genotype on the risk of hemorrhagic complications in African-American and European-American patients on warfarin. *Clin Pharmacol Ther* **2008**, *83*, 312-321, doi:10.1038/sj.clpt.6100290.
3. Rieder, M.J.; Reiner, A.P.; Gage, B.F.; Nickerson, D.A.; Eby, C.S.; McLeod, H.L.; Blough, D.K.; Thummel, K.E.; Veenstra, D.L.; Rettie, A.E. Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose. *N Engl J Med* **2005**, *352*, 2285-2293, doi:10.1056/NEJMoa044503.
4. Caldwell, M.D.; Awad, T.; Johnson, J.A.; Gage, B.F.; Falkowski, M.; Gardina, P.; Hubbard, J.; Turpaz, Y.; Langaee, T.Y.; Eby, C., et al. CYP4F2 genetic variant alters required warfarin dose. *Blood* **2008**, *111*, 4106-4112, doi:10.1182/blood-2007-11-122010.
5. Rieder, M.J.; Reiner, A.P.; Rettie, A.E. Gamma-glutamyl carboxylase (GGCX) tagSNPs have limited utility for predicting warfarin maintenance dose. *J Thromb Haemost* **2007**, *5*, 2227-2234, doi:10.1111/j.1538-7836.2007.02744.x.
6. Henderson, L.M.; Claw, K.G.; Woodahl, E.L.; Robinson, R.F.; Boyer, B.B.; Burke, W.; Thummel, K.E. P450 Pharmacogenetics in Indigenous North American Populations. *J Pers Med* **2018**, *8*, doi:10.3390/jpm8010009.
7. Jaja, C.; Burke, W.; Thummel, K.; Edwards, K.; Veenstra, D.L. Cytochrome p450 enzyme polymorphism frequency in indigenous and native american populations: a systematic review. *Community Genet* **2008**, *11*, 141-149, doi:10.1159/000113876.
8. Fohner, A.E.; Robinson, R.; Yracheta, J.; Dillard, D.A.; Schilling, B.; Khan, B.; Hopkins, S.; Boyer, B.; Black, J.; Wiener, H., et al. Variation in genes controlling warfarin disposition and response in American Indian and Alaska Native people: CYP2C9, VKORC1, CYP4F2, CYP4F11, GGCX. *Pharmacogenet Genomics* **2015**, *25*, 343-353, doi:10.1097/FPC.0000000000000143.
9. Edson, K.Z.; Prasad, B.; Unadkat, J.D.; Suhara, Y.; Okano, T.; Guengerich, F.P.; Rettie, A.E. Cytochrome P450-dependent catabolism of vitamin K: ω-hydroxylation catalyzed by human CYP4F2 and CYP4F11. *Biochemistry* **2013**, *52*, 8276-8285, doi:10.1021/bi401208m.
10. Barrett, J.C.; Fry, B.; Maller, J.; Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **2005**, *21*, 263-265, doi:10.1093/bioinformatics/bth457.
11. Klein, T.E.; Altman, R.B.; Eriksson, N.; Gage, B.F.; Kimmel, S.E.; Lee, M.T.; Limdi, N.A.; Page, D.; Roden, D.M.; Wagner, M.J., et al. Estimation of the warfarin dose with clinical and pharmacogenetic data. *N Engl J Med* **2009**, *360*, 753-764, doi:10.1056/NEJMoa0809329.
12. Abecasis, G.R.; Auton, A.; Brooks, L.D.; DePristo, M.A.; Durbin, R.M.; Handsaker, R.E.; Kang, H.M.; Marth, G.T.; McVean, G.A.; Consortium, G.P. An integrated map of genetic variation from 1,092 human genomes. *Nature* **2012**, *491*, 56-65, doi:10.1038/nature11632.

13. Limdi, N.A.; Wadelius, M.; Cavallari, L.; Eriksson, N.; Crawford, D.C.; Lee, M.T.; Chen, C.H.; Motsinger-Reif, A.; Sagreiya, H.; Liu, N., et al. Warfarin pharmacogenetics: a single VKORC1 polymorphism is predictive of dose across 3 racial groups. *Blood* **2010**, *115*, 3827-3834, doi:10.1182/blood-2009-12-255992.
14. McDonald, M.G.; Rieder, M.J.; Nakano, M.; Hsia, C.K.; Rettie, A.E. CYP4F2 is a vitamin K1 oxidase: An explanation for altered warfarin dose in carriers of the V433M variant. *Mol Pharmacol* **2009**, *75*, 1337-1346, doi:10.1124/mol.109.054833.
15. Shendre, A.; Brown, T.M.; Liu, N.; Hill, C.E.; Beasley, T.M.; Nickerson, D.A.; Limdi, N.A. Race-Specific Influence of CYP4F2 on Dose and Risk of Hemorrhage Among Warfarin Users. *Pharmacotherapy* **2016**, *36*, 263-272, doi:10.1002/phar.1717.
16. Bi, Y.A.; Lin, J.; Mathialagan, S.; Tylaska, L.; Callegari, E.; Rodrigues, A.D.; Varma, M.V.S. Role of Hepatic Organic Anion Transporter 2 in the Pharmacokinetics of R- and S-Warfarin: In Vitro Studies and Mechanistic Evaluation. *Mol Pharm* **2018**, *15*, 1284-1295, doi:10.1021/acs.molpharmaceut.7b01108.
17. Shearer, M.J.; Newman, P. Metabolism and cell biology of vitamin K. *Thromb Haemost* **2008**, *100*, 530-547.
18. Shendre, A.; Parmar, G.M.; Dillon, C.; Beasley, T.M.; Limdi, N.A. Influence of Age on Warfarin Dose, Anticoagulation Control, and Risk of Hemorrhage. *Pharmacotherapy* **2018**, *38*, 588-596, doi:10.1002/phar.2089.
19. Norris, T.; Vines, P.L.; Hoeffel, E.M. The American Indian and Alaska Native Population: 2010. In *2010 Census Briefs*, U.S. Department of Commerce Economics and Statistics Administration U.S. CENSUS BUREAU: 2012.
20. Beyth, R.J.; Milligan, P.E.; Gage, B.F. Risk factors for bleeding in patients taking coumarins. *Curr Hematol Rep* **2002**, *1*, 41-49.

Chapter 5. IDENTIFICATION OF POTENTIAL REGULATORS OF *VKORC1*

5.1 INTRODUCTION

Together, clinical factors and genetic variation in *VKORC1, CYP2C9, CYP4F2,* and *GGCX* account for ~60% of the observed variability in the warfarin dose needed to achieve a therapeutic level of anticoagulation in adult populations [1-7]. *VKORC1* genotype is strongly associated with warfarin dose requirement across different populations [8], explaining approximately 25% of warfarin dose variance in the European population [2], and 34% in the AN/AI population of Southcentral Foundation [9], as discussed in Chapter 4. However, knowledge about the transcriptional and translational regulation of *VKORC1* remains incomplete. The small body of literature on *VKORC1* regulation suggests that the *-1639C>T* variant creates a *c-Myc* transcription factor binding site in the promoter region [10] and that miR-133a [11-13] and miR-137 [13,14] modulate VKOR translation through binding to elements in the 3'UTR region of the gene transcript. This chapter explores associations between *VKORC1* and potential regulators to lay the framework for future studies that test whether a portion of the missing heritability in warfarin dose variance can be partially explained by transcription factors or miRNAs that interact with *VKORC1*.

The human hepatic transcriptome provides a rich source of phenotypic variation, allowing for the identification of genetic variants associated with *VKORC1* expression. RNA sequencing (RNAseq) is a powerful tool to investigate the relationships between genomic variation and gene expression, as well as to generate hypotheses for further testing and evaluation. Accordingly, it was used to characterize the full hepatic transcriptome (mRNA, miRNA and lncRNA) of our human liver samples from two tissue banks: (1) the St. Jude Liver Resource at the St. Jude Children's Research Hospital (Memphis, TN, USA) (n=221), and (2) the University of

Washington Human Liver Bank (Seattle, WA, USA) (n=58). Donor age for the liver samples ranged from 0-87 years (mean 40 years, SD ± 22 years), 42% were female and 58% were male, and ethnicity was 92% Caucasian, 3% African American, <1% Asian, <1% Hispanic, and 5% unknown. Only a limited amount of additional information such as cause of death, medications, and liver pathology was known (<50%). DNA, RNA, and subcellular fractions (cytosol, microsomes, cell plasma membrane and mitochondria) were isolated previously from all liver samples.

A high-quality, deeply sequenced RNAseq dataset was generated from 279 human liver biopsy samples from the UW and St. Jude liver banks. Hepatic mRNA was isolated, purified, and sequenced using University of Washington procedures through the Northwest Genomics Center. The sequencing lab processing pipeline included the following elements: raw base call files, sequence read and base quality, sequences mapped to hg19 with reference transcriptome Ensembl v67, aligned data used for isoform assembly and quantitation, genomic features quantitated, and gene-specific quantitation data for further analysis. Dr. Katrina Claw in the Thummel Lab led quality control assessment of the RNA-Seq data and translation of raw sequencing results into user-friendly datasets. Targeted DNA sequencing and analysis was performed using the PGRN-Seq platform developed for the Pharmacogenomic Research Network [15], with the goal of assessing both common and rare variants in select pharmacogenes, including *VKORC1*. Genome-wide association study (GWAS) data for 150 of these liver samples was also available from St. Jude Children's Research Hospital (Erin Schuetz, PhD and Amarjit Chaudhry, PhD).

To identify trans- and cis-acting factors that influence VKOR synthesis, and therefore warfarin dose requirement, *VKORC1* mRNA fragments per kilobase of transcript per million

mapped reads (FPKM) levels from PGRN-Seq dataset were extracted, then examined for associations through a vitamin K pathway-based association analysis.

5.2 METHODS

5.2.1 *VKORC1 -1639C>T Association with VKORC1 FPKM*

VKORC1 mRNA FPKM levels were extracted from the RNASeq dataset and then correlated with *VKORC1 -1639C>T* genotype using RStudio version 1.2.1335 (RStudio, Inc., Boston, MA). To account for heteroscedasticity, robust standard errors were calculated using the lmtest package in R.

5.2.2 *Pathway-Based Analysis*

The following vitamin K cycle-associated genes were included in the pathway analysis: *VKORC1*, *GGCX*, *CYP4F2*, and *CYP4F11*. Potential transcription factor binding sites were identified in *VKORC1* 5'-flanking region (-2000 bp from transcription start site) using TRANSFAC 7.0 Public Database [16]. Conservative input parameters were assigned, with the minimum matrix conservation set to 80% and similarity of sequence to matrix set to 100%. *C-Myc* was also included as a potential transcription factor, as the *VKORC1 -1639C>T* polymorphism creates an Enhancer Box motif, ACCTG [10] , and *c-Myc* genotype has been associated with stable warfarin dose in a Korean patients with mechanical heart valve replacement [17]. The gene coordinates for *GGCX, CYP4F2, CYP4F11, c-Myc,* and transcription factors with putative binding sites in the 5'-flanking region of *VKORC1* were obtained from the GRCh37/hg19 assembly from the UCSC Genome Browser [18], and the start coordinate was adjusted by 2000 bp to capture more of the 5'-region of the gene. For *VKORC1,* the start coordinate was extended further upstream to capture all of the SNVs used to define haplotypes in Rieder et al. 2005 [2]. Associations between

pathway genes and *VKORC1* FPKM (from the GWAS liver bank dataset) were evaluated with a univariate linear regression model using the GWASTools package in R [19] and Bonferroni adjusted p-values were used to correct for multiple testing.

As the goals of this analysis include generating new hypotheses surrounding *VKORC1* regulation, variants meeting less stringent conditions ($p < 0.01$) were also included as secondary results, to facilitate insights into *VKORC1* regulation with future studies.

5.2.3 *Associations with miRNA*

The MicroRNA Target Prediction Database (miRDB) [20,21] was used to predict potential miRNA targets and subsequent analysis of associations between these miRNAs and *VKORC1* FPKM from the RNASeq dataset was performed using RStudio version 1.2.1335 (RStudio, Inc., Boston, MA). In order to be included in the analysis, the miRNAs of interest were required to have FPKM data for at least 80% of liver samples in the RNASeq dataset.

5.3 RESULTS

5.3.1 *VKORC1 -1639C>T Association with VKORC1 FPKM*

The histograms in **Figure 5.1** show that the distribution of *VKORC1* FPKM in both the RNASeq data set (n=279) and the GWAS dataset (n=142), with the GWAS liver samples included in the RNASeq dataset. **Figure 5.2** depicts the strong association between *VKORC1 -1639C>T* and expression of *VKORC1* mRNA, indicated by FPKM. The difference between reference (wildtype) and heterozygotes (-30 in FPKM, p = 5.7 e-10), reference and homozygote variants (-69 in FPKM, p = 2.2 e-16), and heterozygotes and homozygous variants (-39 in FPKM, p = 2.2 e-16) all significant by Wald t-test of coefficients. The F-statistic for the ANOVA test of the null

hypothesis (e.g. all three *VKORC1 -1639C>T* genotype groups have the same mean FPKM) was 123.3 (p-value < 2.2e-16), indicating a strong correlation between genotype and mRNA levels.

5.3.2 *Pathway-Based Analysis*

The GWAS liver bank dataset contained 142 samples with FPKM values for *VKORC1*. Three transcription factor binding sites were identified using TRANSFAC: *Retinoid X Receptor Beta (RXR-beta)*, *Specificity Protein 1 (Sp1)*, and *Transcription Factor Activating Enhancer Binding Protein 2 Alpha (TFAP2A)*. A total of 1448 variants were evaluated for association with *VKORC1* FPKM: 169 variants in *GGCX*, 226 variants in *TFAP2A*, 66 variants in *RXR-beta*, 75 variants, in *c-Myc*, 329 variants in *Sp1*, 52 variants in *VKORC1*, 170 variants in *CYP4F2,* and 361 variants in *CYP4F11*. A conservative Bonferroni correction for multiple testing of 1448 variants required p < 3.5e-05 to reach statistical significance. A Manhattan plot of the targeted pathway-based analysis results for association with *VKORC1* FPKM is shown in **Figure 5.3**, with the nine variants significantly associated with *VKORC1* expression listed is **Table 5.1**, notably all were in or upstream of *VKORC1*.

Table 5.2 lists variants in *VKORC1*, *GGCX*, *CYP4F2*, *CYP4F11*, *Sp1*, and *TFAP2A* associated with *VKORC1* FPKM under less stringent conditions (p < 0.01) for hypothesis generating research, for follow up with future studies.

5.3.3 *Associations with miRNA*

Of the thirty-six predicted miRNA binding sites in *VKORC1* (**Table 5.3**), a total of six miRNAs met the 80% cutoff with FPKM values in the RNASeq dataset required for inclusion in the analysis: miR-133a1, miR-133a2, miR484, miR1285.1, miR330, and miR499a. miR-133a1 and a2 are different transcripts of the same miRNA, so both were included in the analysis. Only

miR133a1 was significantly associated with *VKORC1* FPKM by Wald t-test of coefficients (p=0.0034), though with a small effect size (-0.0041 in *VKORC1* FPKM) (**Figure 5.4**). P-values are given for all miRNAs tested, no data means that the miRNA was not included in our RNASeq dataset.

5.4 DISCUSSION

This chapter examined associations between potential regulators and *VKORC1* expression to identify trans- and cis-acting factors for future testing and evaluation. In the GWAS liver bank dataset, which contained 142 samples with FPKM values for *VKORC1,* nine variants in *VKORC1* (five intronic and four 5'-flanking) were statistically significant (Bonferroni adjusted p-value < 0.05). The five *VKORC1* SNVs found to be predictive of the low warfarin dose requirement, haplotype A in Rieder et. al 2005 [2], were all significantly associated with *VKORC1* mRNA expression in this analysis: rs7196161 (*-4931T>C*), rs9923231 (*-1639G>A*), rs9934438 (*1173C>T*), rs8050894 (*1542G>C*), rs2359612 (*2255C>T*). These SNVs are in high linkage disequilibrium with one another in European populations, with an $R^2 \geq 0.9$ [2]. When the p-value requirement was increased for leniency due to the hypothesis-generating nature of this analysis, four additional *VKORC1* variants (two 5'-flanking, one intronic, and one 3'-UTR) were found to be associated with *VKORC1* FPKM, as well as five *GGCX* variants (three intronic and two 5'-flanking), one intronic *CYP4F2* variant, and one intronic *CYP4F11* variant (**Table 5.2**). CYP4F2 and CYP4F11 are thought to help maintain vitamin K homeostasis, and therefore, variation in these genes may possibly influence *VKORC1* through a feedback loop.

The proximal *VKORC1* sequence resembles a CpG promoter, with previous work focused on the role of DNA methylation [22], mRNA splicing [22], and influence of miRNAs [11-14]. Using TRANSFAC 7.0 Public Database, three transcription factor binding sites were identified: *RXR-*

beta, *Sp1*, and *TFAP2A. Sp1* and *TFAP2A* were associated with *VKORC1* expression under less stringent conditions ($p < 0.001$) for hypothesis generating research, and future studies should evaluate if these associations are significant with a larger sample size. Sp1, a member of the Cys2His2 zinc finger domain transcription factor class, is thought to be a "housekeeping" transcription factor and binds to GC-rich elements [23]. TFAP2A is a basic helix-span-helix transcription factor which also binds to GC-rich DNA sequences [24]. In the GWAS liver bank dataset, twenty *Sp1* variants (one missense, one 5'-flanking, and eighteen intronic) and five *TFAP2A* variants (four intronic and one missense) were associated with *VKORC1* expression. Notably, one SNV in each of these transcription factors had a p-value < 0.001, rs58778950 in *Sp1* (p-value = 0.00085) and chromosome 6 position 10407670 (no rsID) in TFAP2A (p-value=0.0007).

Thirty-six miRNAs were predicted to have binding sites in the 3'UTR region of *VKORC1,* including miR-133 and miR-137, both of which have been previously associated with *VKORC1* expression [11-14]. However, only miR-133a1 and miR-133a2 were significantly associated with *VKORC1* FPKM in this analysis, while miR-137 did not meet the 80% cutoff required for inclusion in the regression analysis. Further studies with larger sample sizes are needed to establish reproducibility of the miRNA associations in the literature, and SNVs within associated miRNAs should be investigated to determine if polymorphisms in regulatory elements can further explain *VKORC1* expression, and warfarin dose requirement.

While stable dose requirements for warfarin have been associated with two main haplotypes in *VKORC1 [2],* the causal polymorphism remains unknown [22]. Allelic expression imbalance in hepatic tissue predicted that either *-1639 G>A* or *1173 C>T* to be the functional SNV responsible for lower mRNA expression in the *VKORC1* haplotype A group [22]. The *1173 C>T*

SNV did not affect the formation of different *VKORC1* splice variants nor did it affect RNA processing other than splicing [22], making it unlikely that this SNV explains the lower mRNA levels. Chromatin immunoprecipitation with antibodies against acetyl-histone3 and K4-trimethyl-histone3 demonstrated that the *-1639G>A* variant is less likely to associate with transcriptionally active chromatin than the *-1639G* (reference) allele, consistent with the observed decreased mRNA expression, however DNA methylation did not correlate with this variant [22]. While these data support that *-1639G>A* is the causal SNV, the molecular mechanism is still unclear. If *VKORC1* expression is mediated by the *-1639G>A* promoter SNV, due to allele specific chromatin modification, then the introduction of an E-box binding site could bind repressive E-box binding proteins, such as c-Myc [10]. In order for methylation-based silencing to occur, methylating enzymes, mediated by corepressors, are recruited to the promoter of the repressed gene [25], however the level of DNA methylation did not correlate with *-1639G>A* genotype or *VKORC1* mRNA expression [22].

Much remains unknown about the regulation of *VKORC1* expression, with the molecular genetic mechanism likely to help explain a portion of the missing heritability (~40%) in warfarin dose. **Figure 5.5** summarizes the findings of this study, as they pertain to *VKORC1* regulation. This study identified nine variants in *VKORC1* (5'-flanking and intronic) that were significantly associated with *VKORC1* expression, which may affect transcription factor binding, gene transcription, and mRNA stability. Variants in the genes encoding the transcription factors Sp1 and TFAP2A were associated with *VKORC1* FPKM, under less stringent conditions ($p < 0.01$), and future studies should investigate SNVs or environmental factors affecting these transcription factors to evaluate their potential impact *VKORC1* expression. Finally, *miR-133a* was significantly

associated with *VKORC1* expression, therefore SNVs in this miRNA should be studied for their potential to affect transcription processing and stability.

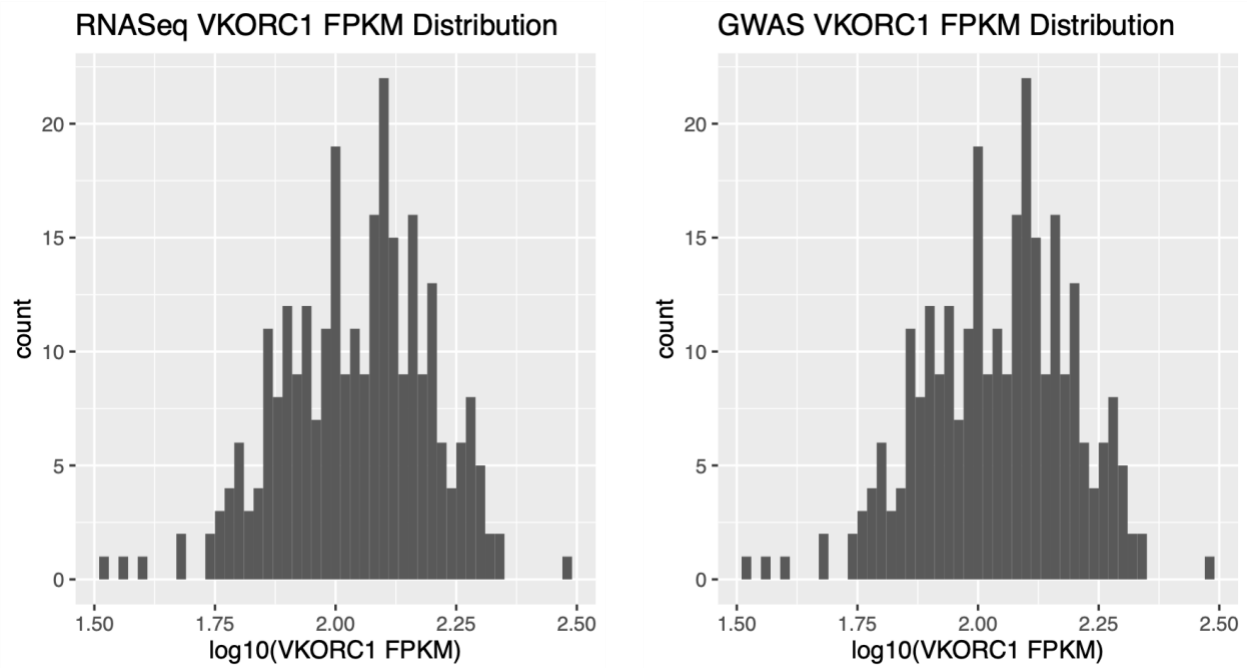

Figure 5.1. Distribution of *VKORC1* FPKM in the RNASeq dataset (n = 279) and GWAS dataset (n = 142).

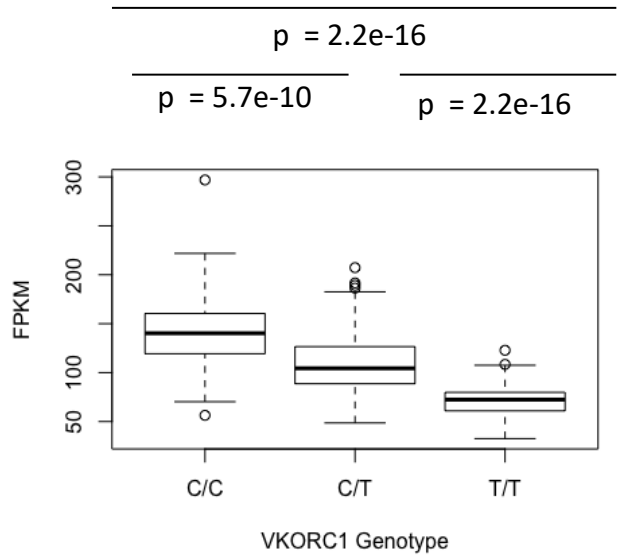

Figure 5.2. *VKORC1 -1639C>T* genotype and expression of *VKORC1* mRNA.

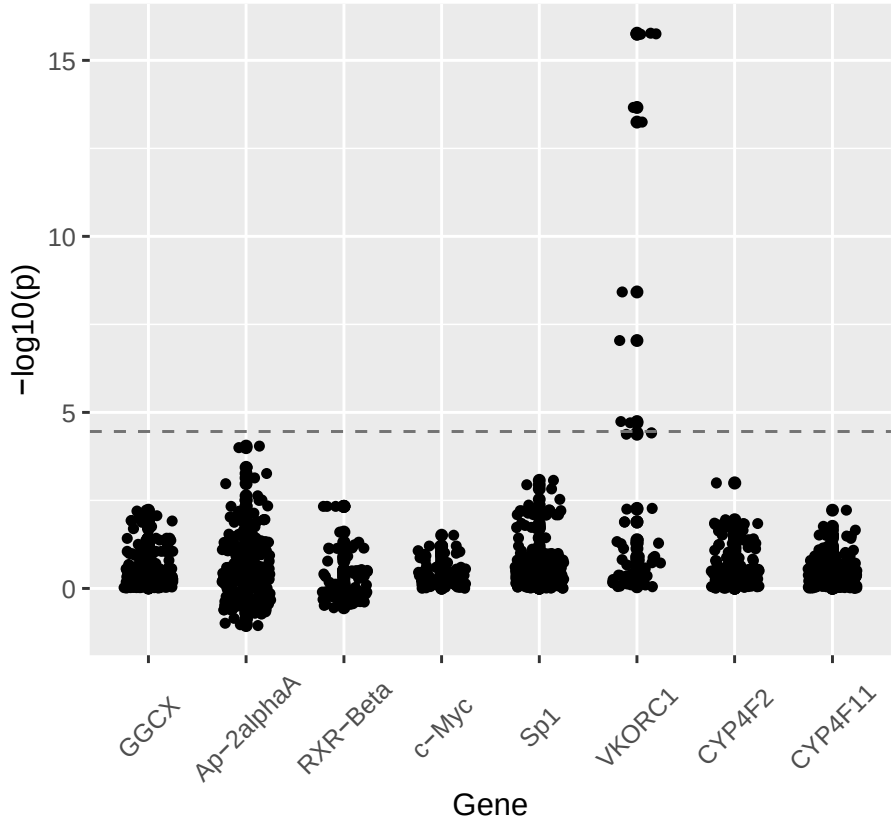

Figure 5.3. Manhattan plot of targeted-pathway analysis results for association with *VKORC1* FPKM.

Table 5.1. Variants Significantly Associated with *VKORC1* FPKM.

Gene	rsID	Bonferroni Adjusted p-value
VKORC1	rs2359612	2.5e-13
VKORC1	rs9923231	2.5e-13
VKORC1	rs9934438	2.6e-13
VKORC1	rs8050894	3.2e-11
VKORC1	rs7196161	8.2e-11
VKORC1	rs2884737	5.5e-06
VKORC1	rs144197331	1.3e-04
VKORC1	rs59924013	2.6e-02
VKORC1	rs59502288	2.9e-02

Table 5.2. Variants Associated with *VKORC1* FPKM (p < 0.001).

Gene	rsID	Location	p-value
VKORC1	rs2359612	intron	1.7e-16
VKORC1	rs9923231	5' flanking	1.8e-16
VKORC1	rs9934438	intron	1.8e-16
VKORC1	rs8050894	intron	2.2e-14
VKORC1	rs7196161	5' flanking	5.7e-14
VKORC1	rs2884737	intron	3.8e-09
VKORC1	rs144197331	intron	9.1e-08
VKORC1	rs59924013	5' flanking	1.8e-05
VKORC1	rs59502288	5' flanking	2.0e-05
VKORC1	rs17880887	5' flanking	3.8e-05
VKORC1	rs7294	3' UTR	4.2e-05
VKORC1	rs17708472	intron	0.0053
VKORC1	rs17881535	5' flanking	0.0056
GGCX	rs75830997	5' flanking	0.0060
GGCX	rs60769490	intron	0.0064
GGCX	rs58768785	5' flanking	0.0086
GGCX	rs75604698 (rs370467631)	intron	0.0086
GGCX	rs77142461	intron	0.0086
CYP4F2	rs148984845	intron	0.0010
CYP4F11	rs11491237	intron	0.0060

Sp1	rs35376163	missense variant (coding)	8.5e-04
Sp1	rs58778950	intron	0.0011
Sp1	rs73311395	5' flanking	0.0015
Sp1	rs784881	intron	0.0029
Sp1	rs113854959	intron	0.0043
Sp1	rs111792083	intron	0.0059
Sp1	rs784880	intron	0.0062
Sp1	rs111698831	intron	0.0062
Sp1	rs2694846	intron	0.0062
Sp1	rs2608299	intron	0.0062
Sp1	rs189070211	intron	0.0062
Sp1	rs2694845	intron	0.0063
Sp1	rs200615705	intron	0.0073
Sp1	rs147778161	intron	0.0074
Sp1	rs2045472	intron	0.0081
Sp1	rs2045471	intron	0.0081
Sp1	rs2460883	intron	0.0081
Sp1	rs784883	intron	0.0081
Sp1	rs2683511	intron	0.0081
Sp1	rs2694854	intron	0.0082
TFAP2A	chr6: 10407670 (position, no rsID)	intron	0.0007
TFAP2A	chr6: 10404492 (position, no rsID)	intron	0.0011
TFAP2A	rs143945720	missense variant (coding)	0.0018
TFAP2A	rs112231225	intron	0.0067
TFAP2A	chr6: 10412242 (position, no rsID)	intron	0.0078

Table 5.3. miRNAs Predicted to Bind to *VKORC1*.

Target Rank	miRNA	Associated with *VKORC1* FPKM?
1	miR-133a-3p	Yes, p<0.01 for 133a-1 transcript
1	miR-133a-3p	Yes, p<0.01 for 133a-2 transcript
2	miR-133b	Did not meet 80% cutoff
3	miR-484	Not significant, p=0.61
4	miR-891a-3p	Did not meet 80% cutoff
5	miR-6772-5p	No data
6	miR-3155a	No data
7	miR-3155b	No data
8	miR-6751-3p	No data
9	miR-1199-5p	No data
10	miR-1276	Did not meet 80% cutoff
11	miR-4292	No data
12	miR-137-3p	Did not meet 80% cutoff
13	miR-6776-5p	No data
14	miR-4284	No data
15	miR-6791-5p	No data
16	miR-4512	No data
17	miR-6807-5p	No data
18	miR-6802-5p	No data
19	miR-1915-5p	Did not meet 80% cutoff
20	miR-5189-5p	No data
21	miR-612	No data
22	miR-1285-3p	1285.1 not significant, p=0.76
22	miR-1285-3p	1285.2 did not meet 80% cutoff
23	miR-6860	No data
24	miR-4313	Did not meet 80% cutoff
25	miR-6865-5p	No data
26	miR-6815-5p	No data
27	miR-4455	No data
28	miR-6835-5p	No data
29	miR-330-3p	Not significant, p=0.69
30	miR-4758-3p	No data
31	miR-499b-5p	Not significant, p=0.36
32	miR-12114	No data
33	miR-4535	No data
34	miR-3688-5p	No data
35	miR-3187-5p	No data
36	miR-6810-5p	No data

miRNAs were required to have FPKM data for at least 80% of liver samples in the RNASeq dataset to be included in the regression analysis and no data means that the miRNA was not included in our RNASeq dataset.

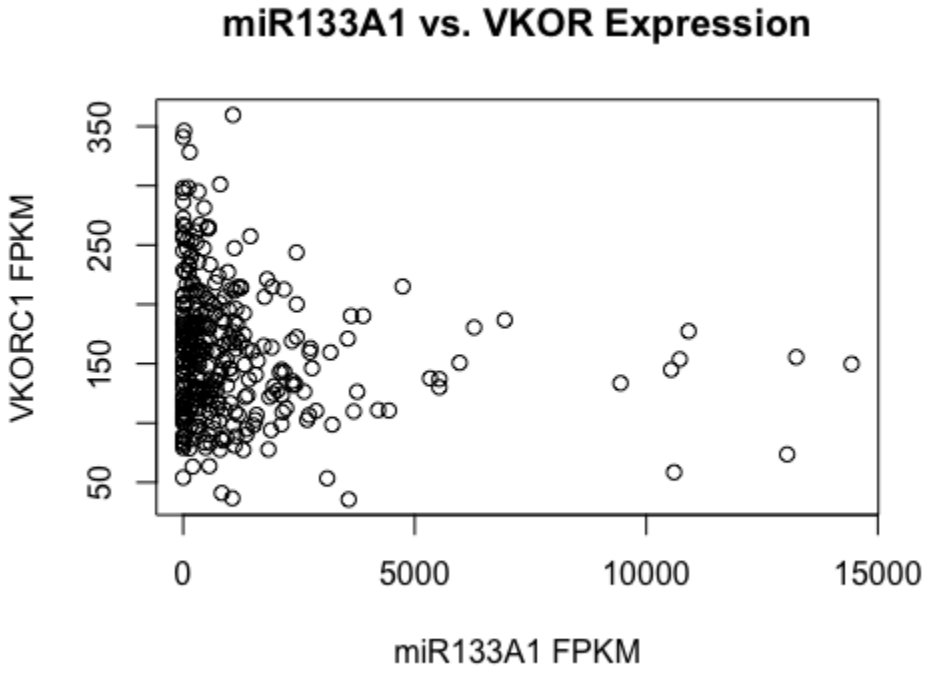

Figure 5.4. miR1133a1 is associated with *VKORC1* FPKM.

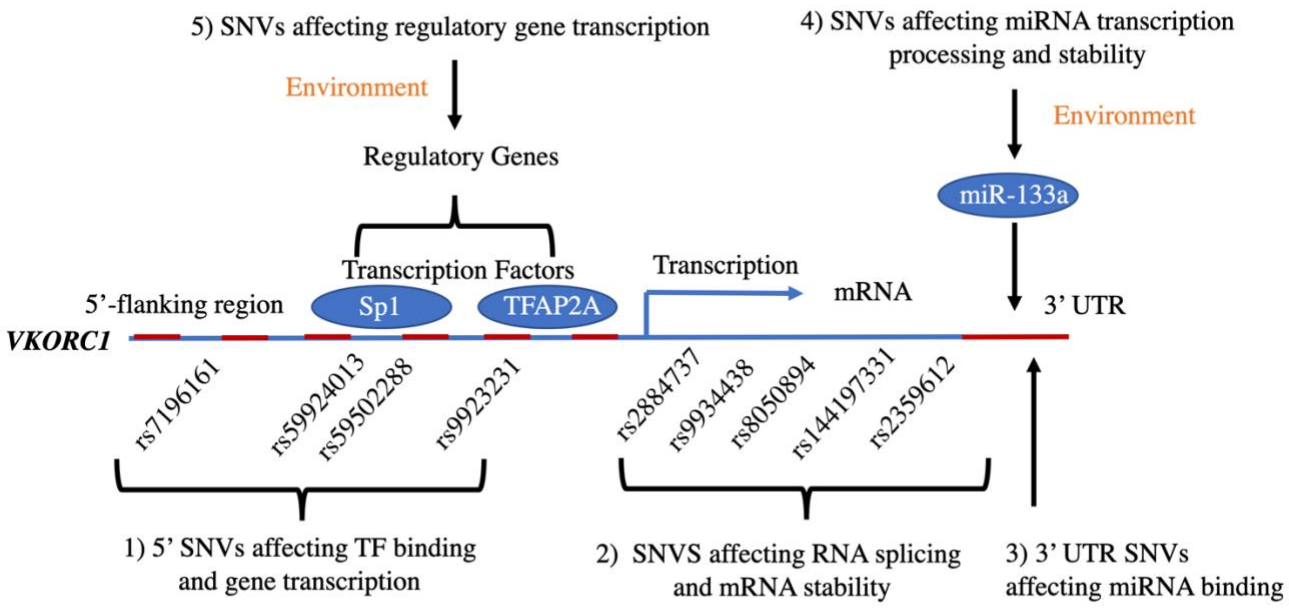

Figure 5.5. Regulatory elements that may affect *VKORC1* transcription and mRNA translation.

5.5 REFERENCES

1. Gage, B.F.; Eby, C.; Milligan, P.E.; Banet, G.A.; Duncan, J.R.; McLeod, H.L. Use of pharmacogenetics and clinical factors to predict the maintenance dose of warfarin. *Thromb Haemost* **2004**, *91*, 87-94, doi:10.1160/TH03-06-0379.
2. Rieder, M.J.; Reiner, A.P.; Gage, B.F.; Nickerson, D.A.; Eby, C.S.; McLeod, H.L.; Blough, D.K.; Thummel, K.E.; Veenstra, D.L.; Rettie, A.E. Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose. *N Engl J Med* **2005**, *352*, 2285-2293, doi:10.1056/NEJMoa044503.
3. Schwarz, U.I.; Stein, C.M. Genetic determinants of dose and clinical outcomes in patients receiving oral anticoagulants. *Clin Pharmacol Ther* **2006**, *80*, 7-12, doi:10.1016/j.clpt.2006.04.008.
4. Caldwell, M.D.; Awad, T.; Johnson, J.A.; Gage, B.F.; Falkowski, M.; Gardina, P.; Hubbard, J.; Turpaz, Y.; Langaee, T.Y.; Eby, C., et al. CYP4F2 genetic variant alters required warfarin dose. *Blood* **2008**, *111*, 4106-4112, doi:10.1182/blood-2007-11-122010.
5. Rieder, M.J.; Reiner, A.P.; Rettie, A.E. Gamma-glutamyl carboxylase (GGCX) tagSNPs have limited utility for predicting warfarin maintenance dose. *J Thromb Haemost* **2007**, *5*, 2227-2234, doi:10.1111/j.1538-7836.2007.02744.x.
6. Au, N.; Rettie, A.E. Pharmacogenomics of 4-hydroxycoumarin anticoagulants. *Drug Metab Rev* **2008**, *40*, 355-375, doi:10.1080/03602530801952187.
7. Johnson, J.A.; Caudle, K.E.; Gong, L.; Whirl-Carrillo, M.; Stein, C.M.; Scott, S.A.; Lee, M.T.; Gage, B.F.; Kimmel, S.E.; Perera, M.A., et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Pharmacogenetics-Guided Warfarin Dosing: 2017 Update. *Clin Pharmacol Ther* **2017**, *102*, 397-404, doi:10.1002/cpt.668.
8. Limdi, N.A.; Wadelius, M.; Cavallari, L.; Eriksson, N.; Crawford, D.C.; Lee, M.T.; Chen, C.H.; Motsinger-Reif, A.; Sagreiya, H.; Liu, N., et al. Warfarin pharmacogenetics: a single VKORC1 polymorphism is predictive of dose across 3 racial groups. *Blood* **2010**, *115*, 3827-3834, doi:10.1182/blood-2009-12-255992.
9. Henderson, L.M.; Robinson, R.F.; Ray, L.; Khan, B.A.; Li, T.; Dillard, D.A.; Schilling, B.D.; Mosley, M.; Janssen, P.L.; Fohner, A.E., et al. VKORC1 and Novel CYP2C9 Variation Predict Warfarin Response in Alaska Native and American Indian People. *Clin Transl Sci* **2019**, *12*, 312-320, doi:10.1111/cts.12611.
10. Pfister, A.; Osman, A. The VKORC1 promoter is occupied by c-Myc transcription factor in HepG2 cells. *Thromb Res* **2010**, *126*, e150-151, doi:10.1016/j.thromres.2010.01.050.
11. Pérez-Andreu, V.; Teruel, R.; Corral, J.; Roldán, V.; García-Barberá, N.; Salloum-Asfar, S.; Gómez-Lechón, M.J.; Bourgeois, S.; Deloukas, P.; Wadelius, M., et al. miR-133a regulates vitamin K 2,3-epoxide reductase complex subunit 1 (VKORC1), a key protein in the vitamin K cycle. *Mol Med* **2013**, *18*, 1466-1472, doi:10.2119/molmed.2012.00062.
12. Ciccacci, C.; Rufini, S.; Politi, C.; Novelli, G.; Forte, V.; Borgiani, P. Could MicroRNA polymorphisms influence warfarin dosing? A pharmacogenetics study on mir133 genes. *Thromb Res* **2015**, *136*, 367-370, doi:10.1016/j.thromres.2015.06.026.
13. Shomron, N. MicroRNAs and pharmacogenomics. *Pharmacogenomics* **2010**, *11*, 629-632, doi:10.2217/pgs.10.26.

14. Tian, Z.; Yang, Y.; Feng, Z.; Wu, D.; Yang, W.; Liu, D. Genetic variant in the promoter region of microRNA-137 reduces the warfarin maintenance dose in patients with atrial fibrillation. *Mol Med Rep* **2019**, *19*, 5361-5367, doi:10.3892/mmr.2019.10205.
15. Gordon, A.S.; Tabor, H.K.; Johnson, A.D.; Snively, B.M.; Assimes, T.L.; Auer, P.L.; Ioannidis, J.P.; Peters, U.; Robinson, J.G.; Sucheston, L.E., et al. Quantifying rare, deleterious variation in 12 human cytochrome P450 drug-metabolism genes in a large-scale exome dataset. *Hum Mol Genet* **2014**, *23*, 1957-1963, doi:10.1093/hmg/ddt588.
16. Matys, V.; Kel-Margoulis, O.V.; Fricke, E.; Liebich, I.; Land, S.; Barre-Dirrie, A.; Reuter, I.; Chekmenev, D.; Krull, M.; Hornischer, K., et al. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* **2006**, *34*, D108-110, doi:10.1093/nar/gkj143.
17. Lee, K.E.; Chang, B.C.; Park, S.; Gwak, H.S. Effects of single nucleotide polymorphisms in c-Myc on stable warfarin doses in patients with cardiac valve replacements. *Pharmacogenomics* **2015**, *16*, 1101-1108, doi:10.2217/pgs.15.37.
18. Kent, W.J.; Sugnet, C.W.; Furey, T.S.; Roskin, K.M.; Pringle, T.H.; Zahler, A.M.; Haussler, D. The human genome browser at UCSC. *Genome Res* **2002**, *12*, 996-1006, doi:10.1101/gr.229102.
19. Gogarten, S.M.; Bhangale, T.; Conomos, M.P.; Laurie, C.A.; McHugh, C.P.; Painter, I.; Zheng, X.; Crosslin, D.R.; Levine, D.; Lumley, T., et al. GWASTools: an R/Bioconductor package for quality control and analysis of genome-wide association studies. *Bioinformatics* **2012**, *28*, 3329-3331, doi:10.1093/bioinformatics/bts610.
20. Liu, W.; Wang, X. Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. *Genome Biol* **2019**, *20*, 18, doi:10.1186/s13059-019-1629-z.
21. Wong, N.; Wang, X. miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res* **2015**, *43*, D146-152, doi:10.1093/nar/gku1104.
22. Wang, D.; Chen, H.; Momary, K.M.; Cavallari, L.H.; Johnson, J.A.; Sadée, W. Regulatory polymorphism in vitamin K epoxide reductase complex subunit 1 (VKORC1) affects gene expression and warfarin dose requirement. *Blood* **2008**, *112*, 1013-1021, doi:10.1182/blood-2008-03-144899.
23. Solomon, S.S.; Majumdar, G.; Martinez-Hernandez, A.; Raghow, R. A critical role of Sp1 transcription factor in regulating gene expression in response to insulin and other hormones. *Life Sci* **2008**, *83*, 305-312, doi:10.1016/j.lfs.2008.06.024.
24. Hilger-Eversheim, K.; Moser, M.; Schorle, H.; Buettner, R. Regulatory roles of AP-2 transcription factors in vertebrate development, apoptosis and cell-cycle control. *Gene* **2000**, *260*, 1-12, doi:10.1016/s0378-1119(00)00454-6.
25. Kouzarides, T. Chromatin modifications and their function. *Cell* **2007**, *128*, 693-705, doi:10.1016/j.cell.2007.02.005.

Chapter 6. CONCLUSIONS AND FUTURE DIRECTIONS

(Part of this chapter was published as "*P450* Pharmacogenetics in Indigenous North American Populations." (2018) Journal of Personalized Medicine 8(1): 9)

6.1 CONCLUSIONS AND FUTURE DIRECTIONS

Until 50 years ago, treatment with pharmaceutical therapies was approached by a "trial and error" iterative process, however the healthcare field now seeks to treat individuals in a more prospective way using personalized, or precision medicine. The goal of precision medicine is to improve patient safety and clinical outcomes by using knowledge of patient-specific biological (including genetic variation) and environmental factors to maximize therapeutic success and minimize the risk of adverse drug reactions or therapeutic failure. The FDA now recommends pharmacogenomic assessment to identify populations in which the appropriate dose of a drug is different from "the norm" due to genetic variation [1]. As the push towards precision medicine grows, there is a heightened need for pharmacogenetic trait discovery and validation in minority groups, particularly minor minorities like AN/AI communities, where population-level differences in the types and frequency of gene variation can exist due to founder effects, selective pressures, and genetic drift [2-4]. A pharmacogenetically guided dosing algorithm developed based on polymorphisms identified in the major world populations may not be as clinically effective for minority sub-populations with different pharmacogene variation profiles. New health care inequalities could arise if populations such as AN/AI are not well represented in pharmacogenetic research.

The allele frequencies of known and recently reported *CYP2C9* variants in the Yup'ik population living in the Yukon-Kuskokwim delta of Alaska illustrate this scenario. As described

in Chapter 1, the *CYP2C9*2* and **3* alleles that define the CYP2C9 PM phenotype in the European population are found at lower frequencies in the Yup'ik population. Moreover, the most common *CYP2C9* variants observed, *M1L* and *N218*, are novel with respect to the rest of the world population. These variants are also found in some of the other Alaska Native tribes. The *in vitro* functional characterization of these novel variants, reported in Chapter 2, found that *N218I* markedly decreased the catalytic efficiency towards four CYP2C9 probe substrates and suggested that *M1L* is a null variant. As shown in Chapter 3, the *in vivo* phenotype of M1L is one of reduced intrinsic formation clearance, as indicated by a 42% decrease in the *(S)*-naproxen O-demethylation formation clearance in genotyped Yup'ik participants, consistent with the *in vitro* results presented in Chapter 2. Considering both *M1L* and *N218I* variants, as well as *CYP2C9*2* and **3* alleles, a substantial portion of the Yup'ik population (~14%) and the more tribally diverse AN/AI population at Southcentral Foundation (~10%) are predicted to be CYP2C9 poor metabolizers. Importantly, if only the allele frequencies known to be clinically relevant in the European population are applied to the AN/AI population, an individual homozygous or heterozygous for the *M1L* or *N218I* variants would be classified as a CYP2C9 EM and improper warfarin dosing could result in adverse events for these individuals. This potential for misclassification and inappropriate drug dosing has also been described of African populations, where the *CYP2C9*8* allele contributes significantly to the PM phenotype [5].

These findings demonstrate the need for studies which specifically evaluate genotype-phenotype relationships in AN/AI populations, which may have different gene-dose associations from European populations due to their unique pharmacogenetic profiles. Chapter 4 describes the first study, to our knowledge, to evaluate associations between warfarin pharmacogene variation in AN/AI people with the warfarin dose needed to achieve a stable, therapeutic level of

anticoagulation. *VKORC1* genotype explained 34% of warfarin dose variability in the AN/AI population, higher than what is observed in European populations (25%). With regard to CYP2C9, three study participants were heterozygous for the *N218I* variant and this was associated with a lower average warfarin dose requirement (1.1 mg/day) in a univariate statistical analysis, although the association just missed statistical significance in an analysis that included other known covariates. A larger study may help clarify its *in vivo* phenotype.

Future research focused on the *N218I* variant could involve a healthy volunteer study using *(S)*-naproxen as a probe for CYP2C9 enzyme activity, a method described and validated in Chapter 3. Furthermore, a pragmatic clinical trial testing the clinical utility of pharmacogenetically-guided warfarin therapy at Southcentral Foundation and the Yukon-Kuskokwim Health Cooperation would inform on whether pharmacogenetic testing provides unique benefit in rural communities with less access to therapeutic monitoring. Based on the data presented in Chapters 2 – 4, a pharmacogenetic algorithm for warfarin dosing in the AN/AI population should consider the impact of the *CYP2C9 M1L* and *N218I* variants, in addition to the lower frequency **2* and **3* variants, as well as the greater impact of *VKORC1* genotype on overall warfarin dose variability in this population.

Associations between *VKORC1* mRNA expression and vitamin K cycle-associated genes as well as transcription factors and miRNAs with predicted binding sites were examined using a high-quality, deeply sequenced RNAseq dataset generated from human liver samples from the UW and St. Jude liver banks, described in Chapter 5. Nine variants in *VKORC1* (5'-flanking and intronic) and *miR-133a* expression were significantly associated with *VKORC1* expression, while variants in *Sp1* and *TFAP2A* were associated under less stringent conditions. Future studies should investigate SNVs or environmental factors affecting *Sp1*, *TFAP2A*, and *miR-133a* expression and

evaluate their potential to modify *VKORC1* expression, in an effort to try to explain part of the missing heritability (40%) of stable warfarin dose requirement.

Further studies are needed to identify and establish the allele frequencies of both known and novel variants for a growing list of pharmacogenes in the AN/AI population, particularly those *P450* genes that encode enzymes that have a clinically significant impact on drug disposition. Moving forward, it will also be important to improve genotyping and sequencing quality as well as increase study sample size, as these will help improve imputation and haplotype estimation in the AN/AI population. This may lead to the discovery of additional functionally important P450 SNVs or structural variants. At the time of this writing, there are no pharmacogenetic studies in AI populations east of the Mississippi, leaving significant uncertainty about P450 genetic diversity for these people and drug phenotype relationships. In 2008, Jaja et al. conducted a systemic review of *P450* variation in Indigenous and Native American Populations that identified ten original studies, of which six of were from Canada, four from North, Central and South America and none in AIAN [6]. An update reported in 2018, partially presented in the introduction of this dissertation, identified twenty-seven original P450 pharmacogenetic studies in Indigenous North American populations, with six in AN/AI, seven in Indigenous people of Canada, and fourteen in Amerindian populations of Mexico [7]. One method for increasing representation of Indigenous people in genetic studies is first to form collaborative research partnerships, in which community partners share control of the research process and apply the values and procedures of community-based participatory research to establish research priorities and acceptable conditions under which the research will occur [8]. The work presented in Chapter 3 is an example of such a partnership, where the selection of *(S)*-naproxen as an *in vivo* CYP2C9 probe substrate was guided by the community's request that the probe drug be a safe, commonly used, over the counter medication.

A minor reduction in probe selectivity (CYP1A2 contribution to the O-demethylation reaction) was accepted to accommodate the request and an increase in sample size provided acceptable power to test the research hypothesis.

The limited data on *P450* genetic variation in populations such as AN/AI translates to missed opportunities for optimizing care. Interestingly, the data that does exist suggests that the AN/AI population has unique genetic variation profiles that may critically impact their response to drug therapy. Without a complete understanding of this population's unique pharmacogene variation profile, AN/AI people may not derive the same benefit from genomics-based precision medicine as those of European descent. The populations included in pharmacogenetic research stand to gain the most from clinical trials findings that establish test validity and utility. A better understanding of the unique *P450* pharmacogenetic variation in the AN/AI population is needed if these communities are to be included in clinical decisions regarding personalized drug therapy and policies surrounding precision medicine.

6.2 REFERENCES

1. Clinical Pharmacogenomics: Premarket Evaluation in Early-Phase Clinical Studies and Recommendations for Labeling. FDA Guidance for Industry. 2013.
2. Arcos-Burgos, M.; Muenke, M. Genetics of population isolates. *Clin Genet* **2002**, *61*, 233-247.
3. Hatzikotoulas, K.; Gilly, A.; Zeggini, E. Using population isolates in genetic association studies. *Brief Funct Genomics* **2014**, *13*, 371-377, doi:10.1093/bfgp/elu022.
4. Peltonen, L.; Palotie, A.; Lange, K. Use of population isolates for mapping complex traits. *Nat Rev Genet* **2000**, *1*, 182-190, doi:10.1038/35042049.
5. Johnson, J.A.; Cavallari, L.H. Warfarin pharmacogenetics. *Trends Cardiovasc Med* **2015**, *25*, 33-41, doi:10.1016/j.tcm.2014.09.001.
6. Jaja, C.; Burke, W.; Thummel, K.; Edwards, K.; Veenstra, D.L. Cytochrome p450 enzyme polymorphism frequency in indigenous and native american populations: a systematic review. *Community Genet* **2008**, *11*, 141-149, doi:10.1159/000113876.
7. Henderson, L.M.; Claw, K.G.; Woodahl, E.L.; Robinson, R.F.; Boyer, B.B.; Burke, W.; Thummel, K.E. P450 Pharmacogenetics in Indigenous North American Populations. *J Pers Med* **2018**, *8*, doi:10.3390/jpm8010009.
8. Boyer, B.B.; Dillard, D.; Woodahl, E.L.; Whitener, R.; Thummel, K.E.; Burke, W. Ethical issues in developing pharmacogenetic research partnerships with American Indigenous communities. *Clin Pharmacol Ther* **2011**, *89*, 343-345, doi:10.1038/clpt.2010.303.

VITA

Lindsay Maritza Henderson was born and raised in California. She received a Bachelor of Science in Pharmacological Chemistry from the University of California, San Diego in 2014.