Precision Approaches for Assessing Complex Pharmacogenomic Traits in Vitamin K Metabolism

Ayoade Nathaniel Alade

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Kenneth Thummel, Chair

Allan Rettie

Nina Isoherranen

Program Authorized to Offer Degree:

Pharmaceutics

University of Washington

Abstract

Precision Approaches for Assessing Complex Pharmacogenomic Traits in Vitamin K Metabolism

Ayoade Nathaniel Alade

Chair of the Supervisory Committee:

Kenneth Thummel, Ph.D.

Pharmaceutics

The term “gene-environment interaction” accounts for various ways in which the effects of environmental exposures can be modified by our genetic variation. Of particular interest for this dissertation is the *CYP4F2*3* coding variant, which has been shown to decrease vitamin K (VK) metabolic clearance *in vitro* and is associated with altered pharmacodynamic effects of coumarin based anticoagulant drugs (e.g., Warfarin) *in vivo*. On average, *CYP4F2*3* carriers receiving coumarin-based anticoagulation therapy require higher doses to achieve a therapeutic INR. It is thought that increased hepatic VK levels caused by reduced VK metabolic clearance in *CYP4F2*3* carriers antagonizes the effect of warfarin, explaining in part the wide range in warfarin dose requirements across the population. However, metabolic contributions from CYP4F11 toward VK metabolism and complex linkage disequilibrium (LD) patterns across the *CYP4F* gene loci make it challenging to quantify the true effect size of individual causal variant alleles. The overall objective of this dissertation was to develop and apply precision approaches

for assessing pharmacogenomic traits in vitamin K metabolism. We hypothesized that the inclusion of diplotype data capturing linked genetic variability across the *CYP4F2* and *CYP4F11* locus can improve phenotype predictions, relative to the *CYP4F2*3* V433M amino acid substitution data alone.

In Chapter 1, we explore the role of VK, its disposition, and interaction with warfarin. In Chapter 2, we propose the use of non-linear mixed effect models (NLME) as an unconventional approach to evaluate the impact that complex genomic traits have on xenobiotics metabolism *in-vitro.* Here, we developed a novel population-based Michaelis-Menten Modeling approach (PopMM) that utilizes strategic sparse sampling allowing for approximation of individual parameter estimates using 50-70% less primary data. In Chapter 3, we evaluated the impact of *CYP4F2/CYP4F11* diplotype on *CYP4F2* and *CYP4F11* mRNA abundance, protein abundance, and metabolic activity towards VK. Lastly, in Chapter 4 we assessed the modifying effect of the singular *CYP4F2*3* variant on the relationship between dietary VK exposure and short-term biomarkers of VK status in healthy human subjects using a LC-MS/MS-based assay. Overall, this dissertation has advanced our understanding of the influence that genetic variability in VK metabolic pathways has on VK status. It applies novel tools to characterize the impact that complex genomic traits have on measures of VK metabolism *in-vitro* and examines the modifying effect that the *CYP4F2*3* variant has on measures of VK exposure following dietary intake *in vivo*.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEGMENTS

First, I would like to thank my Ph.D. advisor Dr. Kenneth Thummel, you have been an amazing mentor, and have guided me through some of the highest peaks and lowest valleys of graduate school. Thank you for always offering support, serving as an exemplary academic role model, and shaping me into the scientist that I am today; I am proud to be one of your graduate students. To my reading committee members, Nina Isoherranen, and Allan Rettie, thank you for your tremendous contributions towards this dissertation, and for helping build these projects into contributions that advance the boundary of our field. To Thummel lab and PK lab members, past and present, thank you for years of moral support, and intellectually satisfying conversations in and out of the lab. To the friends I've made while in graduate school I appreciate all the quality time I got to spend with each of you, and I'm looking forward to what the future holds for us. I also would like to acknowledge the Pharmaceutics, Medicinal Chemistry, and Pharmacy Departments for fostering a welcoming environment and providing a rich network of academic expertise to lean on. Special thanks to the Northwest-Alaska Pharmacogenomics Research Network (NWA-PGRN) for providing an opportunity to collaborate and partner on a variety of meaningful precision medicine projects. Finally, I want to thank my family who have been my primary source of energy while in graduate school, with special recognition for my wife Samone and daughter Avalon. Thank you for believing in me every step of the way.

Chapter 1. INTRODUCTION

1.1 BACKGROUND

1.1.1 *Vitamin K*

Vitamin K (VK) refers to a group of lipid-soluble molecules most notably recognized for their role in regulating blood coagulation. VK exist in two forms, phylloquinone (vitamin K_1, VK_1), found predominantly in leafy greens, and a class of molecules called menaquinones (vitamin K_2, VK_2), found in meats and fermented foods[1] (**Figure 1.1**). While vitamin K_1 has been studied extensively, the physiologic role and dietary recommendations of menaquinones is an emerging area of research[2–5]. In most adults, the recommended dietary intake of vitamin K_1 is between 90-120 μg/day in the United States, and generally 1 μg/(kg • day) for other countries[1]. In tissues, VK is reduced to its active, dihydroquinone form by vitamin K epoxide reductase (VKORC1). Dihydroquinone serves as an obligate cofactor for the enzyme, γ-glutamyl carboxylase (GGCX) that catalyzes the post-translational γ-carboxylation of glutamic acid residues in GLA proteins such as clotting factors II, VII, IX and X [6]. The by-product of this reaction is vitamin K 2,3-epoxide, which is recycled back to the parent form by VKORC1 to be used again in subsequent γ-carboxylation reactions; this recycling process is known as the vitamin K cycle (**Figure 1.2**). Gla-proteins contain an 18 amino acid propeptide that binds directly to GGCX and orients the carboxylation sites (glutamate residues) of the full peptide proximal to the GGCX active site. Once the pro-blood protein is γ-carboxylated it is transported to the Golgi apparatus, where the propeptide is cleaved leaving the blood protein to be secreted into circulation as a zymogen[7,8].

There are seven Gla-containing vitamin K-dependent clotting factors that are involved in hemostasis; factor II (FII), factor VII (FVII), factor IX (FIX), factor X (FX), protein C, protein S,

and protein Z. Of these proteins, FII, FVII, FIX, and FX are involved in the activation of thrombin[9]. Protein C and protein S are anticoagulant proteins that inhibit the coagulation cascade[10]. Protein Z is a protein that helps to stabilize protein C and protein S[11]. The coagulation cascade is initiated when FVII is activated by tissue factor (Factor III), a protein that is exposed when a blood vessel is damaged. FIX is activated by FVIIa and calcium ions. FX is activated by FIXa, FVIIa, and calcium ions. Protein C and protein S inhibit the coagulation cascade by inactivating FVa and FVIIIa. Protein Z also inhibits the coagulation cascade by binding to factor Xa and preventing its activation[12]. A complete schematic of the coagulation cascade, including calcium ion dependent steps can be found outlined in **Figure 1.3**.

In addition to supporting hemostasis, VK also plays an active role in cardiovascular health, bone remineralization, renal function, and cell growth through the γ-carboxylation of other Gla-proteins such as osteocalcin, growth arrest specific 6 protein (Gas6) and Matrix Gla-protein (MGP); which are extrahepatic proteins expressed in regions such as the bone and vascular endothelium[13]. Evidence to date suggests that VK_1 is the principal cofactor for GGCX in the liver, whereas VK_2 support most of the extrahepatic enzyme activity [3,14,15]. Our understanding of the physiologic role of VK has expanded considerably over the past couple of decades, particularly its importance in maintaining cardiovascular health [13,16]. For example, through its activation of MGP, VK is thought to reduce vascular plaque formation and coronary and carotid obstructions that can follow [4].

1.1.2 *Impact of Vitamin K Status on Warfarin Pharmacodynamics*

Warfarin, a VK antagonist, is one of the most widely used oral anticoagulants worldwide with proven efficacy in conditions characterized by increased thromboembolic risk, including atrial fibrillation, deep vein thrombosis, and mechanical heart valve replacement. Warfarin inhibits

VKOR at both the epoxide and quinone reduction steps by inducing a structural change from an open to closed conformation, precluding VK and its epoxide from binding in the VKOR active site. This effectively disrupts VK recycling, leading to the depletion of functional VK reserves and accumulation of vitamin K epoxide. Warfarin is administered as a mixture of the (S)- and (R)-stereoisomers[17] with the (S)-warfarin being 3–5 times more potent inhibitor of the VKOR than (R)-warfarin[18]. Additionally, warfarin is metabolized by multiple cytochrome P450 (CYP) enzymes into a series of hydroxylation metabolites at different positions on the molecule (**Figure 1.4**). These metabolites however do not bind VCOR.

The pharmacological effect of warfarin is assessed by measuring prothrombin activity (Factor II) in the blood using the international normalized ratio test (INR), which is a ratio of *ex vivo* blood clotting activity (prothrombin time) with and without therapy; with the target therapeutic INR falling ideally between 2-3 for most clinical indications. Warfarin is a narrow therapeutic index drug (NTI), meaning small changes in dose or blood concentrations may lead to either increased risk for thromboembolism or excessive bleeding. Patients taking warfarin are subject to continuous therapeutic drug monitoring (TDM) for the duration of their warfarin use. Despite its effectiveness and safety in most patients, long term use of warfarin is associated with morbidity and mortality primarily from major bleeding events that often arise without warning due to its narrow therapeutic window [19].

Initialization of warfarin therapy has historically been approached empirically where a standard 5 mg dose is administered during the first two days of therapy (in otherwise healthy adults) then titrated up or down over the next 5-days depending on the measured INR, with an INR <1.5 indicative of lack of efficacy (warfarin resistance) and an INR >3 indicative of warfarin sensitivity. Additionally, several previously identified patient-related clinical and

pharmacogenetic factors have been found to contribute largely to variability in warfarin dose requirements [20] **(Figure 1.6**) and can inform algorithms for personalizing the initial starting dose[21]. Patients who were concurrently taking CYP2C9-interacting drugs were found to have greater INR variability and lesser time in therapeutic range [22], and nonsteroidal anti-inflammatory drugs (NSAIDs) are known to interact with warfarin and can cause a serious bleeding complication [23]. An important consideration is the impact that genetic variability for the principal VK cycle constituents, VKORC1, CYP4F2 and CYP2C9, have on warfarin dose requirements. Variation in two genes responsible for warfarin pharmacokinetics (*CYP2C9*) and pharmacodynamics (*VKORC1*), most associated with reduced enzyme functions, have repeatedly been found to be significantly associated with warfarin responsiveness, and explain approximately 15 and 25% of the therapeutic dose variability [24,25], respectively. The role of CYP4F enzymes and their impact on warfarin dose will be explained in further detail in the following section.

1.1.3 *The role of CYP4F in Vitamin K Metabolism*

CYP4 family of enzymes (n = 13) catalyze the ω-oxidation of endogenous fatty acids and eicosanoids in humans[26], and are predominantly expressed in the liver and kidney[27–30] (**Table 1.1**). Of these, the CYP4F subfamily consists of 7 members (4F2, 4F3A, 4F3B, 4F8, 4F11, 4F12, and 4F22) whose gene cluster resides on chromosome 19 (**Figure 1.9**). These CYP4F enzymes are responsible for the ω-hydroxylation of medium, long, and very-long chain fatty acids[31]. Both CYP4F2 and CYP4F11 catalyze the metabolic inactivation of VK[30,32] and consequently have a direct effect on systemic exposure, with CYP4F2 putatively playing the dominant role in VK catabolism [32]. CYP4F2 is thought to be regulated by the sterol regulatory element-binding protein (SREBP), which is a DNA-binding protein that regulates transcription of genes involved in cholesterol and fatty acid synthesis[33]. This is supported by studies demonstrating that statin drugs

(i.e., mevastatin and lovastatin), which promote the expression of SREBP, can selectively increase *CYP4F2* mRNA by 3-fold in HepG2 cells and 5-fold in hepatocytes[34]. Alternatively, studies have demonstrated that *CYP4F11* is regulated by nuclear receptor liver X receptor alpha (LXRα), which is a critical signal node connecting inflammation and lipid metabolism. However, the effect of LXRα on expression of the other CYP4F family members and the underlying mechanism remain unclear[35].

Genome wide association studies (GWAS) implicate the *CYP4F2* functional variant, rs2108622 (C>T, V433M *3 allele) as a contributor to variability in warfarin dose requirement, accounting for 1-7% of the variability observed clinically[36]. A previous study conducted at the University of Washington evaluating *CYP4F* genotype-phenotype relationships in a bank of human livers did not find any association between rs2108622 and hepatic *CYP4F2* mRNA abundance, but observed a significant association between the TT genotype at this locus and lower microsomal CYP4F2 protein concentration, paired with reduced vitamin K1 oxidation [30]. A more comprehensive analysis investigating the relationship between genotype and gene expression found that select single nucleotide variants (SNVs) and extended haplotypes derived from variants in the *CYP4F2*, *CYP4F11*, and *CYP4F12* locus were associated with *CYP4F* mRNA abundances. It was speculated that *CYP4F* haplotype and associated diplotype (haplotype pairs) variability might affect warfarin dose requirement [37]. Although intriguing, the impact that having multiple variants in linkage disequilibrium (as a haplotype or diplotype) has on CYP4F protein levels and enzyme activities is not known.

As noted above, the pharmacological effect of warfarin is assessed by measuring prothrombin activity in blood (INR), but this does not provide information about the sensitivity of a patient to changes in the steady-state conditions underlying the INR (e.g., hepatic VK

availability). Plasma concentration of the VK dependent protein PIVKA-II (des-γ-carboxy-prothrombin) is a commonly utilized biomarker of hepatic VK status; with elevated levels (≥ 2 ng/mL) indicative of hepatic VK insufficiency [38] and, more broadly, decreased cofactor support of GGCX-catalyzed reactions (**Figure 1.5**). Due to its relatively long half-life in plasma, PIVKA-II has served as biomarker of long-term hepatic VK status [39,40]. Warfarin suppresses the conversion of PIVKA-II to functional prothrombin, effectively interrupting the clotting factor cascade. Accordingly, increased levels of circulating PIVKA II serve as an indirect biomarker associated with clinical endpoints such as INR [41].

In addition to environmental factors, *CYP4F2*3* has been shown to alter bleeding risk in patients receiving warfarin; specifically, it afforded some protection against major bleeding risk amongst long-term warfarin patients in a Washington state community setting [42,43]. In addition, *CYP4F2*3* homozygous variant carriers are reported to need a higher warfarin dose, approximately +1 mg/day, on average, to achieve a therapeutic INR, compared to individuals homozygous for the reference allele [36]. It has been proposed that reduced metabolic clearance of VK will produce higher hepatic concentrations, at any given level of VK intake, potentially antagonizing the pharmacological effect of warfarin [30]. Diet plays a major predictive role in determining hepatic VK status, and is known to influence warfarin anticoagulation control [44]. Dietary depletion and repletion studies show that long-term changes in VK consumption have a direct impact on markers of hepatic (PIVKA-II) and extrahepatic (undercarboxylated osteocalcin, %ucOC) VK functional status [45]. Populations such as Alaska Native people residing in the Yukon Kuskokwim delta region (Yup'ik people), who have inconsistent access to VK rich foods, have some of the highest frequencies of the VK sparing *CYP4F2*3* variant (>50%) in the world [46]. Moreover, differences in the consumption of VK rich foods (i.e., tundra greens, animal liver, and

fish liver) as well as *CYP4F2*3* in this population were associated with variation in plasma PIVKA-II levels, or long-term VK status [47]. Additionally, in a separate study, the *CYP4F2*3* polymorphism was shown to substantially increase plasma VK_1 exposure in humans in an allele-dependent manner after single-dose supplementation [48]. Although a relationship between long-term dietary VK consumption and warfarin pharmacodynamics is established, there is a gap in our knowledge of the exact modifying effect that *CYP4F2*3* genetic variation has on day-to-day measures of hepatic and extrahepatic VK functional status. Importantly, acute changes in dietary VK intake are suspected of causing significant deviations in INR in patients receiving an otherwise stable, therapeutic warfarin dose. Exactly what the modifying effect of *CYP4F2*3* variation would be in this patient scenario is unclear.

1.1.4 *Assessing Vitamin K Status*

Assessing whole body VK functional status as a single measure is challenging because of the unique tissue distribution properties of phylloquinone (VK_1) and menaquinones (VK_2); plasma VK concentrations alone do not reflect extrahepatic levels of VK, nor hepatic levels under acute VK dosing conditions. Interestingly, VK_1 appears more prominently in the blood whereas menaquinones concentrate in tissues [14]. This is likely due to the fact that the lipophilic character of menaquinones increases substantially as the phytyl chain length increases, which has a significant impact on uptake, tissue storage, plasma transport, and half-life [49]. After dietary consumption, VK is sequestered into cholesterol containing mixed micelles, which are then taken up by enterocytes predominantly in the duodenum [50]. From there, VK is then repackaged into golgi-derived pre-chylomicron vesicles where chylomicron maturation occurs through incorporation of additional triglycerides and lipoproteins [51]. Once maturation is complete, chylomicrons are secreted into the lymph (through diffusion across the lamina propria) where they

remain in transit before entering systemic blood circulation via the thoracic duct. Upon entering the systemic circulation, chylomicron particle sizes are reduced to become a smaller lipoprotein molecule called a chylomicron remnant. Chylomicron remnants are taken up into cells mainly by the low-density lipoprotein receptor (LDLR) expressed by liver hepatocytes and cells in many other tissues (including but not limited to: brain, testes, ovaries, kidneys, skeletal muscle, and intestine) following the classical receptor-mediated pathway of endocytosis [52] (**Figure 1.6**).

Distribution of VK from the liver to extrahepatic tissues follows a similar process, however as opposed to chylomicron packaging, VK is packaged within very low-density lipoprotein (VLDL), low density lipoprotein (LDL), or high-density lipoprotein (HDL). VK_1 is strongly associated with the triglyceride rich lipoprotein (TGRLP) fraction of human plasma in the post-prandial phase of dietary consumption; this fraction encompasses chylomicrons, chylomicron remnants, and VLDL. Since the TGRLP fraction is mainly cleared by the liver, this suggests that initially most of the K-vitamins are transported to the liver [14]. However, in contrast to VK_1 which is rapidly cleared from circulation after it reaches the liver, menaquinones (MK4, MK7, and MK9) are found at higher concentrations in LDL and HDL, which are responsible for transport to and from extrahepatic tissues. This in part helps explain their greater accumulation in organs such as the heart, brain, and pancreas[53].

Once taken into hepatocytes and possibly other cell types, both forms of VK are reduced to active dihydroquinone or undergo the same catabolic reaction – CYP4F2 mediated sequential ω-hydroxylation (primary alcohol to carboxylic acid) followed by successive rounds of β-oxidation resulting in the formation of two chain-shortened carboxylic acid metabolites, K-Acid I and II [30,54] (**Figure 1.8**). A previously established method for quantifying K-Acids in urine relies on electrochemical detection (ECD), which is of limited use in many laboratories and lacks the

selectivity of mass spectrometry. Recently, a novel LC-MS/MS based assay was developed at the University of Washington and provides a minimally invasive yet highly sensitive alternative to ECD for monitoring urinary K-Acid levels in humans [55]. This assay provides a sensitive platform for assessing total systemic exposure to VK, and more specifically to probe interindividual variation in hepatic VK functional status.

Another important determinant of systemic VK exposure and hepatic VK status is the extent of its absorption in the gut from dietary sources. Drug and dietary interactions leading to variability in pathways implicated in transporter-mediated VK_1 absorption have been shown to elicit changes in hepatic VK levels *in vivo* [56]. Studies investigating cholesterol uptake using genome-edited mice, genome-wide association approaches, and gene mutation analyses have revealed several membrane proteins that play crucial roles in intestinal cholesterol absorption (i.e. NPC1L1, SR-B1, and CD36) and are also capable of transporting plant sterols, vitamin E, and VK_1 [50,56–58]. Additionally, the use of the cholesterol lowering drug, ezetimibe, has been shown to enhance the pharmacological effect of warfarin by inhibiting intestinal absorption of VK_1 mediated by Niemann–Pick C1-like protein 1(NPC1L1) in the gut [57]. There is also evidence that VK_1 competes with other fat-soluble compounds such as vitamin E and cholesterol for uptake via the NPC1L1 pathway [59]. Despite our understanding of uptake processes that affect VK_1 absorption, there are no studies evaluating NPC1L1 uptake of VK_2 and whether it also competes for absorption with cholesterol, vitamin E, and VK_1. This is of direct relevance for Yup'ik people in the YK-delta, as menaquinone consumption from fish, a major constituent in their diet, may be substantial if not the primary source of VK.

1.1.5 *Clinical Implication*

Understanding the pathophysiology of VK insufficiency can help us better identify ways to treat and prevent harmful cardiovascular outcomes [13]. VK dependent proteins (VKDPs) include key clotting factors involved in the coagulation cascade (Factors II, VII, IX, X), circulating anti-coagulants (proteins C, S and Z), as well as proteins involved in bone and soft-tissue mineralization like osteocalcin (OC) and Matrix gla-protein (MGP), respectively [19]. It is possible that VK status in specific tissues may be quantifiable by monitoring the abundance of post-translational modification of tissue specific VK dependent proteins, or globally by monitoring levels of urinary VK catabolite excretion. Clinicians and patients need information that can help improve the safety of coumarin based oral anticoagulants during the extended maintenance phase of therapy when variability in environmental exposure, rather than initial dose-ranging, places them at risk of anticoagulation related events. The development and use of superior precision medicine tools than those currently available may help identify patients at greater or reduced risk of bleeding (or clotting) during the maintenance phase of therapy and help guide clinical warfarin dosing decisions.

1.1.6 *Hypothesis and Specific aims*

We hypothesize that genetic variation in the VK catabolic pathway contributes to interindividual variability in VK exposure (systemic, hepatic, and extrahepatic) following dietary intake in a genotype or diplotype dependent manner, potentially leading to variability in dose response for those on coumarin based therapies (i.e., warfarin). The objective of this dissertation research is to clarify the influence that variation in genes controlling VK catabolism have on VK status, with the long-term goal of using this knowledge to advance our understanding of VK related

pathophysiology and drug-drug interactions [13]. To test this hypothesis, our specific aims are threefold:

Specific Aim 1 – Develop a method to efficiently probe the effects of complex genomic traits on drug metabolism *in vitro* at a population level. In this aim, we present a practical application of non-linear mixed effect (NLME) modeling for characterizing *in vitro* kinetic parameters in a population. We then demonstrate the utility of population Michaelis-Menten modeling (PopMM) to facilitate the evaluation of the impact that complex genomic traits have on the metabolism of xenobiotics *in vitro*.

Specific Aim 2 – Reassess the relationship between *CYP4F2* and *CYP4F11* gene **variability and vitamin K_1 metabolism *in vitro* using a haplotype/diplotype-based approach.** In this aim, we sought to apply the PopMM modeling approach (***using a well-characterized human liver bank)*** to determine whether stratification of livers by *CYP4F2/CYP4F11* locus diplotypes improves the identification of poor and extensive metabolizers of VK_1, compared to single allele genotyping alone.

Specific Aim 3 – Investigate the modifying effect of *CYP4F2*3* variation on biomarkers of hepatic vitamin K status *in vivo.* In this aim, the modifying effect of *CYP4F2*3* polymorphism on the relationship between dietary VK exposure and hepatic VK status in healthy human subjects was evaluated. Specifically, a dietary intervention study was conducted to test the effect that VK_1 supplementation has on long-term (plasma Factor II proteoforms) and short-term (urinary K-acid catabolites) biomarkers of hepatic VK status. The effect that supplementation has on systemic VK

exposure (AUC, C_{ss}, C_{max}, etc.), hepatic response (PIVKA-II levels), and extrahepatic response (%ucOC) between *CYP4F2*1/*1* and **3/*3* homozygotes was measured and compared.

1.2 TABLES AND FIGURES

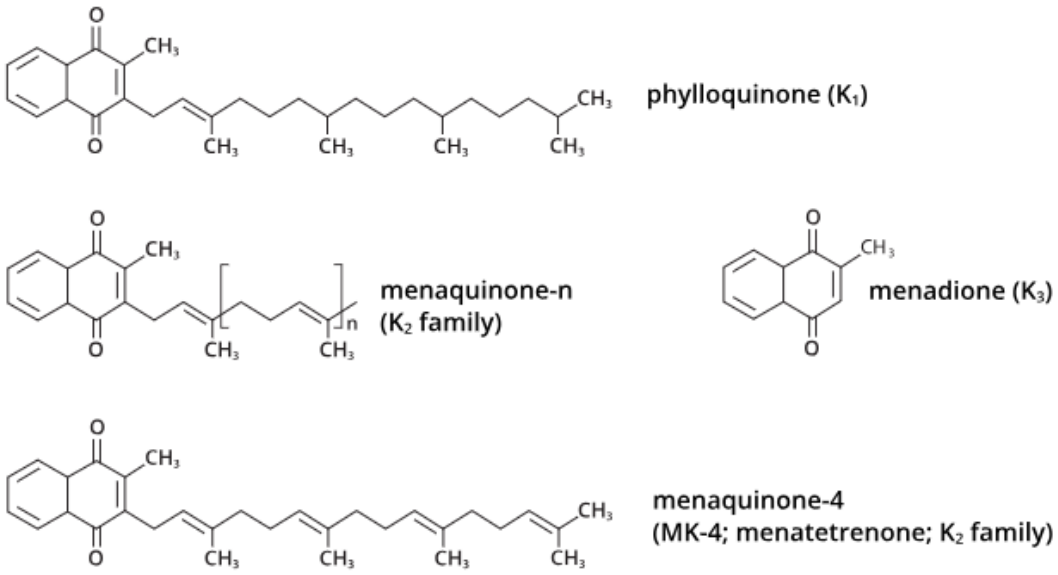

Figure 1.1 Chemical Structures of Vitamin K.

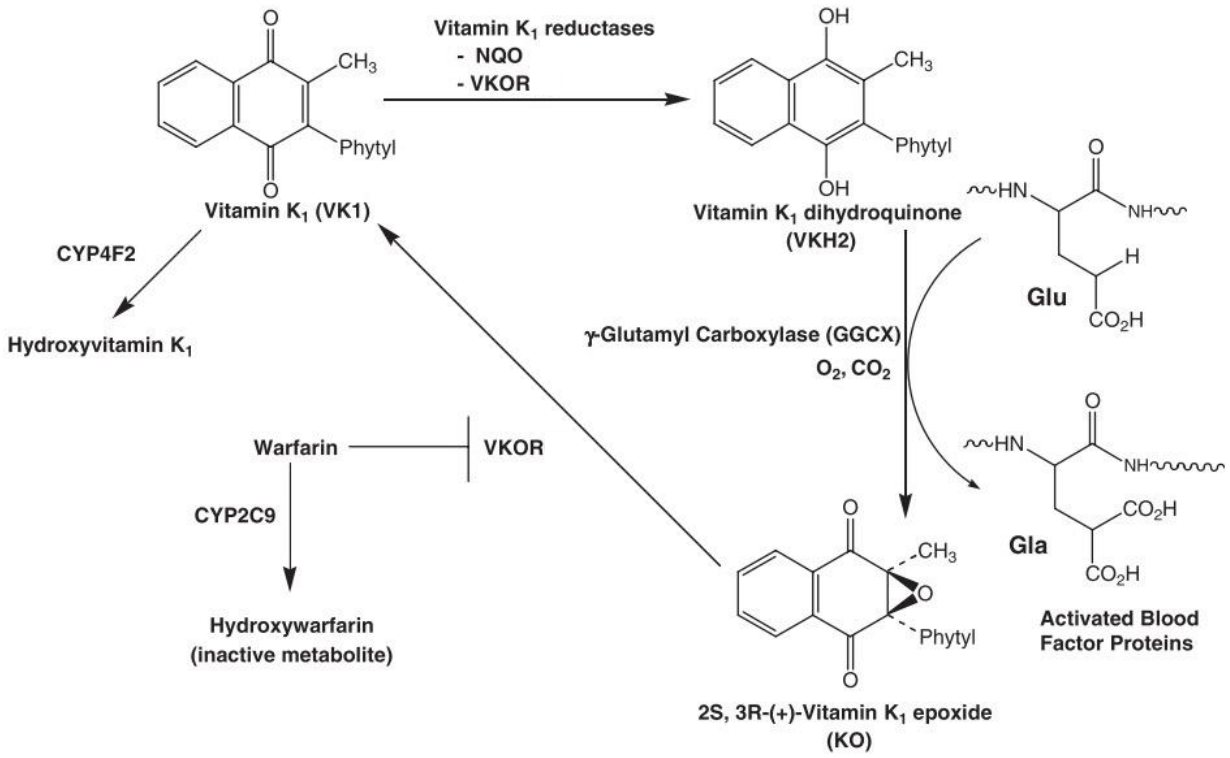

Figure 1.2. The vitamin K cycle, depicted with phylloquinone (VK_1).

Vitamin K is reduced to its active form vitamin K dihydroquinone (VKH2) which is utilized as an essential cofactor in the GGCX mediated γ-carboxylation of vitamin K dependent proteins such as clotting factors (Glu→Gla). The by-product of this reaction is 2,3-vitamin K epoxide (KO) which is recycled back to Vitamin K by the gene product of VKORC1, VKOR, to be used for subsequent γ-carboxylation reaction. Warfarin functions to impede the vitamin K cycle by inhibiting VKORC1 (and the VK reductase(s) and is cleared metabolically by CYP2C9. CYP4F2 functions as the major pathway of elimination for vitamin K by catalyzing the initial ω-hydroxylation step which leads to subsequent catabolism *via* β-oxidation. This figure is sourced from a previously published vitamin K cycle scheme created by Matthew McDonald [30].

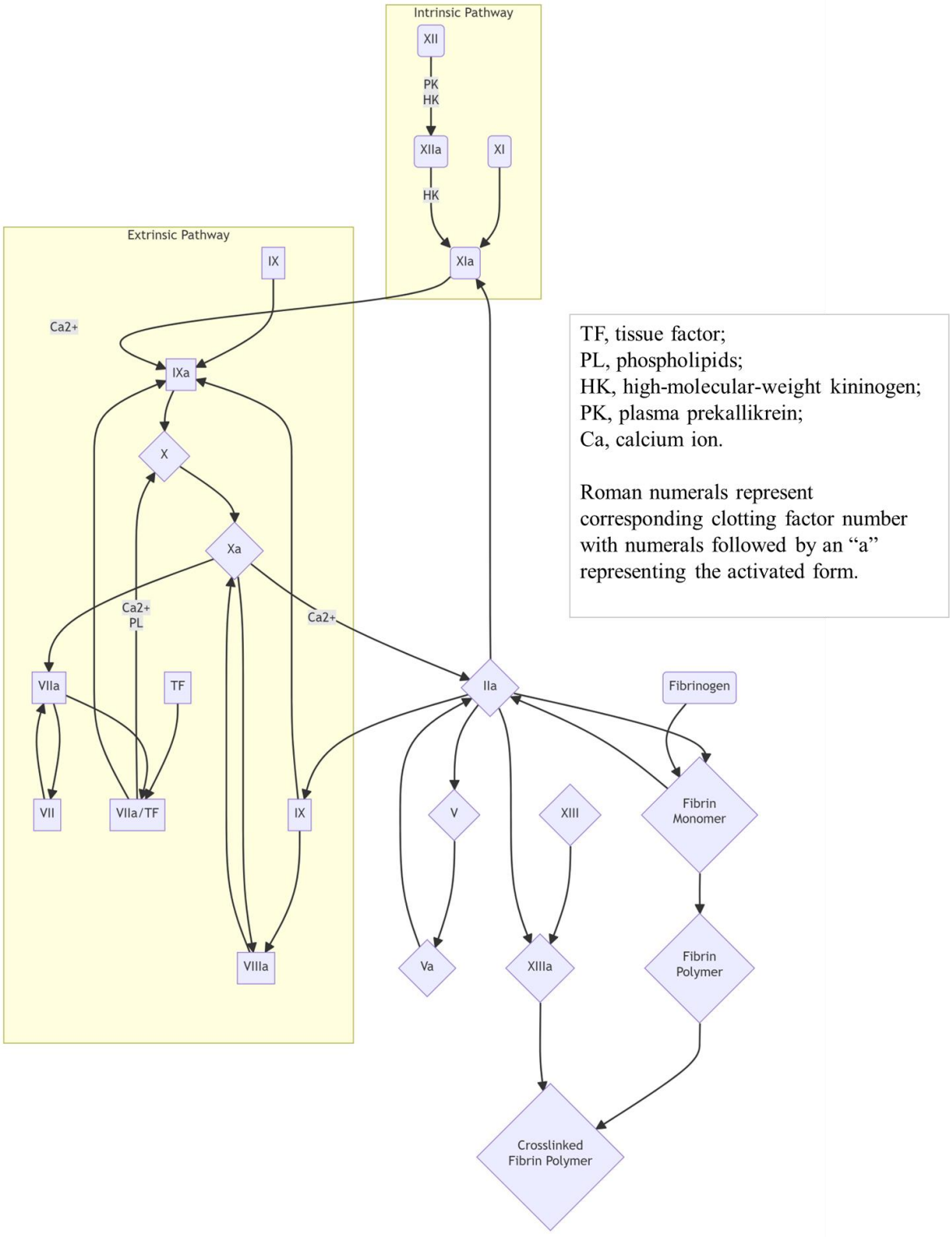

Figure 1.3. Schematic representation of the coagulation cascade.

This figure was adapted from: Ferguson JD, Banning AP., 1998.

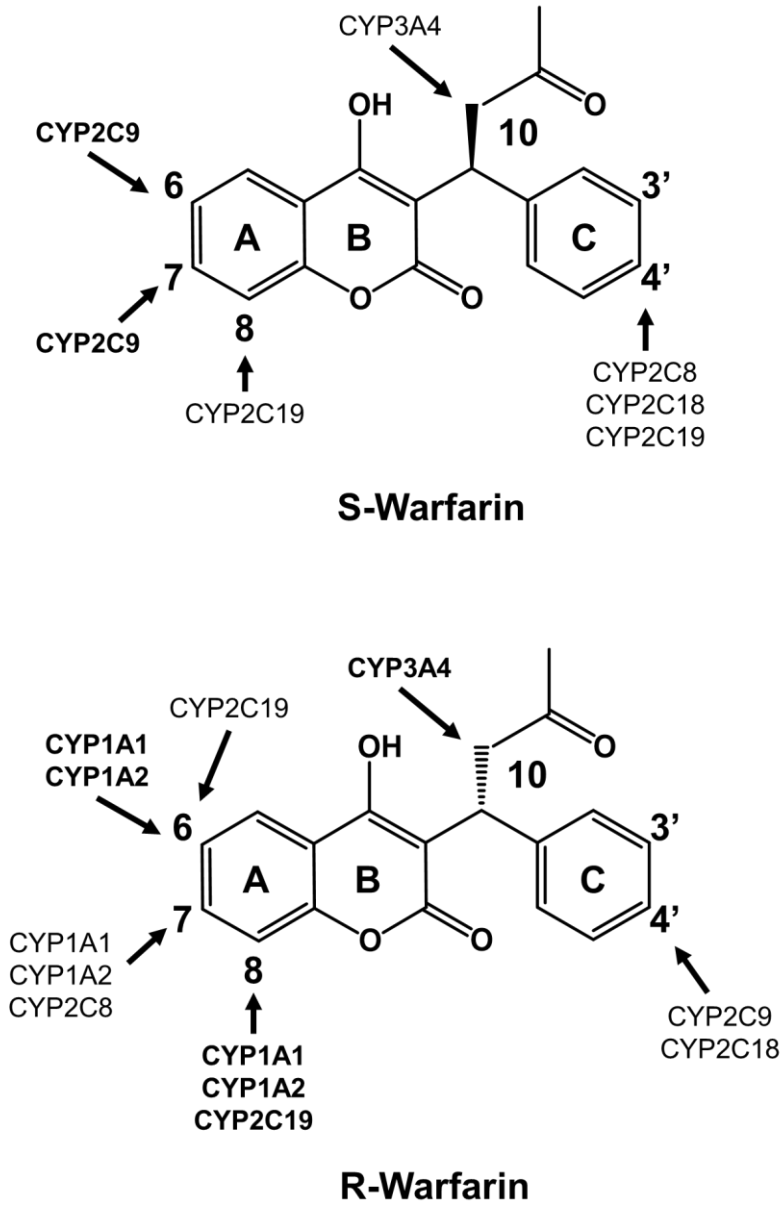

Figure 1.4. Cytochrome P450 catalyzed hydroxylation sites and corresponding CYP isoforms hydroxylating warfarin at the specific sites.

This figure was adapted from from: Jin, S. et al., 2022 [60].

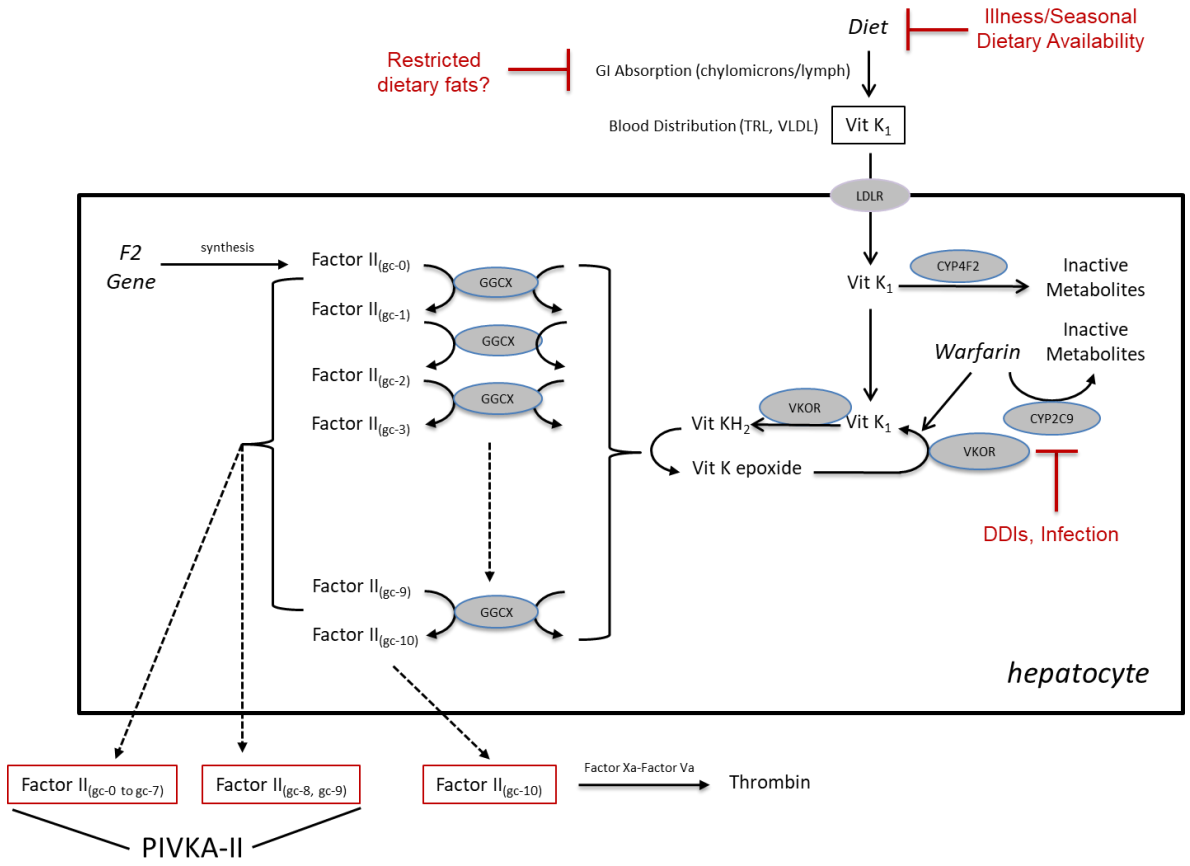

Figure 1.5. Synthesis of γ-carboxylated Factor II (prothrombin) proteoforms and their dependence on dietary vitamin K and GGCX, VKOR, CYP4F2 and CYP2C9 function during warfarin therapy.

This figure is provided courtesy of Kenneth Thummel.

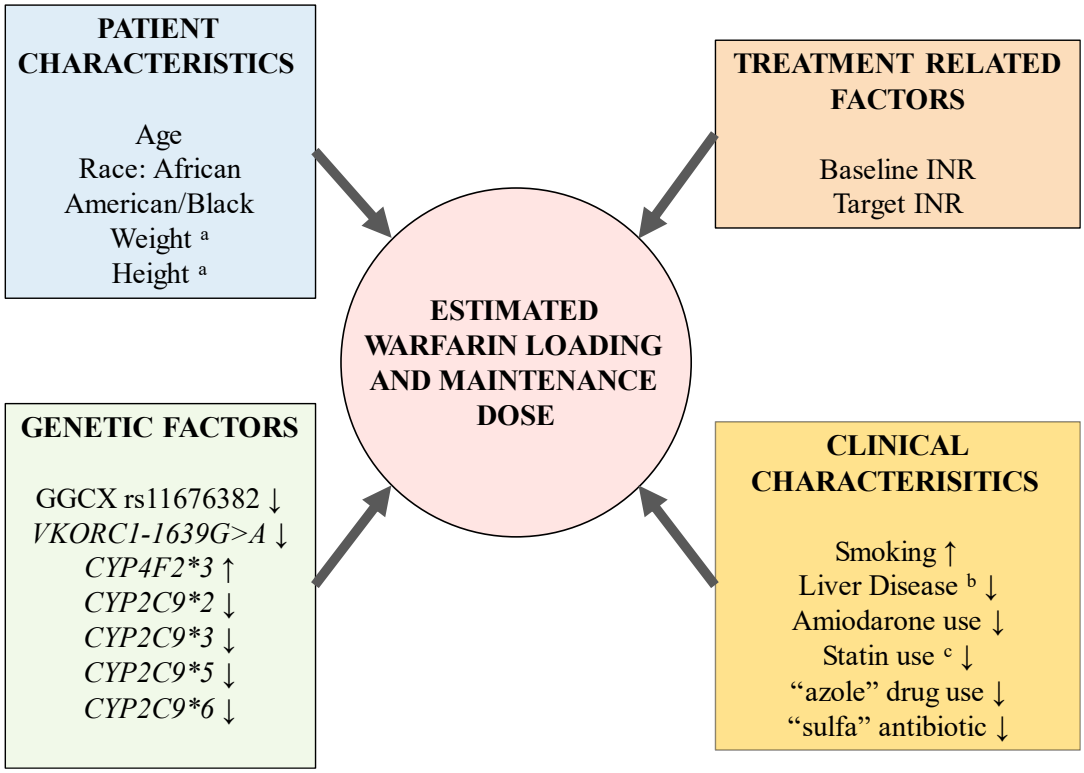

Figure 1.6. Simplified warfarin dosing algorithm incorporating important patient factors, clinical parameters, and genetic information.

A simplified scheme with important patient factors, clinical parameters, and genetic information used to provide an estimation of the loading dose as well as the maintenance dose of warfarin in patients upon initiation. "↓", lower warfarin dose; "↑", higher warfarin dose. [(a)] Higher dose with higher body surface area. [(b)] Severe liver disease may predispose to an elevated international normalized ratio. [(c)] Concomitant statin use (except pravastatin) decreases warfarin maintenance dose. Although this figure describes important factors that are considered when initiating warfarin therapy, it does not include all factors that can potentially lead to variability in warfarin response (i.e., use of herbal supplements, or sudden change in dietary vitamin K intake). This figure is adapted from a previously published warfarin dosing algorithm flow chart [20].

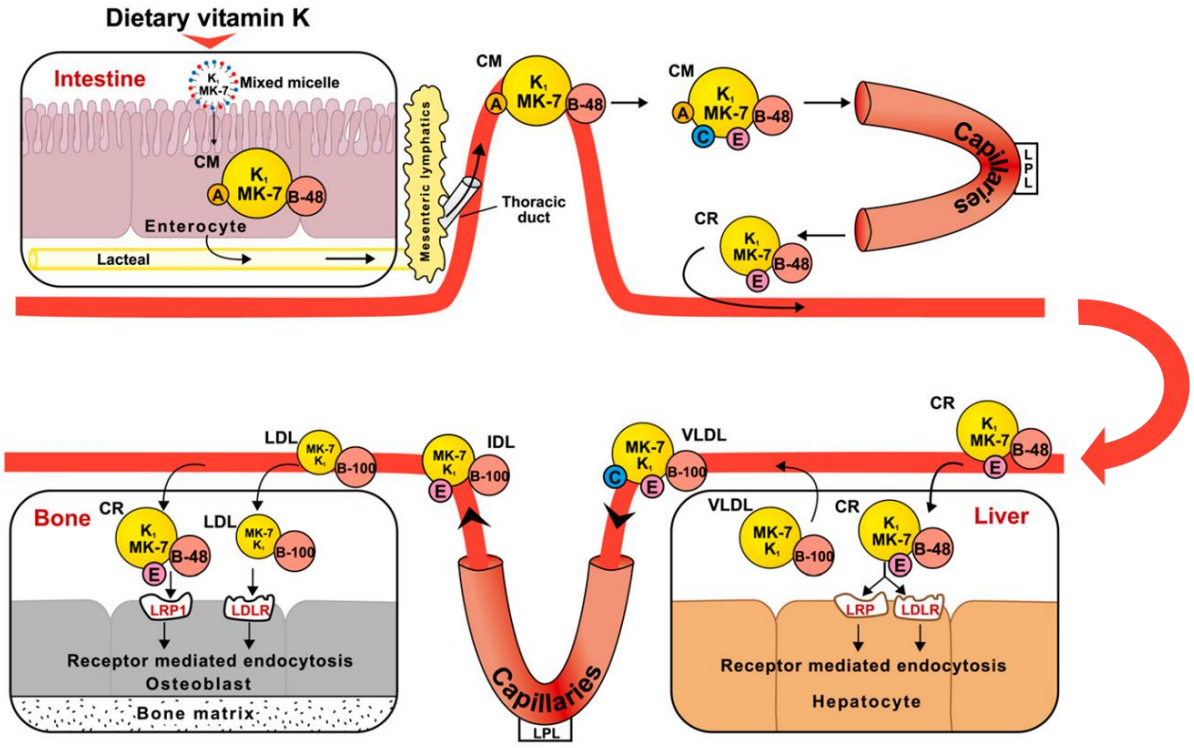

Figure 1.7. Intestinal absorption and distribution of dietary vitamin K_1 and MK-7.

Mixed micelles are taken up by intestinal enterocytes of the small intestine and are incorporated into nascent chylomicrons (CM). Chylomicrons are secreted from within the intestinal villi into the lymphatic capillaries (lacteals), which join larger lymphatic vessels and empty into the blood circulation via the thoracic duct. CM enter the capillary beds of peripheral tissues where they lose much of their TG cargo through the action of LPL, at the same time losing apoA and C. The resulting chylomicron remnants (CR) that reenter the circulation are smaller and have a central lipid core with surface apoB-48 and apoE. CR can interact with cell surface lipoprotein receptors (e.g., LDLR and LRP) and are taken up by target cells by receptor-mediated endocytosis. This figure is adapted from a previously published scheme by Shearer et al., 2008[61].

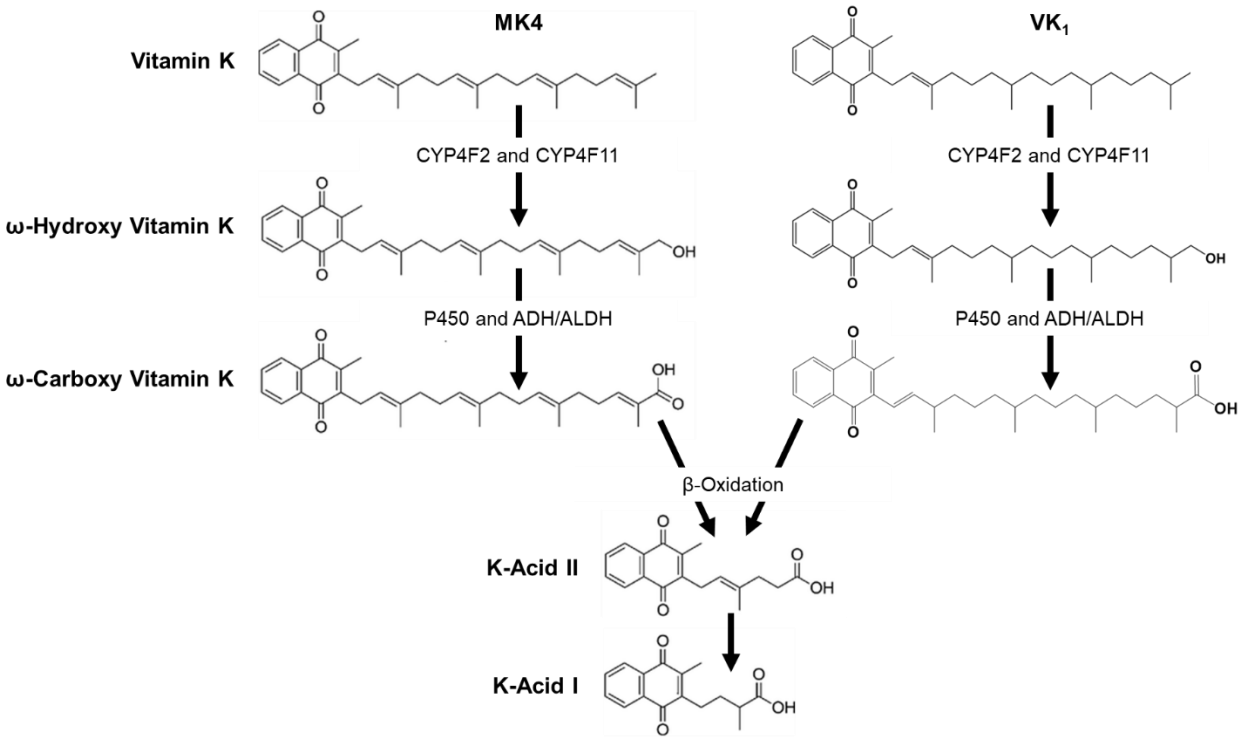

Figure 1.8. Vitamin K metabolic pathway.

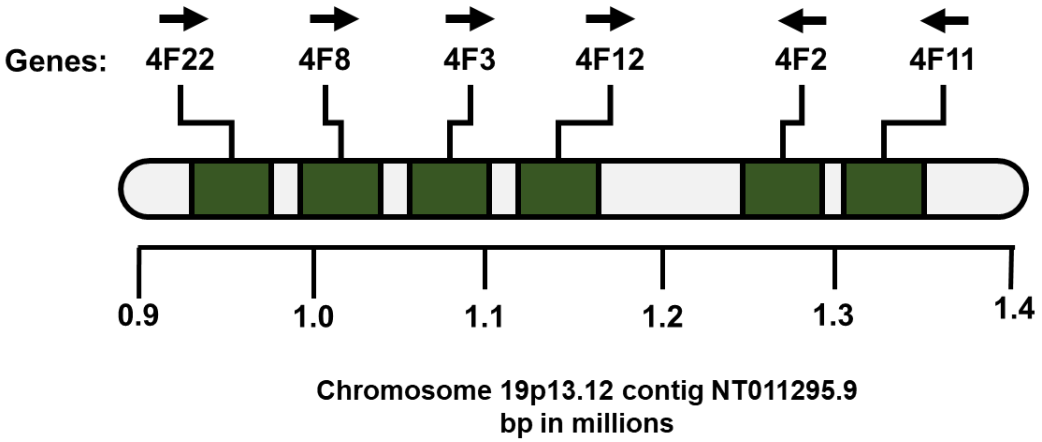

Figure 1.9. The CYP4F gene cluster on chromosome 19.

This figure was adapted from a figure previously published[26] by Edson and Rettie, 2013.

Table 1.1. CYP4 ω-hydroxylase ortholog genes expressed in various human organs[31].

P450 Isoform	Organs	Reference[62]
4A11	Liver, kidney, small intestine, lung, heart, skin, adrenal, prostate, testis, uterus, mammary, placenta.	Hrycay and Bandiera, (2009).
4A22	Liver, kidney.	Hrycay and Bandiera, (2009)
4B1	Small intestine, lung, kidney, heart, skin, spleen, thymus, pancreas, skeletal muscle, eye, adrenal, prostate, urinary bladder, testis, uterus, mammary, placenta.	Hrycay and Bandiera, (2009)
4F2	Liver, small intestine, kidney, brain, skin, prostate, testis.	Hrycay and Bandiera, (2009).
4F3	Liver, small intestine, trachea, kidney, prostate.	Hrycay and Bandiera, (2009).
4F8	Small intestine, lung, stomach, kidney, skin, eye, adrenal, prostate, urinary bladder, testis, uterus.	Hrycay and Bandiera, (2009).
4F11	Liver, colon, heart, brain, skeletal muscle, ovary, placenta, kidney.	Hrycay and Bandiera, (2009).
4F12	Liver, small intestine, stomach, colon, kidney, heart, skin, prostate, ovary, placenta.	Hrycay and Bandiera, (2009).
4F22	Liver small intestine, kidney, brain, skin, skeletal muscle, testis, placenta.	Hrycay and Bandiera, (2009)
4V2	Eye, ovary.	Hrycay and Bandiera, (2009).
4X1	Liver, small intestine, trachea, lung, colon, kidney, heart, brain, skin, spleen, thymus, pancreas, skeletal muscle, prostate, testis, ovary, uterus, mammary, placenta.	Hrycay and Bandiera, (2009).
4Z1	Liver, kidney, skeletal muscle, mammary, ovary.	Hrycay and Bandiera, (2009).

Data sourced from: *Ni K Di, Liu JY. Front Pharmacol. 2021.*

Chapter 2. UTILITY OF NON-LINEAR MIXED EFFECTS MODELS IN THE EVALUATION OF COMPLEX GENOMIC TRAITS IN-VITRO

A version of this chapter has been submitted for publication to

Drug Metabolism and Disposition (2023)

2.1 ABSTRACT

In pharmacogenomic studies, the use of human liver microsomes as a model system to evaluate the impact of complex genomic traits (i.e., linkage-disequilibrium patterns, coding, and non-coding variation, etc.) on the efficiency of drug metabolism is challenging. To accurately predict the true effect size of genomic traits requires large richly sampled datasets representative of the study population. Moreover, the acquisition of this data can be labor-intensive if the study design or bioanalytical methods are not high throughput, and it is potentially unfeasible if the abundance of sample needed for experiments is limited. To overcome these challenges, we developed a novel strategic approach using non-linear mixed effect models (NLME) to determine enzyme kinetic parameters for individual liver specimens using sparse data. This method can facilitate evaluation of the impact that complex genomic traits have on the metabolism of xenobiotics *in vitro* when tissue and other resources are limited. In addition to facilitating the accrual of data, it allows for rigorous testing of covariates as sources of kinetic parameter variability. In this *in silico* study, we present a practical application of such an approach using previously published *in vitro* CYP2D6

data and explore the impact of sparse sampling, and experimental error on known kinetic parameter estimates of CYP2D6 mediated formation of 4-hydroxy-atomoxetine in human liver microsomes.

2.2 INTRODUCTION

Pharmacogenomic studies are a key precision medicine tool for evaluating the impact of genetic variability on drug disposition [63–66]. Major innovations in DNA sequencing technologies now enable exploration and discovery of previously unidentified variants, including new polymorphisms, and complex genomics traits [67]. Despite these advancements, measuring the effect of often rare and/or coinciding genetic variants on drug disposition remains challenging due to either sample size restrictions and/or the inability to distinguish causal variants from others. This is further complicated by interindividual variability (IIV) in drug-metabolizing enzyme (DME) expression associated with other intrinsic factors (i.e., age, sex, disease state, etc.) and extrinsic factors (i.e., diet, concomitant drug use, and environmental exposures) which are known to impact metabolic activity [68]. There are several methods by which human pharmacokinetic (PK) parameters can be predicted from preclinical pharmacokinetic data and/or *in vitro* studies[69]. Of these methods, *in vitro* to *in vivo* extrapolation (IVIVE) using human liver microsomes (HLMs) provides a convenient system for evaluating the impact of genetic variants on drug metabolism and *in vivo* pharmacokinetics.

Variation in genes encoding DMEs can produce altered catalytic activity in several ways [70], including amino-acid substitutions that lead to perturbations in protein structure, stability and/or function, modifications of regulatory elements that control gene transcription, and nucleotide insertions and/or deletions that alter gene translation. The downstream impact that genetic variability has on enzymatic activity can be assessed experimentally by estimating the Michaelis-

Menten parameters K_m, and V_{max} ($k_{cat} \times [E]$, where E = enzyme abundance) *in vitro* and comparing the intrinsic clearance (V_{max}/K_m) values for tissues from donors exhibiting combinations of reference and variant alleles. However, to accurately predict the true effect size of genetic variation on intrinsic clearance (CL_{int}) at a population level using HLMs often requires large richly sampled datasets representative of the study population in question [71,72], and accumulating this data can be labor-intensive and time-consuming if the study design and bioanalytical methods are not high-throughput.

An efficient and commonly employed method to address this involves pooling individual HLMs by genotype and comparing the pool's CL_{int} estimates to a reference group by fitting experimental data using a Michaelis-Menten model [30,73,74]. Although this method provides robust estimates of kinetic parameters, relevant information about the sample population such as between-subject variability (BSV) and its correlation with known intrinsic or extrinsic factors is lost. Alternatively, another commonly used *in vitro* approach approximates individual intrinsic clearances (CL_{int}) by evaluating metabolite formation or parent drug depletion at singular sub-saturating substrate concentrations [75] ($[S] << K_m$); this generally allows for a larger number of individuals to be included in the data analysis, and the retention of covariate information. However, this approach does not provide robust estimates of the underlying kinetic parameters (V_{max}, and K_m) that ultimately determine CL_{int}, and as a result, there is a loss in mechanistic understanding of the impact that genetic variation has on DME function.

Population pharmacokinetic models (PopPK) are powerful tools that utilize a non-linear-mixed effect modeling approach to study both population and individual-level PK and PD relationships *in vivo* [76–79]. Additionally, the impact of known covariates (i.e., intrinsic, and extrinsic factors) on PK parameters (e.g., CL, and V) can be incorporated into PopPK models and quantified

as “fixed-effects,” while potential sources of unexplained variability (e.g., individual subjects) can be quantified as “random effects”. One key advantage of PopPK models is the ability to estimate individual PK parameters based on a complete profile of the entire population instead of its individuals; thus, sparse data can be used at the individual level to generate meaningful estimates of both fixed and random effects. Although top-down approaches are not designed to provide a mechanistic understanding of factors that impact PK parameters, a population-based approach using non-linear mixed-effect modeling *in vitro* can answer these questions while maintaining the previously mentioned advantage of sparse sampling at the individual level; its utility in this regard has been demonstrated extensively in the fields of agronomy and ecology [80,81].

In this Chapter, we further explore the utility of non-linear mixed-effect models and their practical application in the field of pharmacogenomics, for estimating individual level and population-level Michaelis-Menten kinetic parameters (V_{max} and K_m) of known *CYP2D6* variants using a sparse strategic sampling approach. We also explore the utility of mixed effect models for hypothesis testing and characterization of complex genomic traits.

2.3 Materials and Methods

2.3.1 *Virtual Population Design*

A virtual population (n = 9000, 1000 subjects per genotype) consisting of individuals across 9 CYP2D6 profiles: extensive metabolizers (EM: **1/*1, *1/*2, *1/*4, *2/*2, *2/*3, and *2/*4*), ultra-rapid metabolizers (UM: **1/*2x2*), intermediate (IM: **4/*41*) and poor metabolizers (PM: **4/*5*, null variant) was simulated *in silico* using R (v4.2.0). Individual microsomal V_{max} and K_m values were generated randomly and centered by genotype around mean estimates extracted from previously published work [82]. Within-group standard deviation for each parameter was fixed at 25% of the extracted mean estimate to simulate the variability observed within *in-vitro*

population studies. A figure describing how the range of CYP2D6 parameter estimates change with increasing or decreasing CV% can be found in (**Figure 2.1**).

2.3.2 *In-Silico Study Design*

In silico experiments were conducted to simulate CYP2D6 mediated formation of 4-hydroxy-atomoxetine (4-OH-ATX) in human liver microsomes using atomoxetine as a probe substrate. Michaelis-Menten kinetics were simulated across substrate concentrations ranging between 0-2000 µM using V_{max} and K_m parameters obtained from individuals in the virtual population. Two types of sampling designs were employed to define substrate concentrations used to generate the Michaelis-Menten profiles: rich and sparse designs. Rich design experiments were conducted using 9 overlapping atomoxetine (ATX) incubation concentrations per subject (for extensive and ultra-rapid metabolizers) or 16 overlapping atomoxetine (ATX) incubation concentrations per subject (for intermediate and poor metabolizers). To determine an optimal sparse design sampling scheme a pilot study was conducted at the outset to explore the lower limit of a randomized sampling approach. In this pilot study, individuals of a virtual population (n = 21) were randomly assigned to either a 2-point or 3-point sampling scheme with concentrations that were randomly selected across a fixed range (1-100 µM). Results from that study revealed that a 3-point sampling scheme, where substrate concentration landed above, near and below the population K_m values, performed better, in that 2-point sampling schemes gave the least accurate estimates and were subject to a high degree of model shrinkage (data not shown). Based on these results, for the study at hand, the sparse design experiments used 3-4 staggered incubation concentrations (below, near, and above anticipated K_m value) per subject, with both designs covering similar concentration ranges, depending on the genotype. Residual error was added to simulated data using a constant coefficient of variation structure, where the observed metabolic rate (V_{obs}) for the i^{th} individual at

the j^{th} substrate incubation concentration is described by **Equation 1**. The residual errors (ε_i) were normally distributed with a mean of 0 and a standard deviation (σ) set at the following values: 0, 0.05, 0.10, or 0.2.

Equation 1

$$V_{obs\ ij} = V_{ij} + (V_{ij} \times \varepsilon_{ij})$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

2.3.3 *Model Development*

In silico population modeling was performed in R (v4.2.0) using the NLME [83] mixed-effect modeling package and supporting packages for data manipulation and graphics. The final model was built by fitting a structural (base) model and a random-effect model with covariates. The structure of the base model was derived from the Michaelis-Menten equation, with modifications to V_{max} and K_m to incorporate both fixed- and random-effects. Product formation rate for the i^{th} individual at the j^{th} substrate incubation concentration, defined as (V_{ij}), was modeled as a function of a Φ_i and substrate concentration (S_i) plus residual error (ε_{ij}); where Φ_i is a matrix of parameters (Φ_{1i} and Φ_{2i}) each being vectors of fixed and random effects whose sum equates to the predicted individual-level estimates of V_{max} and K_m, respectively (**Equation 2**). Specifically, individual estimates of V_{max} (Φ_{1i}) and K_m (Φ_{2i}) were defined as the sum of a typical population value for V_{max} ($\theta_{1,\ Vmax}$) or K_m ($\theta_{2,\ Km}$), a fixed covariate effect based on individual *CYP2D6* genotype ($X_{1i,\ Genotype\ Vmax}$ or $X_{2i,\ Genotype\ Km}$) and an individual specific random effect on both V_{max} and K_m (η_{1i}, and η_{2i}, respectively) (**Equation 3**). $X_{1i,\ genotype\ Vmax}$ and $X_{2i,\ genotype\ Km}$ were incorporated as additive fixed-effects under the assumption that they do not depend on the magnitude of the typical value (reference group); rather they deviate from the typical value by a fixed magnitude. An additive

variation was used to describe between-subject variability under the assumption that random effects are distributed evenly across a parameter estimate after accounting for the fixed genotype effect. Individual level random effects (η_i) and residual error (ε_{ij}) were assumed to be normally distributed with a mean of 0 and ε_{ij} having a variance of σ^2. V_{max} and K_m were assumed to be uncorrelated (independent), therefore variance (ψ_i) for η_i was structured as a diagonal matrix with covariance set to 0.

Equation 2

$$V_{ij} = f(\Phi_i, S_i) = \left(\frac{\Phi_{1i} \times S_i}{\Phi_{2i} + S_i}\right) + \varepsilon_{ij}$$

Equation 3

$$\Phi_i = \begin{bmatrix}\Phi_{1i}\\ \Phi_{2i}\end{bmatrix} = \begin{bmatrix}\theta_{1,V_{max}}\\ \theta_{2,K_m}\end{bmatrix} + \begin{bmatrix}X_{1i,Genotype\ V_{max}}\\ X_{2i,Genotype\ K_m}\end{bmatrix} + \begin{bmatrix}\eta_{1i}\\ \eta_{2i}\end{bmatrix} = \theta + X_i + \eta_i$$

$$\eta_i = \begin{bmatrix}\eta_{1i}\\ \eta_{2i}\end{bmatrix} \sim N(0, \begin{bmatrix}\psi_1 & 0\\ 0 & \psi_2\end{bmatrix}), \qquad \varepsilon_{ij} \sim N(0, \sigma^2)$$

Intrinsic clearance (CL_{int}) estimates were calculated as the quotient of V_{max}/K_m, with population level intrinsic clearance defined as ($\theta_{1,\ Vmax}$)/ ($\theta_{2,\ Km}$), genotype level defined as ($X_{1i,\ Genotype\ Vmax} + \theta_{1,\ Vmax}$)/ ($X_{2i,\ Genotype\ Km} + \theta_{2,\ Km}$) and individual level intrinsic clearance defined by Φ_{1i} / Φ_{2i} as described in **Equation 4**.

Equation 4

$$Cl_{int\ i} = \frac{\Phi_{1i}}{\Phi_{2i}} = \frac{\theta_{1,V_{max}} + X_{1i,Genotype\ V_{max}} + \eta_{1i}}{\theta_{2,K_m} + X_{2i,Genotype\ K_m} + \eta_{2i}}$$

Selection of the appropriate covariates and error model was based on Akaike criterion (AIC), *F*-tests (ANOVA), diagnostic plots, and the precision of the parameter estimate (95% confidence intervals, relative standard error). The final model was weighted using a power residual error structure.

2.3.4 *Model Evaluation*

Intrinsic clearance (CL_{int}) sensitivity was assessed by comparing standardized residuals (Pearson) values for each CYP2D6 genotype across all experimental designs and conditions (rich and sparse designs, 0-20% CV). Standardized residuals were calculated using **Equation 5**, where the Pearson residual ($R_{p,ijk}$) of the i^{th} genotype and the j^{th} experimental condition for the k^{th} parameter is equal to the difference between the predicted (P_{ijk}) estimate of the i^{th} genotype for the j^{th} condition and expected value for the k^{th} parameter divided by the square-root of the expected value for the k^{th} parameter for the i^{th} genotype.

Equation 5

$$R_{p,ijk} = \frac{P_{ijk} - E_{ik}}{\sqrt{E_{ik}}}$$

Additionally, non-parametric bootstrap analysis based on a total of 120 generated datasets (10 re-sampled populations per experimental condition) was performed to evaluate the precision of the final model parameters. Mean parameter estimates and 95% confidence intervals for each condition were summarized by genotype. Confidence intervals were calculated using **Equation 6,** where confidence interval (CI) is equal to the mean value of the sample population ($\bar{x}$) plus or minus the *t*-score at significance level α (0.05) for *N-1* degrees of freedom, $t_{\left(\frac{\alpha}{2}\right),(N-1)}$ multiplied by the standard error of the sample population mean $S_{\bar{x}}$.

Equation 6

$$CI = \bar{x} \pm t_{\left(\frac{\alpha}{2}\right),(N-1)} \times S_{\bar{x}}$$

2.3.5 *Substrate Depletion Model*

An alternative, and mathematically equivalent, method of estimating intrinsic clearance and Michaelis-Menten parameters is by monitoring depletion of parent drug rather than metabolite formation. Here, we assume (1) reaction rate (v), which is the change in substrate concentration over time ($d[S]_t/dt$), is equal to the change in product formation over time ($-d[P]_t/dt$) at steady-state, and (2) product formation represents the sum of all metabolic formation clearance pathways for the parent in the system. Reaction rate (v) can be modeled as a substrate dependent hyperbolic relationship defined by a theoretical first-order depletion rate constant observed at infinitesimally low substrate concentrations ($k_{dep[S]\rightarrow 0}$) and the Michaelis-Menten constant (K_m); where $k_{dep[S]\rightarrow 0}$ is equal to intrinsic clearance (**Equation 7**). Similar to the Michaelis-Menten equation, a population approach can be applied by re-parameterizing this model to include fixed and random effects (**Equation 8**). The depletion rate for the i^{th} individual at the j^{th} substrate incubation concentration, defined as (v_{ij}), can be modeled as a function of a Φ_i and substrate concentration (S_i) plus residual error (ε_{ij}); where Φ_i is a matrix of parameters (Φ_{1i} and Φ_{2i}) each being vectors of fixed and random effects whose sum equates to the predicted individual-level estimates of V_{max} and K_m, respectively.

Equation 7

$$v = k_{dep}[S] = k_{dep([S]\rightarrow 0)} \times \left(1 - \frac{[S]}{[S] + K_m}\right)$$

$$k_{depl([S]\rightarrow 0)} = \frac{V_{max}}{K_m}$$

Equation 8

$$v_{ij} = f(\Phi_i, S_i) = \frac{\Phi_{1i}}{\Phi_{2i}} \times \left(1 - \frac{[S]}{[S] + \Phi_{2i}}\right) + \varepsilon_{ij}$$

2.4 RESULTS

2.4.1 *Virtual Population Characteristics*

A summary of the model development workflow can be found in **(Figure 2.2).** To generate a virtual population, reported V_{max} and K_m estimates from 21 human liver microsome (HLM) donors [82] were used as a proxy to assign Michaelis-Menten values. Average V_{max} and K_m values across individuals from a total of 9 genotype groups were used as the mean estimates for their respective virtual group. A summary of the extracted parameter estimates can be found in **(Table 2.1).** Within-group variability was simulated (n = 1000 per genotype group) using randomly selected and normally distributed values with a standard deviation of 25% of the mean reported genotype group estimate for both V_{max} and K_m **(Figure 2.3)**. Setting the standard deviation for the underlying distribution of the Michaelis-Menten parameters (V_{max} and K_m) for each genotype to 25% of the original estimate provided a 707-fold range in V_{max} values (3.47 - 2457 pmol/min/mg protein) and a 464-fold range in K_m values (0.26 - 120.53 μM) across the virtual population (n = 9000). Additionally, at 25% CV, within-group variability of Michaelis-Menten parameters exhibited a 6-fold range for each *CYP2D6* genotype. Population geometric means for V_{max} and K_m were 181.29 pmol/min/mg protein, and 2.74 μM, respectively. A comparison of the virtual mean estimates and standard deviation for V_{max} and K_m values *vs.* the actual mean estimate along with

their ranges can be found in (**Table 2.2 and Table 2.3)**. The experimental population was generated by randomly sampling the virtual population with a total of 10 individuals per genotype group (n = 90); a summary of parameter estimates for the sample population (experimental population) can be found in (**Table 2.4**).

2.4.2 *In-Silico Experiments*

To evaluate the utility of estimating Michaelis-Menten parameters using a non-linear mixed effect modeling approach with a minimal dataset, two *in silico* experimental designs (rich design, and sparse design) were evaluated against one another. Due to substantial differences in parameter estimates across genotype groups, in order to characterize the full kinetic profile of 4-OH-ATX formation *in vitro*, ATX incubation concentrations for extensive metabolizer (**1/*1, *1/*2, *1/*4, *2/*2, *2/*3,* and **2/*4*) and ultra-rapid metabolizers (**1/*2x2*) ranged from 0.1-100 μM, while ATX incubation concentrations for intermediate metabolizers (**4/*41*) and poor metabolizers (**4/*5*) ranged from 1-2000 μM (**Figure 2.4)**. Additionally, each experimental design was subjected to 4 levels of experimental error: 0, 5, 10, and 20% CV (**Figure 2.5)**.

2.4.3 *Model Optimization*

A Michaelis-Menten structural model was used to characterize the hyperbolic relationship between the initial reaction rate (V0) of 4-OH-ATX formation and ATX substrate concentration (S). When fixed covariate effects and individual-level random effects (Equation 2 andEquation **3**) were both incorporated into the model there was improvement in model fit when compared to the base model alone for both rich and sparse sampling designs (ANOVA, $p < 0.0001$). Additive, proportional, combined (additive + proportional), and power structures were used to model residual error (data not shown). The power residual error structure was selected because it provided the greatest

improvement across Akaike information criterion (AIC), Bayesian information criterion (BIC), and log-likelihood, in addition to eliminating heteroscedastic patterns in residual plots **(Table 2.5-Table 2.12).** Final UM/EM models and IM/PM models stratified by condition are shown in **Figure 2.6**. Summary tables for all models and experimental conditions can be found in the supplementary data (**Table 2.13-Table 2.16**). Model diagnostic plots and the complete code used to perform the data analysis can be found in the supplemental information.

2.4.4 *Model Evaluation*

Sensitivity analysis was conducted to determine the impact of experimental design on parameter estimates (V_{max} and K_m). Both V_{max} and K_m estimates were within 2-fold of the reference value across all experimental design conditions (**Figure 2.7**, and **Figure 2.8**)**.** Additionally, extracted estimates of intrinsic clearance (V_{max}/K_m) were also evaluated for sensitivity to experimental conditions. Overall, intrinsic clearance estimates for all *CYP2D6* genotype groups across each experimental condition were within 2-fold of reference values and the magnitude of the standardized residual ($R_{p,}$ *P* = Pearson) of estimated intrinsic clearance from reference values for each group fell within an acceptable range ($-2 < R_p < 2$) across all conditions (**Figure 2.9**); with **1/*2* being the only notable exception for sparse 3- and 4-point 20% CV sampling designs (R_p = -2.4 and -2.3, respectively). To assess the sensitivity of the model to individuals within the sample population (n=90 subjects), bootstrap analysis was performed on the original virtual population (n=9000 subjects) which yielded 10 re-sampled populations datasets for each experimental design and condition (n = 120 new datasets). Each dataset was fit to the final non-linear mixed effect covariate model and fixed effects on V_{max} and K_m for each *CYP2D6* genotype were extracted. Bootstrapped parameter estimates and 95% confidence intervals for V_{max} and K_m across all conditions for each *CYP2D6* genotype are plotted in **Figure 2.10,** and **Figure 2.11**; tabulated

summaries can be found in the supplemental data (**Table 2.17**-**Table 2.19**). In summary, bootstrapped Michaelis-Menten estimates, and their 95% confidence intervals fell within 80-125% of reference values for the majority of experimental designs and conditions, specifically 88% and 98% of designs for K_m and V_{max}, respectively. Sparsely sampled designs with high residual error (20% CV) most frequently fell out of this range for K_m (13% of designs) yet infrequently for V_{max} (<1% of designs). All bootstrapped parameter estimates (and confidence intervals) were within 2-fold of their respective reference values.

2.5 DISCUSSION

The concept of conducting population analysis *in vitro* using human liver microsomes as a model system has been demonstrated previously [84] where non-linear mixed effect modeling was used as a tool to identify human cytochrome(s) P450s (CYP) involved in the de-ethylation of phenacetin to acetaminophen. A two-enzyme non-linear mixed effect kinetic model was fit to a sparse data set, generated from incubations containing human liver microsomes (n=19) and phenacetin. Variability in CYP1A2 activity was found to correlate with inter-HLM V_{max} variability, and when CYP1A2 activity was included in the model as a covariate the coefficient of variation improved from 70% to 39%. A noted limitation of that study was a potential data set bias due to the experimental design favoring the identification of the low K_{m1} enzyme (CYP1A2) and lacking the optimum design needed to identify the high K_{m2} enzyme (CYP2C19). In a subsequent study [85] this concept was expanded upon by exploring how a categorical factor with covariate information can be a design variable combined with other design factors. Bogacka et al.[85] demonstrated how incorporating optimum design strategies into experimental planning can lead to substantial savings and concluded that optimized designs can be equally efficient using less experimental material

than is needed in standard practice or, if a similar experimental effort is allowed, can achieve higher efficiency.

To demonstrate the utility of this methodology, we started with an *in vitro* CYP2D6 case study[82] looking at the metabolism of ATX to its primary CYP2D6-mediated metabolite 4-hydroxy ATX using human liver microsomes. ATX is a second-line nonstimulant medication used to control symptoms of attention deficit hyperactivity disorder (ADHD) and is a widely recognized drug whose disposition and response is influenced by pharmacogenomics[86]. Inconsistent therapeutic efficacy has been reported with ATX related to variable CYP2D6-mediated drug clearance[87] and activity score[88]. Our analysis demonstrates that the incorporation of genotype as a categorical factor with covariate information in conjunction with a dynamic strategic sampling approach can allow for accurate measurement of metabolic kinetic parameters using 50-70% less primary data than conventional experimental designs.

Our strategic sampling approach, using sparse data at the individual level, staggered in a manner to create rich data at the covariate level, provided accurate estimates of kinetic parameters associated with *CYP2D6* genotype; with a geometric mean percent error of 2.61% (range 0%-9.33%) for V_{max} and 4.39% (range 0%-55.7%) for K_m across all experimental designs and conditions, when compared to the published estimates. Bootstrap analysis further supported the efficacy of the strategic sampling approach, with mean parameter values and their 95% confidence intervals falling within 80-125% of published estimates for most experimental designs and condition (88% and 98% of conditions for K_m and $V_{max,}$ respectively). Additionally, there were minimal differences in parameter estimates and sensitivity to condition between sparse 3-point and 4-point sampling strategies, even with experimental errors as high as 10% CV.

In a scenario where metabolic products have not been (or cannot be) characterized, our approach can still be utilized by adjusting the experimental design to monitor the elimination of the parent drug by switching from a Michaelis-Menten structural model to a substrate depletion model (**Equation 7**); which has been previously identified as a mathematically equivalent means for monitoring intrinsic clearance of the parent compound *in vitro* particularly when one metabolic pathways dominates clearance [89]. Similarly, a non-linear mixed effect model can be applied to estimate kinetic parameters (V_{max}, and K_m) across a substrate concentration range to approximate intrinsic clearance ($k_{depl([S]\rightarrow 0)}$) at the individual level (**Equation 8**). Although modeling *in-vitro* data using a substrate depletion approach is not the focus of this study, the principles remain the same, hence, provides a suitable example for the broad applicability of our method.

While our strategic sampling approach does not leverage optimal design estimators [90], which statistically determine the best experimental design by using an iterative search algorithm that seeks to minimize the covariance of the parameter estimates for a specified model using optimality criterion (i.e., D, and E-optimality), it provides a practical example of how such an approach can be applied empirically, solely based off a sampling range that would be used in a conventionally designed *in vitro* study. It is important to note that *a priori* knowledge of a genotype-effect (*in vitro* or *in vivo*) is not required when deciding which genotype to include as a covariate in this model, when the goal of a study is to test whether there is a significant difference in activity between a particular variant compared to the reference allele. This is particularly useful for exploratory analysis or hypothesis testing purposes. Moreover, the incorporation of random effects not only allows us to estimate between-subject variability, but also quantify and attribute sources of variability (i.e., inter-day, inter-instrument, and between-analyst variability); many of which are factors unaccounted for when conducting *in vitro* studies [91]. This additional benefit also

provides an opportunity to probe complex characteristics of experimental data, such as an interaction between fixed-effects, and hierarchical random-effect structures [83] (crossed, or nested-random effect designs); both of which could be informative tools for identifying causal variants/conditions found in complex genomic traits [92–94] (i.e., haplotype synergy, and gene-environment interactions).

A notable limitation of the sparse sampling approach is its sensitivity to high experimental error (≥20% CV), particularly in the estimation of K_m, where the initial sample population estimates of K_m were over-predicted, leading to underestimation of intrinsic clearance. The overprediction of K_m was not readily apparent in the bootstrap analysis, suggesting that its occurrence is a phenomenon specific to that sample population. The bootstrap analysis confirms that a 20% CV leads to a decrease in the precision of the K_m estimate; an observation that is more apparent for the 3-pointt and 4-point sampling conditions. This is presumably because variability in the K_m region of the Michaelis-Menten curve can lead to significant changes in the predicted reaction rate, and consequently, the estimated K_m value. Inclusion of more points across the full Michaelis-Menten profile may help anchor and/or center K_m estimates under high variability conditions, as evidenced by the greater precision observed for the rich design models. Additionally, it is assumed that the sampling range needed to capture the full kinetic profile for each genotype is generally known; that may not always be the case. Scenarios where the K_m value for a particular genotype varies drastically from the typical population value warrant adjustments in the sampling range to maximize profile coverage. Finally, the frequencies of an existing gene variant may differ from sub-population to sub-population due to exogenous factors (i.e., genetic drift, population bottlenecks, and founders' effects), therefore, estimates of population level fixed

effects will be heavily dependent on the sub-population they are estimated from, which may not always reflect the total population.

In summary, this study highlights the utility of non-linear mixed effect models as a population based *in vitro* approach for characterizing the impact of pharmacogenetic variation on the underlying kinetic parameters that influence xenobiotic metabolism. Introducing *CYP2D6* genotype as a categorical covariate in our NLME model allowed for an estimation of the impact that each genetic variant had on Michaelis-Menten parameters (V_{max} and K_m) as fixed-effects and provided a means for assessing between-subject variability. Additionally, we provide a method for empirically designing a sampling strategy that can generate robust estimates of covariate effects using sparse data at the individual level. Altogether, we envision significant utility for this approach, not only as an exploratory tool for identifying covariates of interest with *in vitro* studies, but also as a quantitative tool for providing a mechanistic understanding of how genetic variation impacts xenobiotic metabolism at a molecular level.

2.6 TABLES AND FIGURES

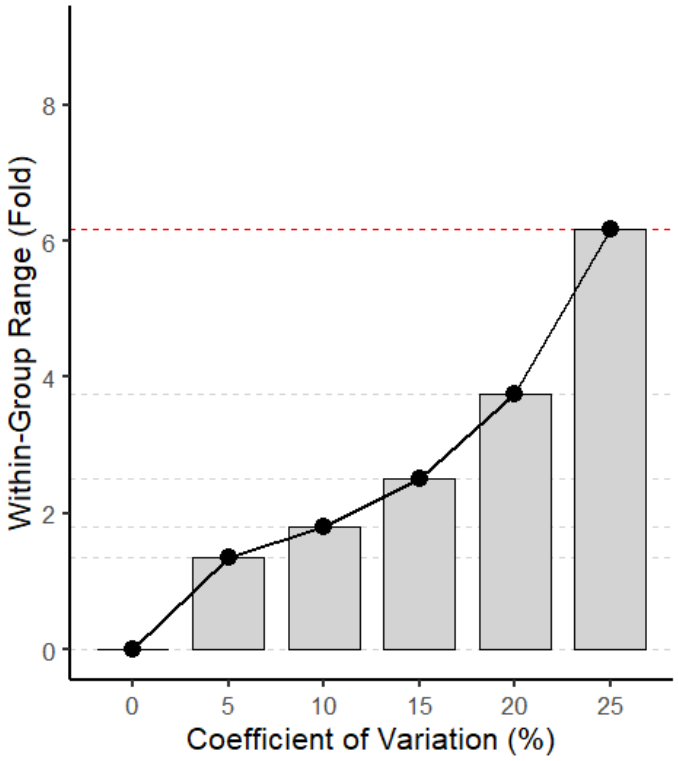

Figure 2.1. The impact of coefficient of variation on within-group variability.

Within-group range represents the fold-change between the smallest and largest Michaelis-Menten parameter value (V_{max} and K_m) for each *CYP2D6* genotype group. Red line represents within-group range of the virtual population (6.18-fold).

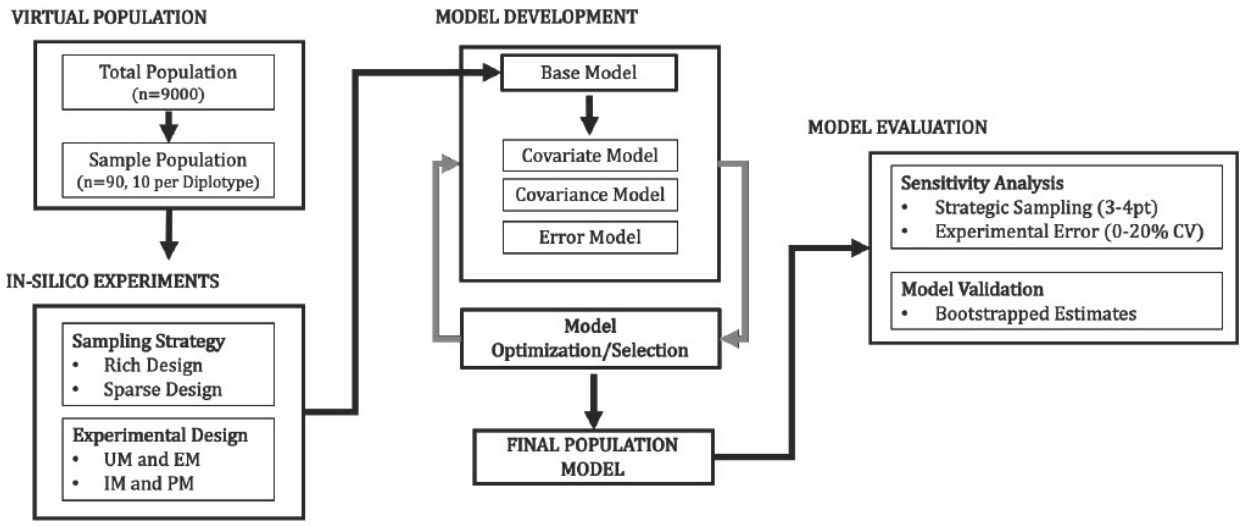

Figure 2.2 NLME Model development workflow.

UM, EM, IM, and PM in the experimental design represent ultra-rapid metabolizer (UM), extensive metabolizer (EM), intermediate metabolizer (IM), and poor metabolizer (PM), respectively.

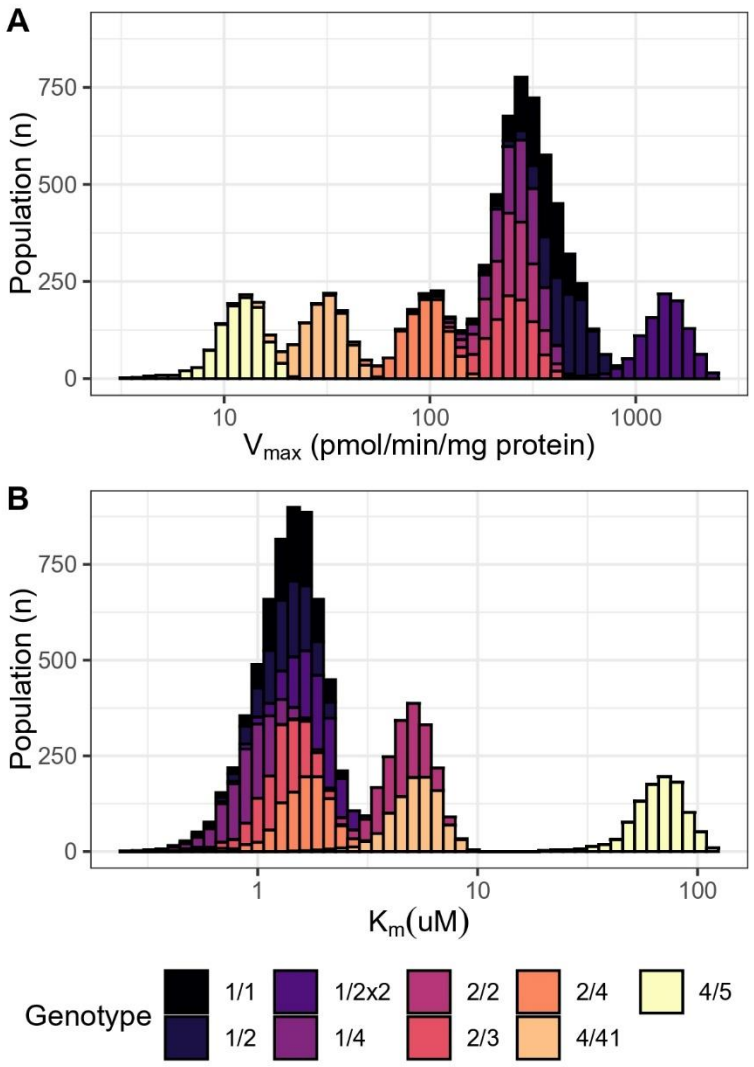

Figure 2.3. Distribution of Michaelis-Menten parameters for virtual population (n=9000) stratified by *CYP2D6* genotype.

K_m represents the Michaelis-Menten constant, and V_{max} represents maximal 4-OH Atomoxetine formation (pmol/min/mg protein) by CYP2D6 in human liver microsomes.

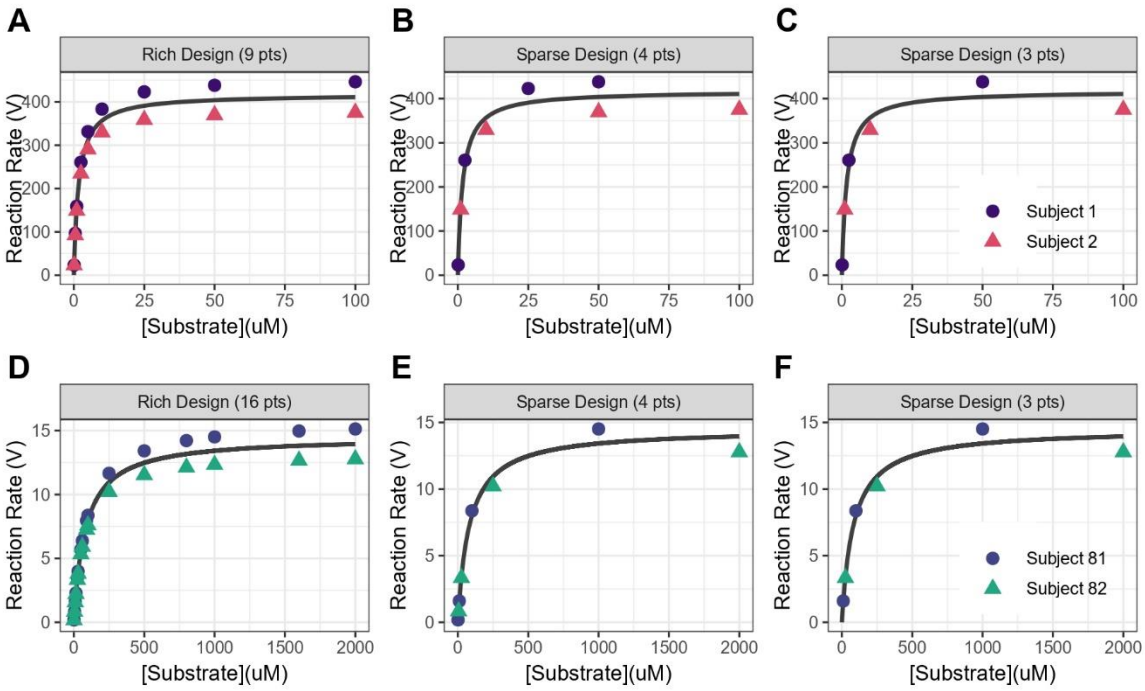

Figure 2.4. Illustration of rich sampling design (9-16 points per subject) vs strategic sparse sampling design (3-4 points per subject) for two subjects with the same genotype and no experimental error (0% CV).

(A-C) Rich, sparse 4-point, and sparse 3-point experimental designs for extensive and ultra-rapid metabolizers; **(D-F)** designs for intermediate and poor metabolizers. Black line represents the non-linear least-squares Michaelis-Menten fit for the genotype group.

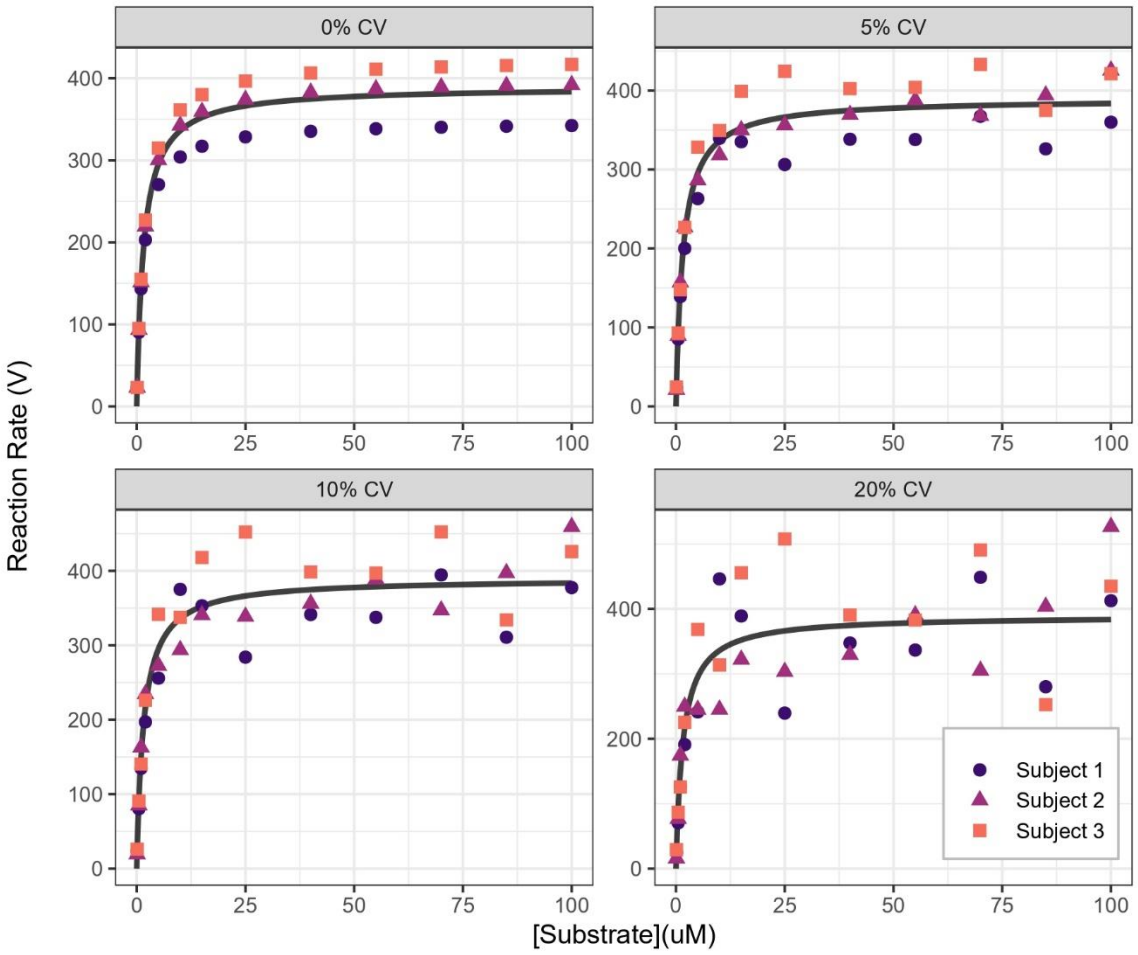

Figure 2.5. Illustration of experimental error scenarios for 3 subjects with the same genotype. Black line represents the non-linear least-squares Michaelis-Menten fit for the genotype group.

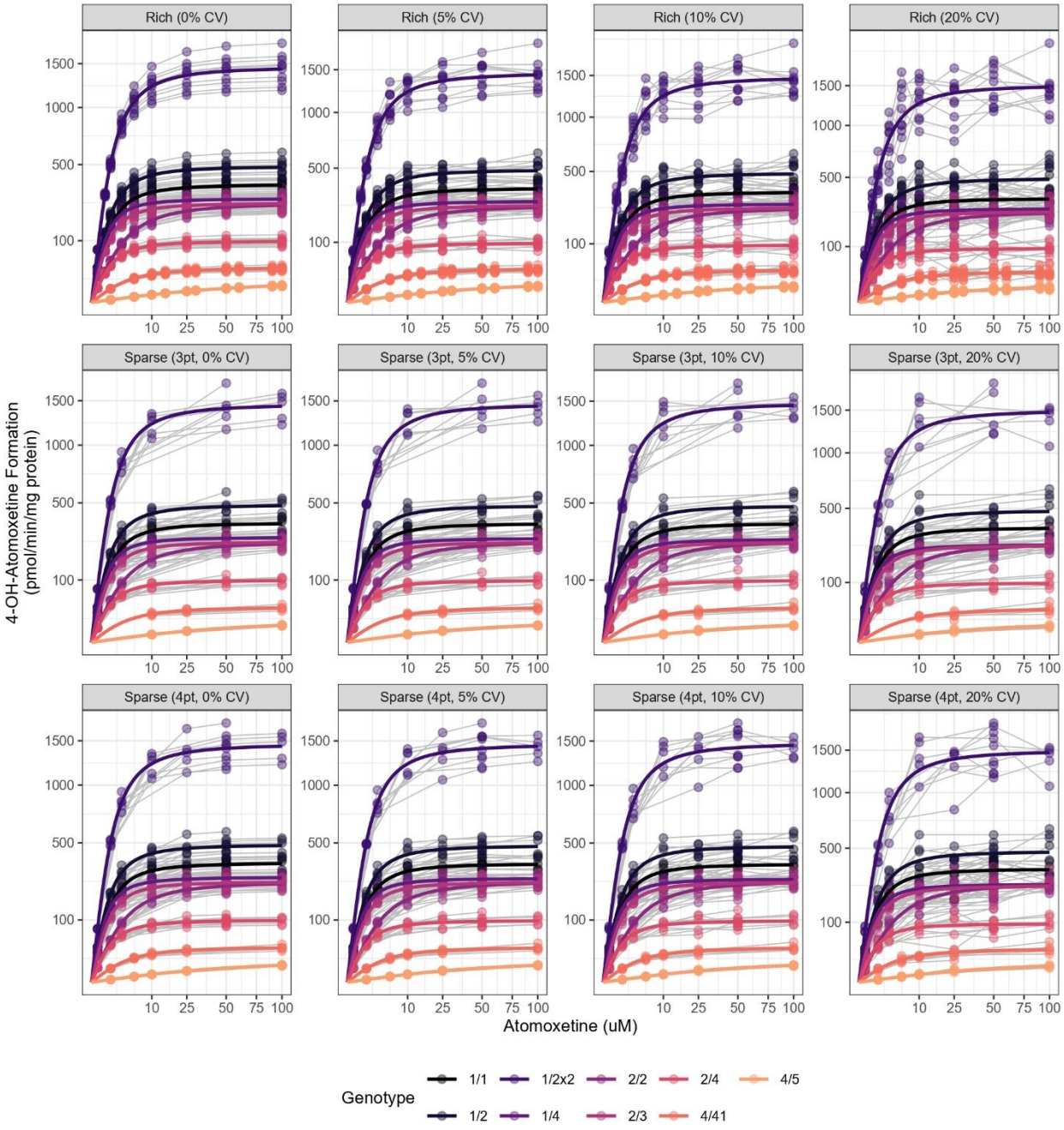

Figure 2.6. In-silico NLME predicted Michaelis-Menten fits of 4-hydroxy-atomoxetine formation using human liver microsomes.

Profiles for CYP2D6 extensive metabolizers (**1/*1, *1/*2, *1/*4, *2/*2, *2/*3, and *2/*4*), ultra-rapid metabolizers (**1/*2x2*), intermediate (**4/*41*), and poor metabolizers (**4/*5*) across all experimental conditions (sparse 3-4pt, 0-20% CV).

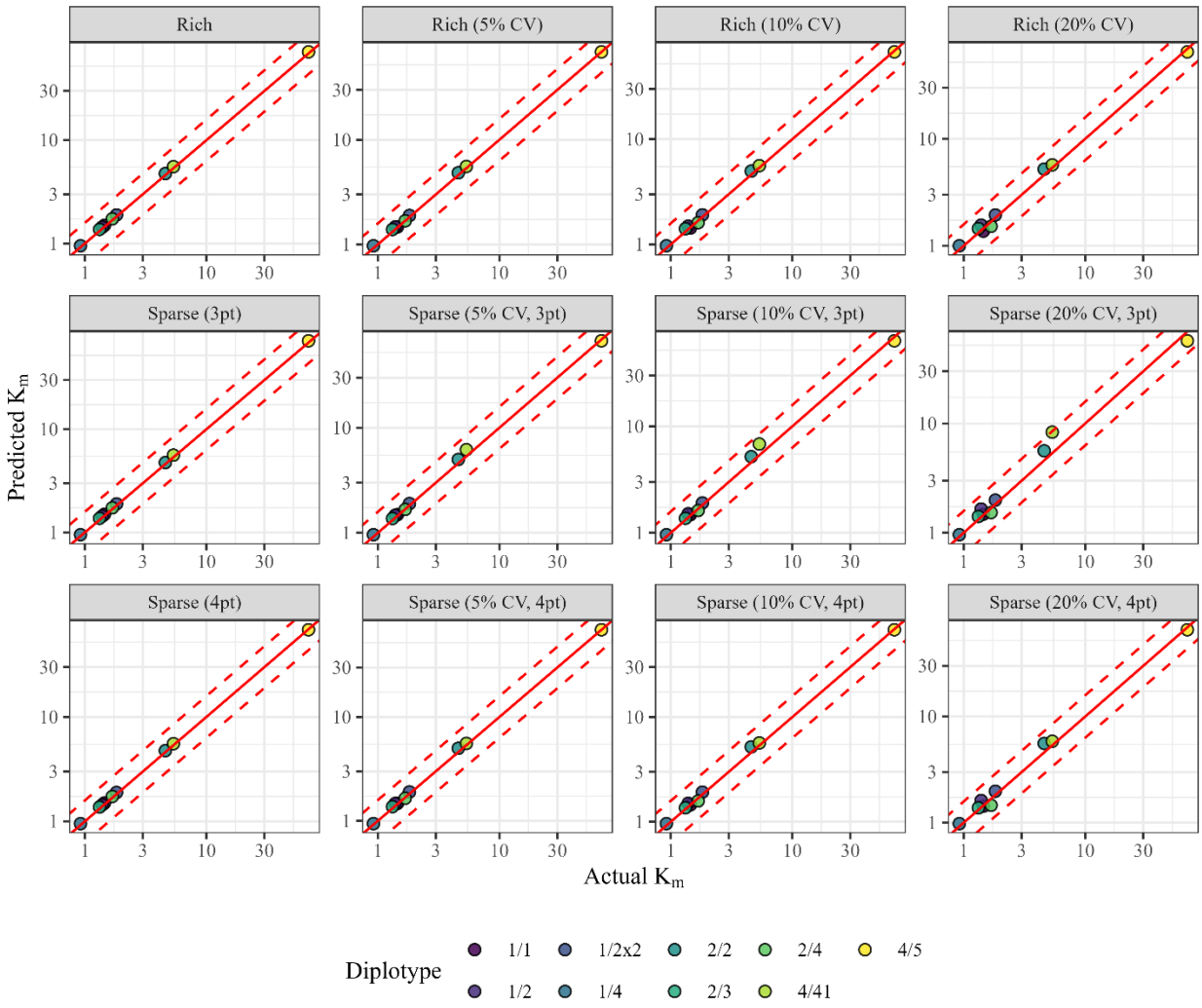

Figure 2.7. Predicted vs. actual K_m estimates, stratified by experimental condition. Dotted red lines represent a 2-fold change from the reference value.

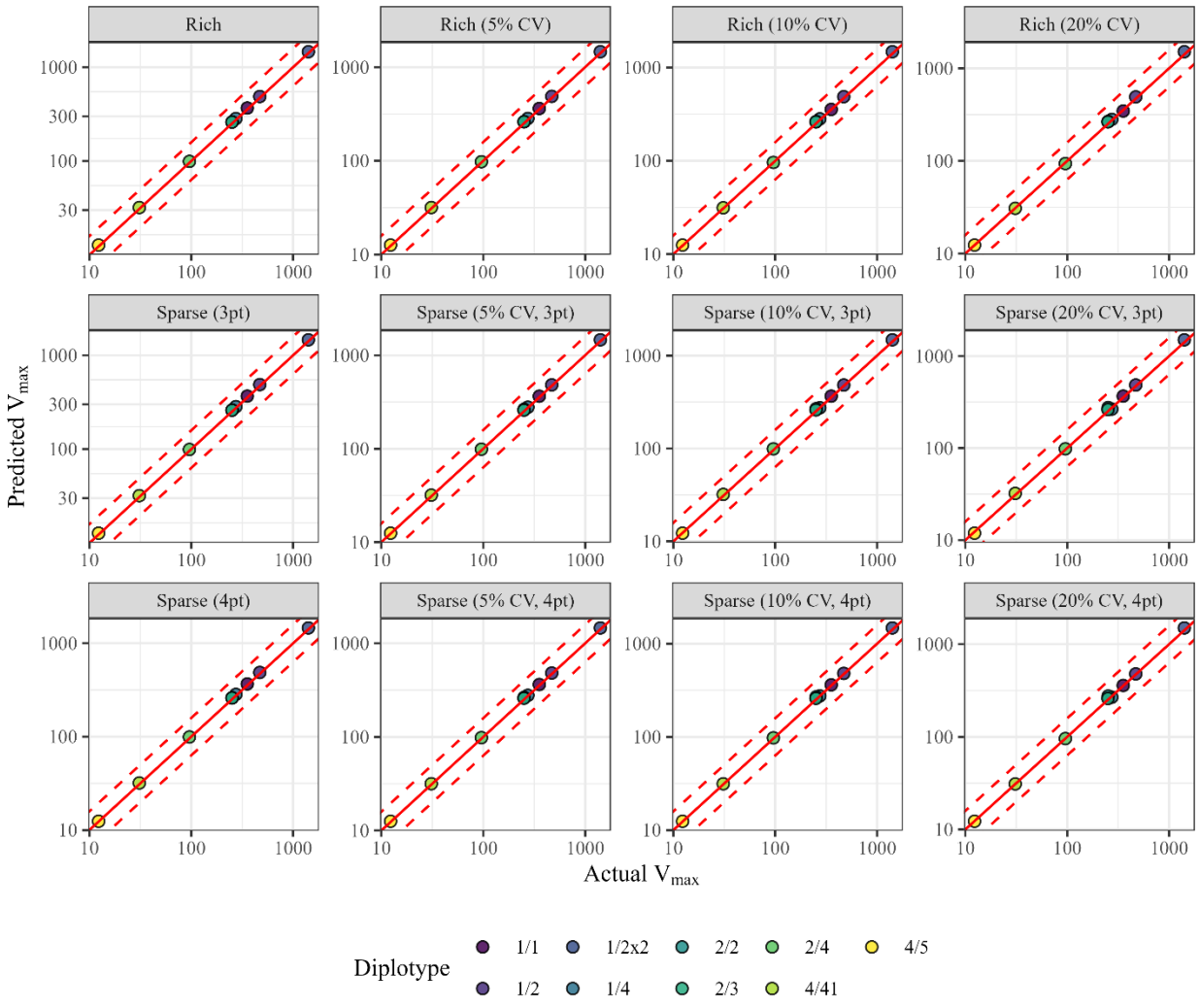

Figure 2.8. Predicted vs. actual V_{max} estimates, stratified by experimental condition. Dotted red lines represent a 2-fold change from the reference value.

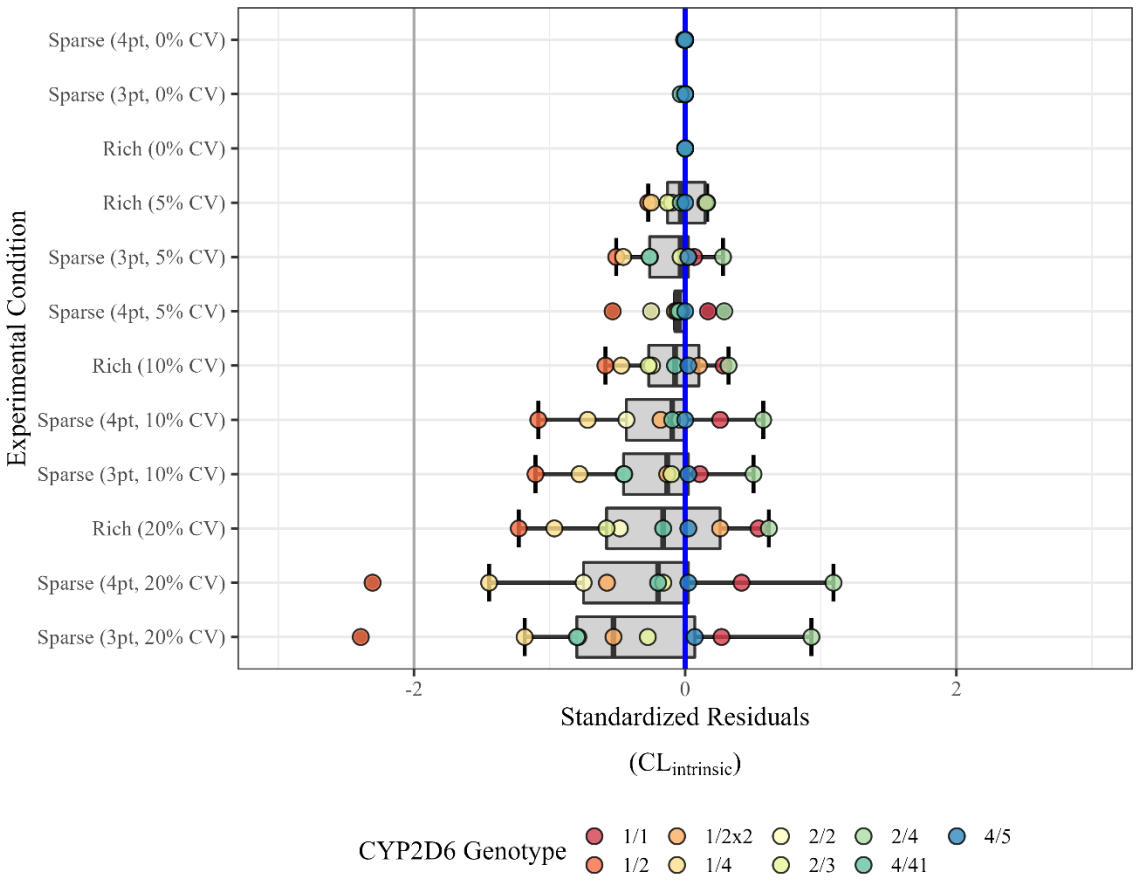

Figure 2.9. Pearson standardized residuals (R_p) of estimated CL_{int} from reference values for CYP2D6 genotypes stratified by experimental condition. Gray lines represent $-2<R_p<2$ bounds.

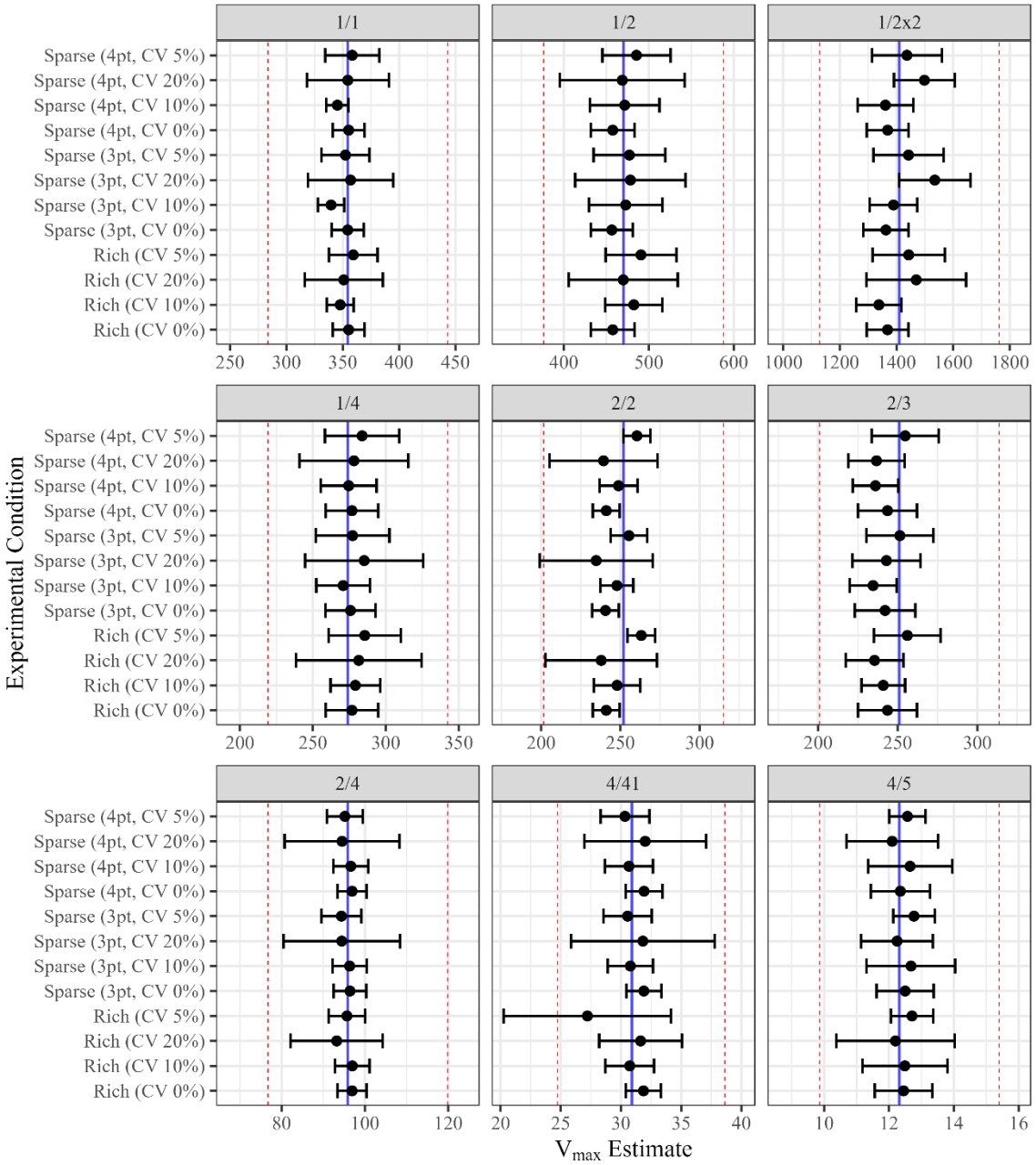

Figure 2.10. Bootstrapped parameter estimates with 95% confidence intervals (CI) for the final model across all experimental conditions (stratified by CYP2D6 genotype).

Black markers with intervals represent the mean estimate and 95% confidence interval of 10 bootstrapped sample populations, blue line represents true estimate of V_{max} for specified genotype group, dotted red lines represent 80%-125% lower-and-upper bounds of the parameter estimate.

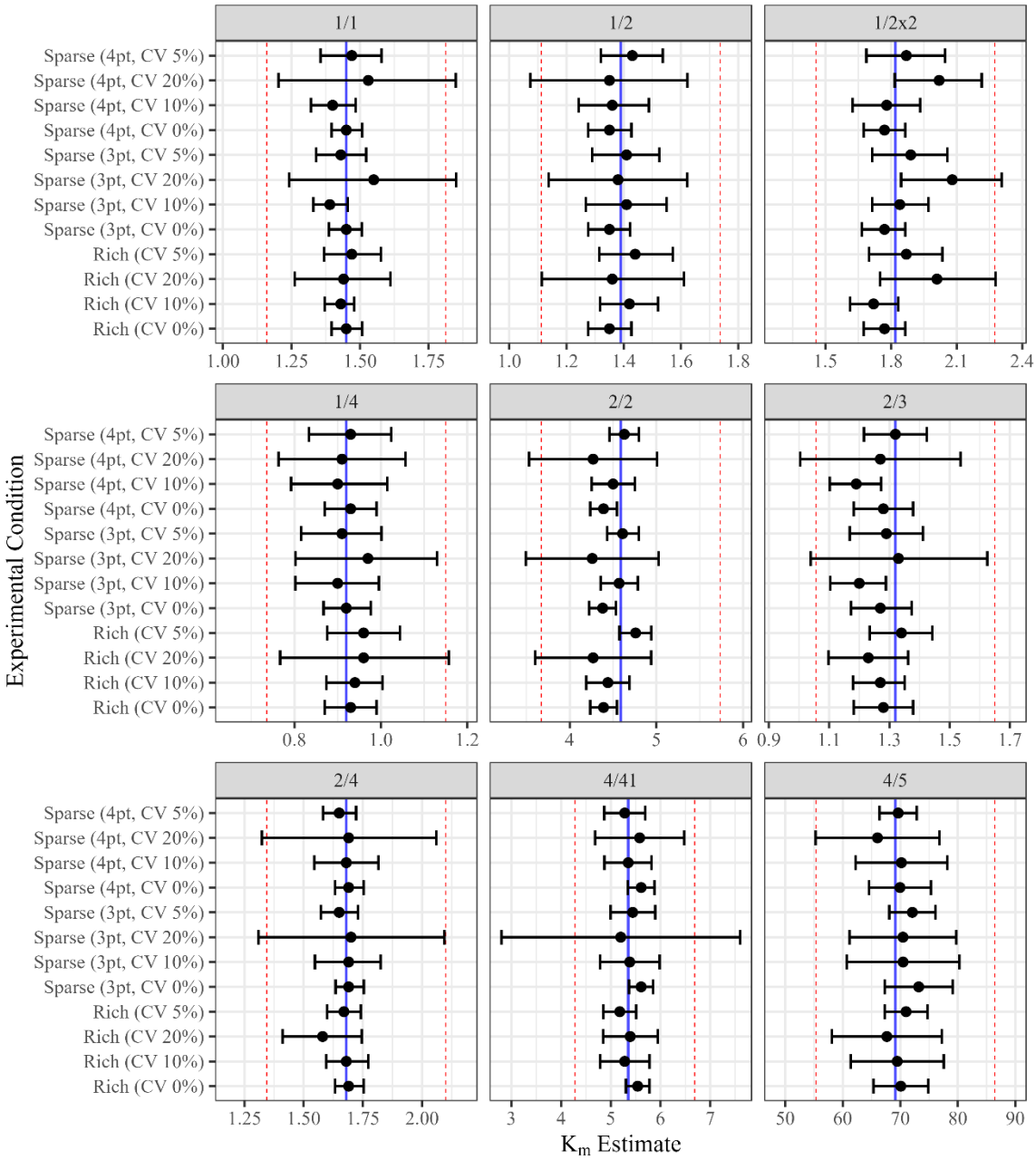

Figure 2.11. Bootstrapped parameter estimates with 95% confidence intervals (CI) for the final model across all experimental conditions (stratified by CYP2D6 genotype).

Black markers with intervals represent the mean estimate and 95% confidence interval of 10 bootstrapped sample populations, blue line represents true estimate of K_m for specified genotype group, dotted red lines represent 80%-125% lower-and-upper bounds of the parameter estimate.

Table 2.1. Extracted CYP2D6 Genotype Michaelis-Menten Parameter Estimates for Atomoxetine Metabolism.

CYP2D6 Genotype	**N**	**Age** (*yr*)	**K_m** (μM)	**V_{max}** (*pmol/min/mg protein*)	**Activity Score**
**1/*1*	4	43 ± 30	1.45	354.50	2
**1/*2*	3	25 ± 27	1.39	470.33	2
**1/*2x2*	2	33 ± 38	1.82	1410.00	3
**1/*4*	4	41 ± 26	0.92	274.00	1
**2/*2*	1	61 ± NA	4.59	252.00	2
**2/*3*	1	20 ± NA	1.32	251.00	1
**2/*4*	3	56 ± 12	1.68	95.90	1
**4/*41*	1	21 ± NA	5.35	30.90	0.5
**4/*5*	2	46 ± 4	69.15	12.32	0

Table 2.2. Virtual Population Michaelis-Menten Parameter Estimate Ranges

CYP2D6 Genotype	V_{max} Range (pmol/min/mg protein)	K_m Range (*u*M)
1/1	100.02 - 617.94	0.41 - 2.53
1/2	132.7 - 819.85	0.39 - 2.42
1/2x2	397.81 - 2457.8	0.51 - 3.18
1/4	77.31 - 477.61	0.26 - 1.6
2/2	71.1 - 439.27	1.3 - 8
2/3	70.82 - 437.52	0.37 - 2.3
2/4	27.06 - 167.16	0.47 - 2.92
4/41	8.72 - 53.86	1.51 - 9.33
4/5	3.48 - 21.48	19.51 - 120.54
All	3.48 - 2457.8 (707-fold)	0.26 - 120.54 (464-fold)

Table 2.3. Comparison of Virtual Population Kinetic Parameter Estimates vs Actual.

	V_{max} Estimate (*pmol/min/mg protein*)		**K_m Estimate** (μM)		
Genotype	**Actual.**	**Simulated (mean, SD)**	**Actual**	**Simulated (mean, SD)**	**n**
**1/*1*	354.50	352.17 ± 88.91	1.45	1.44 ± 0.36	1000
**1/*2*	470.33	467.24 ± 117.97	1.39	1.38 ± 0.35	1000
**1/*2x2*	1410.00	1400.73 ± 353.65	1.82	1.81 ± 0.46	1000
**1/*4*	274.00	272.20 ± 68.72	0.92	0.91 ± 0.23	1000
**2/*2*	252.00	250.34 ± 63.20	4.59	4.56 ± 1.15	1000
**2/*3*	251.00	249.35 ± 62.95	1.32	1.31 ± 0.33	1000
**2/*4*	95.90	95.27 ± 24.05	1.68	1.67 ± 0.42	1000
**4/*41*	30.90	30.70 ± 7.75	5.35	5.31 ± 1.34	1000
**4/*5*	12.32	12.24 ± 3.09	69.15	68.7 ± 17.34	1000

Abbreviations: SD; Standard deviation.

Table 2.4. Summary of Parameter Estimates for Sample Population

CYP2D6 Genotype	V_{max} (*pmol/min/mg protein*)	K_m (μM)	n
**1/*1*	373.70 ± 62.84	1.53 ± 0.26	10
**1/*2*	495.81 ± 83.38	1.47 ± 0.25	10
**1/*2x2*	1486.38 ± 249.95	1.92 ± 0.32	10
**1/*4*	288.84 ± 48.57	0.97 ± 0.16	10
**2/*2*	265.65 ± 44.67	4.84 ± 0.81	10
**2/*3*	264.60 ± 44.5	1.39 ± 0.23	10
**2/*4*	101.09 ± 17	1.77 ± 0.30	10
**4/*41*	32.57 ± 5.48	5.64 ± 0.95	10
**4/*5*	12.99 ± 2.18	72.9 ± 12.26	10

V_{max} and K_m values represent mean ± standard deviation of each genotype group.

Table 2.5. UM and EM Design - Model Optimization – 0% CV

	Model	AIC	BIC	Log-Likelihood	Test	L.Ratio	*p-value*
Rich Design (9 pt, 0% CV)							
Base Model (0% CV)	1	-2377.19	-2354.96	1193.59			
Covariate Model (0% CV)	2	-2801.26	-2725.69	1417.63	1 vs 2	448.08	<.0001
Sparse Design (4 pt, 0% CV)							
Base Model (0% CV)	1	626.40	644.58	-308.20			
Covariate Model (0% CV)	2	202.32	264.11	-84.16	1 vs 2	448.08	<.0001
Sparse Design (3 pt, 0% CV)							
Base Model (3 pt, 0% CV)	3	1200.06	1216.79	-595.03			
Covariate Model (3 pt, 0% CV)	4	775.98	832.88	-370.99	1 vs 2	448.08	<.0001

Abbreviations: AIC: Akaike information criterion, BIC: Bayesian information criterion, L.Ratio: Likelihood Ratio.

Table 2.6. UM and EM Design - Model Optimization – 5% CV

	Model	AIC	BIC	Log-Likelihood	Test	L.Ratio	*p-value*
Rich Design (9pt, 5% CV)							
Base Model (5% CV)	1	6349.28	6371.51	-3169.64			
Covariate Model (5% CV)	2	5959.28	6034.86	-2962.64	1 vs 2	414.00	<.0001
Covariate & Error Model (5% CV)	3	4943.38	5023.40	-2453.69	2 vs 3	1017.90	<.0001
Sparse Design (3pt, 5% CV)							
Base Model (5% CV)	1	2404.27	2421.00	-1197.13			
Covariate Model (5% CV)	2	2084.09	2140.99	-1025.05	1 vs 2	344.18	<.0001
Covariate & Error Model (5% CV)	3	1714.18	1774.43	-839.09	2 vs 3	371.91	<.0001
Sparse Design (4pt, 5% CV)							
Base Model (5% CV)	1	3071.64	3089.81	-1530.82			
Covariate Model (5% CV)	2	2757.41	2819.20	-1361.70	1 vs 2	338.23	<.0001
Covariate & Error Model (5% CV)	3	2374.67	2440.10	-1169.34	2 vs 3	384.74	<.0001

Abbreviations: AIC: Akaike information criterion, BIC: Bayesian information criterion, L.Ratio: Likelihood Ratio.

Table 2.7. UM and EM Design - Model Optimization – 10% CV

	Model	AIC	BIC	Log-Likelihood	Test	L.Ratio	*p-value*
Rich Design (9pt, 10% CV)							
Base Model (10% CV)	1	6998.13	7020.36	-3494.07			
Covariate Model (10% CV)	2	6710.77	6786.35	-3338.39	1 vs 2	311.36	<.0001
Covariate & Error Model (10% CV)	3	5571.44	5651.47	-2767.72	2 vs 3	1141.33	<.0001
Sparse Design (3pt, 10% CV)							
Base Model (10% CV)	1	2530.76	2547.50	-1260.38			
Covariate Model (10% CV)	2	2260.63	2317.53	-1113.31	1 vs 2	294.13	<.0001
Covariate & Error Model (10% CV)	3	1856.38	1916.63	-910.19	2 vs 3	406.25	<.0001
Sparse Design (4pt, 10% CV)							
Base Model (10% CV)	1	3286.37	3304.55	-1638.19			
Covariate Model (10% CV)	2	3025.46	3087.25	-1495.73	1 vs 2	284.91	<.0001
Covariate & Error Model (10% CV)	3	2576.89	2642.32	-1270.44	2 vs 3	450.57	<.0001

Abbreviations: AIC: Akaike information criterion, BIC: Bayesian information criterion, L.Ratio: Likelihood Ratio.

Table 2.8. UM and EM Design - Model Optimization – 20% CV

	Model	AIC	BIC	Log-Likelihood	Test	L.Ratio	*p-value*
Rich Design (9pt, 20% CV)							
Base Model (20% CV)	1	7727.85	7750.08	-3858.92			
Covariate Model (20% CV)	2	7492.45	7568.03	-3729.23	1 vs 2	259.39	<.0001
Covariate & Error Model (20% CV)	3	6295.86	6375.88	-3129.93	2 vs 3	1198.59	<.0001
Sparse Design (3pt, 20% CV)							
Base Model (20% CV)	1	2700.70	2717.44	-1345.35			
Covariate Model (20% CV)	2	2485.75	2542.65	-1225.88	1 vs 2	238.95	<.0001
Covariate & Error Model (20% CV)	3	2077.69	2137.94	-1020.85	2 vs 3	410.06	<.0001
Sparse Design (4pt, 20% CV)							
Base Model (20% CV)	1	3555.14	3573.31	-1772.57			
Covariate Model (20% CV)	2	3334.11	3395.90	-1650.06	1 vs 2	245.02	<.0001
Covariate & Error Model (20% CV)	3	2873.73	2939.16	-1418.87	2 vs 3	462.38	<.0001

Abbreviations: AIC: Akaike information criterion, BIC: Bayesian information criterion, L.Ratio: Likelihood Ratio.

Table 2.9. PM and IM Design - Model Optimization – 0% CV

	Model	AIC	BIC	Log-Likelihood	Test	L.Ratio	*p-value*
Rich Design (16 pt, 0% CV)							
Base Model (0% CV)	1	-2099.50	-2080.66	1054.75			
Covariate Model (0% CV)	2	-2211.98	-2185.60	1112.99	1 vs 2	116.47	<.0001
Covariate & Error Model (0% CV)	3	-2212.96	-2182.82	1114.48	2 vs 3	2.99	0.0839
Sparse Design (3 pt, 0% CV)							
Base Model (0% CV)	1	122.28	132.76	-56.14			
Covariate Model (0% CV)	2	156.83	171.49	-71.41	1 vs 2	30.54	<.0001
Covariate & Error Model (0% CV)	3	148.28	164.98	-66.11	2 vs 3	10.60	0.0011
Rich Design (4 pt, 0% CV)							
Base Model (0% CV)	1	-37.86	-25.95	23.93			
Covariate Model (0% CV)	2	180.59	197.26	-83.30	1 vs 2	214.45	<.0001
Covariate & Error Model (0% CV)	3	132.79	151.85	-58.39	2 vs 3	49.80	<.0001

Abbreviations: AIC: Akaike information criterion, BIC: Bayesian information criterion, L.Ratio: Likelihood Ratio.

Table 2.10. PM and IM Design - Model Optimization – 5% CV

	Model	AIC	BIC	Log-Likelihood	Test	L.Ratio	*p-value*
Rich Design (16pt, 5% CV)							
Base Model (5% CV)	1	1183.66	1202.50	-586.83			
Covariate Model (5% CV)	2	1033.36	1059.74	-509.68	1 vs 2	154.30	<.0001
Covariate & Error Model (5% CV)	3	684.47	714.62	-334.23	2 vs 3	350.89	<.0001
Sparse Design (3pt, 5% CV)							
Base Model (5% CV)	1	376.35	386.82	-183.17			
Covariate Model (5% CV)	2	255.49	270.15	-120.74	1 vs 2	124.86	<.0001
Covariate & Error Model (5% CV)	3	219.63	236.38	-101.81	2 vs 3	37.86	<.0001
Sparse Design (4pt, 5% CV)							
Base Model (5% CV)	1	433.01	444.92	-211.50			
Covariate Model (5% CV)	2	302.04	318.71	-144.02	1 vs 2	134.97	<.0001
Covariate & Error Model (5% CV)	3	176.97	196.03	-80.48	2 vs 3	127.07	<.0001

Abbreviations: AIC: Akaike information criterion, BIC: Bayesian information criterion, L.Ratio: Likelihood Ratio.

Table 2.11. PM and IM Design - Model Optimization – 10% CV

	Model	AIC	BIC	Log-Likelihood	Test	L.Ratio	*p-value*
Rich Design (16pt, 10% CV)							
Base Model (10% CV)	1	1573.83	1592.67	-781.91			
Covariate Model (10% CV)	2	1437.77	1464.15	-711.89	1 vs 2	140.05	<.0001
Covariate & Error Model (10% CV)	3	1027.94	1058.08	-505.97	2 vs 3	411.84	<.0001
Sparse Design (3pt, 10% CV)							
Base Model (10% CV)	1	391.01	401.49	-190.51			
Covariate Model (10% CV)	2	313.12	327.78	-149.56	1 vs 2	81.89	<.0001
Covariate & Error Model (10% CV)	3	258.63	275.38	-121.31	2 vs 3	56.50	<.0001
Sparse Design (4pt, 10% CV)							
Base Model (10% CV)	1	483.27	495.18	-236.63			
Covariate Model (10% CV)	2	384.12	400.79	-185.06	1 vs 2	103.15	<.0001
Covariate & Error Model (10% CV)	3	240.47	259.53	-112.24	2 vs 3	145.65	<.0001

Abbreviations: AIC: Akaike information criterion, BIC: Bayesian information criterion, L.Ratio: Likelihood Ratio.

Table 2.12. PM and IM Design - Model Optimization – 20% CV

	Model	AIC	BIC	Log-Likelihood	Test	L.Ratio	*p-value*
Rich Design (16pt), 20% CV)							
Base Model (20% CV)	1	1978.05	1996.89	-984.02			
Covariate Model (20% CV)	2	1853.25	1879.63	-919.63	1 vs 2	128.80	<.0001
Covariate & Error Model (20% CV)	3	1420.19	1450.34	-702.10	2 vs 3	435.06	<.0001
Sparse Design (3pt, 20% CV)							
Base Model (20% CV)	1	425.78	436.25	-207.89			
Covariate Model (20% CV)	2	380.12	394.78	-183.06	1 vs 2	49.66	<.0001
Covariate & Error Model (20% CV)	3	317.46	334.21	-150.73	2 vs 3	64.66	<.0001
Sparse Design (4pt, 20% CV)							
Base Model (20% CV)	1	531.56	543.47	-260.78			
Covariate Model (20% CV)	2	477.72	494.40	-231.86	1 vs 2	57.84	<.0001
Covariate & Error Model (20% CV)	3	328.93	347.99	-156.47	2 vs 3	150.79	<.0001

Abbreviations: AIC: Akaike information criterion, BIC: Bayesian information criterion, L.Ratio: Likelihood Ratio.

Table 2.13. Final Covariate Model Estimates (w/ 0% Residual Error & 95% CI)

			Rich			Sparse (3pt)			Sparse (4pt)		
Parameter	Genotype	Actual	Estimate	CI (95%)	RSE (%)	Estimate	CI (95%)	RSE (%)	Estimate	CI (95%)	RSE (%)
V_{max} Estimate *(pmol/min/mg protein)*											
V_{max}	**1/*1*	354.50	367.01	(319.69, 414.33)	6.64	367.01	(319.34, 414.67)	6.79	367.01	(319.51, 414.51)	6.73
V_{max}	**1/*2*	470.33	486.93	(420.01, 553.85)	7.08	486.93	(419.52, 554.34)	7.24	486.93	(419.75, 554.11)	7.18
V_{max}	**1/*2x2*	1410.00	1459.75	(1392.84, 1526.67)	2.36	1459.75	(1392.34, 1527.16)	2.42	1459.75	(1392.57, 1526.93)	2.39
V_{max}	**1/*4*	274.00	283.67	(216.75, 350.58)	12.14	283.67	(216.26, 351.08)	12.43	283.67	(216.49, 350.85)	12.32
V_{max}	**2/*2*	252.00	260.89	(193.98, 327.81)	13.21	260.89	(193.48, 328.3)	13.52	260.89	(193.71, 328.07)	13.40
V_{max}	**2/*3*	251.00	259.86	(192.94, 326.77)	13.26	259.86	(192.45, 327.27)	13.57	259.86	(192.68, 327.04)	13.45
V_{max}	**2/*4*	95.90	99.28	(32.37, 166.2)	34.70	99.28	(31.87, 166.69)	35.52	99.28	(32.1, 166.46)	35.20
V_{max}	**4/*41*	30.90	31.99	(30.14, 33.84)	2.96	31.99	(30.28, 33.69)	2.72	31.94	(30.2, 33.67)	2.79
V_{max}	**4/*5*	12.32	12.75	(10.14, 15.37)	10.48	12.74	(10.34, 15.15)	9.65	12.43	(10.64, 14.23)	7.40
K_m Estimate *(uM)*											
K_m	**1/*1*	1.45	1.50	(1.33, 1.67)	5.92	1.50	(1.33, 1.67)	6.06	1.50	(1.33, 1.67)	6.00
K_m	**1/*2*	1.39	1.44	(1.2, 1.68)	8.73	1.44	(1.19, 1.68)	8.93	1.44	(1.19, 1.68)	8.85
K_m	**1/*2x2*	1.82	1.89	(1.65, 2.13)	6.65	1.89	(1.64, 2.14)	6.80	1.89	(1.64, 2.13)	6.74
K_m	**1/*4*	0.92	0.95	(0.71, 1.2)	13.18	0.95	(0.71, 1.2)	13.49	0.95	(0.71, 1.2)	13.37
K_m	**2/*2*	4.59	4.75	(4.51, 5)	2.64	4.75	(4.51, 5)	2.71	4.75	(4.51, 5)	2.68
K_m	**2/*3*	1.32	1.37	(1.12, 1.61)	9.19	1.37	(1.12, 1.61)	9.41	1.37	(1.12, 1.61)	9.32
K_m	**2/*4*	1.68	1.74	(1.49, 1.98)	7.24	1.74	(1.49, 1.98)	7.41	1.74	(1.49, 1.98)	7.34
K_m	**4/*41*	5.35	5.54	(1.69, 9.39)	35.57	5.61	(5.3, 5.92)	2.83	5.54	(5.07, 6.02)	4.38
K_m	**4/*5*	69.15	71.60	(66.15, 77.05)	3.89	71.96	(67.76, 76.15)	2.98	68.35	(65.66, 71.04)	2.02

Abbreviations: RSE: Relative Standard Error, CI: Confidence Interval.

Table 2.14. Final Covariate Model Estimates (w/ 5% Residual Error & 95% CI)

			Rich (5% CV)			Sparse (3pt, 5% CV)			Sparse (4pt, 5% CV)		
Parameter	Genotype	Actual	Estimate	CI (95%)	RSE (%)	Estimate	CI (95%)	RSE (%)	Estimate	CI (95%)	RSE (%)
V_{max} Estimate *(pmol/min/mg protein)*											
V_{max}	**1/*1*	354.50	362.01	(326.7, 397.31)	5.02	364.79	(346.49, 383.1)	2.63	363.04	(343.06, 383.02)	2.86
V_{max}	**1/*2*	470.33	487.84	(437.63, 538.06)	5.30	482.00	(452.75, 511.25)	3.17	481.46	(451.75, 511.16)	3.21
V_{max}	**1/*2x2*	1410.00	1461.41	(1405.98, 1516.83)	1.95	1463.40	(1387.76, 1539.04)	2.70	1460.85	(1405.71, 1515.98)	1.96
V_{max}	**1/*4*	274.00	283.50	(233.72, 333.28)	9.04	278.28	(254.56, 302.01)	4.46	278.34	(250.88, 305.8)	5.13
V_{max}	**2/*2*	252.00	261.57	(211.72, 311.41)	9.81	264.06	(239.64, 288.48)	4.84	264.62	(236.97, 292.26)	5.44
V_{max}	**2/*3*	251.00	261.41	(211.64, 311.17)	9.80	258.97	(235.43, 282.51)	4.75	259.09	(231.73, 286.45)	5.49
V_{max}	**2/*4*	95.90	97.83	(48.22, 147.45)	26.11	98.79	(77.14, 120.44)	11.46	98.20	(71.53, 124.86)	14.13
V_{max}	**4/*41*	30.90	31.70	(30.04, 33.35)	2.67	31.88	(30.1, 33.65)	2.85	31.62	(29.89, 33.36)	2.81
V_{max}	**4/*5*	12.32	12.60	(10.27, 14.92)	9.44	12.47	(10.11, 14.84)	9.70	12.56	(10.72, 14.4)	7.51
K_m Estimate *(uM)*											
K_m	**1/*1*	1.45	1.47	(1.34, 1.59)	4.26	1.49	(1.38, 1.59)	3.74	1.47	(1.36, 1.58)	3.87
K_m	**1/*2*	1.39	1.46	(1.29, 1.63)	6.03	1.46	(1.31, 1.62)	5.41	1.47	(1.31, 1.62)	5.54
K_m	**1/*2x2*	1.82	1.88	(1.71, 2.06)	4.81	1.89	(1.71, 2.08)	5.18	1.90	(1.72, 2.07)	4.87
K_m	**1/*4*	0.92	0.97	(0.8, 1.13)	8.93	0.96	(0.83, 1.09)	6.97	0.94	(0.79, 1.08)	7.96
K_m	**2/*2*	4.59	4.83	(4.61, 5.05)	2.37	4.99	(4.65, 5.32)	3.51	4.99	(4.73, 5.25)	2.71
K_m	**2/*3*	1.32	1.39	(1.22, 1.56)	6.33	1.37	(1.22, 1.51)	5.48	1.37	(1.22, 1.52)	5.72
K_m	**2/*4*	1.68	1.68	(1.5, 1.85)	5.31	1.67	(1.52, 1.81)	4.61	1.65	(1.5, 1.8)	4.73
K_m	**4/*41*	5.35	5.56	(5.17, 5.96)	3.65	6.19	(5.07, 7.31)	9.23	5.57	(5.03, 6.11)	4.97
K_m	**4/*5*	69.15	69.17	(67.09, 71.25)	1.54	67.23	(61.18, 73.28)	4.60	69.26	(65.67, 72.85)	2.66

Abbreviations: RSE: Relative Standard Error, CI: Confidence Interval.

Table 2.15. Final Covariate Model Estimates (w/ 10% Residual Error & 95% CI)

			Rich (10% CV)			Sparse (3pt, 10% CV)			Sparse (4pt, 10% CV)		
Parameter	Genotype	Actual	Estimate	CI (95%)	RSE (%)	Estimate	CI (95%)	RSE (%)	Estimate	CI (95%)	RSE (%)
V_{max} Estimate *(pmol/min/mg protein)*											
V_{max}	**1/*1*	354.50	355.96	(336.28, 375.64)	2.85	365.54	(341.6, 389.48)	3.43	361.21	(339.36, 383.05)	3.15
V_{max}	**1/*2*	470.33	487.46	(457.38, 517.54)	3.18	481.34	(442.9, 519.79)	4.18	479.61	(445.64, 513.58)	3.68
V_{max}	**1/*2x2*	1410.00	1479.15	(1420.74, 1537.56)	2.03	1476.20	(1380.61, 1571.8)	3.39	1474.03	(1400.08, 1547.98)	2.61
V_{max}	**1/*4*	274.00	282.04	(255.33, 308.75)	4.88	273.99	(243.4, 304.57)	5.84	275.03	(246.08, 303.98)	5.48
V_{max}	**2/*2*	252.00	263.89	(236.56, 291.22)	5.33	267.54	(235.45, 299.63)	6.27	268.41	(238.59, 298.23)	5.78
V_{max}	**2/*3*	251.00	262.09	(235.47, 288.71)	5.23	259.86	(229.39, 290.33)	6.13	259.08	(230.28, 287.89)	5.78
V_{max}	**2/*4*	95.90	96.22	(70.87, 121.57)	13.57	98.98	(71.55, 126.42)	14.50	97.58	(70.7, 124.47)	14.33
V_{max}	**4/*41*	30.90	31.37	(29.86, 32.87)	2.45	32.02	(29.53, 34.52)	3.98	31.40	(29.23, 33.57)	3.55
V_{max}	**4/*5*	12.32	12.50	(10.44, 14.56)	8.44	12.25	(9.58, 14.92)	11.12	12.46	(10.1, 14.81)	9.68
K_m Estimate *(uM)*											
K_m	**1/*1*	1.45	1.43	(1.33, 1.53)	3.63	1.48	(1.32, 1.65)	5.85	1.45	(1.3, 1.61)	5.43
K_m	**1/*2*	1.39	1.49	(1.34, 1.63)	5.03	1.51	(1.28, 1.75)	8.22	1.51	(1.29, 1.72)	7.54
K_m	**1/*2x2*	1.82	1.91	(1.74, 2.07)	4.51	1.92	(1.64, 2.2)	7.51	1.92	(1.67, 2.17)	6.74
K_m	**1/*4*	0.92	0.97	(0.85, 1.1)	6.55	0.96	(0.76, 1.17)	10.98	0.96	(0.78, 1.15)	10.12
K_m	**2/*2*	4.59	4.97	(4.63, 5.31)	3.54	5.19	(4.64, 5.74)	5.54	5.19	(4.7, 5.69)	4.98
K_m	**2/*3*	1.32	1.41	(1.26, 1.55)	5.17	1.38	(1.15, 1.6)	8.61	1.37	(1.16, 1.57)	7.95
K_m	**2/*4*	1.68	1.61	(1.46, 1.76)	4.76	1.62	(1.38, 1.86)	7.70	1.59	(1.37, 1.8)	7.19
K_m	**4/*41*	5.35	5.60	(5.2, 5.99)	3.60	6.82	(4.18, 9.46)	19.74	5.66	(4.87, 6.45)	7.16
K_m	**4/*5*	69.15	67.53	(64.29, 70.77)	2.46	63.57	(54.88, 72.26)	6.99	68.20	(62.5, 73.89)	4.28

Abbreviations: RSE: Relative Standard Error, CI: Confidence Interval.

Table 2.16. Final Covariate Model Estimates (w/ 20% Residual Error & 95% CI)

			Rich (20% CV)			Sparse (3pt, 20% CV)			Sparse (4pt, 20% CV)		
Parameter	Genotype	Actual	Estimate	CI (95%)	RSE (%)	Estimate	CI (95%)	RSE (%)	Estimate	CI (95%)	RSE (%)
V_{max} Estimate *(pmol/min/mg protein)*											
V_{max}	**1/*1*	354.50	345.88	(319.99, 371.77)	3.85	366.90	(329.06, 404.73)	5.39	358.47	(326.75, 390.19)	4.60
V_{max}	**1/*2*	470.33	491.31	(448.44, 534.19)	4.49	484.18	(421.45, 546.91)	6.78	477.51	(426, 529.02)	5.61
V_{max}	**1/*2x2*	1410.00	1510.05	(1408.03, 1612.07)	3.48	1501.07	(1339.29, 1662.86)	5.64	1497.51	(1373.43, 1621.59)	4.31
V_{max}	**1/*4*	274.00	280.85	(246.74, 314.97)	6.25	265.12	(218.62, 311.62)	9.18	267.56	(227.18, 307.95)	7.85
V_{max}	**2/*2*	252.00	266.92	(230.87, 302.98)	6.95	275.51	(224.49, 326.53)	9.69	275.49	(232.06, 318.91)	8.20
V_{max}	**2/*3*	251.00	264.38	(230.39, 298.38)	6.62	262.08	(215.29, 308.86)	9.34	259.33	(218.97, 299.7)	8.10
V_{max}	**2/*4*	95.90	93.65	(63.33, 123.97)	16.67	98.20	(58.18, 138.22)	21.32	95.77	(60.19, 131.34)	19.32
V_{max}	**4/*41*	30.90	30.77	(29.18, 32.36)	2.64	32.17	(28.23, 36.11)	6.26	31.02	(27.58, 34.46)	5.68
V_{max}	**4/*5*	12.32	12.33	(10.37, 14.3)	8.16	11.89	(7.68, 16.1)	18.10	12.28	(8.52, 16.03)	15.67
K_m Estimate *(uM)*											
K_m	**1/*1*	1.45	1.37	(1.18, 1.55)	6.92	1.48	(1.15, 1.8)	11.68	1.43	(1.14, 1.72)	10.53
K_m	**1/*2*	1.39	1.56	(1.28, 1.83)	9.13	1.64	(1.16, 2.13)	15.47	1.61	(1.18, 2.04)	13.84
K_m	**1/*2x2*	1.82	1.94	(1.63, 2.25)	8.25	1.98	(1.45, 2.51)	14.06	1.98	(1.51, 2.45)	12.32
K_m	**1/*4*	0.92	1.00	(0.77, 1.23)	11.91	0.96	(0.54, 1.37)	22.61	0.98	(0.62, 1.34)	19.16
K_m	**2/*2*	4.59	5.20	(4.52, 5.88)	6.71	5.62	(4.48, 6.76)	10.62	5.58	(4.53, 6.64)	9.82
K_m	**2/*3*	1.32	1.45	(1.18, 1.72)	9.47	1.41	(0.95, 1.86)	17.00	1.38	(0.98, 1.78)	15.24
K_m	**2/*4*	1.68	1.51	(1.24, 1.79)	9.29	1.53	(1.06, 2)	16.19	1.46	(1.04, 1.88)	14.96
K_m	**4/*41*	5.35	5.70	(4.93, 6.48)	6.96	8.33	(3.77, 12.89)	27.95	5.85	(4.43, 7.28)	12.49
K_m	**4/*5*	69.15	64.82	(59.1, 70.55)	4.51	57.03	(42.81, 71.26)	12.74	65.65	(55.21, 76.09)	8.15

Abbreviations: RSE: Relative Standard Error, CI: Confidence Interval.

Table 2.17. Bootstrapped Estimates (Rich Design)

Genotype	Rich (CV 0%)	Rich (CV 5%)	Rich (CV 10%)	Rich (CV 20%)
Parameter: K_m *(uM)*				
**1/*1*	1.45 (1.4 - 1.51)	1.47 (1.37 - 1.58)	1.43 (1.37 - 1.48)	1.44 (1.26 - 1.61)
**1/*2*	1.35 (1.28 - 1.43)	1.44 (1.31 - 1.57)	1.42 (1.32 - 1.52)	1.36 (1.11 - 1.61)
**1/*2x2*	1.77 (1.68 - 1.86)	1.87 (1.7 - 2.03)	1.72 (1.61 - 1.83)	2.01 (1.75 - 2.28)
**1/*4*	0.93 (0.87 - 0.99)	0.96 (0.88 - 1.04)	0.94 (0.87 - 1)	0.96 (0.77 - 1.16)
**2/*2*	4.39 (4.24 - 4.55)	4.76 (4.57 - 4.94)	4.44 (4.19 - 4.69)	4.27 (3.6 - 4.94)
**2/*3*	1.28 (1.18 - 1.38)	1.34 (1.23 - 1.44)	1.27 (1.18 - 1.35)	1.23 (1.1 - 1.36)
**2/*4*	1.69 (1.63 - 1.75)	1.67 (1.6 - 1.74)	1.68 (1.59 - 1.77)	1.58 (1.41 - 1.75)
**4/*41*	5.54 (5.3 - 5.78)	5.18 (4.85 - 5.51)	5.28 (4.79 - 5.78)	5.39 (4.85 - 5.94)
**4/*5*	70.09 (65.33 - 74.85)	71 (67.29 - 74.71)	69.46 (61.37 - 77.56)	67.66 (58.09 - 77.23)
Parameter: V_{max} *(pmol/min/mg protein)*				
**1/*1*	355.26 (341.18 - 369.33)	359.33 (337.8 - 380.86)	347.74 (335.78 - 359.69)	350.92 (316.18 - 385.66)
**1/*2*	457.58 (431.87 - 483.29)	490.83 (449.17 - 532.5)	482.42 (448.81 - 516.03)	469.97 (405.79 - 534.14)
**1/*2x2*	1368.42 (1294.99 - 1441.85)	1443 (1315.47 - 1570.54)	1337.91 (1258.07 - 1417.75)	1469.67 (1293.6 - 1645.73)
**1/*4*	276.85 (258.85 - 294.85)	285.6 (260.9 - 310.29)	279.18 (262.18 - 296.17)	281.5 (238.41 - 324.6)
**2/*2*	241.07 (232.59 - 249.55)	263.16 (254.51 - 271.82)	247.88 (233.2 - 262.55)	237.89 (202.69 - 273.1)
**2/*3*	243.56 (225.06 - 262.06)	256.02 (235.07 - 276.96)	240.89 (227.19 - 254.59)	235.43 (217.3 - 253.56)
**2/*4*	96.92 (93.45 - 100.38)	95.69 (91.33 - 100.05)	96.97 (92.82 - 101.11)	93.23 (82.19 - 104.28)
**4/*41*	31.87 (30.41 - 33.32)	27.21 (20.26 - 34.15)	30.73 (28.71 - 32.74)	31.63 (28.19 - 35.06)
**4/*5*	12.45 (11.56 - 13.34)	12.71 (12.06 - 13.37)	12.49 (11.18 - 13.8)	12.2 (10.38 - 14.02)

[*]*Mean and 95% confidence interval reported.*

Table 2.18. Bootstrapped Estimates (Sparse 3pt Design)

Genotype	Sparse (3pt, CV 0%)	Sparse (3pt, CV 5%)	Sparse (3pt, CV 10%)	Sparse (3pt, CV 20%)
Parameter: K_m *(uM)*				
**1/*1*	1.45 (1.39 - 1.51)	1.43 (1.34 - 1.52)	1.39 (1.33 - 1.46)	1.55 (1.24 - 1.85)
**1/*2*	1.35 (1.28 - 1.42)	1.41 (1.29 - 1.52)	1.41 (1.27 - 1.55)	1.38 (1.14 - 1.62)
**1/*2x2*	1.77 (1.67 - 1.86)	1.89 (1.71 - 2.06)	1.84 (1.71 - 1.97)	2.08 (1.85 - 2.31)
**1/*4*	0.92 (0.87 - 0.98)	0.91 (0.82 - 1)	0.9 (0.8 - 1)	0.97 (0.8 - 1.13)
**2/*2*	4.38 (4.22 - 4.53)	4.61 (4.43 - 4.8)	4.57 (4.36 - 4.79)	4.26 (3.49 - 5.03)
**2/*3*	1.27 (1.17 - 1.37)	1.29 (1.17 - 1.41)	1.2 (1.1 - 1.29)	1.33 (1.04 - 1.62)
**2/*4*	1.69 (1.63 - 1.75)	1.65 (1.57 - 1.73)	1.69 (1.55 - 1.82)	1.7 (1.31 - 2.09)
**4/*41*	5.61 (5.37 - 5.85)	5.44 (5 - 5.89)	5.38 (4.78 - 5.98)	5.2 (2.8 - 7.6)
**4/*5*	73.2 (67.3 - 79.1)	72.09 (68.08 - 76.09)	70.48 (60.68 - 80.29)	70.46 (61.18 - 79.74)
Parameter: V_{max} *(pmol/min/mg protein)*				
**1/*1*	354.43 (340.11 - 368.74)	352.44 (331.18 - 373.7)	339.63 (327.97 - 351.29)	356.93 (319.02 - 394.84)
**1/*2*	456.5 (431.92 - 481.08)	477.16 (435.01 - 519.32)	472.83 (429.53 - 516.13)	478.35 (413.4 - 543.3)
**1/*2x2*	1362.6 (1283.13 - 1442.07)	1442.35 (1318.39 - 1566.3)	1389.25 (1305.75 - 1472.76)	1535.06 (1409.16 - 1660.95)
**1/*4*	275.84 (258.67 - 293.01)	277.27 (252.01 - 302.53)	270.8 (252.39 - 289.22)	285.23 (244.72 - 325.74)
**2/*2*	240.58 (232.13 - 249.04)	255.37 (243.85 - 266.89)	247.76 (237.41 - 258.11)	234.71 (199.05 - 270.36)
**2/*3*	242.04 (223.04 - 261.05)	251.33 (230.25 - 272.41)	234.49 (219.75 - 249.23)	242.89 (221.55 - 264.23)
**2/*4*	96.42 (92.5 - 100.34)	94.36 (89.55 - 99.17)	96.34 (92.23 - 100.44)	94.45 (80.48 - 108.41)
**4/*41*	31.9 (30.45 - 33.36)	30.55 (28.55 - 32.55)	30.78 (28.9 - 32.67)	31.82 (25.86 - 37.78)
**4/*5*	12.5 (11.62 - 13.38)	12.77 (12.13 - 13.41)	12.68 (11.31 - 14.04)	12.25 (11.14 - 13.35)

[*] *Mean and 95% confidence interval reported.*

.

Table 2.19. Bootstrapped Estimates (Sparse 4pt Design)

Genotype	Sparse (4pt, CV 0%)	Sparse (4pt, CV 5%)	Sparse (4pt, CV 10%)	Sparse (4pt, CV 20%)
Parameter: K_m *(uM)*				
**1/*1*	1.45 (1.4 - 1.51)	1.47 (1.36 - 1.58)	1.4 (1.32 - 1.48)	1.53 (1.2 - 1.85)
**1/*2*	1.35 (1.28 - 1.43)	1.43 (1.32 - 1.54)	1.36 (1.24 - 1.49)	1.35 (1.07 - 1.62)
**1/*2x2*	1.77 (1.68 - 1.86)	1.87 (1.69 - 2.05)	1.78 (1.62 - 1.93)	2.02 (1.82 - 2.21)
**1/*4*	0.93 (0.87 - 0.99)	0.93 (0.83 - 1.02)	0.9 (0.79 - 1.02)	0.91 (0.76 - 1.06)
**2/*2*	4.39 (4.24 - 4.55)	4.63 (4.46 - 4.8)	4.5 (4.25 - 4.75)	4.27 (3.53 - 5.01)
**2/*3*	1.28 (1.18 - 1.38)	1.32 (1.22 - 1.42)	1.19 (1.1 - 1.27)	1.27 (1 - 1.54)
**2/*4*	1.69 (1.63 - 1.75)	1.65 (1.58 - 1.72)	1.68 (1.54 - 1.82)	1.69 (1.32 - 2.06)
**4/*41*	5.61 (5.34 - 5.88)	5.28 (4.87 - 5.69)	5.35 (4.87 - 5.82)	5.58 (4.68 - 6.48)
**4/*5*	69.95 (64.56 - 75.34)	69.63 (66.39 - 72.86)	70.19 (62.23 - 78.15)	66.03 (55.28 - 76.79)
Parameter: V_{max} *(pmol/min/mg protein)*				
**1/*1*	355.26 (341.18 - 369.33)	358.39 (334.39 - 382.38)	345.24 (335.35 - 355.12)	354.55 (318.21 - 390.89)
**1/*2*	457.58 (431.87 - 483.29)	485.51 (445.34 - 525.68)	471.65 (430.7 - 512.6)	468.79 (395.28 - 542.3)
**1/*2x2*	1368.65 (1295.13 - 1442.17)	1436.78 (1313.4 - 1560.16)	1360.81 (1262.49 - 1459.13)	1498.46 (1390.86 - 1606.06)
**1/*4*	276.85 (258.85 - 294.85)	283.78 (258.34 - 309.22)	274.61 (255.47 - 293.74)	278.2 (240.89 - 315.51)
**2/*2*	241.07 (232.59 - 249.55)	260.42 (251.88 - 268.96)	248.8 (236.82 - 260.78)	239.36 (205.37 - 273.34)
**2/*3*	243.56 (225.06 - 262.06)	254.68 (233.64 - 275.71)	236 (221.71 - 250.29)	236.64 (218.99 - 254.29)
**2/*4*	96.92 (93.45 - 100.38)	95.19 (90.88 - 99.5)	96.62 (92.46 - 100.78)	94.51 (80.72 - 108.31)
**4/*41*	31.92 (30.38 - 33.45)	30.33 (28.31 - 32.36)	30.66 (28.66 - 32.67)	32.01 (26.95 - 37.08)
**4/*5*	12.35 (11.44 - 13.26)	12.57 (12.01 - 13.13)	12.65 (11.36 - 13.95)	12.1 (10.69 - 13.52)

[*] *Mean and 95% confidence interval reported.*

Chapter 3. *CYP4F2* AND *CYP4F11* HAPLOTYPE MAPPING AND ASSOCIATION WITH HEPATIC GENE EXPRESSION AND VITAMIN K HYDROXYLATION ACTIVITY

A version of this chapter is currently a draft manuscript in preparation for peer-review.

3.1 ABSTRACT

In this chapter, we evaluated the underlying mechanistic links between genetic variability in vitamin K metabolic pathway genes (*CYP4F2* and *CYP4F11*) and metabolic activity using a genotype- and diplotype-based approaches. Specifically, we investigated the impact of genetic variability in the *CYP4F2*3* background/*CYP4F11* locus and determined the utility of using haplotype, compared to genotype, as a predicator of hepatic intrinsic clearance of vitamin K (VK). To accomplish this, we measured *CYP4F2* and *CYP4F11* mRNA levels, protein abundances and metabolic activities (ω-hydroxy VK1 formation) in matched human liver microsome samples, utilizing a novel in vitro population modeling approach. Our results suggest that accounting for the *CYP4F2*3* allele alone is sufficient to capture most of the genetics-derived variability in metabolic activity observed in vitro. Additionally, our findings highlight the important contribution that CYP4F11 makes towards vitamin K metabolism in the human liver.

3.2 INTRODUCTION

The cytochrome P450 4 (CYP4) enzyme family plays an important role in the metabolism of fatty acids and fat-soluble vitamins (i.e., vitamin D, E, and K),[95] generating in some instances important

signaling molecules, such as arachidonic acid-derived eicosanoids (e.g. prostaglandins, leukotrienes, and thromboxane),[96] which are involved in the regulation of a variety of physiologic processes (i.e., immune response, cardiovascular health, and hemostasis). Importantly, several studies have shown that variation in *CYP4* genes contributes to inter-individual variability in disease susceptibility or altered drug pharmacodynamics. Notable examples include the loss of function *CYP4B1*2* variant allele (rs3215983, D>G) associated with a 1.75-fold increased risk of bladder cancer in Japanese populations[97] due to reduced clearance of pro-carcinogens. Other examples include the *CYP4A11* variant rs1126742 (A>G) allele, which is associated with hypertension in White ancestry populations[98] proposed to be caused by diminished 20-HETE synthesis from arachidonic acid, the *CYP4F3* variant rs4646904 (G>A) allele associated with ulcerative colitis and lung cancer in smokers[99,100], and the *CYP4V2* variant rs13146272 (C>A) allele associated with the risk of deep venous thrombosis and tamoxifen-induced venous thrombosis[101–103]. Of particular interest for this study is the impact of genetic variation at the *CYP4F2* and *CYP4F11* gene locus, and their downstream effects on vitamin K-related pathophysiology and the effects of oral vitamin K antagonists, such as warfarin.

Vitamin K (VK) refers to a group of lipid-soluble molecules recognized for their role in regulating blood coagulation. VK exists in two forms, phylloquinone (VK_1), found predominantly in leafy greens, and a class of molecules called menaquinones (VK_2), found in meats and fermented foods.[1] VK serves as a cofactor for the post-translational γ-carboxylation of glutamic acid residues of Gla proteins, including key clotting factors.[6] In addition to supporting hemostasis, VK and GLA protein production play an active role in cardiovascular health, bone remineralization, renal function, and cell growth[13]. CYP4F2 and CYP4F11 serve as the primary enzymes involved in the

metabolic elimination of VK and have a direct effect on systemic exposure to VK, with CYP4F2 playing the dominant role in the initiation of VK catabolism by ω-hydroxylation[32].

The VK antagonist, warfarin, is a widely used oral anticoagulant with proven efficacy in treating certain thromboembolic disorders, preventing deep vein thrombosis and pulmonary embolism. Despite its efficacy, long-term use of warfarin is associated with morbidity and mortality primarily from major bleeding events that can arise without warning, due to the drug's narrow therapeutic window[19]. Variations in two genes responsible for warfarin pharmacokinetics (*CYP2C9*) and pharmacodynamics (*VKORC1*) have repeatedly been found to be significantly associated with warfarin dose response and explain approximately 15 and 25% of the therapeutic dose variability[24,25], respectively. GWAS studies also implicate a *CYP4F2* functional variant, rs2108622 (C>T, V433M, *CYP4F2*3*), which accounts for 1-7% of the variability in warfarin dose-response [36]. A previous study evaluating genotype-phenotype relationships found a significant association between the rs2108622 TT genotype and lower microsomal CYP4F2 protein concentration, paired with reduced VK_1 oxidation[30,32]. However, this effect may be substrate-dependent. In a follow-up study, although there was a statistically significant positive correlation, microsomal CYP4F2 protein content only explained 13% of the variability seen in VK_2 ω-hydroxylation rate in vitro[32], suggesting that other factors besides CYP4F2 tissue abundance may control VK_2 catabolism (i.e. other pathways of metabolic clearance). A substrate-dependent effect was also observed in a study evaluating the effect of the *CYP4F2*3* variant on 20-HETE production and leukotriene B4 (LTB4) metabolism, where *CYP4F2*3* was associated with a 50% reduction in 20-HETE production but no change in LTB4 ω-hydroxylation[104].

A comprehensive analysis investigating the relationship between genotype and gene expression found that select single nucleotide variants (SNVs) and extended haplotypes in

CYP4F2, *CYP4F11*, and *CYP4F12* were associated with mRNA expression across CYP4F isoforms in human tissues and that this variability was associated with warfarin dose-response. [37] The study observed potentially competing effects of different SNVs within the same gene cluster, resulting in the "canceling out" of independent SNVs effects on mRNA expression and warfarin pharmacodynamics. *CYP4F11* variant rs1060467 was associated with a decrease in the therapeutic warfarin dose; *CYP4F2*3* was associated with an increase in the therapeutic warfarin dose. A haplotype that combines both these variants lacked a significant association.[37] The authors noted that although the overall effect of SNVs in *CYP4F2* and *CYP4F11* on warfarin dose-response variability in their study was very small, populations with different linkage-disequilibrium patterns may yield haplotypes with larger effects.

Taking all of this into account, the objective of this study was to reassess, using a well-characterized human liver bank, the relationship between VK metabolic pathway gene variation and phenotypic markers of VK metabolic activity. We hypothesized that genetic variation at the *CYP4F2* and *CYP4F11* gene locus contributes to interindividual variability in hepatic protein expression and intrinsic clearance of VK_1 in a haplotype/diplotype-dependent manner. Specifically, we sought to determine whether stratification of livers by *CYP4F2/CYP4F11* diplotypes can improve our ability to identify poor and extensive metabolizers of VK_1, compared to single allele genotyping alone.

3.3 Materials and Methods

3.3.1 *Chemicals*

Phylloquinone, lapachol, NADPH, potassium phosphate, EDTA, and methyl-β-cyclodextrin (MβCDX) were purchased from Sigma-Aldrich (St. Louis, MO). Acetonitrile, UPLC-grade water, hexane, iodoacetamide, trypsin, and dithiothreitol were purchased from Thermo Fisher Scientific

(Waltham, MA). Unlabeled and $^{13}C,^{15}N$ isotopically labeled peptides were purchased from New England Peptide (Harvard, MA) and Thermo Fisher (Waltham, MA), respectively.

3.3.2 *CYP4F2 and CYP4F11 DNA Sequencing, RNA Quantification and Variant Selection*

The Pharmacogenetic Research Network (PGRN) next-generation sequencing platform PGRNseqversion1.1 was used to identify *CYP4F2* and *CYP4F11* genomic variation following previously described methods[75,105]. Liver RNA was isolated, purified, and quantified previously using a formerly described and validated method[105].

3.3.3 *Haplotype and Diplotype Identification and Linkage Disequilibrium Analysis*

A total of 32 single nucleotide variants within the *CYP4F2* and *CYP4F11* gene locus were identified. Those with a minor allele frequency of less than 0.05 (n = 19) were excluded from further evaluation and the remainder (n = 13) were included for statistical analysis. Each variant was assessed for Hardy–Weinberg equilibrium (HWE) using a Chi-Square test (χ^2) with threshold p-value set to 0.01 for inclusion in haplotype analysis. Haplotypes and diplotypes were generated for 274 re-sequenced human liver DNA samples using the statistical deconvolution software PHASE 2.1.1[106]; individuals with a haplotype pair probability of <90% were excluded from phenotype association tests. Linkage disequilibrium (LD) was calculated between genetic variants found across *CYP4F2* and *CYP4F11* loci. Pairwise linkage disequilibrium relationships were both calculated and plotted in a correlation matrix using the software Haploview (version 4.2) [107]; LD block and D' values were generated using the default algorithm. A total of 20 unique haplotypes and 50 unique diplotypes were identified in the liver bank population; a summary of variant information for each haplotype and diplotype can be found in Tables 1 and 3.

3.3.4 *Human Liver Microsome (HLM) Preparation*

Liver microsomes were prepared using an adaptation of a previously described method[30]. Human liver tissue (1.6-1.7 g dry weight) was thawed on ice in prechilled buffer (0.25 M sucrose, 1 mM EDTA, 50 mM potassium phosphate, pH 7.4). Fat and connective tissue were removed, and the remaining tissue was cut into fine pieces before homogenization in an Omni Bead Mill Homogenizer (Omni International, Kennesaw GA). The homogenate was fractionated by centrifugation (15,000*g* for 30 min at 4°C), and the supernatant was isolated. The microsomal pellet was collected upon further fractionation by ultracentrifugation (120,000*g* for 70 min at 4°C). The pellet was resuspended in wash buffer (10 mM potassium phosphate, 0.10 mM potassium chloride, 1 mM EDTA), re-homogenized using a Dounce homogenizer, and re-pelleted by ultracentrifugation (120,000*g*, 70 min). The final pellet was resuspended in a minimal volume (200-250 μL) of buffer (0.25 M sucrose, 10 mM EDTA, and 50 mM potassium phosphate, pH 7.4) and stored at -80°C. A bicinchoninic acid assay (BCA) was used to determined total microsomal protein content[108].

3.3.5 *Quantification of Hepatic CYP4F2 and CYP4F11*

The abundance of hepatic CYP4F2 and CYP4F11 in liver microsomes was quantified using an adaption of a previously validated LC-MS/MS-based protocol[32]. Surrogate peptides for CYP4F proteins were selected based on *in silico* criteria[32]. Peptides with single-nucleotide polymorphisms (SNPs), or those susceptible to degradation were not selected. Peptides with contiguous arginine and lysine sequences (RR, RK, KR, and KK) were also excluded from the peptide selection pool. A genome wide BLAST search was performed to identify specific peptides for human CYP4F2 and CYP4F11. Unique peptide sequences for CYP4F2 (SVINASAAIAPK) and CYP4F11 (TLTQLVTTYPQGFK) and stable-isotope-labeled (SIL) forms of these peptides (labeled lysine

residue; two nitrogen isotopes ^{15}N and six isotopically labelled ^{13}C atoms) were used as internal standards for CYP4F2 and CYP4F11, respectively. Human liver microsomes (40 µL, 2 mg protein/ml) were incubated with 4 µL of dithiothreitol (100 mM) and 40 µL of ammonium bicarbonate buffer (100 mM, pH 7.8) for 5 min at 95 °C (denaturation and reduction). Samples were cooled to room temperature and cysteines alkylated by addition of 4 µL of iodoacetamide (200 mM) for 20 min at room temperature in the dark. Ten microliters of a trypsin solution were added (0.16 mg/mL), and samples were digested at 37 °C for 24 h. At the end of the digestion period, the reaction was quenched by adding 30 µL of a 50 nM stable-isotope-labeled (SIL) internal standard solution (contain both CYP4F2 and CYP4F11 peptides in a 50% acetonitrile, 0.1% formic acid, v/v), and the mixture centrifuged at 5,000g for 5 min at 4 °C and stored at −80 °C until analysis. Samples were analyzed using electrospray ionization (positive) under multiple reaction monitoring (MRM) conditions using a Xevo TQ-XS triple-quadrupole mass spectrometer coupled to an Acquity liquid chromatography system (Waters®). Ten microliters of sample were injected onto a C18 1.7 µm, 100 mm ×2.0 mm UPLC column (Waters®).

LC conditions were as follows: the mobile phase consisted of 0.1% formic acid (v/v) in water (A) and acetonitrile (B), at a flow rate of 0.3 mL/min. The mobile phase gradient was as follows: 0−2 min (hold 10% B, v/v), 2-4 min (10−90% B, v/v), 4−6 min (hold 90% B, v/v), 6− 7.5 min (90−10% B, v/v), 7.5−8 min (10% B, v/v). MRM mass transitions that were monitored and the peptide retention times are listed, with *m/z* fragments in brackets: CYP4F2 retention time = 3.71 min, *m/z* 571 > [499.3, 657.4, 728.4]; CYP4F11 retention time = 3.83 min, *m/z* 789 > [576.4, 941, 1040.6] (parent z = 2, fragment z = 1); CYP4F2 SIL retention time = 3.68 min, *m/z* 575.3 > [507.3, 665.4, 736.4]; CYP4F11 SIL retention time = 3.85 min, *m/z* 802.9 > [584.4, 949.6, 1048.6] (parent z = 2, fragment z = 1). Instrument settings for CYP4F2 and CYP4F11 peptides were as

follows, respectively: dwell time, 10 ms for both; cone, 30 V; collision energy, 20 eV. Protein abundance was quantified by calculating peak area ratios (PAR) of unlabeled and labeled (IS) peptide fragment transitions and interpolating concentrations against a linear standard calibration curve. Standard curve was constructed using the PAR of an unlabeled peptide standard to a labeled peptide standard for both CYP4F2 and CYP4F11 with a range of 0.10-100 pmol/mL (R^2>0.99). Quantified samples were normalized to total microsomal protein content per mL of microsomal solution for final units of (pmol CYP) / (mg microsomal protein).

3.3.6 *Microsomal Activity Assay: ω-Hydroxy-Phylloquinone Formation*

Catalytic activity for combined CYP4F2 and 4F11 was determined using selected human liver microsomes (HLMs) and monitoring the formation of ω-hydroxy-phylloquinone over time. Incubations were performed in 0.5 mL of 0.1 M KPi buffer (pH 7.4) with 1 mM EDTA, 1 mM MβCDX, 1-2 mg/mL HLM, and 1 mM NADPH. Substrate (phylloquinone, VK_1) was dissolved in 2-propanol and added to buffer and microsomes to achieve the appropriate nominal concentrations, with a final organic solvent concentration <0.7%. All incubations were performed in duplicate. Samples were pre-incubated in a water bath for 3 min at 37°C and 80 RPM and reactions were initiated by the addition of NADPH. After a 30 min incubation, reactions were quenched by the addition of 2 ml of chilled hexanes followed by 50 μL of internal standard (1 μM, lapachol). Reaction mixtures were extracted (LLE) with hexanes (2 x 2 mL). The hexane layer was evaporated under N_2-gas and the residue reconstituted in 200 μL of 50:50 water and 2-propanol for analysis.

3.3.7 *UPLC-MS/MS Analysis of ω-Hydroxy-Phylloquinone*

Production of ω-hydroxy VK_1 in human liver microsomes was quantified by ultra-performance liquid chromatography tandem mass spectrometry (UPLC-MS/MS) using previously published multiple reaction monitoring conditions[32] (**Figure 3.1**). In summary, liquid-chromatography-mass spectrometry analyses were conducted on ACQUITY Ultra Performance LC™ (UPLC) system with integrated Waters autoinjector coupled to a Waters Xevo TQ-XS tandem quadrupole mass spectrometer (Waters Co., Milford MA.). MS/MS analysis was performed under atmospheric pressure negative chemical ionization ($APCI^-$) with multiple reaction monitoring (MRM). Instrument source temperature was set at 150 °C with a probe temperature set to 650 °C. The following mass transitions (*m/z*) for each analyte were monitored on separate ion channels: *m/z* 466.3 >185 (ω-hydroxy VK_1); *m/z* 241>186 (lapachol). The cone voltages were set at 58 V and 40 V for ω-hydroxy VK_1 and lapachol respectively. Collision energies were set to 30 V and 20 V for ω-hydroxy VK_1 and lapachol, respectively.

Analytes were separated on an Acquity UPLC C18 1.7 μm, 100 × 2.0 mm column (Waters) with water (A) and methanol (B) mobile phase. The mobile phase gradient ran as follows: solvent B held at 40% for 0-1 min, ramped up to 98% between 1 and 3.5 min, held at 98% B from 3.5 to 8.0 min, ramped down to 40% B from 8 to 8.5 min, held at 40% B from 8.5-9.5 mins. Retention times for ω-hydroxy VK_1 and lapachol were 5.27 min and 3.92 min, respectively. ω-Hydroxy VK_1 identity was confirmed using high-resolution mass spectrometry using an AB Sciex 5600 Q-TOF LC-MS/MS system (data not shown), and its formation characterized in human liver microsomes using the probe CYP4F inhibitor HET0016 (**Figure 3.2**). No authentic standard was available and thus data are reported as peak area ratios (PAR) for the metabolite and internal standard. Product formation linearity over time (0-40 mins) was confirmed at low and high substrate concentrations

prior to setting the VK_1 incubation conditions. Analyte detection proportionality under the LC-MS/MS conditions was confirmed across a 250-fold serial dilution of the maximum amount of ω-hydroxy VK_1 formed during the 30 min HLM incubation period. This encompassed all of PAR values generated from individual HLM incubations.

3.3.8 *In-Vitro Study Design*

In-vitro incubation concentrations were selected using a strategic sparse sampling approach where 2 sets of 3 non-overlapping substrate concentrations spanning the range of the Michaelis-Menten profile (Set 1: 0.5, 5, and 50 μM, Set 2: 1, 10, and 100 μM VK_1). For each genotype/diplotype group, half of the subjects were assigned to set 1 and the other half to set 2. Using an optimal design estimator (PopED®) we confirmed in preliminary data from a sparse sampling design (n = 3 pts per subject) that sampling efficiency ratio at the proposed substrate concentrations relative to rich a sampling design (n = 6-9 pts per subject) is estimated to be between 1-1.2, demonstrating equivalency and comparable relative standard error for both set 1 and set 2 concentrations using D-optimality criterion (data not shown).

3.3.9 *Statistical Analysis*

Statistical analysis was performed using R version 4.0 via RStudio. For liver bank mRNA analysis, *CYP4F11* mRNA abundance (fragments per kilobase per million bases read, FPKM) were square root transformed to obtain normality; *CYP4F2* mRNA expression data were normally distributed without transformation (**Figure 3.3**). mRNA data were further normalized to the median expression from the individual genes. To evaluate the association between genotype and mRNA expression a one-way analysis of variance (ANOVA) model was generated with a significance threshold set to a p-value of 0.05 and expression levels were compared across alleles; post hoc

analysis was conducted using a Tukey HSD test. Parametric test assumptions of normality and homogeneity of variance were assessed using a Shapiro-Wilk test and Bartlett test, respectively. Wilcox/Mann-Whitney test was used to compare mean mRNA expression by sex. Genotype and diplotype groups containing only 1 individual were excluded from mRNA statistical analysis. Linear regression models were used to evaluate additive inherited allele effects and covariate interactions on parameter estimates. Coding of individual genotypes for additive inherited allele effects was as follows: WT = 0, heterozygotes = 1, homozygous variants = 2. Coding for haplotypes and diplotype was as follows: 0 = zero copies of specified genotype, 1 = one copy of specified genotype, 2 = two copies of specified genotype. VK activity assay data were evaluated using non-linear mixed effect modeling (see Population Michaelis-Menten Modeling section described below). Age, sex, CYP4F2/4F11 protein and cytochrome P450 reductase abundances data were evaluated as additional covariates during the model optimization steps. Diplotype groups with n<2 (6 /17, 1 /12, 17 /17, 4 /13, and 12 /12) are reported in summary tables but excluded from statistical analysis. Estimates in the text are reported as mean ± standard deviation unless otherwise stated. Post-hoc analysis of diplotype data was conducted using Dunnett's test for multiple comparisons using diplotype 13 /13 as the reference group for comparison. The intrinsic clearance estimates adjusted for CYP4F2 and CYP4F11 protein content were evaluated using multiple linear regression; reported p-values were adjusted for false discovery rate using the Benjamin-Hochberg correction method[109].

3.3.10 *Population Michaelis-Menten Modeling*

Population Michaelis-Menten Modeling (PopMM) was performed in R (v4.2.0) using the non-linear mixed effect modeling package NLME [83] and supporting packages for data manipulation

and data visualization. The final model was built by fitting a base Michaelis-Menten model with a random-effect structure and covariates. The model was derived from the Michaelis-Menten equation with modifications to V_{max} and K_m to incorporate both fixed- and random-effects. Product formation rate for the i^{th} individual at the j^{th} substrate incubation concentration, defined as (V_{ij}), was modeled as a function of a Φ_i and substrate concentration (S_i) plus residual error (ε_{ij}); where Φ_i is a matrix of parameters (Φ_{1i} and Φ_{2i}) each being vectors of fixed and random effects whose sum equates to the predicted individual-level estimates of V_{max} and K_m, respectively (**Equation 9**). Individual estimates of V_{max} (Φ_{1i}) and K_m (Φ_{2i}) were defined as natural log of the sum of a typical population value for V_{max} ($\theta_{1,\,Vmax}$) or K_m ($\theta_{1,\,Km}$) and a fixed covariate effect ($\theta_{2,\,Vmax}$ or $\theta_{2,\,Km}$) based on individual *CYP4F2* genotype or diplotype (X_i), plus an individual specific random effect on both V_{max} and K_m (η_{1i}, and η_{2i}, respectively) (**Equation 10**). Model parameters (θ, X, and η) were incorporated using an exponential variation model to account for non-normal distribution of population estimates. Individual level random effects (η_i) and residual error (ε_{ij}) were assumed to be normally distributed with a mean of 0 and ε_{ij} having a variance of σ^2. V_{max} and K_m were assumed to be uncorrelated (independent), therefore variance (ψ_i) for η_i was structured as a diagonal matrix with covariance set to 0.

Equation 9

$$V_{ij} = f(\Phi_i, S_i) = \left(\frac{\Phi_{1i} \times S_i}{\Phi_{2i} + S_i}\right) + \varepsilon_{ij}$$

Equation 10

$$\Phi_i = \begin{bmatrix}\Phi_{1i}\\ \Phi_{2i}\end{bmatrix} = \left(\begin{bmatrix}\theta_{1,V_{max}}\\ \theta_{1,K_m}\end{bmatrix} + \begin{bmatrix}\theta_{2,V_{max}}\\ \theta_{2,K_m}\end{bmatrix} \times X_i\right) \times e^{\begin{bmatrix}\eta_{1i}\\ \eta_{2i}\end{bmatrix}}$$

or

$$ln(\Phi_i) = \begin{bmatrix}\Phi_{1i}\\ \Phi_{2i}\end{bmatrix} = ln\left(\begin{bmatrix}\theta_{1,V_{max}}\\ \theta_{1,K_m}\end{bmatrix} + \begin{bmatrix}\theta_{2,V_{max}}\\ \theta_{2,K_m}\end{bmatrix} \times X_i\right) + \begin{bmatrix}\eta_{1i}\\ \eta_{2i}\end{bmatrix}$$

where,

$$\eta_i = \begin{bmatrix}\eta_{1i}\\ \eta_{2i}\end{bmatrix} \sim N(0, \begin{bmatrix}\psi_1 & 0\\ 0 & \psi_2\end{bmatrix}), \qquad \varepsilon_{ij} \sim N(0, \sigma^2)$$

Intrinsic clearance (CL_{int}) estimates were calculated as the quotient of V_{max}/K_m, with population level intrinsic clearance defined as ($\theta_{1,\,Vmax}$)/ ($\theta_{1,\,Km}$), and genotype (or diplotype) level defined as the sum of $\theta_{1,Vmax}$ and $\theta_{2,\,Vmax}$ for a given genotype divided by the sum of $\theta_{1,Km}$ and $\theta_{2,\,Km}$ for a given genotype; ($\theta_{1,\,Vmax\,+\,}\theta_{2,\,Vmax,\,Genotype}$)/($\theta_{1,\,Km\,+\,}\theta_{2,\,Km,\,Genotype}$). Selection of the appropriate covariates and error model were based on Akaike criterion (AIC), *F*-tests (ANOVA) and diagnostic plots. The final model was weighted using a power residual error structure based on optimization of AIC and BIC and log-likelihood criteria. The impact of genotype (and diplotype) on intrinsic clearance was evaluated using multiple regression of the PopMM derived intrinsic clearance estimates; with intrinsic clearance defined as Φ_1 / Φ_2 as described in **Equation 11**. For this analysis, intrinsic clearance estimates were natural log-transformed to correct for non-normal distribution. The reference groups (intercept) for the genotype and diplotype multiple regression models were *CYP4F2*1/*1* and *CYP4F* diplotype *13 / 13*, respectively.

Equation 11

$$\beta_{0\,Cl_{int}} = \frac{\Phi_1}{\Phi_2}$$

$$Cl_{int} = \beta_{0\,CLint} + \beta_{1\,CL_{int}} \times X_{Genotype\,CL_{int}} + \varepsilon$$

3.3.11 *Proteomic Based Prediction of In Vitro Intrinsic Clearance*

Vitamin K_1 intrinsic clearance (CL_{int}) and ω-hydroxy vitamin K_1 formation rate (v) for the observed data were predicted based on CYP4F2 and CYP4F11 protein abundance using model-derived estimates of V_{max} and K_m. An estimate of catalytic turnover (k_{cat}) was calculated by dividing model-derived V_{max} values for each *CYP4F2*3* genotype by the total CYP4F (CYP4F2 + CY4F11) protein content (for that specific genotype); individual predictions were made using the following: substrate concentration used in the experiment, individual CYP4F2 protein abundance, individual CYP4F11 protein abundance, genotype-dependent catalytic turnover value ($k_{cat,\ Genotype}$), and a genotype-dependent Michaelis-Menten constant ($K_{m,\ Genotype}$) as independent variables as shown in **Equation 12 and Equation 13**. A linear regression model was fit to the predicted vs. observed data and used to evaluate the predictive value of experimentally determined CYP4F protein abundance data towards metabolic activity.

Equation 12

$$Predicted\ Rate\ (v_i) = \frac{(k_{cat,Genotype} \times [E_{CYP4F2,i} + E_{CYP4F11,i}]) \times [S]}{K_{m,Genotype} + [S]}$$

Equation 13

$$Predicted\ Clearance\ (CL_{int,i}) = \frac{V_{max,i}}{K_{m,i}} = \frac{(k_{cat,Genotype} \times [E_{CYP4F2,i} + E_{CYP4F11,i}])}{K_{m,Genotype}}$$

3.4 RESULTS

3.4.1 *Study Population*

Human liver tissue samples (n = 274) were obtained from two liver banks: University of Washington Human Liver Bank, Seattle, WA (n = 58) and St Jude Liver Resource at the St. Jude Children's Research Hospital in Memphis, TN (n = 216)[73]. Of these 274 subjects, demographic information was available at the following percentages: Age (77%), n = 213; sex (100%, n = 274), 58% male and 41% female. For subjects included in the proteomic and activity analysis (n = 88) demographic data was available at the following percentages: Ages (84%), with mean and standard deviation of 39 ± 20 years; sex (100%) with 59% male and 41% female; ethnicity (100%) with Caucasian/White = 96%, African American/Black = 1%, Asian = 1%, and Hispanic/Latino = 2%. Institutional review boards at both sites approved the collection and use of these samples for research purposes and all links between archived tissues and donors were previously destroyed.

3.4.2 *CYP4F2 and CYP4F11 SNV Identification*

A total of 13 common gene variants (5 missense, 5 synonymous, and 3 intron) in the *CYP4F2-CYP4F11* locus were detected in the population (**Figure 3.25**). Of these, 9 variants were found in the *CYP4F2* gene (rs2108622, rs2074900, rs3093160, rs3093153, rs3093114, rs3093106, rs3093105, rs3093103, rs3093100) and 4 in the *CYP4F11* gene (rs1060463, rs8104361, rs3765070, rs2305801). For *CYP4F2*, rs2108622 (*CYP4F2*3*) was the most prominent non-synonymous variant with a MAF = 0.26 and for *CYP4F11* rs1060463 was the most prominent non-synonymous variant with a MAF = 0.41. Details on all the variants characterized can be found in **Table 3.1**.

3.4.3 *Evaluation of CYP4F2 and CYP4F11 Hepatic mRNA Abundance*

CYP4F2 and *CYP4F11* mRNA abundance in the combined St. Jude and University of Washington Liver Banks (n = 274) was evaluated. Overall, mean *CYP4F2* mRNA expression was higher than *CYP4F11* expression (48 ± 20 FPKM vs 23 ± 28 FPKM, respectively). A summary table of the population characteristics stratified by *CYP4F2*3* genotype can be found in **Table 3.2**. Associations between SNVs found in the *CYP4F2* and *CYP4F11* genes and hepatic mRNA abundances were investigated. There were no statistically significant associations between *CYP4F2* mRNA abundance and *CYP4F2* variants (**Figure 3.4**), including the *CYP4F2*2* and *CYP4F2*3* alleles. The same was true for variants in the *CYP4F11* gene and its mRNA abundance (**Figure 3.5**). However, rs3093153 (missense mutation found in the *CYP4F2* loci) showed a statistically significant association with *CYP4F11* mRNA expression (ANOVA $p = 0.014$) (**Figure 3.6**); post hoc analysis revealed higher hepatic *CYP4F11* mRNA abundance when comparing heterozygous variant to homozygous reference allele (8.1% increase, Tukey HSD $p = 0.036$); additive inheritance effects could not be evaluated due to the low sample size in the homozygous variant group (n = 1). There were no allele dependent effects observed between *CYP4F11* variants (rs1060463, rs2305801, rs3765070, and rs8104361) and *CYP4F2* mRNA abundance (**Figure 3.7**).

3.4.4 *Impact of Age and Sex on Hepatic mRNA Expression*

Sex was identified as a significant covariate with females having modestly higher hepatic *CYP4F2* mRNA abundance than males with normalized FPKM values of 1.08 ± 0.41 and 0.94 ± 0.41, respectively (Wilcox $p = 0.01$); and higher *CYP4F11* mRNA abundance than males at 1.02 ± 0.17 vs. 1.00 ± 0.17, respectively (Wilcox $p = 0.05$) as shown in **Figure 3.8**. Additionally, age was also evaluated for its association with hepatic *CYP4F2* and *CYP4F11* mRNA abundance. Available age data (n = 212) was binned into 3 age categories: <18 years, 19-55 years, and 55+ years then

normalized mRNA abundance was compared by age group (**Figure 3.9**). There was a statistically significant difference in *CYP4F2* mRNA abundance between the <18 years and the 19-55 years age categories (23% lower, Tukey p=0.01), as well as the <18 years and 55+ years age category (23% lower, Tukey p=0.01). Similarly, there was a statistically significant difference in *CYP4F11* mRNA abundance between the <18 years and the 19-55 years age categories (9.5% lower, Tukey p=0.01) as well as the <18 years and 55+ years age category (8.9% lower, Tukey p=0.02). There was no statistically significant difference between 19-55 years and 55+ years age categories for both *CYP4F2* and *CYP4F11*. Additionally, the combined impact of sex and age on mRNA abundance was evaluated using multiple linear regression. There were no statistically significant interactions between sex and age when used as predictors of *CYP4F2* or *CYP4F11* mRNA abundance (*CYP4F2*, ANOVA p = 0.29; *CYP4F11*, ANOVA p = 0.34).

3.4.5 *Haplotype-Mapping of CYP4F2 and CYP4F11 Loci*

Linkage disequilibrium (LD) analysis was conducted to determine associations between SNVs across *CYP4F2* and *CYP4F11* genes. Genomic crosstalk between *CYP4F2* and *CYP4F11* loci was explored by generating a single haplotype block spanning both regions yielding 20 unique haplotypes incorporating the 13 previously mentioned SNVs. Strong associations were present between all SNVs across the locus; a visual representation can be found in **Figure 3.10**. The population frequencies and associations between *CYP4F2/4F11* haplotypes are summarized in **Table 3.3**. For our investigation, we assigned haplotype 13 as the reference *CYP4F* gene locus sequence based on its frequency in the population (31%, highest observed).

3.4.6 *Hepatic CYP4F2 and CYP4F11 Proteomic Analysis*

Individual human liver microsome donors (n = 88) were evaluated for CYP4F2 and CYP4F11 protein abundance using quantitative proteomics (**Figure 3.11, and Table 3.4**). Overall, the fractional contribution of CYP4F2 was greater than CYP4F11 across most of the donors (n=84) with the exception of 4 subjects (HL-149, SJLB335, SJLB174, and SJLB767); all of whom were *CYP4F2*3* carriers and had CYP4F11 protein abundance accounting for 52-59% of the combined CYP4F protein pool. *CYP4F2*1* homozygotes had the greatest CYP4F2 contribution accounting for 83% of the combined CYP4F protein pool, followed by *CYP4F2*3* heterozygotes at 77%, and *CYP4F2*3* homozygotes at 61% fractional abundance on average (**Figure 3.12**). Additionally, when stratified by diplotype, carriers of haplotypes 4 and 6 (*1/*1 carrier) had the highest fractional CYP4F2 contribution accounting for 82-87% of the combined CYP4F protein pool; however total protein content varied (**Figure 3.13**). Diplotype 6 /13 had the highest individual and combined CYP4F protein abundance out of all the groups analyzed, with a 40% higher combined protein abundance (21.73 ± 5.59 pmol/mg microsomal protein) compared to the reference group 13 /13 (15.31 ± 6.4 pmol/mg microsomal protein). Alternatively, Diplotype 20 / 20 (*3/*3 carrier) had the lowest combined CYP4F protein abundance (4.97 ± 2.47 pmol/mg microsomal protein). Both CYP4F2 and CYP4F11 microsomal protein abundances were evaluated by age group (<18 years, 19-55 years, and 55+ years) as shown in **Figure 3.14**. Overall, there were no statistically significant differences in mean protein abundance for CYP4F2 and CYP4F11 across age groups (ANOVA, p = 0.11 and 0.16, respectively), nor was there interaction between sex and age on protein abundance (CYP4F2, ANOVA p=0.57; CYP4F11, ANOVA p = 0.94).

3.4.7 *Population Michaelis-Menten Analysis*

Individual and population Michaelis-Menten parameters were estimated using a previously described non-linear mixed effect modeling method in a sub-population of individuals from the UW and St Jude Liver Banks (n = 88; selected for coverage of the common haplotypes). ω-Hydroxy VK_1 formation over time was monitored across a range of substrate concentrations using a sparse sampling approach (**Figure 3.15**). After fitting a base model incorporating random effects on V_{max} and K_m, the typical population values were 0.14 PAR/min/mg protein, 95% CI [0.12-0.18] and 4.22 μM, 95% CI [3.6-4.95]. A power residual error structure that provided the greatest improvement in objective values and corrected the heteroscedastic nature of the observed data was selected to characterize residual variability. A summary of model diagnostics including residual error plots, individual, and population predicted vs. observed values can be found in **Figure 3.16**. Covariate selection was based on random-effects correlations for both V_{max} and K_m (**Figure 3.17** and **Figure 3.18**)**.** Two additional models, a *CYP4F2* genotype covariate model and a *CYP4F* diplotype covariate model, were also evaluated for goodness of fit compared to the base model (**Table 3.5**). Incorporating *CYP4F2*3* genotype as a fixed effect on V_{max} and K_m improved the model criterion (ΔAIC = -25.77, ANOVA p <0.0001), and yielded the following respective V_{max} and K_m estimates: *CYP4F2*1/*1:* 0.17 PAR/min/mg protein, 95% CI [0.13-0.23] and 4.25 μM, 95% CI [3.5-5.15]; *CYP4F2*1/*3*, 0.17 PAR/min/mg protein, 95% CI [0.1-0.28] and 4.17 μM, 95% CI [2.48-7]; *CYP4F2*3/*3*, 0.07 PAR/min/mg protein, 95% CI [0.04-0.13] and 8.66 μM, 95% CI [5.5-13.65]. Importantly, *CYP4F2*3/*3* was associated with a significantly lower V_{max} (-57%, p=0.004) and higher K_m (+103%, p=0.002). There were no significant differences in either V_{max} or K_m for the *CYP4F2*1/*3* heterozygous variant group compared to the *CYP4F2*1/*1* reference group.

The third model evaluated used *CYP4F* diplotype as an explanatory covariate for deviations in V_{max} estimates. A total of 50 unique diplotype groups were identified in the liver bank population, of which 21 were characterized *in vitro*. Incorporation of *CYP4F* diplotype as a fixed-effect on V_{max} while maintaining *CYP4F2*3* genotype as a fixed-effect on K_m improved model criterion compared to the base model (ΔAIC = -19.72, ANOVA $p<0.0001$), however, it did not provide greater improvement in objective values compared to the incorporation of *CYP4F2*3* genotype as a covariate alone ($AIC_{Genotype\ model}$ = -923.80, $AIC_{Diplotype\ Model}$= -917.83, ANOVA $p=0.038$ in favor of genotype model). Of the diplotypes analyzed, there were 2 groups (20 / 20, and 6 / 13) with statistically significant differences in Michaelis-Menten parameters relative to the reference group (13 / 13). Diplotype 20 / 20 was associated with a 62% lower V_{max} compared to the reference group at 0.07 PAR/min/mg protein, 95% CI [0.03-0.14] and 0.17 PAR/min/mg protein, 95% CI [0.11-0.27], respectively, while diplotype 6 / 13 was associated with 181% higher V_{max} compared to the reference group (0.49 PAR/min/mg protein, 95% CI [0.21-1.15] and 0.17 PAR/min/mg protein, 95% CI [0.11-0.27], respectively). The magnitude and significance of the impact that *CYP4F2*3* genotype had on K_m remained constant after adjusting for diplotype effects.

To further assess the impact that the *CYP4F2*3* variant has on metabolic activity, the genotype based PopMM model was adjusted to account for differential protein expression, by incorporating both CYP4F2 and CYP4F11 protein abundance as covariates affecting V_{max}. After adjusting the model for protein content, both CYP4F2 and CYP4F11 protein abundance were found to be a significant predictor of V_{max}, however incorporation of both proteins' abundances as predictors did not significantly improve model performance compared to incorporation of CYP4F2 abundance alone; AIC = -734.47 (CYP4F2 + CYP4F11) vs AIC -732.90 (CYP4F2 only). Additionally, after protein abundance adjustment, the *CYP4F2*3* variant no longer posed as a

significant predictor of V_{max}, suggesting that variability in V_{max} can be explained by variability in CYP4F2 abundance rather than variability in catalytic turnover (k_{cat}). Adjusting for protein content had no impact on the significant *CYP4F2*3* related changes in K_m estimates. Additional models were evaluated and compared to identify optimal covariates of response. Adding CYP4F2 protein abundance and cytochrome P450 reductase mRNA abundance as covariates for V_{max} while maintaining an allele dependent *CYP4F2*3* effect on K_m provided the best performing model for the observed data; AIC = -742.87.

3.4.8 *Evaluation of In Vitro Hepatic Intrinsic Clearance*

Population estimates of intrinsic clearance generated using the PopMM model were bimodal and non-normally distributed. Multiple linear regression was performed on the natural log-transformed estimates of intrinsic clearance which corrected for the assumption of normality needed for the linear regression model (Shapiro-Wilks, p = 0.367 for natural log-transformed data). *CYP4F2*3/*3* genotype was associated with a 65% lower intrinsic clearance compared to the reference *CYP4F2*1/*1* genotype, however, the intrinsic clearance of the *CYP4F2*1/*3* genotype group was not different from that of the *CYP4F2*1/*1* group. Of the *CYP4F* diplotypes analyzed, 3 groups were associated with statistically significant changes in intrinsic clearance: diplotypes compared to the reference diplotype group: 6 / 13, 17 / 20, and 20 / 20. Diplotype 17 /20 and 20 / 20 (both *CYP4F2*3/*3* carriers) were associated with a 63% and 68% lower hepatic intrinsic clearance, respectively. In contrast, diplotype 6 / 13 was associated with a 175% higher hepatic intrinsic clearance compared to reference diplotype 13 / 13 (**Table 3.6, and Figure 3.19**). Although the genotype-based multiple regression model provided better performance than the diplotype-based model, the difference was marginal (ΔAIC = -8.12, Δ Residual Sum of Squares = 19.55, *p = 0.167*).

Additionally, model derived estimates of *in vitro* activity were predicted and compared to observed data by evaluating ω-hydroxy VK_1 formation rate (v) and VK_1 intrinsic clearance (CL_{int}) based on CYP4F2 and CYP4F11 protein abundance. Overall, incorporating CYP4F2 and CYP4F11 protein abundance data in conjunction with model-derived estimates of k_{cat} and K_m, 49% of the variability in experimentally observed reaction rates and 44% of the variability in experimentally observed CL_{int} values (v/[S], where S << Km) could be explained (**Figure 3.20**). Finally, to investigate the impact that the *CYP4F2*3* and haplotype 6 have on CL_{int} and catalytic efficiency (k_{cat}/K_m), multiple linear regression analysis was performed. To correct for the non-normal distribution of CL_{int}, values were log-transformed prior to regression. The impact of genotype on catalytic efficiency was determined by adjusting the regression model to account for both CYP4F2 and CYP4F11 protein abundances on CL_{int} estimates by including both proteins as covariates. Unadjusted and untransformed data were also provided as a comparison (**Figure 3.21**); objective function and goodness of fit values (AIC, log-likelihood, R^2) are provided for model comparison purposes in **Table 3.7 and Table 3.8**. Overall, the *CYP4F2*3* variant was no longer a significant predictor of metabolic activity when variability in CYP4F2 and CYP4F11 protein abundance was accounted for (**Figure 3.21 D**). Similarly, for haplotype 6, when log transformed CL_{int} for diplotypes 13 /13 (reference), 6 / 13 (heterozygous variant haplotype) and 6 / 6 (homozygous variant haplotype) were analyzed, diplotype 6 /13 was no longer a significant predictor of response after accounting for CYP4F2 and CYP4F11 protein abundance (**Figure 3.21 H**).

3.4.9 *Correlation between mRNA, Protein, and Metabolic Activity*

To identify possible mechanistic links between *CYP4F2* and *CYP4F11* mRNA abundances, CYP4F2 and CYP4F11 protein abundances, and VK_1 ω-hydroxylation activity, Pearson

correlations were estimated and compared. While *CYP4F2* mRNA abundance was not correlated with CYP4F2 protein abundance ($R = 0.19$, $p = 0.078$), there were significant positive correlations (Pearson's $R > 0.3$, $p < 0.005$) between *CYP4F11* mRNA abundance and CYP4F11 protein levels. Both CYP4F2 and CYP4F11 protein abundances were strongly correlated with ω-hydroxy VK_1 intrinsic clearance, as shown in **Figure 3.22**. There were no observable differences in *CYP4F2* mRNA and *CYP4F11* abundance across *CYP4F2*3* genotypes (Kruskal-Wallis, $p = 0.81$ and 0.59 for *CYP4F2* and *CYP4F11* mRNA, respectively). In contrast, a significant allele dependent lower CYP4F2 protein abundance was observed for *CYP4F2*3* carriers, with heterozygous carriers (**1/*3*) associated with a 19.4% lower abundance (10.3 pmol/mg microsomal protein), and homozygous variant carriers (*3/*3) associated with a 73% lower abundance (**Figure 3.23**). However, when total CYP4F2+CYP4F11 protein abundance was accounted for there was no statistically significant difference in abundance between **1/*1* and **1/*3* carriers, and an overall 63% lower abundance between **1/*1* and **3/*3* (**Figure 3.24**).

3.5 DISCUSSION

In this study we assessed the relationship between common variation in the *CYP4F2* and *CYP4F11* genes, protein abundances and metabolic activity towards VK_1 (phylloquinone)**.** To do so, we first characterized variability across *CYP4F2* and *CYP4F11* gene loci and identified 13 genetic variants (9 *CYP4F2*, and 4 *CYP4F11*) with minor allele frequencies greater than 5%. We then performed haplotype mapping analysis across the *CYP4F2* and *CYP4F11* locus, which revealed 50 unique *CYP4F* diplotypes made up of combinations of 20 unique *CYP4F* haplotypes. We then evaluated correlations between mRNA, protein, and metabolic activity using a genotype and diplotype based approach. We hypothesized that stratification of livers by *CYP4F* haplotypes would yield unique and stronger associations than single allele genotypes. However, our data suggests that *CYP4F2*3*

genotype alone is sufficient to account for most of the genetic associated variability observed in VK_1 metabolism. The ω-hydroxy VK_1 intrinsic formation clearance of diplotypes 17 / 20 and 20 / 20 were similarly lower (63 and 68%) than that of the reference diplotype (13 / 13); this is a result of a combination of both a decrease in CYP4F2 protein expression and an increase in apparent K_m. Diplotype 17 / 20 and 20 / 20 are both *CYP4F2*3/*3* containing groups, with diplotype 17 / 20 having a single copy of 8 additional *CYP4F2/4F11* genetic variants on the **3* background (rs3093160, rs3093114, rs3093106 (**2*), rs3093105, rs3093103, rs3093100, rs1060463, and rs3765070), while diplotype 20 /20 having two copies of the additional 8 variants on the **3* background.

The abundance and activity of CYP4F2 protein containing both W12G (**2*) and V433M (**3*) amino acid substitutions was not different than protein with only V433M, implying that M433 causes the change in enzyme abundance and function, which is in agreement with experimental results with recombinantly expressed reference and variant proteins[30]. Hepatic *CYP4F2* and *CYP4F11* mRNA abundances were found to be poor predictors of metabolic activity. Additionally, contrary to a previous report[37] we found no significant association between rs2108622 (*CYP4F2*3*) and hepatic *CYP4F2* or *CYP4F11* mRNA abundance. However, in our analysis we did identify donors <18 years of age had lower *CYP4F2* and *CYP4F11* mRNA abundance compared to adults. While the underlying mechanism of this observation in our system is unclear, there is evidence of postnatal hepatic CYP4F maturation in mouse models demonstrating a similar effect[110].

Interestingly, our study did reveal a unique haplotype-phenotype relationship. Of the 20 unique haplotypes found in the sample population, 6 exist commonly at a frequency >5% (Haplotypes 1, 6, 12, 13, 17, and 20). A total of 5 out of 6 of these haplotypes fall under previously

established *CYP4F2* star allele nomenclature as noted by the Pharmacogene Variation Consortium (PharmVar) under the following definitions: Haplotype 1 and 13 = *CYP4F2*1,* Haplotype 12 = *CYP4F2*6,* Haplotype 17 = *CYP4F2*3,* and Haplotype 20 = *CYP4F2*4.* However, a novel *CYP4F11* variant combination (Haplotype 6) was identified and shown to exhibit differential protein abundance and activity compared to the reference group. The intrinsic clearance of diplotype 6 / 13 was 175% higher than that of the population reference diplotype (13 / 13). CYP4F2 Protein abundance in diplotype 6 / 13 was also 45% higher than that of the reference (18.07 ± 4.86 vs 12.44 ± 5.24 pmol/mg microsomal protein, respectively) and 28% higher for CYP4F11 than the reference diplotype (3.66 ± 0.87 vs. 2.87 ± 1.59 pmol/mg microsomal protein, respectively). However, these differences were not observed for the homozygous diplotype 6 / 6 group. Individuals in diplotype 6 / 13 group are carriers of 2 *CYP4F11* missense mutations and 3 synonymous mutations (rs1060463, rs8104361, rs3765070, rs2305801, and rs2074900, respectively). How these changes might affect CYP4F2 abundance is unclear, but possibly by some cis-acting process.

Curiously, there was lack of a *CYP4F2*3* allele effect on metabolic activity for heterozygous carriers (**1/*3*) compared to homozygous reference allele carriers (**1/*1*), despite the presence of a clear allele dependent decrease in CYP4F2 protein abundance. This can be explained in part by variability in background CYP4F11 protein abundance (**Figure 3.23 D**) in the **1/*3* group. Moreover, when accounting for combined CYP4F2 and CYP4F11 protein content, the allele dependent decrease in protein abundance between *CYP4F2*3* heterozygous carriers (**1/*3*) and homozygous reference (**1/*1*) carriers is smaller and no longer statistically significant (**Figure 3.24**). Additionally, when intrinsic clearance was adjusted for CYP4F protein content (**Figure 3.21**) the data indicates that there were no significant changes in apparent catalytic

turnover at the genotype level for the *CYP4F2*3* variant nor haplotype 6 and that genetic variability in *CYP4F2* at the genotype and diplotype level was associated with variable protein abundance, which in turn is a strong predictor of metabolic activity. This taken together with results from a previous study[111] which identified recombinant CYP4F11 as having similar catalytic efficiency as CYP4F2 for VK_2 (MK_4) ω-hydroxylation, provides evidence that the lack on an allele dependent *CYP4F2*3* effect on the intrinsic clearance is due to comparable levels of total CYP4F2+CYP4F11 protein content in the *CYP4F2*1/*3* and *CYP4F2*1/*1* groups (13.4 ± 4.24 vs. 15.3 ± 5.31 pmol/mg microsomal protein, respectively; $p = 0.083$).

A notable observation from this study was the wide range of protein abundances (25-fold range) and corresponding aggregate metabolic activity (80-fold range) for CYP4F2 and CYP4F11. Although some of this variability can be explained by *CYP4F2*3* genotype (as described in this study), a considerable amount of within genotype variability can be observed for the wild-type *(*1/*1*) group that is not explained by underlying *CYP4F* haplotype. We did observe a significant contribution of cytochrome P450 reductase (CPR) protein abundance (**Figure 3.18**) to the variability in metabolic activity towards VK_1, but the absolute contribution to the observed variability was small. This suggests that non-genetic (environmental) factors may be involved and contributing to the observed phenotypic behavior, particularly for individuals who are carriers of fully functional alleles.

A limitation of this study was the lack of commercially available ω-hydroxy VK_1 standard for metabolite quantitation. However, we expect proportionality between measured PAR values and absolute levels of ω-hydroxy VK_1, based on results of control experiments. Additionally, sample size limitations for diplotype groups restricted our ability to probe the effect size of the *CYP4F2*3* variant alone (diplotype 17 / 17), *CYP4F2*2* variant alone (diplotype 12 / 12). It is

possible that variants in LD with the *CYP4F2*3* allele seen in haplotype 20 have an undetected effect on catalytic activity that would only emerge by comparison with the diplotype 17 / 17 group, but if true, it would likely be of a small magnitude.

In conclusion, we effectively evaluated underlying mechanistic links between genetic variability in VK metabolic pathway genes (*CYP4F2* and *CYP4F11*) and metabolic activity using a diplotype-based approach. Our study highlights the challenge of determining the effect size of causal variants given complex genomics traits such as linkage disequilibrium and demonstrates the practical application of population Michaelis-Menten (PopMM) modeling as an approach for characterizing genomic traits *in vitro*. Results suggest that testing for the *CYP4F2*3* allele alone is sufficient for *in vivo* predictions such as the warfarin dose needed to achieve a therapeutic INR. Results also indicate that both CYP4F2 and CYP4F11 enzymes contribute importantly to the metabolism of VK_1 in human liver.

3.6 TABLES AND FIGURES

Table 3.1 Characteristics of select CYP4F2 and CYP4F11 variants.

RSID	Gene	Location (Chr.19)	MAF	Alleles	Amino Acid Δ	Mutation Type
rs2108622 (*3)	*CYP4F2*	15990431	0.256	C>T	Val/Met	Missense
rs2074900	*CYP4F2*	15996820	0.323	G>A	His	Synonymous
rs3093160	*CYP4F2*	15996907	0.147	C>T	--	Intron
rs3093153	*CYP4F2*	16001215	0.059	C>A	GLY/VAL	Missense
rs3093114	*CYP4F2*	16006413	0.152	G>A	ALA	Synonymous
rs3093106	*CYP4F2*	16008257	0.161	T>C	PRO	Synonymous
rs3093105 (*2)	*CYP4F2*	16008388	0.159	A>C	TRP/GLY	Missense
rs3093103	*CYP4F2*	16008434	0.152	A>G	--	Intron
rs3093100	*CYP4F2*	16008469	0.159	C>G	--	Intron
rs1060463	*CYP4F11*	16025176	0.412	T>C	ASP/ASN	Missense
rs8104361	*CYP4F11*	16034714	0.203	G>A	CYS/ARG	Missense
rs3765070	*CYP4F11*	16040292	0.412	G>A	ILE	Synonymous
rs2305801	*CYP4F11*	16045141	0.206	C>T	GLY	Synonymous

Minor allele frequency (MAF) >0.05 for all variants. Star allele nomenclature denoted next to RSID if available.

Table 3.2. Human liver bank mRNA analysis population demographic summary

Parameter	CYP4F2 Genotype		
	***1/*1**	***1/*3**	***3/*3**
Population (n)	154	100	20
Male/Female (n)	85/69	66/34	10/10
Age (years)	40.05 ± 22.97	41.18 ± 20.23	46.41 ± 17.41
CYP4F2 mRNA (FPKM)‡	46.64 ± 19.14	48.91 ± 20.17	48.47 ± 22.6
CYP4F11 mRNA (FPKM)‡	23.06 ± 7.74	22.2 ± 7.08	22.82 ± 8.32

‡*Abbreviations: FPKM, Fragments per kilobase million.*

Table 3.3 Liver Bank Haplotype Summary

Haplotype Code	Haplotype	RSID	Frequency	
			n	(%)
1	CGCCGTAACTGGC	rs1060463	50	8.96
2	CGCCGTAACTGAC	rs1060463 rs3765070	2	0.36
3	CGCCGTAACCGGT	rs2305801	1	0.18
4	CGCCGTAACCGAC	rs3765070	30	5.38
5	CGCCGTAACCAAC	rs8104361 rs3765070	6	1.08
6	CGCCGTAACCAAT	rs8104361 rs3765070 rs2305801	105	18.82
7	CGCCGCAACCGGT	rs3093106 rs2305801	1	0.18
8	CGCCGCCAGTGGC	rs3093106 rs3093105 rs3093100 rs1060463	2	0.36
9	CGCCGCCAGCGAT	rs3093106 rs3093105 rs3093100 rs3765070 rs2305801	2	0.36
10	CGCCACCGGTGGC	rs3093114 rs3093106 rs3093105 rs3093103 rs3093100 rs1060463	2	0.36
11	CGCCACCGGCGAC	rs3093114 rs3093106 rs3093105 rs3093103 rs3093100 rs3765070	1	0.18
12	CGCAGTAACTGGC	rs3093153 rs1060463	33	5.91
13	CACCGTAACTGGC	rs2074900 rs1060463	172	30.82
14	CACCGTAACTGGT	rs2074900 rs1060463 rs2305801	2	0.36
15	CACCGTAACCGAC	rs2074900 rs3765070	4	0.72
16	CACCGTAACCAAT	rs2074900 rs8104361 rs3765070 rs2305801	2	0.36
17	TGCCGTAACTGGC	rs2108622 rs1060463	59	10.57
18	TGCCGTAACTGGT	rs2108622 rs1060463 rs2305801	2	0.36
19	TGTCACCGGTGGC	rs2108622 rs3093160 rs3093114 rs3093106 rs3093105 rs3093103 rs3093100 rs1060463	4	0.72
20	TGTCACCGGCGAC	rs2108622 rs3093160 rs3093114 rs3093106 rs3093105 rs3093103 rs3093100 rs3765070	78	13.98

Table 3.4. Summary of CYP4F2 and CYP4F11 protein abundances.

	CYP4F2[‡] *(mean ± sd)*	**CYP4F11**[‡] *(mean ± sd)*	**Total CYP4F**[‡] *(mean ± sd)*
Genotype			
**1/*1*	12.6 ± 4.44	2.51 ± 1.20	15.2 ± 5.31
**1/*3*	10.5 ± 3.37	3.22 ± 1.13	13.7 ± 4.08
**3/*3*	3.40 ± 1.38	2.25 ± 1.34	5.65 ± 2.27
Diplotype[◊]			
13 / 13	12.4 ± 5.24	2.87 ± 1.59	15.3 ± 6.40
13 / 17	12.5 ± 5.39	3.12 ± 1.27	15.7 ± 6.53
1 /17	10.3 ± 1.15	3.33 ± 1.19	13.6 ± 2.34
13 / 20	9.82 ± 2.47	3.57 ± 1.37	13.4 ± 3.23
6 / 13	18.1 ± 4.86	3.66 ± 0.87	21.7 ± 5.59
1 / 20	9.06 ± 2.67	2.36 ± 0.45	11.4 ± 3.12
6 / 20	10.0 ± 1.92	2.97 ± 0.28	13.0 ± 1.64
20 / 20	3.01 ± 1.32	1.96 ± 1.49	4.97 ± 2.47
1 / 13	11.9 ± 6.02	3.16 ± 1.29	15.0 ± 7.01
6 / 6	12.1 ± 2.86	1.71 ± 0.59	13.8 ± 3.20
17 / 20	3.98 ± 1.37	2.70 ± 1.03	6.67 ± 1.63
4 / 6	12.5 ± 5.83	1.86 ± 0.88	14.4 ± 6.54
4 / 4	12.3 ± 5.81	2.80 ± 1.54	15.1 ± 7.35
1 / 6	12.3 ± 3.05	2.00 ± 0.68	14.3 ± 3.66
12 / 13	10.9 ± 2.53	2.41 ± 0.96	13.3 ± 3.24
1 / 4	12.7 ± 1.92	2.02 ± 0.45	14.7 ± 1.48

[‡]*Units expressed as pmol/mg microsomal protein.*

[◊]*Diplotype groups containing only <2 subjects were omitted from analysis (6/17, 1/12, 17/17, 4/13, and 12/12)*

Table 3.5. Population Michaelis-Menten model parameter estimate comparisons.

	(1) Base Model		(2) Genotype Covariate Model		(3) Diplotype Covariate Model		Change (%)
	Estimate◊	95% CI	Estimate◊	95% CI	Estimate◊	95% CI	
$\theta_{Typical, Vmax}$	**0.14 *****	**[0.12-0.18]**					
$\theta_{Typical, Km}$	**4.22 *****	**[3.6-4.95]**					
Genotype Estimates							
V_{max} *1/*1			**0.17 *****	**[0.13-0.23]**			
V_{max} *1/*3			0.17	[0.1-0.8]			-2% ↓
V_{max} *3/*3			**0.07 ****	**[0.04-0.13]**			**-57%** ↓
K_m *1/*1			**4.25 *****	**[3.5-5.15]**	**4.24 *****	**[3.48-5.17]**	
K_m *1/*3			2.89	[1.96-4.25]	2.89	[1.94-4.31]	-32% ↓
K_m *3/*3			**8.66 ****	**[5.5-13.65]**	**8.6 ****	**[5.4-13.74]**	**103%** ↑
Diplotype Estimates							
V_{max} 13 / 13					**0.17 *****	**[0.11-0.27]**	
V_{max} 13 / 17					0.17	[0.07-0.41]	-2% ↓
V_{max} 1 / 17					0.32	[0.09-1.1]	82% ↑
V_{max} 13 / 20					0.18	[0.08-0.39]	5% ↑
V_{max} 6 / 13					**0.49 ***	**[0.21-1.15]**	**181%** ↑
V_{max} 1 / 20					0.11	[0.03-0.39]	-36% ↓
V_{max} 6 / 20					0.17	[0.05-0.61]	0%
V_{max} 20 / 20					**0.07 ***	**[0.03-0.14]**	**-62%** ↓
V_{max} 1 / 13					0.28	[0.1-0.8]	61% ↑
V_{max} 6 / 6					0.14	[0.07-0.29]	-21% ↓
V_{max} 17 / 20					0.08	[0.03-0.18]	-63% ↓
V_{max} 4 / 6					0.15	[0.06-0.38]	-14% ↓
V_{max} 4 / 4					0.21	[0.06-0.73]	21% ↑
V_{max} 1 / 6					0.14	[0.05-0.35]	-21% ↓
V_{max} 12 / 13					0.15	[0.07-0.33]	-12% ↓
V_{max} x 1 / 4					0.07	[0.02-0.26]	-58% ↓
Residual (Power model)	0.848		0.835		0.838		
Subjects (n)	88		88		88		
logLik	455.057		471.942		486.917		
AIC	-898.113		-923.884		-917.833		

**** p < 0.001; ** p < 0.01; * p < 0.05.*

◊ Units for V_{max} and K_m are PAR/min/mg microsomal protein and μM, respectively.

Log-likelihood; logLik, Akaike information criterion; AIC

Diplotype groups containing only <2 subjects were omitted from analysis (6/17, 1/12, 17/17, 4/13, and 12/12).

Parameters reported as; Estimate [95% CI]

Table 3.6. Intrinsic clearance estimates comparisons

	(1) Genotype Covariate Model		(2) Diplotype Covariate Model		Change (%)
	Estimate[◊]	95% CI	Estimate[◊]	95% CI	
ln β_0					
Reference[†]	1.18	[0.92-1.51]	1.17	[0.75-1.83]	
ln $\beta_{1,\ Genotype}$					
Genotype *1/*3	1.32	[0.82-2.14			12%↑
Genotype *3/*3	**0.42 *** **	**[0.25-0.7]**			**-65%↓**
ln $\beta_{1,\ Diplotype}$					
Diplotype 13 / 17			1.33	[0.54-3.25]	13% ↑
Diplotype 1 / 17			2.43	[0.66-8.97]	107%↑
Diplotype 13 / 20			1.41	[0.64-3.11]	20%↑
Diplotype 6 / 13			**3.23 ***	**[1.32-7.9]**	**175%↑**
Diplotype 1 / 20			0.9	[0.24-3.32]	-23%↓
Diplotype 6 / 20			1.34	[0.36-4.95]	15%↑
Diplotype 20 / 20			**0.38 ****	**[0.18-0.78]**	**-68%↓**
Diplotype 1 / 13			1.88	[0.63-5.63]	60%↑
Diplotype 6 / 6			0.95	[0.45-2.03]	-19%↓
Diplotype 17 / 20			**0.43 ***	**[0.19-1]**	**-63%↓**
Diplotype 4 / 6			1.03	[0.39-2.74]	-12%↓
Diplotype 4 / 4			1.43	[0.39-5.28]	22%↑
Diplotype 1 / 6			0.95	[0.36-2.52]	-19%↓
Diplotype 12 / 13			1.04	[0.47-2.3]	-11%↓
Diplotype 1 / 4			0.52	[0.14-1.93]	-55%↓
Subjects (n)	88		88		
R^2	0.174		0.398		
logLik	-116.034		-102.094		
AIC	240.068		248.188		

**** p < 0.001; ** p < 0.01; * p < 0.05.*

[◊] Units for intrinsic clearance are (PAR/ mg microsomal protein/ μMol phylloquinone) x L/min.

*[†] Genotype covariate model reference = *1/*1, Diplotype covariate model reference = 13 / 13*

Log-likelihood; logLik, Akaike information criterion; AIC

Diplotype groups containing only <2 subjects were omitted from analysis (6/17, 1/12, 17/17, 4/13, and 12/12).

Parameters reported as; Estimate [95% CI]

Table 3.7. Genotype-specific CL_{int} estimates adjusted for total CYP4F2 and CYP4F11 protein abundance.

	CL_{int}			**CL_{int} (Log-Transformed)**			**CL_{int} (Adjusted for protein)**			**CL_{int} (Log-Transformed, adjusted for protein)**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	1.82	1.41 – 2.22	**<0.001**	0.16	-0.09 – 0.42	0.198	-0.68	-1.66 – 0.30	0.172	-1.30	-1.93 – -0.67	**<0.001**
Genotype [*1/*3]	0.00	-0.79 – 0.80	0.992	0.12	-0.37 – 0.61	0.639	0.19	-0.57 – 0.94	0.627	0.21	-0.27 – 0.69	0.394
Genotype [*3/*3]	-1.30	-2.15 – -0.46	**0.003**	-1.04	-1.56 – -0.52	**<0.001**	0.03	-1.04 – 1.10	0.955	-0.29	-0.96 – 0.39	0.406
CYP4F2 (pmol/mg)							0.13	0.04 – 0.21	**0.007**	0.07	0.01 – 0.13	**0.019**
CYP4F11 (pmol/mg)							0.33	0.05 – 0.61	**0.020**	0.21	0.04 – 0.39	**0.019**
R^2 / R^2 adjusted	0.106 / 0.085			0.174 / 0.155			0.366 / 0.335			0.390 / 0.360		
AIC	324.842			240.058			286.584			204.613		
log-Likelihood	-158.421			-116.029			-137.292			-98.994		

Table 3.8. Haplotype 6 containing diplotype CL_{int} estimates adjusted for total CYP4F2 and CYP4F11 protein abundance.

	CL_{int}			**CL_{int} (Log-Transformed)**			**CL_{int} (Adjusted for protein)**			**CL_{int} (Log-Transformed, adjusted for protein)**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	1.61	0.70 – 2.52	**0.001**	-0.03	-0.59 – 0.54	0.925	0.05	-2.02 – 2.11	0.961	-0.96	-2.25 – 0.32	0.134
Diplotype [6 / 13]	2.47	0.79 – 4.15	**0.006**	1.20	0.16 – 2.23	**0.026**	1.92	0.08 – 3.76	**0.042**	0.82	-0.33 – 1.97	0.153
Diplotype [6 / 6]	-0.56	-2.00 – 0.88	0.428	-0.02	-0.91 – 0.87	0.957	-0.20	-1.81 – 1.41	0.801	0.10	-0.90 – 1.10	0.835
CYP4F2 (pmol/mg)							0.06	-0.13 – 0.25	0.535	0.05	-0.06 – 0.17	0.351
CYP4F11 (pmol/mg)							0.29	-0.41 – 1.00	0.396	0.09	-0.35 – 0.53	0.676
R^2 / R^2 adjusted	0.376 / 0.319			0.230 / 0.160			0.463 / 0.356			0.326 / 0.191		
AIC	96.734			82.955			96.957			83.633		
log-Likelihood	-44.367			-32.330			-42.478			-30.668		

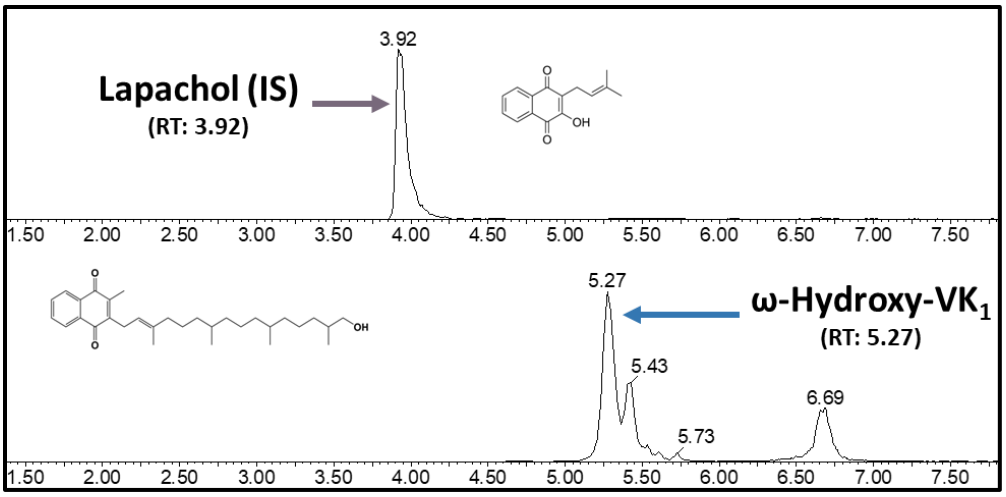

Figure 3.1. UPLC-MS/MS chromatogram of ω-hydroxy vitamin K_1 metabolite (466.3>185 *m/z*) and internal standard (lapachol, 241>186.1 *m/z*). Units for retention times (RT) are in minutes.

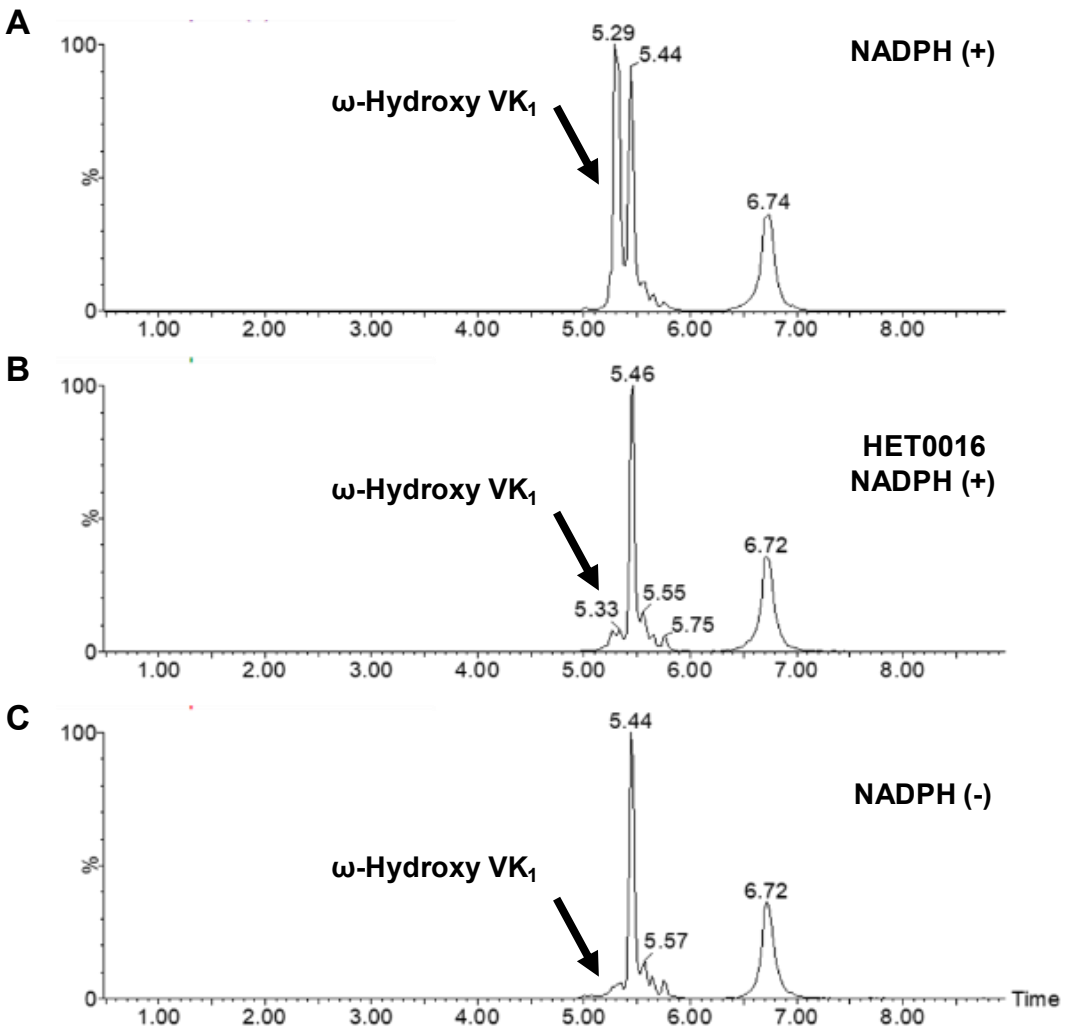

Figure 3.2. UPLC-MS/MS chromatogram of ω-hydroxy vitamin K_1 metabolite formation under **(A)** NADPH (+), **(B)** NADPH (+) with selective CYP4F inhibitor HET0016 (20 μM), and **(C)** NADPH (-) conditions. Units for retention times (RT) are in minutes.

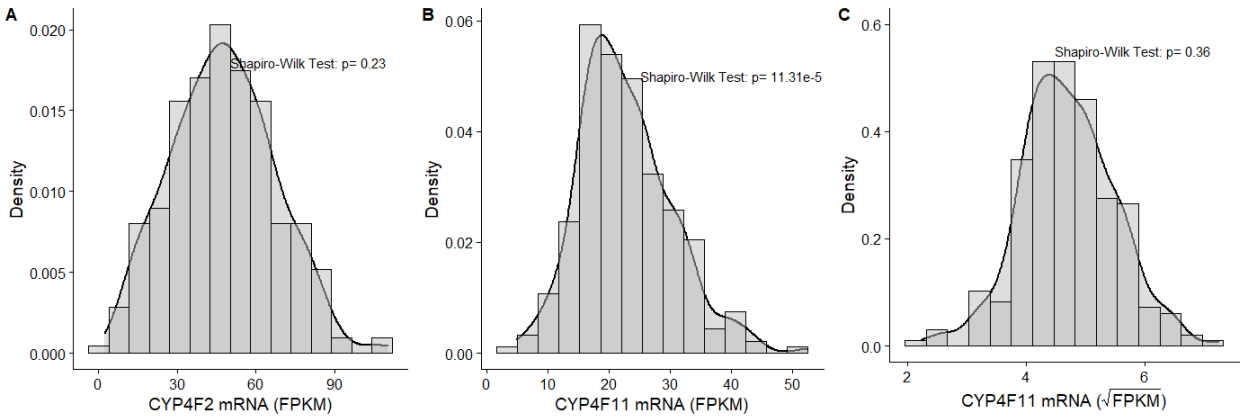

Figure 3.3. Distribution of hepatic mRNA expression of (A) *CYP4F2*, (B) *CYP4F11*, (C) *CYP4F11* after square root transformation. Expression was measured as fragments per kilobase per million bases read (FPKM). Normality was assessed using the Shapiro-Wilks test for normality with *p*-values >0.05 indicating no statistically significant deviation from a normal distribution.

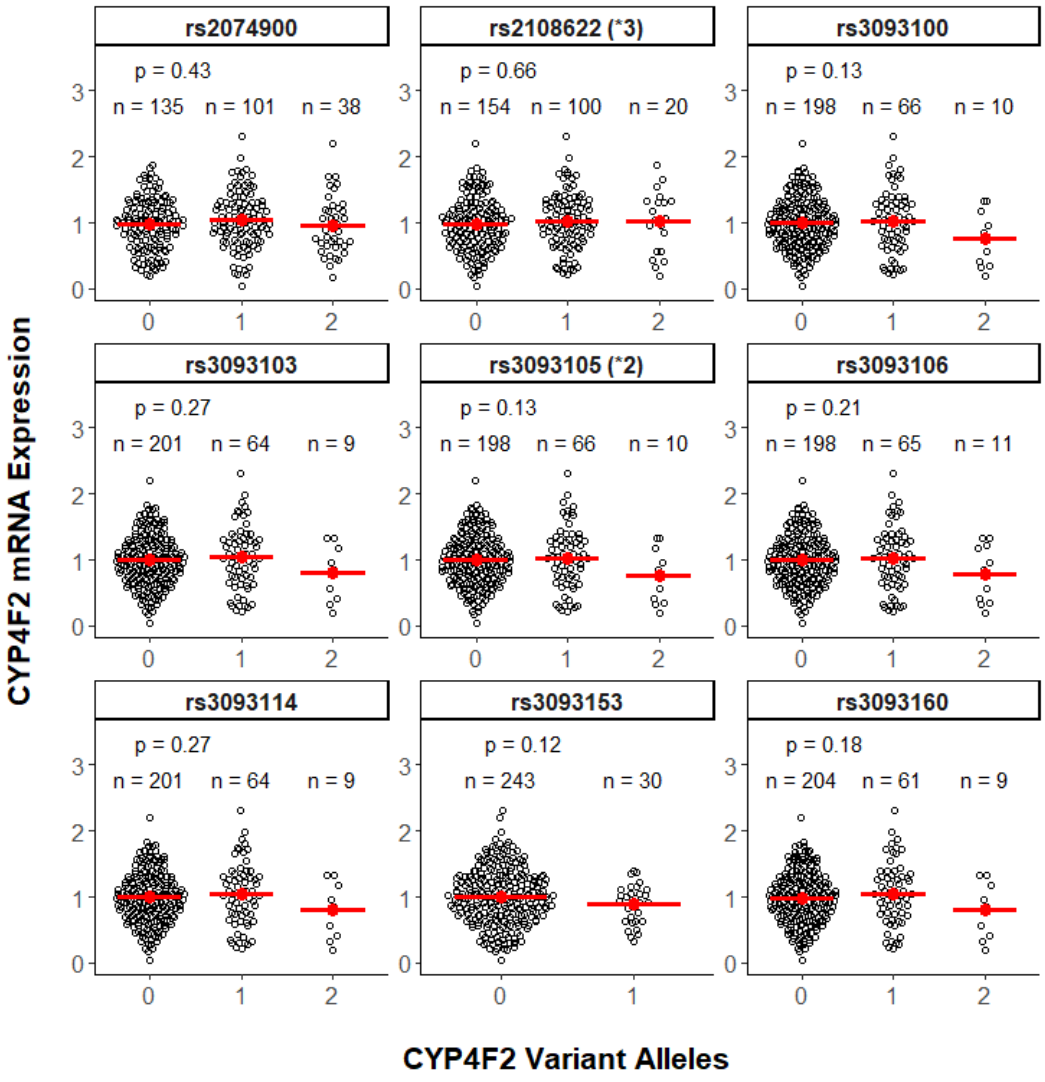

Figure 3.4. Human liver bank (n=274) hepatic *CYP4F2* mRNA abundance stratified by variants found at the *CYP4F2* locus.

The y-axis represents normalized mRNA expression values and the x-axis discreet number of variant alleles. p-values comparing mean expression across allele copy groups were obtained using a one-way ANOVA.

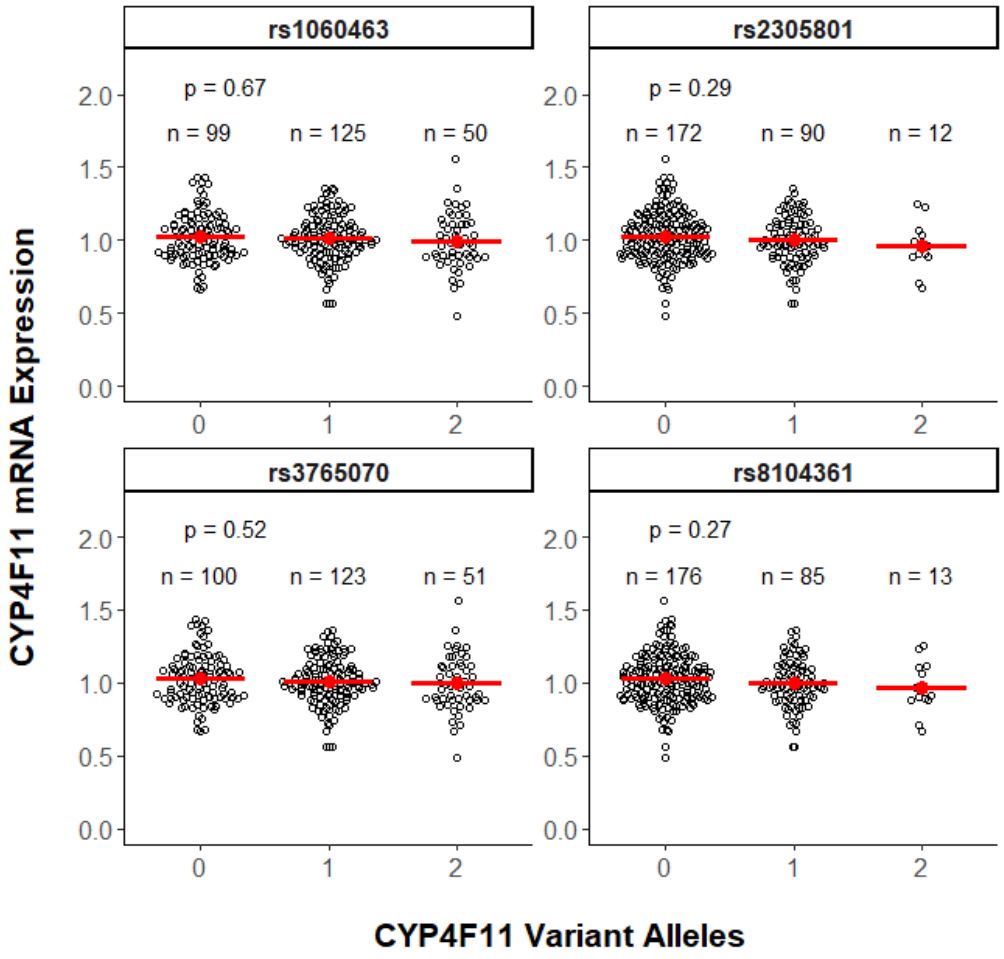

Figure 3.5. Human liver bank (n=274) hepatic *CYP4F11* mRNA abundance stratified by variants found at the *CYP4F11* locus.

The y-axis represents normalized mRNA abundance values and the x-axis discreet number of variant alleles. p-values comparing mean expression across allele copy groups were obtained using a one-way ANOVA.

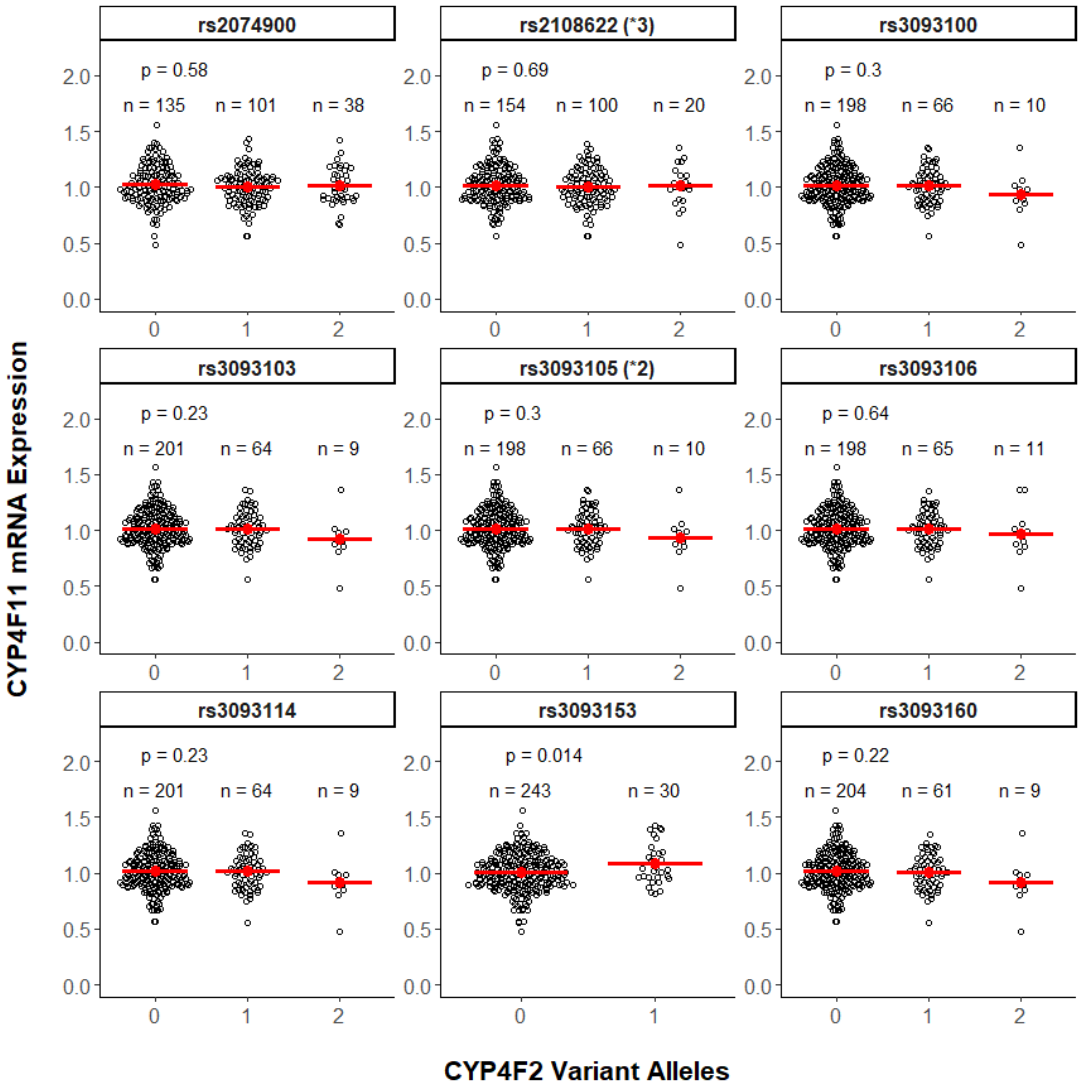

Figure 3.6. Hepatic *CYP4F11* mRNA expression stratified by variants found at the *CYP4F2* locus. The y-axis represents normalized mRNA expression values and the x-axis discreet number of variant alleles. *p*-values comparing mean expression across allele copy groups were obtained using a one-way ANOVA.

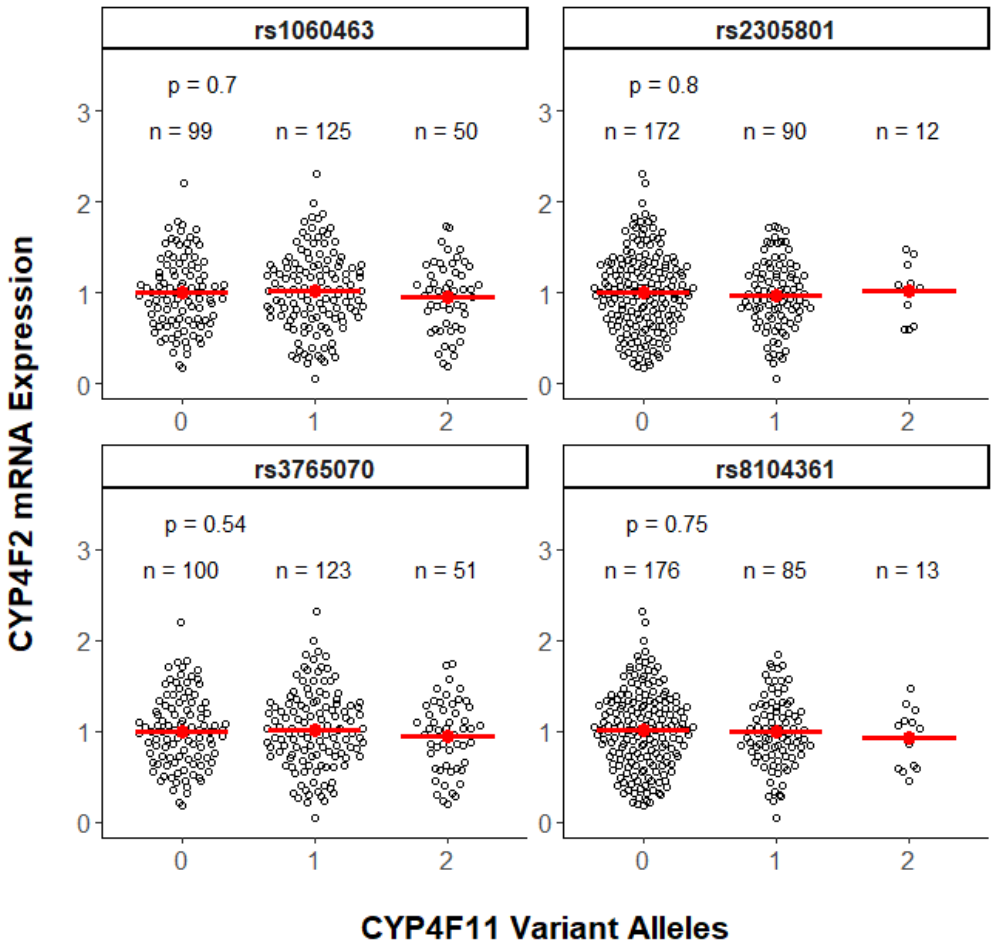

Figure 3.7. Hepatic *CYP4F2* mRNA expression stratified by variants found at the *CYP4F11* locus. The y-axis represents normalized mRNA expression values and the x-axis discreet number of variant alleles. *p*-values comparing mean expression across allele copy groups were obtained using a one-way ANOVA.

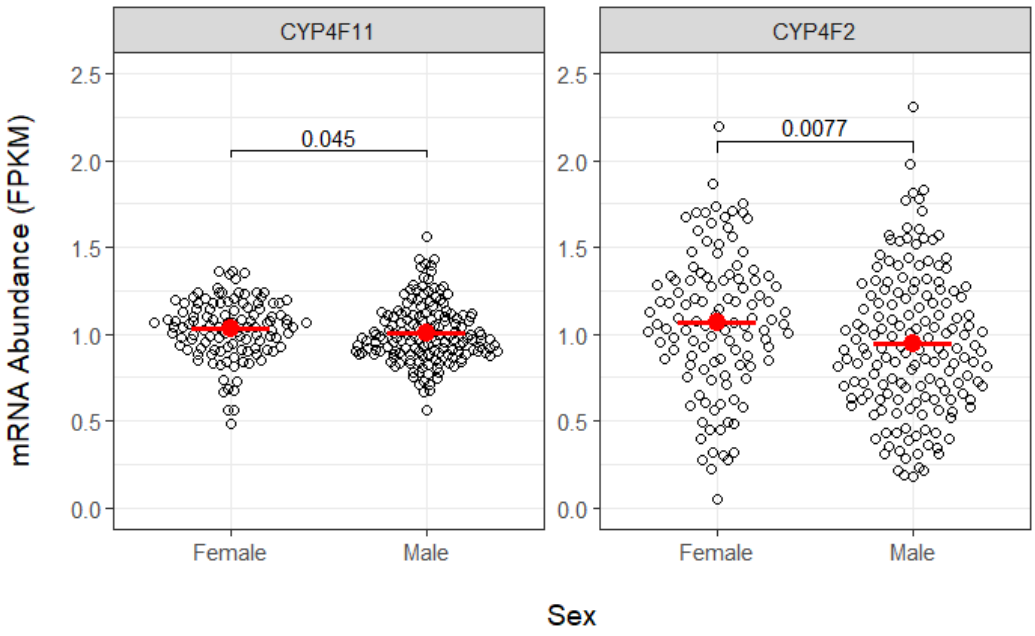

Figure 3.8. Human liver bank (n=274) hepatic *CYP4F2* mRNA abundance stratified by sex. The y-axis represents normalized mRNA expression values and the x-axis sex. *p*-values comparing mean expression across sexes were obtained using a Wilcox test.

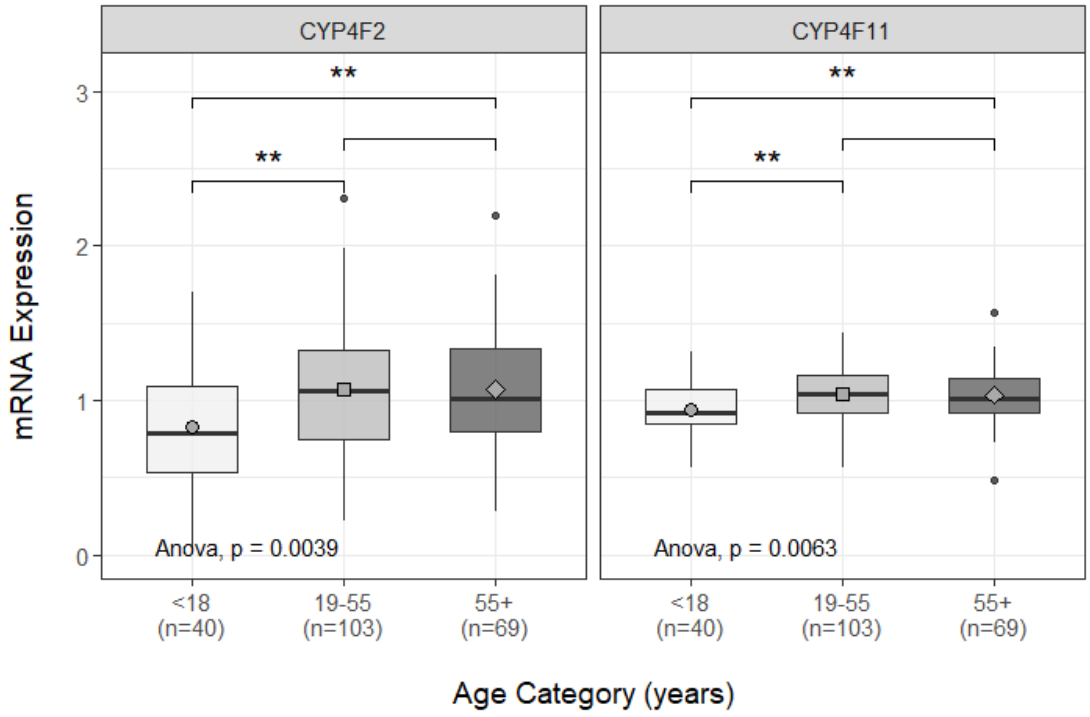

Figure 3.9. *CYP4F2* and *CYP4F11* mRNA expression by age category.

The y-axis represents normalized mRNA expression values. Comparison of mean mRNA expressions across age groups was evaluated using a one-way ANOVA. In the left panel a statistically significant difference in mean normalized *CYP4F2* mRNA expression was found between the following groups: <18 years and 19-55 years age groups, 0.82 ± 0.42 vs. 1.07 ± 0.43, respectively (post-hoc Tukey p=0.01); <18 years and 55+ years age groups at 0.82 ± 0.42 vs. 1.07 ± 0.43, respectively (post-hoc Tukey p=0.01). In the right panel a statistically significant difference in mean normalized *CYP4F11* mRNA expression was found between the following groups: <18 years and 19-55 years age groups, 0.94 ± 0.17 vs. 1.04 ± 0.17, respectively (post-hoc Tukey p=0.01); <18 years and 55+ years age groups at 0.94 ± 0.17 vs. 1.04 ± 0.16, respectively (post-hoc Tukey p=0.02).

Figure 3.10. Linkage-disequilibrium (LD) plot for pairwise D' between variants across *CYP4F2* and *CYP4F11* (Chromosome 19).

D' values are indicated as percentages within squares with strong LD specified by red, while light gray/pink indicate weak association and white as low confidence values.

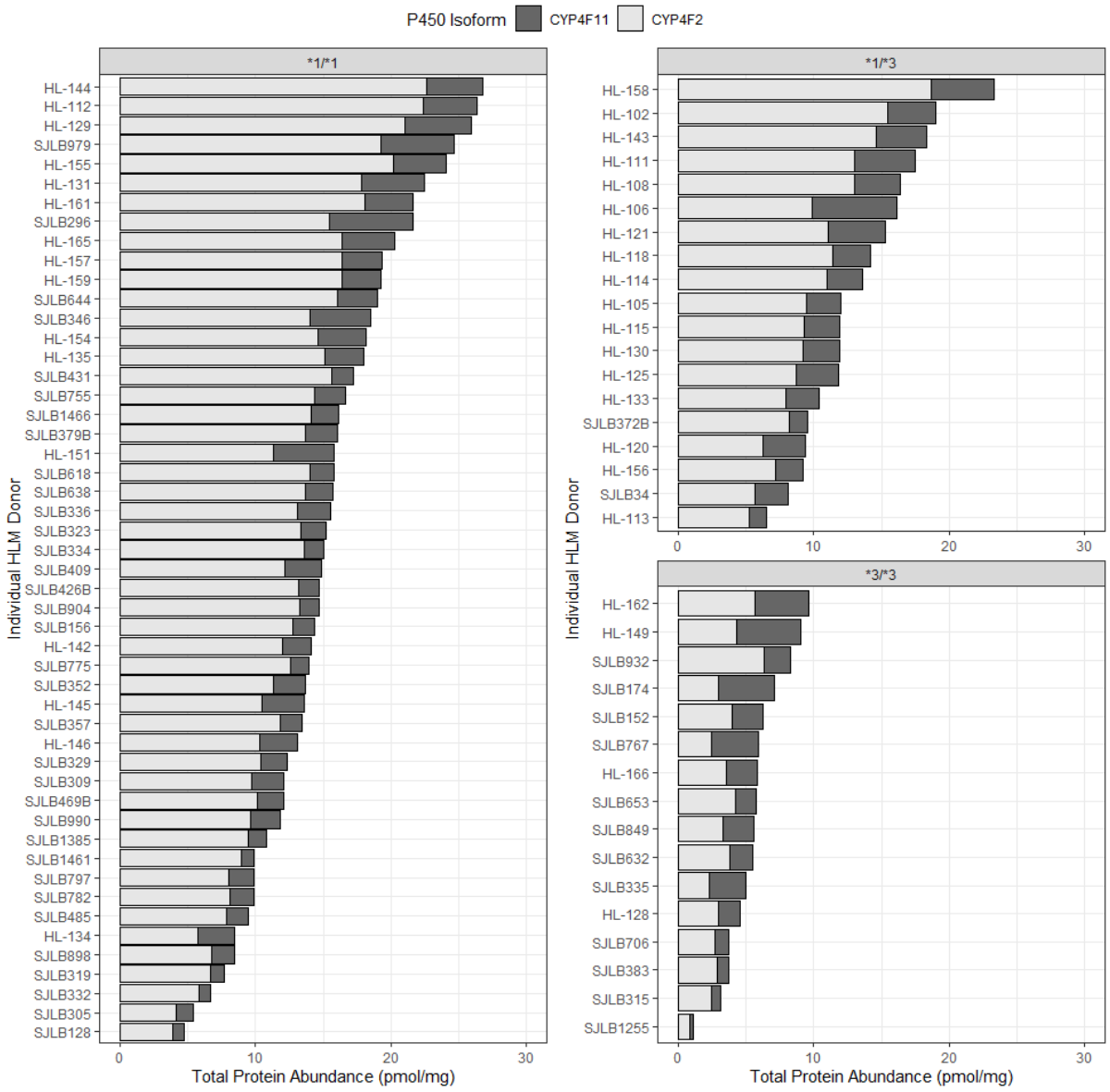

Figure 3.11. The fractional contribution of CYP4F2 and CYP4F11 towards total protein abundance of individual human liver microsomes for (left) wild-type, (top right) heterozygous, and (bottom right) homozygous variant *CYP4F2*3* genotypes.

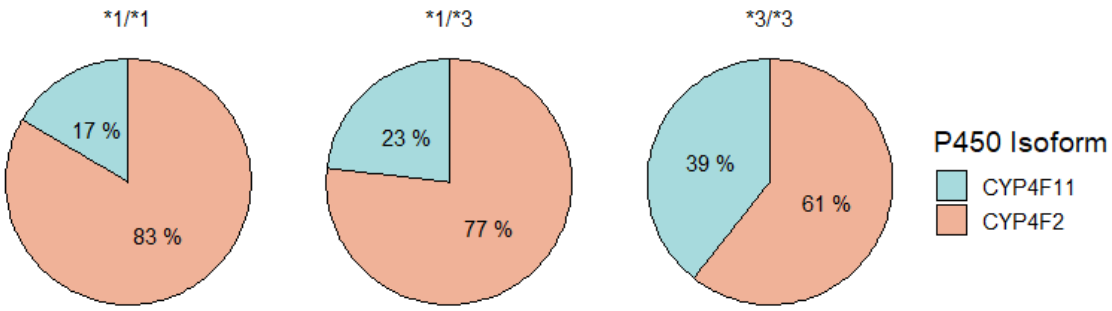

Figure 3.12. Cumulative percent of total protein abundance for CYP4F2 and CYP4F11 by *CYP4F2*3* genotype.

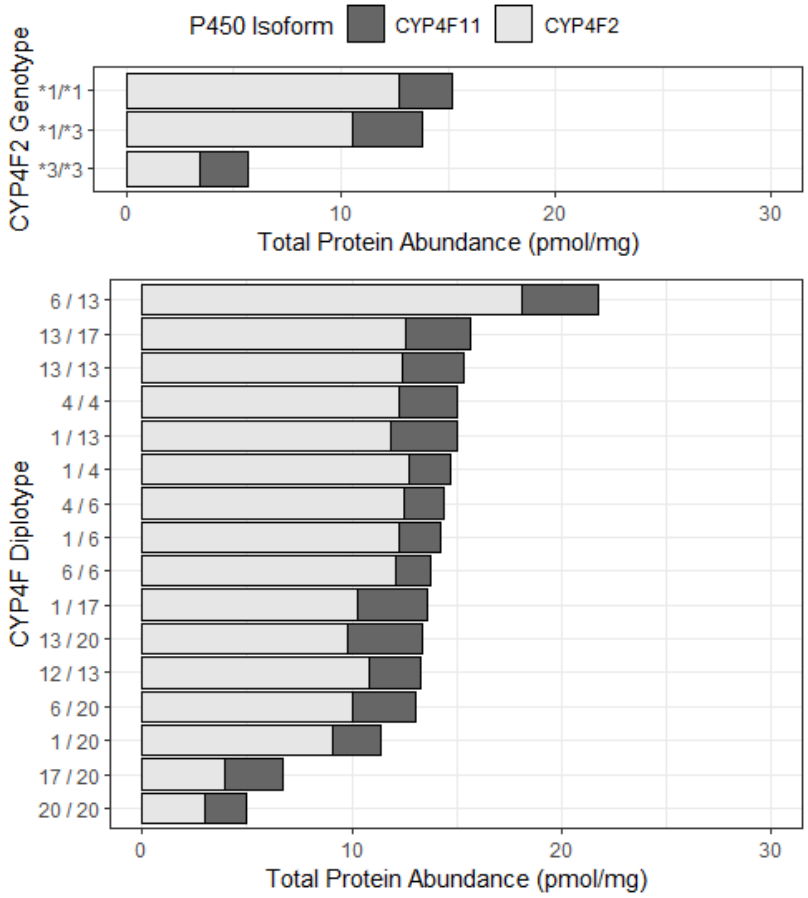

Figure 3.13. The fractional contribution of CYP4F2 and CYP4F11 towards total protein abundance grouped by *CYP4F2*3* genotype (top panel), and *CYP4F* diplotype (bottom-panel). Diplotype groups containing only <2 subjects were omitted from analysis (6/17, 1/12, 17/17, 4/13, and 12/12).

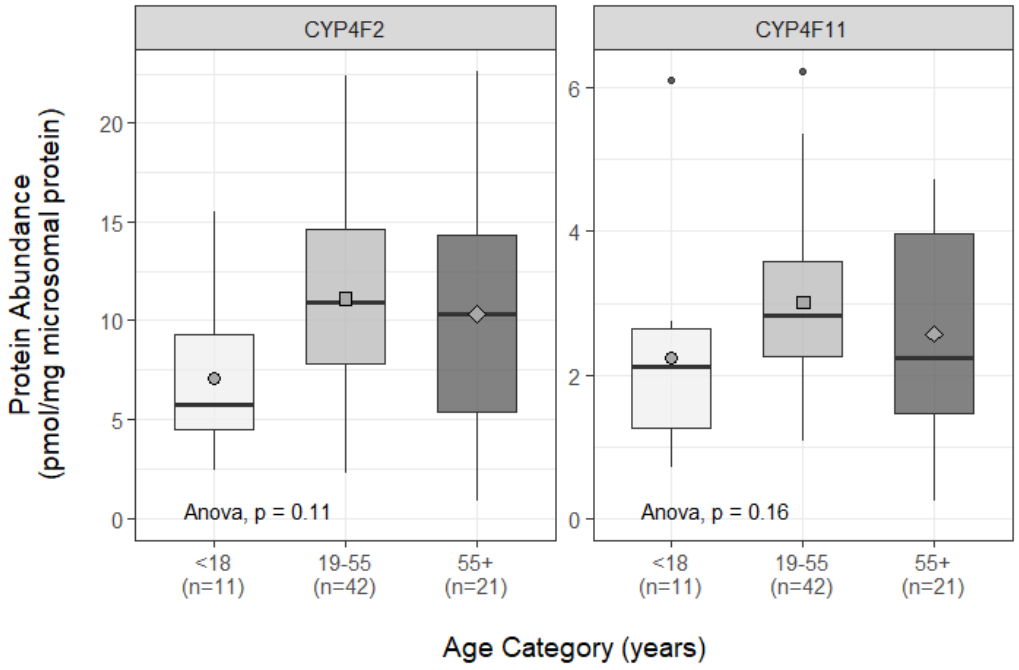

Figure 3.14. CYP4F2 and CYP4F11 protein abundance by age category.

The y-axis represents normalized mRNA expression values. Comparison of mean protein abundance across age groups was evaluated using a one-way ANOVA.

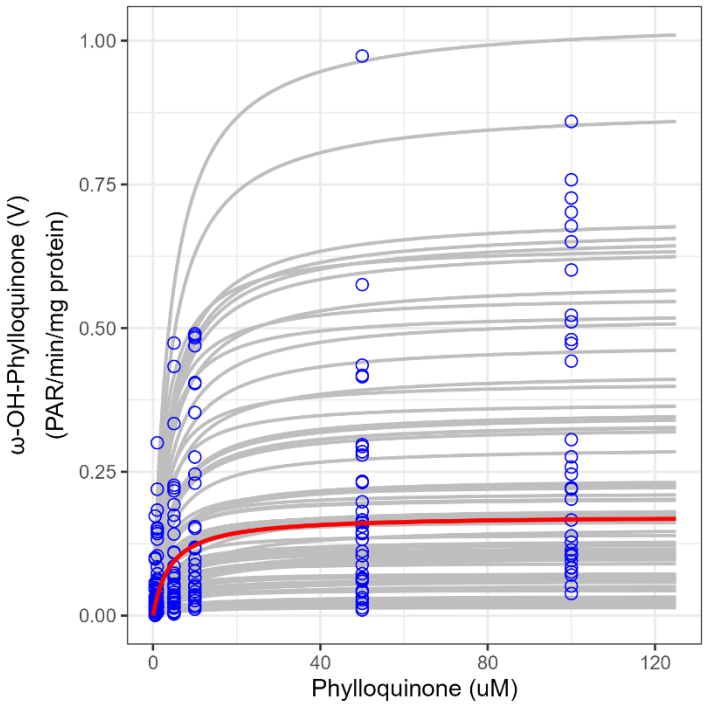

Figure 3.15. NADPH-dependent formation of ω-hydroxy metabolite of vitamin K_1 in human liver microsomal incubations.

Blue circles represent observed data points; solid red and grey lines represent population-predicted and individual-predicted Michaelis-Menten fits (n=88).

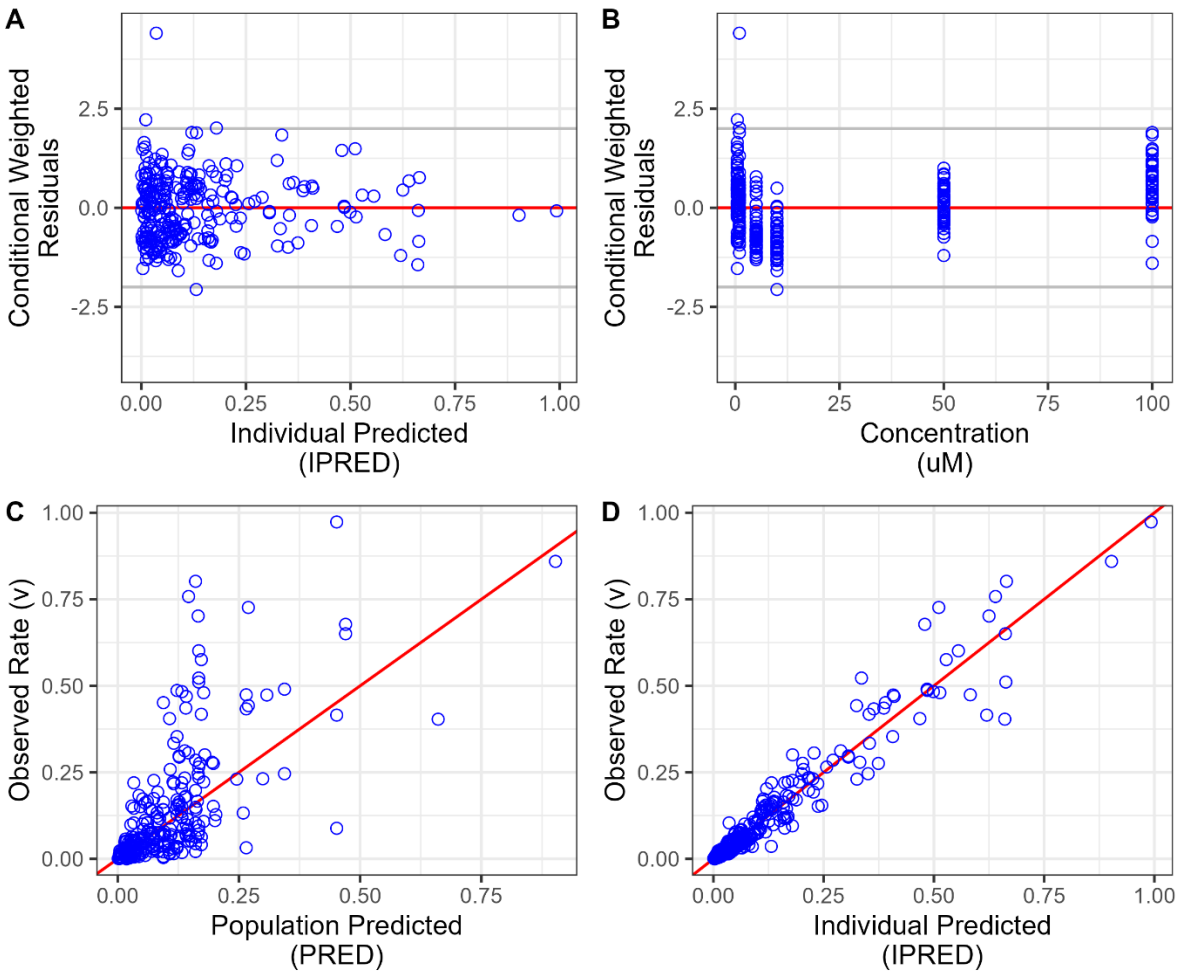

Figure 3.16. Goodness-of-fit plots for the final population Michaelis-Menten (PopMM) model. Blue circles represent observed values and red line represents line of unity.

Top row: conditional weighted residuals vs. (A) individual predicted rate and (B) substrate (phylloquinone) incubation concentration. Bottom row: observed rate vs. (C) population predicted estimate and (D) individual predicted estimate.

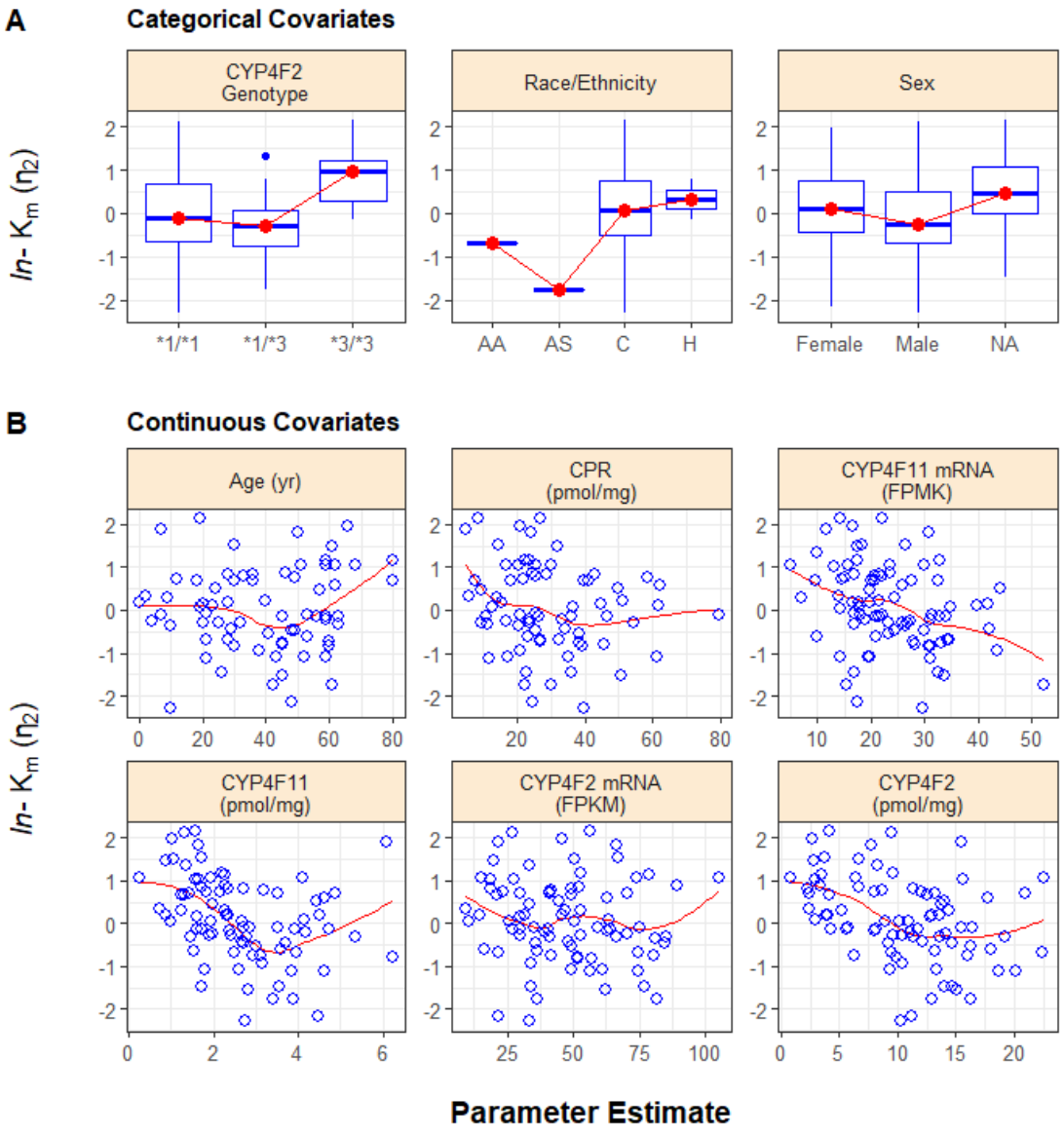

Figure 3.17. Random-effects correlation plots of base PopMM model for natural log estimate of K_m (*ln*-K_m, referred to as η_2 in Equation 2) stratified by (A) categorical covariates, (B) continuous covariates. Abbreviations are defined as the following: AA = African American, AS = Asian, C = Caucasian, H = Hispanic/Latino, CPR = Cytochrome P450 reductase, NA = data unavailable.

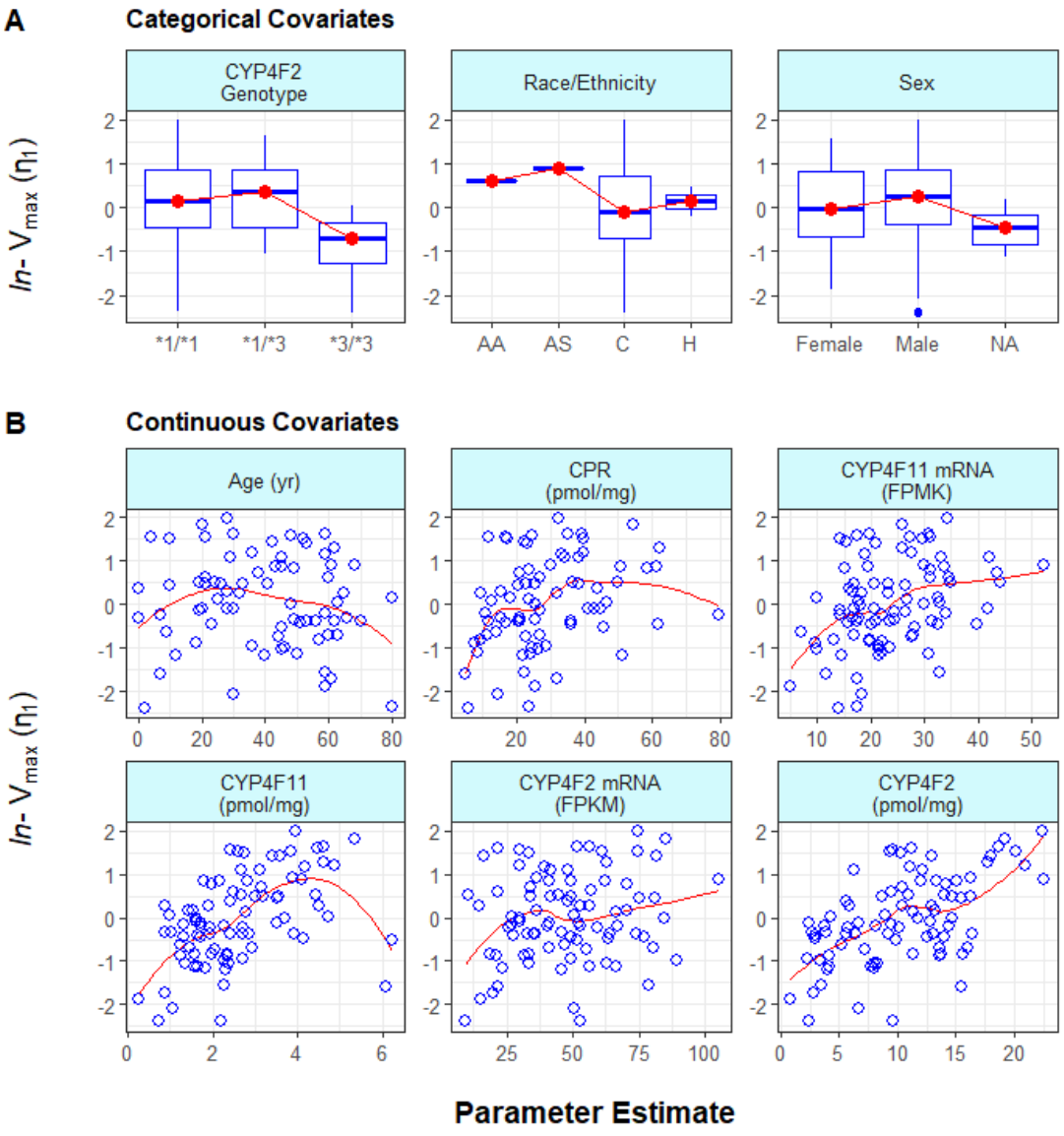

Figure 3.18. Random-effects correlation plots of base PopMM model for natural log estimate of V_{max} (*ln*-V_{max}, referred to as η_1 in Equation 2) stratified by (A) categorical covariates, (B) continuous covariates. Abbreviations are defined as the following: AA = African American, AS = Asian, C = Caucasian, H = Hispanic/Latino, CPR = Cytochrome P450 reductase, NA = data unavailable. CYP4F2 and CYP4F11 protein level were significantly correlated with V_{max}, while cytochrome P450 reductase, CYP4F2 mRNA and CYP411 mRNA were weakly correlated; there was no correlation between age and V_{max}.

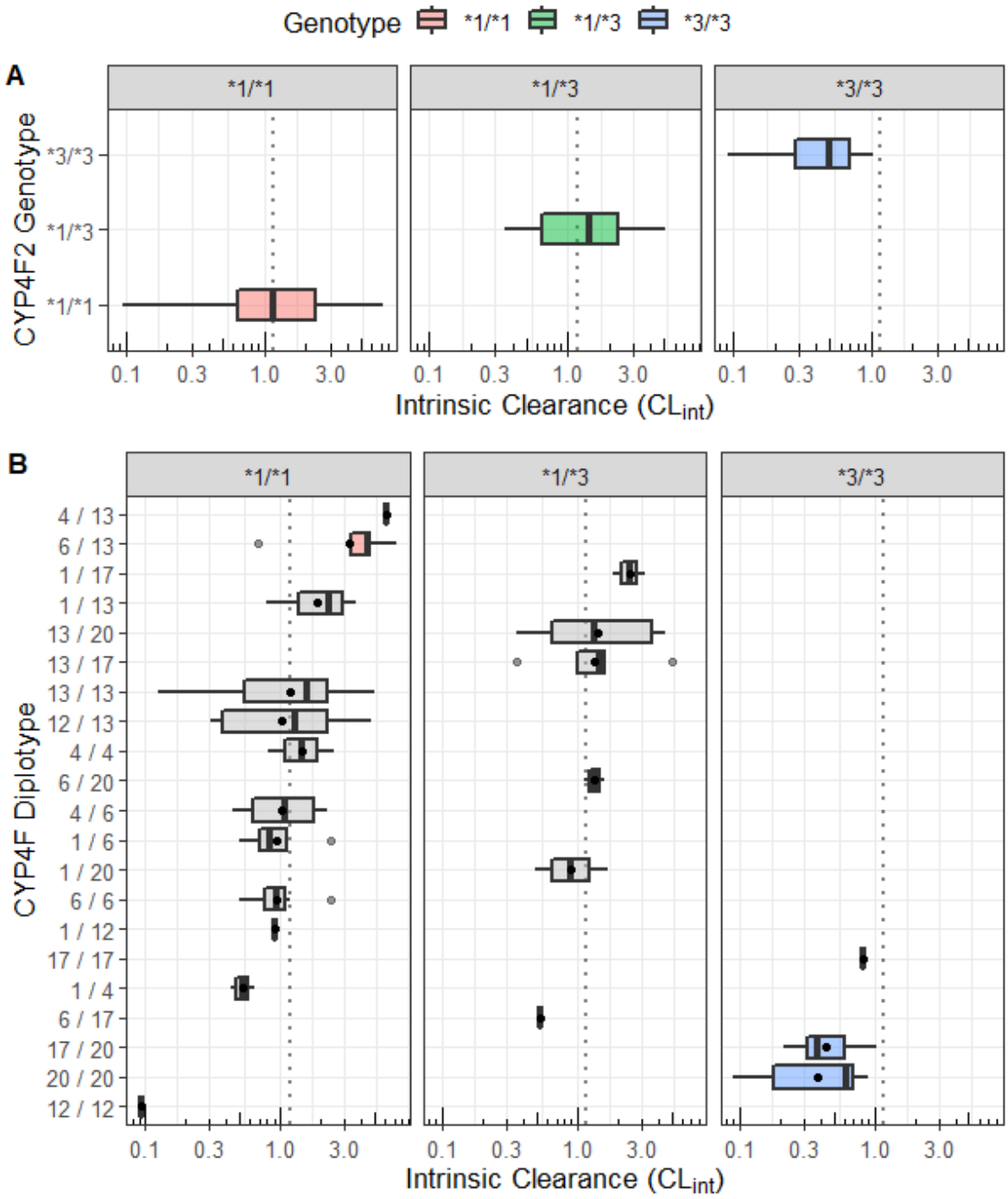

Figure 3.19. Distribution of intrinsic clearance (CL_{int}) estimates stratified by (A) Genotype, and (B) Diplotype. Units for intrinsic clearance are defined as [PAR/ mg microsomal protein/ µMol phylloquinone] x L/min. The dotted line represents intrinsic clearance estimate for reference *CYP4F2* genotype (*1/*1).

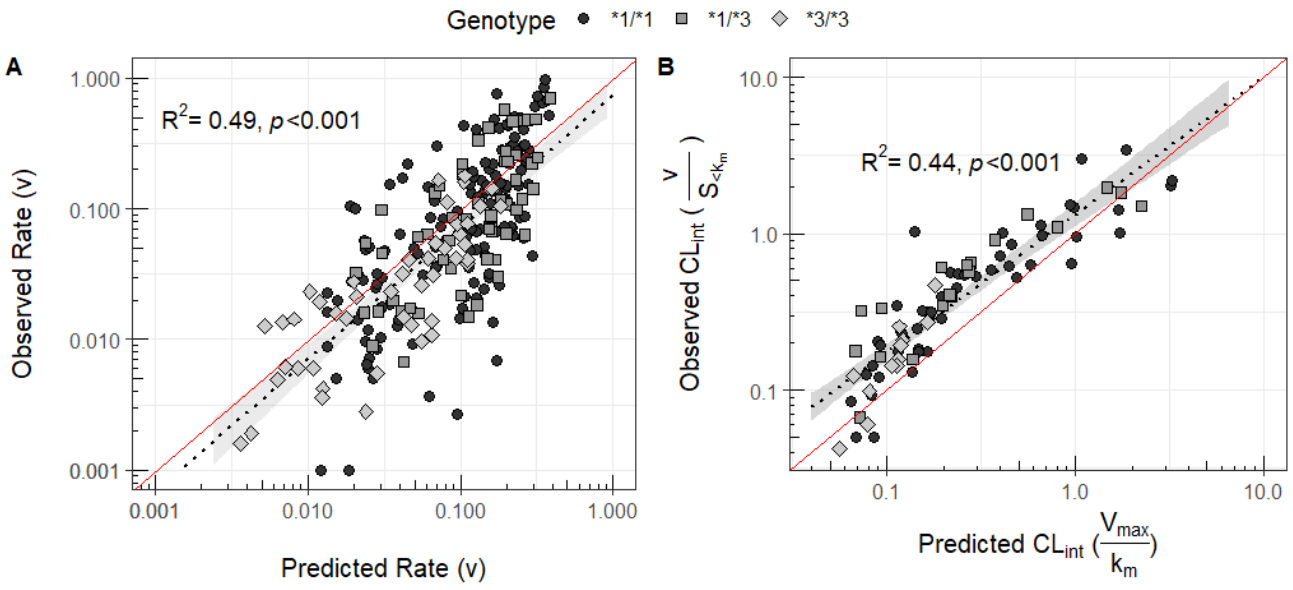

Figure 3.20. Predicted vs experimentally observed reaction rate for ω-hydroxy phylloquinone (VK_1) formation **(A)** and CL_{int} **(B)** in human liver microsomes based on individual CYP4F2 and CYP4F11 protein abundances.

Observed omega-hydroxy VK_1 formation rate for individual human liver microsomes was predicted using the Michaelis-Menten estimates generated from the PopMM analysis. Observed intrinsic clearance was defined as v/[S], where S is a substrate concentration between 3- and 18-fold below expected K_m value. The solid red line represents the line of unity, and the dotted black line represents a linear regression curve with 95% CI (shaded region). Units for parameter estimates are the following: reaction rate (*v*), PAR/min/mg microsomal protein; CL_{int}, (PAR/ mg microsomal protein/ μMol phylloquinone) x L/min.

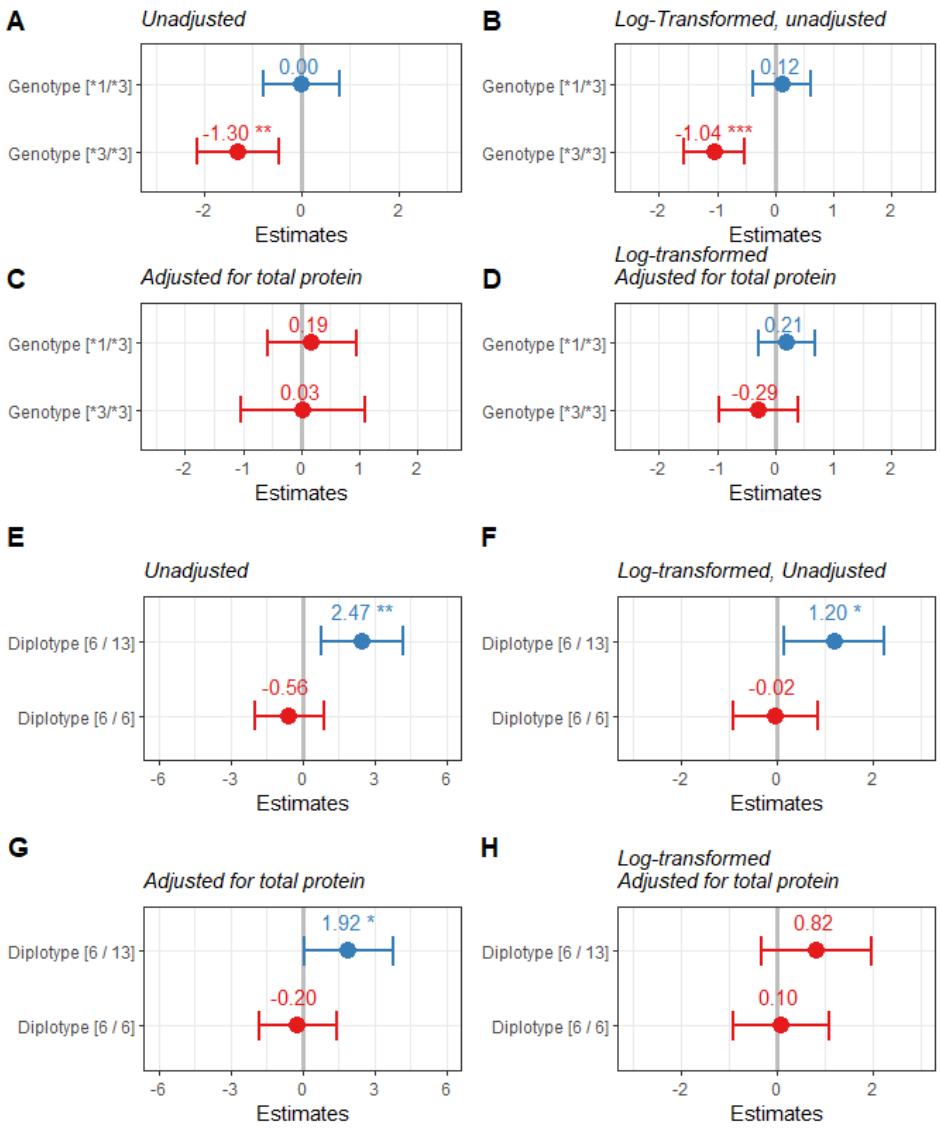

Figure 3.21. Multiple linear regression estimates surveying the impact of *CYP4F2*3* allele **(A-D)** and Haplotype 6 **(E-H)** on CL_{int}.

Conditions: **(A, E)** unadjusted, **(B, F)** log-transformed, **(C, G)** adjusted for CYP4F2 and CYP4F11 protein abundance, and **(D, H)** log-transformed and adjusted for both CYP4F2 and CYP4F11 protein content. Estimates and associated 95% confidence intervals (solid dots with error bars) represent the magnitude of the genotype effect on intrinsic clearance. Units for estimates are (PAR/ mg microsomal protein/ μMol phylloquinone) x L/min. Significance codes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Reference group for CYP4F2 genotype = *1/*1 carriers; CYP4F diplotype = 13 / 13.

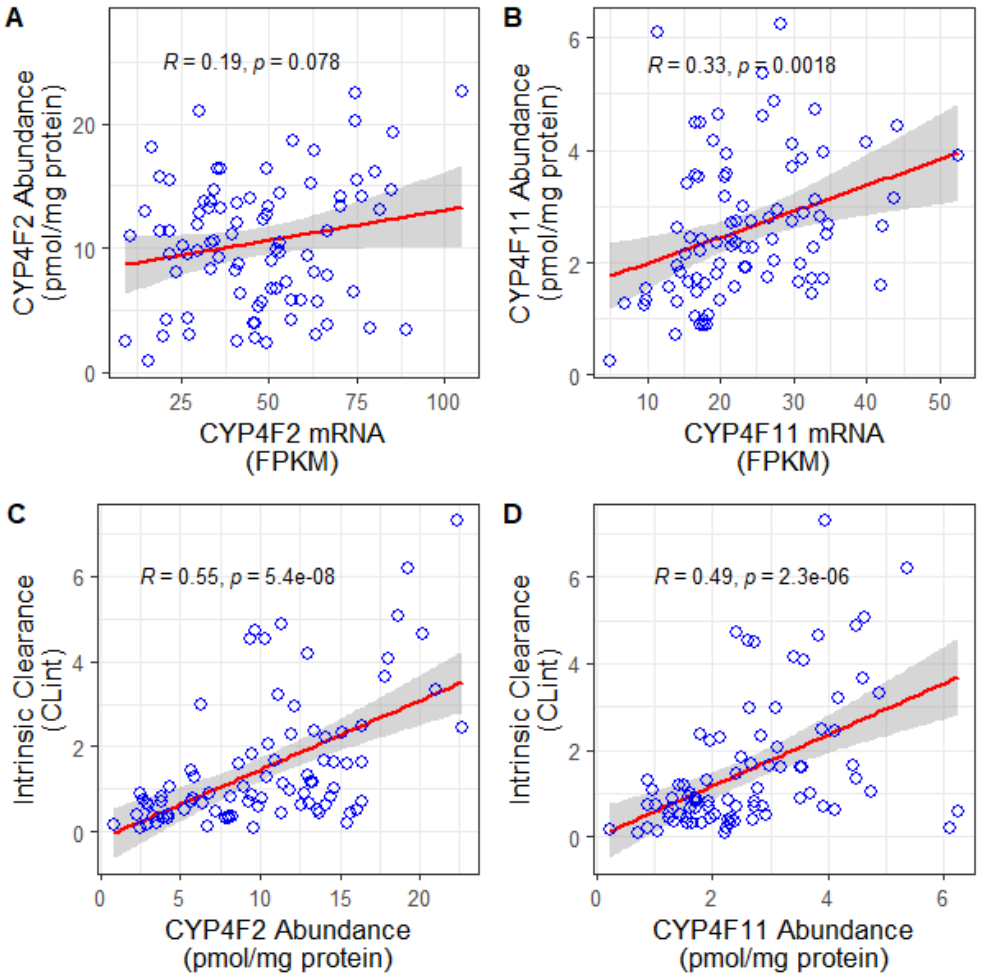

Figure 3.22. Pearson correlation plots across *CYP4F2* and *CYP4F11* mRNA and protein abundance derived from 88 human liver tissue donors.

(A-B) mRNA vs protein abundance, and (C-D) protein abundance vs. intrinsic clearance (CL_{int}). Blue circles represent observed data, solid red line represents unweighted linear regression fit, and grey band represents 95% confidence interval. Units for CL_{int} are (PAR/ mg microsomal protein/ μMol phylloquinone) x L/min.

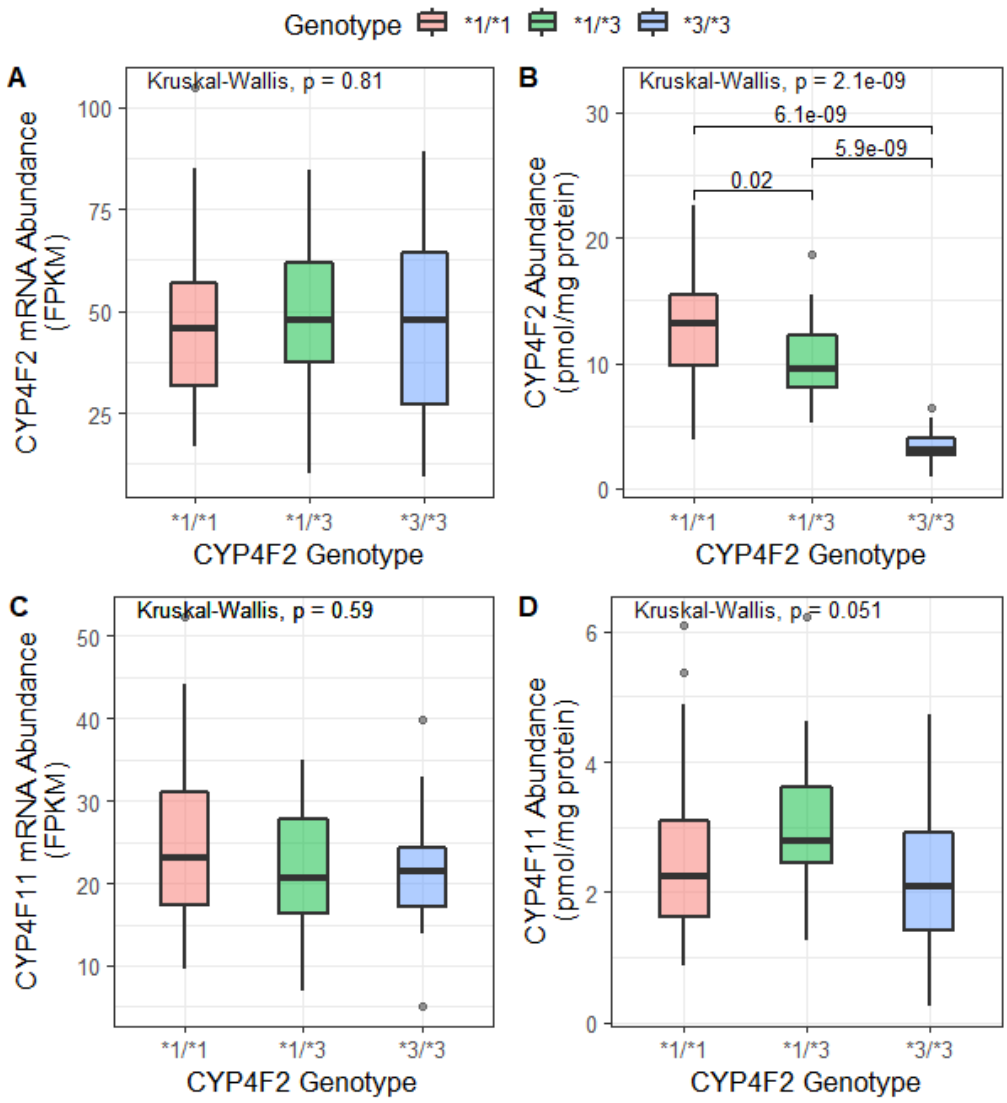

Figure 3.23. *CYP4F2* and *CYP4F11* mRNA and protein abundance derived from 88 human liver tissue donors and relationship to corresponding *CYP4F2*3* genotype (n: **1/*1* = 53, **1/*3* = 19, **3/*3* = 16). (A, and C) mRNA abundance; (B, and D) protein abundance. Analysis of variance between *CYP4F2* genotypes was performed using a Kruskal-Wallis test.

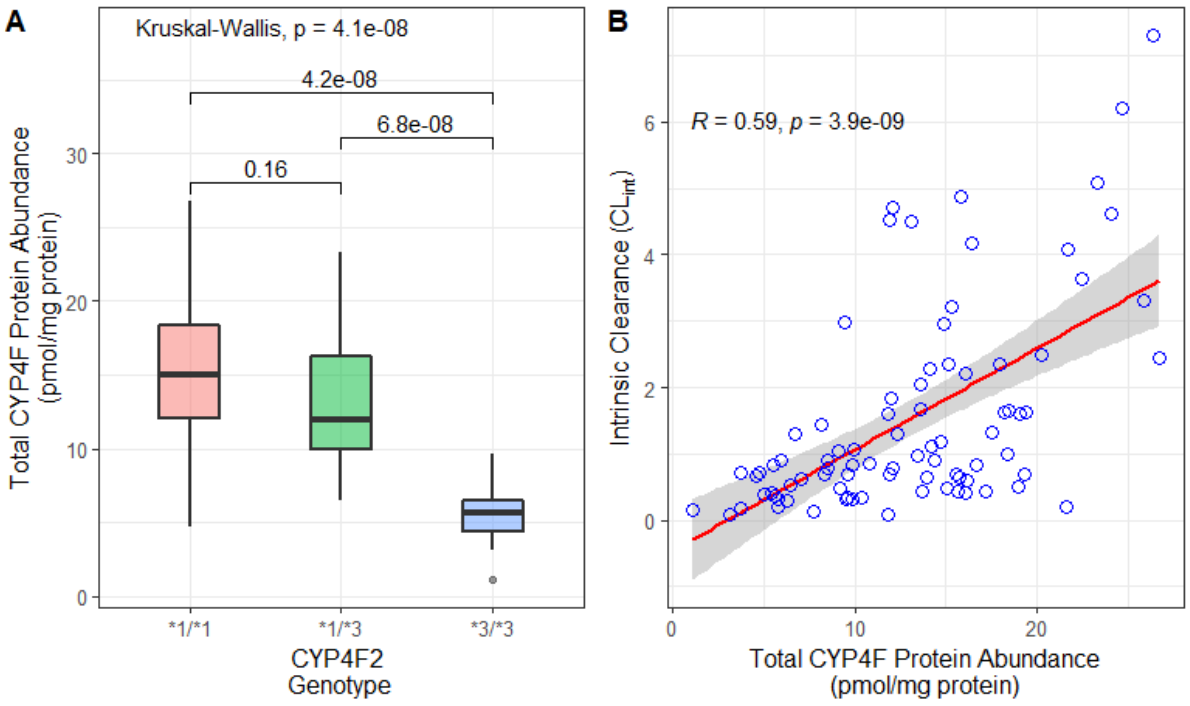

Figure 3.24. Combined CYP4F2 and CYP4F11 protein abundance, and associations with (A) *CYP4F2* genotype and (B) intrinsic clearance.

Samples were derived from 88 human liver tissue donors. Estimates for total protein abundance (mean ± SD) for each *CYP4F2*3* genotype are as follows: **1/*1*, 15.3 ± 5.3 (n=53); **1/*3*, 13.4 ± 4.2 (n=19), **3/*3*, 5.6 ± 2.1 (n=16) pmol/mg microsomal protein. Analysis of variance for protein abundance across *CYP4F2* genotypes was performed using a Kruskal-Wallis test. Blue circles represent observed data, solid red line represents unweighted linear regression fit, and grey band represents 95% confidence interval. Units for CL_{int} are (PAR/ mg microsomal protein/ μMol phylloquinone) x L/min**.**

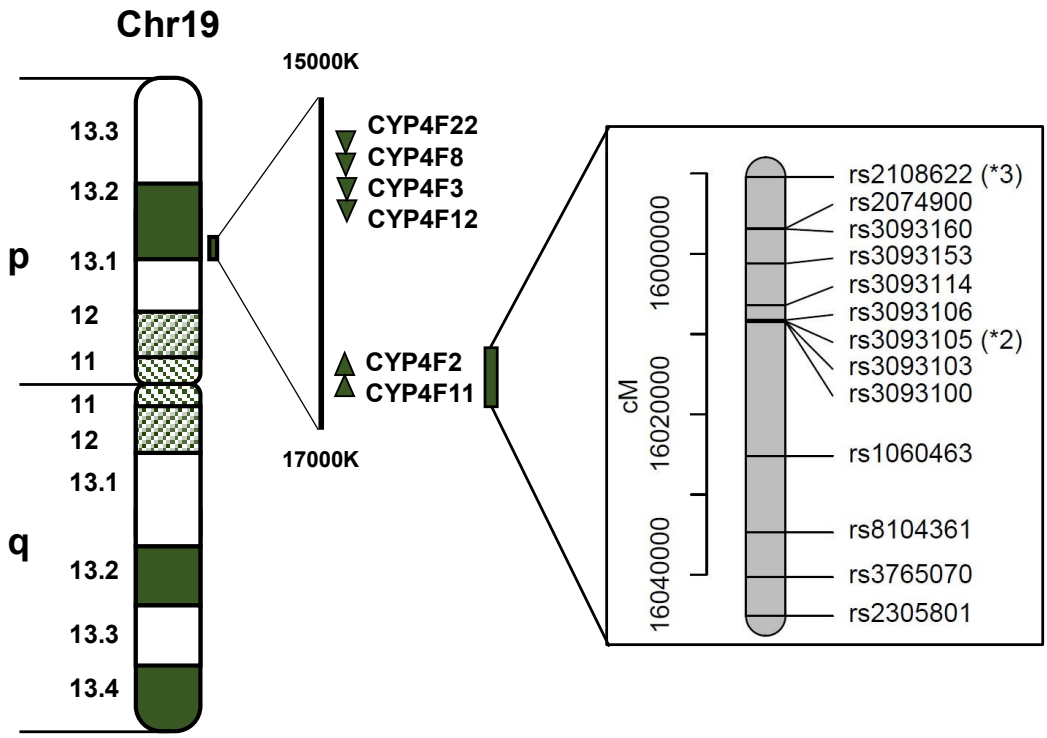

Figure 3.25. Localization of common *CYP4F2* and *CYP4F11* gene variants on chromosome 19.

Locations and directions of the human CYP4 family genes found in the NCBI human genome database are shown by arrowheads on the ideogram of human chromosome 19. The region containing CYP4F genes is magnified to display the relative location of common genetic variants (n=13) found across *CYP4F2* and *CYP4F11* in the UW and St Jude human liver bank samples. This figure is an adapted version of a figure previously published[112] by Kikuta et al., 2002.

Chapter 4. THE IMPACT OF *CYP4F2*3* VARIATION ON SYSTEMIC VITAMIN K STATUS FOLLOWING ORAL SUPPLEMENTATION IN HEALTHY ADULTS

4.1 ABSTRACT

In this chapter, we assessed the modifying effect of the singular *CYP4F2*3* variant on the relationship between dietary vitamin K exposure and short-term biomarkers of vitamin K status in healthy adults using a novel previously published LC-MS/MS-based assay. To do this, we conducted a pilot 10-day dietary intervention study testing the effect that vitamin K_1 supplementation (1 mg orally administered dose, n = 8 subjects) had on short-term biomarkers of vitamin K status – plasma vitamin K, and urinary K-acid catabolites. We then compared the impact that supplementation had on systemic vitamin K exposure (plasma AUC, $C_{ss\ trough}$, urine K-acid excretion) in *CYP4F2*1/*1* and **3/*3* homozygotes (n = 4 each). Overall, this pilot study aimed to evaluate both genotype-dependent effect on vitamin K exposure after dietary consumption, and the utility of using K-Acid catabolites as a quantitative biomarker of vitamin K status.

4.2 INTRODUCTION

Gene-environment interactions play an important role in determining an individual's predisposition to disease, and importantly therapeutic response[113,114]. Of these interactions, the interplay between dietary vitamin K (VK) consumption and warfarin pharmacodynamics remains an important area of research to be explored[44,115–117]. As mentioned in Chapter 1, warfarin is a commonly prescribed oral anticoagulant used to treat and prevent blood clots. Its pharmacological

effect is assessed by measuring prothrombin activity (INR) in the blood; however, this does not provide specific information about a patient's sensitivity to changes in the steady-state hepatic VK availability (VK status). Dietary depletion and repletion studies[45] show that long-term changes in VK consumption directly impact markers of hepatic and extrahepatic VK status; i.e., the ability to support the γ-carboxylation of Gla proteins such as Factor II. Variation in *VKORC1, GGCX* and *CYP4F2* genes can all affect the efficiency of that critical process and be modified by dietary (environmental) VK intake[42,118].

Gene-environment interactions involving *CYP4F2* are important, as the gene product controls the catabolic clearance of VK. Reduced function *CYP4F2*3* homozygous variant carriers are reported to need approximately an additional 1 mg/day of warfarin, on average, to achieve a therapeutic INR, compared to individuals homozygous for the reference allele[36]. It has been proposed that reduced metabolic clearance of VK will produce higher hepatic steady-state concentrations, at any given level of VK intake, antagonizing the pharmacological effect of warfarin[30]. Additionally, in a separate study, the *CYP4F2*3* variant was shown to significantly increase dietary VK_1 exposure in humans in an allele-dependent manner after single-dose supplementation[48]. Finally, a study conducted in patients treated with warfarin receiving care in an integrated health-care system[42] showed that the *CYP4F2*3* variant was associated with a 38% lower bleeding risk suggesting that carriers may have greater tolerance to environmental changes, such as altered dietary VK intake. While diet plays a major predictive role in determining hepatic VK status, and is thought to influence warfarin anticoagulation control[44], nutritional assessment tools such as food frequency questionnaires (FFQs) provide limited accuracy and may not capture all aspects of dietary consumption (i.e., such as portion sizes or cooking methods) that can directly impact measures of VK intake[47,119–121].

Assessing whole-body functional VK status with a single measure is challenging, because the unique distribution properties of VK_1 and menaquinone(s') plasma concentrations alone fail to reflect extrahepatic levels of VK. VK_1 appears more prominently in the blood whereas menaquinones concentrate in extrahepatic tissues[14]. However, both forms of VK undergo the same CYP4F2-mediated sequential ω-hydroxylation followed by successive rounds of β-oxidation resulting, ultimately, in the formation of the same two chain-shortened carboxylic acid metabolites (K-Acid I and II) as shown in **Figure 4.1**[30,54]. A previously established method for quantifying K-Acids in urine relies upon electrochemical detection (ECD), however, this method lacks the selectivity and sensitivity of mass spectrometry-based approaches. Recently, a novel LC-MS/MS assay was developed at the University of Washington that offers a minimally invasive yet highly sensitive alternative to ECD for monitoring urinary K-Acid levels in humans[55] and hence provides a robust method for evaluating systemic VK status.

The aim of this study was to assess the modifying effect of the *CYP4F2*3* variant on the relationship between dietary VK exposure and biomarkers of VK status in healthy human subjects. A 10-day dietary intervention study was conducted to test the effect that VK_1 supplementation has on short-term biomarkers of VK status (plasma VK, and urinary K-acid catabolites) using an LC-MS/MS-based approach. Here we compare the effect that supplementation has on systemic VK exposure (AUAC, C_{ss}, and C_{max}) in *CYP4F2*1/*1* and **3/*3* homozygotes in a healthy adult population.

4.3 Materials and Methods

4.3.1 *Materials*

Organic solvents and Taqman™ qPCR genotyping assays were purchased from Thermo Fisher Scientific (Waltham, MA). Menaquinone-1 (MK1), K-Acid I, and K-Acid II were previously

synthesized[54,122] and gifted to us by Dr. Allan Rettie, (University of Washington, Medicinal Chemistry). VK standards (VK_1 and MK-4) and corresponding internal standards (d_7-phylloquinone and d_7-menaquinone-4) were purchased from Sigma Aldrich (St. Louis, MO). An oral VK supplement was purchased through Nature's Life (Ponte Vedra Beach, FL). Colorimetric assay kits for urinary creatinine analysis were purchased from Cayman Chemical (Ann Arbor, MI). Phlebotomy supplies were purchased through BD Biosciences (San Jose, CA).

4.3.2 *Subject Recruitment and Study Design*

Healthy male and female adults were recruited by posted advertisements at the University of Washington Medical Center (UWMC), Institute for Translational Health Sciences website (www.ITHS.org), www.ClinicalTrials.gov, and a standalone website (www.VitKStudy.com). Participant information was recorded and stored using an electronic data capture tool (ITHS REDCap™). All participants were determined to be healthy based on a detailed questionnaire provided during an initial screening. Participants were automatically disqualified from enrollment if they met any of the following exclusion criteria: <18 years of age, inability to read or understand English, pregnant (or lactating), taking lipid lowering or anticoagulant medications, weighing <110 lbs. All qualified participants provided signed informed consent prior to enrollment. Upon completion of the study all participants were compensated for their time and participation. This study protocol was approved by the University of Washington Institutional Review Board (IRB), Human Subjects Division, Seattle, WA (IRB#: STUDY00001593), and registered at Clinical Trials.gov (NCT04450212).

The overall study was broken into two phases, Phase I: Genetic Screening, and Phase II: Supplementation. During the first phase, DNA from qualified subjects was collected via a buccal swab and extracted using a QIamp mini prep DNA isolation kit; additionally, demographic data

(age, weight, gender, ethnicity, and race) was collected. Buccal swab samples were screened for *CYP4F2*3* genotype using qPCR (TaqMan Kit). Subjects who were heterozygous carriers for the *CYP4F2*3* allele (**1/*3*) were excluded from participation in the supplementation phase of the study, and homozygous *CYP4F2*1/*1*) individuals were enrolled along with demographically matched (by age, sex, and weight) *CY4F2 *3/*3* homozygotes. Enrolled subjects were given a 1.0 mg oral dose of VK_1 (approximately 10-16 μg/kg bodyweight) with 2% milk for 9 consecutive days. A 10 mL fasted blood sample and a spot urine sample (50 mL) were collected between 6:00am and 8:00am from each participant on 7 of the 10 study days (day 1-5, 8, and 10).

4.3.3 *Plasma Vitamin K Analysis*

Prior to analysis, whole blood was collected into commercially available EDTA-treated tubes then centrifuged at 2,000 × g for 10 minutes at 4 °C. The plasma fraction was collected and stored at -80 °C until analysis. A 200 μL plasma sample (prepared in duplicate) was spiked with stable isotopically labelled internal standard (10 μL of 10 μg/mL d_7- VK_1 and d_7-MK4, respectively). Samples were extracted twice, by adding hexanes (2 x 1 mL), vortexing vigorously for 10 minutes, centrifuging at 3,500 × *g* for 5 minutes then combining the organic layers for solvent removal. Samples were dried under nitrogen gas and resuspended with 75 μL of ethanol prior to UPLC-MS/MS analysis.

4.3.4 *Urinary K-Acid Analysis*

VK-acid catabolites were extracted from urine and quantified using a previously established and validated LC-MS/MS method.[55] In summary, a 2.5 mL urine sample (prepared in duplicate) was spiked with internal standard (100 pmol of menaquinone-1; MK-1) and loaded onto a C_{18} SPE cartridge (Agilent Technologies, Bond Elut 3 mL), then washed with 2.5 mL of water and eluted

with 2 mL of methanol. Methanol eluted samples were acidified by adding 500 µL of concentrated hydrochloric acid, then shaken in a water bath at room temperature overnight under low-light conditions. Samples were diluted by adding 1 mL of deionized water and extracted into dichloromethane, which was then washed with 10 mL of water to remove traces of hydrochloric acid. Aglycone catabolites were converted into methyl ester derivatives through the addition of etheric diazomethane (in excess) then dried under a nitrogen gas. Samples were resuspended in 2 mL of hexane and loaded onto a normal phase Bond Elut LRC-Si 500 mg SPE column (Agilent), then, washed with 8 mL of hexane, and eluted with 10 mL of 15% ether in hexane. Samples were dried under a stream of nitrogen gas and dissolved in 100 µL of isopropyl alcohol prior to UPLC-MS/MS analysis. Urinary creatinine concentrations were used to normalize K-Acid concentrations in individual samples and were measured using a commercially available colorimetric based assay (Cayman Chemical, Ann Arbor, MI).

4.3.5 *UPLC-MS/MS Analysis*

Liquid-chromatography-mass spectrometry analyses were conducted on ACQUITY Ultra Performance LCTM (UPLC) system with integrated Waters autoinjector coupled to a Waters Xevo TQ-S tandem quadrupole mass spectrometer (Waters Co., Milford MA.). MS/MS analysis was performed using atmospheric pressure negative chemical ionization ($APCI^-$) with multiple reaction monitoring (MRM). The instrument source temperature was set at 150 °C with a probe temperature set to 650 °C. The following mass transitions (m/z) for each analyte were monitored on separate m/z channels: *m/z* 450 > 184 for VK_1 (phylloquinone), *m/z* 457.2>214.9 VK_1-d_7 (phylloquinone IS), *m/z* 444.4>185 VK_2 (MK4), and *m/z* 451.7>191.9 VK_2-d_7 (MK4 IS). The cone voltages were set at 14 V and 58 V for phylloquinone and MK4, respectively. Collision energies were set to 32 V and 30 V for phylloquinone and menaquinone 4, respectively. Analytes were

separated on an Acquity UPLC C18 1.7 μm, 100 × 2.0 mm column (Waters) with water (A) and methanol (B) mobile phases. The mobile phase gradient ran as follows: solvent B held at 65% for 0-2 min, ramped up to 98% between 2-3.5 min, held at 98% from 3.5-8.0 min, ramped down to 65 from 8-8.5 min, held at 65% from 8.5-9.5 mins. The column temperature was set to 50° C and mobile phase flow rate was 0.35 mL/min. Retention times for VK analytes are as follows: 4.98, and 4.8 for deuterated internal VK standards (VK_1-d_7, and MK4-d_7 , respectively); 5.00 and 4.81 for phylloquinone, and MK4, respectively (**Figure 4.2**).

Urinary K-acids were also analyzed by liquid-chromatography-mass spectrometry using the same system as described previously. The instrument source temperature was set at 150 °C with a probe temperature set to 500 °C. The following mass transitions for the methyl ester derivatives of each metabolite were monitored: *m/z* 240 > 185 (MK1, internal standard), m/z 312 > 195 (K-Acid I), m/z 284 > 195 (K-Acid II). The following settings were applied for each of the analytes: cone voltage set at 50 V for K acid I, II, and 60 V for MK1-acid; collision energies set at 40, 30, and 18 eV, respectively, for K acid I, II, and MK1. Analytes were separated on a Shim-pack XR-ODS 2.2 μm, 2.0 × 75 mm UPLC column (Shimadzu Scientific, Columbia, MD) at a set temperature of 50 °C with 0.10% formic acid (A) and methanol (B) mobile phases after a 10 μL injection (flow-rate set to 0.35 mL/min). The mobile phase gradient ran as follows: solvent B at 65% held for 3 mins, ramped up to 98% between 3-7 min, held at 98% from 7-12 min, ramped down to 65 from 12-13 min, and held at 65% from 13-14 min. Retention times for K-acid analytes are as follows: MK1-acid (internal standard), 6.01 mins; K-acid I, 4.86 mins; and K-acid II, 3.54 mins (**Figure 4.3**).

A linear standard calibration curve with a range of 0.05–300 ng/mL (r^2>0.99) was established for both phylloquinone and MK4. Similarly, a linear standard calibration curve with a

range of 0.3–300 ng/mL (r^2>0.99) was established for both K-acid I and K-acid II. Standard curves were prepared in phosphate buffered saline solution. VK_1, MK4, K-acid I, and K-acid II analytes were quantified by comparing peak area ratios relative to their internal standards (VK_1-d_7 and MK4-d_7) for phylloquinone and menaquinone, respectively; and MK1-acid for K-acid I and II to ratios from the appropriate calibration curve using linear regression. Intra assay coefficients of variation (CV) for VK_1, MK4, K-Acid I and K-Acid II were less than 12%; while inter assay CV across analytes was less than 17%. The lower limit of quantification (LLQ) for K-Acid I and K-Acid II was 5 pmol/mL urine, while MK4 and VK1 was 50 pmol/mL plasma.

4.3.6 *Pharmacokinetic Analysis*

Pharmacokinetic parameters were estimated by non-compartmental analysis via the PKNCA package[123] using R (v4.2.0) in R Studio (RStudio 2023.03.0, Build 386). The area under the concentration versus time curve (Area under the curve, AUC; and area under the amount curve, AUAC) was evaluated directly from observed plasma or urine data to the last measurable time point using the linear trapezoidal rule. Steady-state plasma trough concentration (C_{ss}) was estimated by taking the average of the last 4 concentration time points obtained on days 4, 5, 8, and 10. An apparent clearance (CL/F) was estimated according to **Equation 14**, where R_0 represents dosing frequency (ng • hr^{-1}), and C_{ss} represents steady state concentration (ng • L^{-1}).

Equation 14

$$\frac{CL}{F} = \frac{R_0}{C_{ss}}$$

A urinary K-Acid-to-plasma VK_1 ratio (catabolite excretion ratio) was estimated as an indirect marker of metabolic formation clearance. Upon reaching steady-state (days 4-10), the

partial area under the total creatine normalized urinary K-acid concentration-time curve (AUAC) was divided by the partial AUC of the VK_1 concentrations-time curve (**Equation 15**).

Equation 15

$$Catabolite\ Excretion\ Ratio = \left(\frac{Total\ Normalized\ Urinary\ K_{Acid}\ AUAC_{Day\ 4-10}}{Plasma\ vitamin\ K_1\ AUC_{Day\ 4-10}}\right)$$

4.3.7 *Statistical Analysis*

All statistical analyses were performed with R (v4.2.0) in R Studio (RStudio 2023.03.0, Build 386). Mean comparisons of all parameters were carried out using the Wilcoxon rank sum test. To visualize variability, data estimates in figures, tables and text are expressed as mean values ± SD unless otherwise stated. Associations between K-acid excretion and plasma VK_1 levels were evaluated with a multivariate linear regression model. To account for heteroscedasticity, robust standard errors were calculated using the lmtest package in R. Linear regression models were compared using the following objective function values: residual sum of squares (RSS), AIC, and log-likelihood.

4.4 RESULTS

4.4.1 *Population Characteristics*

A total of 106 subjects were screened for *CYP4F2*3* genotype; frequencies were as follows: **1/*1,* 53% (n = 56); **1/*3*, 37% (n = 39); **3/*3*, 8% (n = 8), 4 of which were lost to follow-up); no genotype call, 2% (n = 3). Of the 106 subjects, 8 (4 homozygous reference and 4 homozygous for the **3* allele) were enrolled in phase II of the study (2 female subjects, and 6 male subjects), all of whom received 9 oral doses of VK_1 and provided blood and urine samples on 7 of the 10 study

visit days (**Figure 4.5**). There were no statistically significant differences in age or weight between the two genotype groups (**Table 4.1**).

4.4.2 *Impact of CYP4F2*3 on Plasma Vitamin K_1 Pharmacokinetics*

To assess the modifying effect that the *CYP4F2*3* variant has on VK status, plasma concentrations of VK were measured over 9 days of supplementation. VK_1 plasma concentrations were elevated post-supplementation (**Figure 4.6**), with geometric mean baseline concentrations of 0.39 ng/mL 95% CI [0.13-1.17] pre-supplementation vs 0.77 ng/mL 95% CI [0.38-1.53] post-supplementation. Overall VK_1 exposure ($AUC_{Day\ 1\text{-}10}$) was 2.1-fold higher for the *CYP4F2 *3/*3* group compared to the **1/*1* group, also a 26% higher baseline VK_1 (C_0) concentration, and a 37% higher baseline adjusted C_{ss} were observed for *CYP4F2 *3/*3* homozygotes compared to **1/*1* homozygotes (**Table 4.1**, **Figure 4.6**). Additionally, there was a significant difference in post steady state $AUC_{Day\ 4\text{-}10}$ between genotype groups (p=0.02), where **3/*3* homozygotes had 2.7-fold higher AUC than **1/*1* homozygotes (**Table 4.1**).

4.4.3 *Impact of CYP4F2*3 on Urinary K-Acid Excretion.*

Urinary K-acid catabolites were measured as an additional marker of VK status. To account for differences in spot urine dilution between subjects, measured K-acid catabolite levels were normalized to urinary creatinine concentrations. Similar to VK_1 concentrations, there was an increase in total K-acid excretion post supplementation, with geometric mean baseline concentrations of 7 ng/g creatinine 95% CI [4-16] vs 25 ng/mg creatinine 95% CI [9-74] post supplementation. Of the two K-acid catabolites, K-acid II was more sensitive to changes in dietary VK_1 intake, with an overall geometric mean fold change from baseline of 3.9-fold vs 2.7-fold for K-acid I (**Figure 4.7**). This observation is likely because unlike K-acid I, which is an intermediate

metabolite, K-acid II is the terminal metabolite in the VK catabolic pathway, and its abundance in the urine is controlled predominantly by its formation clearance. There were no statistically significant differences in $AUAC_{Day\ 1\text{-}10}$, $AUAC_{Day\ 4\text{-}10}$ (steady state AUAC), and C_0 for K-Acid I, II, and total K-acids when stratified by *CYP4F2* genotype (**Figure 4.8**, **Table 4.2**). At steady-state (days 4-10), creatinine normalized total K-acid-to-plasma VK_1 AUAC/AUC ratio was higher for **1/*1* homozygotes at 0.056 ± 0.04 vs 0.020 ± 0.01 for **3/*3* homozygotes (**Table 4.1**), however the variability in the data did not support a statistically significant difference (p=0.38).

Additionally, the association between creatinine normalized K-acid concentrations and plasma VK_1 levels were evaluated using univariate and multivariate linear regression models. The multivariate linear regression model, adjusted for both K-acid I and K-acid II levels, performed better as a predictor of plasma VK_1 concentration than the univariate model using total K-acid concentrations as a predictor. However, further evaluation of the multivariate model revealed that individual K-acids were not a significant predictor of plasma VK_1 concentrations after adjusting for variability at the individual level (**Table 4.3**).

4.5 DISCUSSION

VK status is a difficult variable to control during the maintenance phase of warfarin therapy. This is because VK consumption is a highly individualized parameter that depends heavily on the types of foods being consumed, methods of preparation, frequency of consumption, and quantity. Additionally, not all VK is created equal, as VK_1 and VK_2 have different physiochemical properties that lead to their distinct distribution patterns and even apparent activity[61,124]. This is further confounded by the fact that there is significant inter-individual variability in VK metabolism (and hence exposure), a large component of which is driven by genomic factors such as the *CYP4F2*3* variant. Moreover, assessing systemic VK status also poses a challenge, as

commonly used biomarkers such as PIVKA-II, and osteocalcin are retrospective in nature, and mainly serve as a long-term tissue-dependent indicator of VK related pharmacological activity, which does not necessarily inform directly on total VK levels or metabolism itself.

In this study we evaluated the modifying effect that the *CYP4F2*3* variant has on both plasma and urinary biomarkers of VK status by conducting a 10-day VK_1 supplementation study. There was an increase in plasma VK_1 concentrations over the first 3-days of supplementation before reaching a steady-state, however, a marked decrease in plasma VK_1 concentrations was observed after day 5 followed by a rebound in concentration at day 8. This dip in plasma concentrations was more prominent for the reference group (**1/*1*) than it was for the variant group (**3/*3*). A plausible reason for such a decline could be poor subject compliance, as the period of time where a decline in plasma concentrations was observed was also the same period of time where participants were required to perform supplementation from home (days 6, and 7). This could be caused by inadequate consumption of fat containing milk at the time of supplementation (which helps improve the absorption of VK), or the complete lack of supplement consumption. Notably, this decline was not readily apparent in urinary K-acid excretion levels, suggesting that while VK_1 levels in the plasma depleted rapidly there still remained an available pool of VK to be metabolized and eliminated as K-acid. This available pool is presumably from extrahepatic stores of VK_2.

Although VK_1 is a major dietary form of VK (>90%), MK4 exists in the highest concentrations in tissues of animals and humans[53,125]. It is thought that menaquione-4 is the major storage form of VK and helps support extrahepatic VK activity. Additionally, there is growing body of evidence that dietary VK_1 can be converted into MK4 *in-vivo*[125,126]. To determine whether this conversion would be observable in our system we monitored circulating MK4 levels in the

plasma over the course of supplementation. Unfortunately, circulating MK4 levels were at or below the limit of quantification and did not increase over the course of supplementation (**Figure 4.9**). It has been proposed that MK4 biosynthesis can occur through the release of menadione from VK_1 in the intestine followed by the prenylation with geranylgeranyl phosphate forming MK4 in tissues[15]. This pathway of biosynthesis likely results in no detectable change in circulating MK4 levels as it is the precursor (menadione) that we would expect to begin to accumulate in the plasma, which in part helps explain our results.

At large, our study was subject to a variety of limitations. Due to disruptions in subject recruitment and screening, as a direct consequence of the COVID-19 pandemic, we were unable to complete enrollment resulting in a lack of power for comparison of study arm (**1/*1* vs **3/*3*) results. Additionally, variability in urinary creatinine levels that are not associated with dilution effects (i.e., extensive exercise/strenuous activity, and high protein diet) could have potentially led to underestimation (or overestimation) of urinary K-acid levels; however, creatinine levels across all subjects and timepoints fell within a normal range when stratified by age groups (data not shown). Study participants were informed to avoid taking multivitamins or additional VK containing supplements prior to initiation of phase II of the study, however, data regarding eating habits or preferences was not captured, hence measured values were not controlled for day-to-day dietary VK consumption.

In summary, the present study shows that the homozygous carriage of the *CYP4F2*3* variant is associated with decreased metabolic clearance of VK_1. Additionally, the novel UPLC-MS/MS assay[55] for quantifying VK aglycone catabolites (K-acid I, and K-acid II) was effective at monitoring changes in metabolite excretion over-time given a change in dietary VK exposure. Future directions evaluating the association between urinary K-acid levels and plasma biomarkers

of VK pharmacodynamics may help provide insights on the positive predictive value of urinary K-acids as a biomarker of systemic VK status.

4.6 TABLES AND FIGURES

Table 4.1. Summary of vitamin K_1 pharmacokinetic parameters by *CYP4F2* genotype.

Parameter	***CYP4F2* Genotype**		***p*-value**[†]
	**1/*1* (n=4)	**3/*3* (n=4)	
Age (years)	45 ± 23	25 ± 3	0.25
Weight (lbs)	175 ± 27	147 ± 16	0.11
AUC[††] (ng·hr/mL)	80.2 ± 22.2	167 ± 46.3	0.06
(partial $AUC_{Day4-10}$)	**(34.6 ± 11.4)**	**(94.2 ± 30.8)**	**(0.02*)**
C_0 (ng/mL)	0.39 ± 0.42	0.490 ± 0.400	0.75
C_{ss} VK_1 (ng/mL)	0.61 ± 0.30	0.790 ± 0.260	0.40
CL/F (L/hr)	78.1 ± 29.9	57.2 ± 19.6	0.40
C_{ss} Urinary K-Acids (ng/g CRE)	0.020 ± 0.010	0.019 ± 0.010	0.93
Urinary K-Acid/Plasma VK_1 [‡]	0.056 ± 0.040	0.020 ± 0.010	0.38

[†] *p-value reported for Wilcoxon Rank Sum Test between genotype estimates.*

[††] *AUC determined based on trough sampling from days 1-10 (partial AUC, days 4-10).*

‡ *Parameter was estimated using partial urinary K-Acid (CRE adjusted, AUAC) and Plasma VK_1 AUC values from day 4-10. Abbreviations: AUC = Area under the curve; AUAC = Area under the amount curve; C_0 = Baseline plasma concentration; C_{ss} = Steady-state concentration (trough); CRE = Creatinine. Data are expressed as mean value ± SD.*

Table 4.2. Summary of urinary K-acid pharmacokinetic parameters by *CYP4F2* genotype.

Analyte‡	**AUAC**†† (ng·hr/g creatinine)			**Baseline plasma C_0** (ng/g creatinine)		
	****1/*1*** (n=4)	****3/*3*** (n=4)	***p*-value**†	****1/*1*** (n=4)	****3/*3*** (n=4)	***p*-value**†
K-Acid I (Day 4-10)	1230 ± 1204 (723 ± 753)	525 ± 116 (233 ± 252)	1.00 (0.80)	2.66 ± 1.91	0.84 ± 0.85	0.40
K-Acid II (Day 4-10)	1845 ± 293 (1029 ± 171)	2772 ± 1390 (1671 ± 1408)	0.80 (0.80)	4.06 ± 1.31	7.19 ± 3.32	0.67
Σ K-Acids (Day 4-10)	3075 ± 911 (1752 ± 582)	3358 ± 1431 (1963 ± 1588)	1.00 (1.00)	6.72 ± 0.60	8.04 ± 3.79	1.00

† *p-value reported for Wilcoxon Rank Sum Test between genotype estimates.*

††*AUAC determined based on daily spot urine sampling from days 1-10, while partial AUAC denote by parenthesis represents spot urine sampling from days 4-10.*

‡ *Parameters were estimated using analyte values that were normalized to urinary creatinine levels.*

Abbreviations: AUAC = Area under the amount curve; C_0 = Baseline plasma concentration.

Data are expressed as mean value ± SD.

Table 4.3. Model comparison: Total K-acids vs. K-acid II adjusted for K-acid I as predictors for plasma VK_1 concentrations.

Coefficient‡	**Univariate** *(Total K-Acid)*		**Multivariate** *(K-Acid I & II, adjusted by Subject)*	
	Estimate	*SE*	*Estimate*	*SE*
(Intercept)	0.245	(0.151)	0.313 *	(0.114)
Total Urinary K-Acid	13.2	(7.87)		
Urinary K-Acid II			8.86	(12.7)
Urinary K-Acid I			15.82	(38.9)
logLik	-12.8		1.21	
AIC	31.6		13.57	

**** p < 0.001; ** p < 0.01; * p < 0.05.*
p-value reported from a Wald Chi-Squared Test with robust standard errors.
‡ Parameter was estimated using values that were normalized to urinary creatinine levels
Abbreviations: logLik = Log-Likelihood; AIC, Akaike information criterion.

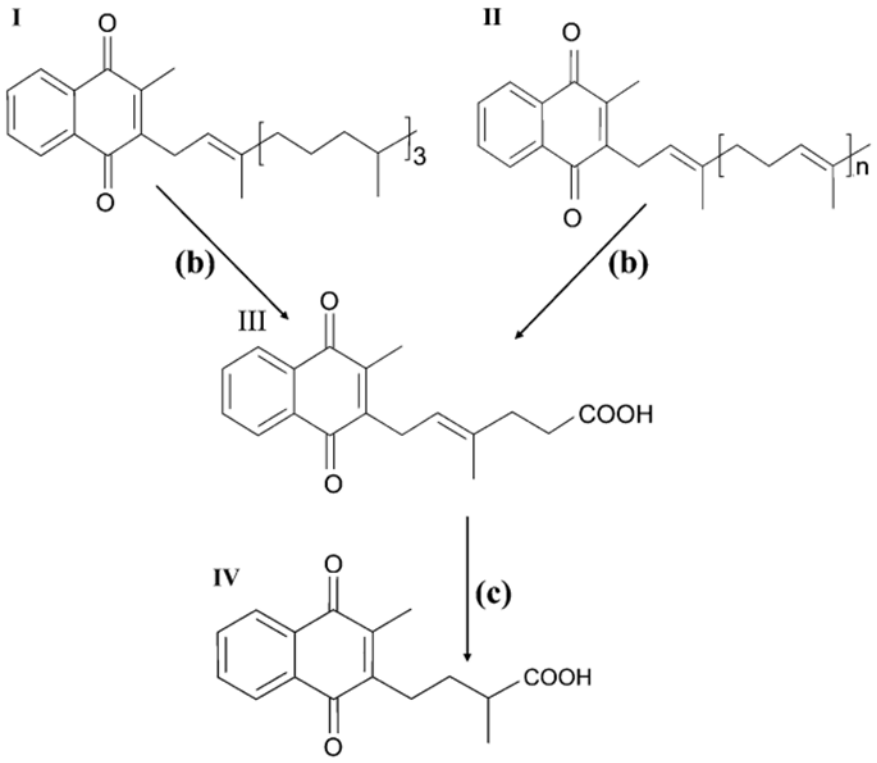

Figure 4.1. Structural formulas of phylloquinone (VK_1), menaquinones (VK_2) and their predominant urinary aglycone metabolites.

(I) VK_1 (II) VK_2; (III) 2-methyl-3-(5`-carboxy-3`-methyl-2`-pentenyl)-1, 4-naphthoquinone (7C-aglycone metabolite of I and II, "K-Acid I"); (IV) 2-methyl-3-(3`-3`-carboxymethylpropyl)-1,4-naphthoquinone (5C-aglycone metabolite of I, II, "K-Acid II"). (b) ω-oxidation and subsequent β-oxidation of the isoprenoid side chain; and (c) further β-oxidation. This scheme was adapted from a figure published by Harrington et al., 2007[127].

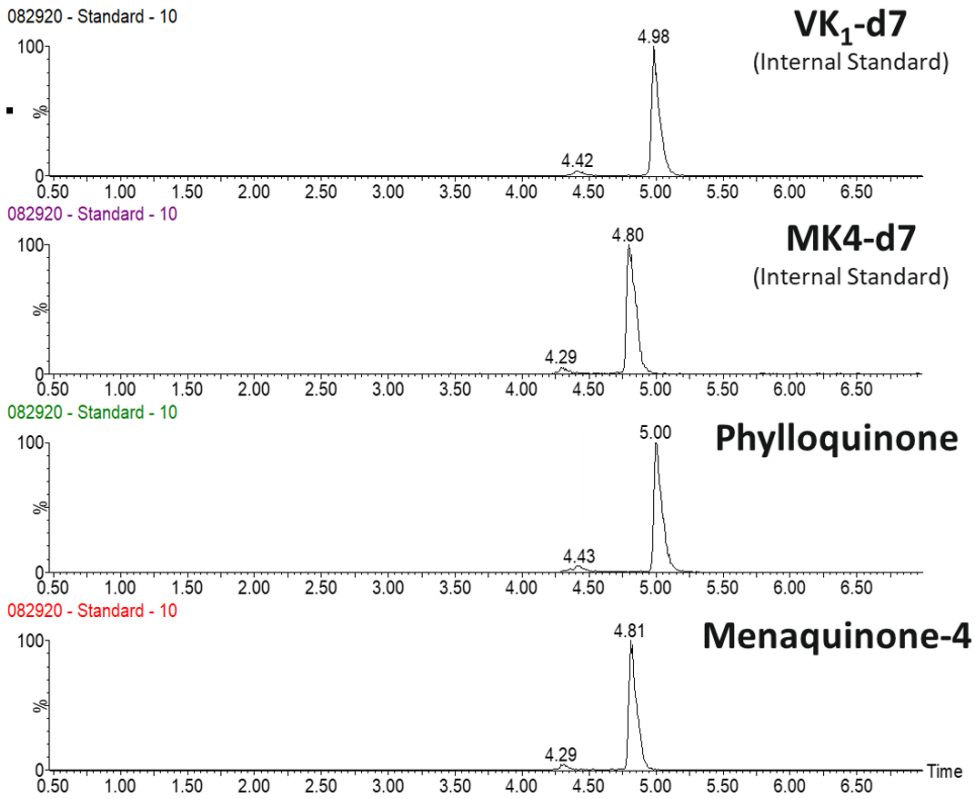

Figure 4.2. UPLC-MS/MS chromatograms of Phylloquinone (VK_1) and Menaquinone-4 (MK4) standards and corresponding internal standards (VK_1-d_7, and MK4-d_7).

Chromatograms represent MRM channel responses of VK analytes after a 10 µL injection of a 10 ng/mL (100 pg) standard cocktail solution.

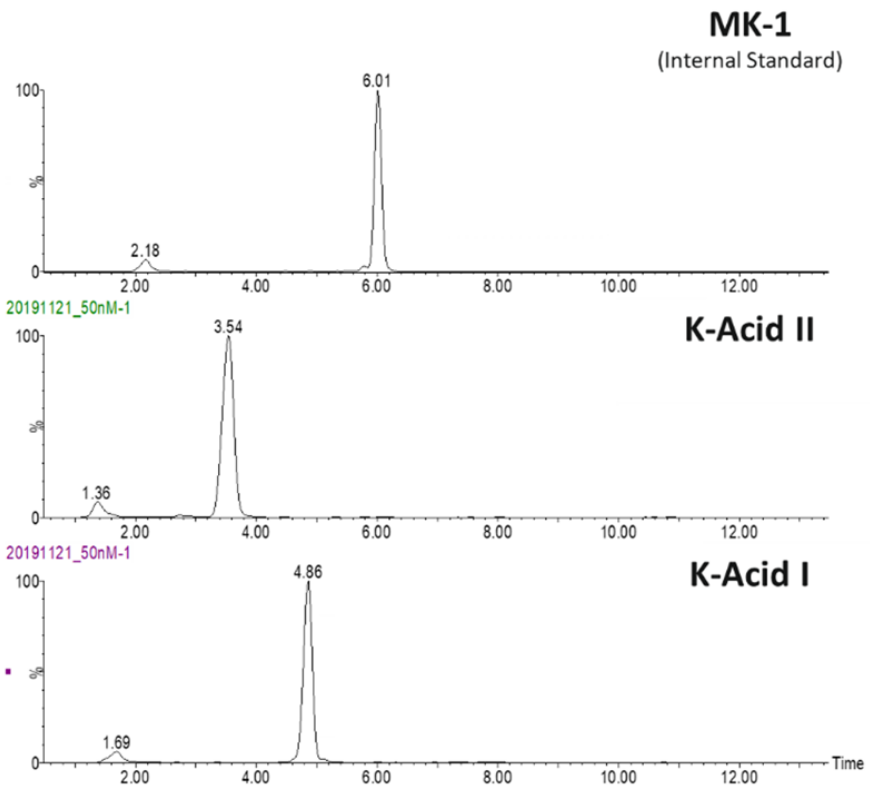

Figure 4.3. UPLC-MS/MS chromatograms of K-Acid I and K-Acid II standards and corresponding internal standard MK1-acid.

Chromatograms represent MRM channel responses of K-acid analytes after a 10 µL injection of a 50 nM (500 fmol) standard cocktail solution.

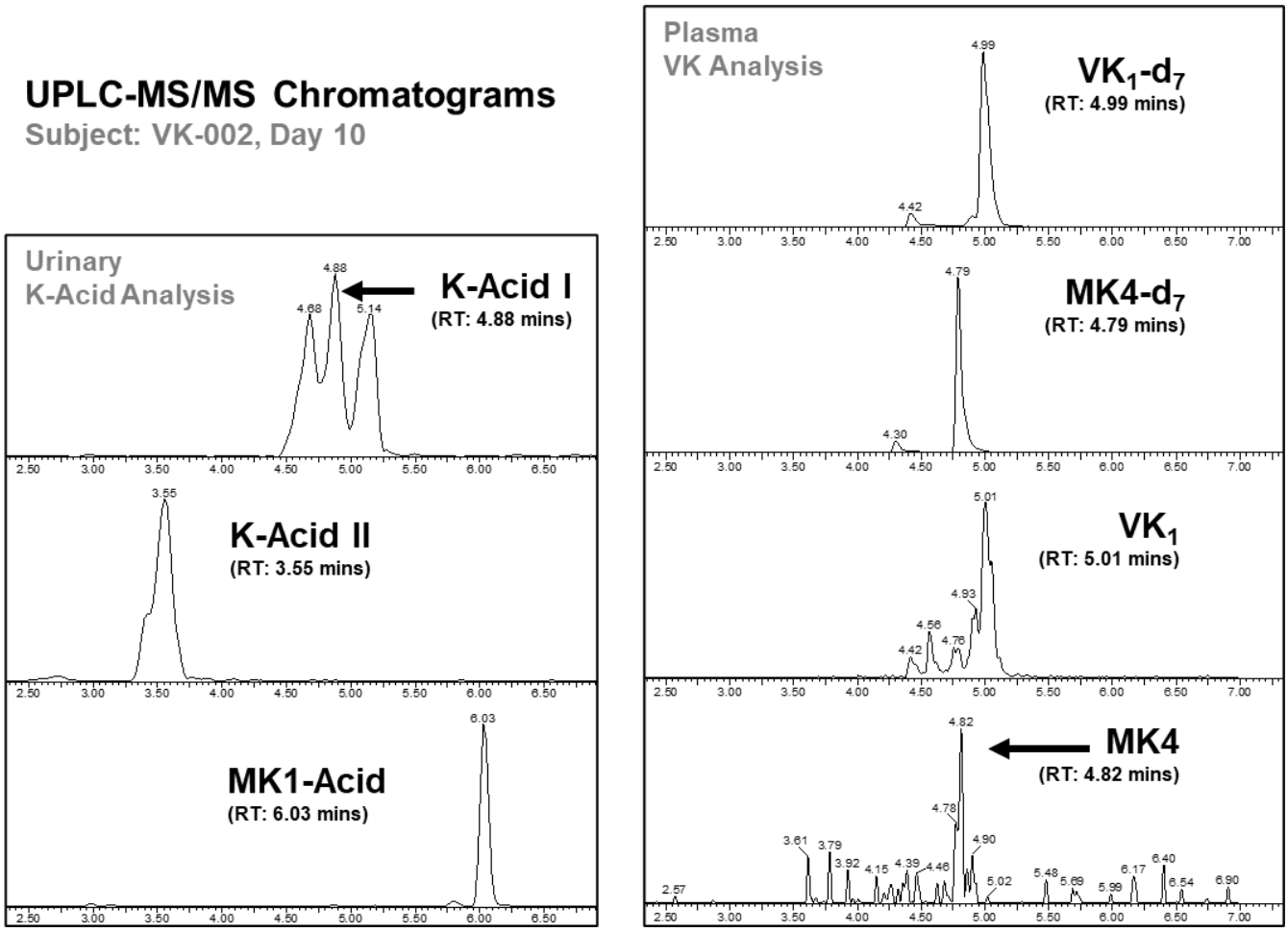

Figure 4.4. Representative UPLC-MS/MS chromatograms of VK and K-acid metabolites from a single study participant (VK-002) post-supplementation.

Left panel: urinary K-acid catabolites; Right panel: plasma VK analytes. Chromatograms represent MRM channel responses of K-acid and VK analytes after a 10 μL injection of extracted sample.

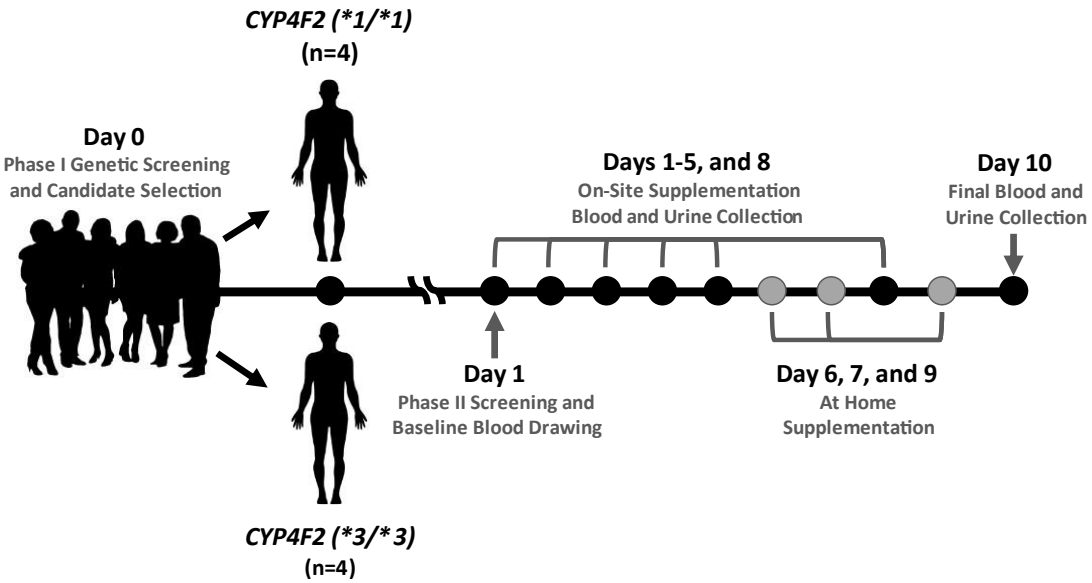

Figure 4.5. Healthy volunteer vitamin K supplementation study design.

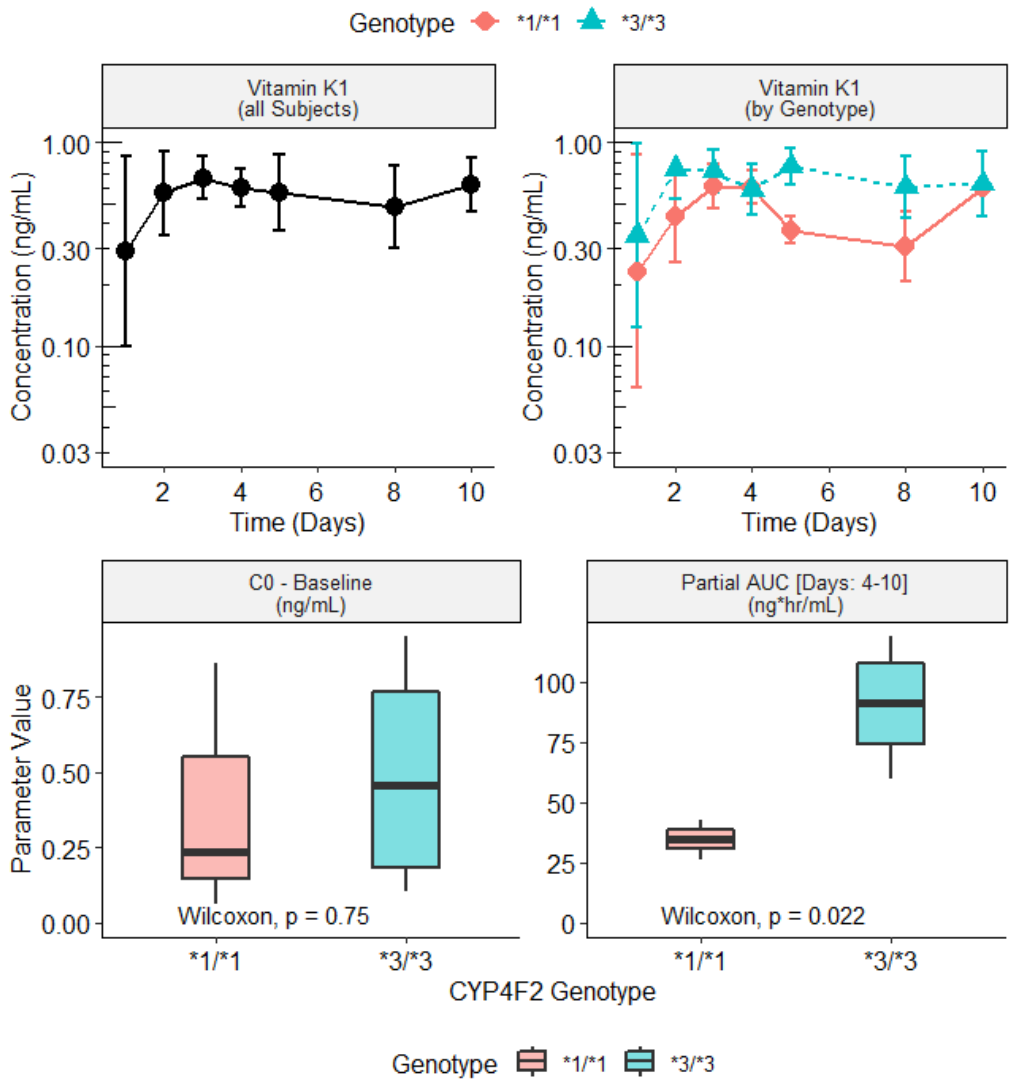

Figure 4.6. Plasma VK_1 concentration vs time profiles.

Top row: Concentration time profiles for VK_1 across all subjects (left) and stratified by *CYP4F2* genotype. Bottom row: distribution of baseline plasma VK_1 concentrations(left) and partial AUC (right) grouped by *CYP4F2* genotype; boxplots represent the median and interquartile range of the data. Partial AUC is based on trough sampling from days 4-10. Data was obtained from 8 study participants (n = 4 per genotype).

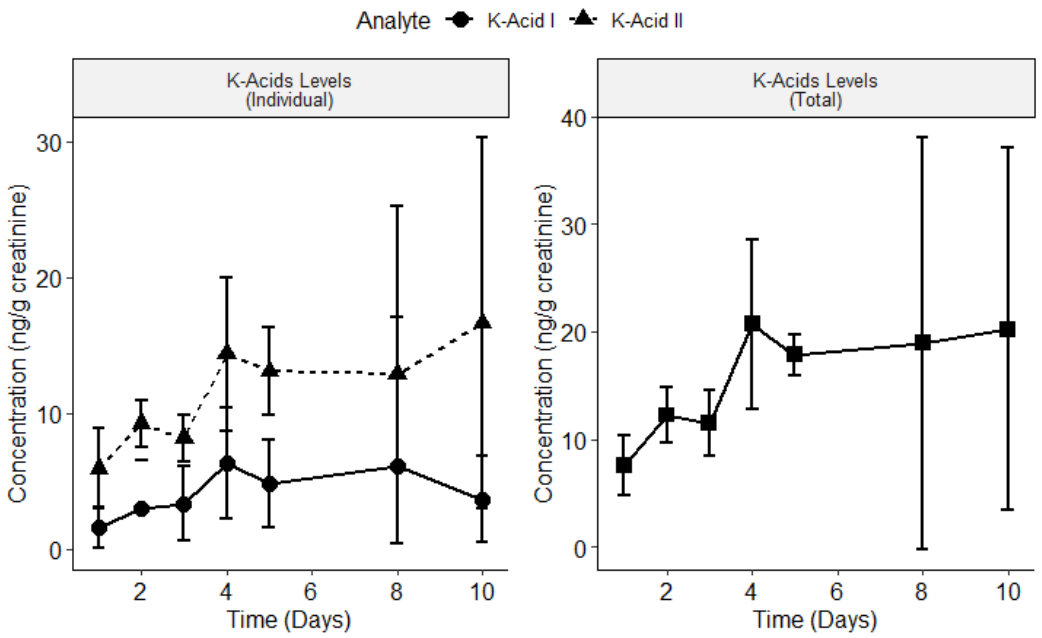

Figure 4.7. Urinary K-acid concentration vs time profiles.

Left panel: Individual K-acid concentration time profiles for K-acid I (solid line with solid circles), and K-acid II (dotted line with solid triangles). Right panel: Total K-acid (K-acid I + K-acid II) concentration time profile (solid line with solid squares).

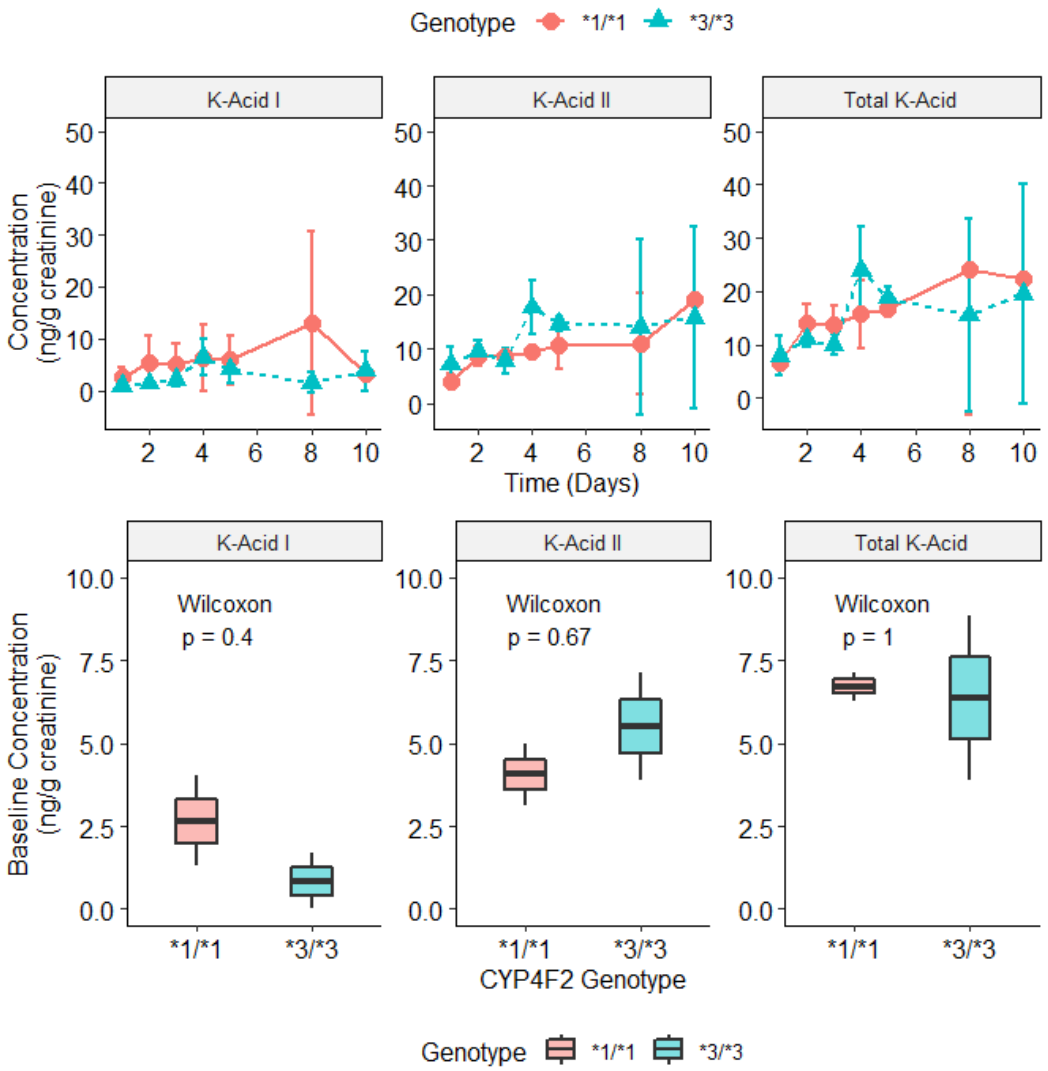

Figure 4.8. Urinary K-acid concentration vs time profiles stratified by *CYP4F2* genotype.

Top row: Urinary concentration over time profiles for K-Acid I, K-Acid II, and total K-Acids (K-Acid I + K-Acid II) stratified by *CYP4F2* genotype. Bottom row: corresponding distribution of baseline urinary K-Acid concentrations by *CYP4F*2 genotype. Boxplots represent the median and interquartile range of the data. Data was obtained from 8 study participants (n = 4 per genotype).

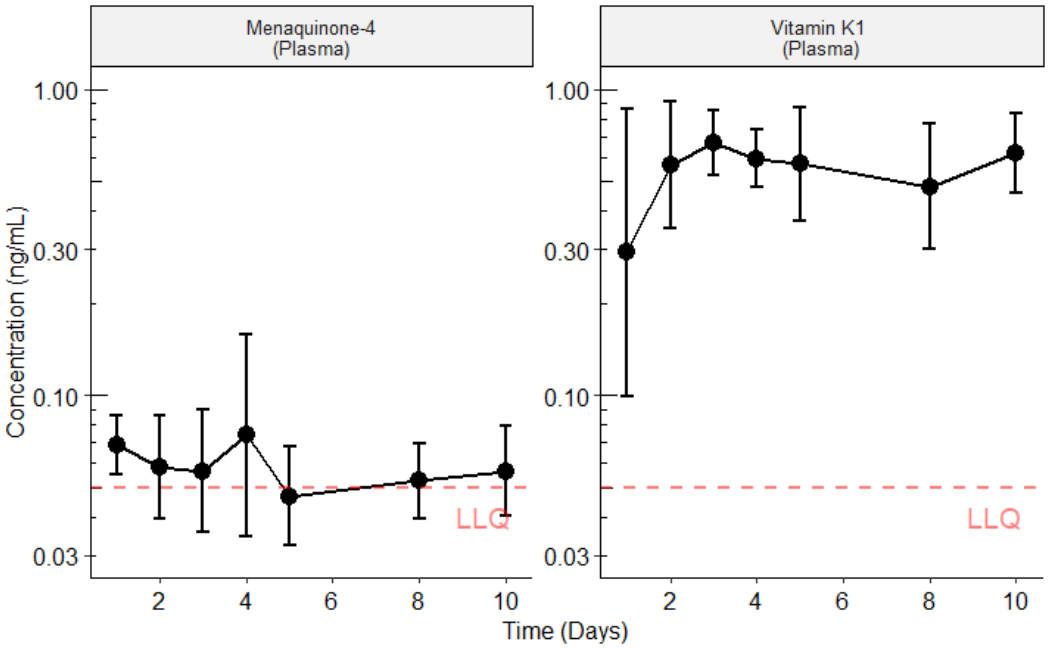

Figure 4.9. VK_1 plasma concentration vs. time profile.

Left panel: MK4 concentration vs. time profile. Right panel: VK_1 concentration vs. time profile. Red dotted line represents lower limit of quantification (LLQ).

Chapter 5. CONCLUSIONS AND FUTURE DIRECTIONS

5.1 CONCLUSIONS AND FUTURE DIRECTIONS

Over the past 60 years, the field of pharmacogenomics has become a cornerstone of precision medicine and is a significant contributor to our understanding of interindividual variability in drug response[128]. Importantly, gene variation-environment interactions can be a key determinant of one's susceptibility to a disease or adverse drug reaction[113,114]. One of the major goals of precision medicine is to identify and understand the genetic and environmental factors that contribute to illness and improve the prevention, diagnosis, and treatment of disease[65]. This dissertation focuses on understanding the influence that genetic variability in VK metabolic pathways has on VK status – the capacity to support VK-dependent metabolic reactions. It applies novel tools to characterize the relationship between complex genomic traits and measures of VK metabolism *in vitro,* and examines through a pilot study the modifying effect that the *CYP4F2*3* variant has on measures of VK exposure following dietary intake *in vivo*.

CYP4F2 and CYP4F11 dominate VK catabolism and are major determinants of hepatic VK status[30,32], regulating hepatic VK concentration and the synthesis of VK-dependent γ-carboxylation of select clotting factors.[47,118,129] VK status contributes significantly to inter-individual differences in pharmacodynamic response to the oral anticoagulant drug warfarin[36,37,41,44,115,130,131]. Although much progress has been made elucidating the effects that single allele genetic variation of VK metabolic pathway genes have on the warfarin dose-response relationship, the impact that gene haplotype combinations (diplotypes) have on mRNA abundance, protein abundance, and hepatic intrinsic clearance has yet to be rigorously evaluated. The specific aims of this dissertation were three-fold: (1) Develop a method to efficiently probe the effects of

complex genomic traits on drug metabolism *in vitro* at a population level; (2) Reassess the relationship between *CYP4F2* and *CYP4F11* genetic variability and VK metabolism *in vitro* using a haplotype/diplotype-based approach; and (3) Investigate the modifying effect of *CYP4F2*3* on biomarkers of systemic VK status *in vivo* following oral VK consumption.

The use of human liver microsomes as a model system to evaluate the impact of complex pharmacogenomic traits on the efficiency of drug metabolism has challenges. To accurately predict the true effect size of phenotypic traits that are modified by gene variation typically requires large, richly sampled datasets representative of the study population. Moreover, the acquisition of this data can be labor-intensive if the study design or bioanalytical methods are not high throughput and is potentially unfeasible if the abundance of tissue needed for experiments is limited. In Chapter 2, we developed a novel strategic approach using non-linear mixed effects (NLME) models to determine enzyme kinetic parameters for individual liver specimens using sparse data; an approach we have referred to as population Michaelis-Menten modeling (PopMM). This method can streamline an evaluation of the impact that complex genomic traits have on the metabolism of xenobiotics *in vitro,* particularly when tissue and other resources are limited. In addition to facilitating the accrual of data, it allows for rigorous testing of covariates as sources of kinetic parameter variability. In this *in silico* study we presented a practical application of this approach using previously published *in vitro*-derived CYP2D6 data, and explored the impact of sparse sampling and experimental error on known kinetic parameter estimates of CYP2D6 mediated formation of 4-hydroxy-atomoxetine in human liver microsomes.

Our analysis demonstrated that the incorporation of genotype as a categorical factor with covariate information in conjunction with a dynamic strategic sampling approach allows for accurate measurement of metabolic kinetic parameters using 50-70% less primary data than

conventional experimental designs. Our strategic sampling approach, using sparse data at the individual level, staggered in a manner to create rich data at the covariate level, provided accurate estimates of kinetic parameters associated with CYP2D6 genotype; with a geometric mean percent error of 2.61% (range 0%-9.33%) for V_{max} and 4.39% (range 0%-55.7%) for K_m across all experimental designs and conditions, when compared to the published estimates. The bootstrap analysis we conducted further supported the efficacy of the strategic sampling approach, with mean parameter values and their 95% confidence intervals falling within 80-125% of published estimates for most experimental designs and conditions (88% and 98% of conditions for K_m and V_{max}, respectively). Overall, our novel, strategically sampled NLME based approach provides a customizable framework that future researchers can use to efficiently assess non-linear relationships and associated covariates at an individual and population level. Additionally, to help facilitate future implementation of this approach, we have provided easily accessible open-source (license free) R code that can be adapted to perform alternative fit-for-purpose parameter estimation using the "nlme" R package[83].

In Chapter 3, we focused on the application of our PopMM approach developed in Chapter 2 to evaluate the relationship between *CYP4F2/CYP4F11* genotype and diplotype on hepatic mRNA abundance, protein abundance, and VK_1 ω-hydroxylation activity. Homozygosity for the *CYP4F2*3* allele was associated with a 57% decrease in V_{max} (73% decrease in protein abundance), a 103% increase in apparent $K_{m,}$ and a 65% decrease in hepatic VK_1 intrinsic clearance. Linkage disequilibrium between *CYP4F2*3* and other *CYP4F2* and *CYP4F11* allele variants yielded two diplotypes associated with comparably reduced metabolic activity (diplotype 17 / 20, and 20 / 20) relative to a reference diplotype (13 / 13) population. Overall, our results suggest that stratification by diplotype offers no additional advantage over single allele *CYP4F2*3*

genotyping for the identification of slow metabolizers of VK_1. However, in our analysis, the *CYP4F2*3* variant rarely exists as a sole polymorphism making it difficult to fully discern the magnitude of the **3* variant effect in the absence of background variability. Additionally, the data indicate a substantial contribution from CYP4F11 to VK_1 catabolism evidenced by the significant correlation between CYP4F11 protein abundance and activity after adjusting for variability in CYP4F2. This observation is especially apparent for *CYP4F2*3* carriers where CYP4F11 constituted ~40% of the combined CYP4F2 and CYP4F11 protein content. To our knowledge, this is the first study to systematically evaluate a *CYP4F* diplotype-phenotype relationships across mRNA abundance, protein abundance, and metabolic activity towards VK_1 in matched liver donor samples. Additionally, this study serves as the first *in vitro* pharmacogenomic application of the PopMM methodology, where *CYP4F* diplotype was used as a categorical covariate to evaluate the impact of background genetic variability on kinetic parameters (V_{max} and K_m).

Carriers of *CYP4F2*3* gene receiving coumarin-based anticoagulation therapy (i.e. warfarin) require, on average, higher doses in order to achieve a therapeutic INR[36,37], and it has been proposed that this is due to an increase in not only hepatic but also systemic VK exposure (i.e., status)[48]. As mentioned in Chapter 1, assessing whole body VK functional status as a single measure is challenging because of the unique tissue distribution properties of VK_1 and menaquinone. Plasma concentration alone fails to reflect hepatic and extrahepatic levels of VK under acute VK dosing conditions. However, a recently developed quantitative LC-MS/MS based assay for the analysis of VK metabolites in human urine has provided us with a minimally invasive, highly sensitive, and analyte-specific alternative for monitoring VK status in humans[55]. In Chapter 4, we assessed the modifying effect of the singular *CYP4F2*3* variant on the relationship between dietary VK exposure and short-term biomarkers of VK status in healthy human subjects using the

previously mentioned novel LC-MS/MS-based assay. To do this, we conducted a pilot 10-day dietary intervention study testing the effect that VK_1 supplementation (1 mg orally administered dose, n = 8 subjects) had on short-term biomarkers of VK status (plasma VK, and urinary K-acid catabolites). We then compared the impact that supplementation had on systemic VK exposure (plasma AUC, $C_{ss\ trough}$; urine K-acid excretion) in *CYP4F2*1/*1* and **3/*3* homozygotes (n = 4, respectively).

Overall, plasma VK_1 concentrations were elevated post-supplementation with a geometric mean baseline concentrations of 0.39 ng/mL 95% CI [0.13-1.17] vs 0.77 ng/mL 95% CI [0.38-1.53] post-supplementation. VK_1 exposure (trough $AUC_{Day\ 1\text{-}10}$) was 2.1-fold higher for the *CYP4F2 *3/*3* group compared to the **1/*1* group, with a 26% higher baseline (C_0) VK_1 concentration, and 37% higher baseline adjusted $C_{ss,\ trough}$ for *CYP4F2 *3/*3* homozygotes compared to **1/*1* homozygotes. Additionally, there was a statistically significant difference in steady state AUC (days 4-10) between genotype groups, where **3/*3* homozygotes had 2.7-fold higher AUC than **1/*1* homozygotes. This observation agrees with the *in vitro* predicted effect on AUC from the analysis conducted in Chapter 3, where a 65% decrease in CL_{int} was observed corresponding to a predicted 2.85-fold increase in AUC. Similar to VK_1 concentrations, there was an increase in total K-acid excretion amounts post supplementation, with geometric mean baseline concentrations of 7 ng/g creatinine 95% CI [4-16] vs 25 ng/mg creatinine 95% CI [9-74] post supplementation. Of the two K-acid catabolites, K-acid II was more sensitive to changes in dietary VK_1 intake, with an overall geometric mean fold change from baseline of 3.9-fold vs 2.7-fold for K-acid I. There were no statistically significant differences in urinary $AUC_{Day\ 1\text{-}10}$, $AUC_{Day\ 4\text{-}10}$ (steady state AUC), and C_0 for K-Acid I, II, and total K-acids (5C-aglycone plus 7C-aglycone metabolites) when stratified by *CYP4F2* genotype. However, at steady-state (days 4-10), creatinine

normalized total urine K-acid-to-plasma VK_1 AUC ratio was higher for **1/*1* homozygotes at 0.056 ± 0.04 vs 0.020 ± 0.01 for **3/*3* homozygotes, however the data did not support a statistically significant difference (p=0.38). Although our analysis was not adequately powered to detect statistically significant differences in other parameter estimates, the overall trends found in our data support differential effects of VK_1 supplementation on biomarkers of VK_1 status between *CYP4F2*1/*1* and **3/*3* homozygotes and warrants further investigation. In addition, although there are earlier published VK feeding studies evaluating VK-acid catabolites as candidate markers of VK dietary exposure and status[127,132], this study serves as the first LC-MS/MS based approach using K-acids as biomarkers for evaluating the modifying effect that the *CYP4F2*3* variant has on daily VK_1 supplementation response that mimics daily dietary VK exposure.

There are several future directions that can be taken to advance the research presented in this dissertation. First, the application of the NLME based approach used in our enzyme kinetics studies is not limited to metabolite formation but, in theory, can also be applied using a substrate depletion approach described by **Equation 7** and **Equation 8** in Chapter 2. This provides an opportunity to evaluate the impact of potential covariates on metabolic drug clearance, while limiting the need to characterize all relevant metabolites, which can be particularly useful in scenarios where no commercially available standards exist. Additionally, the NLME approach presented in Chapter 2 was conducted using maximum-likelihood and restricted maximum likelihood estimation routines[83], however, there are more powerful approaches (particularly when using sparse data[133]) such as first-order conditional estimation with interaction (FOCEI) and the newer stochastic approximation expectation maximization (SAEM) estimation algorithm; both of which can be implemented in commercial software such as NONMEM®, or in the case of FOCEI, Phoenix NLME®. An alternative estimation option is the use of Stan, a probabilistic programming

language that uses Markov chain Monte Carlo (MCMC) algorithms for Bayesian computational analysis[134]. Although less common, Bayesian inference offers a very straightforward means for fitting realistic models to complex data sets with measurement error, censored or missing observations, multilevel or serial correlation structures, and multiple endpoints[135]. Comparing these estimation routines side by side can provide further insight into the advantages and/or limitations of a strategic sampling approach. The application of the PopMM modeling to characterize the impact that genetic variability has on *in vitro* parameter estimates can be extended to models with greater complexity such as PopPK/PD models[77], and in particular PBPK models[136] where variability of *in vivo* processes can be modeled using estimates obtained from *in vitro* data. From a pharmacogenomic perspective this can be particularly useful because haplotype interactions captured *in vitro* may not be as straightforward as the sum of the individual SNV effects that comprise them; and the variability of its effect can be quite broad (as seen in our analysis).

Secondly, although measurement of urinary K-acids provide a highly sensitive alternative for assessing VK status (compared to plasma VK levels alone) their relationship with other functional indicators of VK status such, as PIVKA II, undercarboxylated osteocalcin (ucOC), or the less-commonly measured, urinary free Gla[137] and other VK dependent proteins (VKDPs) remains unclear. Vermeer et al. found in animal models that prothrombin synthesis reaches its maximal level at a relatively low dietary VK intake, however, urinary Gla-excretion was still substantially increased at higher VK intakes[138]; effectively showing a different VK utilization by liver and extrahepatic tissues. Moreover, this observation was found for both VK_1 and MK-4, but not for MK-8, which questions the bioavailability and distribution properties of higher menaquinones and their ability to support Gla formation in extrahepatic tissues. Further studies

evaluating the correlation between K-acids and functional indicators of VK status can help refine the utility of K-acid measures as a surrogate measure of clinically relevant, metabolic activity towards VK.

Finally, fully active prothrombin contains 10 γ-carboxy glutamic acid residues (Gla), and the absence of as little as two carboxylation modifications leads to a 50% reduction in activity, while the absence of three or more lead to the complete loss of activity[139,140]. Currently available antibody-based assays designed to detect the uncarboxylated form of VKDPs (descarboxy-Gla) such as PIVKA-II, and undercarboxylated osteocalcin (ucOC), do not inform on the position or number of des-carboxylation sites in a protein, and thus might not reflect early changes in coagulation status. Despite these shortcomings, in collaboration with our partners from Washington State University (Spokane, WA), we have recently published a novel LC-MS/MS based assay that provides a thorough, high-throughput, and sensitive means for quantifying des-carboxylated, fully-carboxylated, and total prothrombin in plasma[129]. The utility of this assay as a quantification tool for detecting *CYP4F2*3* genotype dependent changes in the degree of γ-carboxylation as a tissue-specific functional marker of VK status has yet to be explored; and provides an exciting area of research for future investigators.

In conclusion, while the protective role of VK towards the prevention of cardiovascular disease development and progression is still heavily debated[116], understanding genomic and environmental factors that contribute to variability in VK status can help us identify patients not only at risk for adverse drug reactions or therapeutic failure while on coumarin based therapies, but also other VK associated pathophysiology[3,141,142]. Overall, this dissertation provides a framework for investigating complex genomic traits *in-vitro* and helps clarify our understanding of the influence that variation in genes controlling VK metabolism has on VK status.

REFERENCES

1. Shearer MJ, Fu X, Booth SL. Vitamin K Nutrition , Metabolism , and Requirements : *Am Soc Nutr*. 2012;3(August):182-195. doi:10.3945/an.111.001800.182
2. Kyla Shea M, Booth SL. Concepts and controversies in evaluating vitamin K status in population-based studies. *Nutrients*. 2016;8(1):1-25. doi:10.3390/nu8010008
3. Kaneki M. Vitamin K2 as a protector of bone health and beyond. *Clin Calcium*. 2005;15(4):605-610.
4. I. K, P. G, A. M-Z, et al. Effect of Vitamin K2 on progression of atherosclerosis and vascular calcification in nondialyzed patients with chronic kidney disease stages 3-5. *Pol Arch Med Wewn*. 2015;125(9):631-640. http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L606377591%0Ahttp://rug.on.worldcat.org/atoztitles/link/?sid=EMBASE&issn=18979483&id=doi:&atitle=Effect+of+Vitamin+K2+on+progression+of+atherosclerosis+and+vascular+calcification+in+.
5. El Asmar MS, Naoum JJ, Arbid EJ. Vitamin K dependent proteins and the role of vitamin K2 in the modulation of vascular calcification: A review. *Oman Med J*. 2014;29(3):172-177. doi:10.5001/omj.2014.44
6. Oldenburg J, Marinova M, Müller-Reible C, Watzka M. The Vitamin K Cycle. *Vitam Horm*. 2008. doi:10.1016/S0083-6729(07)00003-9
7. Vermeer C. γ-Carboxyglutamate-containing proteins and the vitamin K-dependent carboxylase. *Biochem J*. 1990;266(3):625-636. doi:10.1042/bj2660625
8. Presnell SR, Stafford DW. The vitamin K-dependent carboxylase. *Thromb Haemost*. 2002;87(6):937-946. doi:10.1055/s-0037-1613115
9. Göbel K, Eichler S, Wiendl H, Chavakis T, Kleinschnitz C, Meuth SG. The coagulation factors fibrinogen, thrombin, and factor XII in inflammatory disorders-a systematic review. *Front Immunol*. 2018;9(JUL). doi:10.3389/fimmu.2018.01731
10. Caspers M, Pavlova A, Driesen J, et al. Deficiencies of antithrombin, protein C and protein S - practical experience in genetic analysis of a large patient cohort. *Thromb Haemost*. 2012;108(2):247-257. doi:10.1160/TH11-12-0875

11. Presnell SR, Stafford DW. The Vitamin K-dependent Carboxylase *. 2002:937-946.
12. Winter WE, Greene DN, Beal SG, et al. Clotting factors: Clinical biochemistry and their roles as plasma enzymes. *Adv Clin Chem*. 2020;94:31-84. doi:10.1016/BS.ACC.2019.07.008
13. Shiraki M, Tsugawa N, Okano T. Recent advances in vitamin K-dependent Gla-containing proteins and vitamin K nutrition. *Osteoporos Sarcopenia*. 2015;1(1):22-38. doi:10.1016/j.afos.2015.07.009
14. Schurgers LJ, Vermeer C. Differential lipoprotein transport pathways of K-vitamins in healthy subjects. *Biochim Biophys Acta - Gen Subj*. 2002;1570(1):27-32. doi:10.1016/S0304-4165(02)00147-2
15. Okano T, Shimomura Y, Yamane M, et al. Conversion of phylloquinone (vitamin K1) into menaquinone-4 (vitamin K2) in mice: Two possible routes for menaquinone-4 accumulation in cerebra of mice. *J Biol Chem*. 2008;283(17):11270-11279. doi:10.1074/jbc.M702971200
16. Theuwissen E, Smit E, Vermeer C. The Role of Vitamin K in Soft-Tissue Calcification. *Adv Nutr*. 2012;3(2):166-173. doi:10.3945/an.111.001628
17. O'Reilly RA. Studies on the optical enantiomorphs of warfarin in man. *Clin Pharmacol Ther*. 1974;16(2):348-354. doi:10.1002/cpt1974162348
18. Choonara IA, Haynes BP, Cholerton S, Breckenridge AM, Park BK. Enantiomers of warfarin and vitamin K1 metabolism. *Br J Clin Pharmacol*. 1986;22(6):729-732. doi:10.1111/j.1365-2125.1986.tb02966.x
19. Danziger J. Vitamin K-dependent proteins, warfarin, and vascular calcification. *Clin J Am Soc Nephrol*. 2008;3(5):1504-1510. doi:10.2215/CJN.00770208
20. Verschuren JJW, Trompet S, Wessels JAM, Guchelaar H, Maat MPM De, Simoons ML. Controversies in cardiovascular medicine A systematic review on pharmacogenetics in cardiovascular disease : is it ready for clinical application ? 2011;(July). doi:10.1093/eurheartj/ehr239
21. Pirmohamed M, Burnside G, Eriksson N, et al. A Randomized Trial of Genotype-Guided Dosing of Warfarin. *N Engl J Med*. 2013;369(24):2294-2303. doi:10.1056/nejmoa1311386
22. Agrawal S, Heiss MS, Fenter RB, et al. Impact of CYP2C9-Interacting Drugs on Warfarin

Pharmacogenomics. *Clin Transl Sci*. 2020;13(5):941-949. doi:10.1111/cts.12781

23. Choi KH, Kim AJ, Son IJ, et al. Risk Factors of Drug Interaction between Warfarin and Nonsteroidal Anti-Inflammatory Drugs in Practical Setting. *J Korean Med Sci*. 2010;25(3):337. doi:10.3346/JKMS.2010.25.3.337
24. Aithal GP, Day CP, Kesteven PJL, Daly AK. Association of polymorphisms in the cytochrome P450 CYP2C9 with warfarin dose requirement and risk of bleeding complications. *Lancet*. 1999;353(9154):717-719. doi:10.1016/S0140-6736(98)04474-2
25. Yuan H-Y, Chen J-J, Lee MTM, et al. A novel functional VKORC1 promoter polymorphism is associated with inter-individual and inter-ethnic differences in warfarin sensitivity. doi:10.1093/hmg/ddi180
26. Edson KZ, Rettie AE. CYP4 enzymes as potential drug targets: focus on enzyme multiplicity, inducers and inhibitors, and therapeutic modulation of 20-hydroxyeicosatetraenoic acid (20-HETE) synthase and fatty acid ω-hydroxylase activities. *Curr Top Med Chem*. 2013;13(12):1429-1440. doi:10.2174/15680266113139990110
27. Bylund J, Bylund M, Oliw EH. cDna cloning and expression of CYP4F12, a novel human cytochrome P450. *Biochem Biophys Res Commun*. 2001;280(3):892-897. doi:10.1006/bbrc.2000.4191
28. Powell PK, Wolf I, Jin R, Lasker JM. Metabolism of arachidonic acid to 20-hydroxy-5,8,11, 14-eicosatetraenoic acid by P450 enzymes in human liver: involvement of CYP4F2 and CYP4A11. *J Pharmacol Exp Ther*. 1998;285(3):1327-1336.
29. Cui X, Nelson DR, Strobel HW. A novel human cytochrome P450 4F isoform (CYP4F11): cDNA cloning, expression, and genomic structural characterization. *Genomics*. 2000;68(2):161-166. doi:10.1006/geno.2000.6276
30. McDonald MG, Rieder MJ, Nakano M, Hsia CK, Rettie AE. CYP4F2 Is a Vitamin K1 Oxidase: An Explanation for Altered Warfarin Dose in Carriers of the V433M Variant. *Mol Pharmacol*. 2009;75(6):1337-1346. doi:10.1124/mol.109.054833
31. Ni K Di, Liu JY. The Functions of Cytochrome P450 ω-hydroxylases and the Associated Eicosanoids in Inflammation-Related Diseases. *Front Pharmacol*. 2021;12(September):1-14. doi:10.3389/fphar.2021.716801
32. Edson KZ, Prasad B, Unadkat JD, et al. Cytochrome P450-dependent catabolism of vitamin K: ω-hydroxylation catalyzed by human CYP4F2 and CYP4F11. *Biochemistry*.

2013;52(46):8276-8285. doi:10.1021/bi401208m

33. Eberlé D, Hegarty B, Bossard P, Ferré P, Foufelle F. SREBP transcription factors: Master regulators of lipid homeostasis. *Biochimie*. 2004;86(11):839-848. doi:10.1016/j.biochi.2004.09.018
34. Hsu MH, Savas Ü, Griffin KJ, Johnson EF. Regulation of human cytochrome P450 4F2 expression by sterol regulatory element-binding protein and lovastatin. *J Biol Chem*. 2007;282(8):5225-5236. doi:10.1074/jbc.M608176200
35. Zhang T, Zhan Z, Chen Y, et al. Regulation of cytochrome P450 4F11 expression by liver X receptor alpha. *Int Immunopharmacol*. 2021;90(June 2020). doi:10.1016/j.intimp.2020.107240
36. Caldwell MD, Awad T, Johnson JA, et al. CYK4F-2 genetic variant alters required warfarin dose. *Blood*. 2008;111(8):4106-4112. doi:10.1182/blood-2007-11-122010
37. Zhang JE, Klein K, Jorgensen AL, et al. Effect of genetic variability in the CYP4F2, CYP4F11, and CYP4F12 genes on liver mrna levels and warfarin response. *Front Pharmacol*. 2017;8(MAY). doi:10.3389/fphar.2017.00323
38. Dituri F, Buonocore G, Pietravalle A, et al. PIVKA-II plasma levels as markers of subclinical vitamin K deficiency in term infants. *J Matern Neonatal Med*. 2012;25(9):1660-1663. doi:10.3109/14767058.2012.657273
39. Meguro T, Yamada K. A simple and rapid test for Pivka-II in plasma. *Thromb Res*. 1982;25(1):109-114. doi:https://doi.org/10.1016/0049-3848(82)90219-5
40. Bakhshi S, Deorari AK, Roy S, Paul VK, Singh M. Prevention of subclinical vitamin K deficiency based on PIVKA-II levels: Oral versus intramuscular route. *Indian Pediatr*. 1996;33(12):1040-1043.
41. Lee W, Chung HJ, Kim S, et al. PIVKA-II is a candidate marker for monitoring the effects of the oral anticoagulant warfarin. *Clin Biochem*. 2010;43(13-14):1177-1179. doi:10.1016/J.CLINBIOCHEM.2010.06.022
42. Roth JA, Boudreau D, Fujii MM, et al. Genetic risk factors for major bleeding in patients treated with warfarin in a community setting. *Clin Pharmacol Ther*. 2014;95(6):636-643. doi:10.1038/clpt.2014.26
43. Hart R, Veenstra DL, Boudreau DM, Roth JA. Impact of Body Mass Index and Genetics on Warfarin Major Bleeding Outcomes in a Community Setting. *Am J Med*. 2017.

doi:10.1016/j.amjmed.2016.08.017

44. Booth SL. Dietary vitamin K guidance: An effective strategy for stable control of oral anticoagulation. *Nutr Rev*. 2010;68(3):178-181. doi:10.1111/j.1753-4887.2010.00274.x
45. Truong JT, Fu X, Saltzman E, et al. Age Group and Sex Do Not Influence Responses of Vitamin K Biomarkers to Changes in Dietary Vitamin K. *J Nutr*. 2012;142(5):936-941. doi:10.3945/jn.111.154807
46. Henderson LM, Claw KG, Woodahl EL, et al. P450 pharmacogenetics in indigenous North American populations. *J Pers Med*. 2018;8(1). doi:10.3390/jpm8010009
47. Au NT, Ryman T, Rettie AE, 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*. 2018;62(3):1-9. doi:10.1002/mnfr.201700746
48. Park J, Kim K, Park J. Effects of Ketoconazole, a CYP4F2 Inhibitor, and CYP4F2*3 Genetic Polymorphism on Pharmacokinetics of Vitamin K 1 . *J Clin Pharmacol*. 2019. doi:10.1002/jcph.1444
49. Schurgers LJ, Vermeer C. Determination of phylloquinone and menaquinones in food. Effect of food matrix on circulating vitamin K concentrations. *Haemostasis*. 2000;30(6):298-307. doi:10.1159/000054147
50. Goncalves A, Margier M, Roi S, et al. Intestinal scavenger receptors are involved in vitamin K1 absorption. *J Biol Chem*. 2014;289(44):30743-30752. doi:10.1074/jbc.M114.587659
51. Tso P, Balint JA. Formation and transport of chylomicrons by enterocytes to the lymphatics. *Am J Physiol*. 1986;250(6 Pt 1):G715-26. doi:10.1152/ajpgi.1986.250.6.G715
52. Havel RJ. Receptor and non-receptor mediated uptake of chylomicron remnants by the liver. *Atherosclerosis*. 1998;141(SUPPL. 1):S1-S7. doi:10.1016/s0021-9150(98)00211-1
53. Thijssen HHW, Drittij-Reijnders- MJ. Vitamin K status in human tissues: tissue-specific accumulation of phylloquinone and menaquinone-4. *Br J Nutr*. 1996;75(1):121-127. doi:10.1079/bjn19960115
54. Teitelbaum AM, Scian M, Nelson WL, Rettie AE. Efficient Syntheses of Vitamin K Chain-Shortened Acid Metabolites. doi:10.1055/s-0034-1380002
55. McDonald MG, Yeung CK, Teitelbaum AM, et al. A new LC-MS assay for the quantitative analysis of vitamin K metabolites in human urine. *J Lipid Res*.

2019;60(4):892-899. doi:10.1194/jlr.d087916

56. Yamanashi Y, Takada T, Kurauchi R, Tanaka Y, Komine T, Suzuki H. Transporters for the Intestinal Absorption of Cholesterol, Vitamin E, and Vitamin K. *J Atheroscler Thromb*. 2017;24(4):347-359. doi:10.5551/jat.rv16007
57. Takada T, Yamanashi Y, Konishi K, et al. NPC1L1 is a key regulator of intestinal vitamin K absorption and a modulator of warfarin therapy. *Sci Transl Med*. 2015;7(275). doi:10.1126/scitranslmed.3010329
58. Turshudzhyan A. Anticoagulation Options for Coronavirus Disease 2019 (COVID-19)-Induced Coagulopathy. *Cureus*. 2020;2019(5). doi:10.7759/cureus.8150
59. Booth SL, Golly I, Sacheck JM, et al. Effect of vitamin E supplementation on vitamin K status in adults with normal coagulation status. *Am J Clin Nutr*. 2004;80(1):143-148. doi:10.1093/ajcn/80.1.143
60. Jin S, Li Z, Yang Q, et al. Simultaneous Characterization and Determination of Warfarin and Its Hydroxylation Metabolites in Rat Plasma by Chiral Liquid Chromatography-Tandem Mass Spectrometry. *Pharmaceutics*. 2022;14(6). doi:10.3390/pharmaceutics14061141
61. Shearer MJ, Newman P. Metabolism and cell biology of vitamin K. *Thromb Haemost*. 2008;100(10):530-547. doi:10.1160/th08-03-0147
62. Hrycay EG, Bandiera SM. Expression, function and regulation of mouse cytochrome P450 enzymes: comparison with human P450 enzymes. *Curr Drug Metab*. 2009;10(10):1151-1183. doi:10.2174/138920009790820138
63. Woodcock J, Lesko LJ. Pharmacogenetics-Tailoring Treatment for the Outliers. *n engl j med*. 2009;360. http://www.fda.gov/cder/drug/. Accessed December 2, 2022.
64. Huang SM, Chen L, Giacomini KM. Pharmacogenomic mechanisms of drug toxicity. *Atkinson's Princ Clin Pharmacol*. January 2022:303-322. doi:10.1016/B978-0-12-819869-8.00002-1
65. Peck RW. Precision Medicine Is Not Just Genomics: The Right Dose for Every Patient. *https://doi.org/101146/annurev-pharmtox-010617-052446*. 2018;58:105-122. doi:10.1146/ANNUREV-PHARMTOX-010617-052446
66. Reynolds KS. Achieving the Promise of Personalized Medicine. *Clin Pharmacol Ther*. 2012;92(4):401-405. doi:10.1038/CLPT.2012.147

67. Schwarz UI, Gulilat M, Kim RB. The Role of Next-Generation Sequencing in Pharmacogenetics and Pharmacogenomics. 2019. doi:10.1101/cshperspect.a033027
68. Parkinson A, Mudra DR, Johnson C, Dwyer A, Carroll KM. The effects of gender, age, ethnicity, and liver cirrhosis on cytochrome P450 enzyme activity in human liver microsomes and inducibility in cultured human hepatocytes. *Toxicol Appl Pharmacol*. 2004;199(3):193-209. doi:10.1016/j.taap.2004.01.010
69. Obach RS, Baxter JG, Liston TE, et al. The prediction of human pharmacokinetic parameters from preclinical and in vitro metabolism data. *J Pharmacol Exp Ther*. 1997;283(1):46-58.
70. Nagar S, Argikar U, Tweedie D. *Enzyme Kinetics in Drug Metabolism*. Vol 2342. 2nd ed. (Nagar S, Argikar UA, Tweedie D, eds.). New York, NY: Springer US; 2021. doi:10.1007/978-1-0716-1554-6
71. Mohatt G V., Plaetke R, Klejka J, et al. The Center for Alaska Native Health Research Study: a community-based participatory research study of obesity and chronic disease-related protective and risk factors. *Int J Circumpolar Health*. 2007;66(1):8-18. doi:10.3402/ijch.v66i1.18219
72. LaPierre N, Taraszka K, Huang H, He R, Hormozdiari F, Eskin E. Identifying causal variants by fine mapping across multiple studies. Zeggini E, ed. *PLOS Genet*. 2021;17(9):e1009733. doi:10.1371/journal.pgen.1009733
73. Shirasaka Y, Chaudhry AS, McDonald M, et al. Interindividual variability of CYP2C19-catalyzed drug metabolism due to differences in gene diplotypes and cytochrome P450 oxidoreductase content. *Pharmacogenomics J*. 2016;16(4):375-387. doi:10.1038/tpj.2015.58
74. Flora DR, Tracy TS. Development of an In Vitro System with Human Liver Microsomes for Phenotyping of CYP2C9 Genetic Polymorphisms with a Mechanism-Based Inactivator. 2012. doi:10.1124/dmd.111.043372
75. Dalton R, Lee S been, Claw KG, et al. Interrogation of CYP2D6 Structural Variant Alleles Improves the Correlation Between CYP2D6 Genotype and CYP2D6-Mediated Metabolic Activity. *Clin Transl Sci*. 2020;13(1):147-156. doi:10.1111/cts.12695
76. Sheiner LB, Beal SL. Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis-menten model: Routine clinical pharmacokinetic data. *J*

Pharmacokinet Biopharm 1980 86. 1980;8(6):553-571. doi:10.1007/BF01060053

77. Sheiner LB, Rosenberg B, Marathe V V. Estimation of population characteristics of pharmacokinetic parameters from routine clinical data. *J Pharmacokinet Biopharm 1977 55*. 2015;5(5):445-479. doi:10.1007/BF01061728
78. Bonate PL. Recommended Reading in Population Pharmacokinetic Pharmacodynamics. 2005. http://www.aapsj.org. Accessed December 2, 2022.
79. Bhavatharini PA, Deepalakshmi M, Arun KP. Pharmacometrics: The science applied from bench to bedside ARTICLE INFO. *J Appl Pharm Sci*. 2022;12(01):55-064. doi:10.7324/JAPS.2021.120104
80. Archontoulis S V., Miguez FE. Nonlinear regression models and applications in agricultural research. *Agron J*. 2015;107(2):786-798. doi:10.2134/agronj2012.0506
81. Oddi FJ, Miguez FE, Ghermandi L, Bianchi LO, Garibaldi LA. A nonlinear mixed-effects modeling approach for ecological data: Using temporal dynamics of vegetation moisture as an example. *Ecol Evol*. 2019;9(18):10225-10240. doi:10.1002/ece3.5543
82. Dinh JC, Pearce RE, Haandel L Van, Gaedigk A, Leeder JS. Special Section on Pediatric Drug Disposition and Pharmacokinetics Characterization of Atomoxetine Biotransformation and Implications for Development of PBPK Models for Dose Individualization in Children s. 2016;(July):1070-1079.
83. Pinheiro JC, Bates DM. Mixed-Effects Models in S and S-PLUS. *Mix Model S S-PLUS*. 2000. doi:10.1007/B98882/COVER
84. Belle DJ, Ring BJ, Allerheiligen SRB, et al. A population approach to enzyme characterization and identification: Application to phenacetin O-deethylation. *Pharm Res*. 2000;17(12):1531-1536. doi:10.1023/A:1007665310830
85. Bogacka B, Latif MAHM, Gilmour SG, et al. Optimum Designs for Non-Linear Mixed Effects Models in the Presence of Covariates. 2017;(September):927-937. doi:10.1111/biom.12660
86. Yu G, Li GF, Markowitz JS. Atomoxetine: A Review of Its Pharmacokinetics and Pharmacogenomics Relative to Drug Disposition. *J Child Adolesc Psychopharmacol*. 2016;26(4):314-326. doi:10.1089/cap.2015.0137
87. Brown JT, Abdel-Rahman SM, van Haandel L, Gaedigk A, Lin YS, Leeder JS. Single dose, CYP2D6 genotype-stratified pharmacokinetic study of atomoxetine in children with

ADHD. *Clin Pharmacol Ther*. 2016;99(6):642-650. doi:10.1002/cpt.319

88. Dean L. Atomoxetine Therapy and CYP2D6 Genotype. In: Pratt VM, Scott SA, Pirmohamed M, Esquivel B, Kattman BL, Malheiro AJ, eds. Bethesda (MD); 2012.
89. Nath A, Atkins WM. A theoretical validation of the substrate depletion approach to determining kinetic parameters. *Drug Metab Dispos*. 2006;34(9):1433-1435. doi:10.1124/dmd.106.010777
90. Fedorov V V., Leonov SL. Optimal design for nonlinear response models. *Optim Des Nonlinear Response Model*. January 2013:1-369. doi:10.1201/b15054
91. Piehowski PD, Petyuk VA, Orton DJ, et al. Sources of technical variability in quantitative LC-MS proteomics: Human brain tissue sample analysis. *J Proteome Res*. 2013;12(5):2128-2137. doi:10.1021/pr301146m
92. Favé MJ, Lamaze FC, Soave D, et al. Gene-by-environment interactions in urban populations modulate risk phenotypes. *Nat Commun*. 2018;9(1). doi:10.1038/s41467-018-03202-2
93. Slatkin M. Linkage disequilibrium - Understanding the evolutionary past and mapping the medical future. *Nat Rev Genet*. 2008;9(6):477-485. doi:10.1038/nrg2361
94. Goddard ME, Kemper KE, MacLeod IM, Chamberlain AJ, Hayes BJ. Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture. *Proc R Soc B Biol Sci*. 2016;283(1835). doi:10.1098/RSPB.2016.0569
95. Jarrar YB, Lee SJ. Molecular functionality of cytochrome P450 4 (CYP4) genetic polymorphisms and their clinical implications. *Int J Mol Sci*. 2019;20(17). doi:10.3390/ijms20174274
96. Edson KZ, Rettie AE. CYP4 Enzymes as potential drug targets: focus on enzyme multiplicity, inducers and inhibitors, and therapeutic modulation of 20-hydroxyeicosatetraenoic acid (20-HETE) synthase and fatty acid ω-hydroxylase activities.
97. Sasaki T, Horikawa M, Orikasa K, et al. Possible Relationship Between the Risk of Japanese Bladder Cancer Cases and the CYP4B1 Genotype. *Jpn J Clin Oncol*. 2008;38(9):634-640. doi:10.1093/JJCO/HYN081
98. Sirotina S, Ponomarenko I, Kharchenko A, et al. A Novel Polymorphism in the Promoter of the CYP4A11 Gene Is Associated with Susceptibility to Coronary Artery Disease. *Dis*

Markers. 2018;2018. doi:10.1155/2018/5812802

99. Ananthakrishnan AN, Khalili H, Song M, et al. Genetic Polymorphisms in Fatty Acid Metabolism Modify the Association Between Dietary n3:n6 Intake and Risk of Ulcerative Colitis: A Prospective Cohort Study. 2017. doi:10.1097/MIB.0000000000001236
100. Yin J, Liu H, Liu Z, et al. Pathway-analysis of published genome-wide association studies of lung cancer: A potential role for the CYP4F3 locus. 2017;17:18. doi:10.1002/mc.22622
101. Austin H, De Staercke C, Lally C, Bezemer ID, Rosendaal FR, Hooper WC. New gene variants associated with venous thrombosis: A replication study in White and Black Americans. *J Thromb Haemost*. 2011;9(3):489-495. doi:10.1111/J.1538-7836.2011.04185.X
102. Fiatal S, Pikó P, Kósa Z, Sándor J, Ádány R. Genetic profiling revealed an increased risk of venous thrombosis in the Hungarian Roma population. *Thromb Res*. 2019;179:37-44. doi:10.1016/J.THROMRES.2019.04.031
103. Glurich I, Chyou P-H, Engel JM, Cross DS, Onitilo AA. Tamoxifen-Induced Venothromboembolic Events: Exploring Validation of Putative Genetic Association. *Clin Med Res*. 2013;11(1):16-25. doi:10.3121/CMR.2012.1101
104. Stec DE, Roman RJ, Flasch A, Rieder MJ. Functional polymorphism in human CYP4F2 decreases 20-HETE production. 2021;(3):74-81. doi:10.1152/physiolgenomics.00003.2007.
105. Tanner JA, Prasad B, Claw KG, et al. Predictors of variation in CYP2A6 mRNA, protein, and enzyme activity in a human liver bank: Influence of genetic and nongenetic factors. *J Pharmacol Exp Ther*. 2017;360(1):129-139. doi:10.1124/jpet.116.237594
106. Stephens M, Smith NJ, Donnelly P. A New Statistical Method for Haplotype Reconstruction from Population Data. *Am J Hum Genet*. 2001;68:978-989.
107. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinforma Appl NOTE*. 2005;21(2):263-265. doi:10.1093/bioinformatics/bth457
108. Smith PK, Krohn RI, Hermanson GT, et al. Measurement of protein using bicinchoninic acid. *Anal Biochem*. 1985;150(1):76-85. doi:10.1016/0003-2697(85)90442-7
109. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B*. 1995;57(1):289-300.

http://www.jstor.org/stable/2346101.

110. Cui JY, Renaud HJ, Klaassen CD. Ontogeny of novel cytochrome P450 gene isoforms during postnatal liver maturation in mice. *Drug Metab Dispos*. 2012;40(6):1226-1237. doi:10.1124/dmd.111.042697
111. Edson KZ, Prasad B, Unadkat JD, et al. Cytochrome P450-Dependent Catabolism of Vitamin K: ω - Hydroxylation Catalyzed by Human CYP4F2 and CYP4F11. 2013:2-11.
112. Kikuta Y, Kusunose E, Kusunose M. Prostaglandin and leukotriene ω-hydroxylases. *Prostaglandins Other Lipid Mediat*. 2002;68-69:345-362. doi:10.1016/S0090-6980(02)00039-4
113. Virolainen SJ, Kottyan LC. Gene – environment interactions and their impact on human health. 2023;(September 2022). doi:10.1038/s41435-022-00192-6
114. Hunter DJ. Gene-environment interactions in human diseases. *Nat Rev Genet*. 2005;6(4):287-298. doi:10.1038/nrg1578
115. Li RC, Finkelman BS, Chen J, et al. Dietary vitamin k intake and anticoagulation control during the initiation phase of warfarin therapy: A prospective cohort study. *Thromb Haemost*. 2013;110(1):195-196. doi:10.1160/TH13-02-0111
116. Shea MK, Berkner KL, Ferland G, Fu X, Holden RM, Booth SL. Perspective : Evidence before Enthusiasm — A Critical Review of the Potential Cardiovascular Benefits of Vitamin K. *Adv Nutr*. 2021;2(4):1-15. doi:10.1093/advances/nmab004
117. Booth S ~L., Suttie J ~W. Dietary vitamin K1 and stability of prothrombin times during warfarin therapy. *JAMA*. 1998;280(15):1317-1319.
118. Au NT, Reyes M, Boyer BB, et al. Dietary and genetic influences on hemostasis in a Yup'ik Alaska Native population. *PLoS One*. 2017;12(4):1-16. doi:10.1371/journal.pone.0173616
119. Kim HJ, Shin J, Kang Y, Kim D, Park JJ, Kim HJ. Effect of different cooking method on vitamin E and K content and true retention of legumes and vegetables commonly consumed in Korea. *Food Sci Biotechnol*. 2022;32(5):647-658. doi:10.1007/S10068-022-01206-9/METRICS
120. Booth SL, Lichtenstein AH, O'Brien-Morse M, et al. Effects of a hydrogenated form of vitamin K on bone formation and resorption. *Am J Clin Nutr*. 2001;74(6):783-790. doi:10.1093/ajcn/74.6.783

121. Booth SL, McKeown NM, Lichtenstein AH, et al. A hydrogenated form of vitamin K: Its relative bioavailability and presence in the food supply. *J Food Compos Anal*. 2000;13(4):311-317. doi:10.1006/jfca.1999.0863
122. Fujii S, Shimizu A, Takeda N, et al. Systematic synthesis and anti-inflammatory activity of ω-carboxylated menaquinone derivatives—Investigations on identified and putative vitamin K2 metabolites. *Bioorg Med Chem*. 2015;23(10):2344-2352. doi:10.1016/J.BMC.2015.03.070
123. Denny WS, Duvvuri S, Buckeridge C. Simple, Automatic Noncompartmental Analysis: The PKNCA R Package. *J Pharmacokinet Pharmacodyn*. 2015;42(1):11-107.
124. Schurgers LJ, Teunissen KJF, Hamulya K, Knapen MHJ, Vik H, Vermeer C. Vitamin K – containing dietary supplements : comparison of synthetic vitamin K 1 and natto-derived menaquinone-7. 2018;109(8):3279-3284. doi:10.1182/blood-2006-08-040709.The
125. Thijssen HHW, Drittij-Reijnders MJ. Vitamin K distribution in rat tissues: dietary phylloquinone is a source of tissue menaquinone-4. *Br J Nutr*. 1994;72(3):415-425. doi:10.1079/bjn19940043
126. Ellis JL, Fu X, Karl JP, et al. Multiple Dietary Vitamin K Forms Are Converted to Tissue Menaquinone-4 in Mice. *J Nutr*. 2021;4. doi:10.1093/jn/nxab332
127. Harrington DJ, Booth SL, Card DJ, Shearer MJ. Excretion of the Urinary 5C- and 7C-Aglycone Metabolites of Vitamin K by Young Adults Responds to Changes in Dietary Phylloquinone and Dihydrophylloquinone Intakes. *J Nutr*. 2007;137(7):1763-1768. doi:10.1093/jn/137.7.1763
128. Weinshilboum RM, Wang L. Pharmacogenetics and pharmacogenomics: Development, science, and translation. *Annu Rev Genomics Hum Genet*. 2006;7:223-245. doi:10.1146/annurev.genom.6.080604.162315
129. Basit A, Prasad B, Estergreen J, et al. A novel LC-MS/MS assay for quantification of des-carboxy prothrombin and characterization of warfarin induced changes. *Clin Transl Sci*. 2020.
130. Danese E, Montagnana M, Johnson JA, et al. Impact of the CYP4F2 p.V433M polymorphism on coumarin dose requirement: Systematic review and meta-analysis. *Clin Pharmacol Ther*. 2012;92(6):746-756. doi:10.1038/clpt.2012.184
131. Eunice J, Jorgensen AL, Alfirevic A, et al. Effects of CYP4F2 genetic polymorphisms and

haplotypes on clinical outcomes in patients initiated on warfarin therapy. :781-789. doi:10.1097/FPC.0b013e3283311347

132. Harrington DJ, Savidge GF, Shearer MJ, Soper R, Edwards C, Hodges SJ. Determination of the urinary aglycone metabolites of vitamin K by HPLC with redox-mode electrochemical detection. *J Lipid Res*. 2005;46(5):1053-1060. doi:10.1194/jlr.D400033-JLR200
133. Sukarnjanaset W, Wattanavijitkul T, Jarurattanasirikul S. Evaluation of FOCEI and SAEM Estimation Methods in Population Pharmacokinetic Analysis Using NONMEM ® Across Rich, Medium, and Sparse Sampling Data. *Eur J Drug Metab Pharmacokinet*. 2018;43(6):729-736. doi:10.1007/S13318-018-0484-8/METRICS
134. Carpenter B, Lee D, Brubaker MA, et al. Journal of Statistical Software Stan: A Probabilistic Programming Language. http://www.jstatsoft.org/. Accessed May 21, 2023.
135. Dunson DB. Commentary: Practical Advantages of Bayesian Analysis of Epidemiologic Data. *Am J Epidemiol*. 2001;153(12):1222-1226. doi:10.1093/AJE/153.12.1222
136. Jones HM, Rowland-Yeo K. Basic concepts in physiologically based pharmacokinetic modeling in drug discovery and development. *CPT Pharmacometrics Syst Pharmacol*. 2013;2(8):1-12. doi:10.1038/psp.2013.41
137. Ferland G, Sadowski JA, O'Brien ME. Dietary induced subclinical vitamin K deficiency in normal human subjects. *J Clin Invest*. 1993;91(4):1761-1768. doi:10.1172/JCI116386
138. Craciun AM, Groenen-Van Dooren MMCL, Vermeer C. Nutritional vitamin K-intake and urinary γ-carboxyglutamate excretion in the rat. *Biochim Biophys Acta - Gen Subj*. 1997;1334(1):44-50. doi:10.1016/S0304-4165(96)00073-6
139. Ratcliffe J V., Furie B, Furie BC. The importance of specific γ-carboxyglutamic acid residues in prothrombin. Evaluation by site-specific mutagenesis. *J Biol Chem*. 1993;268(32):24339-24345. doi:10.1016/s0021-9258(20)80531-6
140. Malhotra OP, Nesheim ME, Mann KG. The kinetics of activation of normal and γ-carboxyglutamic acid-deficient prothrombins. *J Biol Chem*. 1985;260(1):279-287. doi:10.1016/s0021-9258(18)89728-9
141. Mishima E, Ito J, Wu Z, et al. A non-canonical vitamin K cycle is a potent ferroptosis suppressor. *Nature*. 2022;608(October 2021). doi:10.1038/s41586-022-05022-3
142. Mizobuchi M, Towler D, Slatopolsky E. Vascular Calcification: The Killer of Patients

with Chronic Kidney Disease. *J Am Soc Nephrol*. 2009;20(7):1453-1464. doi:10.1681/asn.2008070692

VITA

Ayoade Nathaniel Alade was born and raised in Southern California. He received a Bachelor's and Master of Science in Biochemistry from California State University, Long Beach in 2014 and 2017, respectively.