Published in final edited form as:

Clin Pharmacol Ther. 2011 March ; 89(3): 408–415. doi:10.1038/clpt.2010.322.

The Missing Association: Sequencing-Based Discovery of Novel SNPs in *VKORC1* and *CYP2C9* That Affect Warfarin Dose in African Americans

MA Perera¹, E Gamazon¹, LH Cavallari², SR Patel², S Poindexter¹, RA Kittles¹, D Nicolae^{1,3}, and NJ Cox¹

¹Department of Medicine, Section of Genetic Medicine, University of Chicago, Chicago, Illinois, USA

²Department of Pharmacy Practice, University of Illinois at Chicago, Chicago, Illinois, USA

³Department of Statistics, University of Chicago, Chicago, Illinois, USA

Abstract

It is well recognized that the genetic variants *VKORC1*-1639, *CYP2C9**2, and *CYP2C9**3 contribute to warfarin dose response. This has led to warfarin dosing algorithms that include these polymorphisms and explain between 47% and 56% of variability in dose in Caucasians. However, these polymorphisms explain significantly less of the variance in dose among African Americans. In order to identify novel variations that affect warfarin dose in African Americans, we used a targeted resequencing strategy that examined evolutionarily conserved sequences and regions of putative transcriptional binding. Through ethnicity-specific warfarin dose model building in 330 African Americans, we identified two novel genetic associations with higher warfarin dose, namely, *VKORC1*-8191 (rs61162043, $P = 0.0041$) and 18786 in *CYP2C9* (rs7089580, $P = 0.035$). These novel finds are independent of the previous associations with these genes. Our regression model, encompassing both genetic and clinical variables, explained 40% of the variability in warfarin dose in African-American subjects, significantly more than any model thus far.

Warfarin, which is used to prevent thrombotic disorders,¹ is available as a racemic mixture, with the majority of the pharmacologic effect attributed to the S-isomer.^{2,3} Given the role of cytochrome P-450 2C9 (*CYP2C9*) in the hydroxylation of (S)-warfarin,^{4,5} the function of this enzyme is crucial to the therapeutic response and adverse-effect profile of this drug. Warfarin produces its therapeutic effect through the inhibition of vitamin K epoxide reductase (encoded by the gene *VKORC1*), which results in decreased production of the reduced form of vitamin K, an essential cofactor in vitamin K-dependent clotting factor activation.

The variations in warfarin dose have been attributed to several factors such as age,^{6,7} diet,⁸⁻¹⁰ and concomitant medications^{11,12} as well as to genetic factors.¹³ Single-nucleotide polymorphisms (SNPs) in both *CYP2C9* and *VKORC1* have reproducibly been found to affect warfarin dose, leading to the development of pharmacogenetically guided warfarin dosing algorithms.^{14,15} The warfarin drug label was recently revised to reflect these

© 2011 American Society for Clinical Pharmacology and Therapeutics

Correspondence: MA Perera (mperera@bsd.uchicago.edu).

CONFLICT OF INTEREST

The authors declared no conflict of interest.

SUPPLEMENTARY MATERIAL is linked to the online version of the paper at <http://www.nature.com/cpt>

pharmacogenetic findings. Functional SNPs in *CYP2C9* and haplogroup tagging SNPs in *VKORC1* account for >30% of the variation in warfarin dose in Caucasians, but for only ~10% of the variation in African Americans. With the addition of clinical and demographic variables, ~47–56% of the variation in dose in Caucasians can be accounted for, as compared with only 25% of the variation in African Americans.¹⁵ Studies have shown population-specific differences in warfarin requirements, with patients of African descent requiring higher mean warfarin doses.¹⁶ The discrepancy between the ethnic groups with respect to genetic association with warfarin dose requirements is probably due in part to lower frequencies of the *CYP2C9**2 and *CYP2C9**3 alleles among African Americans, lower linkage disequilibrium (LD), and the greater range of genetic variation that exists in populations of African ancestry. Therefore, genetic variations that are predominant in or exclusive to African Americans may play a much larger role in determining warfarin dose response in African Americans while remaining difficult or impossible to detect in individuals of other ethnicities.

In the hope of finding novel variations that may explain more of the variability in dose among African Americans, we used a resequencing strategy that prioritized regions by both evolutionary conservation and transcriptional binding prediction. This technique has proven to be a powerful tool in identifying important areas to be investigated.^{17,18} By expanding our SNP discovery efforts to include novel variants, not only in coding regions but also in noncoding regions and in the region 10 kb upstream from the gene, we were able to identify SNPs that significantly affect warfarin dose, independent of the previously identified SNPs. This method also allowed us to find variations that could not be found using other comprehensive association methods, such as genome-wide association.

Results

Patient demographics

The characteristics for both cohorts of the study are shown in Table 1. For the discovery cohort, 140 subjects were recruited from the University of Chicago Anticoagulation clinic between June and December of 2008. Of these, 3 were excluded because of poor quality of the DNA sample and 14 failed to attain stable warfarin dose. Ultimately, 122 patients were included in the discovery cohort. The validation cohort consisted of 207 African-American subjects recruited from the University of Illinois Medical Center in Chicago, as previously described.¹⁹ Of the demographic information collected at both sites, only sex was significantly different between the two cohorts. All subjects were typed for ancestry informative markers to ensure that they were of African descent, with no significant difference in percentage of African ancestry between sites.

Associations in the discovery cohort

On regression analysis, the clinical covariates associated with warfarin dose in the discovery cohort were age ($P = 12\text{e-}06$), weight ($P = 0.00159$), and history of venous thromboembolism ($P = 0.000876$). Similarly, in the validation cohort, the covariates associated were age ($P = 7.22\text{e-}09$), weight ($P = 1.06\text{e-}06$), and indication of deep-vein thrombosis/pulmonary embolism ($P = 0.0425$).

Of the SNPs genotyped in the discovery cohort, six SNPs showed departure from Hardy–Weinberg equilibrium ($P < 0.05$) and were removed from the analysis. Additive linear association in the discovery cohort showed significant associations between warfarin dose and 24 SNPs in both *CYP2C9* and *VKORC1*, without correction for multiple testing ($P < 0.05$). Table 2 shows the genotype/phenotype associations found in this cohort. All the *VKORC1* variants previously shown to have been associated with warfarin dose in

Caucasians (*VKORC1*-1639, 1173, 1542, 7566, and 3730) were replicated in this cohort,^{14,20} in agreement with previous studies in African Americans.²¹ Of the known functional *CYP2C9* variants, only *CYP2C9**3, *8, and *11 showed significant associations. In addition, several novel SNPs were identified as being significantly associated.

Associations in the validation cohort

Because of the limited size of the discovery cohort, all the significant SNPs as well as known functional *CYP2C9* SNPs were genotyped in the validation cohort to validate these genetic associations. Of the original 26 SNPs found in the discovery cohort, 7 SNPs in *CYP2C9* and 8 SNPs in *VKORC1* remained significantly associated with dose (after correction for multiple testing). Again, all the SNPs in *VKORC1* previously known to be associated with warfarin dose remained significant, as well as *CYP2C9**2 and *8, with *CYP2C9**3 and *11 trending toward significance. We were also able to validate the associations of four novel SNPs in *CYP2C9* (15858, 18786, 19879, and 25706) and three novel SNPs in *VKORC1* (*VKORC1*-8868, *VKORC1*-8191, and *VKORC1*-3427). Table 2 highlights these novel SNP associations.

Regression model in the combined data set

To determine the contribution of both genetic and nongenetic variables to warfarin dose, we conducted a univariate analysis followed by a multivariate regression analysis. We combined the *CYP2C9**2, *3, *5, *8, and *11 variants into a single group of *CYP2C9* star variant alleles, on the basis of earlier data.²² Several clinical covariates showed association with warfarin dose; however, after age and weight were added to the model (the most significant clinical predictors), only indication of deep-vein thrombosis/pulmonary embolism remained significant. After the addition of the clinical covariates to the model, only *VKORC1*-1173, the *CYP2C9* star variants, 18786, and *VKORC1*-8191 were retained in the model. Both of the last two SNPs were associated with increased weekly dose, with *VKORC1*-8191 showing a 5.2 mg/week increase in dose for each allele and the *CYP2C9* 18786 SNP showing a 3.7 mg/week increase in dose for each allele. Table 3 shows the entry of each SNP into the model with the univariate r^2 , the multivariate R^2 after entry into the model, and the P value of each SNP upon entry. Through this analysis we were able to explain ~40% of the variability in warfarin dose in African Americans.

The most significant genetic predictor in the regression analysis was *VKORC1*-1173, which explained ~9% of the variability in dose, similar to the results of other studies in African Americans.²² The presence of a *CYP2C9* star variant allele was the next best predictor of warfarin dose, accounting for ~6% of the variability. Our analysis also shows that two novel SNPs (18786 in *CYP2C9* and *VKORC1*-8191) were both highly predictive of warfarin dose ($P = 0.00413$ and 0.0346 , respectively), even after the addition to the model of *VKORC1*-1173 and the *CYP2C9* star variants.

Comparison with the available models

To determine whether our model was significantly better than existing models at estimating the variability in warfarin dose, we first compared it with the algorithm published by the International Warfarin Pharmacogenetic Consortium (IWPC) to determine the amount of variability that this algorithm explained in our data set. This analysis showed that the IWPC algorithm explained 25% of the variability in warfarin dose (Table 3), in agreement with the results of the IWPC's own African-American subgroup analysis (25.9% in the IWPC discovery set and 19% in the validation set).¹⁴ We then investigated whether our findings could be replicated by chance alone. We used the covariates that were significantly associated in our data set as well as the known *VKORC1* SNPs and the *CYP2C9* star variants but permuted the genotype/phenotype relationship between our novel SNPs

(*VKORC1-8191* and 18786) and warfarin dose to determine whether we would find an R^2 of equal or greater value than our observed R^2 . In 10,000 such simulation data sets, we saw that the observed R^2 value equaling the 40% showed a highly significant improvement in percentage of variability explained ($P = 0.0002$). The expected value in 10,000 simulations is 36% (consistent with the maximum percentage of explained variability obtained thus far in prior studies).²² This strongly suggests that the significant increase in percentage is attributable to the inclusion of our novel SNP associations and greatly decreases the possibility that our findings were obtained by chance alone.

DISCUSSION

Most association studies of both *VKORC1* and *CYP2C9* with warfarin dose have been conducted in predominantly Caucasian populations. In the studies that included African Americans, investigators typically genotyped only the variants shown to affect warfarin dose in Caucasians.^{14,21,23-25} Prior to the current study, SNP discovered in African Americans had yet to be used in warfarin pharmacogenetics or in any currently published warfarin algorithms.^{14,15,23} Our study aimed to fill this gap in knowledge.

Previous SNP discovery efforts in *VKORC1*, such as those by Rieder *et al.*, involved resequencing the entire gene in three populations.²⁰ However, all associations with warfarin dose were made with haplotypes found in Caucasians. Groups of haplotypes associated with high and low warfarin doses were labeled haplogroups A and B, respectively. Most subsequent investigations have genotyped only one or two SNPs in *VKORC1* that differentiated between these haplogroups. In a more comprehensive study, Limdi *et al.* investigated the common SNPs and haplotypes in African Americans²¹ found within 5 kb of the gene start site. The findings of this study, which replicated many of those in the original study by Rieder *et al.*, concluded that the best predictors of warfarin dose in African Americans were *VKORC1-1639* or *VKORC1-1173*. However, this study was still restricted to the previously explored regions of *VKORC1* and did not take into account the possibility that regulation may occur further than 5 kb upstream of the gene.

Our study is distinct from previous investigations in that it used a targeted resequencing strategy. This not only allowed for novel SNP discovery but also enabled us to identify SNPs that would not be found through other comprehensive association methods such as genome-wide association analysis. The SNP *VKORC1-8191*, a novel association in *VKORC1*, is not present on any high-throughput SNP platform. Therefore, only through resequencing would this association with warfarin dose be found. In addition, by resequencing we gain a better understanding of the LD structures of these genes in African Americans, outside of the previously investigated regions. Furthermore, our resequencing approach strongly demonstrates that the inclusion of SNPs located further upstream of reproducibly associated genes (as compared with previous studies) significantly increases the proportion of explained variability in warfarin dose requirements.

In the ethnicity-specific model method, we found that two novel SNPs, along with *VKORC1-1173*, *CYP2C9* star variants, and clinical factors, explained 40% of the variability in warfarin dose. Factors such as smoking status and the use of amiodarone or statin drugs showed no association in our cohorts, in contrast to the findings of previous studies.^{14,15} This discrepancy may suggest either that different demographic and clinical factors affect warfarin dose in African Americans or that these are surrogate markers for unknown causative clinical variables that are not as well correlated in African Americans as in Caucasians. However, the discrepancy may also reflect inadequate power to detect these effects in our cohort.

Previous reports have suggested that *VKORC1-1639* is the causative SNP in the highly correlated LD block in Caucasians.²⁶ Because of the high LD between *VKORC1-1173* and *-1639* ($r^2 = 0.9$), we, like others,²⁷ were unable to determine which SNP is the true causative variant in our population. However, our findings are consistent with those of previous studies,²⁶ given that after the inclusion of *VKORC1-1173* no other SNPs in this LD block were added to the model. Additionally, the novel variant *VKORC1-8191*, which was significant in all regression analyses, is not in LD with any of the SNPs in this LD block (Figure 1).

We corrected for multiple testing in the validation cohort by means of the permutation method. The Bonferroni correction is commonly used for this purpose, but it assumes that the tested SNPs are independent. Even after binning SNPs according to LD, residual LD remains, invalidating this assumption in Bonferroni correction. Permutation allows us to preserve the LD between tested SNPs while correcting for multiple testing. While not as conservative as the Bonferroni correction, this method is extensively used in genome-wide association studies.²⁸

The IWPC published the largest cohort study ($n = 4,043$) examining the predictability of warfarin dose in a diverse population. This study contained 353 African Americans and used *VKORC1-1639* and *CYP2C9*2* and **3* as well as age, weight, height, race, enzyme inducer status, and amiodarone to predict dose. The dosing model including these variables was able to predict 47% of the dose variability in the entire IWPC cohort. However, when examined by race, only 25.9% of the variability in dose (19% in the IWPC validation cohort) was accounted for in African Americans.¹⁴ We were able to reproduce these findings in our data set, as seen in Table 3. This lower R^2 value is not surprising, given that *VKORC1-1639* and *CYP2C9*2* and **3* are present at much lower frequencies in African Americans as compared with Caucasians²⁹ and would therefore exert less influence on dose. Also, the IWPC algorithm did not incorporate *CYP2C9*5*, **8*, and **11*, which occur at higher frequency, or even exclusively, in African Americans. The inclusion of African American-specific *CYP2C9* star variants into the regression analysis increased the explained percentage of variability to 37%, which is in agreement with the findings of Cavallari *et al.*²² Even more importantly, we were able to show, through our simulation study ($N = 10,000$), that the $R^2 = 40\%$ explained by our model is significantly greater ($P = 0.0002$) than would be expected by chance alone, with a sample size equivalent to the IWPC's African-American cohort. While this increase may appear modest, it should be noted that the *CYP4F2* variant contributed only a 2-4% increase in multivariate R^2 (refs. 30-32) in Caucasians and has been added to the online database WarfarinDosing.org. In addition, we were able to increase the overall R^2 by 15% over the previously published algorithms, providing a more comprehensive picture of the genetic variability affecting warfarin dose in African Americans. It is worth noting that both novel alleles are associated with increase in dose. Given that African Americans require higher warfarin doses than do Caucasians,²⁴ these SNPs may improve the prediction of higher doses in African Americans. However, further modeling and validation will be needed to determine the accuracy of these SNPs in predicting warfarin dose.

Given that we used a haplotype-tagging SNP approach, we cannot be certain whether SNP *VKORC1-8191* is the true causative SNP. However, our resequencing efforts allowed some interpretation of our findings. SNP *VKORC1-8191* is ~8 kb upstream from the transcriptional start site. This SNP has a minor-allele frequency of prevalence of 47% in African Americans and is associated with higher dose in both cohorts. In the Coriell African-American samples, we found that *VKORC1-8191* is in strong LD ($r^2 = 0.9$) with a novel SNP 9 kb upstream of *VKORC1* (termed SNP 5018 in our study, no rs number assigned). *VKORC1-8191* is also in LD with rs7196161 ($R^2 = 0.7$), found 4.7 kb upstream. However, this known *VKORC1* variant has been shown to have no association with

warfarin dose and shows low LD with *VKORC1-1639* and *VKORC1-1173* in our study (see Figure 2) as well as in previously published work.²¹

SNP18786 (rs7089580) is a novel association found in the intron 3 of *CYP2C9*. This SNP is in LD ($r^2 = 0.98$) with SNP 12251 (intron 1), 15413 (intron 2), and 25706 (intron 5) found in our resequencing study. From our prediction of putative transcriptional binding sites, 25706 is within 19bp of a predicted transcriptional binding cluster in intron 5. These clusters may signal a gene regulatory region—as distinct from single sites, which occur at random throughout the gene. Supplementary Figure S2 online shows the location of these predicted clusters within *CYP2C9*. Further molecular studies are needed to determine whether this site is a functional regulator of transcription. SNP 18786 is not in LD with any known functional *CYP2C9* star variants (see Figure 2).

In conclusion, we have significantly improved the amount of variability that can be accounted for in warfarin dose variability among African Americans. Even the most current dosing algorithms include only the known *CYP2C9*, *VKORC1*, and *CYP4F2* variants, even though numerous studies have shown these variants to be less predictive in populations of African ancestry. By using resequencing data, we go beyond the information obtained through genotyping. We are able to throw light on both haplotype structure and potential causative variants in LD with the associated SNP in African Americans. Having performed this study—the first to investigate novel variations in African Americans using resequencing data—we hope to highlight the need for the use of resequencing along with GWAs to discover new associations, as well as to begin modifying warfarin dose algorithms to better accommodate African-American patients.

METHODS

Patient population

The discovery cohort was recruited from the University of Chicago's Anticoagulation clinic. The validation cohort was recruited from the anticoagulation clinics at the University of Illinois Medical Center at Chicago. The inclusion criteria were age ≥ 18 years, African-American race by self report, and treatment with a stable dose of warfarin, defined as the same dose for at least three consecutive clinic visits that produced an international normalized ratio within the therapeutic range. All international normalized ratio measurements were made on a ProTime Microcoagulation System (ITC, Edison, NJ), and all values >4 were remeasured via the hospital laboratory services. In the discovery cohort, one 10-ml blood sample was taken from each subject by venipuncture and placed in purple-top Vacutainer tubes containing ethylenediaminetetraacetic acid and frozen at -20°C for DNA extraction. Buccal cell samples were collected from participants enrolled from the University of Illinois Medical Center, as previously described.¹⁹ All protocols were reviewed and approved by the institutional review boards of the University of Chicago and the University of Illinois at Chicago. All patients provided written informed consent for the collection of samples and subsequent analysis.

Resequencing of *VKORC1* and *CYP2C9*

Genetic regions to be resequenced within *CYP2C9* were prioritized using both evolutionary conservation and putative transcriptional cluster prediction. We compared six species, covering 300 million years of evolution, to identify areas of deep sequence conservation. In addition, informatics tools such as the Cister: Cis-Element Cluster Finder^{33,34} were used to identify clusters of hepatic transcription factor binding sites within the noncoding sequence. Supplementary Figures S1 and S2 online show the regions selected for resequencing. *VKORC1* has been completely resequenced in several populations, and the results are

available to the public (<http://pga.gs.washington.edu>), albeit limited to 5 kb upstream of the gene. To better understand the role of gene regulation, we examined variations up to 10 kb upstream of the gene.

After determination of the priority regions, 24 African-American Coriell samples were resequenced for both genes via PCR amplification of 1 kb overlapping fragments, used for covering contiguous sections as previously described.³⁵ Primers were designed using Primer3³⁶ when possible. Because of the high sequence similarity between members of the *CYP2C* gene family, some primers were designed by hand to specifically amplify the *CYP2C9* sequence (details of the sequencing primers are in Supplementary Table S1 online).

For DNA sequencing, PCR products were purified using the QIAquick PCR Purification Kit (Qiagen, Valencia, CA), submitted to the Cancer Research Center DNA Sequencing Core Facility at the University of Chicago, and sequenced using ABI 3730XL DNA sequencers (Applied Biosystems, Foster City, CA). The Phred-Phrap-Consed package (Poly-phred, version 6.11) was used to assemble and analyze all sequences, as previously described.³⁵ Supplementary Tables S2 and S3 online show all the SNPs found in the resequencing study.

selecting tag SNPs and genotyping

To provide a more comprehensive approach, we employed a haplotype-tagging SNP design to capture a majority of the variation within *CYP2C9* and *VKORC1*. From the resequencing study, we determined the patterns of LD and haplotype structure using PHASE 2.0.³⁷ We then used the program LDSelect,³⁸ which bins SNPs together using $r^2 = 0.8$, to identify tagging SNPs within each of the genes with a minor-allele cutoff of 5%. Tagging SNPs were then genotyped in the discovery cohort along with the known functional variants *CYP2C9**2 (rs1799853), *CYP2C9**3 (rs1057910), *CYP2C9**5 (rs28371686), *CYP2C9**8 (rs7900194), *CYP2C9**11 (rs28371685), *VKORC1*-9041 (rs7294), *VKORC1*-7566 (rs2359612), *VKORC1*-1542 (rs8050894), *VKORC1*-1173 (rs9934438), and *VKORC1*-1639 (rs9923231). The discovery cohort was genotyped through SeattleSNPs Genotyping Support (http://pga.gs.washington.edu/genotyping_support.html) using a resequencing strategy targeting haplotype-tagging SNPs. The validation cohort was genotyped by PCR, pyrosequencing, and fragment analysis, as previously described,²² and by the Sequenom multiplex method or via direct sequencing (primers and experimental conditions can be found in Supplementary Table S4 and S5 online). One hundred individuals of known genotype were added to the genotyping platforms to ensure genotyping accuracy. Any SNP that showed discordance was removed from the analysis.

statistical analysis

Warfarin dose values were log₂ transformed to improve normality. All covariates were tested in each data set separately for association with warfarin dose, using a multivariate linear regression to determine covariates that were significantly associated with dose and to take into account any interactions between covariates. The covariates tested were age (years); weight (kg); height (cm); smoking status; use of aspirin, amiodarone, carbamazepine, phenytoin, rifampin, rifampicin, any statin drug, or any azole antifungal drug; comorbidities (hypertension and diabetes); indication for warfarin therapy; and percentage of West African ancestry.

SNPs genotyped in both cohorts were tested for departure from Hardy-Weinberg equilibrium, using the χ^2 -test. Any SNP that showed a departure from Hardy-Weinberg equilibrium ($P < 0.05$) was removed from the analysis. The discovery cohort, from the University of Chicago, was tested for additive linear association using a Student's *t*-test,

using age, weight, and indication of deep-vein thrombosis/pulmonary embolism as covariates. SNPs that showed modest association ($P = 0.05$) were then genotyped in the validation cohort and tested for association with $\log_2$ -transformed warfarin dose. All known functional SNPs in both genes were also genotyped in both cohorts to ensure that all associations with known functional variants were tested. All association P values in the validation cohort were corrected for multiple testing using a permutation method conducted in PLINK³⁹ (PLINK, version 1.07; <http://pngu.mgh.harvard.edu/purcell/plink/>) and were considered significant at a P value $= 0.05$. The permutation method preserves the correlations between SNPs (i.e., LD) and yields a less stringent correction for multiple testing than an approach that assumes independence of tests (such as the Bonferroni correction).

To determine the proportion of the phenotype explained by each variable (both genetic and nongenetic), we conducted a stepwise linear regression analysis on the combined data sets ($n = 330$). All SNPs and covariates were first tested for association by means of a univariate analysis. Any variables showing P values < 0.05 were tested in the forward selection stepwise regression, with the addition of significant clinical covariates first, followed by genotypes. All analyses were done using an additive model. Variables that were significant predictors of warfarin dose ($P < 0.05$) were retained in the model. All tests for statistical association were conducted in PLINK or R (<http://www.r-project.org>).

To determine whether our findings represented a significant improvement over previous findings regarding associations, we applied the IWPC model to our data set and calculated the resulting R^2 . For the purpose of assessing the level of significance of the R^2 explained by the novel associations detected by our resequencing strategy and included in our African-American regression model, we conducted simulations ($N = 10,000$) that created permuted data sets sampled under the null to generate an empirical distribution. Such an approach breaks the genotype-phenotype relationship, but preserves the correlations between SNPs in both the observed and permuted data sets. The number of times that a permuted data set's R^2 exceeds the observed R^2 yields an empirical P value for the significance of the improvement obtained from the novel associations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank E.P. Gengler for his editorial advice on this article. We would like to thank SeattleSNPs for providing genotyping support.

References

1. Hirsh J, et al. Oral anticoagulants: mechanism of action, clinical effectiveness, and optimal therapeutic range. *Chest*. 1998; 114:445S–69S. [PubMed: 9822057]
2. O'Reilly RA. The pharmacodynamics of the oral anticoagulant drugs. *Prog. Hemost. Thromb.* 1974; 2:175–213. [PubMed: 4137288]
3. Breckenridge A, Orme M, Wesseling H, Lewis RJ, Gibbons R. Pharmacokinetics and pharmacodynamics of the enantiomers of warfarin in man. *Clin. Pharmacol. Ther.* 1974; 15:424–430. [PubMed: 4821443]
4. Rettie AE, et al. Hydroxylation of warfarin by human cDNA-expressed cytochrome P-450: a role for P-4502C9 in the etiology of (S)-warfarin-drug interactions. *Chem. Res. Toxicol.* 1992; 5:54–59. [PubMed: 1581537]
5. Kaminsky LS, Zhang ZY. Human P450 metabolism of warfarin. *Pharmacol. Ther.* 1997; 73:67–74. [PubMed: 9014207]

6. Kamali F, et al. Contribution of age, body size, and CYP2C9 genotype to anticoagulant response to warfarin. *Clin. Pharmacol. Ther.* 2004; 75:204–212. [PubMed: 15001972]
7. Absher RK, Moore ME, Parker MH. Patient-specific factors predictive of warfarin dosage requirements. *Ann. Pharmacother.* 2002; 36:1512–1517. [PubMed: 12243598]
8. Franco V, Polanczyk CA, Clausell N, Rohde LE. Role of dietary vitamin K intake in chronic oral anticoagulation: prospective evidence from observational and randomized protocols. *Am. J. Med.* 2004; 116:651–656. [PubMed: 15121490]
9. Cushman M, Booth SL, Possidente CJ, Davidson KW, Sadowski JA, Bovill EG. The association of vitamin K status with warfarin sensitivity at the onset of treatment. *Br. J. Haematol.* 2001; 112:572–577. [PubMed: 11260056]
10. Lubetsky A, Dekel-Stern E, Chetrit A, Lubin F, Halkin H. Vitamin K intake and sensitivity to warfarin in patients consuming regular diets. *Thromb. Haemost.* 1999; 81:396–399. [PubMed: 10102468]
11. Holbrook AM, et al. Systematic overview of warfarin and its drug and food interactions. *Arch. Intern. Med.* 2005; 165:1095–1106. [PubMed: 15911722]
12. Gage BF, Fihn SD, White RH. Management and dosing of warfarin therapy. *Am. J. Med.* 2000; 109:481–488. [PubMed: 11042238]
13. Phillips KA, Veenstra DL, Oren E, Lee JK, Sadee W. Potential role of pharmacogenomics in reducing adverse drug reactions: a systematic review. *JAMA.* 2001; 286:2270–2279. [PubMed: 11710893]
14. Klein TE, et al. Estimation of the warfarin dose with clinical and pharmacogenetic data. *N. Engl. J. Med.* 2009; 360:753–764. [PubMed: 19228618]
15. Gage BF, et al. Use of pharmacogenetic and clinical factors to predict the therapeutic dose of warfarin. *Clin. Pharmacol. Ther.* 2008; 84:326–331. [PubMed: 18305455]
16. Blann A, Hewitt J, Siddiqui F, Bareford D. Racial background is a determinant of average warfarin dose required to maintain the INR between 2.0 and 3.0. *Br. J. Haematol.* 1999; 107:207–209. [PubMed: 10520043]
17. Wardrop SL, Brown MA. Identification of two evolutionarily conserved and functional regulatory elements in intron 2 of the human BRCA1 gene. *Genomics.* 2005; 86:316–328. [PubMed: 15990270]
18. Perera MA, et al. Prediction of CYP3A4 enzyme activity using haplotype tag SNPs in African Americans. *Pharmacogenomics J.* 2009; 9:49–60. [PubMed: 18825162]
19. Momary KM, Shapiro NL, Viana MA, Nutescu EA, Helgason CM, Cavallari LH. Factors influencing warfarin dose requirements in African-Americans. *Pharmacogenomics.* 2007; 8:1535–1544. [PubMed: 18034618]
20. Rieder MJ, et al. Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose. *N. Engl. J. Med.* 2005; 352:2285–2293. [PubMed: 15930419]
21. Limdi NA, et al. VKORC1 polymorphisms, haplotypes and haplotype groups on warfarin dose among African-Americans and European-Americans. *Pharmacogenomics.* 2008; 9:1445–1458. [PubMed: 18855533]
22. Cavallari LH, et al. Genetic and clinical predictors of warfarin dose requirements in African Americans. *Clin. Pharmacol. Ther.* 2010; 87:459–464. [PubMed: 20072124]
23. Ferder NS, et al. Ability of VKORC1 and CYP2C9 to predict therapeutic warfarin dose during the initial weeks of therapy. *J. Thromb. Haemost.* 2010; 8:95–100. [PubMed: 19874474]
24. Limdi NA, et al. Influence of CYP2C9 and VKORC1 on warfarin dose, anticoagulation attainment and maintenance among European-Americans and African-Americans. *Pharmacogenomics.* 2008; 9:511–526. [PubMed: 18466099]
25. Schelleman H, et al. Dosing algorithms to predict warfarin maintenance dose in Caucasians and African Americans. *Clin. Pharmacol. Ther.* 2008; 84:332–339. [PubMed: 18596683]
26. Wang D, Chen H, Momary KM, Cavallari LH, Johnson JA, Sadee W. Regulatory polymorphism in vitamin K epoxide reductase complex subunit 1 (VKORC1) affects gene expression and warfarin dose requirement. *Blood.* 2008; 112:1013–1021. [PubMed: 18523153]
27. Limdi NA, et al. Warfarin pharmacogenetics: a single VKORC1 polymorphism is predictive of dose across 3 racial groups. *Blood.* 2010; 115:3827–3834. [PubMed: 20203262]

28. Cantor RM, Lange K, Sinsheimer JS. Prioritizing GWAS results: A review of statistical methods and recommendations for their application. *Am. J. Hum. Genet.* 2010; 86:6–22. [PubMed: 20074509]
29. Limdi N, Goldstein J, Blaisdell J, Beasley T, Rivers C, Acton R. Influence of CYP2C9 Genotype on warfarin dose among African American and European Americans. *Per. Med.* 2007; 4:157–169. [PubMed: 19802360]
30. Caldwell MD, et al. CYP4F2 genetic variant alters required warfarin dose. *Blood.* 2008; 111:4106–4112. [PubMed: 18250228]
31. Sagrieya H, et al. Extending and evaluating a warfarin dosing algorithm that includes CYP4F2 and pooled rare variants of CYP2C9. *Pharmacogenet. Genomics.* 2010; 20:407–13. [PubMed: 20442691]
32. Takeuchi F, et al. A genome-wide association study confirms VKORC1, CYP2C9, and CYP4F2 as principal genetic determinants of warfarin dose. *PLoS Genet.* 2009; 5:e1000433. [PubMed: 19300499]
33. Frith MC, Hansen U, Weng Z. Detection of cis-element clusters in higher eukaryotic DNA. *Bioinformatics.* 2001; 17:878–889. [PubMed: 11673232]
34. Wingender E, et al. TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res.* 2000; 28:316–319. [PubMed: 10592259]
35. Thompson EE, Kuttub-Boulos H, Yang L, Roe BA, Di Rienzo A. Sequence diversity and haplotype structure at the human CYP3A cluster. *Pharmacogenomics J.* 2006; 6:105–114. [PubMed: 16314882]
36. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* 2000; 132:365–386. [PubMed: 10547847]
37. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* 2001; 68:978–989. [PubMed: 11254454]
38. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am. J. Hum. Genet.* 2004; 74:106–120. [PubMed: 14681826]
39. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 2007; 81:559–575. [PubMed: 17701901]
40. Gabriel SB, et al. The structure of haplotype blocks in the human genome. *Science.* 2002; 296:2225–2229. [PubMed: 12029063]

Figure 1.

Linkage disequilibrium (LD) of *VKORC1* single-nucleotide polymorphisms (SNPs) in the combined data set. The LD map shows the LD (reported as a percentage) with the SNPs in *VKORC1* that showed association in the discovery cohort. The shaded boxes represent the locations of the three exons in this gene, with the arrow showing the orientation of the gene. Two blocks of LD are seen, comprising *VKORC1*-1173 (15414) and *VKORC1*-1639 (12603) in one block and *VKORC1*-7566 and *VKORC1*-9041 in the second block (calculated using D' by Gabriel *et al.*).⁴⁰ LD extends between these two blocks in the combined African-American data set. The novel SNP associated with warfarin dose, *VKORC1*-8191, is not in LD with either of these blocks. The LD map color intensity is based on the r^2 , with the darker shades representing higher LD. The values in the boxes represent the pairwise r^2 between markers.

Figure 2.

Linkage disequilibrium (LD) of cytochrome P-450 2C9 (CYP2C9) single-nucleotide polymorphisms (SNPs) in the combined data set. The LD map shows the LD (reported as a percentage) with the SNPs in CYP2C9 that showed association in the discovery cohort. The shaded boxes represent the locations of the nine exons in this gene, with the arrow showing the orientation of the gene. Only one LD block is seen, calculated using D' by Gabriel *et al.*⁴⁰ The SNP 18786 in intron 3 is in strong LD with SNP 25706 in intron 5. There is little or no LD between 18786 and CYP2C9 star variants, CYP2C9*2, *3, *5, *8, and *11. The LD map color intensity is based on the r^2 , with the darker shades representing higher LD and the solid black box representing perfect LD. The values in the boxes represent the pairwise r^2 between markers.

Table 1

Clinical demographics of the discovery and validation cohorts

Variable	Discovery cohort (n = 122)	Validation cohort (n = 207)	P value ^a
<i>Sex</i>			
Male	34%	24%	0.0486
Female	66%	76%	
Age (mean)	59	56	0.186
Weight, kg (mean)	88.3	93.6	0.095
Height, cm (mean)	168.8	167.1	0.156
Warfarin dose, mg/week (mean±SD)	44.4±21.0	44.0±16.9	0.853
<i>Smoking status</i>			
Yes	11.9%	8.6%	0.45
No	88.1%	91.4%	
<i>Indication</i>			
DVT/PE	58%	50%	0.13
Atrial fibrillation	22%	16%	0.11
MVR	8%	7%	0.942
<i>Medications</i>			
CYP2C9 inducers	2%	3%	0.773
Aspirin	28%	28%	0.996
Amiodarone	2%	2%	0.844
Statins	47%	47%	0.279

CYP2C9 inducers: carbamazepine, phenytoin, rifampin, or rifampicin. Statins: simvastatin, atorvastatin, fluvastatin, lovastatin, pravastatin, or rosuvastatin.

CYP2C9, cytochrome P-450 2C9; DVT/PE, deep-vein thrombosis/pulmonary embolism; MVR, mitral valve replacement.

^aP value from the χ^2 -test or paired *t*-test comparing the discovery cohort and the validation cohort.

Table 2

significant associations found in each cohort

Gene	snp name/rs number ^a	SNP number	Location	MAF (%)	Discovery cohort <i>P</i> value ^b	Replication cohort <i>P</i> value ^c
<i>CYP2C9</i>	NA	12479	Intron 1	3	0.0357	0.184
	Rs2860905	15858	Intron 3	22	0.0434	0.00007
	Rs7089580	18786	Intron 3	23	0.0272	0.00019
	Rs6583964	19497	Intron 3	33	0.0354	0.524
	NA	19879	Intron 3	2	0.0444	0.0459
	Rs74150726	25641	Intron 5	4	0.00297	0.238
	Rs35511771	25706	Intron 5	24	0.0316	0.00008
	Rs2017319	62198	Exon 9	14	0.00532	0.342
	NA	65312	3' UTR	2	0.0325	0.344
	CYP2C9*2/rs1799853	15610	Exon 3	2	0.969	0.0041
	CYP2C9*3/rs1057910	54616	Exon 7	1	0.0507	0.0438
	CYP2C9*5/rs28371686	54621	Exon 7	1	0.0930	0.242
	CYP2C9*8/rs7900194	15629	Exon 3	6	0.0514	0.0208
	CYP2C9*11/rs28371685	54544	Exon 7	4	0.0302	0.0791
<i>VKORC1</i>	VKORC1-8868/rs62057090	5374	5' UTR	29	0.000213	0.0051
	NA	5663	5' UTR	2	0.00336	0.517
	VKORC1-8191/rs61162043	6051	5' uTR	47	0.00008	0.0375
	rs7187995	7475	5' UTR	32	0.00208	0.0807
	rs58826038	7507	5' UTR	14	0.0212	0.336
	VKORC1-3427/rs59502288	10995	5' uTR	42	0.00906	0.0486
	VKORC1-1639/rs9923231	12603	5' UTR	11	0.00633	0.00007
	VKORC1 1173/rs9934438	15414	Intron 1	12	0.00451	0.00007
	VKORC1 1542/rs8050894	15783	Intron 2	30	0.00664	0.0002
	VKORC1 7566/rs2359612	16497	Intron 2	21	0.00526	0.0007
	VKORC1 9041/rs7294	17972	3' UTR	44	0.00703	0.0031
	NA	18005	3' UTR	>1	0.00771	0.502

All novel SNP associations in each gene are highlighted in boldface.

CYP2C9, cytochrome P-450 2C9; MAF, minor-allele frequency; NA, not applicable; SNP, single-nucleotide polymorphism; UTR, untranslated region.

^a rs Number is from dbSNP build 37.1.

^b *P*value determined from Student's *t*-test.

^c The *P*value was corrected for multiple testing via the permutation method conducted in PLINK.

Table 3

Comparison of warfarin dosing models in African Americans

Model	Entry into model	Variable	Gene	Univariate r^2	Adjusted R^2 after entry	P
IWPC model	1	Age in decades	NA	—	14.7	5.31 E-13
	2	Height (cm)	NA	4.5	15.4	0.00298
	3	Weight (kg)	NA	9.0	17.6	0.0034
	4	<i>VKORC1</i> -1639	<i>VKORC1</i>	9.1	23.3	4.86 E-6
	5	<i>CYP2C9</i> *2 (rs1799853)	<i>CYP2C9</i>	1.3	24.9	0.047
	6	<i>CYP2C9</i> *3 (rs1057910)	<i>CYP2C9</i>	1.6	25.3	0.11
	7	Race	NA	ND	ND	ND
	8	Enzyme inducer status	NA	2.0	25.6	0.18
	9	Amiodarone status	NA	0.003	25.8	0.24
African-American model	1	Age in years	NA	—	14.7	5.31 E-13
	2	Weight (kg)	NA	9.0	20.5	2.84 E-6
	3	DVT/PE	NA	4.1	22.0	0.00747
	4	<i>VKORC1</i> 1173	<i>VKORC1</i>	9.9	29.4	1.40 E-8
	5	<i>CYP2C9</i> star alleles	<i>CYP2C9</i>	5.6	37.1	1.84e-07
	6	<i>VKORC1</i> -8191 (rs61162043)	<i>VKORC1</i>	3.2	38.6	0.00413
	7	18786 (rs7089580)	<i>CYP2C9</i>	0.4	40.0	0.0346

Both univariate and multivariate coefficients are expressed in term of percentage. Multivariate R^2 and P value were calculated in our combined data set using the variables published in the IWPC study. 14
 Race was not considered since all subjects were African American. Enzyme inducer status = 2 if patient is taking carbamazepine, phenytoin, rifampin or rifampicin, otherwise 1. Amiodarone status = 2 if patient is taking amiodarone, otherwise 1. DVT/PE = 2 if the indication for warfarin use is deep-vein thrombosis and/ or pulmonary embolism, otherwise 1. *CYP2C9* star allele = 2 the presence of any of the following alleles: *CYP2C9**2, *3, *5, *8, or *11, otherwise 1. P values presented in the table are associated with the variable upon entry into the model.

CYP2C9, cytochrome P-450 2C9; IWPC, International Warfarin Pharmacogenetic Consortium; NA, not applicable; ND, not determined.