Worldwide allele frequency distribution of four polymorphisms associated with warfarin dose requirements

Kendra A Ross, Abbigail W Bigham, Melissa Edwards, Agnes Gozdzik, Guilherme Suarez-Kurtz and Esteban J Parra

The aim of the study is to improve our understanding of the worldwide allele frequency distribution of four genetic polymorphisms known to influence warfarin dosing (VKORC1 rs9923231, CYP2C9 rs1799853, CYP2C9 rs1057910 and CYP4F2 rs2108622). These four polymorphisms were genotyped in the Human Genome Diversity Project-Centre Etude Polymorphism Humain (HGDP-CEPH) worldwide sample, as well as in a sample of individuals of European, East Asian and South Asian ancestry living in Canada. The VKORC1 rs9923231 single nucleotide polymorphism (SNP) showed an extensive geographic differentiation, with the derived T allele appearing at very high allele frequencies in East Asian populations. Results from several tests of positive selection indicate that this unusual distribution may be the result of positive selection in East Asia. Understanding the worldwide distribution of markers determining warfarin dosing is important for the future application of pharmacogenomic-based algorithms to different population groups.

Journal of Human Genetics (2010) 55, 582–589; doi:10.1038/jhg.2010.73; published online 17 June 2010

Keywords: algorithm; allele frequency distribution; dose; natural selection; VKORC1 (rs9923231), CYP2C9 (rs1799853 and rs1057910) and CYP4F2 (rs2108622) polymorphisms; warfarin

INTRODUCTION

The anticoagulant warfarin is administered for the prevention and treatment of thromboembolic disorders that have the potential of causing pulmonary embolisms and strokes, and for reducing the risk of death from heart attacks. With a 1.45-fold increase in prescriptions between 1998 and 2004, warfarin has become the most broadly used anticoagulant worldwide. Although warfarin therapy has the potential to be highly effective, there are risks of incorrect dosing that are lethal in some cases. Among patients over 65 years of age, warfarin is one of three drugs responsible for approximately one-third of adverse drug events treated in the emergency room. In addition, US death certificates reviewed by Wysocki et al. indicate that between 2003 and 2004, anticoagulants were the leading cause of death among drugs that cause adverse effects. These statistics can be attributed to the high interindividual variability in warfarin dosing requirements and the combination of genetic and non-genetic factors that contribute to an individual's sensitivity to the medication.

Dosing of warfarin attempts to maintain the international normalized ratio of prothrombin time between 2.0 and 3.0, in between which there is minimal risk of over or under anticoagulation. With a narrow therapeutic window and large interindividual variability, clinical factors, such as age, height, sex, diet and drug–drug interactions explain 12% to 15% of the observed variability in warfarin dosing. Genetic factors have also been shown to affect warfarin maintenance. Previous studies have demonstrated that two genes, VKORC1 and CYP2C9, which are involved in the vitamin K-dependent clotting pathway, explain an additional 30–54% of variance observed in warfarin dosing.

Warfarin exists as a mixture of two stereoisomers, R and S. S-warfarin is approximately three times as active as the R-enantiomer and works by antagonizing the vitamin K-dependent clotting pathway. Vitamin K epoxide reductase subunit 1 (VKORC1), having a major role in the vitamin K pathway, is the target protein of warfarin. Within the VKORC1 gene, the single nucleotide polymorphism (SNP) rs9923231 has been extensively studied among European and Asian populations. Original research by Wadelius showed that VKORC1 predicts 30% of the variance observed in warfarin dose. Another critical gene is CYP2C9, which codes for the isoenzyme responsible for metabolizing S-warfarin. The gene CYP2C9 has been cited as accounting for 80–85% of the elimination of warfarin. Specifically two variants of this highly polymorphic gene, CYP2C9*2 (rs1799853) and CYP2C9*3 (rs1057910) have been shown...
to reduce warfarin metabolism and lower the mean warfarin dose requirements, thereby significantly increasing an individual’s risk of having complications related to bleeding. A review by Schelleman et al. illustrates the large number of CYP2C9*2 and 3 allele frequency studies that have focused on populations of European ancestry, in which these genetic variants are common. Together they have been found to account for 12–18% of the variation in warfarin dose. Recent research has identified a third gene, CYP4F2, as influencing an individual’s sensitivity to warfarin. CYP4F2 codes for a vitamin K oxidase that causes a decrease in the metabolism of vitamin K. Several studies have shown that carriers of the derived T allele at the non-synonymous SNP rs2108622 require higher warfarin doses than subjects with the ancestral C allele. Studies on European populations have indicated that CYP4F2 rs2108622 may account for an additional 1–7% of observed variation in warfarin dosing.

The effect of this CYP4F2 polymorphism seems to be substantially smaller than the effect of the VKORC1 and CYP2C9 variants, and the clinical relevance of CYP4F2 has recently been questioned. The traditional approach for warfarin dosing is based on an initial fixed dose, followed by stabilization through trial and error. Typically, in individuals of Asian ancestry the initial dose is 3.5 mg and in individuals of European ancestry is 5 mg or 10 mg. In this respect, it is important to mention that several studies have characterized individuals of African ancestry as requiring higher average warfarin doses and Asians as requiring lower doses than Europeans.

Research has recognized the influence that genetics has on the metabolism of warfarin and led to a revision of warfarin labels by the US Food and Drug Administration in 2007. New package inserts have been designed to include information on the effects of VKORC1 and CYP2C9 polymorphisms. Replacing the fixed dose approach, multiple regression models and various dosing algorithms, which incorporate VKORC1 and CYP2C9 variants along with clinical and environmental information have been developed and successfully explain anywhere from 53–65% of variation in warfarin dose among Europeans and Asians. In general, there has been reduced success in groups, such as African Americans, in which dosing algorithms have only explained up to 36% of the variation in warfarin dose. This has been attributed to greater genetic diversity among people of African ancestry and the possible influence of other variants in the VKORC1 and CYP2C9 genes. In contrast, a study examining the predictive capability of a new algorithm for warfarin dosing in Brazil has had much more success. When applied to Brazilian patients of self-identified ‘race/color’, researchers observed that it explained 51% of the variability in warfarin dose.

The second set of samples used in this study is comprised of 316 subjects that were recruited between 2007 and 2009 at the Molecular Anthropology Laboratory, University of Toronto at Mississauga (Mississauga, Canada). Geographic origin was determined through a questionnaire whereby questions pertaining to the participant’s, their parent’s and grandparent’s place of birth were recorded. Information provided was used to classify individuals into three broad geographic regions: 121 from Europe, 100 from East Asia and 95 from South Asia. The study was approved by the University of Toronto Health Sciences Research Ethics Board.

**Genotyping**

Four SNPs known to influence the metabolism of warfarin (rs9923231, rs1799853, rs1057910 and rs2108622) were genotyped at KBiosciences (http://www.kbioscience.co.uk/) using a KASPar assay based on fluorescent genotyping and allele-specific PCR. To verify genotyping quality, 170 samples were genotyped as blind duplicates, and the concordance rate between the samples and duplicates was 100%.

**Statistical analysis**

Tests evaluating deviations from Hardy–Weinberg proportions and hierarchical analyses of molecular variance were carried out with the program Arlequin. Tests of allele frequency differences between pairs of populations were done with the program Arlequin, or alternatively, with the program Genepop (http://kimura.univ-montp2.fr/~rousset/Genepop.htm). Three different tests of selection were used to evaluate evidence of positive selection in the HapMap East Asian sample for the gene VKORC1: the locus-specific branch length (LSBL) test, the log of the ratio of heterozygosities (lnRH) test and Tajima’s D test. The results reported here are based on a genome-wide analysis of the HapMap data, which includes approximately 4 million SNPs (http://hapmap.ncbi.nlm.nih.gov/). The LSBL test evaluates whether genetic markers within a genomic region show unusual levels of differentiation with respect to the genome average. This test apportions the genetic variation observed in East Asian, European and West African populations for each SNP, and identifies markers with high levels of genetic differentiation in the East Asian sample. The lnRH test highlights genomic regions with low levels of genetic diversity in the population of interest in comparison with other population groups. This statistic was calculated for a two-way population comparison between East Asians and Europeans, and East Asians and West Africans, using an overlapping sliding window size of 100,000 base pairs (bp) and moving in 25,000-bp increments along a chromosome. Regions of the genome with negative Tajima’s D values (excess of rare alleles with respect to neutral expectations) are also a hallmark of positive selection. However, negative values of D can also result from seven geographic regions: Africa, the Middle East, Europe, South/Central Asia, East Asia, the Americas and Oceania. Additionally, we also analyzed these polymorphisms in 316 DNA samples of individuals of European, East Asian and South Asian ancestry living in Canada. We show that the VKORC1 rs9923231 polymorphism shows a unique allele frequency distribution, with the derived T allele showing very high allele frequencies in East Asian populations. Given this unusual distribution, we carried out three tests of selection in the HapMap East Asian sample, and report that the VKORC1 gene is an outlier for two of the tests, probably as a result of positive selection in East Asian populations.

**MATERIALS AND METHODS**

**DNA samples**

We obtained DNA samples of 963 individuals from 52 world populations from the Foundation Jean Dausset-CEPH in Paris. Each individual’s geographic origin was available through a database containing information on HGDP-CEPH participants. Individuals were grouped into seven geographic regions: Africa (N = 108), Middle East (N = 164), Europe (N = 159), South/Central Asia (N = 204), East Asia (N = 233), America (N = 64) and Oceania (N = 31). Information on sample sizes of individual populations is available in Supplementary Table 1. All samples met the ethical requirements set forth by the HGDP and the Morrison Institute for Population Resource Studies at Stanford University.

The HapMap data, which includes approximately 4 million SNPs (http://www.ncbi.nlm.nih.gov/). The LSBL test evaluates whether genetic markers within a genomic region show unusual levels of differentiation with respect to the genome average. This test apportions the genetic variation observed in East Asian, European and West African populations for each SNP, and identifies markers with high levels of genetic differentiation in the East Asian sample. The lnRH test highlights genomic regions with low levels of genetic diversity in the population of interest in comparison with other population groups. This statistic was calculated for a two-way population comparison between East Asians and Europeans, and East Asians and West Africans, using an overlapping sliding window size of 100,000 base pairs (bp) and moving in 25,000-bp increments along a chromosome. Regions of the genome with negative Tajima’s D values (excess of rare alleles with respect to neutral expectations) are also a hallmark of positive selection. However, negative values of D can also result from seven geographic regions: Africa, the Middle East, Europe, South/Central Asia, East Asia, the Americas and Oceania. Additionally, we also analyzed these polymorphisms in 316 DNA samples of individuals of European, East Asian and South Asian ancestry living in Canada. We show that the VKORC1 rs9923231 polymorphism shows a unique allele frequency distribution, with the derived T allele showing very high allele frequencies in East Asian populations. Given this unusual distribution, we carried out three tests of selection in the HapMap East Asian sample, and report that the VKORC1 gene is an outlier for two of the tests, probably as a result of positive selection in East Asian populations.

**Genotyping**

Four SNPs known to influence the metabolism of warfarin (rs9923231, rs1799853, rs1057910 and rs2108622) were genotyped at KBiosciences (http://www.kbioscience.co.uk/) using a KASPar assay based on fluorescent genotyping and allele-specific PCR. To verify genotyping quality, 170 samples were genotyped as blind duplicates, and the concordance rate between the samples and duplicates was 100%.

**Statistical analysis**

Tests evaluating deviations from Hardy–Weinberg proportions and hierarchical analyses of molecular variance were carried out with the program Arlequin. Tests of allele frequency differences between pairs of populations were done with the program Arlequin, or alternatively, with the program Genepop (http://kimura.univ-montp2.fr/~rousset/Genepop.htm). Three different tests of selection were used to evaluate evidence of positive selection in the HapMap East Asian sample for the gene VKORC1: the locus-specific branch length (LSBL) test, the log of the ratio of heterozygosities (lnRH) test and Tajima’s D test. The results reported here are based on a genome-wide analysis of the HapMap data, which includes approximately 4 million SNPs (http://hapmap.ncbi.nlm.nih.gov/). The LSBL test evaluates whether genetic markers within a genomic region show unusual levels of differentiation with respect to the genome average. This test apportions the genetic variation observed in East Asian, European and West African populations for each SNP, and identifies markers with high levels of genetic differentiation in the East Asian sample. The lnRH test highlights genomic regions with low levels of genetic diversity in the population of interest in comparison with other population groups. This statistic was calculated for a two-way population comparison between East Asians and Europeans, and East Asians and West Africans, using an overlapping sliding window size of 100,000 base pairs (bp) and moving in 25,000-bp increments along a chromosome. Regions of the genome with negative Tajima’s D values (excess of rare alleles with respect to neutral expectations) are also a hallmark of positive selection. However, negative values of D can also result from seven geographic regions: Africa, the Middle East, Europe, South/Central Asia, East Asia, the Americas and Oceania. Additionally, we also analyzed these polymorphisms in 316 DNA samples of individuals of European, East Asian and South Asian ancestry living in Canada. We show that the VKORC1 rs9923231 polymorphism shows a unique allele frequency distribution, with the derived T allele showing very high allele frequencies in East Asian populations. Given this unusual distribution, we carried out three tests of selection in the HapMap East Asian sample, and report that the VKORC1 gene is an outlier for two of the tests, probably as a result of positive selection in East Asian populations.
from demographic events, such as population expansions. For this reason, it is important to compare local values of Tajima's $D$ with the empirical levels observed in the genome. As for the lnRH analysis, Tajima's $D$ was calculated for each population using an overlapping sliding window size of 100,000 bp with a 25,000-bp offset. The statistical significance for each of the LSBL, lnRH and Tajima's $D$ statistics was based on the genome-wide empirical distribution, using the formula $PE(x) = (\text{number of loci} > x)/\text{total number loci}$. The three statistics used in this analysis have been described in more detail in the paper by Bigham et al.\textsuperscript{55}

RESULTS

No significant deviations from the Hardy–Weinberg proportions were observed in any of the HGDP-CEPH samples for the CYP2C9 and CYP4F2 polymorphisms. For the VKORC1 rs9923231 SNP, the She from East Asia ($P=0.047$), Sardinians from Europe ($P=0.049$) and Papuan from Oceania ($P=0.014$) showed departures from Hardy–Weinberg proportions, but the $P$-values were not significant when multiple testing was considered. For the samples collected in Canada, only the individuals of European ancestry showed a nominally significant deviation from Hardy–Weinberg proportions ($P=0.001$).

Four polymorphisms were genotyped in the HGDP-CEPH DNA samples. The allele frequencies of each marker are shown in Figures 1–4 for all 52 populations. Table 1 reports the average allele frequency of each marker for each geographical region. Tables reporting the $P$-values for allele frequency comparisons between major geographic groups and between individual populations are provided as Supplementary information (Supplementary Tables S2–S9). The average allele frequencies are in general agreement with the information available for the HapMap samples (http://hapmap.ncbi.nlm.nih.gov/). The derived T allele of the VKORC1 rs9923231 SNP has been associated with lower required warfarin dosing and higher risks of bleeding in multiple studies. The derived T allele has very low frequencies in African populations (lower than 10%), with the notable exception of the San (33%). The frequencies of the T allele are

---

**Figure 1** Frequency of VKORC1 rs9923231T and C alleles in 52 Human Genome Diversity Project-Centre Etude Polymorphism Humain (HGDP-CEPH) populations.

**Figure 2** Frequency of CYP2C9 rs1799853T and C alleles in 52 Human Genome Diversity Project-Centre Etude Polymorphism Humain (HGDP-CEPH) populations.

**Figure 3** Frequency of CYP2C9 rs1057910 C and A alleles in 52 Human Genome Diversity Project-Centre Etude Polymorphism Humain (HGDP-CEPH) populations.
CYP2C9 and therefore increased risk of bleeding. The distribution of the CYP2C9*3 polymorphism. The CYP2C9*2 allele is primarily restricted to European (2–29%), Middle Eastern (11–20%) and Central/South Asia populations (2–16%), and it is mostly absent in other population groups. Exceptions are the North Eastern Bantu from Africa (4%), the Yakut from East Asia (2%) and the Maya (2%). Similarly, the highest frequencies of the CYP2C9*3 allele are also observed in Europe (4–21%), the Middle East (3–11%) and Central/South Asia (5–15%). The allele is not observed in Africa or most populations from the Americas, except the Pima (7%). In Oceania, the allele is not present in Melanesians, but in Papua New Guinea the frequency is 12%. Finally, in East Asia, the allele is absent in many populations, but reaches frequencies of 10% or higher in some populations, such as the Tu, Tujia and Xibo. Finally, the non-synonymous CYP4F2 rs2108622 polymorphism has been associated with warfarin dosing. Homozygotes for the C allele (coding for valine) require less warfarin than homozygotes for the T allele (coding for methionine), with heterozygotes requiring intermediate doses. The frequency of the minor T allele is on average low in Africa and the Americas (7 and 10%, respectively). In these geographic areas, the allele is absent in some populations (Southern Bantu, Mandenka and Biaka Pygmy in Africa, and Colombians in the Americas), and it is present at frequencies lower than 23% in other populations. In contrast, the T allele is present at higher frequencies in other regions: Europe (20–44%), Middle East (27–45%), Central/South Asia (26–40%), East Asia (6–57%), with the highest average frequencies found in Oceania (61%).

The same four polymorphisms were analyzed in a sample of Canadians of European, East Asian and South Asian ancestry (allele frequencies are reported in Table 2). In general, the frequencies in the Canadian sample are in agreement with those observed in the relevant HGDP-CEPH geographic groups. For VKORC1 rs9923231, the lowest frequencies of the derived T allele are observed in the Canadian South Asian sample (17% vs an average of 28% in the HGDP-CEPH Central/South Asian sample), followed by the European sample (38% vs an average of 51% in the European HGDP-CEPH sample). The highest frequencies for the T allele were observed in the East Asian sample (86 vs 90% in the HGDP-CEPH sample). For the CYP2C9*2 allele (rs1799833T allele), there is excellent agreement between the Canadian and HGDP-CEPH frequencies. The frequency of the T allele is less than 1% in the East Asian Canadian sample (same as in the HGDP-CEPH sample), 5% in the South Asian Canadian sample (vs 7% in the HGDP-CEPH Central/South Asian sample) and 17% in the European sample (vs 13% in the HGDP-CEPH sample). For the CYP2C9*3 allele (rs1057910 A allele), the highest frequency is observed in the Canadian South Asian sample (12 vs 10% in the HGDP-CEPH sample), followed by the European Canadian sample (4 vs 9% in the HGDP-CEPH sample) and the East Asian Canadian sample (2 vs 4% in the HGDP-CEPH sample). Finally, similar to what is observed in the HGDP-CEPH
samples, the highest frequencies for the CYP4F2 rs2108622 minor allele are found in the South Asian Canadian sample (48 vs 35%), and they are lower in the European and East Asian Canadian samples (24 vs 30%, and 22 vs 29%, respectively). Tables reporting the P-values for allele frequency comparisons between the Canadian samples and the relevant HGDP-CEPH populations are provided as Supplementary information (Supplementary Tables S10–S13).

A hierarchical analysis of molecular variance is presented in Table 3. This analysis provides information on the percentage of total variation due to genetic differences between the seven geographic groups (America, East Asia, Central/South Asia, Europe, the Middle East, Oceania and Africa), differences between populations within each of the geographic groups and genetic variation within populations. Calculations were made for all four polymorphisms, and we also report an average over all loci (Table 3). Overall, variation within populations accounted for the largest percentage of variation (80.41%), followed by variation due to differences among the major geographic groups (17.67%). The genetic differences among populations within groups explained a very small percentage of the total variation (1.92%). We observed differences in the patterns of genetic differentiation for the four markers. In particular, the variant rs9923231 within VKORC1 shows much higher genetic differentiation between geographic groups (31.63%) than the other markers. The percentage of variation among groups was around 7% for CYP2C9 rs1799853 and CYP4F2 rs2108622, and lower than 2% for the CYP2C9 rs1057910 polymorphism.

Given the unusual pattern of allele frequencies observed for the VKORC1 rs9923231-derived T allele, which shows very high allele frequencies in East Asian populations, three tests of positive selection were applied to the VKORC1 gene using genome-wide data available for the HapMap East Asian, European and West African samples (approximately four million SNPs). The results of these tests are illustrated in Table 4. The VKORC1 gene showed extreme distributions for each of these population genetic parameters with respect to the genome average. The LSBL test shows that this gene has many markers with extreme levels of differentiation in the East Asian sample. Similarly, the lnRH test indicates that the genetic variation in East Asia for VKORC1 is reduced when compared with that observed in the European sample. The VKORC1 gene is also an outlier for the Tajima’s D statistic that has negative values that depart from the average genomic distribution. The implications of these results are addressed in the discussion.

### DISCUSSION

This study examined the allele frequencies of four genetic polymorphisms known to influence warfarin dosing (VKORC1 rs9923231, CYP2C9 rs1799853, CYP2C9 rs1057910 and CYP4F2 rs2108622) in a total of 963 individuals from the HGDP-CEPH samples and 316 Canadians of diverse geographic ancestry. The ultimate goal was to improve our understanding of the worldwide allele frequency distribution of these polymorphisms and to discuss the results in light of the known effect of these markers on warfarin dosing and the differences in average warfarin dosing that have been reported.

### Table 2 Allele Frequency for four SNPs tested among Canadian samples

<table>
<thead>
<tr>
<th>Gene</th>
<th>Polymorphism</th>
<th>Geographic region</th>
<th>Sample size (n)</th>
<th>Allele frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>VKORC1</td>
<td>rs9923231 (C/T)</td>
<td>East Asia</td>
<td>100</td>
<td>C=0.143, T=0.857</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Central/South Asia</td>
<td>95</td>
<td>C=0.832, T=0.168</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Europe</td>
<td>121</td>
<td>C=0.621, T=0.378</td>
</tr>
<tr>
<td>CYP2C9</td>
<td>rs1799853 (C/T)</td>
<td>East Asia</td>
<td>100</td>
<td>C=0.995, T=0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Central/South Asia</td>
<td>95</td>
<td>C=0.951, T=0.049</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Europe</td>
<td>121</td>
<td>C=0.828, T=0.172</td>
</tr>
<tr>
<td>CYP2C9</td>
<td>rs1057910 (A/C)</td>
<td>East Asia</td>
<td>100</td>
<td>A=0.984, C=0.016</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Central/South Asia</td>
<td>95</td>
<td>A=0.883, C=0.117</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Europe</td>
<td>121</td>
<td>A=0.958, C=0.042</td>
</tr>
<tr>
<td>CYP4F2</td>
<td>rs2108622 (C/T)</td>
<td>East Asia</td>
<td>100</td>
<td>C=0.777, T=0.233</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Central/South Asia</td>
<td>95</td>
<td>C=0.516, T=0.484</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Europe</td>
<td>121</td>
<td>C=0.760, T=0.240</td>
</tr>
</tbody>
</table>

### Table 3 Hierarchical analysis of molecular variance for the four loci analyzed in the study

<table>
<thead>
<tr>
<th>Gene</th>
<th>Polymorphism</th>
<th>Percent variation among groups</th>
<th>Percent variation among populations</th>
<th>Percent variation within populations</th>
</tr>
</thead>
<tbody>
<tr>
<td>VKORC1</td>
<td>rs9923231 (C/T)</td>
<td>31.632</td>
<td>2.647</td>
<td>65.722</td>
</tr>
<tr>
<td>CYP2C9</td>
<td>rs1799853 (C/T)</td>
<td>7.170</td>
<td>1.153</td>
<td>91.677</td>
</tr>
<tr>
<td>CYP2C9</td>
<td>rs1057910 (A/C)</td>
<td>1.982</td>
<td>0.274</td>
<td>97.744</td>
</tr>
<tr>
<td>CYP4F2</td>
<td>rs2108622 (C/T)</td>
<td>7.475</td>
<td>1.666</td>
<td>90.859</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average over four loci</td>
<td>17.67</td>
<td>1.920</td>
</tr>
</tbody>
</table>

### Table 4 Results of three tests of positive selection in the East Asian HapMap sample

<table>
<thead>
<tr>
<th>LSLB</th>
<th>lnRH</th>
<th>Tajima’s D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of significant markers (P &lt; 0.001, P &lt; 0.01, P &lt; 0.05)</td>
<td>Number of significant windows (P &lt; 0.001, P &lt; 0.01, P &lt; 0.05)</td>
</tr>
<tr>
<td></td>
<td>(1, 3, 4)</td>
<td>EAS/EUR (0, 0, 4)</td>
</tr>
<tr>
<td></td>
<td>EAS/WAF (0, 0, 0)</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: lnRH, log of the ratio of heterozygosities; LSLB, locus-specific branch length.
in the literature. For example, Dang et al. reported mean weekly warfarin doses of 24 mg for Asian Americans, 31 mg for Hispanics, 36 mg for whites and 43 mg for African Americans, and a recent study by the International Warfarin Pharmacogenetics Consortium reported average stable therapeutic warfarin doses of 21 mg per week in Asians, 31.5 mg per week in whites and 40 mg per week in blacks. Most of the previous studies on the polymorphisms analyzed in our study have focused on European and East Asian populations, and data for other geographical regions such as South/Central Asia, Africa, the Middle East, Oceania and the Americas are quite limited. Unnecessary bleeding complications during initial warfarin dosing and knowledge that genetics influences an individual’s reaction to warfarin highlight the need for more extensive studies in these areas.

The VKORC1 gene is believed to be the most important individual predictor of warfarin dose. Previous research has shown that the derived T allele of the rs9923231 polymorphism, which is located in the promoter region of the gene, is associated with lower warfarin dose requirements. Moreover, there is evidence indicating that the rs9923231 T allele is associated with a reduction of VKORC1 mRNA levels of up to 70%, in comparison to the wild-type allele. This variation in mRNA levels may therefore be causing some of the variability observed in warfarin dosing among individuals. The derived T allele is present at relatively high frequencies in most populations, and data for other geographical regions such as South/Central Asia, Africa, the Middle East, Oceania and the Americas are quite limited. Unnecessary bleeding complications during initial warfarin dosing and knowledge that genetics influences an individual’s reaction to warfarin highlight the need for more extensive studies in these areas.

The derived T allele is present at relatively high frequencies in most populations, and data for other geographical regions such as South/Central Asia, Africa, the Middle East, Oceania and the Americas are quite limited. Unnecessary bleeding complications during initial warfarin dosing and knowledge that genetics influences an individual’s reaction to warfarin highlight the need for more extensive studies in these areas.

The derived T allele is present at relatively high frequencies in most populations, and data for other geographical regions such as South/Central Asia, Africa, the Middle East, Oceania and the Americas are quite limited. Unnecessary bleeding complications during initial warfarin dosing and knowledge that genetics influences an individual’s reaction to warfarin highlight the need for more extensive studies in these areas.

The derived T allele is present at relatively high frequencies in most populations, and data for other geographical regions such as South/Central Asia, Africa, the Middle East, Oceania and the Americas are quite limited. Unnecessary bleeding complications during initial warfarin dosing and knowledge that genetics influences an individual’s reaction to warfarin highlight the need for more extensive studies in these areas.

The derived T allele is present at relatively high frequencies in most populations, and data for other geographical regions such as South/Central Asia, Africa, the Middle East, Oceania and the Americas are quite limited. Unnecessary bleeding complications during initial warfarin dosing and knowledge that genetics influences an individual’s reaction to warfarin highlight the need for more extensive studies in these areas.

The derived T allele is present at relatively high frequencies in most populations, and data for other geographical regions such as South/Central Asia, Africa, the Middle East, Oceania and the Americas are quite limited. Unnecessary bleeding complications during initial warfarin dosing and knowledge that genetics influences an individual’s reaction to warfarin highlight the need for more extensive studies in these areas.
research confirms the effect of rs2108622 on warfarin dose, it has the potential to be useful in future dosing algorithms.

As described above, the VKORC1 rs9923231 SNP has a very unusual allele frequency distribution. This polymorphism shows very high levels of geographic differentiation. In particular, the derived T allele is almost absent in Africa and has extremely high frequencies in East Asian populations. We applied three tests of positive selection to the East Asian HapMap sample to explore whether this unusual pattern of polymorphism could be the result of positive selection in East Asia. The three tests of selection are based on different characteristics of the data: genetic differentiation (LSBL test), genetic diversity (ln\(D\) test) and allele frequency distribution (Tajima’s \(D\)). Interestingly, we observed significant results for the VKORC1 gene in the three tests (Table 4). These tests are based on a comparison of the values of the aforementioned statistics for the VKORC1 gene with the genome-wide empirical distribution. These analyses show that VKORC1 has an extreme genetic differentiation in the East Asian sample, with eight markers in the top 5% of the empirical distribution. Furthermore, when comparing the genetic diversity (heterozygosity) of the East Asian and the European HapMap samples, the VKORC1 gene shows a relative reduction of genetic diversity in East Asia with respect to what is observed in the rest of the genome. Finally, the analysis of the Tajima’s \(D\) statistic shows an excess of rare alleles (negative Tajima’s \(D\)) in the VKORC1 gene with respect to the genome-wide average. It is important to note that the average Tajima’s \(D\) value in the HapMap East Asian sample is positive (average value for the empirical data=1.74) because the HapMap data set is biased toward common polymorphisms. It is well known that negative Tajima’s \(D\) values may be the result of positive selection or demographic processes such as population expansions. However, demographic processes such as population expansions would be expected to affect patterns of variation at all loci in the genome, whereas natural selection will act only on specific loci. Therefore, conditioning on the genome-wide distribution helps to differentiate between these two possibilities. Interestingly, Carlson et al.\(^9\) also identified a genomic region including the VKORC1 gene as a potential target of selection in East Asians using a sliding window analysis of Tajima’s \(D\) based on a different data set (Perlegen, Mountain View, CA, USA). Although we cannot entirely eliminate the possibility that our results are due to stochastic processes, the extreme distributions observed for three different parameters point to positive selection as a potential explanation. It would be extremely interesting to explore which potential selective factors could have driven the T allele to such high frequencies in East Asian populations. The protein encoded by VKORC1 has a critical role in the vitamin K cycle and the activation of vitamin K-dependent proteins. In addition to the well-known effect of these proteins in hemostasis, recent evidence indicates that vitamin K-dependent proteins are present in numerous tissues and may be important in bone mineralization, regulation of calcification and other biological processes.\(^{40,42}\)

In conclusion, we report the allele frequency distribution of four SNPs known to affect warfarin metabolism in a large sample representing individuals from around the world. To date, most of the studies have been carried out in European and East Asian populations. Therefore, it is important to provide information on samples from other geographic regions to better understand the potential impact of these polymorphisms at the population level. Our data also have interesting implications in terms of the application of warfarin-dosing algorithms in different populations. Differences in allele frequency distribution between groups can have an impact on the amount of variance in required warfarin dosing explained by the algorithms. Given the relative lack of polymorphism of the VKORC1 rs9923231 and the CYP2C9*2 and *3 SNPs in African populations, it is not surprising that application of algorithms including these markers in cohorts of individuals of African ancestry explain less variation in warfarin dose than in cohorts of individuals of European ancestry, as it has been reported in recent studies.\(^{18,30}\) Unfortunately, the VKORC1, CYP2C9 and CYP4F2 genes remain poorly characterized in many population groups, and there may be other variants within these genes affecting warfarin dose that would be relevant to include in the pharmacogenetic-based dosing algorithms. For example, Scott et al.\(^31\) recently described a variant (CYP2C9*8) present in an African-American male with a lower than predicted warfarin dose (based on prior genotyping of CYP2C9 and VKORC1 alleles). On further characterization of this variant in an African-American sample, the authors reported that CYP2C9*8 is the most frequent variant in this sample, and predicted that incorporation of this polymorphism in pharmacogenetic-based dosing algorithms could potentially reclassify the predicted metabolic phenotypes of almost 10% of African Americans. The main goal of our study was to provide worldwide allele frequency information for the most commonly studied variants influencing warfarin dosing (CYP2C9 rs1799853 and rs1057910, and CYP4F2 rs2108622). However, it is important to note that, in addition to the variants analyzed in this study, recent pharmacogenetic-based dosing algorithms now include other polymorphisms such as CYP2C9*5 and *6. Future studies will be required to characterize the frequency of these and other polymorphisms (for example, CYP2C9*8) in worldwide populations. Ideally, existing algorithms should be validated in different populations and efforts directed to identify new functional variants that could be included in pharmacogenetic-based dosing algorithms. The dramatic advances in high-density microarray and next generation sequencing technologies and the consequent reduction in costs will make it possible to obtain a much better understanding of the genetic factors influencing warfarin dosing in different populations and the development of pharmacogenetic-based algorithms that will reduce adverse reactions to warfarin.

---

Polymorphisms associated with warfarin dose

KA Ross et al

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)