

ORIGINAL ARTICLE

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

Kendra A Ross¹, Abigail W Bigham², Melissa Edwards¹, Agnes Gozdzik¹, Guilherme Suarez-Kurtz³ and Esteban J Parra¹

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 Polymorphisme Humain (HGDP-CEPH) worldwide sample ($N=963$), as well as in a sample of individuals of European, East Asian and South Asian ancestry living in Canada ($N=316$). 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 2004,¹ 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.² In addition, US death certificates reviewed by Wysowski *et al.*¹ 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.³

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.^{4,5} With a narrow therapeutic window and large interindividual variability,

clinical factors, such as age, height, sex, diet and drug–drug interactions^{6–8} explain 12⁶ to 15%⁹ of the observed variability in warfarin dosing. Genetic factors have also been shown to affect warfarin maintenance.^{3,10} 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.^{6,9,11}

Warfarin exists as a mixture of two stereoisomers, R and S.^{8,12} S-warfarin is approximately three times as active as the R-enantiomer¹² 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.¹³ Within the *VKORC1* gene, the single nucleotide polymorphism (SNP) rs9923231 has been extensively studied among European and Asian populations. Original research by Wadelius¹⁰ 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.⁵ Specifically two variants of this highly polymorphic gene, *CYP2C9**2 (rs1799853) and *CYP2C9**3 (rs1057910) have been shown

¹Department of Anthropology, University of Toronto, Toronto, Ontario, Canada; ²Department of Pediatrics, University of Washington, Seattle, WA, USA; ³Department of Pharmacology, Instituto Nacional de Câncer, Rio de Janeiro, Brazil

Correspondence: Professor EJ Parra, University of Toronto Mississauga, Department of Anthropology, 3359 Mississauga Road, Room 2002 South Building., Mississauga, Ontario, Canada L5L 1C6.

E-mail: esteban.parra@utoronto.ca

Received 16 March 2010; revised 25 May 2010; accepted 26 May 2010; published online 17 June 2010

to reduce warfarin metabolism and lower the mean warfarin dose requirements,^{14–16} thereby significantly increasing an individual's risk of having complications related to bleeding.¹⁷ A review by Schelleman *et al.*¹⁸ 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 12¹⁰ to 18%⁶ of the variation in warfarin dose. Recent research has identified a third gene, *CYP4F2* as influencing an individual's sensitivity to warfarin.^{9,19,20} *CYP4F2* codes for a vitamin K₁ oxidase that causes a decrease in the metabolism of vitamin K₁.²¹ 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.²² Studies on European populations have indicated that *CYP4F2* rs2108622 may account for an additional 1–7% of observed variation in warfarin dosing.^{9,22,23} 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.²⁴

The traditional approach for warfarin dosing is based on an initial fixed dose, followed by stabilization through trial and error.²⁵ 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.^{25,26} 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.^{3,26} 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.^{7,10,19,25,27} 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.²⁸ 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.¹⁸ 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 total variance in warfarin dose.²⁹ This compares well with the power of other dosing algorithms developed for European and Asian populations, and works equally well among white and black Brazilians.²⁹ The development of a universally useful algorithm will, however, rely on further research into the frequency of *VKORC1*, *CYP2C9* and *CYP4F2* variants in a greater number of populations and characterization of other genetic variants associated with warfarin dosing.

The majority of warfarin association studies have focused on individuals of European, Asian and, to a lesser extent, African ancestry. Improvement in the safety of warfarin therapy requires a comprehensive description of the occurrence of the key alleles known to influence the metabolism and dosing of warfarin. In this study, we present a more complete picture of the worldwide frequency distribution of four genetic polymorphisms known to influence warfarin dosing: *VKORC1* (rs9923231), *CYP2C9* (rs1799853 and rs1057910) and *CYP4F2* (rs2108622). We characterized these SNPs in DNA samples from the Centre Etude Polymorphisme Humain (CEPH) Human Genome Diversity Project (HGDP), which includes 963 individuals

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 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.³¹ 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 (ln*RH*) test and Tajima's *D* test.^{32,33,34} 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 ln*RH* 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 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 $\ln RH$ 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 $LSBL$, $\ln RH$ 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.*³⁵

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

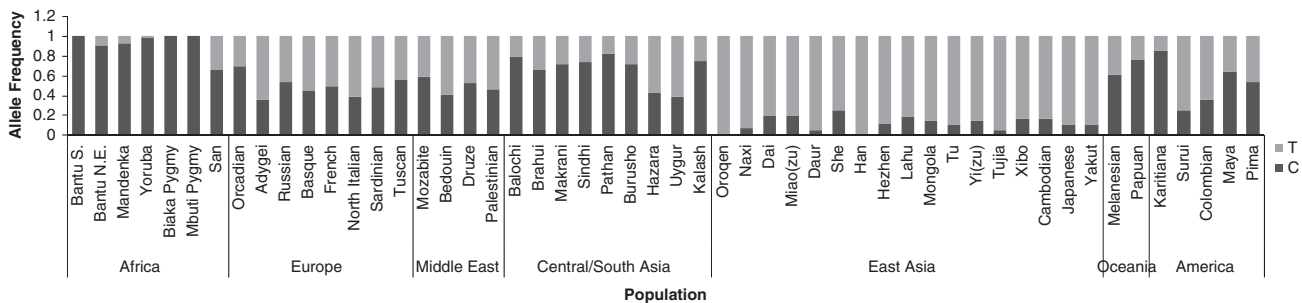


Figure 1 Frequency of *VKORC1* rs9923231T and C alleles in 52 Human Genome Diversity Project-Centre Etude Polymorphisme Humain (HGDP-CEPH) populations.

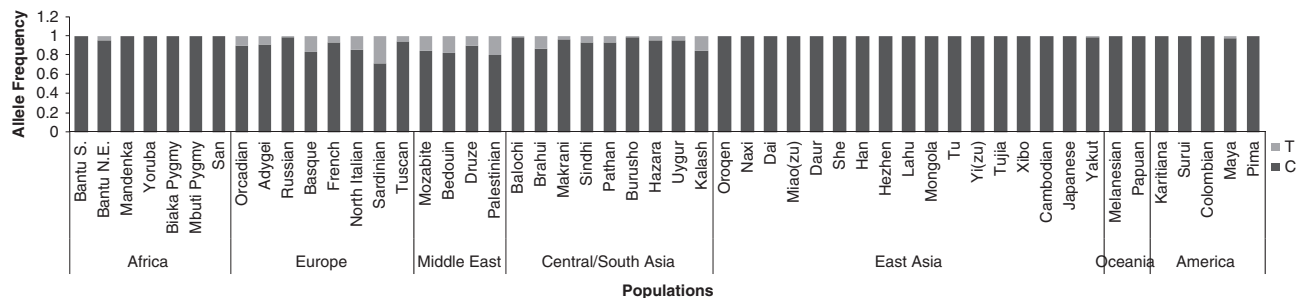


Figure 2 Frequency of *CYP2C9* rs1799853T and C alleles in 52 Human Genome Diversity Project-Centre Etude Polymorphisme Humain (HGDP-CEPH) populations.

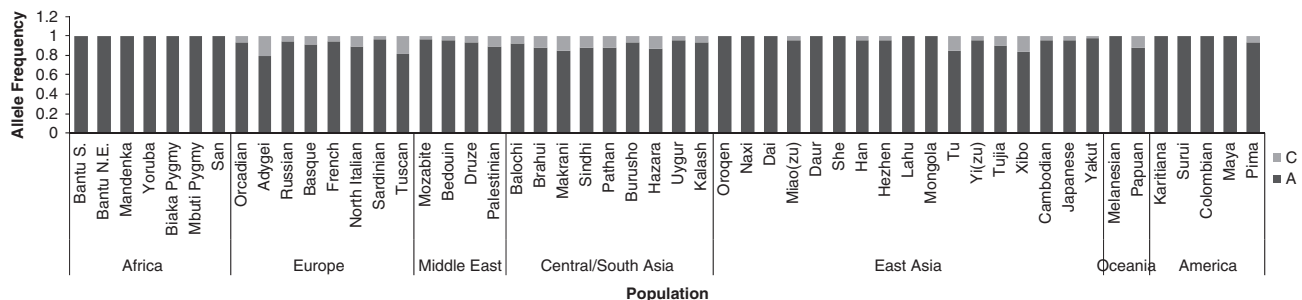


Figure 3 Frequency of *CYP2C9* rs1057910 C and A alleles in 52 Human Genome Diversity Project-Centre Etude Polymorphisme Humain (HGDP-CEPH) populations.

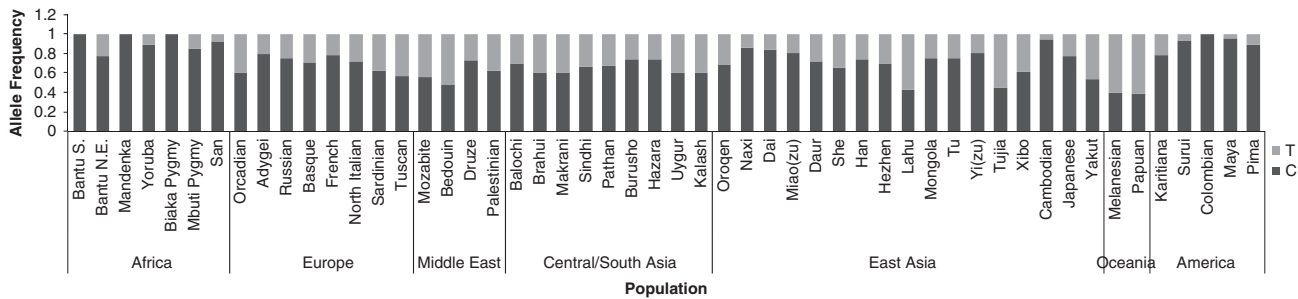


Figure 4 Frequency of *CYP4F2* rs2108622T and C alleles in 52 Human Genome Diversity Project-Centre Etude Polymorphisme Humain (HGDP-CEPH) populations.

Table 1 Average allele frequencies for four SNPs tested among CEPH samples^a

Gene	Polymorphism	Geographic Region	Sample size (n)	Allele frequencies
<i>VKORC1</i>	rs9923231 (C/T)	America	64	C=0.586, T=0.414
		East Asia	233	C=0.102, T=0.898
		Central/South Asia	204	C=0.724, T=0.276
		Europe	159	C=0.494, T=0.506
		Middle East	164	C=0.488, T=0.512
		Oceania	31	C=0.694, T=0.306
		Africa	108	C=0.953, T=0.047
<i>CYP2C9</i>	rs1799853 (C/T)	America	64	C=0.992, T=0.008
		East Asia	233	C=0.998, T=0.002
		Central/South Asia	204	C=0.929, T=0.071
		Europe	159	C=0.873, T=0.127
		Middle East	164	C=0.840, T=0.160
		Oceania	31	C=1.000, T=0.000
		Africa	108	C=0.995, T=0.005
<i>CYP2C9</i>	rs1057910 (A/C)	America	64	A=0.984, C=0.016
		East Asia	233	A=0.957, C=0.043
		Central/South Asia	204	A=0.899, C=0.101
		Europe	159	A=0.914, C=0.086
		Middle East	164	A=0.933, C=0.067
		Oceania	31	A=0.953, C=0.046
		Africa	108	A=1.000, C=0.000
<i>CYP4F2</i>	rs2108622 (C/T)	America	64	C=0.902, T=0.098
		East Asia	233	C=0.708, T=0.292
		Central/South Asia	204	C=0.650, T=0.350
		Europe	159	C=0.703, T=0.297
		Middle East	164	C=0.597, T=0.403
		Oceania	31	C=0.387, T=0.613
		Africa	108	C=0.930, T=0.070

Abbreviations: CEPH, Centre Etude Polymorphisme Humain; SNP, single-nucleotide polymorphism.

^aFor a graphical representation of the individual population frequencies, see Figures 1–4.

intermediate in Europe (30–65%), the Middle East (41–51%), Central/South Asia (17–61%), Oceania (23–39%) and the Americas (14–75%). Finally, the frequencies of the derived allele are very high in East Asian populations, with frequencies ranging from 75 (She) to 100% (Han and Oroqen). The *CYP2C9**2 (rs1799853T) and particularly the *CYP2C9**3 (rs1057910 C) alleles lead to reduced warfarin metabolism, and therefore increased risk of bleeding. The distribution of the *CYP2C9* alleles is very different from that observed for the *VKORC1*

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 Pigmy 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 (rs1799853T 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

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

Abbreviations: SNP, single-nucleotide polymorphism.

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

Gene	Polymorphism	Percent variation among groups	Percent variation among populations	Percent variation within populations
<i>VKORC1</i>	rs9923231 (C/T)	31.632	2.647	65.722
<i>CYP2C9</i>	rs1799853 (C/T)	7.170	1.153	91.677
<i>CYP2C9</i>	rs1057910 (A/C)	1.982	0.274	97.744
<i>CYP4F2</i>	rs2108622 (C/T)	7.475	1.666	90.859
		Average over four loci		
		17.67	1.920	80.410

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

LSBL	<i>lnRH</i>	Tajima's D
Number of significant markers ($P < 0.001$, $P < 0.01$, $P < 0.05$)	Number of significant windows ($P < 0.001$, $P < 0.01$, $P < 0.05$)	Number of significant windows ($P < 0.001$, $P < 0.01$, $P < 0.05$)
(1, 3, 4)	EAS/EUR (0, 0, 4) EAS/WAF (0, 0, 0)	(0, 0, 5)

Abbreviations: *lnRH*, log of the ratio of heterozygosities; LSBL, locus-specific branch length.

in the literature.^{3,26,36,37} For example, Dang *et al.*²⁶ 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 Consortium³⁶ 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.^{29,38} 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.¹³ This variation in mRNA levels may therefore be causing some of the variability observed in warfarin dosing among individuals. The significance of this polymorphism is further evidenced by studies showing that it may be accountable for all of the variability in warfarin metabolism that is due to the *VKORC1* haplotype.³⁰ Among the HGDP-CEPH samples, the derived T allele at the *VKORC1* polymorphism was present in most world populations, except the Southern Bantu and the Pygmies from Africa. The low frequency of the derived T allele observed among African populations (<10% with the exception of the San who had a frequency of 33%) is consistent with previous research indicating that individuals of African ancestry require a higher average warfarin dose than those of European or Asian descent.^{3,26} In contrast, the East Asian populations included in the HGDP-CEPH panel had very high frequencies for the derived T allele, ranging from 75 (among the She) to 100% (Han and Oroqen). Canadians of East Asian ancestry showed a similar frequency for the T allele (86%). These findings are also in agreement with studies indicating that, on average, warfarin dose requirements among East Asian individuals are lower than in African and European subjects.^{3,37} The average frequency of the derived T allele in European populations was 51% in the HGDP-CEPH samples and 38% in the sample of Canadians of European ancestry. These values are in the middle of those observed among the African and East Asian populations and also reflect current population-based trends in warfarin dosing. Therefore, a correlation is observed between average warfarin dose requirements and the observed frequency of rs9923231 alleles in African, European and East Asian populations. With respect to other population groups, the frequencies of the derived T allele in the Middle East are similar to those observed in Europe. The average frequency in the HGDP-CEPH Central/South Asian populations is lower than in Europe, and we also observed that the frequency of the derived T allele in the Canadian sample of South Asian ancestry (17%) is lower than the frequency in the sample of European ancestry (38%). Finally, it is important to mention that the frequency of the derived T allele in Native American populations shows considerable variation (14–75%). Perini *et al.*²⁹ recently analyzed two Amerindian populations, the Guarani and Kaingang from Brazil, and they estimated that 60% of Guarani and 40% of Kaingang individuals have the variant T allele.²⁹ Together with our findings, this suggests that a high percentage of individuals of Native American ancestry may

require reduced warfarin doses.²⁹ Our hierarchical analysis of molecular variance indicates that the *VKORC1* rs9923231 SNP shows an extremely high differentiation between geographic groups, with more than 31% of the total variation due to differences between the seven groups included in the analysis. This geographic differentiation is much more pronounced than that observed for the other three SNPs analyzed in this study (Table 3).

The *CYP2C9* gene was the first gene recognized as influential in the metabolism of warfarin. Two non-synonymous polymorphisms within this gene, *CYP2C9**2 (430C>T causing the substitution R144C) and *CYP2C9**3 (1075A>C leading to an I359L substitution) have been the focus of most studies. Past research has indicated that *CYP2C9**3 (derived C allele) is more influential on warfarin dose requirements than the *CYP2C9**2 (derived T allele) variant.^{15,16,17} The results from genotyping the HGDP-CEPH and Canadian samples are in accordance with literature on the allele frequency of the *CYP2C9**2 and *3 variants among European and East Asian populations. The *CYP2C9**2 allele has been most extensively studied among European populations in which it appears at a frequency of approximately 10%,³⁷ which is comparable to our findings of frequencies between 13% and 17% in the HGDP-CEPH and Canadian European samples, respectively. Previous research has indicated that the *CYP2C9**2 variant is extremely rare if not absent in East Asian populations.³⁷ With the exception of the Yakut (2%), we observed that the derived T allele was absent in all East Asian populations from the HGDP-CEPH sample and appeared at a frequency of 0.5% among the Canadian East Asians. With regards to the *CYP2C9**3 allele in these regions, the frequency of the derived C allele was higher in the European populations (average HGDP-CEPH = 9% and Canadian samples = 4%) than in East Asians (CEPH = 4% and Canadian samples = 1%). These figures are in general agreement with those reported by Kim *et al.*,³⁷ who found that the frequency of the *CYP2C9**3 variant was 8% in Europeans and 2% in East Asians. With respect to the distribution of these polymorphisms in other population groups, it is interesting to note that, whereas the frequency of the *CYP2C9**2 allele is lower in Central/South Asian populations than in Europe (average HGDP-CEPH in Central/South Asia = 7%, and Canadian South Asian sample = 5%), the opposite is true for the *CYP2C9**3 allele (average HGDP-CEPH in Central/South Asia = 10%, and Canadian South Asian sample = 12%). In African, Melanesian and Native American populations *CYP2C9**2 and *CYP2C9**3 tend to be absent or at very low frequencies, with the exception of *CYP2C9**3 in Papuans (12%).

The most recent polymorphism reported to influence warfarin dosing has been the non-synonymous rs2108622 SNP (C/T polymorphism) located within the *CYP4F2* gene. Caldwell *et al.*¹⁹ studied three independent white US cohorts and found a 4–12% increase in warfarin dose per T allele. On the basis of the frequency distributions of the derived T allele (30% among Europeans and Asians vs approximately 7% in Africans), these authors suggested that the expected contribution of this SNP to stable warfarin dose in Africans will be lower than in Europeans and Asians. After correcting for the effects of *VKORC1* and *CYP2C9*, a recent genome-wide association study on Swedish patients has confirmed the association of rs2108622 with warfarin dose.⁹ However, another study by Zhang *et al.*²⁴ in the United Kingdom concluded that there was no clear association between this polymorphism and stable warfarin dose. Our study indicates that the derived T allele is present at relatively high frequencies in most worldwide populations. The main exceptions are African and Native American populations, in which the derived T allele has relatively low frequencies or is absent. The highest frequencies observed for the T allele are found in Melanesian populations (>60%). If further

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 (lnRH 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.*³⁹ 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.*⁴³ 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.

- 1 Wysowski, D. K., Nourjah, P. & Swartz, L. Bleeding complications with warfarin use: a prevalent adverse effect resulting in regulatory action. *Arch. Intern. Med.* **167**, 1414–1419 (2007).
- 2 Budnitz, D. S., Shehab, N., Kegler, S. R. & Richards, C. L. Medication use leading to emergency department visits for adverse drug events in older adults. *Ann. Intern. Med.* **147**, 755–765 (2007).
- 3 Takahashi, H., Wilkinson, G. R., Nutescu, E. A., Morita, T., Ritchie, M. D., Scordo, M. G. *et al.* Different contributions of polymorphisms in *VKORC1* and *CYP2C9* to intra- and inter-population differences in maintenance dose of warfarin in Japanese, Caucasians and African-Americans. *Pharmacogenet. Genomics* **16**, 101–110 (2006).
- 4 Ansell, J., Hirsh, J., Poller, L., Bussey, H., Jacobson, A. & Hylek, E. The pharmacology and management of the vitamin K antagonists: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. *Chest* **126**, 204S–233S (2004).
- 5 Kim, M. J., Huang, S. M., Meyer, U. A., Rahman, A. & Lesko, L. J. A regulatory science perspective on warfarin therapy: A pharmacogenetic opportunity. *J. Clin. Pharmacol.* **49**, 138–146 (2009).
- 6 Carlquist, J. F., Horne, B. D., Muhlestein, J. B., Lappe, D. L., Whiting, B. M., Kolek, M. J. *et al.* Genotypes of the cytochrome p450 isoform, *CYP2C9*, and the vitamin K epoxide reductase complex subunit 1 conjointly determine stable warfarin dose: a prospective study. *J. Thromb. Thrombolysis* **22**, 191–197 (2006).
- 7 Sconce, E. A., Khan, T. I., Wynne, H. A., Avery, P., Monkhouse, L., King, B. P. *et al.* The impact of *CYP2C9* and *VKORC1* genetic polymorphism and patient characteristics upon warfarin dose requirements: proposal for a new dosing regimen. *Blood* **106**, 2329–2333 (2005).
- 8 Wells, P. S., Holbrook, A. M., Crowther, R. & Hirsh, J. Interactions of warfarin with drugs and food. *Ann. Intern. Med.* **121**, 676–683 (1994).
- 9 Takeuchi, F., McGinnis, R., Bourgeois, S., Barnes, C., Eriksson, N., Soranzo, N. *et al.* A genome-wide association study confirms *VKORC1*, *CYP2C9*, and *CYP4F2* as principal genetic determinants of warfarin dose. *PLoS Genet.* **5**, e1000433 (2009).
- 10 Wadelius, M., Chen, L. Y., Lindh, J. D., Eriksson, N., Ghori, M. J., Bumpstead, S. *et al.* The largest prospective warfarin-treated cohort supports genetic forecasting. *Blood* **113**, 784–792 (2009).

- 11 Wadelius, M., Chen, L. Y., Downes, K., Ghorji, J., Hunt, S., Eriksson, N. *et al*. Common VKORC1 and GGCX polymorphisms associated with warfarin dose. *Pharmacogenomics J.* **5**, 262–270 (2005).
- 12 Choonara, I. A., Cholerton, S., Haynes, B. P., Breckenridge, A. M. & Park, B. K. Stereoselective interaction between the R enantiomer of warfarin and cimetidine. *Br. J. Clin. Pharmacol.* **21**, 271–277 (1986).
- 13 Oldenburg, J., Watzka, M., Rost, S. & Muller, C. R. 2007. VKORC1: molecular target of coumarins. *J. Thromb. Haemost.* **5**(Suppl 1), 1–6 (2007).
- 14 Aithal, G. P., Day, C. P., Kesteven, P. J. & Daly, A. K. Association of polymorphisms in the cytochrome P450 CYP2C9 with warfarin dose requirement and risk of bleeding complications. *Lancet* **353**, 717–719 (1999).
- 15 D'Andrea, G., D'Ambrosio, R. L., Di Perna, P., Chetta, M., Santacrose, R., Brancaccio, V. *et al*. A polymorphism in the VKORC1 gene is associated with an interindividual variability in the dose-anticoagulant effect of warfarin. *Blood* **105**, 645–649 (2005).
- 16 Sanderson, S., Emery, J. & Higgins, J. CYP2C9 gene variants, drug dose, and bleeding risk in warfarin-treated patients: a HuGenet systematic review and meta-analysis. *Genet. Med.* **7**, 97–104 (2005).
- 17 Higashi, M. K., Veenstra, D. L., Kondo, L. M., Wittkowsky, A. K., Srinouanprachanh, S. L., Farin, F. M. *et al*. Association between CYP2C9 genetic variants and anticoagulation-related outcomes during warfarin therapy. *JAMA*. **287**, 1690–1698 (2002).
- 18 Schelleman, H., Limdi, N. A. & Kimmel, S. E. Ethnic differences in warfarin maintenance dose requirement and its relationship with genetics. *Pharmacogenomics* **9**, 1331–1346 (2008).
- 19 Caldwell, M. D., Awad, T., Johnson, J. A., Gage, B. F., Falkowski, M., Gardina, P. *et al*. CYP4F2 genetic variant alters required warfarin dose. *Blood* **111**, 4106–4112 (2008).
- 20 Perez-Andreu, V., Roldan, V., Anton, A. I., Garcia-Barbera, N., Corral, J., Vicente, V. *et al*. Pharmacogenetic relevance of CYP4F2 V433M polymorphism on acenocoumarol therapy. *Blood* **113**, 4977–4979 (2009).
- 21 McDonald, M. G., Rieder, M. J., Nakano, M., Hsia, C. K. & Rettie, A. E. CYP4F2 is a vitamin K1 oxidase: An explanation for altered warfarin dose in carriers of the V433M variant. *Mol. Pharmacol.* **75**, 1337–1346 (2009).
- 22 Borgiani, P., Ciccacci, C., Forte, V., Sirianni, E., Novelli, L., Bramanti, P. *et al*. CYP4F2 genetic variant (rs2108622) significantly contributes to warfarin dosing variability in the Italian population. *Pharmacogenomics* **10**, 261–266 (2009).
- 23 Pautas, E., Moreau, C., Gouin-Thibault, I., Gilmard, J. L., Mahe, I., Legendre, C. *et al*. Genetic factors (VKORC1, CYP2C9, EPHX1, and CYP4F2) are predictor variables for warfarin response in very elderly, frail inpatients. *Clin. Pharmacol. Ther.* **87**, 57–64 (2010).
- 24 Zhang, J. E., Jorgensen, A. L., Alfirevic, A., Williamson, P. R., Toh, C. H., Park, B. K. *et al*. Effects of CYP4F2 genetic polymorphisms and haplotypes on clinical outcomes in patients initiated on warfarin therapy. *Pharmacogenet. Genomics* **19**, 781–789 (2009).
- 25 Yin, T. & Miyata, T. Warfarin dose and the pharmacogenomics of CYP2C9 and VKORC1—rationale and perspectives. *Thromb. Res.* **120**, 1–10 (2007).
- 26 Dang, M. T., Hambleton, J. & Kayser, S. R. The influence of ethnicity on warfarin dosage requirement. *Ann. Pharmacother.* **39**, 1008–1012 (2005).
- 27 Miao, L., Yang, J. & Shen, Z. Contribution of age, body weight, and CYP2C9 and VKORC1 genotype to the anticoagulant response to warfarin: proposal for a new dosing regimen in Chinese patients. *Eur. J. Clin. Pharmacol.* **63**, 1135–1141 (2007).
- 28 Cavallari, L. H., Langaee, T. Y., Momary, K. M., Sharp, N. L., Nutescu, E. A., Coty, W. A. *et al*. Genetic and clinical predictors of warfarin dose requirements in African Americans. *Clin. Pharmacol. Ther.* **87**, 459–464 (2010).
- 29 Perini, J. A., Struchiner, C. J., Silva-Assuncao, E., Santana, I. S. C., Rangel, F., Ojopi, E. B. *et al*. Pharmacogenetics of warfarin: development of a dosing algorithm for Brazilian patients. *Clin. Pharmacol. Ther.* **84**, 722–728 (2008).
- 30 Langley, M. R., Booker, J. K., Evans, J. P., McLeod, H. L. & Weck, K. E. Validation of clinical testing for warfarin sensitivity: comparison of CYP2C9–VKORC1 genotyping assays and warfarin-dosing algorithms. *J. Mol. Diagn.* **11**, 216–225 (2009).
- 31 Excoffier, L., Laval, G. & Schneider, S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* **1**, 47–50 (2005).
- 32 Shriver, M. D., Kennedy, G. C., Parra, E. J., Lawson, H. A., Sonpar, V., Huang, J. *et al*. The genomic distribution of population substructure in four populations using 8525 autosomal SNPs. *Hum. Genomics* **1**, 274–286 (2004).
- 33 Storz, J. F., Payseur, B. A. & Nachman, M. W. Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. *Mol. Biol. Evol.* **21**, 1800–1811 (2004).
- 34 Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
- 35 Bigham, A. W., Mao, X., Mei, R., Brutsaert, T., Wilson, M. J., Julian, C. G. *et al*. Identifying positive selection candidate loci for high-altitude adaptation in Andean populations. *Hum. Genomics* **4**, 79–90 (2009).
- 36 Limdi, N. A., Wadelius, M., Cavallari, L., Eriksson, N., Crawford, D. C., Lee, M. T. *et al*. Warfarin pharmacogenetics: a single VKORC1 polymorphism is predictive of dose across 3 racial groups. *Blood* **115**, 3827–3834 (2010).
- 37 Kim, K., Johnson, J. A. & Derendorf, H. Differences in drug pharmacokinetics between East Asians and Caucasians and the role of genetic polymorphisms. *J. Clin. Pharmacol.* **44**, 1083–1105 (2004).
- 38 Aquilante, C. L., Langaee, T. Y., Lopez, L. M., Yarandi, H. N., Tromberg, J. S., Mohuczy, D. *et al*. Influence of coagulation factor, vitamin K epoxide reductase complex subunit 1, and cytochrome P450 2C9 gene polymorphisms on warfarin dose requirements. *Clin. Pharmacol. Ther.* **79**, 291–302 (2006).
- 39 Carlson, C. S., Thomas, D. J., Eberle, M. A., Swanson, J. E., Livingston, R. J., Rieder, M. J. *et al*. Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Res.* **15**, 1553–1565 (2005).
- 40 Berkner, K. L. & Runge, K. W. The physiology of vitamin K nutriture and vitamin K-dependent protein function in atherosclerosis. *J. Thromb. Haemost.* **2**, 2118–2132 (2004).
- 41 Booth, S. L. Roles for vitamin K beyond coagulation. *Annu. Rev. Nutr.* **29**, 89–110 (2009).
- 42 Garcia, A. A. & Reitsma, P. H. VKORC1 and the vitamin K cycle. *Vitam. Horm.* **78**, 23–33 (2008).
- 43 Scott, S. A., Jaremkov, M., Lubitz, S. A., Kornreich, R., Halperin, J. L. & Desnick, R. J. CYP2C9*8 is prevalent among African Americans: implications for pharmacogenetic dosing. *Pharmacogenomics* **10**, 1243–1255 (2009).

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