Aim: Warfarin dose requirement is associated with VKORC1 rs9923231, and we studied whether it is a functional variant. Materials & methods: We selected variants in linkage disequilibrium with rs9923231 that bind transcription factors in an allele-specific way. Representative haplotypes were cloned or constructed, nuclear protein binding and transcriptional activity were evaluated. Results: rs56314408C>T and rs2032915C>T were detected in a liver enhancer in linkage disequilibrium with rs9923231. The rs56314408–rs2032915 C-C haplotype preferentially bound nuclear proteins and had higher transcriptional activity than T-T and the African-specific T-C. A motif for TFAP2A/C was disrupted by rs56314408T. No difference in transcriptional activity was detected for rs9923231G>A.

Conclusion: Our results supported an activating role for rs56314408C, while rs9923231G>A had no evidence of being functional.

First draft submitted: 15 December 2015; Accepted for publication: 26 January 2016; Published online: 5 February 2016

Keywords: regulatory sequences • vitamin K epoxide reductases • warfarin

Warfarin is still the most commonly prescribed oral anticoagulant for thrombotic disorders and atrial fibrillation [1]. Individual response varies, and to maintain adequate anticoagulation, warfarin dosing needs to be adjusted according to the prothrombin clotting time international normalized ratio (INR). Despite INR monitoring, patients sensitive to warfarin spend more time over-anticoagulated and have an increased risk of bleeding during the first 90 days of treatment [2]. Variation in genes associated with warfarin metabolism and response is a major cause of sensitivity to warfarin, but dose requirements are also influenced by age, gender, clinical status, body size, diet and interacting medications [3].

CYP2C9 encodes the main enzyme responsible for inactivating warfarin, cytochrome P450 2C9 [3]. The coding CYP2C9 variants *2 and *3 have higher allele frequencies and thus explain a larger proportion of variance in dose than in Africans and Asians [5]. It has been suggested that it would be beneficial to include additional genetic variants when predicting warfarin dose in non-European populations [6]. CYP2C9*6, *8 and the intergenic CYP2C single nucleotide polymorphism (SNP) rs12777823 have, for example, been shown to improve dose predictions in African–American patients [7–9].

Warfarin acts through the inhibition of vitamin K epoxide reductase [10] encoded by VKORC1 [11]. Rare coding variants of VKORC1 cause warfarin resistance [11], while common noncoding variants are associated with warfarin sensitivity [12]. These noncoding variants are thought to act through regulation of VKORC1 gene expression [5]. They exist in an extended region with high linkage disequilibrium (LD) including gene regulatory elements and genes flanking VKORC1. The two main VKORC1 haplotypes – the ‘normal’ and the warfarin sensitive haplotype...
– are generally represented by rs9923231 (c.-1639G>A) that is detected by clinical VKORC1 tests and included in genotype-guided warfarin dosing algorithms [13]. VKORC1 rs9923231 has been shown to explain up to a twofold difference in dose on an individual level [4], but the proportion of variance in dose explained (R²) differs greatly on a population level [5]. The frequency of rs9923231 A is approximately 37% in Europeans, 86% in East Asians and 10% in African–Americans [5,14]. In Europeans and East Asians, the variance explained is driven by allele frequencies (R² = 22.5% and R² = 18.4%, respectively), but they cannot fully explain the low R² in African–Americans (R² = 4.2%). It has been hypothesized that rs9923231 is a poor proxy for underlying functional variants in populations with high genetic diversity and low LD in the VKORC1 region.

Studies to determine VKORC1 functional variants have so far been contradictory and inconclusive [15–19]. In a study on VKORC1 expression, mRNA levels were about three times as high in liver homozygous for the normal dose haplotype, defined by rs9923231 G, compared with liver homozygous for the low-dose haplotype containing rs9923231 A [16]. In a subsequent study, a twofold allelic mRNA expression imbalance was observed between rs9923231 G/A in human liver and a 1.5-fold difference in the model cell line HepG2 [17]. Using luciferase reporter gene assays, a study showed that a construct containing rs9923231 G had 44% higher activity than one containing rs9923231 A [17], but this was not confirmed by two other studies [15,18]. There were higher signals for histone modifications associated with high transcriptional activity at rs9923231 G in the liver model cell line HepG2 and in liver samples that are heterozygous at this position [16]. However, as concluded by the authors, the allelic difference in these marks could be driven by other functional polymorphisms in high LD with rs9923231. At least five expression quantitative trait loci on the haplotype have been associated with VKORC1 gene expression based on analysis in adult human livers [20], but they were detected with statistical methods and their possible functionality has not been confirmed experimentally. Given the high LD around VKORC1, other SNPs on the haplotype are also potential drivers of the effect on VKORC1 gene expression.

The confusion surrounding whether rs9923231 regulates VKORC1 was enhanced by two large randomized clinical trials to evaluate the benefits of genotype-guided initiation of warfarin [21,22]. Genotype-guided dosing improved time in therapeutic INR range in the EU-PACT trial [21], but the American COAG trial failed to demonstrate this, and patients of African origin actually spent less time in therapeutic INR range after genotype-guided initiation of warfarin [22]. Notably, standard genotyping assays do not interrogate African-specific minor alleles of CYP2C9 or rarer haplotypes of VKORC1 not distinguished by rs9923231 [1,23]. Subsequently, it was shown that rs9923231 is associated with significantly lower average dose reductions per A allele in African–Americans than in European Americans [6]. This further illustrates that rs9923231 is a poor proxy for functional VKORC1 variants in a genetically diverse African–American population.

Whenever functional variants driving genome-wide association studies (GWAS) of disease or gene expression have been detected, a common feature is that they change the binding of transcription factors (TFs) in gene regulatory elements [24]. TF binding sites can be identified across the genome by chromatin immunoprecipitation followed by large-scale sequencing (ChIP-seq). Allele-specific TF binding can be detected by counting the number of ChIP-seq reads derived from either allele in heterozygous positions [25–27], for example, from public HepG2 data in the ENCYclopedia Of DNA Elements (ENCODE) project [28].

The aim of this study was to determine whether the conventionally analyzed VKORC1 SNP rs9923231 is located in a regulatory element in liver cells, and whether it is likely to drive the effect on warfarin dose requirements. If not, we aimed to find underlying candidate functional variants regulating VKORC1 in the liver model cell line HepG2.

### Methods

#### Allele-specific SNP definition

An allele specific SNP was defined by our established SNP selection pipeline [M Cavalli and C Wadelius, Unpublished Data] which identifies heterozygous SNPs that show a statistically significant difference in ChIP-seq read count over the two alleles. The pipeline followed these steps: publicly available raw reads (.fastq) from ChIP-seq for several TFs in HepG2 were obtained from the ENCODE project [28] database [38]. They were aligned to the reference University of California Santa Cruz human genome 19 assembly (UCSC hg19) based on the Genome Reference Consortium Human genome build 37 (GRCh37), but excluding random and unplaced contigs. They were also aligned to the HepG2-specific alternate allele, built using the FastaAlternateReferenceMaker GenomeAnalysisTK utility that generates an alternative reference sequence replacing the reference bases at variation sites with the bases supplied by a SNP collection. The HepG2 genome was sequenced to 10x coverage using 100 base pair (bp) paired-end reads, which in combination with reads from published HepG2 experiments and from our nucleosome sequencing and ChIP-seq experiments gave an average 55x genome coverage [29]. Illumina reads were aligned...
to the human genome using the alignment software package BWA aln. Sequencing by oligonucleotide ligation and detection (SOLiD) reads were aligned using the blat-like fast accurate search tool BFAST. Duplicate reads were removed from each sample. SNP calling was done using the GenomeAnalysisTK unified genotyper and high accuracy was verified by genotyping a GWAS array. Reads mapped specifically to the reference allele 1 or the alternate HepG2 allele 2 were counted at the heterozygous SNPs. SNPs with ‘0’ reads mapped on either allele were discarded. To determine whether the read counts differed significantly between the two alleles, a binomial test was applied against the null hypothesis of an equal coverage between the alleles. After correcting for multiple testing (Benjamini & Hochberg or FDR), allele specific SNPs with p < 0.05 were selected. Allele-specific SNPs were then intersected with the 1000 Genomes SNP collection in order to retrieve allele frequencies. Extensive filtering of the selected allele-specific SNPs was performed to remove those that were in centromeric or telomeric regions (UCSC Gap table ± 1 Mb), in blacklisted ENCODE regions (± 100 bp) or copy number variants. Pruned allele specific SNPs in LD (r^2 > 0.8) with the GWAS SNP rs9923231 (proxy SNPs calculated via the SNP annotation and proxy (SNAP) tool [30]) were selected as candidate functional allele-specific SNPs for experimental validation.

Cell cultures
HepG2 cells were cultured in RPMI 1640 medium (Sigma-Aldrich) supplemented with 10% non-inactivated fetal bovine serum (FBS), L-glutamine, and penicillin-streptomycin solution (PEST; Sigma-Aldrich, MO, USA) at 37°C with 5% CO_2.

Construction of cloning plasmids & luciferase report assays
All luciferase expression constructs were built based on pGL4.23 renilla luciferase reporter vectors from Promega (Promega, WI, USA). The control of cell death B toxin (ccdB) expression cassette was inserted into restriction enzyme KpnI and EcoRV sites of pGL4.23 to construct pGL4.23-ccdB, which was used as a basal vector to diminish false-positive signal during the cloning process. Genomic sequences surrounding the GWAS SNP rs9923231 and the haplotype rs56314408–rs2032915 were amplified by Phusion Hot Start Flex DNA polymerase (New England Biolabs). The control of cell death B toxin (ccdB) expression cassette was inserted into restriction enzyme KpnI and EcoRV sites of pGL4.23 to construct pGL4.23-ccdB, which was used as a basal vector to diminish false-positive signal during the cloning process. Genomic sequences surrounding the GWAS SNP rs9923231 and the haplotype rs56314408–rs2032915 were amplified by Phusion Hot Start Flex DNA polymerase (New England Biolabs). The control of cell death B toxin (ccdB) expression cassette was inserted into restriction enzyme KpnI and EcoRV sites of pGL4.23 to construct pGL4.23-ccdB, which was used as a basal vector to diminish false-positive signal during the cloning process. Genomic sequences surrounding the GWAS SNP rs9923231 and the haplotype rs56314408–rs2032915 were amplified by Phusion Hot Start Flex DNA polymerase (New England Biolabs) and the concentration was determined by Qubit® Protein Assay Kit (Life Technologies, CA, USA). For each haplotype tested, oligonucleotide probes (45 bp: 10 bp + SNP + 23 bp + SNP + 10 bp) were designed in both the cold and 5′-biotinylated form (Integrated DNA Technologies, IA, USA). Biotinylated and unbio-
tinylated oligonucleotides were annealed with reverse complementary oligonucleotides (95–25°C temperature stepdown). For the binding reaction, 3–6 µg of nuclear extract was incubated for 40 min on ice with 200 fmol of each biotinylated double-stranded DNA probe in 10 mM TrisHCl, 30 mM KCl, 1mM DTT, 1 µg of Poly (dI-dC), 7.5% glycerol, 0.063% NP-40, 2 mM MgCl_2 and 0.1 mM EDTA. For the competition assays, 20 pmol of unlabelled double-stranded DNA probes were added to the binding reaction. Samples were electrophoresed in Criterion™ 5% tris boric acid EDTA (TBE) Precast Gels (Bio-Rad, CA, USA), and electro-transferred into a Genescreen plus™ hybridization transfer membrane (Perkin Elmer, MA, USA). DNA–protein complexes were crosslinked using UV light and detected by chemiluminescence using LightShift® Chemiluminescent EMSA Kit (Thermo Scientific, MA, USA).

Electrophoresis mobility shift assay
Nuclear extracts were prepared from HepG2 cells using the NucBuster™ Protein Extraction kit (Novagen, WI, USA) and the concentration was determined by Qubit® Protein Assay Kit (Life Technologies, CA, USA). For each haplotype tested, oligonucleotide probes (45 bp: 10 bp + SNP + 23 bp + SNP + 10 bp) were designed in both the cold and 5′-biotinylated form (Integrated DNA Technologies, IA, USA). Biotinylated and unbio-
tinylated oligonucleotides were annealed with reverse complementary oligonucleotides (95–25°C temperature stepdown). For the binding reaction, 3–6 µg of nuclear extract was incubated for 40 min on ice with 200 fmol of each biotinylated double-stranded DNA probe in 10 mM TrisHCl, 30 mM KCl, 1mM DTT, 1 µg of Poly (dI-dC), 7.5% glycerol, 0.063% NP-40, 2 mM MgCl_2 and 0.1 mM EDTA. For the competition assays, 20 pmol of unlabelled double-stranded DNA probes were added to the binding reaction. Samples were electrophoresed in Criterion™ 5% tris boric acid EDTA (TBE) Precast Gels (Bio-Rad, CA, USA), and electro-transferred into a Genescreen plus™ hybridization transfer membrane (Perkin Elmer, MA, USA). DNA–protein complexes were crosslinked using UV light and detected by chemiluminescence using LightShift® Chemiluminescent ENSA Kit (Thermo Scientific, MA, USA).
Results
Evaluation of whether rs9923231 is a regulatory functional variant

Data available so far from the ENCODE [28] and Epigenome Roadmap [32] projects do not provide evidence that the GWAS top hit for association with warfarin dose rs9923231 (chr 16:31107689 G>A) is located in a regulatory element in liver cells (Figure 1 & Supplementary Figure 1). The lack of difference in transcriptional activity in luciferase reporter assays between the two alleles further supports that rs9923231 may not be the functional variant driving the allelic difference in VKORC1 activity detected in liver samples and HepG2. All luciferase assays in this study were performed in six replicates per transfection – see the ‘Methods’ section.

---

Figure 1. Evaluation of the functional activity of rs9923231. UCSC genome browser view of the genomic context for rs9923231. In the insert, the enhancer activity of the two alleles of rs9923231 was tested by luciferase assay with the empty vector as control. ns: Not significant; RLU: Relative luciferase unit.
Novel regulatory variant detected on the VKORC1 haplotype that is associated with warfarin dose

**Research Article**

Regulatory functional variants in LD with rs9923231

We utilized genome-wide TF binding data from ENCODE ChIP-seq experiments [28] in HepG2 cells to identify variants that regulate nearby genes (cis-acting regulatory variants). SNPs that showed allele-specific TF binding (allele-specific SNPs) were suggested to be functional, and in total we found 3001 such variants in HepG2. We identified an allele-specific SNP in high LD ($r^2$ = 0.8–0.9) with rs9923231, namely rs56314408 (chromosome 16:31117389 C>T) located in an enhancer upstream of the gene BCKDK. Enhancers are known to regulate the activity of genes located several kilobases away and sometimes over distances of hundreds of kilobases, so rs56314408 clearly is at a position from which it could regulate the activity of VKORC1. One nearby SNP rs2032915 (chr 16:31117413 C>T) was located 24 bp away in the same enhancer. In contrast to rs9923231, the putative functional SNP pair shows evidence of TF binding sites defined by ChIP-seq, of being sensitive to cleavage by the DNase I enzyme (DNAse I hypersensitive), and of active enhancer histone marks according to ENCODE [28] and Epigenome Roadmap data [32] (Supplementary Figure 2).

Counting ChIP-seq reads across all TFs aligned to the two alleles of rs56314408, we found more reads (156) from the T-allele than from the C-allele (64; $p < 1e-4$). There were also more reads from the T-allele of rs2032915 (149) than the C-allele (98; $p < 1e-4$). Both these signals were significantly different at a genome-wide level. This suggests that one or both of these SNPs are functional and, according to ENCODE ChIP-seq data [28], the rs56314408 T-rs2032915 T-haplotype is preferentially bound by the TFs so far analyzed in ENCODE.

Evaluation of enhancer activity with luciferase assay

We cloned a fragment of the regulatory element in HepG2 (381 bp in blue in Figure 2B) which contains most of the currently annotated regulatory sequences at this element. We obtained the two rs56314408-rs2032915 haplotypes C-C and T-T that are common in people of European and East Asian ancestry (Table 1). The C-C haplotype had a significantly higher activity than the TT haplotype and both of the haplotypes were found to have higher enhancer activities than the fragment containing rs9923231. We then focused on a smaller fragment (186 bp in orange in Figure 2B) centered at the two SNPs that cause allele-specific activity. The T-C haplotype, which is almost exclusively found in African populations (Table 1), was built by site-directed mutagenesis. The C-T haplotype is extremely rare in all populations and was therefore not studied.

We tested the three rs56314408–rs2032915 haplotypes C-C, T-T and T-C in reporter gene luciferase assays. We observed a significantly higher transcriptional activity with the 186 bp construct containing the C allele of rs56314408 ($p < 1e-4$), but no difference in activity between the C and T alleles of rs2032915 (Figure 3A). The rs56314408 C variant is on the same haplotype as the high dose rs9923231 G variant in Europeans and Asians.

Using the motif-based sequence analysis tools (MEME suite) [33] and the Catalog of Inferred Sequence Binding Preferences (CIS-BP) [34], we scanned the DNA fragments tested in luciferase for TF binding sites. We identified, among others, binding sites for USF1 and YY1 that also showed significant ChIP-seq peaks in this region in HepG2, however, neither of these binding sites overlapped with any of the two SNPs. We overexpressed USF1 and YY1 in HepG2 cells together with the 186 bp luciferase constructs of the three haplotypes, respectively. There was a significant increase in expression for all haplotypes and a highly significant difference in activity between the C and T alleles of rs56314408 ($p < 1e-4$) (Figure 3C). When YY1 was overexpressed, there was also a small difference between the C and T alleles of rs2032915 ($p < 0.0003$) suggesting that this SNP may partially contribute to the haplotype effect in some conditions.

---

**Table 1. Haplotype frequencies for rs56314408–rs2032915 in Europeans, East Asians and Africans according to 1000 Genomes Phase III data.**

<table>
<thead>
<tr>
<th>Haplotype frequencies</th>
<th>Europeans (n = 503)</th>
<th>East Asians (n = 504)</th>
<th>Africans (n = 661)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-C</td>
<td>60.7%</td>
<td>11.5%</td>
<td>51.3%</td>
</tr>
<tr>
<td>T-T</td>
<td>38.0%</td>
<td>88.5%</td>
<td>1.8%</td>
</tr>
<tr>
<td>T-C</td>
<td>1.2%</td>
<td>0.0%</td>
<td>46.9%</td>
</tr>
<tr>
<td>C-T</td>
<td>0.1%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Total</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Data taken from [37].

---

[Figure 2B] [Figure 3A] [Figure 3C]
Prediction of allele-specific TF binding sites
We used the online library of transcription factors and their DNA-binding motifs CIS-BP, and the transcription factor affinity prediction (TRAP) [35] Web Tools to test the SNPs for potential differential binding of TFs. The results suggest that rs56314408 C is preferentially bound by TFAP2 (A/B/C/E), and that rs2032915 C is bound by NKX2 (-2/-8). Both TFs are expressed in the liver, but have not been analyzed by ChIP-seq in HepG2. TFAP2A and TFAP2C have, however, been analyzed in HeLaS3 and both showed a significant enrichment in this region. rs56314408 is located in the predicted motif of TFAP2A and TFAP2C with the C allele at position 9 (Figure 3D) conforming better to the consensus.

Evaluation of nuclear protein binding with electrophoretic mobility shift assays
In an electrophoresis mobility shift assay (EMSA), all nuclear protein expressed in the cell are assayed simultaneously. The three haplotypes were tested by EMSA with nuclear protein extract from HepG2 cells. We found that the C allele of rs56314408 was preferentially bound by nuclear proteins (Figure 3B) as indicated by the arrow. Combined with the predicted binding of TFAP2A and TFAP2C to this allele described above it supports an activating role for the C-C haplotype.

Discussion
Common genetic variants of warfarin’s target vitamin K epoxide reductase (VKOR) have a large effect
Figure 3. Evaluation of different functional activity and binding of the rs56314408–rs2032915 haplotypes. (A) Luciferase assay testing the enhancer activity of the 186 bp constructs of the three most common haplotypes for rs56314408–rs2032915. The empty vector pGL4.23 is included as a control. (B) EMSA testing nuclear protein binding to the three different haplotypes. (C) Luciferase assay testing the variation in enhancer activity of the 186 bp constructs of the three haplotypes from panel A upon overexpression of YY1 and USF1. (D) TFAP2 motif disrupted by rs56314408 (arrow).

* p < 0.01; ** p < 0.001; *** p < 0.0001.

EMSA: Electrophoresis mobility shift assay; ns: Not significant; RLU: Relative luciferase unit.

Novel regulatory variant detected on the VKORC1 haplotype that is associated with warfarin dose

Research Article

on dose requirements [12,16]. These variants are non-coding and believed to act by changing the expression of the VKORC1 gene and protein, but it has not been resolved how the effect is mediated.

There is prior evidence against rs9923231 being the functional polymorphism regulating VKORC1. According to ENCODE [20] and Epigenome Roadmap [32] data so far, rs9923231 is not located in a regulatory element active in HepG2 and adult liver cells. This is further supported by our luciferase assay that shows no difference in activity between rs9923231 alleles. Our data suggest that rs56314408 could be a functional variant regulating VKORC1, perhaps with a minor contribution from rs2032915. In the European population, both of these SNPs are in high LD with rs9923231. They are located in a regulatory element active in HepG2 and liver cells, and display difference in TF binding between alleles in ChIP-seq. EMSA experiments of the rs56314408–rs2032915 haplotypes showed preferential nuclear protein binding to the rs56314408 C allele. In functional luciferase assays, there is a significant difference in activity between rs56314408 C and T alleles. The element is activated by overexpression of TFs known to bind there, resulting in a highly significant difference between the rs56314408 alleles. According to ChIP-seq signals in...
HeLaS3 cells, the rs56314408 T allele disrupts a motif for the TFs TFAP2A and TFAP2C that are expressed in the liver. According to ENCODE ChIP-seq data [28], there were more ChIP-seq reads from the rs56314408 T allele, which could occur if, ENCODE by chance had tested TFs acting as repressors at this element out of the hundreds of TFs that are present in the cell.

The allele frequencies of rs9923231, rs56314408 and rs2032915 differ markedly between populations [36], resulting in distinct haplotype and LD patterns (Table 1 & Figure 4). It is interesting to note that allele frequencies of rs56314408 C are very similar in European (CEU) and African Yoruban (YRI) populations, 0.56 and 0.53, respectively. In Europeans, LD is high between rs9923231 and rs56314408 (LD = 0.8–0.9), rs9923231 and rs2032915 (LD = 0.8–0.9), and between rs2032915 and rs56314408 (LD = 0.9–1.0). In East Asians, LDs are almost equally high, 0.8, 0.7 and 0.9, respectively, and genetic variation is low at these sites. This suggests that rs56314408 would not improve dose predictions in a European or East Asian population.

On the other hand, in African–Americans, LD is high between rs9923231 and rs2032915 (LD = 0.8), but low between rs9923231 and rs56314408 (LD = 0.2), and rs2032915 and rs56314408 (LD = 0.2). If rs56314408 drives VKORC1, the conventionally analyzed rs9923231 would only predict warfarin dose when in high LD with rs56314408. It is therefore warranted to genotype rs56314408 in warfarin-treated African–Americans to test whether it could improve warfarin dose predictions.

**Conclusion**

The conventionally analyzed rs9923231 has no evidence of being a functional variant. We propose that rs9923231 predicts warfarin dose well only when in high LD with rs56314408 or other possible functional variants.

**Author contributions**

M Wadelius and C Wadelius conceived the study, C Wadelius and M Cavalli designed the work, M Cavalli, G Pan and H Nord performed the analysis, and interpreted the data together with C Wadelius, M Wadelius and N Eriksson. M Cavalli drafted the manuscript, C Wadelius and M Wadelius revised it critically, and all authors approved the final version to be published.

**Financial & competing interests disclosure**

This study was supported by Uppsala University, the Swedish Research Council (Science and Technology 621-2011-6052, Medicine 521-2010-3505, Medicine 521-2011-2440 and 521-2014-3370), the Swedish Heart and Lung Foundation (20120557 and 20140291), Selander’s foundation, Thureus’...
Novel regulatory variant detected on the VKORC1 haplotype that is associated with warfarin dose

Research Article

Foundation and the Clinical Research Support (ALF) at Uppsala University. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project b2010003 and b2011107. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

Supplementary data

Additional supporting information may be found in the online version of this article: Supplementary Figure 1. UCSC genome browser view for the GWAS SNP rs9923231; Supplementary Figure 2. UCSC genome browser view for the SNPs pair rs56314408–rs2032915; Supplementary Table 1. Oligonucleotide sequences used in this study.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/

Executive summary

Background

- Warfarin dose requirements vary between individuals, and are influenced by coding variants in CYP2C9 and presumed VKORC1 regulatory variants, but due to high linkage disequilibrium (LD) in the VKORC1 region the causative variants have not been found.
- We aimed to find single nucleotide polymorphisms (SNPs) variants in LD with rs9923231 that are candidates to regulate VKORC1.

Findings

- According to ENCODE data, the SNP rs56314408 binds transcription factors in an allele-specific way, and disrupts a motif for TFAP2A/C.
- rs56314408 is located in a liver enhancer in LD with rs9923231 together with the neighboring SNP rs2032915.
- The rs56314408-rs2032915C-C haplotype preferentially bound nuclear proteins in electrophoresis mobility shift assays, and had higher transcriptional activity in luciferase reporter assays than T-T and the African-specific T-C haplotype.
- rs9923231 is not located in a regulatory element according to ENCODE data, and no difference in transcriptional activity was detected.

Conclusion

- The conventionally analyzed VKORC1 rs9923231 has no evidence of being a functional variant.
- We propose that rs9923231 predicts warfarin dose well only when in high LD with rs56314408 or other possible functional variants.

References


