MEASURES OF EXPOSURE IMPACT GENETIC ASSOCIATION STUDIES: AN EXAMPLE IN VITAMIN K LEVELS AND VKORC1

DANA C. CRAWFORD

Institute for Computational Biology, Department of Epidemiology and Biostatistics, Case Western Reserve University, Wolstein Research Building, 2103 Cornell Road, Suite 2527, Cleveland, OH 44106, USA
Email: dana.crawford@case.edu

KRISTIN BROWN-GENTRY

Center for Human Genetics Research, Vanderbilt University, 519 Light Hall, 2215 Garland Avenue, Nashville, TN 37232, USA
Email: Kristin.Gentry@healthspring.com

MARK J. RIEDER

Adaptive Biotechnologies Corporation, 1551 Eastlake Avenue East, Suite 200, Seattle, WA 98102, USA
Email: mrieder@adaptivebiotech.com

Studies assessing the impact of gene-environment interactions on common human diseases and traits have been relatively few for many reasons. One often acknowledged reason is that it is difficult to accurately measure the environment or exposure. Indeed, most large-scale epidemiologic studies use questionnaires to assess and measure past and current exposure levels. While questionnaires may be cost-effective, the data may or may not accurately represent the exposure compared with more direct measurements (e.g., self-reported current smoking status versus direct measurement for cotinine levels). Much like phenotyping, the choice in how an exposure is measured may impact downstream tests of genetic association and gene-environment interaction studies. As a case study, we performed tests of association between five common *VKORC1* SNPs and two different measurements of vitamin K levels, dietary (n=5,725) and serum (n=348), in the Third National Health and Nutrition Examination Studies (NHANES III). We did not replicate previously reported associations between *VKORC1* and vitamin K levels using either measure. Furthermore, the suggestive associations and estimated genetic effect sizes identified in this study differed depending on the vitamin K measurement. This case study of *VKORC1* and vitamin K levels serves as a cautionary example of the downstream consequences that the type of exposure measurement choices will have on genetic association and possibly gene-environment studies.

1. Introduction

Complex human diseases and traits are shaped both by genetics and the environment. The development of dense genotyping arrays in the past decade has enabled genome-wide association studies in large epidemiologic and clinic-based collections, and these studies have been successful in identifying thousands of common genetic variants associated with common disease [1]. The more recent advent of relatively

cost-effective sequencing technologies is now contributing towards knowledge of rare genetic variants contributing to human disease and traits [2].

While genomic technologies have made tremendous advancements in the number of variants that can be assayed or the number of reads that can be sequenced, the methods to analyze such complex data have not evolved as rapidly. Also, very few studies have attempted to incorporate environmental exposures or gene-environment interactions. Indeed, most studies that have examined the impact of gene-environment interaction on common diseases or traits have been limited to candidate gene studies [e.g., 3] as there is little consensus on how to best test for gene-environment interactions on the genome-wide scale [4].

Another challenge faced in examining the effects on environmental exposures on human health is how to measure these data. Most large-scale epidemiologic collections use questionnaires to assess past and present exposures, which can vary in accuracy and be associated with substantial biases depending on the exposure being measured [e.g., 5]. Clinic-based collections are further hampered by the fact that most clinics do not routinely collect exposure data in a standardized manner [6]. To further complicate the field, substantial across-study differences exist in how exposure data is collected, making data harmonization efforts difficult [7,8]. Oftentimes, the simplest or most frequently used measure of exposure found across studies is chosen for harmonization across studies despite the availability of more accurate exposure measures albeit at the cost of statistical power.

It is presently unclear what impact the common practices of exposure harmonization across studies have on genetic association study findings. To document potential genetic association study differences due to differences in exposures measures, we examined the association of vitamin K epoxide reductase complex subunit 1 (*VKORC1*) common genetic variation and two different measures of vitamin K levels in the Third National Health and Nutrition Examination Surveys (NHANES III).

Vitamin K is a fat-soluble vitamin essential for blood clotting and bone formation. Vitamin K is found in two forms, K_1 (phylloquinone) and K_2, the former of which is synthesized by green leafy plants such as kale, spinach, and collards. Vitamin K is a required co-enzyme for the γ-carboxylation of three glutamic acid (Glu) residues in osteocalcin, converting them to gamma-carboxyglutamic acid (Gla) as part of the vitamin cycle. VKORC1 is a co-enzyyme in this vitamin cycle required for the recycling of vitamin K and the eventual Glu to Gla conversion [9,10]. The anticoagulant drug Warfarin targets VKORC1 and effectively blocks the recycling of vitamin K, which in turn blocks the blood clotting process.

It is well established that *VKORC1* genetic variants are associated with warfarin dosing [11,12]. Previous candidate gene studies have also suggested that *VKORC1* common variants are associated with vitamin K levels. VKORC1 is a rate-controlling enzyme in the vitamin cycle and common genetic variants within *VKORC1* have previously been associated with vitamin K levels [13,14]. Data presented here suggest that *VKORC1* associations differ for vitamin levels measured from dietary questionnaires compared with serum and serve as a cautionary case study that may be applicable to further gene-environment studies.

2. Methods

2.1. *Study population*

The study population and DNA samples described here are from the Third National Health and Nutrition Examination Surveys (NHANES III) conducted by the National Center for Health Statistics (NCHS) at the Centers for Disease Control and Prevention (CDC). NHANES III was conducted between 1988 and 1994 in two phases, and biospecimens for DNA extraction were collected in phase 2 (1991-1994). NHANES is a series of cross-sectional surveys designed to document the health status of Americans at the time of ascertainment; as such, participants are ascertained regardless of health status. NHANES III was a complex survey design where minority (non-Hispanic blacks and Mexican Americans) participants as well as the elderly were oversampled. Health data on NHANES III participants is collected through questionnaires, a physical examination by health professionals, and laboratory measures. All physical examinations are performed in the Mobile Examination (MEC) unless the participant is physically unable, in which case the examinations are conducted in the participant's home. Design and operations of NHANES III has previously been described [15].

The present study was approved by the CDC Ethics Review Board. Because the study investigators did not have access to personal identifiers, this study was considered non-human subjects research by the Vanderbilt University Internal Review Board.

2.2. *Genotyping*

We selected five *VKORC1* tagSNPs as previously described [16,17] and allele frequencies for NHANES III have been previously published [12]. Briefly, tagSNPs were chosen to tag common *VKORC1* genetic variation for mostly European-descent populations. SNP rs2884737 was genotyped using Sequenom's iPLEX® Gold coupled with MassARRAY MALDI-TOF MS detection (San Diego, CA). SNPs rs9923231, rs9934438, rs8050894, and rs2359612 were genotyped using Applied Biosystem's (now Thermo Scientific) TaqMan® SNP Genotyping Assays (Foster City, CA). All SNPs were genotyped by the Vanderbilt University's Center for Human Genetics Research DNA Resources Core. All SNPs were in Hardy Weinberg Equilibrium. In additional to genotyping experimental NHANES III DNA samples, we genotyped 368 blinded duplicates required by CDC for additional quality control. All genotypes have been deposited into CDC's Genetic NHANES database and are available for secondary analysis.

2.3. *Vitamin K*

Vitamin K was collected on NHANES III participants in two ways. First, total nutrient intake of vitamin K (mcg) was collected from the 24-hour dietary recall performed in the MEC [15]. These data were collected in collaboration with the University of Minnesota's Nutrition Coordinating Center (NCC). Second, serum vitamin K (phylloquinone; ng/ml) was measured using reverse phase HPLC [18] in non-Hispanic white women ages 6-29 years in Phase 2 of NHANES III. According to NHANES documentation, the lower

limit of detection is 0.05 and the range of serum vitamin K levels in the subset of NHANES samples tested was 0.05ng/ml-6.799ng/ml.

2.4. *Statistical methods*

Single SNP tests of association were performed unadjusted and adjusted using linear regression assuming an additive genetic model stratified by self-identified race/ethnicity. Two dependent variables were tested for an association with each SNP: dietary vitamin K levels and serum vitamin K levels, both log transformed. Serum vitamin K adjusted models included age, body mass index, current smoking status, dietary calcium, phosphorous, magnesium, iron, zinc, copper, sodium, potassium, protein, carbohydrates, fiber, total vitamin A, total carotenes, total alpha-tocopherol equivalents, vitamin C, vitamin B_6, vitamin B_{12}, folic acid, and total calories as covariates. Body mass index (continuous variable) was calculated based on height and weight measured at the MEC. Current smoking status (binary variable) was defined by "do you smoke cigarettes now?" or cotinine levels > 15ng/ml. The remaining dietary covariates (continuous traits) were available in NHANES III from the 24-hour dietary recall performed in the MEC. Dietary vitamin K models included the same covariates as serum vitamin K models with the addition of sex as a covariate. All analyses were conducted remotely in SAS v9.2 (SAS Institute, Cary, NC) and SUDAAN (Research Triangle Institute, Research Triangle Park, NC) using the Analytic Data Research by Email (ANDRE) portal of the CDC Research Data Center in Hyattsville, MD. All analyses presented here were performed weighted to account for the complex survey design. Results of tests of association wee visualized using Synthesis View [19].

3. Results

Study population characteristics are given in Table 1. On average, study participants with dietary vitamin K levels were more likely to be female among non-Hispanic whites and non-Hispanic blacks. The average ages for non-Hispanic blacks and Mexican Americans were younger than non-Hispanic whites and both populations on average overweight (body mass index >25 mg/k^2) compared with non-Hispanic whites. Mean dietary vitamin K levels were similar across all three racial/ethnic groups.

Given that vitamin K levels from dietary recall data are more readily available than serum levels, we first performed single SNP tests of association with these data stratified by race/ethnicity and adjusted for demographic and relevant covariates (see Methods). From all tests of association, only rs8050894 was significantly associated with dietary vitamin K levels in non-Hispanic whites (beta = 0.06; p=0.01; Figure 1). No test of association was significant all race/ethnicities for the same *VKORC1* SNP.

We then performed tests of association for all SNPs among non-Hispanic white females with serum vitamin K levels and compared these data with tests of association performed among non-Hispanic whites with dietary vitamin K levels (Figure 2). Similar to dietary vitamin K levels, only one SNP in non-Hispanic whites was associated with serum vitamin K levels (beta = 0.16; p = 0.03; Figure 2). However, the SNP associated with serum vitamin K levels (rs2359612) is not the same SNP associated with dietary vitamin K levels (rs8050894). The genetic effect sizes (betas) estimated for dietary vitamin K levels were

lower than those estimated for serum vitamin K levels with the exception of the genetic effect size estimated for rs9923231 (Figure 2).

Based on the disparate results obtained in models where the dependent variable was either dietary or serum vitamin K levels, we tested for a correlation between the variables. A total of 229 non-Hispanic white females have both vitamin K measurements available in NHANES III. The Pearson Correlation Coefficient was 0.11 between the two vitamin K estimates, which was not statistically significant (p=0.08).

Table 1. NHANES III study population characteristics. Unweighted descriptive statistics are shown for basic demographic variables (sex, age, and body mass index) as well as the two measures of vitamin K levels. Sample sizes shown are for participants with dietary vitamin K levels available. Serum vitamin K levels were only measured in non-Hispanic white women ages 6 – 29 years in Phase 2 of NHANES III (n=348). Abbreviations: standard deviation (SD), natural log (ln).

	Non-Hispanic whites	Non-Hispanic blacks	Mexican Americans
	(n=2,344)	(n=1,675)	(n=1,706)
% female	61	58	50
Mean (±SD) age in years	53.46 (20.32)	40.79 (16.71)	41.15 (17.42)
Mean (±SD) body mass index (kg/m^2)	26.66 (5.6)	27.3 (369.6)	27.1 (422.6)
Mean (±SD) ln(dietary vitamin K) (mcg)	4.05 (0.97)	4.02 (1.29)	3.79 (0.99)
Mean (±SD) ln(serum vitamin K) (mg/dl)	-1.30 (0.79)	-	-

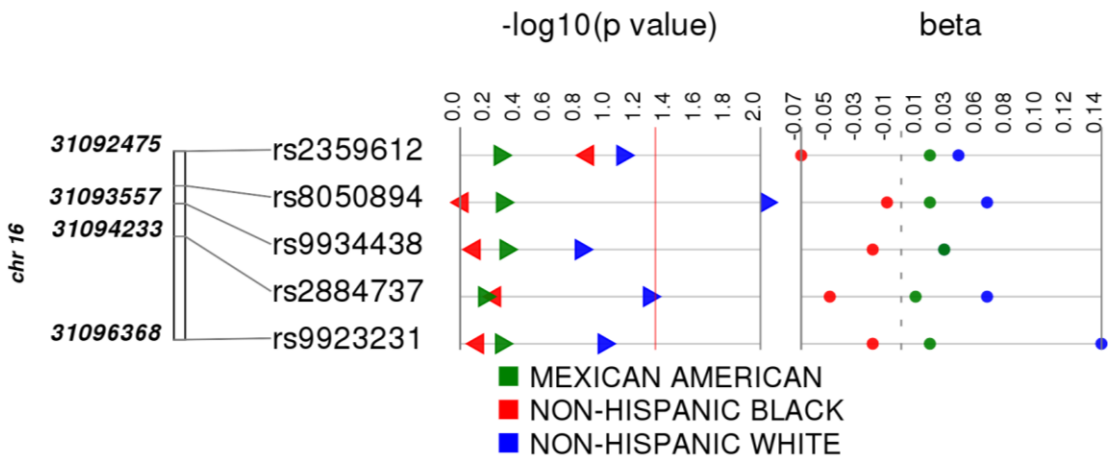
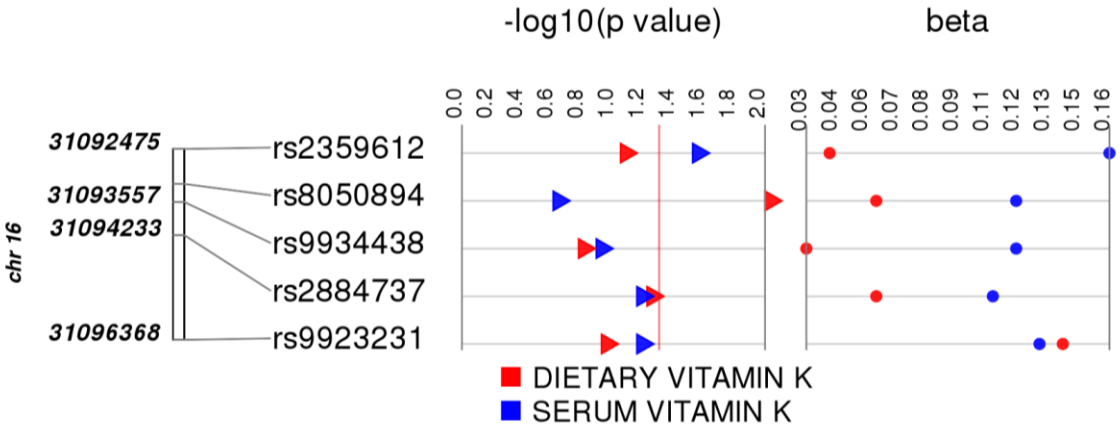


Figure 1. Results of tests of association between *VKORC1* common variation and dietary vitamin K levels stratified by race/ethnicity. Tests of association were performed using linear regression assuming an additive genetic model and adjusted for sex, age, body mass index, current smoking status, dietary calcium, phosphorous, magnesium, iron, zinc, copper, sodium, potassium, protein, carbohydrates, fiber, total vitamin A, total carotenes, total alpha-tocopherol equivalents, vitamin C, vitamin B_6, vitamin B_{12},



folic acid, and total calories. Results (p-values and betas) are plotted by rs number and race/ethnicity (coded by color) using Synthesis View. The direction of the triangles represents the direction of the genetic effect. The red line represents a p-value threshold of 0.05.

Figure 2. Comparison of results of tests of association between *VKORC1* common variation and vitamin K levels by vitamin K measurement. Tests of association were performed using linear regression assuming an additive genetic model and adjusted for sex (for model with dietary vitamin K

levels), age, body mass index, current smoking status, dietary calcium, phosphorous, magnesium, iron, zinc, copper, sodium, potassium, protein, carbohydrates, fiber, total vitamin A, total carotenes, total alpha-tocopherol equivalents, vitamin C, vitamin B_6, vitamin B_{12}, folic acid, and total calories. Results (p-values and betas) are plotted by rs number and vitamin K measurement (coded by color) using Synthesis View. The direction of the triangles represents the direction of the genetic effect. The red line represents a p-value threshold of 0.05.

4. Discussion

To illustrate potential genetic association differences dependent on exposure measurement differences, we tested five *VKORC1* for an association with both dietary and serum-measured vitamin K levels. We identified suggestive associations for both dietary and serum-measured vitamin K levels; however, the results were not concordant between the two vitamin K measurements. Moreover, the genetic effect sizes estimated for each test of association differed between dietary and serum-measured vitamin K levels. Finally, we did not directly replicate previously reported associations between *VKORC1* common genetic variants and vitamin K levels in NHANES III.

Previously studies have suggested that vitamin K levels are associated with *VKORC1* genetic variants. More specifically, Nimptsch and colleagues [13] demonstrated in 548 German males and females that dietary vitamin K levels were inversely correlated with serum undercarboxylated osteocalcin/ total intact osteocalcin ratio levels dependent on rs2359612 genotypes. Crosier and colleagues [14] describe an association between plasma phylloquinone levels and rs8050894 in 416 older (60-80 years) men and women primarily of European descent. Interestingly, both rs2359612 and rs8050894 were suggestively associated with serum and dietary levels of vitamin K, respectively, at $p<0.05$ in this study. However, the tests of association and results described here are not a replicate of the tests performed by Nimptsch et al [13] and Crosier et al [14] given the different modeling assumptions measures of vitamin K associated with each SNP.

We have found that within NHANES III, serum levels of vitamin K are not significantly correlated with dietary measures of vitamin K levels. Unlike the present study, previous studies have suggested a correlation between dietary intake and serum levels of vitamin K levels [20, 21]. The lack of correlation observed in NHANES III may explain in part the differences in observed genetic associations. Indeed, the two measures of vitamin K levels examined here are likely measuring different traits with different underlying genetic architectures. Dietary vitamin K levels are likely measuring phylloquinone levels (K_1) [10]. For serum vitamin K levels, previous family studies have suggested that plasma phylloquinone levels have non-significant heritability [22].

This study has several limitation and strengths. A major limitation of the current study is sample size and power. Although we were properly powered to detect associations with dietary vitamin K levels with >5,000 participants, we have many fewer participants with serum vitamin K levels (n=348). Furthermore, serum vitamin K levels were only measured in a subset of non-Hispanic white females, which may impact the generalizability of these tests of association and comparison across studies. Also, NHANES III did not

collect data for serum undercarboxylated osteocalcin/ total intact osteocalcin ratio levels, another measure of vitamin K levels, on any of the participants. NHANES III also only measured vitamin K levels once per participant for both serum levels as well as dietary intake. It is possible that seasonal variations in vitamin K levels exist and may have impacted this study. Finally, neither ancestry informative markers nor GWAS-level data are available in NHANES III. It is possible that population stratification may have impacted the results observed here particularly for admixed populations such as Mexican Americans.

Despite these limitations, a major strength of NHANES III is the availability of both questionnaire-based and laboratory-based exposure data even for only a fraction of the study population. Most epidemiologic studies collect only questionnaire-based data given it is more cost-effective than the alternative. Self-reported data may be not be as accurate as a laboratory assay, the latter of which is more attractive for quantitative trait genetic association studies. Even within this limited dataset, our results suggest that choice of exposure measurement may have an impact on the results and interpretation of a genetic association study.

5. Acknowledgments

This work was supported, in part, by NIH grant NS053646 (MJR). We would like to thank Jody McLean and Dr. Geraldine McQuillan from the National Center for Health Statistics at the Centers for Disease Control and Prevention for their assistance with the NHANES III genetic data. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the National Institutes for Health or the Centers for Disease Control and Prevention. The Vanderbilt University Center for Human Genetics Research, Computational Genomics Core provided computational and/or analytical support for this work.

References

1. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. *Proc Natl Acad Sci USA.* **106**(23):9362-7 (2009).

2. TG and HDL Working Group of the Exome Sequencing Project, National Heart, Lung, and Blood Institute, Crosby J, Peloso GM, Auer PL, Crosslin DR, Stitziel NO, Lange LA, Lu Y, Tang ZZ, Zhang H, Hindy G, Masca N, Stirrups K, Kanoni S, Do R, Jun G, Hu Y, Kang HM, Xue C, Goel A, Farrall M, Duga S, Merlini PA, Asselta R, Girelli D, Olivieri O, Martinelli N, Yin W, Reilly D, Speliotes E, Fox CS, Hveem K, Holmen OL, Nikpay M, Farlow DN, Assimes TL, Franceschini N, Robinson J, North KE, Martin LW, DePristo M, Gupta N, Escher SA, Jansson JH, Van Zuydam N, Palmer CN, Wareham N, Koch W, Meitinger T, Peters A, Lieb W, Erbel R, Konig IR, Kruppa J, Degenhardt F, Gottesman O, Bottinger EP, O'Donnell CJ, Psaty BM, Ballantyne CM, Abecasis G, Ordovas JM, Melander O, Watkins H, Orho-Melander M, Ardissino D, Loos RJ, McPherson R, Willer CJ, Erdmann J, Hall AS, Samani NJ, Deloukas P, Schunkert H, Wilson JG, Kooperberg C, Rich SS, Tracy RP, Lin DY, Altshuler D, Gabriel S, Nickerson DA,

Jarvik GP, Cupples LA, Reiner AP, Boerwinkle E, Kathiresan S. *N Engl J Med*. **371**(1):22-31 (2014).

3. Dumitrescu L, Carty CL, Franceschini N, Hindorff LA, Cole SA, Bůžková P, Schumacher FR, Eaton CB, Goodloe RJ, Duggan DJ, Haessler J, Cochran B, Henderson BE, Cheng I, Johnson KC, Carlson CS, Love SA, Brown-Gentry K, Nato AQ, Quibrera M, Shohet RV, Ambite JL, Wilkens LR, Le Marchand L, Haiman CA, Buyske S, Kooperberg C, North KE, Fornage M, Crawford DC. *Hum Genet*. **132**(12):1427-31 (2013).

4. Thomas D. *Nat Rev Genet*. **11**(4):259-72 (2010).

5. Mossavar-Rahmani Y, Tinker LF, Huang Y, Neuhouser ML, McCann SE, Seguin RA, Vitolins MZ, Curb JD, Prentice RL. *Nutr J*. **12**:63 (2013).

6. Wiley LK, Shah A, Xu H, Bush WS. *J Am Med Inform Assoc*. **20**(4):652-8 (2013).

7. Hamilton CM, Strader LC, Pratt JG, Maiese D, Hendershot T, Kwok RK, Hammond JA, Huggins W, Jackman D, Pan H, Nettles DS, Beaty TH, Farrer LA, Kraft P, Marazita ML, Ordovas JM, Pato CN, Spitz MR, Wagener D, Williams M, Junkins HA, Harlan WR, Ramos EM, Haines J. *Am J Epidemiol*. **174**(3):253-60 (2011).

8. Pan H, Tryka KA, Vreeman DJ, Huggins W, Phillips MJ, Mehta JP, Phillips JH, McDonald CJ, Junkins HA, Ramos EM, Hamilton CM. *Hum Mutat.* **33**(5):849-57 (2012).

9. Stafford DW.*J Thromb Haemost*. **3**(8):1873-8 (2005).

10. Booth SL. *Food Nutr Res*. **56** (2012).

11. Rieder MJ, Reiner AP, Gage BF, Nickerson DA, Eby CS, McLeod HL, Blough DK, Thummel KE, Veenstra DL, Rettie AE. *N Engl J Med*. **352**(22):2285-93 (2005).

12. Limdi NA, Wadelius M, Cavallari L, Eriksson N, Crawford DC, Lee MT, Chen CH, Motsinger-Reif A, Sagreiya H, Liu N, Wu AH, Gage BF, Jorgensen A, Pirmohamed M, Shin JG, Suarez-Kurtz G, Kimmel SE, Johnson JA, Klein TE, Wagner MJ; International Warfarin Pharmacogenetics Consortium. *Blood*. **115**(18):3827-34 (2010).

13. Nimptsch K, Nieters A, Hailer S, Wolfram G, Linseisen *J. Br J Nutr*. **101**(12):1812-20 (2009).

14. Crosier MD, Peter I, Booth SL, Bennett G, Dawson-Hughes B, Ordovas JM. *J Nutr Sci Vitaminol (Tokyo)* **55**(2):112-9 (2009).

15. http://www.cdc.gov/nchs/nhanes/nh3data.htm

16. Crawford DC, Ritchie MD, Rieder MJ. *Pharmacogenomics*. **8**(5):487-96 (2007).

17. Crawford DC, Brown-Gentry K, Rieder MJ. *PLoS One*. **5**(12):e15088 (2010).

18. Haroon Y, Bacon D, Sadowski J. Liquid-chromatographic determination of vitamin K 1 plasma with fluorometric detection. *Clin Chem* **32**:1925-9 (1986).

19. Pendergrass SA, Dudek SM, Crawford DC, Ritchie MD. *BioData Min*. **3**:10 (2010).

20. Thane CW, Bates CJ, Shearer MJ, Unadkat N, Harrington DJ, Paul AA, Prentice A, Bolton-Smith C. *Br J Nutr* **87**: 615-22 (2002).

21. Thane CW, Wang LY, Coward WA. *Br J Nutr* **96**:1116-11124 (2006).

22. Shea MK, Benjamin EJ, Dupuis J, Massaro JM, Jacques PF, D'Agostino RB, Sr, Ordovas JM, O'Donnell CJ, Dawson-Hughes B, Vasan RS, Booth SL. Genetic and non-genetic correlates of vitamins K and D. *Eur J Clin Nutr*. **63**(4):458-64. (2009).