|  |  |
| --- | --- |
| **External Collaborator Proposal** *for* **eMERGE Network Analysis**  Project/Manuscript Concept Sheet | |
| **Reference Number** | NT262 |
| **Submission Date** | 12/04/2017 |
| **Tentative Lead Investigator** *(first author with contact information and affiliation)* | Kathryn E. Hatchell, Department of Population Health Sciences, University of Wisconsin-Madison School of Medicine and Public Health, Madison, WI, USA  Contact information: [khatchgell@wisc.edu](mailto:khatchgell@wisc.edu); 920-858-4500 |
| **Tentative Senior Author**  *(last author)* | Corinne D. Engelman, Department of Population Health Sciences, University of Wisconsin-Madison School of Medicine and Public Health, Madison, WI, USA |
| **eMERGE Site Sponsor & Contact** | Marshfield Clinic  Dr. Murray Brilliant |
| **Project Title** | Development of ancestry specific polygenic risk scores and exploration of genetic-environmental interactions affecting vitamin D concentrations |
| **All Other Authors** | Dr. Alexis Frazier-Wood |
| **Other eMERGE Sites Involved** | N/A |
| **Background / Significance** | Vitamin D inadequacy affects more than 50% of adults in the United States. Low vitamin D concentrations have been associated with increased risk of autoimmune diseases, hypertension, dyslipidemia, cardiovascular events, and cardiovascular mortality (1-3). Additionally, recent Mendelian randomization studies have suggested a causal relationship between low vitamin D concentrations and increased risk of ovarian cancer, hypertension, lower cognitive function, multiple sclerosis, and all cause and cancer mortality (4-9). Furthermore, some clinical trials have shown that vitamin D and calcium supplementation are important in the prevention of fractures (10-12). Clinical trials of vitamin D alone have found that increasing vitamin D intake may lower risk of cancers, diabetes and depression (13-15). Therefore, knowledge of future risk of vitamin D inadequacy is invaluable in prevention of vitamin D associated morbidity and mortality. Clinical trials of vitamin D supplementation have shown that individual response to vitamin D supplementation is highly variable (16). Vitamin D concentrations are influenced by genetic factors and genetic variants may determine how much vitamin D intake is required to reach an optimal vitamin D blood concentration (17-20). Therefore, knowledge of the genetic determinants of vitamin D concentrations is invaluable in prevention of vitamin D associated morbidity and mortality. Several studies have uncovered vitamin D associated single nucleotide polymorphisms (SNPs) (1, 21, 22). However, these SNPs account for a small portion of the variation in vitamin D concentrations (1). Understanding the complete set of genetic factors that contribute to vitamin D concentrations and how they function together and with environmental factors will allow for personalized treatment of vitamin D inadequacy, instead of the current (and ineffective) one size fits all treatment regimen.To more fully understand the biology of vitamin D concentrations, we propose uncovering a more complete set of SNPs as well as interactions with the two vitamin D sources: sun exposure and dietary intake. |
| **Outline of Project** | **Aim 1: Use results from existing ancestry specific meta-analyses of vitamin D genome-wide association studies (GWASs) to determine the optimal p-value threshold for the set of independent SNPs that best distinguishes between adequate and inadequate vitamin D concentrations in our independent sample and calculate the polygenic risk score (PRS) for this set of SNPs.**  H1: The set of SNPs that best distinguishes between adequate and inadequate vitamin D levels will include more SNPs (from more genes) than previous GWASs.  **Aim 2: Calculate the SNP heritability of vitamin D concentration under three scenarios: 1) using Genome-Wide Complex Trait Analysis (GCTA) software for all imputed SNPs, 2) using linear regression with the set of SNPs reported in the literature and 3) using linear regression with the set of SNPs reported in the literature and included in the PRS in Aim 1.**  H2: The proportion of SNP heritability accounted for by the set of SNPs from Aim 1 will be higher than that for the SNPs currently reported in the literature.  **Aim 3: We will stratify participants by vitamin D intake and, separately, season of blood draw to examine modification of the relationship between the PRS and vitamin D concentration. If qualitative modification of the effect is observed we will test for a statistical interaction between the PRS and vitamin D intake and/or season of blood draw.**  H3a: The PRS effect on vitamin D concentration will vary by level of vitamin D intake and season; there will be a stronger PRS effect on vitamin D concentration with increased dietary intake of vitamin D and in summer.  H3b: Those with a higher PRS will require more vitamin D from diet and sun exposure to maintain optimal vitamin D concentrations. |
| **Desired Variables**  *(essential for analysis*  *indicated by* ***\*****)* | Genomic data\*  25[OH]D\*  Age\*  Sex\*  Month of blood draw\*  Race\*  BMI  Dietary vitamin D intake  Vitamin D supplement use  Physical activity |
| **Desired Data** | GWAS Imputed, Merged set (all SNPs) |
| **Planned Statistical Analyses** | Aim 1:  Results from the TRANSCEN-D ancestry specific GWAS meta-analyses performed using METAL (manuscript submitted; Hatchell and Engelman co-first author and co-senior author, respectively) will be leveraged for determination of a fuller set of SNPs and computation of the PRS in an independent sample combining participants from ARIC, MESA, FHS and WHI. This will be done by ancestry (i.e. European or African). The PRS will be calculated using the profile function in PLINK. The profile function computes a weighted risk score of independent SNPs; the weight will be determined using results from the TRANSCEN-D meta-analysis by converting the z-scores outputted from METAL into effect sizes using the deterministic relationship: = z / , p is the allele frequency for the SNP, and N is the total sample size from TRANSCEN-D. PRS will be calculated for each participant as follows: PRS*i*=, where *i* represents the participant whose PRS is calculated by summing over *n* SNPs. is the risk allele effect size from the TRANSCEN-D meta-analysis for each SNP and *C* is the individual’s count of risk alleles for that SNP. SNPs included in the PRS will be independent (r2 < 0.5; as determined by the clumping function in PLINK) and will have a p-value at or below a threshold. The threshold will be determined by iterative testing of the PRS over a range of p-values (the GWAS significance threshold of p = 5 x 10-8, p = 0.0001, p = 0.001, p = 0.01, p = 0.05, p = 0.10) to find the PRS that best discriminates between adequate and inadequate vitamin D concentrations, while controlling for age, sex, season and principal components for ancestry(23).  Aim 2:  First, we will use Genome-Wide Complex Trait Analysis (GCTA) to calculate the SNP heritability of vitamin D concentration. This will be done by calculating the variance in the phenotype jointly explained by all the SNPs; GCTA fits all inputted SNPs simultaneously as random effects in a mixed linear model (24). Next, using linear regression, we will calculate the proportion of SNP heritability that SNPs previously reported in the literature (or their proxies) account for (using R2 as the metric). We will then calculate, using linear regression, the proportion of SNP heritability accounted for by the SNPs included in the PRS computed in Aim 1, which will include the SNPs, or their proxies, previously reported in the literature (using R2 as the metric).  Specifically, GCTA uses genomic-relatedness-based restricted maximum-likelihood (GREML), similar to restricted maximum likelihood analysis (REML) and relies on the genetic relationship matrix (GRM) of SNPs from unrelated samples (25). Heritability is calculated as the proportion of phenotypic variance that is due to additive genetic effects (26). In GCTA, the GRM is calculated specific to the set of SNPs used as input. The method involves four steps: (1) calculate a linkage-disequilibrium (LD) score for each segment, (2) stratify SNPs based on their LD score, (3) compute the GRM with the stratified SNPs, and (4) perform REML analysis using the GRM (25, 27). After the GRM is generated, phenotypic variance explained by the SNPs is calculated using REML. Prior to running REML, vitamin D concentrations will be adjusted for age, gender, cohort (to remove batch effects, and to control for differing vitamin D assays used) and season (24). Models will be stratified by gender to account for the difference in variances of certain covariates (i.e. BMI) by gender; which cannot be adjusted for in the REML model. Prior to running REML, vitamin D concentration residuals from the stratified models will be standardized (mean 0, standard deviation 1). Due to the differing assays used by the cohorts being combined, standardization must be carried out to allow for combination of the datasets. Then REML is run using a linear mixed model with genetic effects as random effects. This two-stage approach is done to reduce the computation burden and to get more precise heritability estimates. This will be repeated three times: (1) with all imputed SNPs as input, (2) with only SNPs previously reported in the literature and (3) with SNPs previously reported in the literature in tandem with novel SNPs discovered in Aim 1. Expected outcomes are that SNPs currently reported in the literature account for a small fraction of the SNP heritability of vitamin D levels and that more of the SNP heritability will be captured through the PRS computed in Aim 1.  Closely related individuals pose a problem for GCTA, as GCTA calculates heritability for unrelated individuals. Prior to all analyses, one individual from each related pair (IBD > 0.38) from the non-family studies (ARIC, MESA and WHI) will be excluded and one participant per family will be included at random from FHS. Given that GCTA only uses conventionally unrelated subjects, the heritability calculated may differ from that calculated in a pedigree analysis as it is strictly genetic and does not included shared environmental effects and it might not capture all causal variants that a pedigree analysis may capture (i.e. if they were not genotyped) (24). To reduce bias that could result from using GCTA to calculate heritability, the GRM will be generated using all SNPs to capture the largest number of causal variants (24).  Aim 3:  The effect of the PRS computed in Aim 1 will be explored in a combined sample of participants from ARIC, FHS MESA and WHI stratifying by season and quartile of vitamin D intake. Analyses will be performed separately by race. This effect will be determined using a linear model that controls for: age, BMI, and season (for the vitamin D intake models) or vitamin D intake (for the season models). In accord with previous findings, it is expected that the impact of the PRS on vitamin D levels will vary by level of vitamin D intake and season; specifically, it is expected that the PRS will have a larger effect on vitamin D levels with increased dietary intake of vitamin D and in summer (i.e. with higher UV exposure) (18). We also expect that those with a higher PRS will require higher doses of vitamin D (through diet, supplements or UV exposure) to maintain adequate levels. |
| **Ethical Considerations** | None |
| **Available Funding or Resources** | Kathryn funded through: NLM training grant to the Computation and Informatics in Biology and Medicine Training Program grant NLM 5T15LM007359  Resources:   1. Social Science Computing Collaborative at UW-Madison |
| **Milestones\*\*** | December 1, 2017: Approval and data received  December 2017-January 2019: Analysis and writing  February 2019: First draft  March 2019: Second draft  April 2019: Paper submission |

***\*\**** *This section should include the timeline for completion of project, including: approval, project duration, first and second draft of the paper and submission.*

**REFERENCES**

1. Wang TJ, Zhang F, Richards JB, Kestenbaum B, van Meurs JB, Berry D, Kiel D, Streeten EA, Ohlsson C, Koller DL, Palotie L, Cooper JD, O'Reilly PF, Houston DK, Glazer NL, Vandenput L, Peacock M, Shi J, Rivadeneira F, McCarthy MI, Anneli P, de Boer IH, Mangino M, Kato B, Smyth DJ, Booth SL, Jacques PF, Burke GL, Goodarzi M, Cheung C-L, Wolf M, Rice K, Goltzman D, Hidiroglou N, Ladouceur M, Hui SL, Wareham NJ, Hocking LJ, Hart D, Arden NK, Cooper C, Malik S, Fraser WD, Hartikainen A-L, Zhai G, Macdonald H, Forouhi NG, Loos RJF, Reid DM, Hakim A, Dennison E, Liu Y, Power C, Stevens HE, Jaana L, Vasan RS, Soranzo N, Bojunga J, Psaty BM, Lorentzon M, Foroud T, Harris TB, Hofman A, Jansson J-O, Cauley JA, Uitterlinden AG, Gibson Q, Järvelin M-R, Karasik D, Siscovick DS, Econs MJ, Kritchevsky SB, Florez JC, Todd JA, Dupuis J, Hypponen E, Spector TD. Common genetic determinants of vitamin D insufficiency: a genome-wide association study. Lancet. 2010;376(9736):180-8. doi: 10.1016/S0140-6736(10)60588-0. PubMed PMID: PMC3086761.

2. Holick MF. High Prevalence of Vitamin D Inadequacy and Implications for Health. Mayo Clinic Proceedings.81(3):353-73. doi: 10.4065/81.3.353.

3. Holick MF. Vitamin D Deficiency. New England Journal of Medicine. 2007;357(3):266-81. doi: 10.1056/NEJMra070553.

4. Kunutsor SK, Burgess S, Munroe PB, Khan H. Vitamin D and high blood pressure: causal association or epiphenomenon? European Journal of Epidemiology. 2014;29(1):1-14. doi: 10.1007/s10654-013-9874-z.

5. Afzal S, Brøndum-Jacobsen P, Bojesen SE, Nordestgaard BG. Genetically low vitamin D concentrations and increased mortality: mendelian randomisation analysis in three large cohorts. BMJ. 2014;349. doi: 10.1136/bmj.g6330.

6. Mokry LE, Ross S, Ahmad OS, Forgetta V, Smith GD, Leong A, Greenwood CMT, Thanassoulis G, Richards JB. Vitamin D and Risk of Multiple Sclerosis: A Mendelian Randomization Study. PLoS Med. 2015;12(8):e1001866. doi: 10.1371/journal.pmed.1001866.

7. Ong J-S, Cuellar-Partida G, Lu Y, Fasching PA, Hein A, Burghaus S, Beckmann MW, Lambrechts D, Van Nieuwenhuysen E, Vergote I, Vanderstichele A, Anne Doherty J, Anne Rossing M, Chang-Claude J, Eilber U, Rudolph A, Wang-Gohrke S, Goodman MT, Bogdanova N, Dörk T, Dürst M, Hillemanns P, Runnebaum IB, Antonenkova N, Butzow R, Leminen A, Nevanlinna H, Pelttari LM, Edwards RP, Kelley JL, Modugno F, Moysich KB, Ness RB, Cannioto R, Høgdall E, Høgdall CK, Jensen A, Giles GG, Bruinsma F, Kjaer SK, Hildebrandt MAT, Liang D, Lu KH, Wu X, Bisogna M, Dao F, Levine DA, Cramer DW, Terry KL, Tworoger SS, Stampfer M, Missmer S, Bjorge L, Salvesen HB, Kopperud RK, Bischof K, Aben KKH, Kiemeney LA, Massuger LFAG, Brooks-Wilson A, Olson SH, McGuire V, Rothstein JH, Sieh W, Whittemore AS, Cook LS, Le ND, Gilks CB, Gronwald J, Jakubowska A, Lubiński J, Kluz T, Song H, Tyrer JP, Wentzensen N, Brinton L, Trabert B, Lissowska J, McLaughlin JR, Narod SA, Phelan C, Anton-Culver H, Ziogas A, Eccles D, Campbell I, Gayther SA, Gentry-Maharaj A, Menon U, Ramus SJ, Wu AH, Dansonka-Mieszkowska A, Kupryjanczyk J, Timorek A, Szafron L, Cunningham JM, Fridley BL, Winham SJ, Bandera EV, Poole EM, Morgan TK, Risch HA, Goode EL, Schildkraut JM, Pearce CL, Berchuck A, Pharoah PDP, Chenevix-Trench G, Gharahkhani P, Neale RE, Webb PM, MacGregor S. Association of vitamin D levels and risk of ovarian cancer: a Mendelian randomization study. International Journal of Epidemiology. 2016;45(5):1619-30. doi: 10.1093/ije/dyw207.

8. Vimaleswaran KS, Cavadino A, Berry DJ, LifeLines Cohort Study i, Jorde R, Dieffenbach AK, Lu C, Alves AC, Heerspink HJL, Tikkanen E, Eriksson J, Wong A, Mangino M, Jablonski KA, Nolte IM, Houston DK, Ahluwalia TS, van der Most PJ, Pasko D, Zgaga L, Thiering E, Vitart V, Fraser RM, Huffman JE, de Boer RA, Schöttker B, Saum K-U, McCarthy MI, Dupuis J, Herzig K-H, Sebert S, Pouta A, Laitinen J, Kleber ME, Navis G, Lorentzon M, Jameson K, Arden N, Cooper JA, Acharya J, Hardy R, Raitakari O, Ripatti S, Billings LK, Lahti J, Osmond C, Penninx BW, Rejnmark L, Lohman KK, Paternoster L, Stolk RP, Hernandez DG, Byberg L, Hagström E, Melhus H, Ingelsson E, Mellström D, Ljunggren Ö, Tzoulaki I, McLachlan S, Theodoratou E, Tiesler CMT, Jula A, Navarro P, Wright AF, Polasek O, International Consortium for Blood Pressure CfH, Aging Research in Genomic Epidemiology consortium GBPGc, Hayward C, Wilson JF, Rudan I, Salomaa V, Heinrich J, Campbell H, Price JF, Karlsson M, Lind L, Michaëlsson K, Bandinelli S, Frayling TM, Hartman CA, Sørensen TIA, Kritchevsky SB, Langdahl BL, Eriksson JG, Florez JC, Spector TD, Lehtimäki T, Kuh D, Humphries SE, Cooper C, Ohlsson C, März W, de Borst MH, Kumari M, Kivimaki M, Wang TJ, Power C, Brenner H, Grimnes G, van der Harst P, Snieder H, Hingorani AD, Pilz S, Whittaker JC, Järvelin M-R, Hyppönen E. Association of vitamin D status with arterial blood pressure and hypertension risk: a mendelian randomisation study. The lancet Diabetes & endocrinology. 2014;2(9):719-29. doi: 10.1016/S2213-8587(14)70113-5. PubMed PMID: PMC4582411.

9. Kueider AM, Tanaka T, An Y, Kitner-Triolo MH, Palchamy E, Ferrucci L, Thambisetty M. State- and trait-dependent associations of vitamin-D with brain function during aging. Neurobiology of Aging. 2016;39:38-45. doi: <http://dx.doi.org/10.1016/j.neurobiolaging.2015.11.002>.

10. Weaver CM, Alexander DD, Boushey CJ, Dawson-Hughes B, Lappe JM, LeBoff MS, Liu S, Looker AC, Wallace TC, Wang DD. Calcium plus vitamin D supplementation and risk of fractures: an updated meta-analysis from the National Osteoporosis Foundation. Osteoporosis International. 2016;27(1):367-76. doi: 10.1007/s00198-015-3386-5.

11. Bischoff-Ferrari HA, Willett WC, Wong JB, Giovannucci E, Dietrich T, Dawson-Hughes B. Fracture prevention with vitamin d supplementation: A meta-analysis of randomized controlled trials. JAMA. 2005;293(18):2257-64. doi: 10.1001/jama.293.18.2257.

12. Boonen S, Lips P, Bouillon R, Bischoff-Ferrari HA, Vanderschueren D, Haentjens P. Need for Additional Calcium to Reduce the Risk of Hip Fracture with Vitamin D Supplementation: Evidence from a Comparative Metaanalysis of Randomized Controlled Trials. The Journal of Clinical Endocrinology & Metabolism. 2007;92(4):1415-23. doi: 10.1210/jc.2006-1404.

13. Lappe JM, Travers-Gustafson D, Davies KM, Recker RR, Heaney RP. Vitamin D and calcium supplementation reduces cancer risk: results of a randomized trial. The American Journal of Clinical Nutrition. 2007;85(6):1586-91.

14. Jorde R, Sneve M, Figenschau Y, Svartberg J, Waterloo K. Effects of vitamin D supplementation on symptoms of depression in overweight and obese subjects: randomized double blind trial. Journal of Internal Medicine. 2008;264(6):599-609. doi: 10.1111/j.1365-2796.2008.02008.x.

15. Bertone-Johnson ER, Powers SI, Spangler L, Brunner RL, Michael YL, Larson JC, Millen AE, Bueche MN, Salmoirago-Blotcher E, Liu S, Wassertheil-Smoller S, Ockene JK, Ockene I, Manson JE. Vitamin D intake from foods and supplements and depressive symptoms in a diverse population of older women. The American Journal of Clinical Nutrition. 2011;94(4):1104-12. doi: 10.3945/ajcn.111.017384. PubMed PMID: PMC3173027.

16. Aloia JF, Patel M, DiMaano R, Li-Ng M, Talwar SA, Mikhail M, Pollack S, Yeh JK. Vitamin D intake to attain a desired serum 25-hydroxyvitamin D concentration. The American Journal of Clinical Nutrition. 2008;87(6):1952-8.

17. Engelman CD, Fingerlin TE, Langefeld CD, Hicks PJ, Rich SS, Wagenknecht LE, Bowden DW, Norris JM. Genetic and Environmental Determinants of 25-Hydroxyvitamin D and 1,25-Dihydroxyvitamin D Levels in Hispanic and African Americans. The Journal of Clinical Endocrinology and Metabolism. 2008;93(9):3381-8. doi: 10.1210/jc.2007-2702. PubMed PMID: PMC2567851.

18. Engelman CD, Meyers KJ, Iyengar SK, Liu Z, Karki CK, Igo RP, Truitt B, Robinson J, Sarto GE, Wallace R, Blodi BA, Klein ML, Tinker L, LeBlanc ES, Jackson RD, Song Y, Manson JE, Mares JA, Millen AE. Vitamin D Intake and Season Modify the Effects of the GC and CYP2R1 Genes on 25-Hydroxyvitamin D Concentrations. The Journal of Nutrition. 2013;143(1):17-26. doi: 10.3945/jn.112.169482. PubMed PMID: PMC3521459.

19. Nimitphong H, Saetung S, Chanprasertyotin S, Chailurkit L-o, Ongphiphadhanakul B. Changes in circulating 25-hydroxyvitamin D according to vitamin D binding protein genotypes after vitamin D3 or D2supplementation. Nutrition Journal. 2013;12(1):39. doi: 10.1186/1475-2891-12-39.

20. Wjst M. Linking vitamin D, the microbiome and allergy. Allergy. 2017;72(3):329-30. doi: 10.1111/all.13116.

21. Ahn J, Yu K, Stolzenberg-Solomon R, Simon KC, McCullough ML, Gallicchio L, Jacobs EJ, Ascherio A, Helzlsouer K, Jacobs KB, Li Q, Weinstein SJ, Purdue M, Virtamo J, Horst R, Wheeler W, Chanock S, Hunter DJ, Hayes RB, Kraft P, Albanes D. Genome-wide association study of circulating vitamin D levels. Human Molecular Genetics. 2010;19(13):2739-45. doi: 10.1093/hmg/ddq155. PubMed PMID: PMC2883344.

22. Engelman CD, Meyers KJ, Ziegler JT, Taylor KD, Palmer ND, Haffner SM, Fingerlin TE, Wagenknecht LE, Rotter JI, Bowden DW, Langefeld CD, Norris JM. Genome-wide association study of vitamin D concentrations in Hispanic Americans: The IRAS Family Study. The Journal of steroid biochemistry and molecular biology. 2010;122(4):186-92. doi: 10.1016/j.jsbmb.2010.06.013. PubMed PMID: PMC2949505.

23. Mormino EC, Sperling RA, Holmes AJ, Buckner RL, De Jager PL, Smoller JW, Sabuncu MR, Initiative FtAsDN. Polygenic risk of Alzheimer disease is associated with early- and late-life processes. Neurology. 2016. doi: 10.1212/wnl.0000000000002922.

24. Gondro C, Werf Jvd, Hayes B. Genome-wide association studies and genomic prediction2013.

25. Visscher PM, Hemani G, Vinkhuyzen AAE, Chen G-B, Lee SH, Wray NR, Goddard ME, Yang J. Statistical Power to Detect Genetic (Co)Variance of Complex Traits Using SNP Data in Unrelated Samples. PLoS Genetics. 2014;10(4):e1004269. doi: 10.1371/journal.pgen.1004269. PubMed PMID: PMC3983037.

26. Lee Sang H, Wray Naomi R, Goddard Michael E, Visscher Peter M. Estimating Missing Heritability for Disease from Genome-wide Association Studies. American Journal of Human Genetics. 2011;88(3):294-305. doi: 10.1016/j.ajhg.2011.02.002. PubMed PMID: PMC3059431.

27. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A Tool for Genome-wide Complex Trait Analysis. American Journal of Human Genetics. 2011;88(1):76-82. doi: 10.1016/j.ajhg.2010.11.011. PubMed PMID: PMC3014363.