VITAMIN D: GENETIC AND ENVIRONMENTAL PREDICTORS OF STATUS AND ASSOCIATIONS WITH PULMONARY OUTCOMES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Joyanna Gilmour Hansen

August 2013

© 2013 Joyanna Gilmour Hansen

VITAMIN D: GENETIC AND ENVIRONMENTAL PREDICTORS OF STATUS AND ASSOCIATIONS WITH PULMONARY OUTCOMES

Joyanna Gilmour Hansen

Cornell University 2013

Vitamin D, a pleiotropic hormone essential for calcium homeostasis, has generated widespread interest due to associations with numerous health outcomes. Cross-sectional studies of vitamin D and lung function reported strong, positive associations, but representative, longitudinal population-based studies are lacking, and biological mechanisms are unclear. Lung function decline is the primary characteristic of chronic obstructive pulmonary disease (COPD), the 3rd leading cause of mortality in the United States; given limited treatments to delay progression, identifying preventative approaches is critical. This work aims to elucidate determinants of vitamin D status, and investigate the role of vitamin D as a determinant of lung function.

First, we explored genetic and non-genetic determinants of serum vitamin D [25(OH)D] in African Americans. Approximately 25% of 25(OH)D variability was explained by non-genetic factors, and multivitamin supplement use was the strongest predictor. A single nucleotide polymorphism (SNP) in the vitamin D binding protein modified the effect of multivitamin supplement use on 25(OH)D. About 23% of 25(OH)D variability was estimated to be attributable to genetic

variation, with replication in a separate cohort. However, the influence of genetic ancestry made an exact estimate impossible; further exploration of genetic determinants of 25(OH)D in African Americans is needed.

Second, potential mechanisms for vitamin D—lung health associations were explored through a cross-sectional study of SNPs in 13 candidate vitamin D-responsive genes. SNPs in *SGPP2,* a phosphatase in the sphingosine-1-phosphate signaling pathway, were associated with lung function and COPD risk. Further, we identified an association between SNPs in *SGPP2* and lung-tissue specific expression of *SGPP2*. While specific mechanisms remain to be investigated, *SGPP2* is a promising vitamin D-responsive candidate gene.

Finally, associations between variants in vitamin D metabolic genes, serum 25(OH)D and lung function were explored in the Framingham Heart Study. SNPs in four vitamin D metabolic genes were associated with rate of change in FEV_1, but there was no association between 25(OH)D and rate of change in FEV_1 in the Third Generation cohort, a group of largely vitamin D sufficient middle-aged adults. Future studies should consider the influence of baseline nutritional status and underlying genetic variation on vitamin D—disease associations.

BIOGRAPHICAL SKETCH

Joyanna Gilmour Hansen grew up just outside of New York City in Bloomfield, New Jersey, with her parents, Paul and Kathy Gilmour, and her five siblings. Joyanna attended Liberty University for her undergraduate studies and graduated *summa cum laude* in 2009 with a Bachelor of Science in Biochemistry and Molecular Biology. As an undergraduate, Joyanna completed summer research in Dr. Kathryn Boor's laboratory in Cornell University's Department of Food Science, sparking her interest in pursuing graduate research. Immediately following her undergraduate work, she matriculated in Cornell's doctoral program in Human Nutrition in the Division of Nutritional Sciences.

Joyanna's goal of developing a deep understanding of associations between diet and chronic disease drew her to Dr. Pat Cassano's research group, where she ultimately focused on epidemiologic studies investigating predictors of vitamin D status and effects of vitamin D on lung health. Joyanna was particularly interested in the emerging research on genetic susceptibility to disease resulting from completion of the Human Genome Project, and integrated genetic epidemiology into her dissertation work. Joyanna has presented her research at four international scientific conferences while at Cornell, and was awarded the David Bates award in 2013 for Promising Investigation in the Field of Environmental and Occupational Health by the American Thoracic Society. As a graduate student, Joyanna pursued her interest in science communication by writing for the *Cornell Chronicle* and blogging for the American Society for

Nutrition. Joyanna is grateful to have been funded by a Cornell University Presidential Life Sciences Fellowship and by an NIH Training Grant Fellowship throughout her time at Cornell.

While at Cornell, Joyanna met and married Nate Hansen, a Cornell Ph.D. alumnus in Chemical Engineering. They have thoroughly enjoyed their experiences in Ithaca and are excited for new adventures in Portland, Oregon, where Joyanna will be pursuing a dietetic internship at the Oregon Health & Sciences University. Looking forward, Joyanna hopes to use her interdisciplinary training in nutrition, epidemiology, and genomics to contribute to public health and nutrition education.

This work is dedicated to:

- My parents, who have shown unconditional love, support, and encouragement through the frustrations and exhilarations of graduate school. I would not be who I am today without you.
- Nate, the best and most unexpected gift I found at Cornell. You inspire me every day to be true to myself, pursue my dreams, and enjoy the journey.
- God, for your grace shown to me each and every day.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Pat Cassano, for her incredible mentorship over the past four years. My development as a scientist and epidemiologist is largely due to your patient guidance and training. You have taught me so much about how to think critically, approach research from a hypothesis-oriented perspective, and clearly present my ideas. Your gracious approach to every person and situation you encounter and your high standards of excellence have truly been an inspiration to me, and your genuine care about my personal development has meant so much. I will use what I've learned at Cornell for my whole career, and will always be grateful for the tremendous impact both personally and professionally that you have had on my life.

Patsy Brannon, thank you for serving both as my committee field member and as a co-trainer for the NIH Training Grant fellowship. I always look forward to our meetings, and leave with a notepad full of new ideas to pursue; your insight into vitamin D and the DRI process has been invaluable. I am very grateful for your wise questions, listening ear, and careful advice as I considered post-graduate career options, and I am inspired by the significant impact you have had on the field of nutrition.

Andy Clark, thank you for bringing genetics to life in your fascinating Human Genetics course – I looked forward to every lecture! Thank you also for generously sharing your time and expertise to answer questions and provide new insights and perspectives throughout the course of my dissertation research.

Jason Mezey, thank you for stepping in as a replacement committee member. Your willingness to collaborate on eQTL projects added an exciting new dimension to my research, and I've learned so much from working with you and your research group.

Thanks to Rob Strawderman who served on my committee for 2 years, and whose excellent introductory statistics course provided the statistical foundation I've relied on throughout my dissertation research.

Thank you to the Cassano research group members, past and present. You have truly made my graduate experience amazing through conversations, collaborations, and celebrations. I'm so lucky to have worked with you all!

Thanks to the Cornell Statistical Consulting Unit, particularly Francoise Vermeylen and Haim Bar, who graciously answered *many* statistical questions.

Thank you to my collaborators and the study participants in the Health Aging, and Body Composition study and the Framingham Heart Study, who made this dissertation research possible.

I am very grateful to Cornell University, the Division of Nutritional Sciences, and the NIH Training Grant for financial support of this work.

Last but not least, a huge thank you to all my family and friends who have stood behind me and encouraged me throughout this whole journey. I appreciate you all more than you know!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Vitamin D: Background & Metabolism

Vitamin D is a pleiotropic hormone classically associated with calcium homeostasis and bone health (1, 2). Observational studies demonstrating associations between vitamin D status and extra-skeletal health outcomes were the motivation for an ongoing, large-scale randomized controlled trial of vitamin D supplementation in relation to cancer and cardiovascular endpoints (3). It is biologically plausible that vitamin D plays a role in multiple health outcomes, as approximately 3% of the genome is directly or indirectly responsive to the effects of this metabolite (4). Vitamin D status, assessed via the circulating biomarker 25-hydroxyvitamin D [25(OH)D], is of clinical interest both as a biomarker of *status*, which reflects total endogenous and exogenous sources of vitamin D, and as a biomarker of *exposure*, which can be studied in relation to health outcomes (1) in the search for causal associations to guide preventative action.

The primary sources of vitamin D are diet, particularly intake of vitamin D-fortified foods and supplements, and skin synthesis in response to sun exposure. Endogenous skin production of vitamin D is influenced by a number of factors including skin pigmentation, season, and latitude of residence (2, 5). Vitamin D is hydroxylated to 25(OH)D in the liver, and hydroxylated a second time to the active vitamin D metabolite, 1,25-dihydroxyvitamin D [$1,25(OH)_2D$], in the kidney. $1,25(OH)_2D$ is also synthesized locally in tissues throughout the body, including lung tissue (2, 6). In the nucleus, $1,25(OH)_2D$ forms a complex with the vitamin D receptor (VDR) and the retinoid X receptor (RXR), subsequently binding to the genome to

regulate gene transcription. Hepatic production of 25(OH)D is largely unregulated and this molecule has a long half-life, making 25(OH)D the best circulating biomarker of vitamin D status (7). Genetic variation in vitamin D metabolic genes has been associated with serum 25(OH)D in large, genome-wide association studies (GWAS) in Caucasian populations (8, 9) and family-based studies estimate that vitamin D heritability is between 29-80% (10, 11), suggesting a strong genetic influence on serum 25(OH)D levels.

Dietary Reference Intakes for Vitamin D

In response to enormous public interest and emerging research on the health benefits of vitamin D, the Institute of Medicine (IOM) conducted a comprehensive review of the evidence for benefits of vitamin D on health outcomes. The IOM review, completed in 2010, determined that a causal role for serum 25(OH)D can only be definitively established for bone health outcomes at present. Based on bone health outcome data, the IOM concluded that individuals with serum 25(OH)D concentrations <12 ng/mL are at risk of vitamin D deficiency, and individuals with serum 25(OH)D concentrations <20 ng/mL are at risk of vitamin D inadequacy. According to the IOM, "practically all persons" are sufficient at serum 25(OH)D concentrations of ≥ 20 ng/mL and above (5). However, significant controversy persists regarding optimal vitamin D status (12). Dietary Reference Intakes (DRI) of vitamin D to achieve 25(OH)D sufficiency are 600 IU/day for ages 1-70, and 800 IU day for ages 70+ (5); the upper limit is 4,000 IU for individuals greater than 9 years. Up to 1/3 of the United States population has a serum 25(OH)D concentration <20 ng/mL according to national survey data, and non-Hispanic black and Mexican-American persons were more likely to be at risk of vitamin D deficiency or risk of inadequacy compared to non-Hispanic white persons (13).

Vitamin D and Lung Function

A 2005 study demonstrated a strong, positive association between serum 25(OH)D and cross-sectional lung function in 14,000 participants of the National Health and Nutrition Examination Survey (NHANES) III. These intriguing results sparked interest in further exploration of vitamin D and lung function (14). Decreased lung function is the primary characteristic of chronic obstructive pulmonary disease (COPD), which is currently the 3rd leading cause of mortality in the United States and affects over 14 million Americans (15, 16). COPD is characterized by irreversible airflow limitation and chronic inflammation in the lung, and it is reliably diagnosed by a test of lung function, specifically the forced expiratory volume in the first second (FEV_1) and the ratio of FEV_1 to forced vital capacity (FVC) (17, 18). Although cigarette smoking is the strongest risk factor for COPD, about 15% of COPD arises in persons without obvious smoke exposure, and not all smokers develop COPD, confirming the importance of other mechanisms including genetic variation and nutritional status (19). There are few effective treatments for COPD beyond smoking cessation; thus, identifying nutritional therapies that could slow the progression or even prevent COPD is a critical public health need.

Following the NHANES study, several other cross-sectional studies demonstrated similar associations of 25(OH)D and lung function (14, 20, 21), although one study in the Hertfordshire cohort did not replicate cross-sectional associations (22). Additionally, COPD patients have a higher prevalence of vitamin D deficiency compared to healthy controls (23) and lower vitamin D status was associated with a higher risk of respiratory infections in two cohort studies (21, 24). Longitudinal evidence for vitamin D—lung function associations has been inconclusive; studies in COPD populations showed no association between baseline 25(OH)D and subsequent rate of

lung function decline (25), or baseline 25(OH)D and risk of COPD exacerbation (acute worsening of symptoms) over a 1-year follow-up (25). However, a recent study in an elderly male cohort reported steeper lung function decline in current smokers with 25(OH)D ≤ 20 ng/mL compared to smokers with higher 25(OH)D (26), and high-dose vitamin D supplementation reduced exacerbations in COPD patients with severe vitamin D deficiency (27), suggesting there are population subgroups with a greater potential to benefit from improved vitamin D status. However, all published longitudinal studies are limited because they specifically studied only smokers, COPD patients, restricted age groups, or males, and associations between vitamin D and the rate of change in lung function in representative population-based cohorts remain unclear.

Biological Mechanisms for Vitamin D—Lung Function Associations

There is strong biological plausibility for a role of vitamin D in the lung compartment (28, 29). The active vitamin D metabolite, $1,25OH_2D$, is constitutively synthesized from 25(OH)D in lung epithelial cells *in vitro* (6) and is involved in biological processes critical to lung health including inflammation and airway remodeling (28, 30, 31). Cigarette smoke and other irritants trigger the innate immune system, which is enhanced by $1,25OH_2D$-mediated expression of vitamin D-responsive genes, including the Toll-like co-receptor CD14 and the antimicrobial peptide cathelicidin (6, 32). Additionally, $1,25OH_2D$ modulates the adaptive immune response through regulation of dendritic and T cell activation and differentiation (31). Expression of several metalloproteinases, important for lung tissue remodeling, is regulated by $1,25OH_2D$ *in vitro* (33). Treatment of bronchial smooth muscle cells with $1,25(OH)_2D$ reduced the expression of metalloproteinases MMP-9 and ADAM-33 (34), which is of interest because

COPD patients have higher levels of MMP-9 compared to healthy controls (35, 36). Finally, genetic variants in the vitamin D binding protein, encoded by the *GC* gene, are associated with COPD risk (23, 37-40), and *GC* plays a role in alveolar macrophage activation (39).

Overview of Dissertation Aims

This dissertation explores 25(OH)D as a biomarker of *status* by investigating predictors of 25(OH)D in populations at high risk of inadequacy, and as a biomarker of *exposure* by studying the association of 25(OH)D with lung function outcomes. A research priority identified by the IOM is to explore the "causal role for vitamin D in non-skeletal health outcomes," (5), and the three projects presented in this dissertation contribute to filling this evidence gap. Figure 1 presents an epidemiologic framework for the dissertation research reported herein.

Study 1 investigates the determinants of circulating 25(OH)D in the Health, Aging, and Body Composition (Health ABC) cohort, an elderly African American population at high risk for vitamin D inadequacy. Environmental predictors, including diet, demographic, and physical activity variables, are assessed for associations with 25(OH)D, GWAS-associated variants from studies in Caucasian populations are evaluated for associations with 25(OH)D in African Americans, and gene by nutrient interactions are explored. A genome-wide complex trait analysis investigates the association of all genotyped SNPs jointly with 25(OH)D. Results are described in Chapter 2.

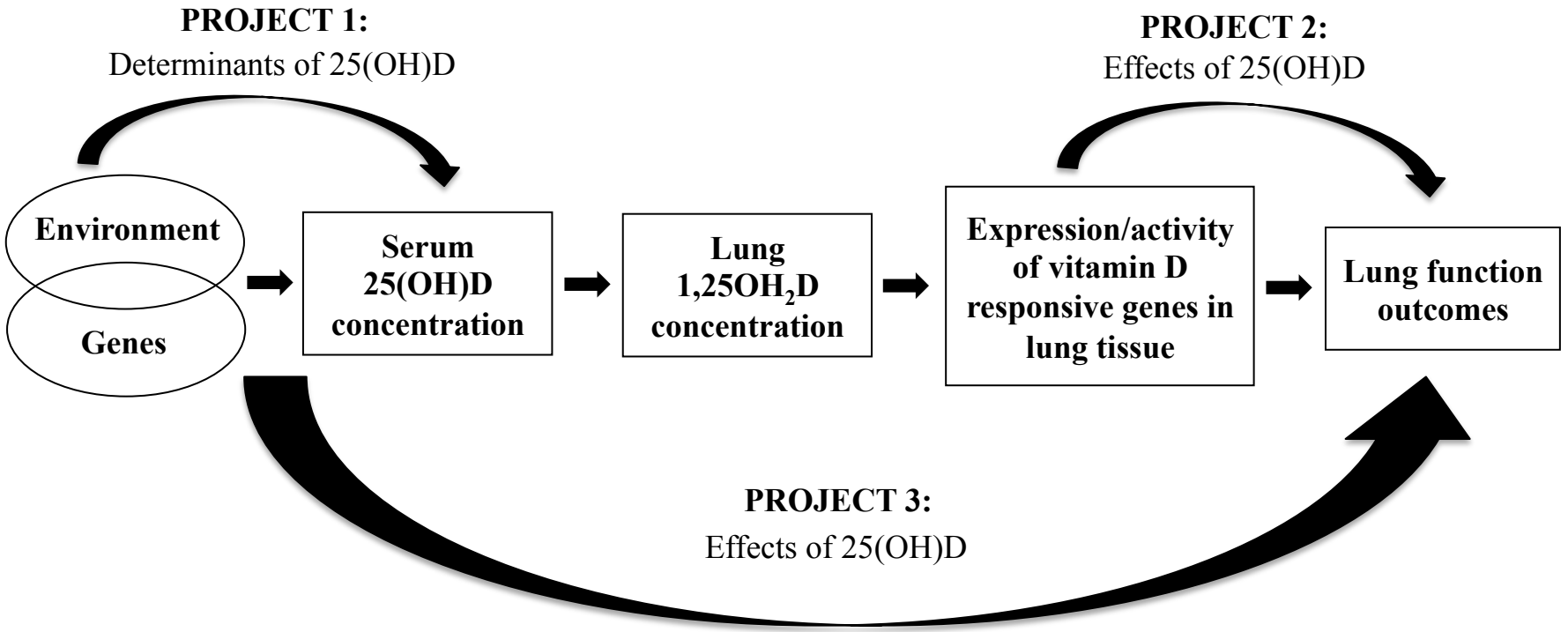
Study 2 explores potential mechanisms for the cross-sectional vitamin D—lung function associations through a hypothesis-oriented candidate gene study. This study builds on previously completed *in vivo* gene expression work, which identified genes differentially expressed in human lung epithelial cells by serum 25(OH)D status. Using a cross-sectional study design,

genetic variants in identified vitamin D-responsive genes are examined for association with lung function in the Health ABC cohort. Additionally, an expression quantitative trait loci (eQTL) analysis explores genetic variation across the candidate vitamin D-responsive genes in association with gene expression in a separate study population. Results of this project are described in Chapter 3.

Lastly, **study 3** extends the existing literature on vitamin D—lung function associations by investigating associations between vitamin D status and both cross-sectional and rate of change in lung function in the Framingham Heart Study. Both genetic variants in vitamin D metabolism genes, hypothesized to influence usual serum 25(OH)D status, and serum 25(OH)D are considered in association with lung outcomes. Given the limitations described earlier of published longitudinal vitamin D—lung function studies, this study contributes to an existing research gap by investigating this question in a large population-based family study encompassing both genders, a wide age range, and a mix of smokers and non-smokers. Results of this study are described in Chapter 4.

Overall, this dissertation research contributes to a more complete understanding of predictors of serum 25(OH)D in African American populations, and further elucidates the association of vitamin D with cross-sectional and longitudinal lung function with an exploration of potential mechanisms. While these studies were not designed to assess directly the causality of vitamin D—lung function associations, taken together they present important data that can inform the design and interpretation of future randomized clinical trials, and ultimately contribute to inferences on causality needed to make public health program and policy decisions.

Figure 1.1 Epidemiologic framework for dissertation projects



REFERENCES

1. Prentice A, Goldberg GR, Schoenmakers I. Vitamin d across the lifecycle: Physiology and biomarkers. *Am J Clin Nutr* 2008;88:500S-506S.

2. Holick MF. Vitamin d deficiency. *N Engl J Med* 2007;357:266-281.

3. Manson JE, Bassuk SS, Lee IM, Cook NR, Albert MA, Gordon D, Zaharris E, Macfadyen JG, Danielson E, Lin J, Zhang SM, Buring JE. The vitamin d and omega-3 trial (vital): Rationale and design of a large randomized controlled trial of vitamin d and marine omega-3 fatty acid supplements for the primary prevention of cancer and cardiovascular disease. *Contemp Clin Trials* 2012;33:159-171.

4. Harvey BG, Heguy A, Leopold PL, Carolan BJ, Ferris B, Crystal RG. Modification of gene expression of the small airway epithelium in response to cigarette smoking. *J Mol Med (Berl)* 2007;85:39-53.

5. Ross AC, Institute of Medicine (U. S.). Committee to Review Dietary Reference Intakes for Vitamin D and Calcium. Dietary reference intakes for calcium and vitamin d. Washington, DC: National Academies Press; 2011.

6. Hansdottir S, Monick MM, Hinde SL, Lovan N, Look DC, Hunninghake GW. Respiratory epithelial cells convert inactive vitamin d to its active form: Potential effects on host defense. *J Immunol* 2008;181:7090-7099.

7. Zerwekh JE. Blood biomarkers of vitamin d status. *Am J Clin Nutr* 2008;87:1087S-1091S.

8. Wang TJ, Zhang F, Richards JB, Kestenbaum B, van Meurs JB, Berry D, Kiel DP, Streeten EA, Ohlsson C, Koller DL, Peltonen 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 CL, Wolf M, Rice K, Goltzman D, Hidiroglou N, Ladouceur M, Wareham NJ, Hocking LJ, Hart D, Arden NK, Cooper C, Malik S, Fraser WD, Hartikainen AL, Zhai G, Macdonald HM, Forouhi NG, Loos RJ, 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 JO, Cauley JA, Uitterlinden AG, Gibson Q, Jarvelin MR, 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;376:180-188.

9. 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. *Hum Mol Genet* 2010.

10. 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* 2009;63:458-464.

11. Wjst M, Altmuller J, Braig C, Bahnweg M, Andre E. A genome-wide linkage scan for 25-oh-d(3) and 1,25-(oh)2-d3 serum levels in asthma families. *J Steroid Biochem Mol Biol* 2007;103:799-802.

12. Holick MF, Binkley NC, Bischoff-Ferrari HA, Gordon CM, Hanley DA, Heaney RP, Murad MH, Weaver CM. Evaluation, treatment, and prevention of vitamin d deficiency: An endocrine society clinical practice guideline. *J Clin Endocrinol Metab* 2011;96:1911-1930.

13. Looker AC JC, Lacher DA, et al. Vitamin d status: United states, 2001–2006. NCHS data brief, no 5. Hyattsville, MD: National Center for Health Statistic; 2011.

14. Black PN, Scragg R. Relationship between serum 25-hydroxyvitamin d and pulmonary function in the third national health and nutrition examination survey. *Chest* 2005;128:3792-3798.

15. CDC. Chronic obstructive pulmonary disease (copd). February 18, 2011 [cited 2011 April 23, 2011]. Available from: http://www.cdc.gov/nchs/fastats/copd.htm.

16. Zhang W, Duan S, Kistner EO, Bleibel WK, Huang RS, Clark TA, Chen TX, Schweitzer AC, Blume JE, Cox NJ, Dolan ME. Evaluation of genetic variation contributing to differences in gene expression between populations. *Am J Hum Genet* 2008;82:631-640.

17. Lin R, Nagai Y, Sladek R, Bastien Y, Ho J, Petrecca K, Sotiropoulou G, Diamandis EP, Hudson TJ, White JH. Expression profiling in squamous carcinoma cells reveals pleiotropic effects of vitamin d3 analog eb1089 signaling on cell proliferation, differentiation, and immune system regulation. *Mol Endocrinol* 2002;16:1243-1256.

18. Fromer L, Cooper CB. A review of the gold guidelines for the diagnosis and treatment of patients with copd. *Int J Clin Pract* 2008;62:1219-1236.

19. Eisner MD, Anthonisen N, Coultas D, Kuenzli N, Perez-Padilla R, Postma D, Romieu I, Silverman EK, Balmes JR. An official american thoracic society public policy statement: Novel risk factors and the global burden of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2010;182:693-718.

20. Sutherland ER, Goleva E, Jackson LP, Stevens AD, Leung DY. Vitamin d levels, lung function, and steroid response in adult asthma. *Am J Respir Crit Care Med* 2010;181:699-704.

21. Berry DJ, Hesketh K, Power C, Hypponen E. Vitamin d status has a linear association with seasonal infections and lung function in british adults. *Br J Nutr* 2011;106:1433-1440.

22. Shaheen SO, Jameson KA, Robinson SM, Boucher BJ, Syddall HE, Aihie Sayer A, Cooper C, Holloway JW, Dennison EM. Relationship of vitamin d status to adult lung function and copd. *Thorax* 2011;66:692-698.

23. Janssens W, Bouillon R, Claes B, Carremans C, Lehouck A, Buysschaert I, Coolen J, Mathieu C, Decramer M, Lambrechts D. Vitamin d deficiency is highly prevalent in copd and correlates with variants in the vitamin d-binding gene. *Thorax* 2010;65:215-220.

24. Ginde AA, Mansbach JM, Camargo CA, Jr. Association between serum 25-hydroxyvitamin d level and upper respiratory tract infection in the third national health and nutrition examination survey. *Arch Intern Med* 2009;169:384-390.

25. Kunisaki KM, Niewoehner DE, Connett JE. Vitamin d levels and risk of acute exacerbations of chronic obstructive pulmonary disease: A prospective cohort study. *Am J Respir Crit Care Med* 2012;185:286-290.

26. Lange NE, Sparrow D, Vokonas P, Litonjua AA. Vitamin d deficiency, smoking, and lung function in the normative aging study. *Am J Respir Crit Care Med* 2012;186:616-621.

27. Lehouck A, Mathieu C, Carremans C, Baeke F, Verhaegen J, Van Eldere J, Decallonne B, Bouillon R, Decramer M, Janssens W. High doses of vitamin d to reduce exacerbations in chronic obstructive pulmonary disease: A randomized trial. *Ann Intern Med* 2012;156:105-114.

28. Pfeffer PE, Hawrylowicz CM. Vitamin d and lung disease. *Thorax* 2012;67:1018-1020.

29. Litonjua AA. Vitamin d and the lung : Mechanisms and disease associations. New York: Humana Press; 2012.

30. Bikle DD. Vitamin d regulation of immune function. *Vitam Horm* 2011;86:1-21.

31. Hansdottir S, Monick MM. Vitamin d effects on lung immunity and respiratory diseases. *Vitam Horm* 2011;86:217-237.

32. Liu PT, Stenger S, Li H, Wenzel L, Tan BH, Krutzik SR, Ochoa MT, Schauber J, Wu K, Meinken C, Kamen DL, Wagner M, Bals R, Steinmeyer A, Zugel U, Gallo RL, Eisenberg D, Hewison M, Hollis BW, Adams JS, Bloom BR, Modlin RL. Toll-like receptor triggering of a vitamin d-mediated human antimicrobial response. *Science* 2006;311:1770-1773.

33. Wang TT, Tavera-Mendoza LE, Laperriere D, Libby E, MacLeod NB, Nagai Y, Bourdeau V, Konstorum A, Lallemant B, Zhang R, Mader S, White JH. Large-scale in silico and microarray-based identification of direct 1,25-dihydroxyvitamin d3 target genes. *Mol Endocrinol* 2005;19:2685-2695.

34. Song Y, Qi H, Wu C. Effect of 1,25-(oh)2d3 (a vitamin d analogue) on passively sensitized human airway smooth muscle cells. *Respirology* 2007;12:486-494.

35. Shea MK, Houston DK, Tooze JA, Davis CC, Johnson MA, Hausman DB, Cauley JA, Bauer DC, Tylavsky F, Harris TB, Kritchevsky SB. Correlates and prevalence of insufficient 25-hydroxyvitamin d status in black and white older adults: The health, aging and body composition study. *J Am Geriatr Soc* 2011;59:1165-1174.

36. Beeh KM, Beier J, Kornmann O, Buhl R. Sputum matrix metalloproteinase-9, tissue inhibitor of metalloprotinease-1, and their molar ratio in patients with chronic obstructive pulmonary disease, idiopathic pulmonary fibrosis and healthy subjects. *Respir Med* 2003;97:634-639.

37. Ito I, Nagai S, Hoshino Y, Muro S, Hirai T, Tsukino M, Mishima M. Risk and severity of copd is associated with the group-specific component of serum globulin 1f allele. *Chest* 2004;125:63-70.

38. Schellenberg D, Pare PD, Weir TD, Spinelli JJ, Walker BA, Sandford AJ. Vitamin d binding protein variants and the risk of copd. *Am J Respir Crit Care Med* 1998;157:957-961.

39. Wood AM, Bassford C, Webster D, Newby P, Rajesh P, Stockley RA, Thickett DR. Vitamin d-binding protein contributes to copd by activation of alveolar macrophages. *Thorax* 2011.

40. Costa JL, Eijk PP, van de Wiel MA, ten Berge D, Schmitt F, Narvaez CJ, Welsh J, Ylstra B. Anti-proliferative action of vitamin d in mcf7 is still active after sirna-vdr knock-down. *BMC Genomics* 2009;10:499.

CHAPTER 2

GENETIC AND ENVIRONMENTAL PREDICTORS OF SERUM 25-HYDROXYVITAMIN D IN AFRICAN AMERICANS IN THE HEALTH, AGING, AND BODY COMPOSITION STUDY

JG Hansen[1], W Tang[1], KC Hootman[1], PM Brannon[1], DK Houston[2], SB Kritchevsky[2], T Harris[3], M Garcia[3], K Lohman[4], Y Liu[5] , IH de Boer[6, 7], BR Kestenbaum[6,7], C Robinson-Cohen[7], DS Siscovick[8], and PA Cassano[1,9]

[1]Division of Nutritional Sciences, Cornell University, Ithaca, NY 14853 USA

[2]Sticht Center on Aging, Wake Forest School of Medicine, Winston-Salem, NC 27157 USA

[3]Intramural Research Program, National Institute on Aging, Laboratory of Epidemiology, Demography, and Biometry, Gateway Building, 3C309, 7201 Wisconsin Avenue, Bethesda, MD 20892 USA

[4]Department of Biostatistical Sciences, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC 27157 USA

[5]Department of Epidemiology & Prevention, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC 27157 USA

[6]Division of Nephrology, Department of Medicine, University of Washington, Seattle, WA 98195 USA

[7]Kidney Research Institute, University of Washington, Seattle, WA 98104 USA

[8]Cardiovascular Health Research Unit, Departments of Medicine and Epidemiology, University of Washington, Seattle, WA

[9]Department of Public Health, Division of Biostatistics and Epidemiology, Weill Cornell Medical College, New York, NY, USA

Address correspondence to: Dr. Patricia A. Cassano, Division of Nutritional Sciences, Cornell University, 209 Savage Hall, Ithaca, NY 14853, pac6@cornell.edu, (607) 255-7551 (ph), (607) 255-1033 (fax).

Keywords: Vitamin D, 25-Hydroxyvitamin D, supplements, *GC,* GCTA, rs7041

ABSTRACT

Background Low circulating 25-hydroxyvitamin D [25(OH)D] is prevalent in African Americans, but predictors of vitamin D status are understudied compared to Caucasian populations.

Objective We investigated environmental and genetic predictors of circulating 25(OH)D in a population of approximately 1,000 elderly African Americans participating in the Health, Aging, and Body Composition (Health ABC) study.

Design Regression analysis estimated the cross-sectional association of non-genetic (environmental) predictors with 25(OH)D. Single nucleotide polymorphisms (SNPs) associated with 25(OH)D in Caucasian genome-wide association studies (GWAS) were analyzed for association with serum 25(OH)D. Genome-Wide Complex Trait Analysis (GCTA) evaluated the association of all genotyped SNPs with serum 25(OH)D in Health ABC with replication in a separate cohort.

Results Gender, study site, season of blood draw, body mass index, dietary supplement use, dairy and cereal consumption, Healthy Eating Index score, and walking > 180 minutes/week were associated with 25(OH)D at $P<0.05$, jointly explaining 25% of the variation in circulating 25(OH)D. Up to 23% (95% CI: 0-52%) of phenotypic variation was estimated to be explained by total additive genetic variation, and this finding was replicated in a separate cohort. Although GWAS-identified SNPs from studies of Caucasians were not replicated in Health ABC African Americans, a gene x nutrient interaction was identified: the *TT* genotype of rs7041, a non-synonymous SNP in *GC*, increased the odds of vitamin D insufficiency in multivitamin supplement users.

Conclusion Modifiable dietary and lifestyle predictors of serum 25(OH)D were identified in African Americans, and a gene x environment interaction was identified between the most significant predictor, multivitamin use, and the rs7041 genotype.

INTRODUCTION

In addition to well-known roles in calcium absorption and skeletal outcomes, vitamin D regulates over 900 genes involved in physiological functions throughout the body (1, 2). Serum 25(OH)D, the major circulating biomarker of vitamin D status (1), is converted to active vitamin D, 1,25OH_2D, primarily in the kidney, but also in extra-renal tissues throughout the body (3). Serum 25(OH)D is derived from dietary intake (food or supplements) and skin exposure to ultraviolet radiation (1, 4).

Approximately 30% of Americans are at risk of inadequate or deficient serum 25(OH)D, according to a recent NHANES report (5). Both African American and elderly populations are at high risk for vitamin D deficiency, due in part to a lower capacity for endogenous synthesis of vitamin D from sunlight (4, 6). In NHANES, 32% of non-Hispanic black adults were at risk of vitamin D deficiency [serum 25(OH)D <12 ng/mL] in comparison to only 3% of non-Hispanic white adults (5).

While genetic and non-genetic predictors of 25(OH)D have been well-described in Caucasian populations (7-13), fewer studies have focused specifically on determinants of 25(OH)D in African Americans. Several modifiable predictors of serum 25(OH)D have been identified in African Americans, including intake of vitamin D-containing foods and supplements, sun exposure, and body mass index (BMI) (10, 14-20). Vitamin D status is heritable, and heritability estimates from twin and family-based studies range from 28-80%; while most estimates derive from Caucasian populations (21-24), a study in 42 African American families reported a heritability coefficient of 28% (16). Genome-wide association studies (GWAS) in Caucasians have identified genetic predictors of vitamin D status that explain between 1 and 4% of phenotypic variability (7, 8). In African Americans, only one published population-based

candidate gene study, limited by small sample size, investigated genetic predictors of vitamin D status (25).

To date, no studies have investigated the respective contribution of both genetic and non-genetic predictors to variability in serum 25(OH)D in elderly African Americans. Given that this population is at significant risk of vitamin D inadequacy, it is important to understand the relative contributions of modifiable and non-modifiable predictors of serum 25(OH)D. For this study, we hypothesized that both environmental and genetic factors contribute to circulating serum 25(OH)D status in an elderly African American population, and our objective was to estimate the variability explained by genes and environment, respectively.

SUBJECTS AND METHODS

Study Population

The primary analyses were in the Health, Aging, and Body Composition (Health ABC) cohort, which comprises 3,075 participants recruited between April 1997 and June 1998, aged 70-79 at baseline and selected as a random sample of Whites and all Black Medicare-eligible residents of ZIP codes in and around Memphis, TN and Pittsburgh, PA. Eligibility criteria included the ability to walk one-quarter of a mile, climb 10 stairs, and perform activities of daily living without difficulty. Additionally, eligible participants were required to be free of life-threatening disease with the intent to stay in the area for 3 or more years(19). The Institutional Review Boards (IRB) at the University of Memphis, Tennessee, and the University of Pittsburgh granted approval to conduct the Health ABC Study, and all participants provided written informed consent. The Cornell University Committee on Human Subjects approved the study reported herein.

Health ABC comprised 1,281 African American participants. For the current study, those who did not return for the 12-month follow-up exam (N=46) were excluded due to missing data on both serum 25(OH)D and dietary intake. Additional exclusion criteria include missing a serum 25(OH)D measurement (N=126), abnormally high serum 25(OH)D (defined as 25(OH)D ≥ 150 ng/mL; N=1), and end-stage kidney failure (defined as glomerular filtration rate <15; N=3). Participants missing key dietary data (N=116) were excluded from non-genetic analyses, leaving a total of 989 participants. 980 participants had genotype and serum 25(OH)D data, comprising the sample for genetic analyses.

Data Collection

Participant data on gender, education, smoking status, and other covariates were collected from a baseline survey administered by trained interviewers. Body Mass Index (BMI) and physical activity, assessed via self-report of minutes spent walking each week, were obtained from data collected at the 12-month follow-up visit.

Trained interviewers assessed dietary intake at the 12-month follow-up visit, using a Block food-frequency questionnaire (FFQ) modified for the Health ABC Study by Block Dietary Data Systems (Berkeley, CA). Nutrient intakes and daily servings of food groups were estimated, and food group information was used to calculate a Healthy Eating Index (HEI) score ranging from 0-100 for each individual (26). The HEI estimates how well each participant's diet matches U.S. Dietary Guidelines recommendations; Health ABC HEI scores were calculated based on compliance with the 1992 USDA Food Guide Pyramid (26, 27) and were grouped into "good" (HEI score >81), "needs improvement" (51-80), and "poor" (<51) (26). Dietary supplement use was also assessed at the 12-month follow-up visit. Further details are provided elsewhere (19, 26).

In Health ABC, serum 25(OH)D was measured in fasting blood samples collected at the 12 month follow-up visit. A two-step radioimmunoassay kit was used to measure 25(OH)D concentrations (25-HydroxyvitaminD 125I RIA Kit, DiaSorin, Stillwater, MN), with an interassay coefficient of variation of 6.78% (19). Season of blood draw was defined as winter (December-February), spring (March-May), summer (June-August), and fall (September-November).

The Illumina Human 1M-Duo custom chip was used for genotyping in Health ABC; race-specific genotype imputation was performed with MACH version 1.0.16, using reference panel data from HapMap release 22 Build 36 (28). Studied SNPs were required to have minor allele frequency (MAF) >1%, Hardy-Weinberg p-values > $1x10^{-6}$, imputation quality scores greater than 0.3, and call rates > 95%.

Statistical Analysis

Bivariate regression analysis explored associations of demographic, dietary, and environmental predictors hypothesized to contribute to variation in serum 25(OH)D. Predictors associated with log-transformed 25(OH)D at $P<0.05$ were further evaluated in multivariate models to determine a final set of variables jointly associated with log-transformed 25(OH)D. All multivariate models were adjusted for age, gender, study site, and season of blood draw.

Normality plots revealed a slightly right-skewed distribution of serum 25(OH)D. However, multivariate model results were equivalent for natural log-transformed and untransformed serum 25(OH)D phenotypes; significant predictor variables were the same, residuals from final models were approximately normally distributed, and model R^2 differed by <1% between the two models. Thus, regression results from modeling the untransformed serum 25(OH)D outcome are presented for ease of interpretation.

Single SNPs previously associated with 25(OH)D in GWAS of Caucasians (7, 8) (Online Supplement for details of candidate SNP selection) were tested for associations with serum 25(OH)D in ordinary least squares linear regression models adjusted for age, gender, study site, season of blood draw, and ancestry principal components. Given results for log-transformed versus untransformed serum 25(OH)D phenotypes did not differ significantly, genetic analysis results are presented for the untransformed serum 25(OH)D phenotype for ease of interpretation. Models estimating genetic—serum 25(OH)D associations were adjusted for age, gender, season, site, and principal components only for two reasons: to maximize sample size, and to avoid adjusting for potential intermediate variables on the causal pathway between exposure (genetics) and outcome [serum 25(OH)D].

SAS version 9.3 (SAS Institute, Inc., Cary, North Carolina) was used for all genetic and non-genetic regression analyses, and all statistical tests were two-sided.

In analyses estimating the overall genetic contribution to serum 25(OH)D variability, all genotyped SNPs with a MAF >1% were tested jointly for association with serum 25(OH)D. The analysis used genome-wide complex trait analysis (GCTA) software, which is based on a linear mixed model approach (v. 1.04; for details of the method see (29)). GCTA uses the restricted maximum likelihood (REML) method to fit a linear mixed model estimating the variance in serum 25(OH)D explained by additive genetic variation. 91 distantly related Health ABC participants (genetic relatedness score >0.05) were removed from the data set prior to analysis, leaving a total of 889 participants for the GCTA analysis. The GCTA analysis was replicated in 1,198 African Americans in MESA, after excluding 136 distantly related participants. Models in both cohorts were adjusted for age, gender, study site, season of blood draw, and were run with and without further control for ancestry principal components; the log-transformed serum

25(OH)D phenotype was used for comparability across cohorts. GCTA estimates from the two cohorts were meta-analyzed with METAL (30), using an inverse-variance weighted meta-analysis model.

RESULTS

The mean serum 25(OH)D concentration in Health ABC was 20.7 ng/mL (**Table 2.1**), and the prevalence of Health ABC participants potentially at risk for 25(OH)D inadequacy (defined as serum 25(OH)D < 20 ng/mL) was 55%. Serum vitamin D characteristics in the MESA replication cohort were nearly identical (mean 19.0 ng/mL and prevalence of participants potentially at risk for vitamin D inadequacy was 60%; assayed by HPLC-tandem mass spectrometry, Waters Xevo TQ mass spectrometer, Milford, MA; further details provided elsewhere (31)). Health ABC participant characteristics are further described in **Table 2.1**.

Environmental Predictors of Circulating 25(OH)D

Twenty-five percent of the variation in serum 25(OH)D in Health ABC African Americans was explained by a single multivariate model (**Table 2.2**), including predictor variables significantly associated with 25(OH)D at $P<0.05$ in single variable models. Although age had little or no association with serum 25(OH)D, male gender, residence in Memphis, and summer season of blood draw were all positively associated with vitamin D status. BMI had a non-linear association with serum 25(OH)D such that the strongest inverse association of BMI with 25(OH)D status was observed in obese individuals.

The strongest predictor of serum 25(OH)D status in the Health ABC population was multivitamin supplement use, which accounted for about 8% of the variability; the use of either vitamin D or calcium supplements also made significant contributions to the model. Both cereal

consumption and dairy consumption were significantly associated with serum vitamin D; furthermore, participants categorized as good on the HEI had about 3 ng/mL higher serum 25(OH)D concentration compared to participants categorized as poor. Overall physical activity had little or no association with serum 25(OH)D, but walking briskly more than 180 minutes/week was positively associated with serum status. Vitamin D intake (micronutrient data estimated from FFQ) was associated with serum 25(OH)D in bivariate models, but the association of this variable was captured by other dietary variables (e.g. dairy intake); thus, vitamin D intake did not remain significantly associated with serum 25(OH)D in multivariate models. We did not observe a significant association of smoking status with continuous serum 25(OH)D.

In a sensitivity analysis, 145 individuals at risk of vitamin D deficiency (defined as serum 25(OH)D < 12 ng/mL) were removed from the analysis to explore the extent to which results were driven by individuals. Season of blood draw, cereal consumption, HEI category, and walking more than 180 minutes/week were no longer statistically significantly associated with serum 25(OH)D in multivariate models. Other model coefficients were somewhat attenuated although all were in the same direction and remained statistically significant (data not shown), and the model explained 21% of the variation in serum 25(OH)D.

Genetic Predictors of Serum 25(OH)D

Two recent GWAS of serum 25(OH)D in Caucasians identified genome-wide significant single nucleotide polymorphisms (SNPs) in or near *CYP2R1, GC,* and *DHCR7/NADSYN1*; we investigated the most significant SNPs from each study (total of 10 SNPs, only 8 of which were available in Health ABC data) in relation to serum 25(OH)D in Health ABC African Americans. None of the SNPs identified in past GWAS of Caucasians were associated with serum 25(OH)D

in this population at a statistical significance level of $P<0.05$ (**Table 2.3**), but the rs7041 SNP in *GC* was near the significance threshold ($P=0.08$).

Given the borderline statistically significant association of rs7041, a functional, nonsynonymous SNP in *GC,* with serum 25(OH)D, we explored gene x nutrient interactions between rs7041 and multivitamin supplement use, the strongest non-genetic predictor of serum 25(OH)D status in Health ABC African-Americans. Other GWAS-associated SNPs, including the two additional SNPs in *GC,* have no known functional role and showed little to no association with the 25(OH)D phenotype in Health ABC African-Americans; thus, these SNPs were not considered in gene x nutrient analyses. Due to the low prevalence of the rs7041 *G* minor allele, the SNP was modeled as homozygous dominant (*TT* genotype versus *GT/GG* genotype). The mean serum 25(OH)D was 3.7 ng/mL higher in multivitamin supplement users with the rs7041 *GG/GT* genotype compared to supplement users with the *TT* genotype, while there was little difference in mean 25(OH)D by genotype among non-supplement users (**Figure 2.1**). We further explored whether the odds of vitamin D insufficiency (defined as serum 25(OH)D <20 ng/mL) associated with the rs7041 genotype varied by multivitamin supplement use, and the SNP by multivitamin use regression coefficient was statistically significant ($P_{Interaction}=0.0073$). Among multivitamin supplement users (N=237), the odds of vitamin D insufficiency was nearly 4 times higher in the *TT* genotype group, compared to participants with a *GG* or *GT* genotype (OR=3.8, 95% CI: 1.7, 8.5; model adjusted for age, gender, study site, season of blood draw, and ancestry principal components). In non-multivitamin supplement users (N=743) there was little or no association of genotype with the odds of vitamin D insufficiency, but the mean serum 25(OH)D in this group was in the insufficient category (19.0 ± 8.3 ng/mL).

Taking a broader approach, novel methods that estimate the genome-wide additive genetic variation in a phenotype were applied to this study of the serum 25(OH)D phenotype in African Americans. Initial models estimated that about 25% (95% CI: 0%, 74%) of the serum 25(OH)D variance is attributed to additive genetic variation in Health ABC (**Table 2.4**); these models do not include adjustments for principal components due to concerns about over-adjusting for population of origin effects that could be proxies for skin color and hence UV absorption. In models adjusting for population substructure using the first two principal components, the estimate was reduced to a near-null value of about 0.6% (95% CI 0%, 65%).

We replicated the GCTA findings in a second cohort of African Americans from the Multiethnic Study of Atherosclerosis (MESA). In MESA, about 21% of serum 25(OH)D variance is attributed to genetic variation (95% CI: 0%, 59%). Similar to Health ABC findings, in models adjusting for population substructure using the first ten principal components the GCTA estimate was reduced to 0%.

Meta-analysis of the findings from the two cohorts led to an estimate of 23% for the variance in serum 25(OH)D explained by additive genetic variation (95% CI: 0%, 52%), after adjustment for age, gender, study site, and season of blood draw (**Table 2.4).**

DISCUSSION

To our knowledge, this is the first study to examine comprehensively both environmental and genetic predictors of 25(OH)D in elderly African Americans. About 25% of the total variability in serum 25(OH)D was explained by environmental determinants. Using a novel analytic approach, we estimated that up to 25% of serum 25(OH)D variability is explained by additive genetic variation in Health ABC African Americans, and replication in the MESA

African American cohort confirmed this estimate, although we were unable to disentangle the influence of population substructure from direct genetic influences on vitamin D nor directly consider the joint effects of genes and environment.

Multivitamin supplement use was the most significant non-genetic predictor of serum 25(OH)D, and supplement users had 6.3 ng/mL higher serum 25(OH)D concentration compared to non-users. GWAS-identified SNPs predictive of serum 25(OH)D in Caucasians were not replicated in Health ABC African American participants. However, rs7041, a GWAS-identified non-synonymous SNP in *GC,* was borderline associated with 25(OH)D and modified the serum response to multivitamin supplementation such that the odds of vitamin D insufficiency were significantly higher in supplement users with the *TT* genotype.

Consideration of non-genetic predictors in the current study builds on a prior study that considered predictors of vitamin D insufficiency in African Americans (19); we found that predictors previously shown to be associated with insufficiency had associations over the full serum range of vitamin D concentrations. We identified additional dietary predictors of circulating 25(OH)D, including frequency of cereal and dairy consumption, suggesting that vitamin D-fortified dairy foods and breakfast cereals may be an important contributor to serum 25(OH)D status when consumed regularly. Beyond multivitamin and vitamin D supplement use, we found that use of calcium supplements was associated with serum 25(OH)D. This is consistent with the hypothesis that calcium supplementation contributes to the maintenance of adequate serum calcium, thereby preventing an increase in PTH and the subsequent renal conversion of 25(OH)D to $1,25(OH)_2D$ and increased activation of the vitamin D-degrading enzyme *CYP24A1* (32, 33). Higher BMI was associated with lower serum 25(OH)D, with evidence for non-linearity (steeper inverse associations at higher BMI); this association is

hypothesized to be due to an increased storage of vitamin D in adipose tissue (34, 35). Although in Caucasians aging skin is associated with a decreased ability to synthesize vitamin D (36), age was not significantly associated with 25(OH)D in our population, likely reflecting the limited 9-year age range of Health ABC participants. We did not identify an association between continuous serum 25(OH)D and smoking status, but a prior Health ABC study found that current smokers had higher odds of 25(OH)D < 20 ng/mL (19), suggesting smoking may be associated primarily with risk of vitamin D inadequacy. The association of season of blood draw with serum vitamin D primarily reflected associations with risk of vitamin D deficiency, and as expected, serum 25(OH)D levels were lowest during the winter months (Dec.-Feb.). Finally, physical activity showed a limited association with 25(OH)D compared to study site and season of blood draw, suggesting that the latter two variables may be better proxies for sun exposure in this elderly population.

Although past genotype—serum 25(OH)D associations demonstrated in European-Americans did not reach significance thresholds in this study of Health ABC African Americans, we demonstrated that rs7041 genotype significantly increases the odds of vitamin D insufficiency among multivitamin supplement users. This suggests that rs7041 genotype modifies the serum response to supplementation, and that individuals with the rs7041 *TT* genotype may need higher supplement doses to achieve vitamin D sufficiency, although the lack of dosage data in Health ABC precludes definitive conclusions. The lack of association of rs7041 genotype and serum 25(OH)D in non-supplement users likely reflects the fact that 63% of these participants were already vitamin D insufficient, thus there was a limited range of vitamin D status. Rs7041 is a non-synonymous SNP in *GC* that causes an amino acid change from aspartic acid to glutamic acid, and was previously associated with lower serum 25(OH)D (*T* allele) in

both African Americans (16) and Caucasians (37-40). The rs7041 *T* allele is more common in ancestral African populations (41); the frequency of the *T* allele was 42% and 90% in HapMap CEU and YRI populations, respectively (42), and 82% in Health ABC African Americans. The higher prevalence of the rs7041 *T* allele in African ancestry populations may contribute to the overall lower serum 25(OH)D status in these populations.

We demonstrated that in this Health ABC population, approximately 25% of the variation in serum 25(OH)D is estimated to be due to additive genetic variation, with replication in the MESA population producing a similar estimate of 21%. We meta-analyzed the results from both cohorts to obtain a final estimate of 23% of 25(OH)D variation due to additive genetic factors.

The GCTA analysis was performed both with and without adjustment for ancestry principal components; when adjusting for ancestry, the GCTA estimate was essentially null. Skin pigmentation affects endogenous skin synthesis of 25(OH)D (43), and ancestry has been shown to correlate strongly with skin pigmentation in African Americans (44). Furthermore, a recent study showed an association between African ancestry, calculated from 276 ancestry informative markers, and serum 25(OH)D (45); thus, including principal components as model covariates may be an over-adjustment for traits that co-vary with population ancestry, such as skin pigmentation. Following this reasoning, the true estimate of direct genetic influences on serum 25(OH)D is likely to be less than 23%, but we were unable to arrive at a more valid estimate. These results are consistent with a recent GCTA meta-analysis in Caucasians, which estimated that 9% (95% CI: 0-22%) of 25(OH)D variation was attributable to additive genetic variation (11). Interestingly, this estimate was adjusted for principal components, possibly reflecting a weaker influence of population substructure on GCTA estimates in non-African populations (11).

Although we demonstrated a consistent estimate of the effect of genetic variation on serum 25(OH)D in two independent African American cohorts, we did not observe significant associations between GWAS-associated SNPs and 25(OH)D. SNP frequencies vary by ancestry (**Supplemental Table 2.5**), limiting the possibility that specific genetic associations will be replicable across races. Furthermore, candidate genes associated with 25(OH)D in African Americans may differ from those in Caucasians, reflecting divergent genetic adaptations to ancestral environments. For instance, genes related to skin pigmentation may be more strongly linked to serum 25(OH)D levels in African Americans than genes related to vitamin D metabolism. Considering patterns of genetic variation more broadly, Africans are typically more genetically diverse than Caucasians; thus, African ancestry populations have more rare SNPs, lower levels of linkage disequilibrium (LD), and shorter haplotype blocks compared to non-Africans (46), and larger sample sizes may be needed to demonstrate associations. In Health ABC, the *GC* gene had lower levels of LD compared to European-Americans (**Supplemental Figure 1**), which may explain the lack of association of GWAS-associated SNPs in this gene (7, 8). In summary, our estimate of the amount of variability in 25(OH)D attributed to genetic variation is not inconsistent with a true effect, and highlights the need to further study the genetic architecture of 25(OH)D in African Americans in larger cohorts.

There are a number of significant strengths in this study, which included a large sample size of elderly African Americans to investigate genetic and non-genetic predictors of 25(OH)D. Extensive data on diet and supplement use, in addition to genotype data, allowed for hypothesis-driven investigations of genetic and environmental predictors of vitamin D. Although supplement use data were available, dosage data were limited, which precluded consideration of dose-response associations. Neither were there direct data on sun exposure, sunscreen use, or

outdoor physical activity, although study site, season of blood draw, and time spent walking were proxies for UV exposure. Given that Health ABC dietary assessment was performed in 1998-99, the older version of the Healthy Eating Index was used for analysis, and re-analysis of the data with a newer HEI version was not technically feasible. Although we explained a significant amount of variation in serum 25(OH)D, there was unexplained variability in 25(OH)D that could be attributable to differences in sun exposure or other unmeasured variables, or to gene x environment interactions that were not considered.

In conclusion, we identified several modifiable factors including diet and supplement use that explain variability in serum 25(OH)D in an elderly African American population. Additionally, the identification of a common genotype that affects serum 25(OH)D response to multivitamin supplements and risk of vitamin D insufficiency in multivitamin supplement users can inform ongoing and future clinical trials of vitamin D supplementation. Finally, we utilized a novel genetic analysis to estimate that up to 23% of the variability in 25(OH)D can be attributed to additive genetic variation, supporting further studies in African Americans to elucidate the causal genetic variants associated with serum 25(OH)D.

Table 2.1 Population Characteristics of African American (N=989) participants of the Health ABC Cohort*

Variable	Mean (SD) or %
Serum 25(OH)D (ng/mL)	20.7 (9.0)
Age, years	74.5 (2.9)
Gender (% female)	57.3
Site (% Pittsburgh)	55.2
Season of blood draw (%)	–
Winter	23.6
Spring	31.9
Summer	17.4
Fall	27.2
Current smokers (%)	14.6
BMI	28.6 (5.5)
BMI Category (%)	–
<25	26.1
25-30	38.7
>30	35.2
Dietary vitamin D intake, IU/d	197.3 (143.2)
Dietary calcium intake, mg/d	768.9 (413.2)
Vitamin D supplement (%)	5.8
Multivitamin (%)	23.9
Calcium supplement (%)	10.9
Dairy Consumption (%)	–
No dairy	32.3
1-3 servings/day >3	63.6
servings per day	4.2
Cereal Consumption (%)	–
No cereal	12.6
1-4 times/month	32.9
>1/week	54.5
Healthy Eating Score (%)	–
Poor	10.7
Needs Improvement	76.2
Good	13.0
Minutes walking/week	106.9 (222.4)
Walks >180 min/week (%)	3.3

*Population characteristics for the 980 participants in the genetic analyses are nearly identical, and thus are not presented here.

Table 2.2 Regression coefficients for predictors of serum 25(OH)D in Health ABC African Americans (N=989)

Variable:	β *	95% Confidence Interval	*P* **	R^2†
Age	0.03	-0.1, 0.2	0.70	0.01%
Gender (Male)	2.2	1.2, 3.3	<0.0001	1.3%
Site (Memphis)	1.7	0.7, 2.7	0.0010	0.8%
Season	–	–	0.0229	0.7%
Summer	Referent			
Winter	-2.0	-3.6, -0.5	–	–
Spring	-0.7	-2.2, 0.7	–	–
Fall	-0.1	-1.6, 1.5	–	–
BMI††	-0.07	-0.2, 0.03	0.1642	0.1%
BMI^2††	-0.01	-0.03, 0.0	0.0071	0.6%
Multivitamin use	6.3	5.1, 7.5	<0.0001	8.3%
Vitamin D supplement use	5.2	2.5, 7.9	0.0002	1.1%
Calcium supplement use	3.9	1.9, 6.0	0.0002	1.1%
Healthy Eating Index	–	–	0.0278	0.6%
Poor, <51	Referent			
Needs Improvement, 51-80	2.0	0.3, 3.6	–	–
Good, >81	2.8	0.7, 4.9	–	–
Dairy Consumption	–	–	0.0008	1.1%
No Dairy	Referent			
1-3 servings per day	2.0	0.9, 3.1	–	–
>3 servings per day	3.3	0.7, 5.9	–	–
Cereal Consumption	–	–	0.0037	0.9%
No Cereal	Referent			
Up to 1x/week	1.1	-0.6, 2.7	–	–
>1x/week	2.5	0.9, 4.1	–	–
Brisk walking 180 min/week (>25 min/day)	3.1	0.4, 5.9	0.0268	0.9%
				Total model R^2: 25.4%

* Estimated change in 25(OH)D per unit increase in predictor variable, adjusted for all other covariates in model

** *P* for trend, adjusted for all other covariates in model

† R^2 for individual variables, adjusted for all other covariates in model; due to correlation between some variables, the sum of the individual variable R^2 values does not add up to the total R^2.

†† Coefficients for mean-centered BMI variables presented for ease of interpretation

Table 2.3 SNP Association in African American Participants of the Health ABC Cohort for SNPs Reported in Published GWAS of Serum 25(OH)D Phenotype in Caucasians

SNP	Gene	Chr	Position	Coded allele	Freq.	β* ± SE	*P*
rs7041**†	*GC*	4	72837198	T ††	0.82	-0.93 ± 0.53	0.08
rs2282679**†	*GC*	4	72827247	G	0.10	0.07 ± 0.71	0.92
rs1155563**†	*GC*	4	72862352	C	0.11	0.12 ± 0.66	0.85
rs2060793†	*CYP2R1*	11	14871886	A	0.37	-0.10 ± 0.42	0.81
rs10741657**	*CYP2R1*	11	14871454	A	0.28	-0.04 ± 0.46	0.92
rs1993116**†	*CYP2R1*	11	14866810	A	0.28	-0.06 ± 0.46	0.90
rs12785878**	*DHCR7/ NADSYN1*	11	70845097	G	0.73	-0.28 ± 0.46	0.55
rs3829251†	*DHCR7/ NADSYN1*	11	70872207	A	0.23	-0.23 ± 0.48	0.63

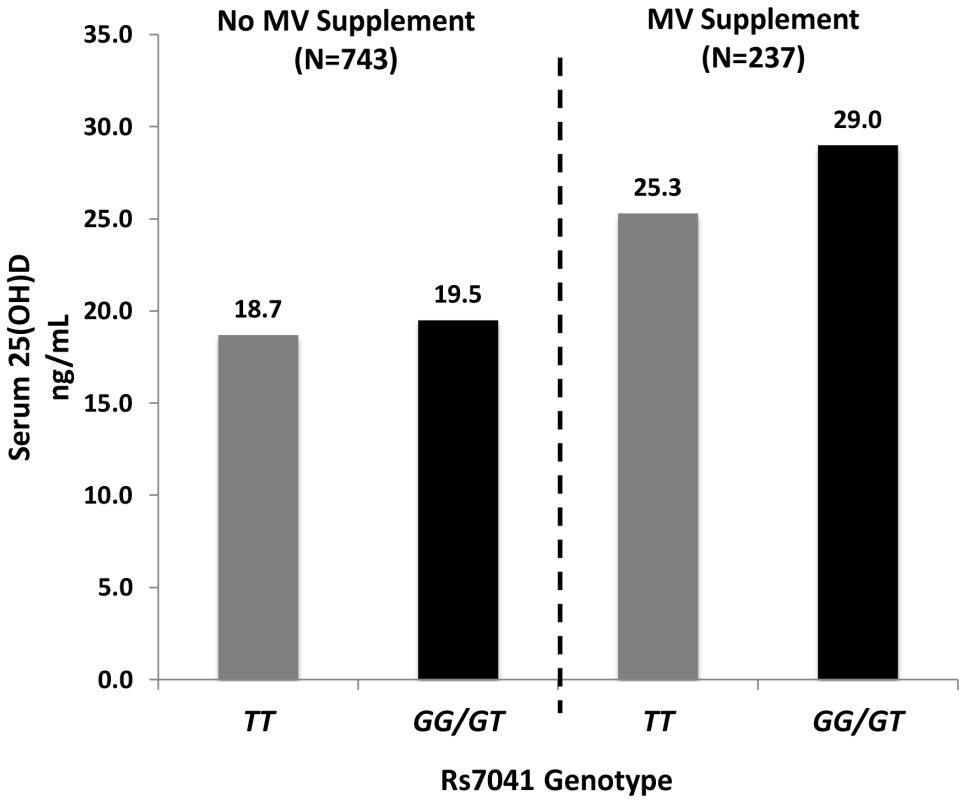
* Estimated change in serum 25(OH)D per copy of **coded** allele
**Associated with serum 25(OH)D in Wang *et al*
†Associated with serum 25(OH)D in Ahn *et al*
*** rs11234027 in *DHCR7/NADSYN1,* associated with 25(OH)D in Ahn *et al*, and rs6013897 in *CYP24A1*, associated with 25(OH)D in Wang *et al*, were not genotyped or imputed in the Health ABC African American population
†† Rs7041 was coded in Health ABC as an A/C SNP, but in this manuscript it is referred to as the equivalent T/G SNP to maintain consistency with the literature

Table 2.4 Estimate of variance in log-transformed 25(OH)D explained by all genome-wide autosomal SNPs *

Cohort	Cohort-specific estimates			Meta-analysis estimates (N=2,087)			
	N	h_g^2 **	S.E.	h_g^2	S.E.	*P* ***	95% CI †
Health ABC	889	0.25	0.25	0.23	0.15	0.14	0, 52
MESA	1,198	0.21	0.19				

*Total number of genotyped SNPs in Health ABC (prior to exclusion for MAF <0.01%): 1,024,986; **GCTA model covariates**: age, gender, study site, and season of blood draw
** Estimated proportion of phenotypic variance explained by additive genetic variation
*** Meta-analyzed *P* value for h_g^2 estimate
† Lower CI bound set at 0 because the genetic association with 25(OH)D serum concentrations cannot be <0

Figure 2.1 Serum 25(OH)D concentrations by multivitamin supplement use and rs7041 genotype (graph shows raw data, unadjusted).



Funding: This research was supported by National Institutes of Health, National Heart Lung and Blood Institute R03 HL095414 (PAC), by RC1-AG035835 (SK, PI; PAC, subcontract PI), and by NRSA Institutional Research Training Grant T32-DK-7158-36 (JGH). This research was also supported by NIA contracts N01AG62101, N01AG62103, and N01AG62106. The genome-wide association study was funded by NIA grant R01-AG032098-01A1 to Wake Forest University Health Sciences and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C. This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging.

The MESA study was supported by grant R01DK088762 and by N01HC95159 through N01HC95169 from the National Heart, Lung, and Blood Institute. The funding organizations and sponsors had no role in the design and conduct of the study; the collection, management analysis, and interpretation of the data; or the preparation review or approval of the manuscript.

Acknowledgements:

Health ABC: The authors thank the Health ABC staff, other investigators and the participants of the Health ABC study for their valuable contributions. The Health ABC website is at http://www.grc.nia.nih.gov/branches/leps/healthabc/index.htm .

MESA: The authors thank the other investigators, the staff, and the participants of the MESA Study for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa.nhlbi.org .

REFERENCES

1. Ross AC, Institute of Medicine (U. S.). Committee to Review Dietary Reference Intakes for Vitamin D and Calcium. Dietary reference intakes for calcium and vitamin d. Washington, DC: National Academies Press; 2011.

2. Wang TT, Tavera-Mendoza LE, Laperriere D, Libby E, MacLeod NB, Nagai Y, Bourdeau V, Konstorum A, Lallemant B, Zhang R, Mader S, White JH. Large-scale in silico and microarray-based identification of direct 1,25-dihydroxyvitamin d3 target genes. *Mol Endocrinol* 2005;19:2685-2695.

3. Zehnder D, Bland R, Williams MC, McNinch RW, Howie AJ, Stewart PM, Hewison M. Extrarenal expression of 25-hydroxyvitamin d(3)-1 alpha-hydroxylase. *The Journal of clinical endocrinology and metabolism* 2001;86:888-894.

4. Holick MF. Vitamin d deficiency. *The New England journal of medicine* 2007;357:266-281.

5. Looker AC, Johnson CL, Lacher DA, Pfeiffer CM, Schleicher RL, Sempos CT. Vitamin d status: United states, 2001-2006. *NCHS data brief* 2011:1-8.

6. MacLaughlin J, Holick MF. Aging decreases the capacity of human skin to produce vitamin d3. *The Journal of clinical investigation* 1985;76:1536-1538.

7. Wang TJ, Zhang F, Richards JB, Kestenbaum B, van Meurs JB, Berry D, Kiel DP, Streeten EA, Ohlsson C, Koller DL, Peltonen 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 CL, Wolf M, Rice K, Goltzman D, Hidiroglou N, Ladouceur M, Wareham NJ, Hocking LJ, Hart D, Arden NK, Cooper C, Malik S, Fraser WD, Hartikainen AL, Zhai G, Macdonald HM, Forouhi NG, Loos RJ, 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 JO, Cauley JA, Uitterlinden AG, Gibson Q, Jarvelin MR, 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;376:180-188.

8. 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.

9. Jacques PF, Felson DT, Tucker KL, Mahnken B, Wilson PW, Rosenberg IH, Rush D. Plasma 25-hydroxyvitamin d and its determinants in an elderly population sample. *The American journal of clinical nutrition* 1997;66:929-936.

10. McCullough ML, Weinstein SJ, Freedman DM, Helzlsouer K, Flanders WD, Koenig K, Kolonel L, Laden F, Le Marchand L, Purdue M, Snyder K, Stevens VL, Stolzenberg-Solomon R, Virtamo J, Yang G, Yu K, Zheng W, Albanes D, Ashby J, Bertrand K, Cai H, Chen Y, Gallicchio L, Giovannucci E, Jacobs EJ, Hankinson SE, Hartge P, Hartmuller V, Harvey C, Hayes RB, Horst RL, Shu XO. Correlates of circulating 25-hydroxyvitamin d: Cohort consortium vitamin d pooling project of rarer cancers. *American journal of epidemiology* 2010;172:21-35.

11. Hiraki LT, Major JM, Chen C, Cornelis MC, Hunter DJ, Rimm EB, Simon KC, Weinstein SJ, Purdue MP, Yu K, Albanes D, Kraft P. Exploring the genetic architecture of circulating 25-hydroxyvitamin d. *Genetic epidemiology* 2013;37:92-98.

12. Hypponen E, Power C. Hypovitaminosis d in british adults at age 45 y: Nationwide cohort study of dietary and lifestyle predictors. *The American journal of clinical nutrition* 2007;85:860-868.

13. van Dam RM, Snijder MB, Dekker JM, Stehouwer CD, Bouter LM, Heine RJ, Lips P. Potentially modifiable determinants of vitamin d status in an older population in the netherlands: The hoorn study. *The American journal of clinical nutrition* 2007;85:755-761.

14. Benjamin A, Moriakova A, Akhter N, Rao D, Xie H, Kukreja S, Barengolts E. Determinants of 25-hydroxyvitamin d levels in african-american and caucasian male veterans. *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA* 2009;20:1795-1803.

15. Chan J, Jaceldo-Siegl K, Fraser GE. Determinants of serum 25 hydroxyvitamin d levels in a nationwide cohort of blacks and non-hispanic whites. *Cancer causes & control : CCC* 2010;21:501-511.

16. 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;93:3381-3388.

17. Harris SS, Soteriades E, Coolidge JA, Mudgal S, Dawson-Hughes B. Vitamin d insufficiency and hyperparathyroidism in a low income, multiracial, elderly population. *The Journal of clinical endocrinology and metabolism* 2000;85:4125-4130.

18. Egan KM, Signorello LB, Munro HM, Hargreaves MK, Hollis BW, Blot WJ. Vitamin d insufficiency among african-americans in the southeastern united states: Implications for cancer disparities (united states). *Cancer causes & control : CCC* 2008;19:527-535.

19. Shea MK, Houston DK, Tooze JA, Davis CC, Johnson MA, Hausman DB, Cauley JA, Bauer DC, Tylavsky F, Harris TB, Kritchevsky SB. Correlates and prevalence of insufficient 25-hydroxyvitamin d status in black and white older adults: The health, aging and body composition study. *Journal of the American Geriatrics Society* 2011;59:1165-1174.

20. Freedman DM, Cahoon EK, Rajaraman P, Major JM, Doody MM, Alexander BH, Hoffbeck RW, Kimlin MG, Graubard BI, Linet MS. Sunlight and other determinants of circulating 25-hydroxyvitamin d levels in black and white participants in a nationwide u.S. Study. *American journal of epidemiology* 2013;177:180-192.

21. Wjst M, Altmuller J, Braig C, Bahnweg M, Andre E. A genome-wide linkage scan for 25-oh-d(3) and 1,25-(oh)2-d3 serum levels in asthma families. *The Journal of steroid biochemistry and molecular biology* 2007;103:799-802.

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* 2009;63:458-464.

23. Orton SM, Morris AP, Herrera BM, Ramagopalan SV, Lincoln MR, Chao MJ, Vieth R, Sadovnick AD, Ebers GC. Evidence for genetic regulation of vitamin d status in twins with multiple sclerosis. *The American journal of clinical nutrition* 2008;88:441-447.

24. Hunter D, De Lange M, Snieder H, MacGregor AJ, Swaminathan R, Thakker RV, Spector TD. Genetic contribution to bone metabolism, calcium excretion, and vitamin d and parathyroid hormone regulation. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research* 2001;16:371-378.

25. Signorello LB, Shi J, Cai Q, Zheng W, Williams SM, Long J, Cohen SS, Li G, Hollis BW, Smith JR, Blot WJ. Common variation in vitamin d pathway genes predicts circulating 25-hydroxyvitamin d levels among african americans. *PloS one* 2011;6:e28623.

26. Lee JS, Kritchevsky SB, Tylavsky FA, Harris T, Everhart J, Simonsick EM, Rubin SM, Newman AB. Weight-loss intention in the well-functioning, community-dwelling elderly: Associations with diet quality, physical activity, and weight change. *The American journal of clinical nutrition* 2004;80:466-474.

27. Promotion. UCfNPa. The healthy eating index. Washington, DC: USDA Center for Nutrition Policy and Promotion; 1995.

28. Artigas MS, Loth DW, Wain LV, Gharib SA, Obeidat M, Tang W, Zhai G, Zhao JH, Smith AV, Huffman JE, Albrecht E, Jackson CM, Evans DM, Cadby G, Fornage M, Manichaikul A, Lopez LM, Johnson T, Aldrich MC, Aspelund T, Barroso I, Campbell H, Cassano PA, Couper DJ, Eiriksdottir G, Franceschini N, Garcia M, Gieger C, Gislason GK, Grkovic I, Hammond CJ, Hancock DB, Harris TB, Ramasamy A, Heckbert SR, Heliovaara M, Homuth G, Hysi PG, James AL, Jankovic S, Joubert BR, Karrasch S, Klopp N, Koch B, Kritchevsky SB, Launer LJ, Liu Y, Loehr LR, Lohman K, Loos RJ, Lumley T, Al Balushi KA, Ang WQ, Barr RG, Beilby J, Blakey JD, Boban M, Boraska V, Brisman J, Britton JR, Brusselle GG, Cooper C, Curjuric I, Dahgam S, Deary IJ, Ebrahim S, Eijgelsheim M, Francks C, Gaysina D, Granell R, Gu X, Hankinson JL, Hardy R, Harris SE, Henderson J, Henry A, Hingorani AD, Hofman A, Holt PG, Hui J, Hunter ML, Imboden M, Jameson KA, Kerr SM, Kolcic I, Kronenberg F, Liu JZ, Marchini J, McKeever T, Morris AD, Olin AC, Porteous DJ, Postma DS, Rich SS, Ring SM, Rivadeneira F, Rochat T, Sayer AA, Sayers I, Sly PD, Smith GD, Sood A, Starr JM, Uitterlinden AG, Vonk JM, Wannamethee SG, Whincup PH, Wijmenga C, Williams OD, Wong A, Mangino M, Marciante KD, McArdle WL, Meibohm B, Morrison AC, North KE, Omenaas E, Palmer LJ, Pietilainen KH, Pin I, Pola Sbreve Ek O, Pouta A, Psaty BM, Hartikainen AL, Rantanen T, Ripatti S, Rotter JI, Rudan I, Rudnicka AR, Schulz H, Shin SY, Spector TD, Surakka I, Vitart V, Volzke H, Wareham NJ, Warrington NM, Wichmann HE, Wild SH, Wilk JB, Wjst M, Wright AF, Zgaga L, Zemunik T, Pennell CE, Nyberg F, Kuh D, Holloway JW, Boezen HM, Lawlor DA, Morris RW, Probst-Hensch N, Kaprio J, Wilson JF, Hayward C, Kahonen M, Heinrich J, Musk AW, Jarvis DL, Glaser S, Jarvelin MR, Ch Stricker BH, Elliott P, O'Connor GT, Strachan DP, London SJ, Hall IP, Gudnason V, Tobin MD. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* 2011.

29. Yang J, Lee SH, Goddard ME, Visscher PM. Gcta: A tool for genome-wide complex trait analysis. *American journal of human genetics* 2011;88:76-82.

30. Willer CJ, Li Y, Abecasis GR. Metal: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010;26:2190-2191.

31. Sachs MC, Shoben A, Levin GP, Robinson-Cohen C, Hoofnagle AN, Swords-Jenny N, Ix JH, Budoff M, Lutsey PL, Siscovick DS, Kestenbaum B, de Boer IH. Estimating mean annual 25-hydroxyvitamin d concentrations from single measurements: The multi-ethnic study of atherosclerosis. *The American journal of clinical nutrition* 2013;97:1243-1251.

32. Jones G, Prosser DE, Kaufmann M. 25-hydroxyvitamin d-24-hydroxylase (cyp24a1): Its important role in the degradation of vitamin d. *Archives of biochemistry and biophysics* 2012;523:9-18.

33. Gropper SAS, Smith JL, Groff JL. Advanced nutrition and human metabolism. Belmont, CA: Thomson/Wadsworth; 2004.

34. Vimaleswaran KS, Berry DJ, Lu C, Tikkanen E, Pilz S, Hiraki LT, Cooper JD, Dastani Z, Li R, Houston DK, Wood AR, Michaelsson K, Vandenput L, Zgaga L, Yerges-Armstrong LM, McCarthy MI, Dupuis J, Kaakinen M, Kleber ME, Jameson K, Arden N, Raitakari O, Viikari J, Lohman KK, Ferrucci L, Melhus H, Ingelsson E, Byberg L, Lind L, Lorentzon M, Salomaa V, Campbell H, Dunlop M, Mitchell BD, Herzig KH, Pouta A, Hartikainen AL, Streeten EA, Theodoratou E, Jula A, Wareham NJ, Ohlsson C, Frayling TM, Kritchevsky SB, Spector TD, Richards JB, Lehtimaki T, Ouwehand WH, Kraft P, Cooper C, Marz W, Power C, Loos RJ, Wang TJ, Jarvelin MR, Whittaker JC, Hingorani AD, Hypponen E. Causal relationship between obesity and vitamin d status: Bi-directional mendelian randomization analysis of multiple cohorts. *PLoS medicine* 2013;10:e1001383.

35. Blum M, Dolnikowski G, Seyoum E, Harris SS, Booth SL, Peterson J, Saltzman E, Dawson-Hughes B. Vitamin d(3) in fat tissue. *Endocrine* 2008;33:90-94.

36. Maclaughlin J, Holick MF. Aging decreases the capacity of human-skin to produce vitamin-d3. *Journal of Clinical Investigation* 1985;76:1536-1538.

37. Sinotte M, Diorio C, Berube S, Pollak M, Brisson J. Genetic polymorphisms of the vitamin d binding protein and plasma concentrations of 25-hydroxyvitamin d in premenopausal women. *The American journal of clinical nutrition* 2009;89:634-640.

38. Janssens W, Bouillon R, Claes B, Carremans C, Lehouck A, Buysschaert I, Coolen J, Mathieu C, Decramer M, Lambrechts D. Vitamin d deficiency is highly prevalent in copd and correlates with variants in the vitamin d-binding gene. *Thorax* 2010;65:215-220.

39. Abbas S, Linseisen J, Slanger T, Kropp S, Mutschelknauss EJ, Flesch-Janys D, Chang-Claude J. The gc2 allele of the vitamin d binding protein is associated with a decreased postmenopausal breast cancer risk, independent of the vitamin d status. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2008;17:1339-1343.

40. Wood AM, Bassford C, Webster D, Newby P, Rajesh P, Stockley RA, Thickett DR. Vitamin d-binding protein contributes to copd by activation of alveolar macrophages. *Thorax* 2011.

41. Kamboh MI, Ferrell RE. Ethnic variation in vitamin d-binding protein (gc): A review of isoelectric focusing studies in human populations. *Human genetics* 1986;72:281-293.

42. Database of single nucleotide polymorphisms (dbsnp). dbSNP accession:rs7041, (dbSNP Build ID: 137) ed. Bethesda (MD): National Center for Biotechnology Information, National Library of Medicine.

43. Armas LA, Dowell S, Akhter M, Duthuluru S, Huerter C, Hollis BW, Lund R, Heaney RP. Ultraviolet-b radiation increases serum 25-hydroxyvitamin d levels: The effect of uvb dose and skin color. *Journal of the American Academy of Dermatology* 2007;57:588-593.

44. Parra EJ, Kittles RA, Shriver MD. Implications of correlations between skin color and genetic ancestry for biomedical research. *Nat Genet* 2004;36:S54-60.

45. Signorello LB, Williams SM, Zheng W, Smith JR, Long J, Cai Q, Hargreaves MK, Hollis BW, Blot WJ. Blood vitamin d levels in relation to genetic estimation of african ancestry. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2010;19:2325-2331.

46. Campbell MC, Tishkoff SA. African genetic diversity: Implications for human demographic history, modern human origins, and complex disease mapping. *Annual review of genomics and human genetics* 2008;9:403-433.

SUPPLEMENTAL MATERIALS

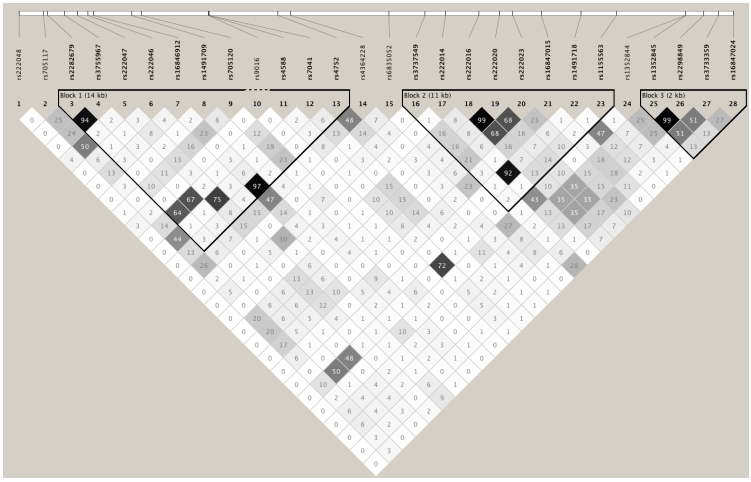
Candidate GWAS-associated SNP selection

Two genome-wide associations studies in Caucasians were reviewed to develop a list of candidate 25(OH)D-associated SNPs (7, 8). From the study by Ahn *et al*, we included the following SNPs associated with 25(OH)D at a GWAS significance level: rs2282679 in *GC*, rs3829251 in *DHCR7/NADSYN1*, and rs2060793 in *CYP2R1*. The association of rs6599638 in c10orf88 with 25(OH)D was not confirmed in a replication sample, and consequently not included in the current study. In the second GWAS study by Wang *et al*, the most significant SNPs from each associated gene were included in our preliminary list: rs2282679 in *GC*, rs12785878 in *DHCR7/NADSYN1*, rs10741657 in *CYP2R1*, and rs6013897 in *CYP24A1*. Finally, SNPs strongly associated with 25(OH)D at GWAS or near-GWAS level significance and in strong LD with the most significant SNPs in Ahn *et al* and Wang *et al* were included: rs7041 and rs1155563 in *GC,* rs1993116 in *CYP2R1*, and rs11234027 in *NADSYN1*.

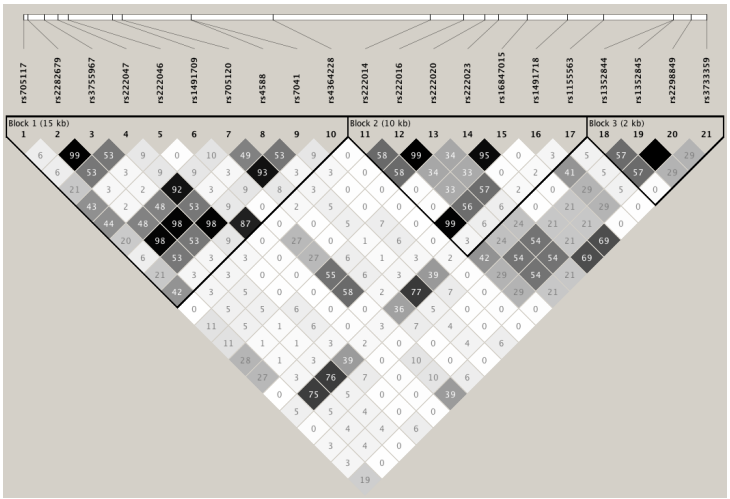
Two SNPs (rs6013897 and rs11234027) were not in the Health ABC African-American database, leaving a final set of 8 candidate SNPs in three genomic loci: rs7041, rs2282679, and rs1155563 in *GC*; rs2060793, rs10741657, rs1993116 in *CYP2R1*; and rs12785878 and rs3829251 in *DHCR7/NADSYN1*.

Supplemental Figure 2.2

a) Linkage disequilibrium (LD) plot of *GC* in Health ABC African-Americans. Shading represents linkage (R^2), where black shading represents complete linkage between SNPs and white shading represents no linkage. Inset triangles represent LD blocks.



b) LD plot of *GC* in Health ABC European Americans.



Supplemental Table 2.5 Comparison of frequencies for GWAS-associated SNPs between Caucasians and Africans (gene frequencies from HapMap populations, CEU and YRI, as reported in dbSNP)

SNP	Allele*	Gene	CEU Frequency	YRI Frequency	Health ABC Frequency
rs7041	A/T	*GC*	0.43	0.90	0.82
rs2282679	G/C	*GC*	0.26	0.04	0.10
rs1155563	C	*GC*	0.29	0.05	0.11
rs2060693	A	*CYP2R1*	0.39	0.35	0.37
rs10741657	A	*CYP2R1*	0.37	0.22	0.28
rs1993116	A/T	*CYP2R1*	0.40	0.26	0.28
rs12785878	G	*DHCR7/NADSYN1*	0.27	0.84	0.73
rs3829251	A	*DHCR7/NADSYN1*	0.17	0.27	0.23

*For three SNPs, the coded allele in Health ABC differed from the allele reported in dbSNP; in these instances, allele frequencies for the complementary allele are reported.

CHAPTER 3

VITAMIN D-RESPONSIVE *SGPP2* VARIANTS ASSOCIATED WITH LUNG CELL EXPRESSION AND LUNG FUNCTION

B Reardon*[1], JG Hansen*[1], RG Crystal[2], DK Houston[3], SB Kritchevsky[3], T Harris[4], K Lohman[5], Y Liu[6], GT O'Connor[7,8], JB Wilk[8,9], J Mezey[2,10], C Gao[10] and PA Cassano[1,11]

*These authors contributed equally to this work. Specifically, Brian Reardon conducted the gene expression study described in the first three paragraphs of "Subjects and Methods" and the first two paragraphs of "Results", and contributed the data presented in Table 3.1, Supplemental Table 3.6, and Supplemental Table 3.8. All other data presented in this chapter represents original research conducted by Joyanna Hansen.

[1]Division of Nutritional Sciences, Cornell University, Ithaca, NY 14853 USA

[2] Department of Genetic Medicine and Division of Pulmonary and Critical Care Medicine, Department of Medicine, Weill Cornell Medical College, New York, NY, USA

[3]Sticht Center on Aging, Wake Forest School of Medicine, Winston-Salem, NC 27157 USA

[4]Intramural Research Program, National Institute on Aging, Laboratory of Epidemiology, Demography, and Biometry, Gateway Building, 3C309, 7201 Wisconsin Avenue, Bethesda, MD 20892 USA

[5]Department of Biostatistical Sciences, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC 27157 USA

[6]Department of Epidemiology & Prevention, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC 27157 USA

[7]Section of Pulmonary, Allergy & Critical Care Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA

[8]The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA, USA

[9]Division of Aging, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

[10]Biological Statistics and Computational Biology, Cornell University, Ithaca NY 14853 USA

[11]Department of Public Health, Division of Biostatistics and Epidemiology, Weill Cornell Medical College, New York, NY, USA

Running Head: Vitamin D Responsive *SGPP2* Variants and Lung

Address correspondence to: Dr. Patricia A. Cassano, Division of Nutritional Sciences, Cornell University, 209 Savage Hall, Ithaca, NY 14853, pac6@cornell.edu, (607) 255-7551 (ph), (607) 255-2691 (fax).

Keywords: Vitamin D, airflow obstruction, FEV_1, *SGPP2*, FEV_1/FVC

ABSTRACT

Background: Vitamin D is associated with lung health in epidemiologic studies, but mechanisms mediating observed associations are poorly understood. This study explores mechanisms for an effect of vitamin D in lung through an *in vivo* gene expression study, an expression quantitative trait loci (eQTL) analysis in lung tissue, and a population-based cohort study of sequence variants.

Methods: Microarray analysis investigated the association of gene expression in small airway epithelial cells with serum 25(OH)D in adult non-smokers. Sequence variants in candidate genes identified by the microarray were investigated in a lung tissue eQTL database, and also in relation to cross-sectional pulmonary function in the Health, Aging, and Body Composition (Health ABC) study, stratified by race, with replication in the Framingham Heart Study (FHS).

Results: Thirteen candidate genes had significant differences in expression by serum 25(OH)D (nominal $p<0.05$), and a genome-wide significant eQTL association was detected for *SGPP2.* In Health ABC, *SGPP2* SNPs were associated with FEV_1 in both European- and African Americans, and the gene-level association was replicated in European-American FHS participants. SNPs in 5 additional candidate genes (*DAPK1, FSTL1, KAL1, KCNS3,* and *RSAD2*) were associated with FEV_1 in Health ABC participants.

Conclusions: *SGPP2,* a sphingosine-1-phosphate phosphatase, is a novel vitamin D-responsive gene associated with lung function. Pending confirmation by follow-up studies, this study implicates lipid-signaling molecules as a key factor in inter-individual variation in cross-sectional lung function and points to a new direction for future research

INTRODUCTION

Vitamin D is of interest in relation to a number of health outcomes, with putative function beyond its classical role in maintaining bone health. The active form of vitamin D, 1,25-dihydroxyvitamin D [1,25$(OH)_2$D], when bound to the vitamin D receptor (VDR), regulates the expression of genes in many molecular pathways, including inflammation, cell proliferation, cell death, and tissue-remodeling pathways (1). Serum 25-hydroxyvitamin D [25(OH)D] is the primary circulating biomarker of vitamin D status, and recent national survey data in the U.S. indicate 32% of Americans are at risk of vitamin D inadequacy or deficiency, defined as <50 nmol/L and <30 nmol/L serum 25(OH)D, respectively (2, 3).

Chronic obstructive pulmonary disease (COPD) is the third leading cause of death in the United States, and is a large and growing burden on health care (4). While smoking is the primary risk factor for rapid lung function decline and development of COPD, about 15% of individuals who have never smoked develop COPD and not all smokers succumb, implicating other factors, such as genetic, dietary, and lifestyle factors, in lifetime lung function patterns and disease risk (5).

Recent evidence indicates that vitamin D, as a steroid hormone capable of influencing gene expression, may be a determinant of lung function (6). A cross-sectional study in the National Health and Examination Survey (NHANES) III reported a strong positive association between serum 25(OH)D and lung function, with clinically relevant effect sizes for forced expiratory volume in the first second (FEV_1) and forced vital capacity (FVC) (7). However, a subsequent cross-sectional study in the U.K. reported no association between serum 25(OH)D and FEV_1 (8). Causal inferences are limited in the cross-sectional design, effect estimates may be biased by uncontrolled confounders such as physical activity, and, furthermore, comparisons are limited by differences in the range in serum 25(OH)D between studies. Investigations of serum 25(OH)D or

high-dose vitamin D supplementation in relation to the risk of exacerbations in COPD patients reported overall null findings (9, 10). However, vitamin D supplementation led to a statistically significant reduction in COPD exacerbations in the subgroup with severe vitamin D deficiency at the study baseline (serum 25(OH)D < 10 ng/mL) (9), underscoring the importance of considering the potential to benefit in studies of nutritional supplementation.

In vitro animal and cell culture studies demonstrate that vitamin D-responsive genes play a role in airway remodeling and inflammation, which are key processes in the pathogenesis of COPD (11, 12). However, few studies directly investigate mechanisms for vitamin D's effect *in vivo,* which would strengthen the causal inference of population-level association studies. Furthermore, most experimental work to date has focused on effects of the active metabolite of vitamin D, 1,25-dihydroxyvitamin D. This metabolite is generated in the kidney for systemic circulation, and in many tissues, including lung (13). It is not yet established whether the population-level range in serum 25-hydroxyvitamin D, the primary biomarker for vitamin D status in humans, is associated with effects similar to those seen *in vitro* for 1,25-hydroxyvitamin D.

We used an interdisciplinary approach to investigate the mechanisms through which vitamin D affects lung function. Genes with *in vitro* evidence of vitamin D regulation were studied to assess whether serum 25(OH)D concentration was associated with gene expression in lung epithelial tissues sampled from free-living humans. Identified genes were investigated in a study of expression quantitative trait loci (eQTL) in human lung epithelial cells to assess if genetic variation affects gene expression. Also, identified genes were investigated in an epidemiologic cohort study in relation to pulmonary function phenotypes. We hypothesized that serum 25(OH)D affects expression of vitamin D-responsive genes by modulating levels of active $1,25(OH)_2D$ in lung

tissue, and that variants in candidate genes directly regulated by $1,25(OH)_2D$ in lung tissue are associated with FEV_1 and FEV_1/FVC, the key parameters used for COPD diagnosis and staging.

SUBJECTS AND METHODS

Gene Expression Study

Twenty-six healthy nonsmoker adult volunteers (**Supplemental Table 3.6**) were recruited and evaluated at the Weill Cornell Medical College (WCMC) General Clinical Research Center under protocols approved by the WCMC Institutional Review Board, as described elsewhere (14). Frozen sera samples were assayed for 25(OH)D by liquid chromatography-tandem mass spectrometry at the Division of Laboratory Sciences, Centers for Disease Control and Prevention (Atlanta, GA). Airway epithelial cells were collected by brushing during bronchoscopy (14), and first and second strand cDNA were synthesized from 6 μg of RNA, *in vitro* transcribed, and fragmented according to Affymetrix protocols; samples were hybridized to the Affymetrix HG-U133 Plus 2.0 array (14). (Supplemental Methods for further details)

The microarray analysis considered 156 genes, which were identified *a priori* based on evidence of regulation by 1,25-dihydroxyvitamin D in squamous epithelial cells (1) and evidence for at least one predicted binding site for VDR (a DR3 or ER6 response element with up to 1 base mismatch from the consensus sequence) (1).

The statistical significance of fold-changes in expression between the first and third tertile of serum 25(OH)D was calculated using a t-test with Bayesian correction (Limma). Given that the purpose of the microarray study was to identify candidate genes to take forward to both the eQTL and the population-based cohort analysis, a statistical significance threshold of nominal $P < 0.05$

was used. Linear regression coefficients and the variance (R^2) in gene expression explained by serum 25(OH)D were calculated, and included the full range of 25(OH)D concentrations.

eQTL Study

The Expression Quantitative Trait Loci (eQTL) study was conducted using lung small airway epithelium tissue samples from 116 individuals (Supplemental Methods for details). Tissue samples were collected under protocols approved by the WCMC Institutional Review Board.

Population-based Cohort Study

The Health, Aging and Body Composition (Health ABC) cohort study enrolled a random sample of European-Americans and all African American Medicare-eligible residents, aged 70-79 at baseline (1997) and residing in the ZIP codes in and around Memphis, TN and Pittsburgh, PA (n=3,075). The Institutional Review Boards at the University of Memphis, Tennessee, and the University of Pittsburgh granted approval to conduct the Health ABC Study. The Institutional Review Board at Cornell University and the Health ABC Publications Committee approved the use of Health ABC data for this study. The Framingham Heart Study (FHS) cohort (n=7,694; includes individuals from the original, offspring, and third generation cohorts) (15) served as a replication cohort for cross-sectional SNP—lung function associations discovered in Health ABC European-Americans (Supplemental Methods for further details on both cohort studies). The Institutional Review Board at the Boston University Medical Campus granted approval for the FHS.

Spirometry met American Thoracic Society criteria for acceptability (16) (17). Participants with missing covariate data were excluded from further consideration (~ 300 in each ancestry group). Participants with an FEV_1 measurement and an FEV_1/FVC ratio below the Lower Limit of Normal were considered to have prevalent airflow obstruction (17, 18). The Illumina Human 1M-Duo custom chip was used for genotyping in Health ABC (19). All assayed SNPs in the 13

candidate genes (identified by the expression study) with a minor allele frequency > 5% and in Hardy Weinberg equilibrium were analyzed, comprising 313 SNPs in European-Americans and 355 SNPs in African Americans (**Supplemental Table 3.7**).

Ordinary least squares linear regression models examined the relation between SNPs and FEV_1 and FEV_1/FVC in sequential regressions (using SAS 9.2). An additive genetic model was used to estimate the main effect of each SNP; SNPs with a nominal $P \leq 0.02$ were further tested in dominant and recessive genetic models to refine effect estimates (Supplemental Methods for details). In genetic studies, the risk of false positives must be minimized without ruling out true associations (20). GWAS-scale multiple testing adjustments are not appropriate for the hypothesis-based investigation of the 13 genomic regions nominated by the gene expression study. Thus, SNPs with nominally significant p-values are presented, and False Discovery Rate (FDR) multiple testing correction was applied (21). Models were adjusted for age, height, cigarette smoking (smoking status and pack-years), gender, study site, and ancestry principal components.

Sensitivity analyses were performed on the top findings for the FEV_1 phenotype by repeating analyses after excluding individuals with prevalent airflow obstruction or individuals with lower quality spirometry (lower reproducibility scores). Exploratory SNP x serum 25(OH)D interaction analyses are presented in Appendix A.

RESULTS

Gene Expression by Serum 25-Hydroxyvitamin D

Healthy, non-smoking adults (n=26) were divided into tertiles of serum 25(OH)D (range of serum 25(OH)D: 2.3-39.7 ng/mL); the lowest tertile boundary corresponded to the cutpoint for deficiency (< 12 ng/mL), and the upper tertile included only vitamin D sufficient individuals (all ≥

20 ng/mL), thus further analysis compared these two groups. Expected associations were confirmed; serum vitamin D concentrations were lower in African American participants, and slightly higher in males (**Supplemental Table 3.6**).

Among the 156 genes studied, thirteen genes (8.3%) had statistically significant (nominal $p < 0.05$) differences in expression between the first and third tertiles of serum 25-hydroxyvitamin D (**Table 3.1**). To characterize further the relation of serum 25-hydroxyvitamin D with the 13 nominally significant genes, the linear association of gene expression with continuous serum 25-hydroxyvitamin D was estimated (**Table 3.1**); the percent of variance (R^2, from linear regression) explained by serum 25-hydroxyvitamin D ranged from 8 to 40%, and *FSTL1* had the highest R^2.

eQTL Analysis

All 13 vitamin D-responsive genes were queried in the eQTL data, but only 12 genes had available data (no data for *RSAD2*). A highly statistically significant *cis* eQTL reaching genome-wide significance thresholds was identified for *SGPP2*; a cluster of SNPs in the 3' region of *SGPP2* was associated with *SGPP2* gene expression in lung tissue (the lead SNP, rs13009608 had a nominal p-value of 2.99×10^{-09}). Figure 2 shows gene-level results and Supplemental Figures 1 and 2 show genome-wide Q-Q and Manhattan plots, respectively. The association of rs13009608 with *SGPP2* gene expression was replicated (p-value: 7.0×10^{-18}) in a publically available eQTL database of lymphoblastoid cell lines (22).

Population-level SNP—Lung Function Associations

All 13 vitamin D-responsive genes identified by the microarray screen were further studied in a population-based candidate gene association study. After excluding participants with missing covariate data, 1,502 European-Americans and 996 African Americans (81% of full cohort) had an acceptable FEV_1 and were included in the FEV_1 analysis. 1,472 European-Americans and 943

African Americans (79% of cohort) had an acceptable FEV_1/FVC, and were included in the ratio analysis (**Table 3.2**).

Five SNPs in two genes (*DAPK1* and *SGPP2)* were associated with FEV_1 at a nominal P<0.02 in European-American participants (P-value range: $2.88x10^{-03}$ to $1.92x10^{-02}$; **Table 3.3**). A SNP in *DAPK1* (rs11141878) had the largest effect; participants with two copies of the minor allele (recessive genotype) were 104 mL lower on FEV_1. In African Americans, 18 SNPs in 6 genes (*DAPK1, FSTL1, KAL1, KCNS3, RSAD2,* and *SGPP2*) were associated with FEV_1 at nominal P<0.02 (range: $1.11x10^{-04}$ to $1.65x10^{-02}$; **Table 3.4**). A group of 3 linked SNPs in a linked 5' block of *SGPP2* were associated with a decreased FEV_1 and a reduced FEV_1/FVC ratio in African Americans with nominal P-values <0.02 and FDR q-values <0.05 (Figure 1; **Table 3.5**). A fourth SNP in *SGPP2*, rs4597517, was borderline significantly associated with FEV_1 in African Americans in the additive model ($p = 2.16x10^{-2}$), and statistically significantly associated with FEV_1 ($p = 4.28x10^{-4}$) in the recessive genetic model. A SNP in *KCNS3* (rs3747515) had the largest effect on FEV_1 in African Americans; participants with the recessive genotype were 244 mL higher on FEV_1. Due to linkage, some SNP associations were redundant; thus, SNPs in the same gene with an $R^2 \geq 0.9$ (indicating strong linkage) are assumed to represent the same effect and redundant SNPs are presented in the supplement only (**Supplemental Table 3.10**).

In European-Americans, 1 SNP in *KLF4* was associated with the FEV_1/FVC ratio (P-value $1.15x10^{-2}$; **Supplemental Table 3.11**). In African Americans, 14 SNPs in 3 genes (*FSTL1, KAL1,* and *SGPP2*) were associated with the ratio at a nominal P<0.02 (range: $1.32x10^{-03}$ to $1.27x10^{-02}$; **Supplemental Table 3.11**).

A sensitivity analysis explored whether the SNP—FEV_1 associations primarily reflected effects of genetic variation on risk of COPD; analyses were repeated after excluding 110 European

Americans and 66 African Americans with prevalent airflow obstruction (as an indicator of COPD). For European-Americans there was little or no difference in analyses with and without prevalent cases. A Bland-Altman analysis showed that for SNPs in *SGPP2,* the effect estimates for African Americans were attenuated after excluding cases of prevalent airflow obstruction (data not shown). Thus, the *SGPP2* SNPs that had statistically significant associations with FEV_1 were further tested in logistic regression models to assess the *SGPP2*—outcome association in African Americans. Individuals with two copies of the SNP most statistically significantly associated with FEV_1, rs4528748, had a 2.6-fold increased risk of airflow obstruction. All 3 *SGPP2* SNPs had odds ratios above 2 for the SNP—COPD association, and all confidence intervals excluded 1 (**Table 3.5**), supporting a role for *SGPP2* in mediating COPD risk.

There was consistency of findings across both phenotypes and both ancestry groups for 2 genes, namely *SGPP2* and *DAPK1*. SNPs in *SGPP2* and *DAPK1* were associated with FEV_1 in both European-Americans and African Americans, and SNPs in *SGPP2* were also associated with FEV_1/FVC and with risk of prevalent airflow obstruction in African Americans.

Genes containing SNPs significantly associated with FEV_1 or FEV_1/FVC in Health ABC European-Americans, namely *DAPK1, KLF4,* and *SGPP2*, were further evaluated in the FHS cohort. Gene-level replication was observed for *DAPK1* and *SGPP2;* 23 out of 340 SNPs in *DAPK1* (6.8%) and 23 out of 145 SNPs (15.8%) in *SGPP2* were associated with cross-sectional FEV_1 at a nominal P-value <0.05 in the FHS cohort, although these comprised different SNPs than the ones associated with lung function in Health ABC (**Supplemental Table 3.9**).

DISCUSSION

Using an interdisciplinary genomics approach we investigated vitamin D and lung outcomes. *SGPP2,* a phosphatase involved in the sphingosine-1-phosphate signaling pathway, was identified in all stages of the study as a promising candidate gene contributing to vitamin D-mediated associations with lung function. *SGPP2* is differentially expressed *in vivo* in lung epithelial cells by serum 25(OH)D. eQTL analysis demonstrates that sequence variants in *SGPP2* are associated with lung cell gene expression. Although the eQTL finding does not prove that vitamin D regulation affects gene expression, the location of associated variants in regulatory regions supports the hypothesis of vitamin D regulation. Furthermore, a group of 3 linked SNPs in the *SGPP2* promoter region are associated with lower FEV_1, a reduced FEV_1/FVC ratio, and a 2-3 fold increased risk of airflow obstruction in African Americans, suggesting that a causal variant in this region may affect *SGPP2* function and, consequently, lung outcomes. Additionally, a SNP in *SGPP2* is associated with FEV_1 in Health ABC European-Americans and *SGPP2* variants were also associated with FEV_1 in the Framingham Heart Study, confirming effects across racial groups and in two cohort studies. This multi-faceted approach identifies putative mechanistic pathways for observed vitamin D—lung function associations while reducing the chance of false positive results.

SGPP2 plays a key role in the sphingolipid signaling pathway through dephosphorylation of sphingosine-1-phosphate (S1P) to sphingosine, which is then converted to ceramide or back to sphingosine-1-phosphate by other enzymes (23). Sphingosine-1-phosphate acts as both an intracellular and extracellular signaling molecule, and regulates critical cell processes including apoptosis, cell growth, and immune function (23, 24). Altered sphingolipid concentrations have important ramifications for lung function; ceramide concentrations are elevated in COPD, contributing to lung alveolar destruction (23). Little research exists on *SGPP2,* although a 2006 paper showed that *SGPP2* is up-regulated in response to inflammatory stimuli in endothelial cells,

suggesting a possible role in mediating inflammation in lung tissue (25). However, *SGPP2*'s biological function to alter sphingosine-1-phosphate concentrations suggests that this gene contributes to the regulation of sphingolipid signaling pathways in lung tissue.

We identified several additional genes, namely *DAPK1, KCNS3,* and *FSTL1,* and all three had mechanistic links to lung function identified through gene ontology analysis and literature reviews (**Supplemental Table 3.8**). Expression of all three genes was strongly associated with serum 25(OH)D, and variants in these genes were associated with pulmonary function in the Health ABC cohort study. However, variants were not replicated in the Framingham Heart Study, nor were there observed eQTL associations. *DAPK1,* which is down-regulated by $1,25(OH)_2D$ both *in vivo* and *in vitro,* is a pro-apoptotic kinase linked to cytoskeletal remodeling and regulation of inflammatory gene expression in macrophages (26, 27). SNPs in *KCNS3,* which encodes a voltage-gated potassium channel protein, were associated with airway hyperresponsiveness in past studies (28), which is of interest given postulated associations of airways hyperresponsiveness with an accelerated rate of FEV_1 decline and risk of COPD (29). *FSTL1* up-regulates pro-inflammatory cytokines; in mice, the highest expression level is in lung (30). Dexamethasone, which is a glucocorticoid used to treat both asthma and COPD, is associated with expression of both *KCNS3* and *FSTL1*(31, 32)*;* interestingly, there are striking similarities in the effects of dexamethasone and 1,25-dihydroxyvitamin on the expression of these genes. The combination of 1,25-dihydroxyvitamin D with dexamethasone was investigated *in vitro* as an anti-inflammatory treatment (33); our results suggest the strong possibility of synergistic effects for this treatment combination.

A major strength of this study is that it translates *in vitro* animal and cell culture studies to an *in vivo* study, and then extends to study population-level SNP associations with lung

phenotypes, which are partially replicated in an independent cohort. The multi-stage approach identified *SGPP2* as a promising vitamin D-responsive gene for further study. The demonstration of differential gene expression in lung tissue associated with the physiologic range of 25-hydroxyvitamin D in a diverse sample of free-living humans confirms *in vitro* studies, and, while our study does not manipulate vitamin D, the *in vivo* evidence of association is novel. The Health ABC population-based cohort study included high-quality spirometry, detailed information on confounding factors such as smoking and population stratification, and comprised 40% African American participants, thus allowing consideration of this understudied population in genomic research. FEV_1 is a predictor of all-cause mortality (34), and thus SNP—FEV_1 associations are clinically relevant. Although associations between SNPs and the FEV_1/FVC ratio were also investigated, the associations were not as strong as for FEV_1. Thus, vitamin D may have a stronger association with overall lung health versus the risk of COPD. This study identifies plausible biological mechanisms that support a true effect of vitamin D on lung function, and will help to guide the design and analysis of randomized controlled intervention trials of the role of vitamin D in lung disease.

Given that the microarray analysis was used exclusively as a candidate screen, limitations including the lack of qPCR confirmation (not possible due to sample volume limitations), use of nominal P values, and the lack of race-stratified analysis (not possible due to sample size limitations) are less of a concern. As expected, the proportion of participants in the race/ethnicity groups varied by tertile of serum 25(OH)D given the role of skin pigmentation in vitamin D synthesis in response to sunlight (2). Race may either confound the serum 25(OH)D—gene expression association, or, race may be a causal antecedent variable that, in part, causes serum 25(OH)D concentration and, in turn, differences in gene expression; adjusting for race may be an

over-adjustment. Of note, in regressions adjusted for race the regression coefficients for the serum 25(OH)D—gene expression association were similar to unadjusted analyses.

While the studies were all cross-sectional, which limits causal inference, the harmony of findings across different designs partly mitigates this concern. Although it would have been ideal to use the same samples in all studies (that is, expression, eQTL and SNP—lung function studies), practical limitations led to the use of different samples in each phase. Finally, although gene-level replication was observed for *SGPP2* and *DAPK1*, the specific SNPs associated with FEV_1 in Health ABC did not reach statistical significance in FHS. We hypothesize that the *SGPP2* SNPs identified in the two cohort studies may be tagging the same unknown causal variant(s) or there may be multiple *SGPP2* regulatory regions associated with lung function. Additionally, the strongest SNP—lung function associations in Health ABC were in African Americans, and, because FHS includes only European Americans, the replication was partial. In summary, SNPs in *SGPP2* were statistically significantly associated with lung outcomes after FDR multiple testing adjustment and a highly statistically significant lung eQTL was identified for *SGPP2*; *SGPP2* emerged as a clear candidate in all stages of this work.

CONCLUSIONS

This study establishes for the first time that physiological concentrations of serum 25(OH)D are associated with differences in gene expression in human lung tissue, and that candidate vitamin D responsive genes are associated with pulmonary function outcomes. We hypothesize that genetic variants associated with pulmonary function in our study affect binding of the VDR/RXR heterodimer to the genome; however, further studies are needed to map lung tissue-specific regulatory regions. Recent evidence shows that vitamin D regulatory elements (VDREs) are located both proximal and distal to vitamin D-responsive genes at promoter regions and enhancer regions,

respectively, and that VDR/RXR binding is cell-type specific (35). This emphasizes the importance of genome-wide VDR/RXR mapping in lung cells to identify regulatory regions (35). Additionally, *in vitro* studies of bronchial epithelial cells to directly assess gene expression changes due to vitamin D would contribute to the current understanding. Overall, the results of our study identify putative mechanisms through which vitamin D may affect lung function and, suggest a physiological range for 25-hydroxyvitamin D at which differential responses occur at the molecular level. Demonstrated associations strengthen the evidence for monitoring serum 25(OH)D concentrations in individuals at risk of rapid decline in lung function.

Table 3.1 Fold Change in Expression and P-value of 13 Genes Reaching Nominal P-value Threshold ($p<0.05$) in Expression Study

Gene	Chromosome	Fold Change*	P-value	$R^{2§}$
KCNS3	2	-1.62	0.00084	28%
FSTL1	3	-1.55	0.00163	40%
DAPK1	9	-2.06	0.00381	17%
RSAD2	2	1.41	0.01103	16%
CST6	11	1.79	0.01516	20%
KAL1	X	-1.38	0.01840	28%
SLITRK6	13	-1.52	0.02482	25%
TMEM40	3	1.55	0.02518	23%
EMB	5	1.52	0.03099	23%
PTGER2	14	1.36	0.03574	9%
DTX4	11	-1.34	0.03812	15%
KLF4	9	1.66	0.03901	9%
SGPP2	2	1.69	0.04491	24%

*Fold change in high versus low tertile serum 25-hydroxyvitamin D
§R-squared calculated in linear regression, considering the full range of serum 25-hydroxyvitamin D, thus equals the proportion of variance in expression accounted for serum 25(OH)D

Table 3.2 Characteristics of Health, Aging and Body Composition Study Participants Included in the FEV_1 Phenotype* Analysis, Stratified by Race

Covariate	African Americans (N=996)	European-Americans (N=1,502)
Age, years**	73.4 (2.9)	73.7 (2.8)
Women (%)	553 (55.5)	708 (47.1)
Memphis, TN site (%)	464 (46.6)	759 (50.5)
Former Smokers (%)	398 (40)	746 (49.7)
Current Smokers (%)	167 (16.8)	99 (6.6)
Pack-years	29.5 (24.1)	36.5 (31.9)
FEV_1, mL	1948.7 (569.4)	2305.4 (654.3)
FEV_1/FVC	75.5 (9.3)	74.4 (7.9)
Height, cm	165.7 (9.4)	167 (9.3)
Mean 25(OH)D (ng/mL)***	20.9 (10.6)	29 (11)
COPD, defined by LLN (%)	66 (7.0)	110 (7.5)

* All participants in table have FEV_1 data; approximately 50 fewer individuals have FEV_1/FVC ratio data, but participant characteristics are the same for both phenotypes.

**Data shown are mean (SD) or number (%)

***Serum 25(OH)D measured for 1,412 (94%) European-Americans and 864 (87%) African Americans with the FEV_1 phenotype, and for 1,383 European-Americans and 864 African Americans with the FEV_1/FVC phenotype.

Table 3.3 The Association of SNPs in Vitamin D-Responsive Genes (nominal P<$2.0x10^{-02}$) with FEV_1 (mL) for European-Americans in the Health, Aging and Body Composition Study (sorted by gene)*

Gene	RS#	Chr	Coded Allele	MAF (%)	β**	SE	Nominal P	Model
DAPK1	rs11141878	9	A	36	-103.98	36.3	$4.26x10^{-03}$	R
	rs4877361†	9	G	14	72.47	27.4	$8.17x10^{-03}$	D
	rs4878089	9	A	46	39.68	16.9	$1.92x10^{-02}$	A
SGPP2	rs4674656	2	A	25	-58.70	19.7	$2.88x10^{-03}$	A

†one redundant SNP not shown

*Abbreviations: Chr, chromosome; MAF, minor allele frequency; β, regression coefficient; SE, standard error; A=additive genetic model, D=dominant model, R=recessive model

**Beta-coefficient estimates the association of allele with FEV_1, based on genetic model shown and adjusted for age, height, smoking, gender, study site, and ancestry principal components.

Table 3.4 The Association of SNPs in Vitamin D-Responsive Genes (nominal P<$2.0x10^{-02}$) with FEV_1 (mL) for African Americans in the Health, Aging and Body Composition Study (sorted by gene)*

Gene	RS#	Chr	Coded Allele	MAF (%)	β**	SE	Nominal P	Model
DAPK1	rs3128491	9	G	33	51.48	21.4	$1.65x10^{-02}$	A
FSTL1	rs4676781	3	T	8	-110.13	35.3	$1.88x10^{-03}$	A
	rs13100865	3	G	9	-105.96	35.0	$2.54x10^{-03}$	A
	rs13097755†	3	T	28	-60.46	21.6	$5.20x10^{-03}$	A
KAL1	rs6530200	23	T	47	-45.28	16.8	$7.20x10^{-03}$	A
	rs974655	23	A	49	79.23	30.3	$9.14x10^{-03}$	D
KCNS3	rs1031771†	2	A	16	243.76	83.5	$3.60x10^{-03}$	R
RSAD2	rs4669114††	2	G	10	-119.55	36.2	$9.93x10^{-04}$	D
	rs6431837	2	C	47	-101.06	33.6	$2.66x10^{-03}$	R
	rs7570384	2	C	38	-55.35	20.1	$5.88x10^{-03}$	A
	rs4669111	2	A	41	-49.75	20.1	$1.34x10^{-02}$	A
SGPP2	rs4528748††	2	C	27	-209.95	54.1	$1.11x10^{-04***}$	R

*Abbreviations: Chr, chromosome; MAF, minor allele frequency; β, regression coefficient; SE, standard error; A=additive genetic model, D=dominant model, R=recessive model

**Beta-coefficient estimates the association of allele with FEV_1, based on genetic model shown, adjusted for age, height, smoking, gender, study site, and ancestry principal components.

*** FDR q-value <0.05

† one redundant SNP not shown

†† two redundant SNPs not shown

Table 3.5 Associations of SNPs in *SGPP2* with Risk of Prevalent COPD* in African Americans in the Health, Aging and Body Composition Study

SNP**	Odds Ratio	95% Confidence Interval	Nominal P
rs4528748	2.63	1.19, 5.80	$1.64x10^{-02}$
rs7556867	2.71	1.23, 5.99	$1.35x10^{-02}$
rs6758392	2.34	1.07, 5.11	$3.33x10^{-02}$

*COPD defined as FEV_1 and FEV_1/FVC ratio below the Lower Limit of Normal

*All SNPs modeled as recessive (two copies of the minor allele) to reflect the most significant coding from Table 3, and models adjusted for age, height, smoking, gender, study site, and ancestry principal components.

Figure 3.1 Association between SNPs and FEV_1 in *SGPP2*. This figure shows all SNPs tested for association with FEV_1 in African Americans (red markers) and European-Americans (blue markers) in Health ABC. The top graph shows the p-values for each SNP on a negative log scale. The threshold for significance, nominal $P = 2x10^{-02}$, is shown as a line in the figure. Effect estimates (β_{SNP}) for FEV_1 (in mL) for each ancestry group are shown underneath the P-values (dotted line shows null value of 0). Effect estimates and p-values are from recessive, dominant, or additive genetic models for SNPs with p<0.02, and from an additive genetic model for all other SNPs. Finally, the linkage disequilibrium structure of *SGPP2* in the Health ABC European-American population is shown at the bottom, with darker shading representing higher R^2.

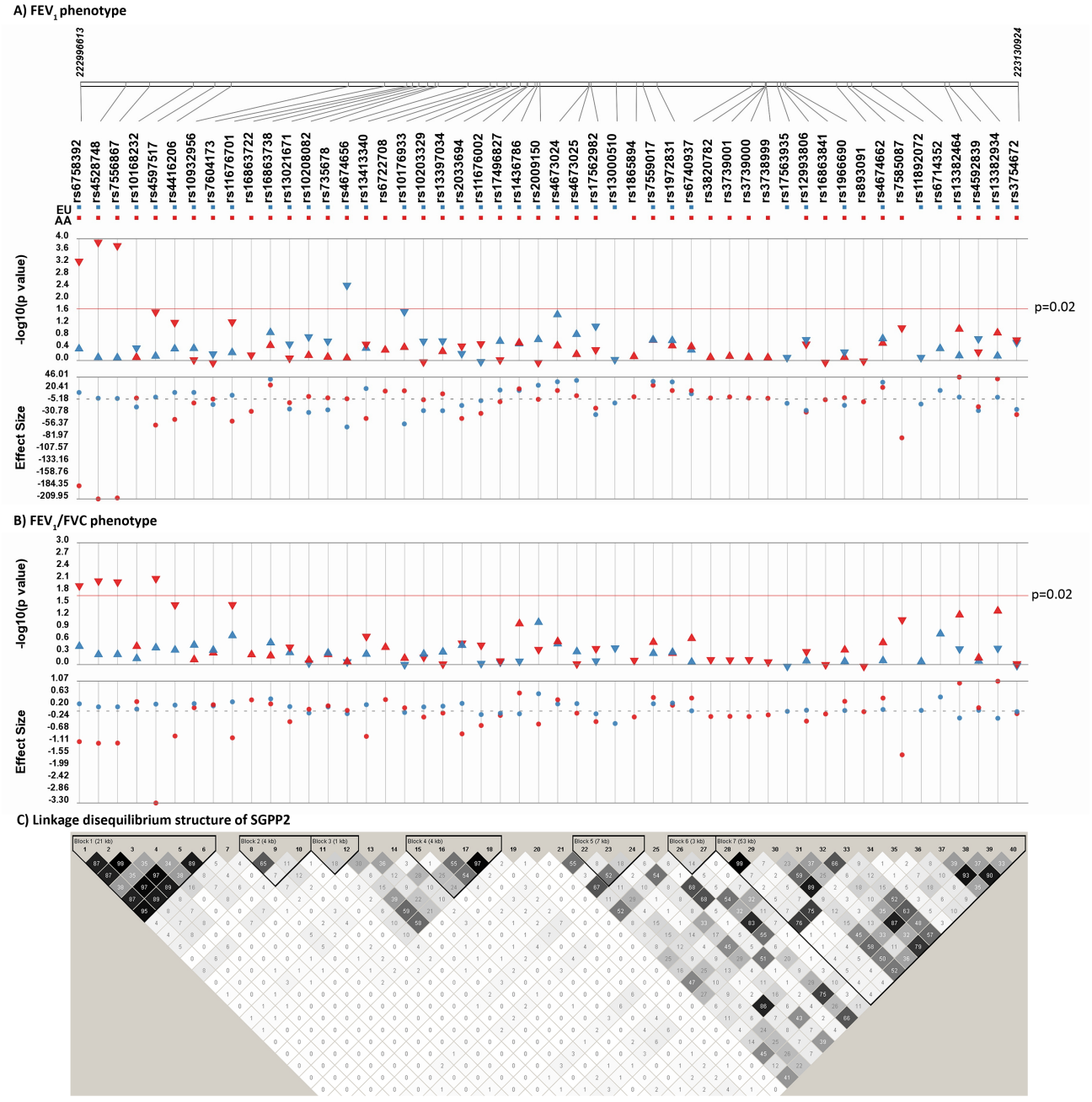
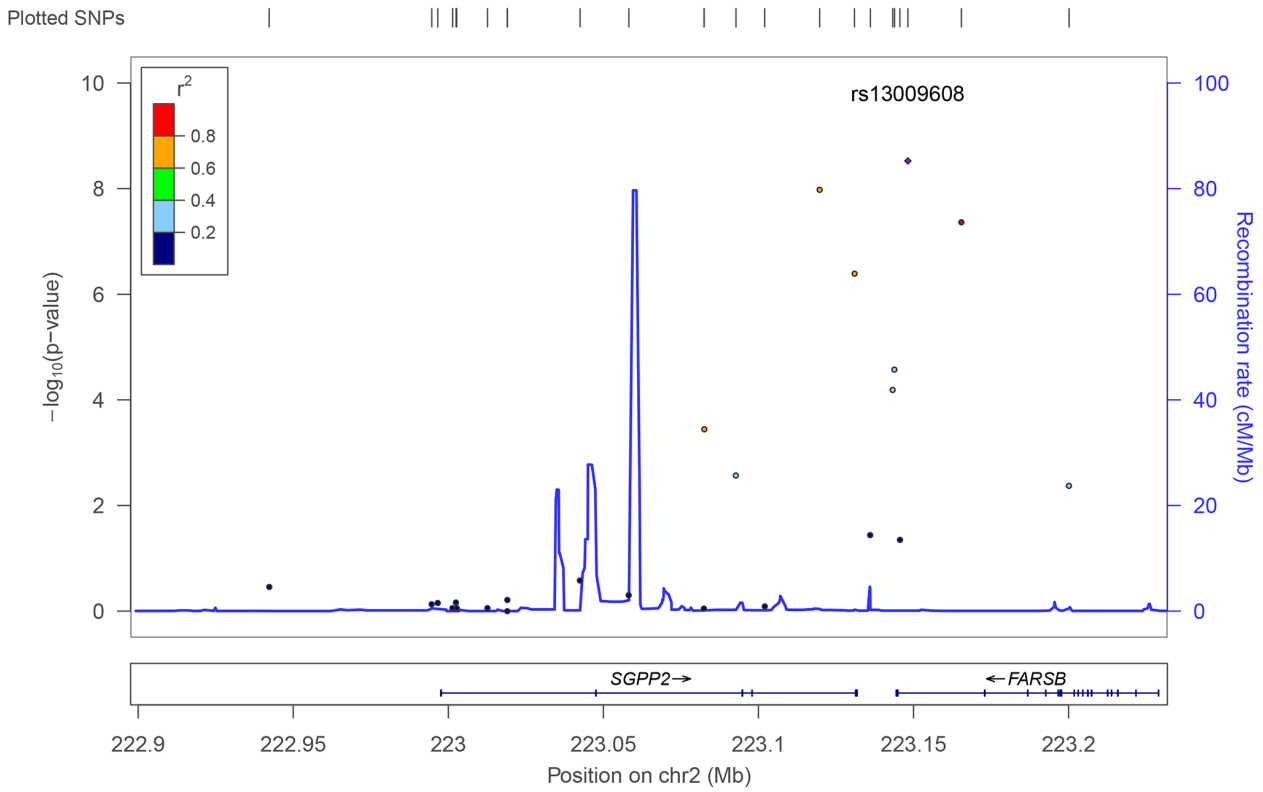


Figure 3.2 Locus Zoom plot of *SGPP2* eQTL associations



Funding: This research was supported by National Institutes of Health, National Heart Lung and Blood Institute R03 HL095414 (PAC) and P50 HL084936 (RGC), by RC1-AG035835 (SK, PI; PAC, PI subcontract), and by NRSA Institutional Research Training Grant T32-DK-7158-36 (JGH). This research also was supported by R01-AG029364, by NIA contracts N01AG62101, N01AG62103, and N01AG62106, and by the Ashken Foundation (RGC). The genome-wide association study was funded by NIA grant R01-AG032098-01A1 to Wake Forest University Health Sciences and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C. This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging. Research was conducted in part using data and resources from the Framingham Heart Study of the NHLBI of the NIH and Boston University School of Medicine. The analyses reflect intellectual input and resource development from the Framingham investigators participating in the SNP Health Association Resource (SHARe) project. This work was partially supported by the NHLBI's Framingham Heart Study (Contract No. N01-HC-25195) and its contract with Affymetrix, Inc. for genotyping services (Contract No. N02-HL-6-4278). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center.

Contributor Statement: All authors satisfy the requirements for authorship and contributorship. BR, RC and PAC designed and conducted the expression study; YL, KL, SK and TH conducted the Health ABC GWAS study, which provided data for this paper; JGH, BR and PAC designed the Health ABC SNP study and JGH and PAC conducted the SNP study; JGH, PAC, JW and GO'C

conducted the replication analysis in FHS, and JGH, PAC, JM and CG conducted the eQTL analysis and interpretation. All coauthors read and edited the final manuscript.

REFERENCES

1. Wang TT, Tavera-Mendoza LE, Laperriere D, Libby E, MacLeod NB, Nagai Y, Bourdeau V, Konstorum A, Lallemant B, Zhang R, Mader S, White JH. Large-scale in silico and microarray-based identification of direct 1,25-dihydroxyvitamin d3 target genes. *Mol Endocrinol* 2005;19:2685-2695.

2. Ross AC, Institute of Medicine (U. S.). Committee to Review Dietary Reference Intakes for Vitamin D and Calcium. Dietary reference intakes for calcium and vitamin d. Washington, DC: National Academies Press; 2011.

3. Looker AC JC, Lacher DA, et al. Vitamin d status: United states, 2001–2006. NCHS data brief, no 5. Hyattsville, MD: National Center for Health Statistic; 2011.

4. Minino AM. Death in the united states, 2009. *NCHS Data Brief* 2011:1-8.

5. Eisner MD, Anthonisen N, Coultas D, Kuenzli N, Perez-Padilla R, Postma D, Romieu I, Silverman EK, Balmes JR. An official american thoracic society public policy statement: Novel risk factors and the global burden of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2010;182:693-718.

6. Finklea JD, Grossmann, R.E., and Tangpricha, V. Vitamin d and chronic lung disease: A review of molecular mechanisms and clinical studies. *Advances in Nutrition* 2011;2:244-253.

7. Black PN, Scragg R. Relationship between serum 25-hydroxyvitamin d and pulmonary function in the third national health and nutrition examination survey. *Chest* 2005;128:3792-3798.

8. Shaheen SO, Jameson KA, Robinson SM, Boucher BJ, Syddall HE, Aihie Sayer A, Cooper C, Holloway JW, Dennison EM. Relationship of vitamin d status to adult lung function and copd. *Thorax* 2011;66:692-698.

9. Lehouck A, Mathieu C, Carremans C, Baeke F, Verhaegen J, Van Eldere J, Decallonne B, Bouillon R, Decramer M, Janssens W. High doses of vitamin d to reduce exacerbations in chronic obstructive pulmonary disease: A randomized trial. *Ann Intern Med* 2012;156:105-114.

10. Kunisaki KM, Niewoehner DE, Connett JE. Vitamin d levels and risk of acute exacerbations of chronic obstructive pulmonary disease: A prospective cohort study. *Am J Respir Crit Care Med* 2012;185:286-290.

11. Wittke A, Chang A, Froicu M, Harandi OF, Weaver V, August A, Paulson RF, Cantorna MT. Vitamin d receptor expression by the lung micro-environment is required for maximal induction of lung inflammation. *Arch Biochem Biophys* 2007;460:306-313.

12. Bosse Y, Maghni K, Hudson TJ. 1alpha,25-dihydroxy-vitamin d3 stimulation of bronchial smooth muscle cells induces autocrine, contractility, and remodeling processes. *Physiol Genomics* 2007;29:161-168.

13. Hansdottir S, Monick MM, Hinde SL, Lovan N, Look DC, Hunninghake GW. Respiratory epithelial cells convert inactive vitamin d to its active form: Potential effects on host defense. *J Immunol* 2008;181:7090-7099.

14. Harvey BG, Heguy A, Leopold PL, Carolan BJ, Ferris B, Crystal RG. Modification of gene expression of the small airway epithelium in response to cigarette smoking. *J Mol Med (Berl)* 2007;85:39-53.

15. Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, Marciante KD, Franceschini N, van Durme YM, Chen TH, Barr RG, Schabath MB, Couper DJ, Brusselle GG, Psaty BM, van Duijn CM, Rotter JI, Uitterlinden AG, Hofman A, Punjabi NM, Rivadeneira F, Morrison AC, Enright PL, North KE, Heckbert SR, Lumley T, Stricker BH, O'Connor GT, London SJ. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet* 2010;42:45-52.

16. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, Crapo R, Enright P, van der Grinten CP, Gustafsson P, Jensen R, Johnson DC, MacIntyre N, McKay R, Navajas D, Pedersen OF, Pellegrino R, Viegi G, Wanger J. Standardisation of spirometry. *Eur Respir J* 2005;26:319-338.

17. Waterer GW, Wan JY, Kritchevsky SB, Wunderink RG, Satterfield S, Bauer DC, Newman AB, Taaffe DR, Jensen RL, Crapo RO. Airflow limitation is underrecognized in well-functioning older people. *J Am Geriatr Soc* 2001;49:1032-1038.

18. Mohamed Hoesein FA, Zanen P, Lammers JW. Lower limit of normal or fev1/fvc < 0.70 in diagnosing copd: An evidence-based review. *Respir Med* 2011;105:907-915.

19. Artigas MS, Loth DW, Wain LV, Gharib SA, Obeidat M, Tang W, Zhai G, Zhao JH, Smith AV, Huffman JE, Albrecht E, Jackson CM, Evans DM, Cadby G, Fornage M, Manichaikul A, Lopez LM, Johnson T, Aldrich MC, Aspelund T, Barroso I, Campbell H, Cassano PA, Couper DJ, Eiriksdottir G, Franceschini N, Garcia M, Gieger C, Gislason GK, Grkovic I, Hammond CJ, Hancock DB, Harris TB, Ramasamy A, Heckbert SR, Heliovaara M, Homuth G, Hysi PG, James AL, Jankovic S, Joubert BR, Karrasch S, Klopp N, Koch B, Kritchevsky SB, Launer LJ, Liu Y,

Loehr LR, Lohman K, Loos RJ, Lumley T, Al Balushi KA, Ang WQ, Barr RG, Beilby J, Blakey JD, Boban M, Boraska V, Brisman J, Britton JR, Brusselle GG, Cooper C, Curjuric I, Dahgam S, Deary IJ, Ebrahim S, Eijgelsheim M, Francks C, Gaysina D, Granell R, Gu X, Hankinson JL, Hardy R, Harris SE, Henderson J, Henry A, Hingorani AD, Hofman A, Holt PG, Hui J, Hunter ML, Imboden M, Jameson KA, Kerr SM, Kolcic I, Kronenberg F, Liu JZ, Marchini J, McKeever T, Morris AD, Olin AC, Porteous DJ, Postma DS, Rich SS, Ring SM, Rivadeneira F, Rochat T, Sayer AA, Sayers I, Sly PD, Smith GD, Sood A, Starr JM, Uitterlinden AG, Vonk JM, Wannamethee SG, Whincup PH, Wijmenga C, Williams OD, Wong A, Mangino M, Marciante KD, McArdle WL, Meibohm B, Morrison AC, North KE, Omenaas E, Palmer LJ, Pietilainen KH, Pin I, Pola Sbreve Ek O, Pouta A, Psaty BM, Hartikainen AL, Rantanen T, Ripatti S, Rotter JI, Rudan I, Rudnicka AR, Schulz H, Shin SY, Spector TD, Surakka I, Vitart V, Volzke H, Wareham NJ, Warrington NM, Wichmann HE, Wild SH, Wilk JB, Wjst M, Wright AF, Zgaga L, Zemunik T, Pennell CE, Nyberg F, Kuh D, Holloway JW, Boezen HM, Lawlor DA, Morris RW, Probst-Hensch N, Kaprio J, Wilson JF, Hayward C, Kahonen M, Heinrich J, Musk AW, Jarvis DL, Glaser S, Jarvelin MR, Ch Stricker BH, Elliott P, O'Connor GT, Strachan DP, London SJ, Hall IP, Gudnason V, Tobin MD. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* 2011.

20. Cooper GM, Shendure J. Needles in stacks of needles: Finding disease-causal variants in a wealth of genomic data. *Nat Rev Genet* 2011;12:628-640.

21. Storey JD. A direct approach to false discovery rates. *J R Statist Soc B* 2002;64:479–498.

22. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KC, Taylor J, Burnett E, Gut I, Farrall M, Lathrop GM, Abecasis GR, Cookson WO. A genome-wide association study of global gene expression. *Nat Genet* 2007;39:1202-1207.

23. Yang Y, Uhlig S. The role of sphingolipids in respiratory disease. *Ther Adv Respir Dis* 2011;5:325-344.

24. Chi H. Sphingosine-1-phosphate and immune regulation: Trafficking and beyond. *Trends Pharmacol Sci* 2011;32:16-24.

25. Mechtcheriakova D, Wlachos A, Sobanov J, Kopp T, Reuschel R, Bornancin F, Cai R, Zemann B, Urtz N, Stingl G, Zlabinger G, Woisetschlager M, Baumruker T, Billich A. Sphingosine 1-phosphate phosphatase 2 is induced during inflammatory responses. *Cell Signal* 2007;19:748-760.

26. Houle F, Poirier A, Dumaresq J, Huot J. Dap kinase mediates the phosphorylation of tropomyosin-1 downstream of the erk pathway, which regulates the formation of stress fibers in response to oxidative stress. *J Cell Sci* 2007;120:3666-3677.

27. Mukhopadhyay R, Ray PS, Arif A, Brady AK, Kinter M, Fox PL. Dapk-zipk-l13a axis constitutes a negative-feedback module regulating inflammatory gene expression. *Mol Cell* 2008;32:371-382.

28. Hao K, Niu T, Xu X, Fang Z. Single-nucleotide polymorphisms of the kcns3 gene are significantly associated with airway hyperresponsiveness. *Hum Genet* 2005;116:378-383.

29. Brutsche MH, Downs SH, Schindler C, Gerbase MW, Schwartz J, Frey M, Russi EW, Ackermann-Liebrich U, Leuenberger P. Bronchial hyperresponsiveness and the development of asthma and copd in asymptomatic individuals: Sapaldia cohort study. *Thorax* 2006;61:671-677.

30. Miyamae T, Marinov AD, Sowders D, Wilson DC, Devlin J, Boudreau R, Robbins P, Hirsch R. Follistatin-like protein-1 is a novel proinflammatory molecule. *J Immunol* 2006;177:4758-4762.

31. Wu Y, Zhou S, Smas CM. Downregulated expression of the secreted glycoprotein follistatin-like 1 (fstl1) is a robust hallmark of preadipocyte to adipocyte conversion. *Mechanisms of development* 2010;127:183-202.

32. Kinyamu HK, Collins JB, Grissom SF, Hebbar PB, Archer TK. Genome wide transcriptional profiling in breast cancer cells reveals distinct changes in hormone receptor target genes and chromatin modifying enzymes after proteasome inhibition. *Molecular carcinogenesis* 2008;47:845-885.

33. Antoniucci DM, Black DM, Sellmeyer DE. Serum 25-hydroxyvitamin d is unaffected by multiple freeze-thaw cycles. *Clin Chem* 2005;51:258-261.

34. Schunemann HJ, Dorn J, Grant BJ, Winkelstein W, Jr., Trevisan M. Pulmonary function is a long-term predictor of mortality in the general population: 29-year follow-up of the buffalo health study. *Chest* 2000;118:656-664.

35. Hinson ER, Joshi NS, Chen JH, Rahner C, Jung YW, Wang X, Kaech SM, Cresswell P. Viperin is highly induced in neutrophils and macrophages during acute and chronic lymphocytic choriomeningitis virus infection. *J Immunol* 2010;184:5723-5731.

36. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. Snap: A web-based tool for identification and annotation of proxy snps using hapmap. *Bioinformatics* 2008;24:2938-2939.

37. Barrett JC. Haploview: Visualization and analysis of snp genotype data. *Cold Spring Harb Protoc* 2009;2009:pdb ip71.

38. Gao C, Tignor N, Salit J, Strulovici-Barel Y, Hackett N, Crystal RG, Mezey JG. Heft: Eqtl analysis of many thousands of expressed genes while simultaneously controlling for hidden factors. *Manuscript submitted for publication*.

39. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. Locuszoom: Regional visualization of genome-wide association scan results. *Bioinformatics* 2010;26:2336-2337.

SUPPLEMENTAL METHODS

Gene Expression Study

Serum 25(OH)D Assays

Frozen serum samples from the 26 individuals recruited for the gene expression study were sent to the Division of Laboratory Sciences at the Centers for Disease Control and Prevention to be assayed for 25(OH)D by liquid chromatography-tandem mass spectrometry. Samples were evaluated in parallel with NIST standard SRM 972, and results were within 2 standard deviations of NIST target values. External quality assurance was provided through participation in the Vitamin D External Quality Assessment Scheme.

Two individuals were excluded because serum and cell collections occurred in different seasons and more than 120 days apart (their serum measurements were in the second tertile of the distribution). The average time between serum and epithelial cell collection was 42 ± 40 days, within the 10-week half-life of serum 25-hydroxyvitamin D.

Microarrays

Image files for the arrays were assessed for quality of hybridization by comparing 3' to 5' intensity of transcripts for actin and GAPDH (ratio < 3). Normalization was carried out by GeneChip Robust Multi-Array Average using MADMAX software (https://madmax.bioinformatics.nl). Quality control of normalized data was evaluated using plots of relative log expression and normalized unscaled standard errors to identify array artifacts. Only probe sets with an interquartile range (IQR) of $\log_2$ normalized values < 0.5 were included in analysis.

Gene Ontology Analysis

Gene ontology annotations were obtained from the UniProtKb-GOA database (http://www.ebi.ac.uk/QuickGO/), with preference given to IDA annotations (inferred from direct assay) or TAS (traceable author statement) evidence codes. IEA (inferred from electronic annotation) evidence codes were used if no other information was available.

Population-based Cohort Study

Participants, Data Collection and Statistical Approach

Participants: To be eligible for the Health ABC cohort study, participants were required to be ambulatory, that is, to have no difficulty walking ¼ mile or climbing 10 stairs without resting, and to be able to independently perform basic activities. Additionally, participants were required to have no history of active cancer treatment at baseline, and no plans to leave the area within 3 years after study baseline. A total of 3,075 individuals were enrolled.

Data Collection: Spirometry was conducted by trained personnel using a dry rolling seal spirometer (SensorMedics Corporation, Yorba Linda, CA) connected to a personal computer. Pulmonary function tests from the baseline clinic visit meeting American Thoracic Society (ATS) criteria for acceptability were included in this study.(17)

Statistical Approach: In all statistical models, SNPs were coded as the number of minor alleles an individual had at a specific genetic locus, based on Health ABC-specific allele frequency data. All models adjusted for population substructure using principal components, which were computed separately by race across all markers. Redundancy between SNP associations was assessed using SNAP (36) in CEU and YRI HapMap populations. SNP-pulmonary function associations were visualized using SynthesisView (37), and linkage disequilibrium in the Health ABC cohort was evaluated in Haploview 4.2.

Genetic Models for SNP—FEV_1 Regression Analyses**:** In models estimating the main effect of SNPs on cross-sectional FEV_1, an additive genetic model was used to estimate the main effect of each SNP; SNPs with a nominal $P \leq 0.02$ were further tested in dominant and recessive genetic models to refine effect estimates. SNPs are modeled in terms of the *effect of the minor SNP allele*. The regression model is presented below with an explanation of additive, dominant, and recessive SNP coding. The SNP used as an example is rs4528748 in the *SGPP2* gene; in African-Americans, rs4528748 had the strongest association with cross-sectional FEV_1.

Regression Model:

$$\mathbf{FEV_1} = \beta_0 + \beta_1\text{Height} + \beta_2\text{Age} + \beta_3\text{Gender} + \beta_4\text{Site} + \beta_5\text{Pack-years} + \beta_6\text{Smoking status} + \beta_7\text{Principal component 1} + \beta_8\text{Principal component 2} + \mathbf{\beta_9 SNP} + \varepsilon$$

Additive genetic model:

The SNP beta coefficient (β_9 from model above) represents the estimated effect per copy of the minor allele on cross-sectional FEV_1**.** For the example SNP, rs4528748, the SNP beta coefficient represents the estimated effect per copy of the *C* minor allele on cross-sectional FEV_1.

Recessive genetic model:

In the recessive model, individuals with two copies of the minor allele are compared to heterozygotes and homozygous individuals with zero copies of the minor allele, combined. For rs4528748, the SNP beta coefficient represents the estimated association with cross-sectional FEV_1 of the *CC* genotype compared to the *CT* + *TT* genotypes.

Dominant genetic model:

In the dominant model, individuals with one or two copies of the minor allele (heterozygotes and homozygous recessive individuals, combined) are compared to individuals with zero copies of the minor allele. For rs4528748, the SNP beta coefficient represents the estimated association with cross-sectional FEV_1 of the *CT + CC* genotypes compared to *TT*.

Sensitivity Analyses**:** In a sensitivity analysis to explore the effect of spirometry quality on the findings, individuals with acceptable tests that were lower quality by reproducibility criteria were excluded and the SNP—FEV_1 association was assessed in the subset. The direction and magnitude of all effect estimates were similar to the full sample results (data not shown).

Replication in Framingham Heart Study

The Framingham Heart Study (FHS) cohort (n=7,694; includes individuals from the original, offspring, and third generation cohorts) was used as a replication cohort to examine the genes with cross-sectional SNP associations (nominal $p<0.02$) in Health ABC European-Americans. FHS genotyping used the Affymetrix GeneChip Human Mapping 500K Array and 50K Human Gene Focused Panel(15).

eQTL Analysis:

Data Collection and Statistical Approach

Associations between SNPs and gene expression of 13 vitamin D-responsive genes in lung small airway epithelium tissue were analyzed. Tissue samples were taken from a diverse cohort of smokers and non-smokers of different genders and ancestries (see Table 1, Gao *et al*(38)). Details of the sample collection are published elsewhere,(14) and details on normalization of gene expression values are available in Gao *et al.*(38) SNPs were assayed using Affymetrix 500k arrays, which provided data on 191,959 genotypes; only SNPs with MAF of > 0.1 were analyzed for associations with gene expression. Thus, there were far fewer SNPs available in

the eQTL study in comparison to the Health ABC GWAS study, and although very few of the exact SNPs studied in Health ABC were in the eQTL database, the eQTL SNPs tagged the sequence variation in each gene.

SNPs within 100kb of the 13 candidate genes (**Supplemental Table 3.7** for gene names) were tested for association with gene expression using PLINK v1.07. Quantile-quantile plots were generated in R and Locus Zoom(39) plots were generated to visually examine P-value distributions. Genome-wide Q-Q plot and Manhattan plot were also examined.

Supplemental Table 3.6 Characteristics of 26 Non-smoking Human Volunteers in the Gene Expression Study, by Tertile of Serum 25-Hydroxyvitamin D Concentration

	Serum 25-Hydroxyvitamin D		
Variable	**Tertile I (n=9)**	**Tertile II (n=9)**	**Tertile III (n=8)**
Serum 25-OH-D, ng/mL (range)	8.99 (2.3 - 11.8)	20.9 (12.7 - 26.7)	33.3 (27.9 - 39.7)
Age, years (median)	36.9 (38)	44.1 (45)	50.6 (46.5)
Males (%)	6 (67%)	6 (67%)	7 (87%)
Race/Ethnicity (%)			
African American	5 (56%)	6 (67%)	1 (13%)
European	1 (11%)	3 (33%)	7 (87%)
Hispanic	2 (22%)	0 (0%)	0 (0%)
Asian	1 (11%)	0 (0%)	0 (0%)

*mean (standard deviation), unless noted

Supplemental Table 3.7 The Distribution of Studied SNPs in Thirteen Vitamin D-Responsive Genes for European- and African American Ancestry Groups in the Health ABC Cohort Study

Gene	Chromosomal Position	EntrezGene ID	Size (bp)*	# SNPs in European-Americans**	# SNPs in African Americans**
CST6	11q13	1474	7513	2	2
DAPK1	9q34.1	1612	216792	124	121
DTX4	11q12.1	23220	42200	11	10
EMB	5q11.1	133418	48724	11	11
FSTL1	3q13.33	11167	62698	22	24
KAL1	Xp22.32	3730	209311	35	47
KCNS3	2p24	3790	60279	22	25
KLF4	9q31	9314	10620	1	1
PTGER2	14q22	5732	20207	24	37
RSAD2	2p25.2	91543	26567	9	11
SGPP2	2q36.1	130367	140294	40	46
SLITRK6	13q31.1	84189	12561	7	9
TMEM40	3p25.2	55287	31416	5	11
				Total = 313	355

*Includes 3,000 bp at 3' and 5' ends of gene

** SNPs filtered for Minor Allele Frequency and Hardy-Weinberg Equilibrium

Supplemental Table 3.8 Gene Ontology of Thirteen Nominally Significant Candidate Genes from the UniProtKb-GOA Database (http://www.ebi.ac.uk/QuickGO/)

Gene	Gene Name	Function(s)	Pathway(s)	Location(s)
CST6	Cystatin E/M	cysteine-type endopeptidase inhibitor	anatomical structure morphogenesis	cornified envelope, extracellular region
DAPK1	Death-associated protein kinase 1	ATP and calmodulin binding	intracellular protein kinase cascade, apoptosis regulation	actin cytoskeleton
DTX4	Deltex homolog 4	zinc ion binding	Notch signaling pathway	cytoplasm
EMB	Embigin	N/A	cell adhesion	integral membrane protein
FSTL1	Follistatin-like 1	calcium ion binding, heparin binding	Bone morphogenetic protein signaling pathway	Extracellular space
KAL1	Kallmann syndrome 1 sequence	extracellular matrix structural component, serine-type endopeptidase inhibitor	axon guidance, chemotaxis, cell movement, cell adhesion	cell surface, extracellular space
KCNS3	Potassium voltage-gated channel, delayed-rectifier, subfamily S, member 3	delayed-rectifier potassium channel	potassium ion transport, regulation of insulin secretion	Golgi and plasma membrane
KLF4	Kruppel-like factor 4	transcription repressor activity	regulation of cell proliferation, mesodermal cell fate determination	nuclear
PTGER2	Prostaglandin E receptor 2 (subtype EP2)	G protein coupled receptor for prostaglandin E	GPCR signaling, regulation of cell proliferation	integral to plasma membrane
RSAD2	Radical S-adenosyl methionine domain containing 2	iron-sulfur cluster binding, metal ion binding	defense response to virus	endoplasmic reticulum
SGPP2	Sphingosine-1-phosphate phosphatase 2	sphingosine-1-phosphate phosphatase activity	sphingosine metabolic process	endoplasmic reticulum membrane
SLITRK6	SLIT and NTRK-like family, member 6	N/A	axonogenesis	integral membrane protein
TMEM40	Transmembrane protein 40	N/A	N/A	integral membrane protein

Supplemental Table 3.9 Gene-Level Replication of Health ABC European-American SNP Associations with the FEV_1 phenotype in the Framingham Heart Study Cohort

Gene	Total # of FHS SNPs	# SNPs with p<0.05	Most Significant SNP in Gene			
			RS#	MAF (%)	Beta (mL)*	Nominal P
SGPP2	145	23	rs10932956	21	29.2	2.23×10^{-02}
DAPK1	340	23	rs7025760	21	23.6	1.86×10^{-02}

*all models use additive genetic coding

Supplemental Table 3.10 The most statistically significant associations (nominal $P<2.0x10^{-02}$) between single nucleotide polymorphisms in vitamin D-responsive genes and FEV_1 for a) European-Americans and b) African Americans (all SNPs, including redundant SNPs are shown).

b) European-Americans

Gene	RS#	Chr	Coded Allele	Coded Allele Freq. (%)	β (mL)	SE (mL)	Nominal *P*	Model
DAPK1	rs11141878	9	A	36	-103.98	36.33	$4.26x10^{-03}$	R
	rs4877361	9	G	14	72.47	27.36	$8.17x10^{-03}$	D
	rs17477827	9	A	14	70.66	27.33	$9.82x10^{-03}$	D
	rs4878089	9	A	46	39.68	16.93	$1.92x10^{-02}$	A
SGPP2	rs4674656	2	A	25	-58.70	19.67	$2.88x10^{-03}$	A

b) African Americans

Gene	RS#	Chr	Coded Allele	Coded Allele Freq. (%)	β (mL)	SE (mL)	Nominal *P*	Model
DAPK1	rs3128491	9	G	33	51.48	21.44	$1.65X10^{-02}$	A
FSTL1	rs4676781†	3	T	8	-110.13	35.34	$1.88X10^{-03}$	A
	rs13100865†	3	G	9	-105.96	35.02	$2.54X10^{-03}$	A
	rs13097755†	3	T	28	-60.46	21.59	$5.20X10^{-03}$	A
	rs2272515†	3	C	28	-60.46	21.59	$5.20X10^{-03}$	A
KAL1	rs6530200	23	T	47	-45.28	16.81	$7.20X10^{-03}$	A
	rs974655	23	A	49	79.22	30.33	$9.14X10^{-03}$	D
KCNS3	rs3747515	2	T	16	243.92	83.47	$3.56X10^{-03}$	R
	rs1031771	2	A	16	243.76	83.52	$3.60X10^{-03}$	R
RSAD2	rs4669114	2	G	10	-119.55	36.20	$9.93X10^{-04}$	D
	rs10495546	2	C	10	-119.08	36.18	$1.03X10^{-03}$	D
	rs4669113	2	C	10	-119.08	36.18	$1.03X10^{-03}$	D
	rs6431837	2	C	47	-101.06	33.55	$2.66X10^{-03}$	R
	rs7570384	2	C	38	-55.35	20.05	$5.88X10^{-03}$	A
	rs4669111	2	A	41	-49.75	20.07	$1.34X10^{-02}$	A
SGPP2	rs4528748†	2	C	27	-209.95	54.10	$1.11X10^{-04}$*	R
	rs7556867†	2	G	27	-207.89	54.48	$1.44X10^{-04}$*	R
	rs6758392†	2	T	28	-182.36	51.94	$4.67X10^{-04}$*	R

Abbreviations: Chr=chromosome; Freq=frequency; β=beta coefficient; SE=standard error; A=additive model; D=Dominant model; R=recessive model

Model adjusted for age, height, smoking, gender, study site, and ancestry principal components.

* = FDR q-value $<5.0x10^{-02}$

† SNP is nominally significant ($P<2.0x10^{-02}$) for both FEV_1 and FEV_1/FVC phenotypes in African Americans

Supplemental Table 3.11 The most statistically significant associations (nominal P<$2.0x10^{-02}$) between single nucleotide polymorphisms in vitamin D-responsive genes and the FEV_1/FVC ratio for a) European-Americans and b) African Americans in the Health ABC cohort

a) European-Americans

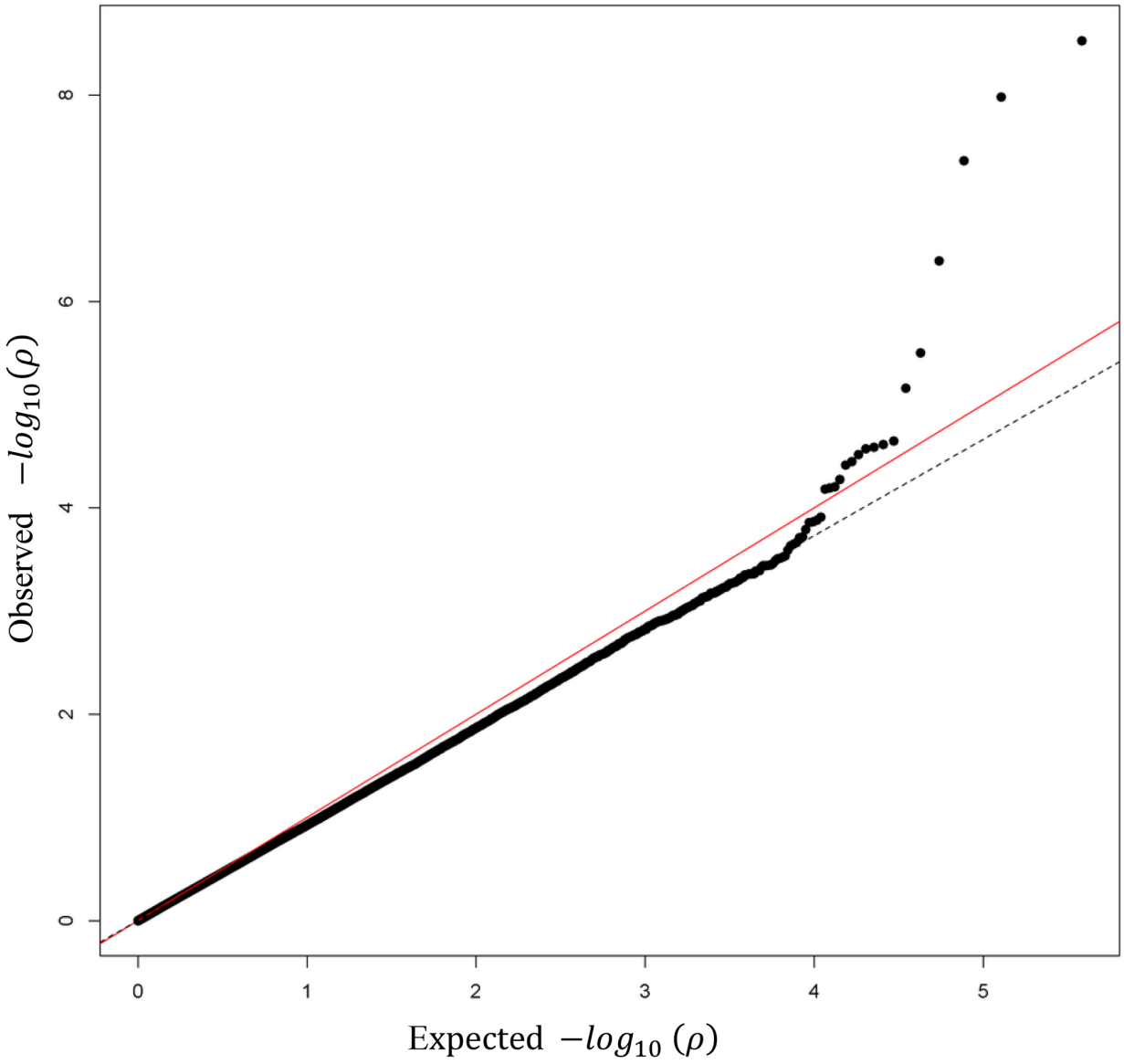
Gene	RS#	Chr	Coded Allele	Coded Allele Freq. (%)	β (%)	SE (%)	Nominal P	Model
KLF4	rs2236599	9	A	19	-0.85	0.33	$1.15x10^{-02}$	A

b) African Americans

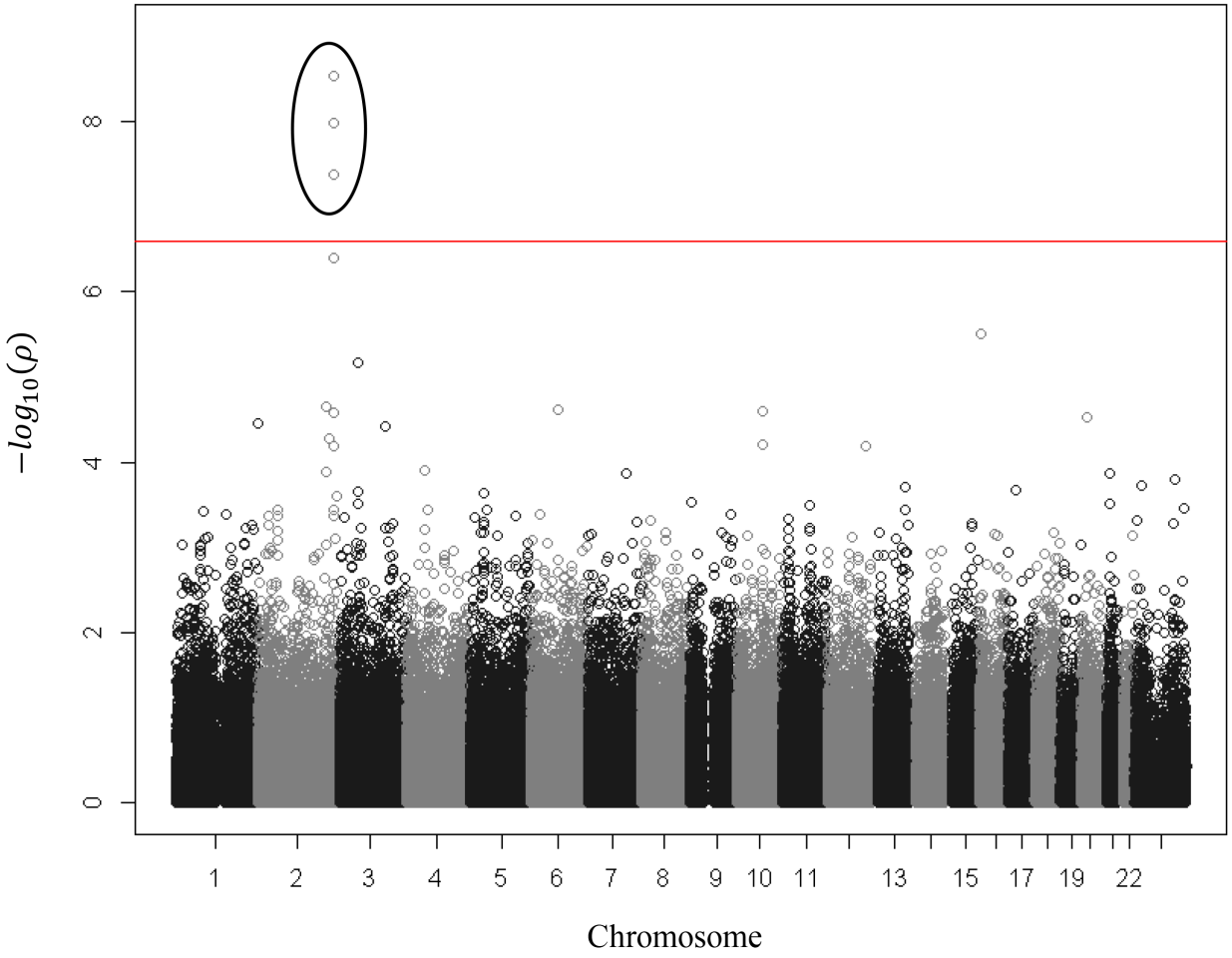
Gene	RS#	Chr	Coded Allele	Coded Allele Freq. (%)	β (%)	SE (%)	Nominal P	Model
FSTL1	rs4676781	3	T	8	-1.92	0.67	$4.47x10^{-03}$	A
	rs13100865	3	G	9	-1.81	0.67	$6.65x10^{-03}$	A
	rs13097755	3	T	28	-1.03	0.41	$1.27x10^{-02}$	A
	rs2272515	3	C	28	-1.03	0.41	$1.27x10^{-02}$	A
KAL1	rs5933668	23	T	19	-2.76	0.86	$1.32x10^{-03}$	R
	rs1859867	23	C	40	0.95	0.32	$3.13x10^{-03}$	A
	rs1079854	23	G	38	1.83	0.63	$3.65x10^{-03}$	R
	rs2108400	23	C	38	1.81	0.63	$4.11x10^{-03}$	R
	rs4830593	23	A	38	1.80	0.63	$4.22x10^{-03}$	R
	rs16998683	23	C	12	1.82	0.70	$9.75x10^{-03}$	D
SGPP2	rs4597517	2	A	23	-3.30	1.21	$6.71x10^{-03}$	R
	rs4528748	2	C	27	-1.15	0.43	$7.70x10^{-03}$	A
	rs7556867	2	G	27	-1.15	0.43	$8.18x10^{-03}$	A
	rs6758392	2	T	28	-1.10	0.43	$1.01x10^{-02}$	A

Abbreviations: Chr=chromosome; Freq=frequency; β=beta coefficient; SE=standard error
Model adjusted for age, height, smoking, gender, study site, and ancestry principal components.
A=additive model, D=Dominant model, R=recessive model

Supplemental Figure 3.3 Genome-wide Quantile-Quantile Plot for *SGPP2* eQTL findings (shows results for all genotyped SNPs).



Supplemental Figure 3.4 Genome-wide Manhattan Plot for *SGPP2* eQTL findings. Red line: Bonferroni *P*-value for 191,959 markers.



CHAPTER 4

25-HYDROXYVITAMIN D STATUS AND GENETIC VARIATION IN THE VITAMIN D METABOLIC PATHWAY IN ASSOCIATION WITH FEV_1 IN THE FRAMINGHAM HEART STUDY

JG Hansen[1], W Gao[2], J Dupuis[2,3], GT O'Connor[3,4], W Tang[1], M Kowgier[5], A Sood[6], S Gharib[7], L Palmer[5,8,9], M Fornage[10,11], S Heckbert[12,13,14], B Psaty[12,13,14,15], SUNLIGHT[16], and PA Cassano[1,17]

[1] Division of Nutritional Sciences, Cornell University, Ithaca, NY 14853 USA

[2] Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA

[3] The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA, USA

[4] Pulmonary Center, Department of Medicine, Boston University School of Medicine, Boston, MA, USA

[5] Ontario Institute for Cancer Research, Toronto, Ontario, Canada

[6] University of New Mexico, NM, USA

[7] Computational Medicine Core, Center for Lung Biology, Division of Pulmonary & Critical Care Medicine, Department of Medicine, University of Washington, Seattle, WA, USA

[8] Epidemiology and Obstetrics & Gynaecology, University of Toronto, Toronto, Ontario, Canada

[9] Samuel Lunenfeld Research Institute, Toronto, Ontario, Canada

[10] Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX, USA

[11] Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA

[12] Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA

[13] Department of Epidemiology, University of Washington, Seattle, WA, USA

[14]Group Health Research Institute, Group Health Cooperative, Seattle, WA, USA

[15] Department of Medicine, University of Washington, Seattle, WA, USA

[16] The SUNLIGHT Consortium (Study of Underlying Genetic Determinants of Vitamin D and Highly Related Traits)

[17] Department of Public Health, Division of Biostatistics and Epidemiology, Weill Cornell Medical College, New York, NY, USA

Address correspondence to: Dr. Patricia A. Cassano, Division of Nutritional Sciences, Cornell University, 209 Savage Hall, Ithaca, NY 14853, pac6@cornell.edu, (607) 255-7551 (ph), (607) 255-2691 (fax).

ABSTRACT

Background Strong cross-sectional vitamin D—lung function associations have stimulated interest, but provide weak evidence for causal inference. Investigations of 25(OH)D and rate of change in lung function are needed.

Methods Using a linear mixed-effects model, we investigated genetic variants in vitamin D metabolic genes, hypothesized to influence usual 25(OH)D status, in relation to rate of change in forced expiratory volume in the first second (FEV_1) in 3,230 Framingham Heart Study (FHS) Offspring participants. We also estimated the 25(OH)D—rate of change in FEV_1 association in FHS Third Generation participants.

Results Variants in four vitamin D metabolic genes were associated with FEV_1 rate of change ($P_{nominal}<0.05$). The associations showed consistent direction of effect in a meta-analyzed set of 4 independent cohorts ($P_{Binominal}=0.06$). 2 SNPs, rs11819875 (*CYP2R1)* and rs842999 (*GC)*, were close to the replication significance threshold; both SNPs were statistically significant in a meta-analysis including FHS. For 3 of 4 SNPs, the SNP—serum 25(OH)D association in SUNLIGHT was consistent with FEV_1 associations. While cross-sectional 25(OH)D—lung function associations were replicated, there was little or no association with rate of change in FEV_1 in FHS Third Generation participants ($P=0.97$).

Conclusions SNP markers of 25(OH)D status were associated with rate of change in FEV_1, supporting a causal role for vitamin D in lung health during aging. The Third Generation study was comprised of healthy, middle-aged adults with sufficient serum 25(OH)D; lack of an association between 25(OH)D and rate of change in FEV_1 highlights the importance of background nutriture on these associations.

INTRODUCTION

Decreased lung function due to airflow obstruction is the primary characteristic of chronic obstructive pulmonary disease (COPD), the 3rd leading cause of mortality in the United States (1). Vitamin D status, assessed via the circulating serum biomarker 25-hydroxyvitamin D [25(OH)D], plays a well-known role in bone health, and is also associated with non-skeletal outcomes including lung function (2, 3). National surveys estimate that over 30% of Americans are at risk for insufficient vitamin D (defined as serum 25(OH)D <20 ng/mL) (3, 4). Vitamin D is obtained by sun exposure and diet (2), and genome-wide association studies (GWAS) have identified single nucleotide polymorphisms (SNPs) in vitamin D metabolic genes that are significantly associated with serum 25(OH)D concentrations (5, 6).

The active vitamin D metabolite, $1,25OH_2D$, is constitutively synthesized from 25(OH)D in lung epithelial cells *in vitro* (7) and is involved in biological processes critical to lung function including inflammation and airway remodeling (8-10). Several cross-sectional, population-based observational studies have demonstrated strong, positive associations between vitamin D and lung function (11-13), although one study in the Hertfordshire cohort did not replicate cross-sectional associations (14). Additionally, vitamin D deficiency is common in COPD patients (15), higher vitamin D is associated with reduced risk of respiratory infections (13, 16), and high-dose vitamin D supplementation reduced COPD exacerbations in patients with severe vitamin D deficiency (17). The few existing reports of longitudinal associations between vitamin D and lung function have been inconclusive; an observational study in COPD patients reported no association between serum 25(OH)D and longitudinal lung outcomes (18), but a recent population-based study in an elderly male cohort reported steeper lung function decline in current smokers with serum 25(OH)D ≤ 20 ng/mL compared to smokers with higher 25(OH)D

(19). Genetic variants in the vitamin D binding protein, encoded by the *GC* gene, are associated with COPD risk (15, 20-23), and *GC* may be an important mediator of hypothesized vitamin D effects on lung function (24).

We investigated the association of vitamin D with lung function outcomes in two generational cohorts of the Framingham Heart Study (FHS). First, we investigated single nucleotide polymorphisms (SNPs) in vitamin D metabolic genes in association with rate of change in FEV_1 in FHS Offspring participants. To strengthen our analysis, we pursued replication in four independent cohorts, and also investigated the SNPs in relation to serum 25(OH)D status in the SUNLIGHT consortium. Second, we investigated the association of serum 25(OH)D with cross-sectional and rate of change in FEV_1 in a subset of the FHS Offspring and in FHS Third Generation participants. As previous studies were limited by consideration of smokers, restricted age groups, or males only, this study provides a comprehensive exploration of vitamin D associations with FEV_1 in a healthy, adult population-based sample including both males and females.

METHODS

Study Population and Ethics

Study participants were from the Offspring and Third Generation cohorts of the Framingham Heart Study (FHS), a longitudinal family-based study established in 1948 in Framingham, MA. The FHS Offspring cohort, consisting of original cohort offspring and their spouses, began in 1971 (25). The FHS Third Generation cohort was initiated in 2002, enrolling children of the Offspring cohort (26). Self-reported ethnicity across all FHS cohorts was >99% Caucasian (26).

3,230 FHS Offspring participants (63% of all Offspring) with genotype data and spirometry measurements from Exams 5-8 were included in SNP—rate of change in FEV_1 analyses. 1,435 FHS Offspring participants (28% of all Offspring) with serum 25(OH)D measured between Exams 6 (1995-1998) and 7 (1998-2001) and *subsequent* spirometry measurements (from Exams 6, 7, or 8) were available for the serum 25(OH)D—FEV_1 analyses. FHS Third Generation participants (N=3,599; 88% of full cohort) with serum 25(OH)D measurements from Exam 1 (2002-2005) and spirometry measurements from Exams 1 and 2 (2008-2010) were included in serum 25(OH)D— FEV_1 analyses (**Supplemental Figure 1**).

All study participants provided written informed consent for this study, and local institutional review boards approved the study protocols.

Measures

Genotyping was performed using the Affymetrix 500K SNP array with a supplemental Affymetrix 50K gene-focused array. Genotyping and imputation methods are described in detail elsewhere (6, 27).

241 imputed SNPs in six candidate genes with well-established roles in vitamin D metabolism and transport [*CYP24A1, CYP27A1, CYP27B1, CYP2R1, DHCR7/NADSYN1* (these two genes considered jointly as a candidate genomic locus due to prior GWAS associations), and *GC*] were analyzed (+/- 5KB region on either end of genes included; **Supplemental Table 4.6** for details).

In the FHS Offspring 5^{th}, 6^{th}, and 7^{th} examinations, spirometry was performed using a Collins Survey II spirometer (Collins Medical, Inc. Braintree, MA) calibrated daily and connected to a computer running software developed by S&M Instruments, Doylestown, PA (28, 29). For Offspring Exam 8 and Third Generation Exams 1 and 2, spirometry was performed

using a Collins CPL system (nSpire Health Inc., Longmont, CO) calibrated daily (29). Acceptable pulmonary function test measurements (as defined by American Thoracic Society standards (30)) were used.

Serum 25(OH)D was assayed separately in the Offspring and Third Generation cohorts using radioimmunassay (DiaSorin Inc, Stillwater, MN, USA) (6, 31), and log-transformed values were used in all analyses. Offspring serum samples for 25(OH)D assays were collected between 1998-2001 (31) and Third Generation serum samples between 2001-2005 (6). The Diasorin RIA assay was reformulated in 1998; however, all FHS samples were analyzed after 1998, so assay drifts due to the reformulated RIA assay, described for the NHANES data (32), do not affect 25(OH)D measurements in Framingham. In addition to the RIA assay reformulation, drifts in assay performance were noted in NHANES, affecting the comparability of NHANES 25(OH)D measurements between 2003-2006. While it is not known if similar assay variation affects comparability between FHS Offspring and Third Generation serum 25(OH)D measurements, the effect of the assay drift was relatively small (statistical adjustment of mean 25(OH)D for assay drift in NHANES 2003-2004 and 2005-2006 resulted in mean 25(OH)D differences of 1-2 ng/mL(33)).

Statistical Analysis

Linear mixed effects models were used for all analyses in R (version 2.15.3), with adjustment for FHS family structure. In SNP—rate of change in FEV_1 analyses, the coefficient of interest was the interaction of SNP x time (time elapsed between each FEV_1 measurement and baseline), which estimated the effect of genotype on rate of change in FEV_1. In serum 25(OH)D—rate of change in FEV_1 analyses, the coefficient of interest was the interaction of 25(OH)D x time, which estimated the effect of serum 25(OH)D on rate of change in FEV_1. All

models were adjusted for baseline age, gender, height, current smoking status, smoking pattern during follow-up and the interaction of smoking pattern x time, and baseline pack-years. Genetic models were further adjusted for the first two ancestry principal components to account for population substructure. Serum 25(OH)D— FEV_1 models were further adjusted for month of 25(OH)D measurement, body mass index (BMI), and FHS cohort (cross-sectional analysis only). Smoking pattern was defined as: persistent smoker (current smoker, all time points during follow-up), intermittent smoker (current smoker at ≥ 1 time point), former smoker (former smoker, all time points), and never smoker (never smoker, all time points). The cross-sectional serum 25(OH)D—FEV_1 association was estimated using the coefficient for serum 25(OH)D from the above-described models. Smoothing spline analyses evaluated log-transformed 25(OH)D by residual FEV_1, after adjustment for described covariates to examine linearity of the association.

SNPs associated with rate of change in FEV_1 at $P<0.05$ in FHS were assessed for replication in four independent cohorts with ≥ 3 FEV_1 measurements per participant (Online Supplement for details). The combined replication cohorts included a total of 10,476 participants and 32,054 spirometry observations. The FHS results were meta-analyzed with the other four cohorts to provide final overall estimates.

The SNPs associated with rate of change in FEV_1 at $P<0.05$ were further evaluated for association with serum 25(OH)D using data from the SUNLIGHT consortium, which comprised 33,996 individuals of European ancestry, including FHS participants, with GWAS data for the serum 25(OH)D phenotype (6).

RESULTS

Average serum 25(OH)D in FHS Offspring participants was 18.0 ng/mL, compared to 34.5 ng/mL in the Third Generation; similarly, the proportion of FHS participants at risk of vitamin D deficiency (defined as serum 25(OH)D <12 ng/mL) was 14.4% and 1.2% in the Offspring and Third Generation, respectively (**Table 4.1**). As expected, Offspring participants were older, had lower baseline FEV_1, had slightly higher body mass index, and were more likely to be former smokers compared to the Third Generation participants. Average spirometry follow-up time for Offspring and Third Generation participants with serum 25(OH)D measurements ranged from 6-7 years (considering only spirometry measurements subsequent to serum 25(OH)D measurement); average spirometry follow-up time for Offspring participants contributing to the SNP—rate of change in FEV_1 analysis was 14.7 years because this analysis used all available spirometry data.

Vitamin D Metabolic Gene SNPs and rate of change in FEV_1

We explored the association of SNPs in vitamin D metabolic genes with rate of change in FEV_1, and found that SNPs in 4 genes, namely *CYP27B1, CYP2R1, DHCR7,* and *GC*, were associated at a nominal P <0.05 (most significant SNP per gene presented in **Table 4.2**; full results in **Supplemental Table 4.7**). The most significant SNP association in *DHCR7* was for rs1790349; the minor allele of this SNP was associated with 2.2 mL/year steeper FEV_1 decline (P=0.0015). The most significant SNP associations in *CYP27B1* and *CYP2R1* were for rs10877013 and rs11819875, respectively; the minor alleles of these SNPs were also associated with steeper FEV_1 decline. However, the minor allele of rs842999, the most significant SNP in *GC,* was associated with about 1 mL/yr attenuation in rate of change in FEV_1. No SNPs in *CYP24A1* and *CYP27A1* met the threshold for statistical significance (P<0.05), and thus these genes were not considered further.

We explored replication of the most significant SNP in *CYP27B1, CYP2R1, DHCR7,* and *GC* in a meta-analyzed set of four independent cohorts, namely the Health, Aging and Body Composition Study (Health ABC), the Coronary Artery Risk Development in Young Adults Study (CARDIA), the Busselton Health Study (BHS), and the Cardiovascular Health Study (CHS), using a *P*-value threshold of $p<0.0125$ (representing the Bonferroni correction for 4 tests with overall $\alpha = 0.05$). Although no SNPs reached the replication threshold, all four SNPs had a direction of effect that was consistent with findings in FHS (*P*-value for binomial test = 0.06), and rs11819875 (*CYP2R1)* and rs842999 (*GC*) were close to the statistical threshold (*P*-values <0.09; **Table 4.2**). A second analysis, combining the 4 independent cohorts with the FHS data, thus combining data across all five cohorts, showed that all four SNPs were associated with rate of change in FEV_1 at $P < 0.05$, and rs11819875 ($P=2.20 \times 10^{-3}$) and rs842999 ($P=7.48 \times 10^{-3}$) had the strongest evidence for association (**Figure 4.1; Table 4.2** and **Supplemental Table 4.8** for replication results).

We investigated the association of the four above-mentioned SNPs with serum 25(OH)D in the SUNLIGHT consortium. The minor alleles of rs10877013, rs1790349, and rs11819875 were associated with both *steeper* FEV_1 decline and *lower* 25(OH)D. However, for rs842999 in *GC,* the minor allele was associated with an attenuated rate of FEV_1 decline and a lower serum 25(OH)D **(Table 4.3).**

25(OH)D Associations with Cross-sectional and Longitudinal FEV_1

25(OH)D had a positive association with cross-sectional FEV_1 in the combined sample of Offspring and Third Generation participants, such that a 1-unit increase in log-transformed 25(OH)D was associated with a 45 mL increase in FEV_1 (*P*=0.0035) **(Table 4.4)**. A consistent direction of association was observed when modeling vitamin D as a dichotomous variable at

thresholds of <12 ng/mL (considered risk of vitamin D deficiency) or <20 ng/mL (considered risk of vitamin D inadequacy), but coefficients for the dichotomous serum vitamin D variables did not reach the significance threshold of $P<0.05$. Visual inspection of the serum 25(OH)D—FEV_1 association in the spline analysis revealed an approximately linear positive association in the range of serum 25(OH)D < 12 ng/mL; in the 12-40 ng/mL range, the serum 25(OH)D—FEV_1 association was positive and linear, but attenuated. However, the serum 25(OH)D—FEV_1 association reached a plateau above a threshold of about 40 ng/mL (**Figure 4.2**).

In longitudinal models including both Offspring and Third Generation participants to assess the serum 25(OH)D association with rate of change in FEV_1, a highly significant association of cohort with rate of change in FEV_1 was observed ($P=1.23x10^{-30}$). This association indicated that the Offspring cohort had a less steep rate of change in FEV_1 compared to the Third Generation cohort, an unexpected finding given Offspring participants are about 20 years older than Third Generation participants. Thus, the average rate of change in FEV_1 in Offspring participants with serum 25(OH)D measured was -14 mL/year; in comparison, in the full Offspring cohort, using all available data (i.e., not limited to participants with serum 25(OH)D measurement nor to lung function measurements subsequent to the serum assay), the average rate of change in FEV_1 was -26 mL/year (personal communication). This definitive evidence of significant selection bias in the 28% subset of Offspring participants with 25(OH)D measurements precluded further analysis of longitudinal 25(OH)D—FEV_1 associations in the Offspring.

There was little or no association between serum 25(OH)D and rate of change in FEV_1 in the Third Generation cohort ($P = 0.97$; **Table 4.5).** The average rate of decline in the Third

Generation participants was -28 mL/year, similar to the rate of decline in the Offspring participants included in the genetic analyses.

DISCUSSION

In this population-based cohort study, we investigated associations between vitamin D and both cross-sectional and rate of change in FEV_1, considering genetic variants in vitamin D metabolic genes and serum 25(OH)D as exposures. SNPs in four vitamin D metabolic genes, *CYP2R1, CYP27B1, DHCR7,* and *GC*, were associated with rate of change in FEV_1 in the FHS Offspring cohort, and the most significant SNP from each gene showed evidence of replication in a meta-analysis of four independent cohort studies. In three genes, SNPs associated with steeper FEV_1 decline in FHS were also associated with lower 25(OH)D in the SUNLIGHT consortium, supporting the hypothesis that SNPs influencing usual 25(OH)D status are associated with lung function. SNPs in vitamin D metabolic genes are randomly assigned at conception and are hypothesized to reflect usual serum 25(OH)D status; thus, investigating genetic variants in association with pulmonary function reduces potential bias from lifestyle confounding and reverse causality.

We demonstrated an association of serum 25(OH)D with cross-sectional FEV_1 consistent with previously published studies. Serum 25(OH)D was not associated with FEV_1 decline in FHS Third Generation participants, which is likely due to the fact that > 90% of Third Generation participants had sufficient serum 25(OH)D [defined as 25(OH)D ≥ 20 ng/mL]; conversely, only 8.6% and 1.2% were considered to be at risk of inadequacy (<20 ng/mL) and deficiency (< 12 ng/mL), respectively (thresholds defined by the IOM (3)). These findings support the assertion that longitudinal associations between 25(OH)D and lung function are non-

linear, and may be limited to individuals with low 25(OH)D. Neither was there any evidence for a differential effect by smoking status (19). We were unable to estimate serum 25(OH)D—rate of change in FEV_1 associations in the Offspring cohort given selection bias. An ideal study would investigate the association between SNPs in vitamin D metabolic genes, serum 25(OH)D, and FEV_1 in the same study population, but this was not feasible in the FHS data.

Rs11819875 in *CYP2R1* (*G* allele) was associated with steeper FEV_1 decline in FHS, had the strongest evidence for association in the meta-analysis of 4 replication cohorts, and was associated with lower serum 25(OH)D in SUNLIGHT. *CYP2R1* is a key hepatic 25-hydroxylase enzyme (34), and variants in this gene are consistently associated with 25(OH)D in GWAS (5, 6). Rs11819875 is located less than 2.5 KB away from rs10741657, the SNP in *CYP2R1* most strongly associated with 25(OH)D concentrations in SUNLIGHT (6), but these two SNPs are in low linkage disequilibrium (R^2=0.13).

Rs842999 in *GC* (*G* allele) was associated with an attenuated rate of change in FEV_1 in FHS and in the meta-analysis of 4 replication cohorts. Contrary to our hypothesis, the rs842999 *G* allele was associated with lower serum 25(OH)D in the SUNLIGHT consortium. Rs842999 is in strong linkage disequilibrium (R^2=0.9) with rs7041, a nearby non-synonymous SNP in *GC*. The rs842999 *G* allele has the same minor allele frequency as the rs7041 *T* allele, which was also associated with attenuated FEV_1 decline in FHS (*P*=0.08) and in the meta-analysis of 4 replication cohorts (*P*=0.04); furthermore, the rs7041 *T* allele was similarly associated with lower serum 25(OH)D in SUNLIGHT. The *GC* gene exists in three common isoforms, GC1F, GC1S, and GC2, based on the alleles present at rs7041 and rs4588, a second non-synonymous *GC* SNP. Importantly, the rs7041 *T* allele is part of both GC2 (rs7041 = *T* allele, rs4588 = *A* allele) and GC1F (rs7041 = *T* allele, rs4588 = *C* allele). GC2 was associated with lower risk of

COPD in several studies (implying attenuated rate of decline), while GC1F was associated with higher risk of COPD and a steeper rate of FEV_1 decline (20-23, 35). GC2 is associated with reduced macrophage activation compared to GC1 variants, providing a hypothesized mechanism for this effect (22), although, paradoxically, the GC2 alleles were associated with lower serum 25(OH)D in Caucasian populations (15, 22, 36, 37). Our finding for rs842999 could be tagging the effect of the GC2 haplotype, which is associated with both lower serum 25(OH)D levels and a protective effect on lung function. Because rs4588 was not imputed in FHS we cannot directly investigate GC2 frequency or the haplotype association with FEV_1; however, GC2 has a higher prevalence (vs. GC1f) in Caucasian populations (38, 39).

Both rs1790349 (*DHCR7)*, and rs10877013 (*CYP27B1)* had a consistent direction of effect on FEV_1 decline in FHS and in the meta-analysis of 4 replication cohorts, although the effect sizes were close to the null value in the replication. Rs1790349 was strongly associated with serum 25(OH)D concentration (5), and the *C* allele was associated with lower 25(OH)D in SUNLIGHT. *DHCR7* encodes the 7-dehydrocholesterol reductase enzyme, which converts pro-vitamin D to cholesterol, thus removing the substrate for endogenous 25(OH)D production(6). Similarly, the *T* allele of rs10877013 in *CYP27B1*, associated with steeper lung function decline in FHS, was associated with lower serum 25(OH)D in SUNLIGHT; *CYP27B1* is a 1-α-hydroxylase that converts 25(OH)D to the active vitamin D metabolite, $1,25(OH)_2D$, and is expressed in airway epithelial cells (7).

Our study has a number of strengths as well as several limitations worth mentioning. A major strength is use of the Framingham Heart Study, a large, population-based study including both male and female smokers and non-smokers. The FHS Offspring participants included in the genetic analysis had an average spirometry follow-up of 14.7 years leading to increased accuracy

of the rate of decline estimates. A limitation of this study is that serum 25(OH)D was measured only once; however, a recent study showed that 25(OH)D measurements 12 months apart had a correlation of 0.8 (40). As the follow-up period for serum 25(OH)D—FEV_1 analyses was 6-7 years, we assumed baseline 25(OH)D was a good approximation of 25(OH)D status throughout follow-up. While spirometry follow-up in the Third Generation cohort averaged 6.1 years, longer follow-up may be needed to identify associations between vitamin D and rate of change in FEV_1. A related limitation is that supplement use data was not available, thus we could not investigate whether low baseline 25(OH)D was associated with subsequent supplement use.

CONCLUSIONS

The SNP—FEV_1 findings show that genetic variants influencing usual 25(OH)D status are associated with rate of change in lung function over time, which is suggestive of a true association, although associations between genetic variants in *GC* and FEV_1 may not be straightforward. Further, these findings suggest that populations with a low prevalence of vitamin D inadequacy are unlikely to demonstrate serum 25(OH)D—lung function association, suggesting important study design considerations regarding potential to benefit.

Table 4.1 Baseline* population characteristics of Framingham Heart Study participants

	SNP—FEV_1 analysis	**25(OH)D—FEV_1 analysis**	
	Offspring Cohort (N=3,230)	Offspring Cohort (N=1,435)	Third Generation Cohort (N=3,599)
Follow-up duration, yr	14.7	7.2	6.1
N of FEV_1 measurements	11,275	3,093	6,493
FEV_1, L	3.0 (0.8)	2.7 (0.8)	3.6 (0.8)
Baseline age, yr	50.9 (10.3)	59.9 (9.2)	40.2 (8.7)
Male, %	47	48	47
Height, cm	165.5 (9.5)	168.0 (9.1)	170.6 (9.3)
Baseline pack-years	25.4 (21.3)	26.0 (22.7)	13.7 (14.2)
Current smokers**, %	24.6	12.8	15.2
Former smokers, %	39.8	50.8	27.0
BMI	Not available	28.0 (5.1)	26.9 (5.4)
25(OH)D†, ng/ml	Not available	18.0 (1.5)	34.5 (1.5)
N of 25(OH)D deficient (<12 ng/mL)	Not available	207 (14.4%)	44 (1.2%)
N of 25(OH)D insufficient (<20 ng/mL), %	Not available	801 (55.8%)	311 (8.6%)

* Baseline measurements for Offspring participants in SNP—FEV_1 analysis are from Exam 5. Baseline measurements for the Offspring participants included in 25(OH)D—FEV_1 analyses are from the exam closest to time of vitamin D measurement (either Exam 6 or 7). Baseline measurements for the Third Generation participants are from Exam 1.

**Current smokers at baseline; former smokers at all time points

† Geometric mean of 25(OH)D

Table 4.2 Association of the most statistically significant SNP per gene with the rate of change in FEV_1 in FHS and in the meta-analyzed replication cohorts (sorted by gene)

SNP	Chr	Position	Gene	Coded Allele*	Freq	FHS (N=3,230)			Replication cohorts, excluding FHS (N=7,246)			Replication cohorts, including FHS (N=10,476)		
						β	SE	*P*	β	SE	*P*	β	SE	*P*
rs10877013	12	56451352	*CYP27B1*	T	0.30	-1.3	0.5	0.0210	-0.4	0.5	0.3996	-0.7	0.4	0.0424
rs11819875	11	14873873	*CYP2R1*	G	0.18	-1.9	0.7	0.0043	-1.0	0.6	0.0851	-1.3	0.4	0.0022
rs1790349	11	70819998	*DHCR7*	C	0.15	-2.2	0.7	0.0015	-0.1	0.6	0.8869	-0.9	0.5	0.0447
rs842999	4	72830554	*GC*	G	0.44	+1.1	0.5	0.0300	+0.8	0.4	0.0784	+0.9	0.3	0.0075

Abbreviations: Chr = chromosome; SNP = single nucleotide polymorphism; β = beta coefficient for SNP x time effect; SE = standard error; *P* = P-value

Adjusted for: baseline age, gender, height, smoking pattern over follow-up and its interaction with time, baseline smoking pack-years, and the first two principal components for genetic ancestry

*Coded allele and frequency for the Framingham Heart Study (FHS). All effect estimates presented in terms of FHS coded allele. Coded allele frequencies between FHS and replication cohorts were nearly identical.

Table 4.3 Association of the most statistically significant SNP per gene with serum 25(OH)D in the SUNLIGHT Consortium

Gene	Chr	SNP	Coded Allele*	Freq	SUNLIGHT	
					Effect on 25(OH)D	*P* **
CYP27B1	12	rs10877013	T	0.30	Decrease	0.1067
CYP2R1	11	rs11819875	G	0.18	Decrease	0.1458
DHCR7	11	rs1790349*	C	0.15	Decrease	2.28×10^{-8}
GC	4	rs842999	G	0.44	Decrease	2.12×10^{-45}

Abbreviations: Chr = chromosome; SNP = single nucleotide polymorphism; β = beta coefficient for SNP x time effect; SE = standard error; *P* = P-value

*Coded allele and frequency for the Framingham Heart Study (FHS)

** *P*-value from combined SUNLIGHT discovery and replication cohorts

Table 4.4 Cross-sectional Association of 25(OH)D and FEV_1 (mL) in the Offspring and Third Generation cohorts, combined (N=5,034)

Model parameterization of vitamin D:	β	SE	*P*
Continuous log-transformed 25(OH)D	45.2	15.5	0.0035
At risk of vitamin D deficiency (<12 ng/mL) vs. not	-46.8	26.7	0.079
At risk of vitamin D inadequacy (<20 ng/mL) vs. not	-30.6	16.6	0.065

Abbreviations: β = beta coefficient; SE = standard error; *P* = P-value

Adjusted for: baseline age, gender, height, smoking pattern, current smoking status, baseline pack-years, FHS cohort, baseline BMI, and month of blood draw; all coefficients show expected direction of effect

Table 4.5 Association of 25(OH)D and Rate of Change in FEV_1 (mL/yr) in the Third Generation Cohort

	Third Generation Cohort (N=3,599)		
Model parameterization of vitamin D:	β	SE	*P*
Continuous log-transformed 25(OH)D	-0.06	1.7	0.97
At risk of vitamin D deficiency (<12 ng/mL) vs. not*	-0.02	6.9	0.997
At risk of vitamin D inadequacy (<20 ng/mL) vs. not	2.0	2.5	0.41

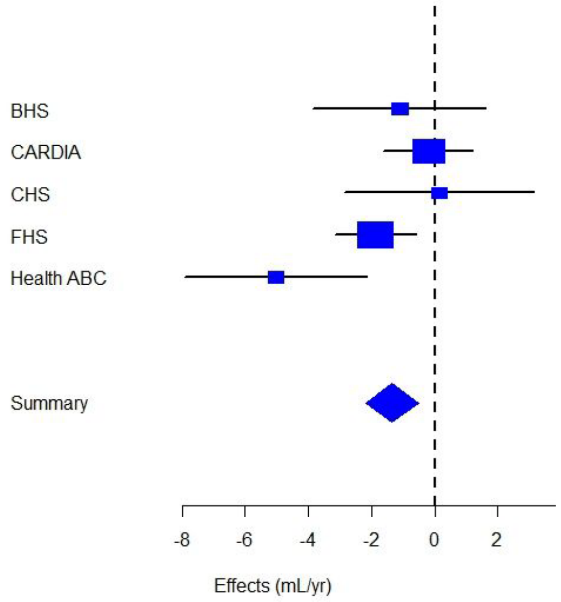
Abbreviations: β = beta coefficient; SE = standard error; *P* = P-value

*Interpretation: Third Generation participants at risk of vitamin D deficiency have a 0.02 mL/yr steeper rate of decline compared to Third Generation participants not at risk of deficiency

Adjusted for: baseline age, gender, height, smoking pattern over follow-up and its interaction with time, baseline pack-years, current smoking at each time point, BMI, and month of blood draw

Figure 4.1 Forest plots for rs11819875 and rs842999, where the size of the square for each study represents its contributing weight to the meta-analyzed replication results.

A) rs11819875



B) rs842999

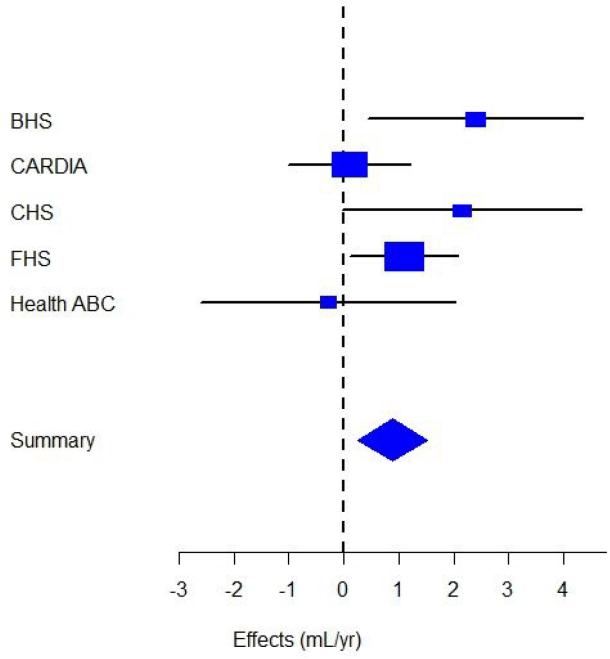
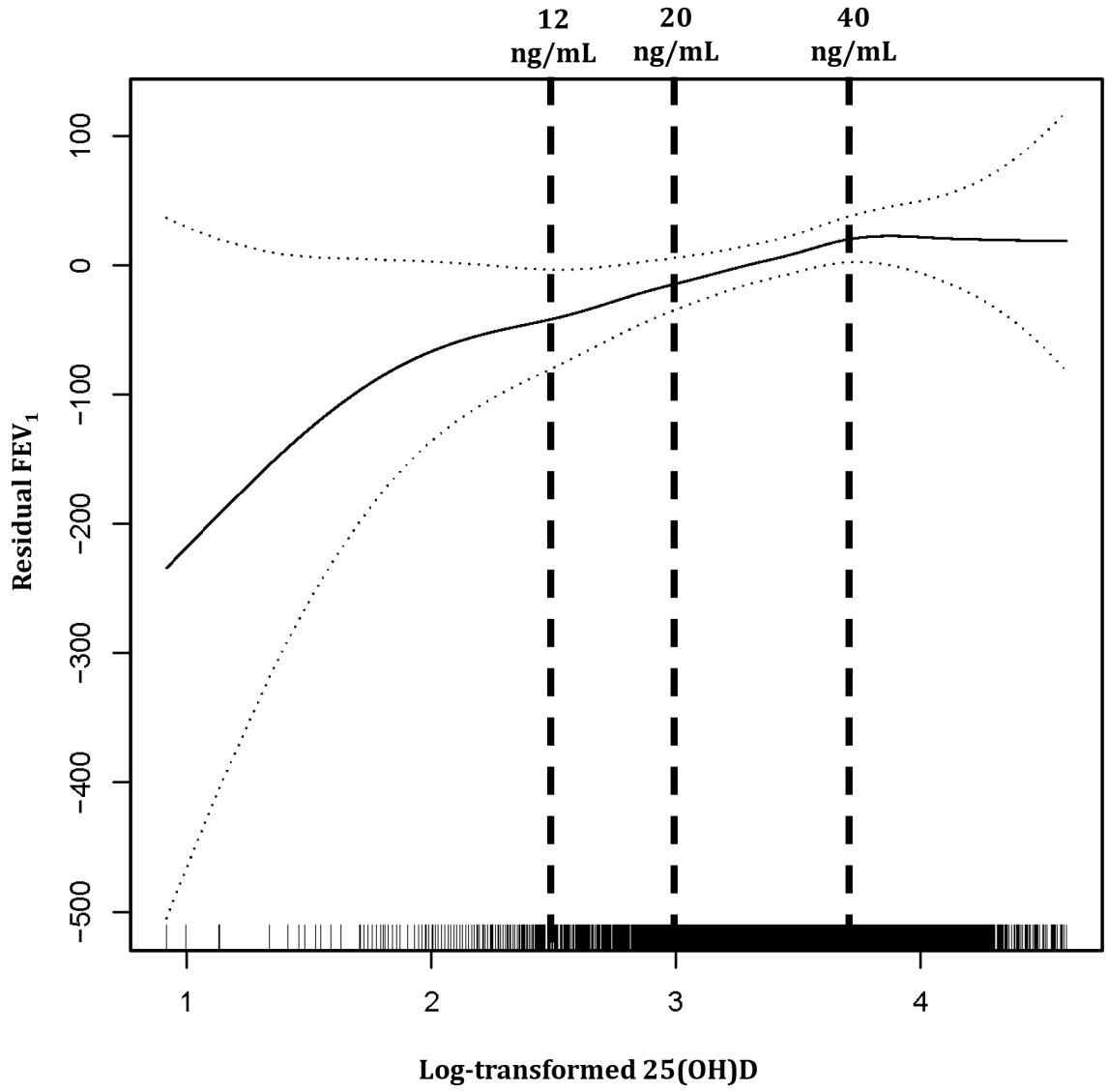


Figure 4.2 Spline analysis of log-transformed 25(OH)D by residual FEV_1



Residual FEV_1 values from model adjusted for baseline age, height, smoking status, baseline pack-years, BMI, FHS cohort, and month of blood draw

REFERENCES

1. Minino AM. Death in the united states, 2009. *NCHS Data Brief* 2011:1-8.

2. Holick MF. Vitamin d deficiency. *N Engl J Med* 2007;357:266-281.

3. Ross AC, Institute of Medicine (U. S.). Committee to Review Dietary Reference Intakes for Vitamin D and Calcium. Dietary reference intakes for calcium and vitamin d. Washington, DC: National Academies Press; 2011.

4. Looker AC JC, Lacher DA, et al. Vitamin d status: United states, 2001–2006. NCHS data brief, no 5. Hyattsville, MD: National Center for Health Statistic; 2011.

5. 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. *Hum Mol Genet* 2010;19:2739-2745.

6. Wang TJ, Zhang F, Richards JB, Kestenbaum B, van Meurs JB, Berry D, Kiel DP, Streeten EA, Ohlsson C, Koller DL, Peltonen 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 CL, Wolf M, Rice K, Goltzman D, Hidiroglou N, Ladouceur M, Wareham NJ, Hocking LJ, Hart D, Arden NK, Cooper C, Malik S, Fraser WD, Hartikainen AL, Zhai G, Macdonald HM, Forouhi NG, Loos RJ, 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 JO, Cauley JA, Uitterlinden AG, Gibson Q, Jarvelin MR, 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;376:180-188.

7. Hansdottir S, Monick MM, Hinde SL, Lovan N, Look DC, Hunninghake GW. Respiratory epithelial cells convert inactive vitamin d to its active form: Potential effects on host defense. *J Immunol* 2008;181:7090-7099.

8. Bikle DD. Vitamin d regulation of immune function. *Vitam Horm* 2011;86:1-21.

9. Hansdottir S, Monick MM. Vitamin d effects on lung immunity and respiratory diseases. *Vitam Horm* 2011;86:217-237.

10. Pfeffer PE, Hawrylowicz CM. Vitamin d and lung disease. *Thorax* 2012;67:1018-1020.

11. Black PN, Scragg R. Relationship between serum 25-hydroxyvitamin d and pulmonary function in the third national health and nutrition examination survey. *Chest* 2005;128:3792-3798.

12. Choi CJ SM, Choi WS, Kim KS, Youn SA, Lindsey T, Choi YJ, Kim CM. Relationship between serum 25-hydroxyvitamin d and lung function among korean adults in korea national health and nutrition examination survey (knhanes), 2008-2010. *J Clin Endocrinol Metab* 2013;98:1703-1710.

13. Berry DJ, Hesketh K, Power C, Hypponen E. Vitamin d status has a linear association with seasonal infections and lung function in british adults. *Br J Nutr* 2011;106:1433-1440.

14. Shaheen SO, Jameson KA, Robinson SM, Boucher BJ, Syddall HE, Aihie Sayer A, Cooper C, Holloway JW, Dennison EM. Relationship of vitamin d status to adult lung function and copd. *Thorax* 2011;66:692-698.

15. Janssens W, Bouillon R, Claes B, Carremans C, Lehouck A, Buysschaert I, Coolen J, Mathieu C, Decramer M, Lambrechts D. Vitamin d deficiency is highly prevalent in copd and correlates with variants in the vitamin d-binding gene. *Thorax* 2010;65:215-220.

16. Ginde AA, Mansbach JM, Camargo CA, Jr. Association between serum 25-hydroxyvitamin d level and upper respiratory tract infection in the third national health and nutrition examination survey. *Arch Intern Med* 2009;169:384-390.

17. Lehouck A, Mathieu C, Carremans C, Baeke F, Verhaegen J, Van Eldere J, Decallonne B, Bouillon R, Decramer M, Janssens W. High doses of vitamin d to reduce exacerbations in chronic obstructive pulmonary disease: A randomized trial. *Ann Intern Med* 2012;156:105-114.

18. Kunisaki KM, Niewoehner DE, Singh RJ, Connett JE. Vitamin d status and longitudinal lung function decline in the lung health study. *Eur Respir J* 2011;37:238-243.

19. Lange NE, Sparrow D, Vokonas P, Litonjua AA. Vitamin d deficiency, smoking, and lung function in the normative aging study. *Am J Respir Crit Care Med* 2012;186:616-621.

20. Ito I, Nagai S, Hoshino Y, Muro S, Hirai T, Tsukino M, Mishima M. Risk and severity of copd is associated with the group-specific component of serum globulin 1f allele. *Chest* 2004;125:63-70.

21. Schellenberg D, Pare PD, Weir TD, Spinelli JJ, Walker BA, Sandford AJ. Vitamin d binding protein variants and the risk of copd. *Am J Respir Crit Care Med* 1998;157:957-961.

22. Wood AM, Bassford C, Webster D, Newby P, Rajesh P, Stockley RA, Thickett DR. Vitamin d-binding protein contributes to copd by activation of alveolar macrophages. *Thorax* 2011.

23. Horne SL, Cockcroft DW, Dosman JA. Possible protective effect against chronic obstructive airways disease by the gc2 allele. *Human heredity* 1990;40:173-176.

24. Chishimba L, Thickett DR, Stockley RA, Wood AM. The vitamin d axis in the lung: A key role for vitamin d-binding protein. *Thorax* 2010;65:456-462.

25. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The framingham offspring study. Design and preliminary data. *Preventive medicine* 1975;4:518-525.

26. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, D'Agostino RB, Sr., Fox CS, Larson MG, Murabito JM, O'Donnell CJ, Vasan RS, Wolf PA, Levy D. The third generation cohort of the national heart, lung, and blood institute's framingham heart study: Design, recruitment, and initial examination. *Am J Epidemiol* 2007;165:1328-1335.

27. Artigas MS, Loth DW, Wain LV, Gharib SA, Obeidat M, Tang W, Zhai G, Zhao JH, Smith AV, Huffman JE, Albrecht E, Jackson CM, Evans DM, Cadby G, Fornage M, Manichaikul A, Lopez LM, Johnson T, Aldrich MC, Aspelund T, Barroso I, Campbell H, Cassano PA, Couper DJ, Eiriksdottir G, Franceschini N, Garcia M, Gieger C, Gislason GK, Grkovic I, Hammond CJ, Hancock DB, Harris TB, Ramasamy A, Heckbert SR, Heliovaara M, Homuth G, Hysi PG, James AL, Jankovic S, Joubert BR, Karrasch S, Klopp N, Koch B, Kritchevsky SB, Launer LJ, Liu Y, Loehr LR, Lohman K, Loos RJ, Lumley T, Al Balushi KA, Ang WQ, Barr RG, Beilby J, Blakey JD, Boban M, Boraska V, Brisman J, Britton JR, Brusselle GG, Cooper C, Curjuric I, Dahgam S, Deary IJ, Ebrahim S, Eijgelsheim M, Francks C, Gaysina D, Granell R, Gu X, Hankinson JL, Hardy R, Harris SE, Henderson J, Henry A, Hingorani AD, Hofman A, Holt PG, Hui J, Hunter ML, Imboden M, Jameson KA, Kerr SM, Kolcic I, Kronenberg F, Liu JZ, Marchini J, McKeever T, Morris AD, Olin AC, Porteous DJ, Postma DS, Rich SS, Ring SM, Rivadeneira F, Rochat T, Sayer AA, Sayers I, Sly PD, Smith GD, Sood A, Starr JM, Uitterlinden AG, Vonk JM, Wannamethee SG, Whincup PH, Wijmenga C, Williams OD, Wong A, Mangino M, Marciante KD, McArdle WL, Meibohm B, Morrison AC, North KE, Omenaas E, Palmer LJ, Pietilainen KH, Pin I, Pola Sbreve Ek O, Pouta A, Psaty BM, Hartikainen AL, Rantanen T, Ripatti S, Rotter JI, Rudan I, Rudnicka AR, Schulz H, Shin SY, Spector TD, Surakka I, Vitart V, Volzke H, Wareham NJ, Warrington NM, Wichmann HE, Wild SH, Wilk JB, Wjst M, Wright AF, Zgaga L, Zemunik T, Pennell CE, Nyberg F, Kuh D, Holloway JW, Boezen HM, Lawlor DA, Morris RW, Probst-Hensch N, Kaprio J, Wilson JF, Hayward C, Kahonen M, Heinrich J, Musk AW, Jarvis DL, Glaser S, Jarvelin MR, Ch Stricker

BH, Elliott P, O'Connor GT, Strachan DP, London SJ, Hall IP, Gudnason V, Tobin MD. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* 2011.

28. Kohansal R, Martinez-Camblor P, Agusti A, Buist AS, Mannino DM, Soriano JB. The natural history of chronic airflow obstruction revisited: An analysis of the framingham offspring cohort. *Am J Respir Crit Care Med* 2009;180:3-10.

29. Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, Marciante KD, Franceschini N, van Durme YM, Chen TH, Barr RG, Schabath MB, Couper DJ, Brusselle GG, Psaty BM, van Duijn CM, Rotter JI, Uitterlinden AG, Hofman A, Punjabi NM, Rivadeneira F, Morrison AC, Enright PL, North KE, Heckbert SR, Lumley T, Stricker BH, O'Connor GT, London SJ. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet* 2010;42:45-52.

30. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, Crapo R, Enright P, van der Grinten CP, Gustafsson P, Jensen R, Johnson DC, MacIntyre N, McKay R, Navajas D, Pedersen OF, Pellegrino R, Viegi G, Wanger J. Standardisation of spirometry. *Eur Respir J* 2005;26:319-338.

31. 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* 2009;63:458-464.

32. Yetley EA, Pfeiffer CM, Schleicher RL, Phinney KW, Lacher DA, Christakos S, Eckfeldt JH, Fleet JC, Howard G, Hoofnagle AN, Hui SL, Lensmeyer GL, Massaro J, Peacock M, Rosner B, Wiebe D, Bailey RL, Coates PM, Looker AC, Sempos C, Johnson CL, Picciano MF. Nhanes monitoring of serum 25-hydroxyvitamin d: A roundtable summary. *J Nutr* 2010;140:2030S-2045S.

33. CDC. Revised analytical note for nhanes 2000-2006 and nhanes iii (1998-1994) 25-hydroxyvitamin d analysis. 2010.

34. Cheng JB, Levine MA, Bell NH, Mangelsdorf DJ, Russell DW. Genetic evidence that the human cyp2r1 enzyme is a key vitamin d 25-hydroxylase. *Proc Natl Acad Sci U S A* 2004;101:7711-7715.

35. Litonjua AA. Vitamin d and the lung : Mechanisms and disease associations. New York: Humana Press; 2012.

36. Abbas S, Linseisen J, Slanger T, Kropp S, Mutschelknauss EJ, Flesch-Janys D, Chang-Claude J. The gc2 allele of the vitamin d binding protein is associated with a decreased postmenopausal breast cancer risk, independent of the vitamin d status. *Cancer Epidemiol Biomarkers Prev* 2008;17:1339-1343.

37. Sinotte M, Diorio C, Berube S, Pollak M, Brisson J. Genetic polymorphisms of the vitamin d binding protein and plasma concentrations of 25-hydroxyvitamin d in premenopausal women. *Am J Clin Nutr* 2009;89:634-640.

38. Kamboh MI, Ferrell RE. Ethnic variation in vitamin d-binding protein (gc): A review of isoelectric focusing studies in human populations. *Hum Genet* 1986;72:281-293.

39. Speeckaert M, Huang G, Delanghe JR, Taes YE. Biological and clinical aspects of the vitamin d binding protein (gc-globulin) and its polymorphism. *Clin Chim Acta* 2006;372:33-42.

40. Jorde R, Sneve M, Hutchinson M, Emaus N, Figenschau Y, Grimnes G. Tracking of serum 25-hydroxyvitamin d levels during 14 years in a population-based study and during 12 months in an intervention study. *Am J Epidemiol* 2010;171:903-908.

ONLINE SUPPLEMENT

Supplemental Methods

Replication Cohorts

Three cohorts from the CHARGE (Cohorts for Heart and Aging Research in Genomic Epidemiology) consortium and one cohort from the SpiroMeta consortium were used for replication of the SNP—FEV_1 findings in the Framingham Heart Study. Cohorts included for the replication had ≥ 3 FEV_1 measurements per participant, namely the Busselton Health Study (BHS), the Coronary Artery Risk Development in Young Adults (CARDIA), the Cardiovascular Health Study (CHS), the Health, Aging, and Body Composition Study (HABC), and the Framingham Heart Study (FHS). Further details on each cohort provided elsewhere (manuscript in preparation).

Supplemental Results

Association of 25(OH)D with Rate of Change in FEV_1 in the Offspring Cohort

In longitudinal analyses of 25(OH)D with rate of change in FEV_1 , we observed a statistically significant interaction of cohort by time such that the Offspring cohort had a significantly attenuated rate of FEV_1 decline compared to the younger Third Generation cohort, as reported earlier. We further explored this finding in two sensitivity analyses. First, we excluded Offspring participants with COPD, defined as GOLD stages 1-4 (222 participants excluded), but the cohort x time interaction remained significant ($P=1.74\text{x}10^{-25}$). Second, we excluded all Offspring Exam 8 spirometry measurements given there was a change in the type of spirometer used at Exam 8 (186 participants excluded); however, the cohort x time interaction remained significant in this analysis as well ($P=2.32\text{x}10^{-30}$).

We hypothesize that the attenuated rate of decline in the Offspring compared to the Third Generation participants reflects a "healthy survivor" bias in the 1,435 Offspring cohort participants with serum 25(OH)D data, leading to systematic differences in rate of decline between cohorts. A previously published study examining FEV_1 change in the Framingham Offspring cohort from Exams 1-6 demonstrated that the oldest male participants had a slight increase in FEV_1, which the authors attributed either to healthy survivor bias or measurement variability (28).

Supplemental Table 4.6 Imputed SNPs in Vitamin D Metabolic Genes in FHS

Gene Symbol	Chr	Function	Imputed SNPs in FHS
CYP24A1	20	Degradation of $1,25(OH)_2D$	42
CYP27A1	2	Vitamin D 25-hydroxylase	23
CYP27B1	12	25(OH)D 1-α-hyroxylase	5
CYP2R1	11	Vitamin D 25-hydroxylase	15
DHCR7/NADSYN1	11	*DHCR7* converts vitamin D_3 substrate to cholesterol; *NADSYN1* is a flanking gene	105
GC	4	Vitamin D binding protein	51

Supplemental Table 4.7 SNPs in *CYP27B1, CYP2R1, DHCR7/NADSYN1,* and *GC* associated with the rate of change in FEV_1 in FHS at $P < 0.05$ (sorted by gene)

Gene	SNP	Chr	Position	Coded Allele	Freq	β	SE	*P*
CYP27B1	rs10877013	12	56451352	T	0.30	-1.3	0.5	0.0210
	rs703842	12	56449006	G	0.30	-1.2	0.5	0.0228
	rs1048691	12	56439215	T	0.21	1.4	0.6	0.0241
CYP2R1	rs11819875	11	14873873	G	0.18	-1.9	0.7	0.0043
	rs16930625	11	14874884	G	0.10	-2.1	0.8	0.0075
	rs16930609	11	14872484	C	0.10	-1.9	0.8	0.0170
DHCR7/ NADSYN1	rs1790349	11	70819998	C	0.15	-2.2	0.7	0.0015
	rs7120029	11	70876605	A	0.15	-2.1	0.7	0.0024
	rs3829251	11	70872207	A	0.15	-2.1	0.7	0.0026
	rs10898193	11	70874731	T	0.15	-2.1	0.7	0.0026
	rs10898203	11	70881084	T	0.14	-1.9	0.7	0.0051
	rs11233933	11	70887718	T	0.14	-1.9	0.7	0.0060
	rs10898211	11	70888325	C	0.14	-1.9	0.7	0.0060
	rs736894	11	70829906	T	0.20	-1.6	0.6	0.0103
	rs1630498	11	70828433	C	0.20	-1.5	0.6	0.0107
	rs1790324	11	70828168	G	0.20	-1.5	0.6	0.0115
	rs1790345	11	70823589	A	0.20	-1.5	0.6	0.0120
	rs1792229	11	70857043	G	0.19	-1.5	0.6	0.0166
	rs1792234	11	70859666	C	0.19	-1.5	0.6	0.0166
	rs1540127	11	70856686	A	0.19	-1.5	0.6	0.0168
	rs1790343	11	70854855	C	0.19	-1.5	0.6	0.0169
	rs1792226	11	70854222	T	0.19	-1.5	0.6	0.0171
	rs1629220	11	70852201	T	0.19	-1.5	0.6	0.0174
	rs2002064	11	70841068	G	0.19	-1.4	0.6	0.0194
GC	rs842999	4	72830554	G	0.44	1.1	0.5	0.0300
	rs705120	4	72833004	A	0.42	1.0	0.5	0.0498

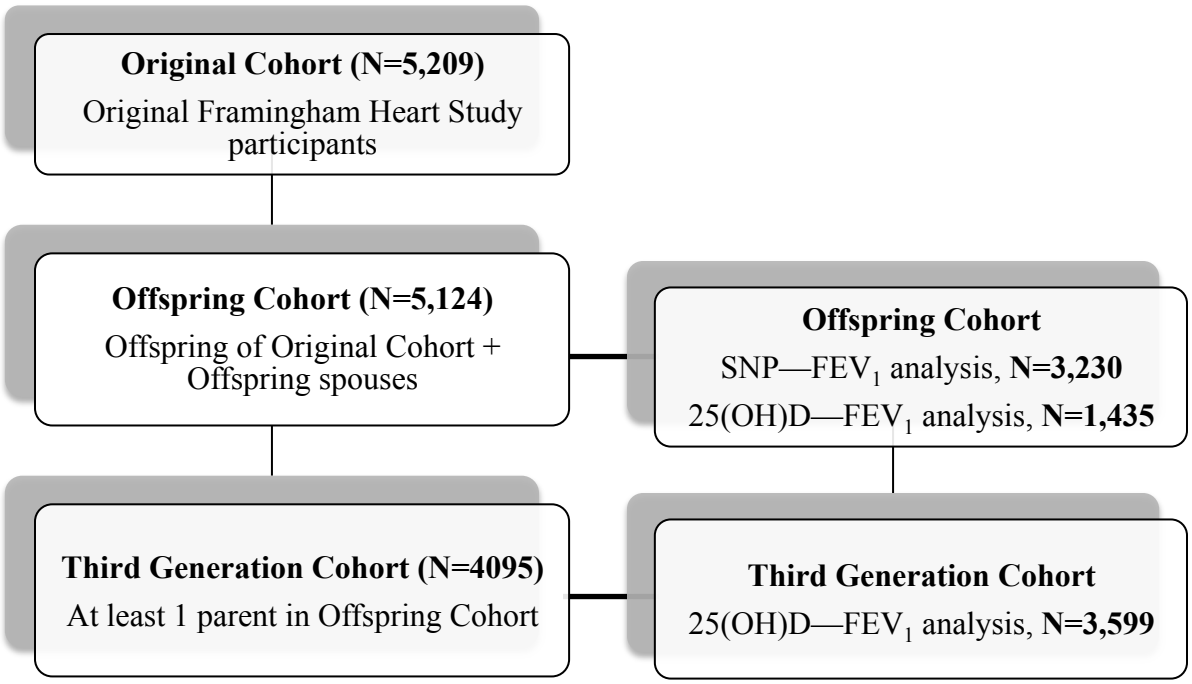
Supplemental Table 4.8 Replication cohort associations of the most significant SNPs per gene with the rate of change in FEV_1 (sorted by gene)

SNP	Gene	Coded Allele	BHS (N=3,230)			CARDIA			CHS			Health ABC		
			β	SE	*P*	β	SE	*P*	β	SE	*P*	β	SE	*P*
rs10877013	*CYP27B1*	T	-1.2	1.1	0.2574	-0.2	0.6	0.6941	+1.1	1.2	0.3751	-1.5	1.2	0.2117
rs11819875	*CYP2R1*	G	-1.1	1.4	0.4279	-0.2	0.7	0.7924	+0.2	1.5	0.9188	-5.0	1.5	0.0006
rs1790349	*DHCR7*	C	-1.8	1.5	0.2184	+0.5	0.7	0.4593	+0.1	1.7	0.9354	-1.3	1.6	0.4423
rs842999	*GC*	G	+2.4	1.0	0.0168	+0.1	0.6	0.8621	+2.2	1.1	0.0523	-0.3	1.2	0.8117

Abbreviations: BHS = Busselton Health Study; CARDIA = Coronary Artery Risk Development in Young Adults Study; CHS = Cardiovascular Health Study; Health ABC = Health, Aging, and Body Composition Study; SNP = single nucleotide polymorphism; β = beta coefficient for SNP x time effect; SE = standard error; *P* = P-value

Adjusted for: baseline age, gender, height, smoking pattern over follow-up and its interaction with time, baseline smoking pack-years, and the first two principal components for genetic ancestry

Supplemental Figure 4.3 Overview of Framingham Heart Study Cohorts



CHAPTER 5

CONCLUSION

Vitamin D has generated enormous interest following studies showing associations between vitamin D status and numerous health outcomes including cancer, cardiovascular disease, and diabetes. National surveys estimate that approximately 1/3 of the United States population is at risk for vitamin D insufficiency (defined as serum 25(OH)D < 20 ng/mL) (1), suggesting that many individuals could benefit from improved vitamin D status, although the precise meaning of deficiency in relation to health outcomes is yet to be fully characterized. Indeed, definitive causal associations with vitamin D have been established only for bone health outcomes, and the Institute of Medicine has identified further elucidation of vitamin D associations with non-skeletal outcomes as a research priority (2).

Understanding vitamin D—lung function associations is a growing research area following cross-sectional observational studies that demonstrated strong, positive associations between serum 25(OH)D and lung outcomes. Decline in lung function is the primary characteristic of chronic obstructive pulmonary disease (COPD), which is currently the 3rd leading cause of death in the U.S. and a significant burden on healthcare resources (3). However, representative, longitudinal, population-based studies of vitamin D and lung outcomes are lacking, and to date there are no published randomized controlled trials investigating vitamin D effects on lung function.

The three projects described in this dissertation were designed and conducted to assess the determinants of serum 25(OH)D status in a population at high risk of inadequacy, and to

investigate associations between serum 25(OH)D and pulmonary function outcomes with an exploration of mechanisms. The three completed studies are as follows: 1) genetic and environmental determinants of serum 25(OH)D were explored in elderly African American participants in the Health, Aging, and Body Composition (Health ABC) study to understand the relative contributions of modifiable and non-modifiable factors to vitamin D status; 2) genetic variants in vitamin D-responsive genes were evaluated in association with cross-sectional lung function in Health ABC to elucidate potential mechanisms for vitamin D—lung function effects; and, 3) the cross-sectional and longitudinal associations of vitamin D with lung function were explored in the Framingham Heart Study (FHS), considering both serum status and genetic variants in vitamin D metabolic pathway genes.

These distinct yet complementary projects investigate the role of vitamin D in lung, providing important information about a non-bone health outcome and addressing gaps in the published literature. A brief summary of each project is presented below.

Genetic and Environmental Predictors of Serum 25-Hydroxyvitamin D in African Americans in the Health, Aging, and Body Composition Study (Chapter 2)

No published studies to date explore both genetic and non-genetic determinants of serum 25(OH)D status in elderly African Americans, a group at high risk of vitamin D insufficiency due to the dual factors of advanced age and skin pigmentation. In studying about 1,000 African American Health ABC participants, we estimated that 25% of the variation in serum 25(OH)D was explained by non-genetic factors; the use of multivitamin supplements was the strongest predictor of status. Up to 23% of 25(OH)D variability was estimated to be attributable to additive genetic variation in Health ABC, and this finding was replicated in a separate cohort of African

Americans. However, we were unable to separate the effects of population ancestry from other genetic effects on 25(OH)D. We studied SNPs associated with 25(OH)D in GWAS of Caucasians, to assess whether these identified SNPs were associated with 25(OH)D in African Americans, and found that none were associated with 25(OH)D at $P<0.05$. However, rs7041, a non-synonymous SNP in the vitamin D binding protein (*GC*), had a borderline statistically significant association with 25(OH)D ($P=0.08$). Given the biological function of *GC*, we further explored SNP interactions with multivitamin supplement use. The rs7041 *TT* genotype, associated with lower 25(OH)D status in previous studies, modified the 25(OH)D response to multivitamin supplementation such that supplement users with the *TT* genotype had lower 25(OH)D and increased odds of vitamin D insufficiency compared to supplement users with the *GG/GT* genotype. There was little or no association of genotype with 25(OH)D in participants who did not use multivitamin supplements.

Overall, we identified an effect of genetic variation on 25(OH)D status in two cohorts of African Americans, consistent with a true effect, but further exploration of the genetic architecture of 25(OH)D in African Americans is needed. Modifiable predictors of 25(OH)D, including dietary patterns, supplement use, physical activity, and BMI were identified. Finally, we demonstrated that the rs7041 SNP modifies the effect of multivitamin supplement use on serum 25(OH)D, highlighting an important consideration for clinical trials of vitamin D supplementation.

Vitamin D-Responsive SGPP2 Variants Associated with Lung Cell Expression and Lung Function (Chapter 3)

The second study investigated the cross-sectional associations between genetic variants in vitamin D-responsive genes and lung function in Health ABC, with the goal of identifying lung tissue-specific mechanisms for observed 25(OH)D—lung function associations. Although vitamin D has a role in biological processes important for lung health, translational studies directly investigating vitamin D's effects *in vivo* are lacking. This study expanded on a previously completed gene expression study, which identified 13 genes differentially expressed by 25(OH)D in human lung epithelial cells. Using a cross-sectional design, we investigated genetic variation in these 13 genes in association with pulmonary function in Health ABC European Americans and African Americans. The strongest finding was for the *SGPP2* gene, a phosphatase that catalyzes the degradation of the signaling molecule sphingosine-1-phosphate (S1P). *SGPP2* has also been implicated in inflammatory signaling. SNPs in *SGPP2* were associated with forced expiratory volume in 1 second (FEV_1), a key parameter of lung function, in both Health ABC racial groups; also, a linked group of SNPs was associated with increased COPD risk in African Americans. Gene-level replication of *SGPP2* in association with cross-sectional lung function was observed in the Framingham Heart Study.

We hypothesized that these risk-associated SNPs would influence gene expression, and performed an expression quantitative trait loci (eQTL) analysis to evaluate the association between genetic variants and gene expression. We identified a highly significant eQTL association, although it was in a different region of *SGPP2* as the previous risk-associated variants. *SGPP2* SNPs associated with FEV_1 or expression may affect binding of $1,25(OH)_2D$/VDR or the co-regulatory complexes required for gene expression and translation, but we could not directly assess specific mechanisms. *SGPP2* remains a promising candidate gene for future study.

25-Hydroxyvitamin D Status and Genetic Variation in the Vitamin D Metabolic Pathway in Association with FEV_1 in the Framingham Heart Study (Chapter 4)

We investigated associations between 25(OH)D and rate of change in FEV_1 in the Framingham Heart Study, a representative population-based cohort of healthy adult participants. Both genetic variation in vitamin D metabolic genes, hypothesized to influence usual vitamin D status, and serum 25(OH)D were considered in association with rate of change in FEV_1. The cross-sectional association of serum 25(OH)D and FEV_1 was also evaluated. Given some variation in available data, these analyses considered data from the Offspring cohort participants (offspring of original study members) and the Third Generation cohort participants (children of offspring).

Genetic variants in four vitamin D metabolic genes were associated with rate of change in FEV_1, and showed an overall consistent direction of effect on rate of change in FEV_1 in 4 independent replication cohorts. We further investigated the association of these identified SNPs with serum 25(OH)D concentrations in the SUNLIGHT consortium. Rs11819875 (*CYP2R1*) demonstrated the strongest association with rate of change in FEV_1; the *G* allele of rs111819875 was associated with steeper FEV_1 decline and, in SUNLIGHT, with lower 25(OH)D. Rs842999 (*GC*) was strongly associated with attenuated FEV_1 decline, and, in SUNLIGHT, with lower 25(OH)D (contrary to expectation). Rs842999 is in strong linkage disequilibrium (R^2=0.9) with rs7041, and the *T* allele of rs7041 had the same direction of effect on FEV_1 and 25(OH)D as the *G* allele of rs842999. The rs7041 *T* allele is part of two common GC isoforms, GC1F and GC2; both are associated with lower 25(OH)D, but GC1F is associated with an increased risk of COPD

while GC2 has a protective effect. The reason for the opposite direction of association with the lung outcomes is unclear.

We confirmed the well-known cross-sectional association of serum 25(OH)D with FEV_1, and found that the association was steepest in the insufficient range of vitamin D [25(OH)D <20 ng/mL]. There was no association between 25(OH)D and rate of change in FEV_1 in the Third Generation cohort; however, the average 25(OH)D was 34.5 ng/mL, and only 8.6% of participants had insufficient 25(OH)D status. Due to evidence of selection bias, we were unable to evaluate associations between 25(OH)D and rate of change in FEV_1 in the Offspring cohort with 25(OH)D measurements (a limited subset of all Offspring cohort data).

Overall, SNP analyses support an effect of 25(OH)D on lung function change over time. Although no association was observed between 25(OH)D and rate of FEV_1 change in the Third Generation cohort, the average serum 25(OH)D in this cohort was in the sufficient range, limiting our ability to detect associations with low 25(OH)D.

Emerging Themes & Future Directions

Several important themes emerge in this work that are relevant for future vitamin D-related research. First, we show that rs7041, a genetic variant in the vitamin D binding protein (*GC*), affects serum 25(OH)D status in multivitamin supplement users. This is an important consideration for vitamin D supplementation trials, particularly as frequency of rs7041 varies across race/ethnicity. Further, we found that a SNP in strong linkage disequilibrium with rs7041 was associated with both lower 25(OH)D in SUNLIGHT and attenuated rate of lung function decline in FHS. This highlights that associations between 25(OH)D and health outcomes may depend on underlying genetic variation. Finally, vitamin D effects may be restricted to the

insufficient ranges of 25(OH)D, and baseline 25(OH)D status is an important consideration for future studies.

Further research is needed to characterize fully the association between serum and lung tissue 25(OH)D concentrations, including whether an increase in serum 25(OH)D similarly increases lung tissue 25(OH)D concentrations. Along the same vein, a better understanding of vitamin D uptake into the lung is needed. The major pathway for vitamin D uptake to the kidney is via receptor-mediated endocytosis, specifically through the action of the megalin and cubulin proteins (4, 5). Megalin and cubulin are expressed in lung alveolar cells (6), and this pathway of vitamin D uptake has been demonstrated in mammary cells (7). However, no studies to date have directly explored mechanisms of *in vivo* 25(OH)D uptake into lung tissue. Alternatively, free 25(OH)D (not bound to the vitamin D binding protein) may diffuse across the cell membrane, as described by the "free hormone hypothesis" (8). However, given that >99% of circulating 25(OH)D is bound to the vitamin D binding protein (4), it is unlikely that diffusion is the only and/or the primary mechanism for vitamin D uptake in non-renal tissues. As discussed in Chapter 4, there are three major isoforms of the vitamin D binding protein; these isoforms have different vitamin D binding affinities (9), which may also affect 25(OH)D tissue uptake.

Large-scale genomics studies are needed to further elucidate how VDR binds to the genome in lung tissue and interacts with co-regulatory complexes to influence gene expression and protein synthesis. ChIP-seq analyses of genome-wide VDR binding in human lymphoblastoid and osteoblast cell lines have revealed critical information about VDR gene regulation (10, 11); however, associations may be cell- and tissue-specific (12), and there are no published ChIP-seq studies of VDR binding in lung tissue.

Finally, further longitudinal studies examining the association between serum 25(OH)D and rate of change in lung function are needed, particularly studies with sufficient representation of healthy individuals with insufficient serum 25(OH)D status. An ongoing vitamin D supplementation trial, VITAL (VITamin D and OmegA-3 Trial), which studies cardiovascular and cancer endpoints, has an ancillary study, lungVITAL (Clinical Trials.gov Identifer: NCT01728571), which registers a subgroup to study the effect of supplementation on rate of change in lung function. While this trial is expected to contribute answers to some of the questions raised above, the exact contribution awaits a better understanding of the proportion of vitamin D insufficient participants registered in the trial. An important understanding from the work reported herein is that if such persons are under-sampled, the trial may be less likely to contribute new information to the ongoing debate on the role of vitamin D in non-bone health outcomes.

REFERENCES

1. Looker AC JC, Lacher DA, et al. Vitamin d status: United states, 2001–2006. NCHS data brief, no 5. Hyattsville, MD: National Center for Health Statistic; 2011.

2. Ross AC, Institute of Medicine (U. S.). Committee to Review Dietary Reference Intakes for Vitamin D and Calcium. Dietary reference intakes for calcium and vitamin d. Washington, DC: National Academies Press; 2011.

3. Minino AM, Murphy SL. Death in the united states, 2010. *NCHS Data Brief* 2012:1-8.

4. Nykjaer A, Dragun D, Walther D, Vorum H, Jacobsen C, Herz J, Melsen F, Christensen EI, Willnow TE. An endocytic pathway essential for renal uptake and activation of the steroid 25-(oh) vitamin d3. *Cell* 1999;96:507-515.

5. Nykjaer A, Fyfe JC, Kozyraki R, Leheste JR, Jacobsen C, Nielsen MS, Verroust PJ, Aminoff M, de la Chapelle A, Moestrup SK, Ray R, Gliemann J, Willnow TE, Christensen EI. Cubilin dysfunction causes abnormal metabolism of the steroid hormone 25(oh) vitamin d(3). *Proc Natl Acad Sci U S A* 2001;98:13895-13900.

6. Kolleck I, Wissel H, Guthmann F, Schlame M, Sinha P, Rustow B. Hdl-holoparticle uptake by alveolar type ii cells: Effect of vitamin e status. *Am J Respir Cell Mol Biol* 2002;27:57-63.

7. Rowling MJ, Kemmis CM, Taffany DA, Welsh J. Megalin-mediated endocytosis of vitamin d binding protein correlates with 25-hydroxycholecalciferol actions in human mammary cells. *J Nutr* 2006;136:2754-2759.

8. Mendel CM. The free hormone hypothesis: A physiologically based mathematical model. *Endocr Rev* 1989;10:232-274.

9. Arnaud J, Constans J. Affinity differences for vitamin d metabolites associated with the genetic isoforms of the human serum carrier protein (dbp). *Hum Genet* 1993;92:183-188.

10. Meyer MB, Goetsch PD, Pike JW. Genome-wide analysis of the vdr/rxr cistrome in osteoblast cells provides new mechanistic insight into the actions of the vitamin d hormone. *J Steroid Biochem Mol Biol* 2010;121:136-141.

11. Ramagopalan SV, Heger A, Berlanga AJ, Maugeri NJ, Lincoln MR, Burrell A, Handunnetthi L, Handel AE, Disanto G, Orton SM, Watson CT, Morahan JM, Giovannoni G, Ponting CP, Ebers GC, Knight JC. A chip-seq defined genome-wide map of vitamin d receptor binding: Associations with disease and evolution. *Genome Res* 2010;20:1352-1360.

12. Pike JW, Meyer MB, Bishop KA. Regulation of target gene expression by the vitamin d receptor - an update on mechanisms. *Reviews in endocrine & metabolic disorders* 2012;13:45-55.

APPENDIX

A. Exploratory Gene-Environment Interactions in the Health ABC Cohort

METHODS

Serum 25(OH)D measurements were completed on a majority of Health ABC participants, which allowed the consideration of gene by nutrient interaction. Serum 25(OH)D was measured in stored serum samples from the 12-month follow-up visit using a 2-step radioimmunoassay (25-Hydroxyvitamin D 125I RIA Kit, DiaSorin, Stillwater, Minn., USA); the interassay coefficient of variation was 6.8% for log transformed 25(OH)D values.[34] Although the serum 25(OH)D measurements are from the 12-month follow-up visit, a recent study reported a correlation of 0.8 between vitamin D measurements taken a year apart[35], supporting the assumption that measured 25(OH)D is an excellent representation of vitamin D serum status at study baseline.

Genotype by serum 25(OH)D interactions were assessed for the FEV_1 and FEV_1/FVC phenotypes in an additive model by including a product term between each SNP and serum 25(OH)D, adjusting for season of vitamin D measurement; a less stringent nominal P-value threshold ($P<0.05$) was used for interaction analyses because of lower power to detect effects. In gene-nutrient interaction analyses, participants with missing serum 25(OH)D data were excluded.

RESULTS

In European-Americans, the genotype—FEV_1 association was modified by serum 25(OH)D for 10 SNPs in 4 genes (*DAPK1, KAL1, SGPP2,* and *SLITRK6)* at nominal $P<5.0x10^{-02}$. In African-Americans, the genotype—FEV_1 association was modified by serum

25(OH)D for 43 SNPs in 9 genes (*DAPK1, DTX4, EMB, FSTL1, KAL1, KCNS3, PTGER2, SGPP2, SLITRK6)* at nominal $P<5.0x10^{-02}$ (**Appendix Table 1**).

In European-Americans, the genotype—FEV_1/FVC association was modified by serum 25(OH)D for 11 SNPs in 4 genes (*KAL1, PTGER2, SGPP2,* and *TMEM40*) at nominal $P<5.0x10^{-02}$. In African-Americans, the genotype—FEV_1/FVC association was modified by serum 25(OH)D for 26 SNPs in 8 genes (*DAPK1, DTX4, EMB, FSTL1, KAL1, KCNS3, PTGER,* and *SGPP2*) at nominal $P<5.0x10^{-02}$ (**Appendix Table 2**).

The *SGPP2*—lung function association was consistently modified by serum 25(OH)D. Thus, the SNP—phenotype association for SNPs in the *SGPP2* gene was modified by serum 25(OH)D in European- and African-Americans for both the FEV_1 and the FEV_1/FVC phenotypes, with consistent direction of effect for the interaction effect (higher serum vitamin D concentrations attenuated genotype—phenotype associations). Similarly, SNPs in *DAPK1, DTX4, EMB, FSTL1, KAL1,* and *PTGER2* showed consistent evidence of genotype—serum 25(OH)D interactions for both FEV_1 and FEV_1/FVC in African-Americans.

For statistically significant genotype x serum 25(OH)D interactions (nominal $P<5.0x10^{-02}$) the mean FEV_1 was estimated (from model coefficients) for serum 25(OH)D concentrations of 12 ng/mL, 20 ng/mL, and 30 ng/mL, corresponding to typical definitions of deficient, sufficient, and optimal levels of vitamin D nutriture, respectively. At each level of serum 25(OH)D, individuals with 0 copies of the variant allele (wild-type homozygotes) were compared to individuals with 1 or 2 copies of the variant allele. In European-American participants, gene x nutrient interactions were consistent such that the allele—FEV_1 association was stronger at the "deficient" level of serum 25(OH)D (12 ng/ml). In African-

Americans, findings were mixed; interaction results went in both directions (**Appendix Table 1**). The pattern of findings for the ratio phenotype is similar (**Appendix Table 2**).

Appendix Table 1 SNP by 25(OH)D interactions associated with the FEV_1 phenotype in a) European-Americans, and b) African-Americans.

a) European-Americans

				Predicted FEV_1 difference (mL) by serum 25(OH)D		
Gene	**SNP**	**Interaction Coefficient $\beta_{Interaction}$****	**Nominal P-value***	**12 ng/ml**	**20 ng/ml**	**30 ng/ml**
DAPK1	rs2378753	-6.98	$4.00x10^{-03}$	108.6***	52.8	-17.0
	rs3095747	-7.68	$5.17x10^{-03}$	111.2	49.8	-26.9
KAL1	rs5933673	5.77	$9.70x10^{-03}$	-121.7	-75.5	-17.8
SGPP2	rs13021671†	-4.97	$1.90x10^{-02}$	52.5	12.8	-36.9
	rs2009150	5.91	$2.42x10^{-02}$	-67.1	-19.9	39.2
	rs6714352†	6.05	$3.15x10^{-02}$	-86.2	-37.7	22.7
	rs735678†	-8.31	$4.45x10^{-02}$	108.0	42.0	-41.0
SLITRK6	rs1337267	-3.05	$4.54x10^{-02}$	58.3	33.9	3.3
	rs356279	-3.25	$4.82x10^{-02}$	53.6	27.6	-5.0
	rs631906	-3.05	$4.82x10^{-02}$	58.3	33.9	3.3

b) African-Americans

				Predicted FEV_1 difference (mL) by serum 25(OH)D		
Gene	**SNP**	**$\beta_{Interaction}$****	**Nominal P-value***	**12 ng/ml**	**20 ng/ml**	**30 ng/ml**
DAPK1	rs1056719	5.76	$2.87x10^{-02}$	-67.96***	-21.89	35.70
	rs11141934	14.62	$2.04x10^{-02}$	-91.70	25.23	171.40
	rs3128519	5.97	$2.07x10^{-02}$	-61.26	-13.49	46.24
	rs10868609	-8.43	$3.47x10^{-02}$	95.62	28.21	-56.04
	rs3128477	-6.11	$4.04x10^{-02}$	49.62	0.70	-60.44
	rs10512187	5.57	$4.36x10^{-02}$	-79.76	-35.24	20.42
DTX4	rs12284698†	9.52	$2.70x10^{-02}$	-54.63	21.50	116.65
	rs1048444	-5.27	$5.43x10^{-03}$	44.21	2.04	-50.67
	rs656163	-5.67	$8.44x10^{-03}$	66.63	21.30	-35.37
EMB	rs13159894	-10.10	$2.77x10^{-02}$	97.08	16.25	-84.79
	rs16879113	-10.10	$2.77x10^{-02}$	97.08	16.25	-84.79
	rs7729211†	11.74	$2.82x10^{-03}$	-84.71	9.24	126.69
FSTL1	rs2673704	10.71	$3.75x10^{-02}$	-21.76	63.90	170.97
KAL1	rs7051071†	-8.50	$3.17x10^{-02}$	105.89	37.88	-47.14
	rs10127300†	-5.04	$3.24x10^{-02}$	43.06	2.75	-47.63
	rs5978935†	-5.03	$3.30x10^{-02}$	41.42	1.16	-49.16
	rs5933677	7.53	$6.61x10^{-03}$	-56.64	3.61	78.92

				Predicted FEV_1 difference (mL) by serum 25(OH)D		
Gene	**SNP**	$\beta_{Interaction}$**	**Nominal P-value***	**12 ng/ml**	**20 ng/ml**	**30 ng/ml**
	rs5978934	5.88	$7.81x10^{-03}$	-61.70	-14.65	44.16
	rs5933668	8.37	$3.06x10^{-03}$	-122.50	-55.57	28.08
	rs6530187†	6.79	$7.15x10^{-03}$	-66.38	-12.02	55.93
	rs5978943	-8.07	$8.10x10^{-03}$	83.65	19.07	-61.66
KCNS3	rs1461949	-7.77	$4.91x10^{-02}$	69.93	7.79	-69.89
	rs1870822	-8.17	$2.16x10^{-02}$	35.66	-29.69	-111.37
	rs4832524	-8.10	$4.11x10^{-02}$	75.68	10.91	-70.04
	rs7583266	-8.53	$5.18x10^{-03}$	51.12	-17.12	-102.41
PTGER2	rs10136396	-8.87	$2.89x10^{-02}$	91.80	20.88	-67.78
	rs10136414	-8.94	$2.82x10^{-02}$	91.94	20.41	-69.01
	rs10151916	-8.91	$2.83x10^{-02}$	89.87	18.61	-70.46
	rs11851457	-8.42	$3.66x10^{-02}$	91.94	24.58	-59.63
	rs12587363	-8.68	$4.88x10^{-02}$	86.47	17.06	-69.71
	rs12590616	-8.18	$2.84x10^{-02}$	79.80	14.32	-67.53
	rs1254598†	8.11	$1.30x10^{-02}$	-78.28	-13.41	67.68
	rs708499	-10.02	$1.53x10^{-02}$	106.51	26.33	-73.89
	rs708498	-9.30	$2.19x10^{-02}$	109.09	34.66	-58.37
	rs10142849	-9.11	$2.51x10^{-02}$	96.62	23.75	-67.33
	rs28613641	-8.26	$4.14x10^{-02}$	84.68	18.61	-63.98
	rs12587410	-8.87	$4.25x10^{-02}$	91.80	20.88	-67.78
SGPP2	rs7559017	6.40	$4.13x10^{-02}$	-29.40	21.82	85.85
	rs10176933	-8.41	$4.52x10^{-02}$	77.59	10.35	-73.70
	rs4674662	6.03	$5.00x10^{-02}$	-23.96	24.29	84.60
	rs4673024†	9.02	$1.40x10^{-03}$	-45.95	26.22	116.42
	rs1436786	7.84	$2.40x10^{-04}$	-28.37	34.34	112.74
SLITRK6	rs431057	10.52	$3.28x10^{-02}$	-76.38	7.77	112.95

* Nominal p-values are from additive models, adjusted for age, height, smoking, gender, study site, ancestry principal components, season of vitamin D measurement, and serum 25(OH)D.

**Interaction regression coefficient compares individuals heterozygous or homozygous for the minor allele (≥1 copy of the minor allele) to individuals with the homozygous wild-type genotype (i.e., no copies of the minor allele)

***Illustrative interpretation: In participants with serum 25(OH)D of 12 ng/mL, participants ≥1 copy of the minor allele had an estimated mean FEV_1 **68 mL lower** than homozygous wild-type individuals

† This SNP has a significant SNP by serum 25(OH)D interaction for both the FEV_1 and FEV_1/FVC phenotypes (within racial group)

Appendix Table 2 SNP by serum 25(OH)D interactions in association with the FEV_1/FVC phenotype in a) European-Americans, and b) African-Americans.

a) European-Americans

				Predicted FEV_1/FVC difference by serum 25(OH)D		
Gene	**SNP**	**$\beta_{Interaction}$****	**Nominal P-value***	**12 ng/mL**	**20 ng/mL**	**30 ng/mL**
KAL1	rs1079854	0.074	$4.59X10^{-02}$	-1.86^{***}	-1.27	-0.53
	rs11095490	0.078	$3.37X10^{-02}$	-1.88	-1.26	-0.48
	rs12840575	0.074	$4.59X10^{-02}$	-1.86	-1.27	-0.53
	rs1859867	0.068	$3.84X10^{-02}$	1.45	0.84	2.72
PTGER2	rs2229187	0.164	$3.87X10^{-02}$	-2.62	-1.31	0.33
SGPP2	rs13021671	-0.090	$1.62X10^{-02}$	1.29	0.57	-0.33
	rs4416206	0.088	$3.04X10^{-02}$	-1.48	-0.78	0.10
	rs6714352	0.105	$2.12X10^{-02}$	-1.19	-0.35	0.70
	rs6758392	0.085	$2.98X10^{-02}$	-1.31	-0.63	0.23
	rs735678	-0.118	$5.10X10^{-03}$	1.94	0.99	-0.19
TMEM40	rs9876483	-0.122	$1.58X10^{-02}$	2.21	1.24	0.02

b) African-Americans

				Predicted FEV_1/FVC difference by serum 25(OH)D		
Gene	**SNP**	**$\beta_{Interaction}$****	**Nominal P-value***	**12 ng/mL**	**20 ng/ml**	**30 ng/mL**
DAPK1	rs3118867	-0.088	$2.68x10^{-02}$	0.22^{***}	-0.48	-1.36
	rs3818584	-0.135	$3.79x10^{-02}$	0.55	-0.52	-1.87
	rs4878115	-0.133	$2.03x10^{-02}$	0.58	-0.48	-1.82
	rs1927975	-0.128	$3.17x10^{-02}$	0.25	-0.78	-2.06
	rs2274605	-0.138	$3.48x10^{-02}$	0.62	-0.48	-1.86
	rs943855	-0.131	$4.10x10^{-02}$	0.50	-0.55	-1.85
DTX4	rs12284698	0.159	$3.77X10^{-02}$	-0.04	1.23	2.82
EMB	rs7729211	0.173	$1.33x10^{-02}$	-1.91	-0.52	1.21
FSTL1	rs1105220	-0.148	$2.68x10^{-02}$	1.78	0.60	-0.88
	rs1624195	-0.117	$3.45x10^{-02}$	1.24	0.31	-0.86
	rs4533682	-0.143	$3.57x10^{-02}$	1.53	0.39	-1.04
KAL1	rs6530187	0.155	$3.68x10^{-02}$	-1.07	0.17	1.72
	rs5978934	-0.144	$2.22x10^{-02}$	0.97	-0.18	-1.62
	rs6640194	-0.104	$2.52x10^{-02}$	-0.01	-0.84	-1.88
	rs5978943	0.143	$3.11x10^{-02}$	-0.85	0.29	1.73

				Predicted FEV_1/FVC difference by serum 25(OH)D		
	rs10127300	-0.078	$4.04x10^{-02}$	0.96	0.34	-0.43
	rs5978935	-0.077	$4.11x10^{-02}$	0.93	0.31	-0.46
	rs7887099	-0.094	$4.18x10^{-02}$	0.08	-0.67	-1.61
	rs7051071	-0.184	$4.42x10^{-02}$	0.77	-0.70	-2.54
KCNS3	rs3747516	0.145	$2.02x10^{-02}$	-1.48	-0.32	1.13
PTGER2	rs1254581	-0.150	$2.90x10^{-02}$	0.31	-0.89	-2.39
	rs1495785	-0.098	$4.99x10^{-02}$	0.08	-0.70	-1.68
	rs1254598	0.121	$2.42x10^{-02}$	-1.12	-0.15	1.06
SGPP2	rs17562982	-0.118	$3.68x10^{-02}$	0.41	-0.54	-1.72
	rs2009150	0.164	$1.18x10^{-02}$	-1.95	-0.65	0.99
	rs4673024	0.181	$1.41x10^{-03}$	-1.20	0.25	2.05

* Nominal p-values are from additive models, adjusted for age, height, smoking, gender, study site, ancestry principal components, season of vitamin D measurement, serum 25(OH)D.

**Interaction regression coefficient compares individuals heterozygous or homozygous for the minor allele (≥1 copy of the minor allele) to individuals with the homozygous wild-type genotype (i.e., no copies of the minor allele)

***Illustrative interpretation: In participants with serum 25(OH)D of 12 ng/mL, participants ≥1 copy of the minor allele had an estimated mean FEV_1 **0.22 higher** than homozygous wild-type individuals