**Title**
Identification of genetic and epigenetic factors in autoimmune disease etiology and treatment response



**Author**
Adams, Cameron

**Publication Date**
2022

Peer reviewed|Thesis/dissertation

# Identification of Genetic and Epigenetic Factors in Autoimmune Disease Etiology and Treatment Response

By

Cameron Jeffrey Adams

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Epidemiology

and the Designated Emphasis

in

Computational and Genomic Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Lisa F. Barcellos, Chair
Professor Alan Hubbard
Professor Priya Moorjani

Summer 2022

## Abstract

Identification of genetic and epigenetic factors in autoimmune disease etiology and treatment response

by

Cameron Jeffrey Adams

Doctor of Philosophy in Epidemiology

Designated Emphasis in Computational and Genomic Biology

University of California, Berkeley

Professor Lisa F. Barcellos, Chair

Multiple sclerosis and rheumatoid arthritis are complex autoimmune disease of unknown etiology. Each are result of exposure to a combination of genetic and environmental risk factors. Effective treatments that reduce the rate that normal tissues are attacked by the immune system have been developed; however, the mechanisms underlying disease pathogenesis and treatment response are not completely understood. In this dissertation, I used a combination of epidemiologic, bioinformatic, and computational methods to study the role of gene and environmental interactions, the vitamin D pathway, and DNA methylation in multiple sclerosis and rheumatoid arthritis. Chapter one introduces the genetic, environmental, and epigenetic factors in multiple sclerosis and rheumatoid arthritis. Chapter two describes results from an investigation into gene and environment interaction between genetic risk factors and pregnancy for multiple sclerosis susceptibility finding no evidence for effect modification between genetic susceptibility and pregnancy. Chapter three uses mendelian randomization methods to identify evidence that variation in vitamin D receptor binding is associated with MS susceptibility. Chapter four shows that changes cell-specific DNA methylation are associated with response to treatment with methotrexate among treatment naïve rheumatoid arthritis patients.

For Sarah and Harry, my Mom and Dad, and Becca.

**Acknowledgements**

First, thank you to my excellent advisor Lisa Barcellos. Her guidance and mentorship have been instrumental in my development as a scientist. Meeting her changed the trajectory of my studies and I will be forever grateful that she invited me join her lab and gave me the opportunity to do research in genetics, genomics, and health. Thank you to my dissertation committee members Alan Hubbard and Priya Moorjani for being available and supportive. Thank you to the many collaborators that made this research possible: Lindsey Criswell at UCSF, Catherine Schaefer and Annette Langer-Gould at Kaiser Permanente, Steffan D. Bos and Marte Wendel-Haga at the University of Oslo in Norway, and Lars Alfredsson, Ingrid Kockum, and Ali Manouchehrinia at the Karolinska Institue in Stockholm Sweden. Thank you to Kate Chase, Xuan Quach, Lauren Krupa, and Janene Martinez for helping me navigating completing requirements for Epidemiology PhD, Biostatistics MA, and computational biology designated emphasis. Hong and Diana Quach for managing the lab and providing me with excellent data, and Indro Fedrigo for his continuing efforts to improve our computing cluster.

Thank you to my past and current lab mates: Brooke Rhead, Xiaorong Shao, Olivia Solomon, Calvin Chi, Milena Gianfrancesco, Giovanna Cruz, Amanda Mok, and Mary Horton for being brilliant, fun, and patient. I learned so much from all of you and you made conferences a lot of fun. Thank you to Chris Rowe for joining me in completing the Biostats MA concurrently with our PhD, I would not have survived it without you. Thank you to Sean Wu for letting me ask you dumb questions about statistics and being a friend. Thank you to my 2017 epi cohort: Eli Michaels, Lauren Hunter, Ruvani Jayaweera, Wendy Qi, Isaac Ergas, Moon Choi, and Sabrina Boyce and thank you to past and present computational biology students. You all made the PhD more enriching and fun. Thank you to Katie Marker, Nick Ramirez, Jordan Keen, and Rosey Stone for your thoughtful work with me on your master's and undergraduate projects.

Thank you to Dr. Josh Bamberger for starting me on my public health and epidemiology journey. Thank you to my partner Sarah Dobbins, I truly could not have done this without your love, support, and guidance and thank you to Harrison Adams for making every day great. Finally, thank you to the National Institutes of Health for funding me.

**Table of Contents**

# Chapter 1 - Introduction

Multiple sclerosis (MS) and rheumatoid arthritis (RA) are autoimmune diseases with adult onset that primarily affect women.[1,2] MS is a demyelinating disease of the central nervous system (CNS) with a worldwide prevalence of approximately 0.1% and is a leading cause of neurological disability in young to middle-aged adults.[2,3] RA is the most common systemic autoimmune disease, with a worldwide prevalence approaching 1%.[1] It exhibits substantial clinical heterogeneity with the potential to cause substantial disability, primarily as a result of the erosive and deforming process in joints.[4,5]

Both MS and RA susceptibility have a substantial genetic component that includes genes within the major histocompatibility complex (MHC) region.[6,7] The contribution of genetics to MS and RA is estimated to comprise approximately 48% and 50%-60% of disease risk, respectively.[6,7] HLA-DRB1*15:01, located in MHC class II region, is the largest individual genetic risk factor for MS and 200 non-MHC risk variants have also been identified.[6] In RA, shared epitope (SE) alleles are the strongest genetic risk factors for developing RA and more than 60 non-MHC risk variants have been identified in GWAS.[7,8]

In addition to the substantial genetic contribution to disease susceptibility, several environmental risk factors have been identified for both MS and RA. Common risk factors include tobacco smoke, obesity, and low-vitamin D.[9–13] There is evidence of interaction between environmental risk factors and HLA gene alleles in MS susceptibility including tobacco smoking, Epstein-Barr Virus infection, and adolescent obesity.[14] The effect of pregnancy on MS susceptibility is unknown with studies yielding mixed results,[15] but among women with MS, pregnancy often results in temporary remission or improvement of symptoms but is followed by relapse or worsening of symptoms 3-6 months after childbirth.[16] Vitamin D insufficiency has been found to be implicated in risk of several autoimmune diseases, including MS.[17]

Epigenetic modifications are regulatory mechanisms that influence gene expression that are not the result of changes in the DNA sequence.[18] These modifications have been found to be of particular importance in the pathogenesis of RA.[19] Differences in DNA methylation (DNAm) at the global and CpG associated with RA cases status and disease severity has been identified in T lymphocytes, B cells, and peripheral blood.[20–25] Emerging evidence has found a role for DNAm as a mediator in response to methotrexate (MTX), the recommended first line of treatment in RA.[26–29] However, recent studies investigating DNAm in blood and response to MTX have yielded no genome-wide significant findings.[30,31] Peripheral blood contains several different cell types, each of which have different methylation profiles. DNAm measurements from blood are a combination of DNAm in the constituent cell-types and adjustment for global cell-proportions is critical when performing epigenome wide association studies.[32,33] Cell-specific differential DNAm may be obscured if, for example, the differential DNAm at a CpG for a given phenotype is present in only one cell-type or if the direction of differential DNAm is in opposing directions between certain cell-types. Recently, methods have been developed to estimate cell-specific differential DNAm from blood at the CpG level.[34]

In the second chapter, we investigated interaction between pregnancy before disease onset and MS genetic risk factors for MS. Previous studies had found mixed results for evidence of an association between pregnancy and MS.[15,16] In MS, pregnancy appears to have short term

beneficial effects on existing MS, but there is no agreement about the effect of pregnancy on MS risk.[35–37] We used a case-only $G \times E$ study design to estimate interaction between pregnancy and MS genetic risk factors including a polygenic risk score comprised of non-MHC risk variants, and the two largest individual risk factors (presence of HLA-DRB1*15:01 alleles and absence of HLA-A*02:01 alleles). Data on female MS cases with complete genetic and reproductive history from four MS studies was used for analyses: Kaiser Permanente Northern California MS Study, Kaiser Permanente Southern California MS Sunshine Study, two Swedish population-based case-control studies, and the Oslo MS Registry.

Previous research has found evidence that high levels of serum 25-hydroxyvitamin D (25(OH)D)[38], sunlight exposure[39,40], vitamin D supplements and diets rich in vitamin D[17] are associated with a decreased risk of MS susceptibility. Studies using mendelian randomization (MR) methods have identified a causal link between lower serum 25(OH)D and MS suscepbility.[13,41] It is well established that 25(OH)D signals through the nuclear vitamin D receptor (VDR), a ligand-regulated transcription factor which mediates all genomic actions of 25(OH)D. The contribution of genetic instruments suggests that more of the human genome is involved in explaining variation in serum 25(OH)D levels, and that different aspects of the vitamin D pathway, specifically transcription and expression mediated by VDR DNA binding are involved in the relationship between vitamin D and MS. In the third chapter, we investigated effect of vitamin D receptor binding at a locus on MS susceptibility. Previous research identified single nucleotide polymorphisms (SNPs) associated with variation in vitamin D receptor (VDR) DNA binding (VDR-BV).[42] We used these VDR-BVs as genetic instruments to conduct a MR study to estimate the association between VDR binding at a locus.

In the final chapter, we used DNAm measurements from before and after treatment with methotrexate (MTX) to identify signatures of DNAm associated with response to MTX in RA patients. We also used new computational methods to estimate cell-specific signatures of DNAm associated with MTX response at the CpG level.[34] MTX is a disease-modifying antirheumatic drug that is the recommended first treatment for RA.[43] Response to the first treatment regime is an important indicator of long-term prognosis[44–47], however, only 30-40% of patients continue with MTX treatment after two years.[27,48] Previous research suggests that medications, including MTX, alter patterns DNAm.[49–52] This is of interest for several reasons, including the fact that treatment-associated changes in the epigenome may explain, at least in part, the mechanism by which MTX exert its effects. Further, DNAm patterns prior to treatment and changes in DNAm associated with treatment may serve as predictors of treatment response.

## References


1. Almutairi KB, Nossent JC, Preen DB, Keen HI, Inderjeeth CA. The Prevalence of Rheumatoid Arthritis: A Systematic Review of Population-based Studies. The Journal of Rheumatology. 2021 May 1;48(5):669–76.

2. Walton C, King R, Rechtman L, Kaye W, Leray E, Marrie RA, et al. Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition. Mult Scler. 2020 Dec 1;26(14):1816–21.

3. Ascherio A, Munger K. Epidemiology of multiple sclerosis: from risk factors to prevention. Semin Neurol. 2008 Feb;28(1):17–28.

4. Suarez-Almazor ME, Soskolne CL, Saunders LD, Russell AS. Outcome in rheumatoid arthritis. A 1985 inception cohort study. The Journal of rheumatology. 1994 Aug;21(8):1438–46.

5. Barton JL, Trupin L, Schillinger D. Racial and ethnic disparities in disease activity and function among persons with rheumatoid arthritis from university-affiliated clinics.

6. International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. Science. 2019 Sep;365(6460):eaav7188–eaav7188.

7. Kim K, Bang SY, Lee HS, Cho SK, Choi CB, Sung YK, et al. High-density genotyping of immune loci in Koreans and Europeans identifies eight new rheumatoid arthritis risk loci. Ann Rheum Dis. 2015 Mar;74(3):e13.

8. van Drongelen V, Holoshitz J. Human Leukocyte Antigen-Disease Associations in Rheumatoid Arthritis. Rheumatic diseases clinics of North America. 2017 Aug;43(3):363–76.

9. Prevention of autoimmune rheumatic disease: state of the art and future perspectives | Annals of the Rheumatic Diseases [Internet]. [cited 2022 Jun 1]. Available from: https://ard.bmj.com/content/69/12/2062.short?casa_token=laKcbeUWXIYAAAAA:YsMuPv8R7lbQYVNQg01neeaHEbbWT-03D7-mdTWipDfSwrm8ygNcR6O4A6kH5ggqQMq8db6gmVo

10. Briggs FBS, Acuna B, Shen L, Ramsay P, Quach H, Bernstein A, et al. Smoking and Risk of Multiple Sclerosis. Epidemiology. 2014 Jul;25(4):605–14.

11. Hedström AK, Bomfim IL, Barcellos LF, Briggs F, Schaefer C, Kockum I, et al. Interaction between passive smoking and two HLA genes with regard to multiple sclerosis risk. International Journal of Epidemiology. 2014 Dec;43(6):1791–8.

12. Gianfrancesco MA, Acuna B, Shen L, Briggs FBS, Quach H, Bellesis KH, et al. Obesity during childhood and adolescence increases susceptibility to multiple sclerosis after

accounting for established genetic and environmental risk factors. Obesity Research & Clinical Practice. 2014 Sep;8(5):e435–47.

13. Mokry LE, Ross S, Ahmad OS, Forgetta V, Smith GD, Leong A, et al. Vitamin D and Risk of Multiple Sclerosis: A Mendelian Randomization Study. Muraro PA, editor. PLOS Medicine. 2015 Aug;12(8):e1001866–e1001866.

14. Olsson T, Barcellos LF, Alfredsson L. Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. Nat Rev Neurol. 2017 Jan;13(1):25–36.

15. Hedström AK, Hillert J, Olsson T, Alfredsson L. Reverse causality behind the association between reproductive history and MS. 2014 Apr;20(4). Available from: http://dx.doi.org/10.1177/1352458513498126

16. Adams Waldorf KM, Nelson JL. Autoimmune disease during pregnancy and the microchimerism legacy of pregnancy. Immunol Invest. 2008;37(5):631–44.

17. Ascherio A, Munger KL, Simon KC. Vitamin D and multiple sclerosis. The Lancet Neurology. 2010 Jun;9(6):599–612.

18. Klein K, Gay S. Epigenetics in rheumatoid arthritis. Current Opinion in Rheumatology. 2015 Jan;27(1):76–82.

19. Nemtsova MV, Zaletaev DV, Bure IV, Mikhaylenko DS, Kuznetsova EB, Alekseeva EA, et al. Epigenetic Changes in the Pathogenesis of Rheumatoid Arthritis. Frontiers in Genetics [Internet]. 2019 [cited 2022 Jun 1];10. Available from: https://www.frontiersin.org/article/10.3389/fgene.2019.00570

20. Glossop JR, Emes RD, Nixon NB, Packham JC, Fryer AA, Mattey DL, et al. Genome-wide profiling in treatment-naive early rheumatoid arthritis reveals DNA methylome changes in T and B lymphocytes. Epigenomics. 2016 Feb;8(2):209–24.

21. de Andres MC, Perez-Pampin E, Calaza M, Santaclara FJ, Ortea I, Gomez-Reino JJ, et al. Assessment of global DNA methylation in peripheral blood cell subpopulations of early rheumatoid arthritis before and after methotrexate. Arthritis Res Ther. 2015 Aug 29;17(1):233.

22. Cribbs AP, Kennedy A, Penn H, Read JE, Amjadi P, Green P, et al. Treg Cell Function in Rheumatoid Arthritis Is Compromised by CTLA-4 Promoter Methylation Resulting in a Failure to Activate the Indoleamine 2,3-Dioxygenase Pathway. Arthritis & Rheumatology. 2014;66(9):2344–54.

23. Liu F. A Revisit to Sample Size and Power Calculations for Testing Odds Ratio in Two Independent Binomials. Biometrics. 2013 Jun;69(2):530–6.

24. Mok A, Rhead B, Holingue C, Shao X, Quach HL, Quach D, et al. Hypomethylation of CYP2E1 and DUSP22 Promoters Associated With Disease Activity and Erosive Disease Among Rheumatoid Arthritis Patients. Arthritis & Rheumatology. 2018;70(4):528–36.

25. Rhead B, Holingue C, Cole M. DNA methylation profiles that distinguish rheumatoid arthritis (RA) from osteoarthritis in fibroblast-like synoviocytes can be detected in immune cells from peripheral blood (abstract 460T. In: 64th Annual Meeting of the American Society of Human Genetics. in San Diego, CA; 2014.

26. Cribbs AP, Kennedy A, Penn H, Amjadi P, Green P, Read JE, et al. Methotrexate Restores Regulatory T Cell Function Through Demethylation of the FoxP3 Upstream Enhancer in Patients With Rheumatoid Arthritis. Arthritis & Rheumatology. 2015;67(5):1182–92.

27. Verstappen SM, Owen SA, Hyrich KL. Prediction of response and adverse events to methotrexate treatment in patients with rheumatoid arthritis. International Journal of Clinical Rheumatology. 2012;7(5):559.

28. Glossop JR, Nixon NB, Emes RD, Sim J, Packham JC, Mattey DL, et al. DNA methylation at diagnosis is associated with response to disease-modifying drugs in early rheumatoid arthritis. Epigenomics. 2017 Apr;9(4):419–28.

29. Nair N, Wilson AG, Barton A. DNA methylation as a marker of response in rheumatoid arthritis. Pharmacogenomics. 2017 Sep;18(14):1323–32.

30. Gosselt HR, Vallerga CL, Mandaviya PR, Lubberts E, Hazes JMW, de Jonge R, et al. Epigenome wide association study of response to methotrexate in early rheumatoid arthritis patients. PLoS One. 2021 Mar 10;16(3):e0247709.

31. Nair N, Plant D, Plant D, Verstappen SM, Verstappen SM, Isaacs JD, et al. Differential DNA methylation correlates with response to methotrexate in rheumatoid arthritis. Rheumatology (United Kingdom). 2020 Jun;59(6):1364–71.

32. Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlén SE, Greco D, et al. Differential DNA Methylation in Purified Human Blood Cells: Implications for Cell Lineage and Studies on Disease Susceptibility. PLoS One. 2012 Jul 25;7(7):e41361.

33. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012 May 8;13:86.

34. Rahmani E, Schweiger R, Rhead B, Criswell LA, Barcellos LF, Eskin E, et al. Cell-type-specific resolution epigenetics without the need for cell sorting or single-cell biology. Nature Communications. 2019 Dec;10(1).

35. Waubant E, Lucas R, Mowry E, Graves J, Olsson T, Alfredsson L, et al. Environmental and genetic risk factors for MS: an integrated review. Annals of Clinical and Translational Neurology. 2019 Sep;6(9):1905–22.

36. McCombe PA. The Short and Long-Term Effects of Pregnancy on Multiple Sclerosis and Experimental Autoimmune Encephalomyelitis. Journal of clinical medicine [Internet]. 2018 Nov;7(12). Available from: http://www.ncbi.nlm.nih.gov/pubmed/30486504

37. Houtchens MK, Edwards NC, Schneider G, Stern K, Phillips AL. Pregnancy rates and outcomes in women with and without MS in the United States. Neurology. 2018;91(17):E1559–69.

38. Munger KL, Levin LI, Hollis BW, Howard NS, Ascherio A. Serum 25-Hydroxyvitamin D Levels and Risk of Multiple Sclerosis. JAMA. 2006 Dec;296(23):2832–2832.

39. van der Mei IAF, Ponsonby AL, Dwyer T, Blizzard L, Simmons R, Taylor BV, et al. Past exposure to sun, skin phenotype, and risk of multiple sclerosis: case-control study. BMJ. 2003 Aug;327(7410):316–0.

40. Kampman MT, Wilsgaard T, Mellgren SI. Outdoor activities and diet in childhood and adolescence relate to MS risk above the Arctic Circle. Journal of Neurology. 2007 Apr;254(4):471–7.

41. Rhead B, Bäärnhielm M, Gianfrancesco M, Mok A, Shao X, Quach H, et al. Mendelian randomization shows a causal effect of low vitamin D on multiple sclerosis risk. Neurology Genetics. 2016 Oct;2(5):e97–e97.

42. Gallone G, Haerty W, Disanto G, Ramagopalan SV, Ponting CP, Berlanga-Taylor AJ. Identification of genetic variants affecting vitamin D receptor binding and associations with autoimmune disease. Human Molecular Genetics. 2017 Jun;26(11):2164–76.

43. Lopez-Olivo MA, Siddhanamatha HR, Shea B, Tugwell P, Wells GA, Suarez-Almazor ME. Methotrexate for treating rheumatoid arthritis. Cochrane Database of Systematic Reviews [Internet]. 2014 [cited 2022 Apr 25];(6). Available from: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD000957.pub2/full

44. Lindqvist E, Jonsson K, Saxne T, Eberhardt K. Course of radiographic damage over 10 years in a cohort with early rheumatoid arthritis. Ann Rheum Dis. 2003 Jul;62(7):611–6.

45. Rezaei H, Saevarsdottir S, Forslind K, Albertsson K, Wallin H, Bratt J, et al. In early rheumatoid arthritis, patients with a good initial response to methotrexate have excellent 2-year clinical outcomes, but radiological progression is not fully prevented: data from the methotrexate responders population in the SWEFOT trial. Ann Rheum Dis. 2012 Feb;71(2):186–91.

46. Vastesaeger N, Xu S, Aletaha D, St Clair EW, Smolen JS. A pilot risk model for the prediction of rapid radiographic progression in rheumatoid arthritis. Rheumatology. 2009 Sep 1;48(9):1114–21.

47. Breedveld FC, Weisman MH, Kavanaugh AF, Cohen SB, Pavelka K, van Vollenhoven R, et al. The PREMIER study: A multicenter, randomized, double-blind clinical trial of combination therapy with adalimumab plus methotrexate versus methotrexate alone or adalimumab alone in patients with early, aggressive rheumatoid arthritis who had not had previous methotrexate treatment. Arthritis Rheum. 2006 Jan;54(1):26–37.

48. Romão VC, Canhão H, Fonseca JE. Old drugs, old problems: where do we stand in prediction of rheumatoid arthritis responsiveness to methotrexate and other synthetic DMARDs? BMC Medicine. 2013 Jan 23;11(1):17.

49. Brown PM, Pratt AG, Isaacs JD. Mechanism of action of methotrexate in rheumatoid arthritis, and the search for biomarkers. Nat Rev Rheumatol. 2016 Dec;12(12):731–42.

50. Kim YI, Logan JW, Mason JB, Roubenoff R. DNA hypomethylation in inflammatory arthritis: reversal with methotrexate. The Journal of laboratory and clinical medicine. 1996;128:165–72.

51. Ellis JA, Munro JE, Chavez RA. Genome-scale case-control analysis of CD4+ T- cell DNA methylation in juvenile idiopathic arthritis reveals potential targets involved in disease. Clinical epigenetics. 2012;4(20).

52. Plant D, Wilson AG, Barton A. Genetic and epigenetic predictors of responsiveness to treatment in RA. Nature Reviews Rheumatology. 2014 Feb;10:329–329.

## Chapter 2 - Pregnancy does not modify risk of multiple sclerosis in genetically susceptible women


### Abstract

*Objective*: To utilize the case-only gene-environment (G×E) interaction study design to estimate interaction between pregnancy before onset of multiple sclerosis (MS) symptoms and established genetic risk factors for MS among white adult females.

*Methods*: We studied 2,497 female MS cases from four cohorts in the United States, Sweden, and Norway with clinical, reproductive, and genetic data. Pregnancy exposure was defined in two ways: 1) ≥ 1 live-birth pregnancy before onset of MS symptoms, and 2) Parity before onset of MS symptoms. We estimated interaction between pregnancy exposure and established genetic risk variants, including a weighted genetic risk score (wGRS) and both HLA and non-HLA variants, using logistic regression and proportional odds regression within each cohort. Within-cohort associations were combined using inverse variance meta-analyses with random effects. The case-only G×E independence assumption was tested in 7,067 individuals without MS.

*Results*: Evidence for interaction between pregnancy exposure and established genetic risk variants, including the strongly associated *HLA-DRB1*15:01* allele and a weighted genetic risk score, was not observed. Results from sensitivity analyses were consistent with observed results.

*Conclusion*: Our findings indicate that pregnancy before symptom onset does not modify risk of MS in genetically susceptible white females.

## Introduction

Multiple sclerosis (MS) is a demyelinating autoimmune disease of the central nervous system with both environmental and genetic risk factors.[3] This progressive disease results in significant disability and decreased quality of life.[4] MS is more prevalent among females than males, and symptoms typically emerge during child-bearing ages, often soon after pregnancy.[4,5] This has led many to hypothesize that female-specific exposures, such as those related to reproduction, pregnancy, and lactation, have a role in MS. Pregnancy appears to have short term beneficial effects on existing MS symptoms,[6] but there is no agreement in the scientific literature about the effect of pregnancy on MS risk in general, or among women with genetic susceptibility to MS.[3,7]

Gene-environment (G×E) interactions, for which the effect of an environmental exposure is modified by specific genotype(s), are believed to contribute substantially to complex disease risk and discovery of these interactions can identify subgroups with higher risk of disease.[8] Studies of pregnancy and MS risk have yielded conflicting results; however, no studies to date, have investigated interaction between pregnancy and genetic susceptibility for risk of MS.[9–18]

We used case-only G×E methods to evaluate interaction between pregnancy *before* symptom onset and known genetic risk factors for MS, including *HLA-DRB1*15:01*. Study participants included 2,497 white individuals from four established MS cohorts based in California, Sweden, and Norway. Analyses were conducted separately within each cohort and then combined with meta-analytic methods. A separate cohort of females without MS was used to test the case-only G×E independence assumption.

## Methods

### *Study Participants*

Participants were selected from the Kaiser Permanente Northern California (KPNC) MS Research Program, the Kaiser Permanente Southern California MS Sunshine Study, the Norwegian MS Registry and Biobank (NOR), and two Swedish MS studies, the Epidemiological Investigation of Multiple Sclerosis (EIMS) and the Genes and Environment in Multiple Sclerosis (GEMS) study.[19–22] MS diagnoses were confirmed by independent neurologists according to the 2005 revised (KPNC, NOR, EIMS/GEMS) or the 2010 revised McDonald diagnostic criteria (MS Sunshine).[23,24] Additional non-MS female members of KPNC were also studied; these individuals were derived from the Genetic Epidemiology Research on Adult Health and Aging (GERA) study.[25] KPNC and KPSC are integrated health services systems and are the largest healthcare providers in California. Their membership is largely representative of their catchment area populations; however, persons of lower socioeconomic status are underrepresented.[26,27]

### *Standard protocol approvals, registrations, and patient consents*

All study participants provided written informed consent and the Institutional Review Boards of KPNC, KPSC, regional ethical review boards in Norway and Sweden, and the University of California, Berkeley approved the study protocols.

*KPNC Participants*

Study recruitment is described in detail elsewhere.[19] Briefly, study participants were recruited from KPNC membership between 2006-2014. Prevalent MS cases were the focus of recruitment. Participants were aged between 18-69 years and were KPNC members at initial contact.

*MS Sunshine Participants*

MS Sunshine is a multi-ethnic case-control study of incident MS and first demyelinating event.[20] Participants in this study were recruited from KPSC membership between 2011-2014. At the time of initial contact, participants were KPSC members, age≥18, and diagnosed with MS within 1.5 years or had MS symptom onset within the three years prior to recruitment.

*NOR Participants*

NOR participants were recruited from the Oslo MS Registry and DNA biobank in 2011-2012.[22] The Oslo MS Registry and biobank was established in 1990 and includes clinical data and DNA samples from a population based MS cohort.

*EIMS/GEMS Participants*

EIMS and GEMS are Swedish population-based case-control studies.[21] At enrollment, EIMS participants were age 18-70 years and had recently (within 2 years) confirmed MS. GEMS participants were identified from the Swedish National MS registry and recruited between 2009-2011. All EIMS participants were distinct from GEMS study.

*GERA Participants (Non-cases)*

GERA participants are a subsample of the Research Program on Genes, Environment, and Health cohort.[25] Participants were respondents to a mailed survey sent in 2007 to individuals who had been members of KPNC for at least two years and were age≥18. After providing informed consent, participants were asked to submit a saliva sample for DNA genotyping. Additional health data was obtained from KPNC electronic health records. Participants included in this study were confirmed to not have MS or other autoimmune disease.

*Ancestry Determination*

KPNC, MS Sunshine, and GERA participants with average European ancestry proportions greater than 80% estimated from ancestry informative markers who did not did not self-report Hispanic ethnicity were included.[28] All NOR and EIMS/GEMS samples were self-reported White and population outliers were identified with principal components analysis (PCA) and excluded.[29]

*Genotype and Exposure Assessment*

All participants in this study completed an interview or self-report questionnaires related to MS disease events, reproductive history, and environmental exposures.[19–22] Participants provided

blood or saliva samples for genotyping. Genome-wide single nucleotide polymorphism (SNP) genotyping was performed using the Illumina Infinium 660K BeadChip Array and Human Omni Express Array (KPNC, MS Sunshine), Illumina ImmunoArray BeadChip Array (NOR, EIMS/GEMS), and Affymetrix Axiom Array (GERA). Genome-wide SNP imputation was performed using SHAPEIT2 and IMPUTE2.[30] HLA alleles were imputed using SNP2HLA.[31] Participants with missing genotypes that met QC thresholds (info score > 0.8, missingness per SNP < 0.05, missingness per cohort < 0.05, and minor allele frequency (MAF) > 0.05) were imputed using the average MAF within each cohort.

Pregnancy exposure was evaluated as a dichotomous variable, having *at least one live-birth pregnancy* before reported age of first MS symptom onset, and as an ordinal variable, the *number of live-birth pregnancies* before reported age of first MS symptom onset (Table 1). Age of pregnancy was defined as mother's age at birth and age of symptom onset for MS cases was determined through review of medical records and/or comprehensive clinical histories for each participant collected through interview or questionnaire.

*Candidate Genes and Weighted Genetic Risk Score*

To maximize power and to identify variants with functional relevance to MS, we conducted a two-tiered analysis. In the primary analysis, we assessed G×E interaction between pregnancy exposure and 1) a weighted genetic risk score (wGRS) comprised of recently established non-HLA MS genome wide association study (GWAS) variants, 2) *HLA-DRB1*15:01* alleles, and 3) *HLA-A*02:01* alleles.[1] We also evaluated evidence for effect modification of the interaction between *HLA-A*02:01* and *HLA-DRB1*15:01* for MS susceptibility by pregnancy exposure.[32] HLA variants were excluded from the wGRS because our objective was to use the wGRS to evaluate the combined contribution of non-HLA MS GWAS variants to the G×E interaction with pregnancy for MS susceptibility. In the secondary analysis, we individually tested established MS-associated variants (HLA: 2 variants, non-HLA: 144 variants) that passed QC criteria in all four cohorts. Individual genotypes were modeled assuming a linear effect of each additional risk allele (0, 1, or 2 alleles).

The wGRS was derived by multiplying the log odds-ratio for each risk allele from recent GWAS by the number of risk alleles carried by each participant and summing across GWAS variants (Table e-1).[1] Scores for each individual were calculated using non-HLA risk variants that passed genotype QC thresholds within each cohort (KPNC, MS Sunshine, and EIMS/GEMS: 182 non-HLA variants, NOR: 144 non-HLA variants, GERA: 161 non-HLA variants). Within each cohort the wGRS was modeled as a continuous variable and with quartiles.

*Statistical Analysis*

Genome-wide SNP genotyping data were available for 990 KPNC, 409 MS Sunshine, 153 NOR, and 1,462 EIMS/GEMS female MS cases. Following exclusion of participants with European ancestry proportions <80% (KPNC, MS Sunshine) and population outliers identified from PCA analysis (NOR, EIMS/GEMS), 814 KPNC, 151 MS Sunshine, 119 NOR, and 1,413 EIMS/GEMS participants had age of onset ≥18 years, complete pregnancy history data, and genome-wide genetic profiles. The final dataset for G×E analysis included 2,497 participants (Table 1, Figure e-1).

Case-only G×E models rely on the assumption that the genotype and environmental exposure are uncorrelated in the source population. If that assumption is valid, the association between genotype and exposure in a case-only model estimates the departure of the joint of effects on the odds-ratio scale from case-control regression. If the disease is rare and the above assumption is valid, the case-only model estimates the departure of joint effects on the risk-ratio scale.[33] An OR>1 indicates an increase in the risk of MS above what is expected given exposure to both G and E factors and an OR<1 indicates a decrease in the risk of MS given exposure to both G and E factors.
We used logistic regression to model having at least one live-birth pregnancy before symptom onset as a function of genotypes or wGRS:

$$logit[P(E = 1|G)] = \beta_0 + \beta_{GE}G.$$

$E$ is an indicator for having at least one live-birth pregnancy before symptom onset or not, $G$ is 0, 1, or 2 alleles for a risk variant or continuous wGRS, and $exp(\beta_{GE})$ is the estimate of the interaction parameter measuring the departure of the joint effects of $E$ and $G$ on the multiplicative scale.[33]

Proportional odds regression was used to model parity before symptom onset,[33] where the probability of an equal or less than number of pregnancies prior to symptom onset, $E \leq k$, to the probability of more than that number of pregnancies, $E > k$, as a function of genotypes or wGRS:

$$log\frac{P(E \leq k|G)}{P(E > k|G)} = \gamma_k - \gamma_{GE}G, \text{for } k \text{ in } 0,1,2,3,4.$$

$E$ is the number of live-birth pregnancies before MS symptom onset, $G$ is 0, 1, or 2 alleles for a risk variant or continuous wGRS, $k$ is the threshold for live-birth pregnancies before symptom onset for each ordinal comparison, and $exp(\gamma_{GE})$ is the estimate of the interaction parameter measuring the departure of the joint effects of $E$ and $G$ on the multiplicative scale.[33]

G×E interaction was estimated within each study and combined estimates of G×E interaction were obtained with random effects meta-analysis using restricted maximum likelihood estimation with weights proportional to the inverse of the variance for each cohort-specific association. $I^2\%$ was used to assess heterogeneity between cohort-specific associations. All models were adjusted for age of MS onset and population stratification using components from PCA. wGRS quartiles were modeled using dummy variables with the first quartile as the reference category. Secondary discovery analysis p-values were adjusted for the false discovery rate using the Benjamini-Hochberg method and adjusted $p$-values<0.05 were considered significant.[34]

Effect modification of the interaction between *HLA-A*02:01* and *HLA-DRB1*15:01* by pregnancy exposure was evaluated separately by stratifying case-only regression models for pregnancy exposure and *HLA-DRB1*15:01* by the absence and presence of *HLA-A*02:01* alleles adjusting for age of MS onset and population stratification.

We tested the assumption of G×E independence between genotype and pregnancy using healthy female GERA participants (N=7,067).[8] Logistic and proportional odds regression models described above were used to test for associations between MS genetic risk factors, including wGRS and *HLA-DRB1*15:01*, and having ≥ 1 pregnancy or not and parity. All statistical analyses were conducted using Plink 1.9 and R 3.5.1.

*Sensitivity Analyses*

We estimated GxE association between *live or non-live birth pregnancy* before symptom onset (binary and ordinal models) and genetic variants among KPNC and MS Sunshine participants (data not available for NOR and EIMS/GEMS) to identify bias from exclusion of non-live birth pregnancies in primary analyses. To determine if there was a short-term effect of pregnancy on MS risk, we altered the definition of pregnancy exposure to only consider first pregnancies that occurred within 5-years prior to age of onset as "exposed". Next, to rule out reverse-causality from MS disease latency or recall bias, we considered age of symptom onset at 5 years prior to reported age of onset. To check for bias from differing numbers of variants used in wGRSs between cohorts, GxE interactions for wGRSs derived from risk variants only found in NOR data (144 variants, Table e-1) were estimated for each cohort and combined with meta-analysis.

*Data availability statement*

Anonymized data from KPNC, MS Sunshine, NOR, and EIMS/GEMS used in this study will be shared on request from any qualified investigator pending Institutional Review Board approval at each site. GERA data is publicly availability on dbGaP (phs000674.v2.p2).

**Results**

Characteristics of MS cases are summarized in Table 1. Age of MS onset among the cohorts that recruited predominantly prevalent cases (KPNC, NOR, and EIMS/GEMS ) was similar (mean = 33.1, 33.4 and 34.4 years, respectively), but occurred later among the cohort that recruited incident cases only (MS Sunshine, mean = 39.9 years). Year of symptom onset occurred earlier among KPNC and NOR (median = 1992), than EIMS/GEMS (median=2005) and MS Sunshine (median=2011). The average age of first pregnancy (live birth) ranged between 24.6 to 27.2 years with KPNC and MS Sunshine participant pregnancies occurring earlier than NOR and EIMS/GEMS. More than half of participants were *HLA-DRB1*15:01* carriers; NOR and EIMS/GEMS had a higher proportion of carriers (60% and 56%, respectively) than KPNC and MS Sunshine (54% and 47%, respectively). The median wGRS was similar across KPNC, MS Sunshine and EIMS/GEMS. NOR participants had lower scores than the others, which is likely because fewer SNPs were available to calculate the NOR wGRS. Approximately half (49.8%) of the participants had a live birth pregnancy before onset of MS symptoms. Parous participants were most likely to have two live-births before symptom onset.

In this study, the odds-ratios from case-only regression models estimate the departure of the multiplicative joint effects of E and G on the risk-ratio scale for susceptibility to MS. We did not find evidence for G×E interaction between pregnancy exposure and primary genetic risk factors (wGRS, *HLA-DRB1*15:01*, and *HLA-A*02:01)* (Table 2, Figure e-2). Point estimates were close

to the null or had confidence intervals that contained the null. Estimates were similar for both pregnancy exposures. Significant evidence for effect modification of interaction between *HLA-DRB1*15:01* and pregnancy exposure by carriage of *HLA-A*02:01* alleles was not observed although estimates indicate possible protective effects (absence *HLA*02:01*: OR: 0.89, 95% CI 0.70, 1.13; presence *HLA*02:01*: OR: 1.09, 95% CI 0.71, 1.66). No variants tested in the secondary discovery analysis had multiple testing adjusted p-values<0.05 (Tables e-2 and e-3).

Among GERA controls, the wGRS and *HLA-DRB1*15:01* were not associated with having at least one pregnancy or with parity (Table 4) and none of the MS GWAS loci were significantly associated with pregnancy in GERA after correcting for multiple testing (Results not shown). Results from sensitivity analyses investigating live or non-birth pregnancy, pregnancy within 5 years of MS onset, and bias from latent disease were consistent with observed results (Table 5). Estimates for interaction between pregnancy exposures and the wGRSs derived from variants present in NOR data were similar to observed results (Table e-2 and e-3).

**Discussion**

We hypothesized that pregnancy before MS symptom onset modifies the risk of MS in genetically susceptible females. Using data from four study populations, we did not find evidence to support G×E interaction between established genetic risk factors for MS and exposure to pregnancy before symptom onset. The *HLA-A*02:01* +/- stratified point estimates suggest a protective effect of *HLA-A*02:01* alleles in the relationship between *HLA-DRB1*15:01* and pregnancy exposure, but our study was not sufficiently powered to detect these associations. Evidence for interaction between pregnancy and non-HLA or HLA variants considered individually was also not observed.

Although we considered pregnancy as a single environmental exposure, pregnancy is a complex and heterogeneous combination of physiologic changes that result in weight gain, increases in lipid levels, and changes in basal metabolic rate, among others.[7] These physiologic changes are the product of pregnancy induced modifications in hormones, such as estriol, progesterone, prolactin, early pregnancy growth factor, alpha-fetoprotein, and leptin as well as elevated levels of other growth factors. There are increases in circulated regulatory T-cells and B-cells, increased Th2 responses, and decreased Th1-and Th-17 immune responses.[7] These immune changes are important for fetal-tolerance, as the maternal immune system and endocrine pathways respond to fetal-antigens that circulate in the mother.[35] Following pregnancy, hormone levels and immune adaptions quickly return to pre-pregnancy states.[36] The reduction of MS relapse rate during pregnancy is attributed to the dynamic immune and endocrine alterations that result from maternal—fetal crosstalk during pregnancy.[37] Little is known about how physiologic changes during pregnancy affect the risk of developing MS, but it is hypothesized that the pregnancy induced changes in endocrine pathways and immune system have protective effects.[7]

Epidemiologic studies investigating the effect of pregnancy on MS risk have reported conflicting results.[3] Five studies reported a protective association between parity and risk of MS; however, two of these studies attributed their protective associations to reverse causality from reduced fertility and increased likelihood of miscarriage among women with latent MS.[9,10,12,13,17] Five additional studies reported no association between pregnancy and risk of MS.[11,14–16,18] A recent

study investigating breastfeeding, ovulatory years, and risk of MS found evidence that cumulative duration of breastfeeding is associated with decreased risk of MS.[38] The authors suggest that breastfeeding duration confounds the association between parity/pregnancy and risk of MS, and may explain previously reported conflicting findings.

Genes within the HLA complex contribute substantially to MS, with the *HLA-DRB1*15:01* allele conferring the largest known genetic risk for disease.[1] Interactions between this allele and environmental exposures such as tobacco smoking, Epstein-Barr Virus infection, and adolescent obesity have previously demonstrated large effect sizes.[3] A recent GWAS identified approximately 200 non-HLA variants associated with MS risk.[1]

Previous research reported that protective associations between pregnancy and MS onset were attributable to reverse causality and that pregnancy was more likely among women with less severe disease.[9,10] However, a registry based study in Norway found that pregnancies among women with MS prior to symptom onset have birthweights and outcomes similar to those in non-MS women.[39] Furthermore, changes in counseling and availability of safer disease modifying therapies as well as new diagnostic criteria that allow for earlier MS diagnosis have likely contributed to a recent reported increase in the rate of pregnancy among women with MS.[40,41] Results from our sensitivity analyses did not demonstrate evidence for bias from latent disease or from pregnancy within 5-years before symptom onset. Including data on non-live births in the pregnancy exposure moved point estimates further from the null, but the confidence intervals still included the null.

Case-only G×E methodologies were developed to address a primary challenge for studying interactions, statistical power.[8] Since they were first introduced, advances in case-only G×E methodology have focused on combining the increased power from case-only methods with evidence for the G×E independence assumption from case-control data.[8] Healthy controls with reproductive history matched on case symptom onset were not available for our large combined dataset of MS cases; however, we utilized non-MS female participants from GERA to formally test for evidence of independence between MS genetic risk factors and pregnancy. Results from this analysis support the validity of our findings. Our models were adjusted for age of MS onset and population stratification. While there may be additional variables that confound the relationship between both pregnancy and MS and genetic risk factors and MS, case-only G×E analyses only need to adjust for confounders of the G×E relationship. We cannot rule out that our findings may be due to confounding from unknown factors.

This was the first study to investigate G×E interaction between pregnancy and established MS genetic risk variants in MS. A primary strength of the current study was the large sample size with complete genetic and reproductive history data. Additionally, cases in this study were largely representative of their respective populations, providing support for external validity, and identification of cases from integrated health services delivery systems and national registries reduced the probability of selection bias. Further, we found evidence to support the independence of G and E factors among the KPNC source population. Results from sensitivity analyses suggest that our findings are not attributable to reverse causality; however, given the uncertainty regarding disease latency and first MS symptoms we cannot conclusively rule out reverse causality. Our study was subject to some limitations. Due to the absence of suitable

controls, we were unable to evaluate the association between pregnancy and MS. Our G×E associations were consistent between studies; however, differences between the study populations may have biased our findings. With the exception of MS Sunshine, the other studies relied on older MS diagnostic criteria, which may have excluded milder MS cases. Due to the case-only study design, we were not able to assess interaction on the additive scale and our null findings on the multiplicative scale cannot rule out the presence of interaction on the additive scale. Additionally, we were not able to assess G and E independence in non-KPNC source populations. Future studies should investigate the effects of pregnancy on MS risk among non-European populations. The current study was focused on white individuals to achieve the statistical power required for analyses. It is possible that interaction between MS genetic risk factors and pregnancy may differ by ethnic/ancestral group. If controls with reproductive data matched to case age of symptom onset are available, G×E interaction should be assessed among cases and controls with methods used to combine case-only and case-control interactions.

Our findings suggest that genetic susceptibility to MS does not modify the association between pregnancy and MS. This information may be useful for counseling women who have genetic susceptibility to MS about decisions to pursue pregnancy, although further research is needed to determine the effect of pregnancy on MS susceptibility.

## References


1. International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. Science. 2019 Sep 27;365(6460):eaav7188.

2. Kim K, Bang SY, Lee HS, Cho SK, Choi CB, Sung YK, et al. High-density genotyping of immune loci in Koreans and Europeans identifies eight new rheumatoid arthritis risk loci. Ann Rheum Dis. 2015 Mar;74(3):e13.

3. Waubant E, Lucas R, Mowry E, Graves J, Olsson T, Alfredsson L, et al. Environmental and genetic risk factors for MS: an integrated review. Vol. 6, Annals of Clinical and Translational Neurology. Wiley-Blackwell; 2019. p. 1905–22.

4. Nelson LM, Wallin MT, Marrie RA, Culpepper WJ, Langer-Gould A, Campbell J, et al. A new way to estimate neurologic disease prevalence in the United States: Illustrated with MS. Neurology. 2019 Mar 5;92(10):469–80.

5. Wallin MT, Culpepper WJ, Nichols E, Bhutta ZA, Gebrehiwot TT, Hay SI, et al. Global, regional, and national burden of multiple sclerosis 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet Neurology. 2019 Mar 1;18(3):269–85.

6. Houtchens MK, Edwards NC, Phillips AL. Relapses and disease-modifying drug treatment in pregnancy and live birth in US women with MS. Neurology. 2018;91(17):E1570–8.

7. McCombe PA. The Short and Long-Term Effects of Pregnancy on Multiple Sclerosis and Experimental Autoimmune Encephalomyelitis. Journal of clinical medicine. 2018 Nov 28;7(12).

8. Gauderman WJ, Mukherjee B, Aschard H, Hsu L, Lewinger JP, Patel CJ, et al. Update on the State of the Science for Analytical Methods for Gene-Environment Interactions. American Journal of Epidemiology. 2017 Oct 1;186(7):762–70.

9. Nielsen NM, Jørgensen KT, Stenager E, Jensen A, Pedersen B V, Hjalgrim H, et al. Reproductive history and risk of multiple sclerosis. Epidemiology (Cambridge, Mass). 2011 Jul;22(4):546–52.

10. Hedström AK, Hillert J, Olsson T, Alfredsson L. Reverse causality behind the association between reproductive history and MS. 2014 Apr;20(4).

11. Khashan AS, Kenny LC, Laursen TM, Mahmood U, Mortensen PB, Henriksen TB, et al. Pregnancy and the risk of autoimmune disease. PLoS ONE. 2011 May;6(5):e19658.

12. Ponsonby AL, Lucas RM, van der Mei IA, Dear K, Valery PC, Pender MP, et al. Offspring number, pregnancy, and risk of a first clinical demyelinating event: The AusImmune Study. Neurology. 2012 Mar 20;78(12):867–74.

13. Magyari M, Koch-Henriksen N, Pfleger CC, Sørensen PS. Reproduction and the risk of multiple sclerosis. Multiple Sclerosis Journal. 2013;19(12):1604–9.

14. Alonso Á, Jick SS, Olek MJ, Ascherio A, Jick H, Hernán MA. Recent use of oral contraceptives and the risk of multiple sclerosis. Archives of Neurology. 2005 Sep;62(9):1362–5.

15. Hernán MA, Hohol MJ, Olek MJ, Spiegelman D, Ascherio A. Oral contraceptives and the incidence of multiple sclerosis. Neurology. 2000 Sep 26;55(6):848–53.

16. Leibowitz U, Antonovsky A, Kats R, Alter M. Does pregnancy increase the risk of multiple sclerosis? Journal of neurology, neurosurgery, and psychiatry. 1967;30(4):354–7.

17. Runmarker B, Andersen O. Pregnancy is associated with a lower risk of onset and a better prognosis in multiple sclerosis. Brain. 1995 Feb;118(1):253–61.

18. Villard-Mackintosh L, Vessey MP. Oral contraceptives and reproductive factors in multiple sclerosis incidence. Contraception. 1993;47(2):161–8.

19. Briggs FBS, Acuña BS, Shen L, Bellesis KH, Ramsay PP, Quach H, et al. Adverse socioeconomic position during the life course is associated with multiple sclerosis. J Epidemiol Community Health. 2014 Jul;68(7):622–9.

20. Langer-Gould A, Brara SM, Beaber BE, Zhang JL. Incidence of multiple sclerosis in multiple racial and ethnic groups. Neurology. 2013 May;80(19):1734–9.

21. Hedström AK, Hillert J, Olsson T, Alfredsson L. Alcohol as a Modifiable Lifestyle Factor Affecting Multiple Sclerosis Risk. JAMA Neurology. 2014 Mar 1;71(3):300.

22. Gustavsen MW, Page CM, Moen SM, Bjølgerudl A, Berg-Hansen P, Nygaard GO, et al. Environmental exposures and the risk of multiple sclerosis investigated in a Norwegian case-control study. BMC Neurology. 2014 Dec 3;14(1):196.

23. Polman CH, Reingold SC, Edan G, Filippi M, Hartung HP, Kappos L, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". Annals of neurology. 2005 Dec;58(6):840–6.

24. Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria. Annals of Neurology. 2011 Feb;69(2):292–302.

25. Banda Y, Kvale MN, Hoffmann TJ, Hesselson SE, Ranatunga D, Tang H, et al. Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the genetic epidemiology research on adult health and aging (GERA) cohort. Genetics. 2015 Aug 1;200(4):1285–95.

26. Krieger N. Overcoming the absence of socioeconomic data in medical records: validation and application of a census-based methodology. Am J Public Health. 1992 May;82(5):703–10.

27. Koebnick C, Langer-Gould AM, Gould MK, Chao CR, Iyer RL, Smith N, et al. Sociodemographic characteristics of members of a large, integrated health care system: comparison with US Census Bureau data. Perm J. 2012;16(3):37–41.

28. Chen CY, Pollack S, Hunter DJ, Hirschhorn JN, Kraft P, Price AL. Improved ancestry inference using weights from external reference panels. Bioinformatics. 2013;29(11):1399–406.

29. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics. 2006 Aug;38(8):904–9.

30. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. G3: Genes, Genomes, Genetics. 2011 Nov 1;1(6):457–70.

31. Jia X, Han B, Onengut-Gumuscu S, Chen WM, Concannon PJ, Rich SS, et al. Imputing amino acid polymorphisms in human leukocyte antigens. PLoS ONE. 2013 Jun;8(6):e64683.

32. Moutsianas L, Jostins L, Beecham AH, Dilthey AT, Xifara DK, Ban M, et al. Class II HLA interactions modulate genetic risk for multiple sclerosis. Nat Genet. 2015 Oct;47(10):1107–13.

33. Clarke GM, Morris AP. A Comparison of sample size and power in case-only association studies of gene-environment interaction. American Journal of Epidemiology. 2010;171(4):498–505.

34. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological). 1995;57(1):289–300.

35. Kinder JM, Stelzer IA, Arck PC, Way SS. Immunological implications of pregnancy-induced microchimerism. Nature Reviews Immunology. 2017 May 8;17(8):483–94.

36. Arck PC, Hecher K. Fetomaternal immune cross-talk and its consequences for maternal and offspring's health. Nature Medicine. 2013 May 1;19(5):548–56.

37. Patas K, Engler JB, Friese MA, Gold SM. Pregnancy and multiple sclerosis: feto-maternal immune cross talk and its implications for disease activity. Journal of Reproductive Immunology. 2013 Mar;97(1):140–6.

38. Langer-Gould A, Smith JB, Hellwig K, Gonzales E, Haraszti S, Koebnick C, et al. Breastfeeding, ovulatory years, and risk of multiple sclerosis. Neurology. 2017 Aug 8;89(6):563–9.

39. Dahl J, Myhr KM, Daltveit AK, Gilhus NE. Pregnancy, delivery and birth outcome in different stages of maternal multiple sclerosis. Journal of Neurology. 2008;

40. Langer-Gould AM. Pregnancy and family planning in multiple sclerosis. Vol. 25, CONTINUUM Lifelong Learning in Neurology. Lippincott Williams and Wilkins; 2019. p. 773–92.

41. Houtchens MK, Edwards NC, Schneider G, Stern K, Phillips AL. Pregnancy rates and outcomes in women with and without MS in the United States. Neurology. 2018;91(17):E1559–69.

## Tables

**Table 1.** Characteristics of n=2,497 participants.

| | KPNC | MS Sunshine | NOR | EIMS/GEMS |
|---|---|---|---|---|
| n | 814 | 151 | 119 | 1413 |
| Age Onset mean (SD) | 33.1 (9.1) | 39.9 (11.5) | 33.4 (9.4) | 34.4 (10.1) |
| Year of Onset median [IQR] | 1991 [1983, 1997] | 2011 [2009, 2013] | 1992 [1986, 1997] | 2005 [2001, 2008] |
| Age 1st pregnant[a] mean (SD) | 24.6 (5.3) | 25.9 (5.8) | 27.1 (6.0) | 27.2 (5.1) |
| ≥1 pregnancy[a] before symptom onset n (%) | 428 (51.7) | 95 (60.1) | 69 ( 58.0) | 668 (47.3) |
| No. pregnancies[a] before symptom onset n (%) | | | | |
| 0 | 400 (48.3) | 63 (39.9) | 50 ( 42.0) | 745 (52.7) |
| 1 | 146 (17.6) | 26 (16.5) | 25 ( 21.0) | 180 (12.7) |
| 2 | 198 (23.9) | 41 (25.9) | 30 ( 25.2) | 350 (24.8) |
| 3 | 67 ( 8.1) | 21 (13.3) | 14 ( 11.8) | 106 ( 7.5) |
| ≥4 | 17 ( 2.1) | 7 ( 4.4) | 0 ( 0.0) | 32 ( 2.3) |
| *HLA-DRB1*15:01* carrier n (%) | 446 (53.9) | 74 (46.8) | 71 ( 59.7) | 789 (55.8) |
| wGRS[b] median [IQR] | 23.3 [22.7, 23.9] | 23.2 [22.7, 23.8] | 18.0 [17.5, 18.6] | 22.6 [22.0, 23.3] |

Abbreviations: EIMS, Environment in Multiple Sclerosis Study; GEMS, Genes Environment in Multiple Sclerosis; HLA, Human Leukocyte Antigen region; IQR, Interquartile range; KPNC, Kaiser Permanente Northern California; MS, Multiple Sclerosis; No., number; NOR, Norway; wGRS, weighted genetic risk score.

[a] Pregnancy defined as pregnancy resulting in live-birth.

[b] wGRS calculated from 182 MS risk loci for KPNC, MS Sunshine, EIMS/GEMS and 144 MS risk loci for NOR participants.

**Table 2.** Results from case-only meta-analyses estimating multiplicative interaction between *HLA-DRB1*15:01, HLA-A*02:01*, and weighted genetic risk scores with pregnancy before symptom onset.

| | Crude | | Adjusted | |
|---|---|---|---|---|
| Variant | OR[a] | 95% CI | OR[a] | 95% CI |
| *≥1 live-birth before symptom Onset*[b] | | | | |
| *HLA-DRB1*15:01* | 0.93 | 0.82, 1.07 | 0.98† | 0.77, 1.25 |
| *HLA-A*02:01* | 1.06 | 0.93, 1.21 | 0.93 | 0.79, 1.09 |
| wGRS[c] | 0.95 | 0.87, 1.04 | 1.04† | 0.89, 1.21 |
| wGRS Q1[d] | ref | | ref | |
| wGRS Q2 | 1.07 | 0.86, 1.34 | 1.15 | 0.88, 1.5 |
| wGRS Q3 | 1.07 | 0.85, 1.33 | 1.25 | 0.95, 1.64 |
| wGRS Q4 | 0.90 | 0.72, 1.12 | 1.11 | 0.84, 1.45 |
| *Parity before symptom onset*[c] | | | | |
| *HLA-DRB1*15:01* | 0.96 | 0.85, 1.07 | 1.01† | 0.84, 1.2 |
| *HLA-A*02:01* | 1.04 | 0.93, 1.18 | 0.93 | 0.81, 1.06 |
| wGRS[c] | 0.97 | 0.89, 1.05 | 1.05 | 0.96, 1.15 |
| wGRS Q1[d] | ref | | ref | |
| wGRS Q2 | 1.09 | 0.88, 1.34 | 1.18 | 0.94, 1.48 |
| wGRS Q3 | 1.12 | 0.91, 1.37 | 1.29 | 1.00, 1.66 |
| wGRS Q4 | 0.90 | 0.73, 1.11 | 1.06 | 0.84, 1.33 |

Abbreviations: OR, Gene-environment interaction odds ratio; CI, confidence interval; HLA, Human Leukocyte Antigen; MS, Multiple Sclerosis; wGRS, weighted genetic risk score; †: $25\% \leq I^2 < 50\%$, ††: $50\% \leq I^2 < 75\%$, †††: $I^2 > 75\%$.

[a] ORs estimate the departure of the multiplicative joint effects of pregnancy and risk variants on the risk-ratio scale for susceptibility to MS. Cohort specific associations were combined with inverse variance meta-analysis with random effects.

[b] ≥ 1 pregnancy before symptom onset cohort specific ORs and 95% CIs estimated with logistic regression models. Parity before symptom onset cohort specific ORs and 95% CIs estimated with proportional odds regression models. Adjusted models included age of MS onset and principal components for genetic ancestry.

[c] wGRS and modeled as a continuous variable. For KPNC, MS Sunshine, and EIMS/GEMS participants wGRS calculated from 182 non-HLA MS risk variants and for NOR wGRS calculated from 144 non-HLA risk variants.

[d] wGRS modeled with categorical quartiles with the first quartile used as the reference category.

**Table 3.** Estimates of multiplicative interaction between *HLA-DRB1*15:01* and pregnancy exposure stratified by presence and absence of *HLA-A*02:01 alleles*.

| | | Crude | | Adjusted | |
|---|---|---|---|---|---|
| | *HLA-A*02:01* Alleles | OR[a] | 95% CI | OR[a] | 95% CI |
| *≥1 live-birth before symptom Onset*[b] | | | | | |
| *HLA-DRB1*15:01* | - | 1.08†† | 0.78, 1.48 | 1.09†† | 0.71, 1.66 |
| | + | 0.85 | 0.70, 1.03 | 0.89 | 0.70, 1.13 |
| *Parity before symptom onset*[b] | | | | | |
| *HLA-DRB1*15:01* | - | 1.01† | 0.83, 1.23 | 1.04†† | 0.77, 1.41 |
| | + | 0.91 | 0.76, 1.09 | 0.97 | 0.80, 1.18 |

Abbreviations: OR, multiplicative interaction odds ratio; CI, 95% Confidence Interval; HLA, Human Leukocyte Antigen; MS, Multiple Sclerosis; †: 25% ≤ $I^2$ < 50%, ††: 50% ≤ $I^2$ < 75%, †††: $I^2$ > 75%.

[a] ORs estimate the departure of the multiplicative joint effects of pregnancy exposure and *HLA-DRB1*15:01* on the risk-ratio scale for susceptibility to MS stratified by absence and presence of *HLA-A*02:01* alleles. Cohort specific associations were combined with inverse variance meta-analysis with random effects.

[b] ≥ 1 pregnancy before symptom onset cohort specific ORs and 95% CIs estimated with logistic regression models. Parity before symptom onset cohort specific ORs and 95% CIs estimated with proportional odds regression models. Adjusted models included age of MS onset and principal components for genetic ancestry.

**Table 4.** Evidence for GxE independence among n=7,067 healthy Genetic Epidemiology Research on Aging participants. Results from crude and adjusted regression analyses for association between weighted genetic risk score and *HLA-DRB1*15:01* and having at least one live-birth pregnancy or not and number of live-birth pregnancies.

| | Crude | | | Adjusted | | |
|---|---|---|---|---|---|---|
| | OR[a] | 95% CI | p | OR[a] | 95% CI | p |
| ≥1 Live birth pregnancy or not[b] | | | | | | |
| *HLA-DRB1*15:01* | 0.93 | 0.81, 1.11 | 0.37 | 0.89 | 0.77, 1.04 | 0.15 |
| wGRS[c] | 1.02 | 0.93, 1.11 | 0.71 | 1.04 | 0.95, 1.13 | 0.41 |
| Number of live-birth pregnancies[b] | | | | | | |
| *HLA-DRB1*15:01* | 0.95 | 0.86, 1.04 | 0.23 | 0.92 | 0.84, 1.01 | 0.08 |
| wGRS[c] | 0.95 | 0.90, 1.00 | 0.04 | 0.96 | 0.91, 1.01 | 0.14 |

Abbreviations: OR, Gene-environment interaction odds ratio; CI, confidence interval; HLA, Human Leukocyte Antigen; MS, Multiple Sclerosis; wGRS, weighted genetic risk score.

[a] ORs estimate the departure of the multiplicative joint effects of pregnancy exposure and wGRS / *HLA-DRB1*15:01* on the risk-ratio scale for susceptibility to MS among healthy female controls from Genetic Epidemiology Research on Aging participants.

[b] ≥ 1 pregnancy before symptom onset ORs and 95% CIs estimated with logistic regression models. Parity before symptom onset ORs and 95% CIs estimated with proportional odds regression models. Adjusted models included principal components for genetic ancestry.

[c] wGRS calculated 161 non-HLA risk variants in GERA data after QC and modeled as a continuous variable.

**Table 5.** Sensitivity analysis results for pregnancy defined as live or non-live birth pregnancy, first pregnancy within 5 years prior to symptom onset, and latent MS onset models.

| Variant/Risk Score | Model | ≥1 live-birth before symptom Onset | | Parity before symptom onset | |
|---|---|---|---|---|---|
| | | OR[a] | 95% CI | OR[a] | 95% CI |
| *HLA-A*02:01* | Observed[b] | 0.93 | 0.79, 1.09 | 0.92 | 0.81, 1.05 |
| | Live or non-live[c] | 0.84† | 0.53, 1.34 | 0.93 | 0.76, 1.13 |
| | Short-term[d] | 0.96 | 0.76, 1.21 | 0.84 | 0.69, 1.03 |
| | Latent[e] | 0.96 | 0.81, 1.14 | 0.98 | 0.85, 1.13 |
| *HLA-DRB1*15:01* | Observed[b] | 0.98† | 0.77, 1.25 | 1.01† | 0.84, 1.21 |
| | Live or non-live[c] | 0.79 | 0.62, 1.00 | 0.85 | 0.71, 1.03 |
| | Short-term[d] | 1.03† | 0.73, 1.45 | 0.97 | 0.78, 1.21 |
| | Latent[e] | 0.97 | 0.82, 1.14 | 1.00 | 0.87, 1.15 |
| wGRS[f] | Observed[b] | 1.04† | 0.89, 1.21 | 1.05 | 0.96, 1.15 |
| | Live or non-live[c] | 0.94 | 0.79, 1.11 | 1.06† | 0.84, 1.33 |
| | Short-term[d] | 1.02 | 0.87, 1.19 | 1.07 | 0.94, 1.21 |
| | Latent[e] | 1.08†† | 0.83, 1.41 | 1.05 | 0.96, 1.16 |

Abbreviations: OR, Gene-environment interaction odds ratio; CI, confidence interval; HLA, Human Leukocyte Antigen; MS, Multiple Sclerosis; wGRS, weighted genetic risk score; †: $25\% \leq I^2 < 50\%$, ††: $50\% \leq I^2 < 75\%$, †††: $I^2 > 75\%$.

[a] ORs estimate the departure of the multiplicative joint effects of pregnancy and risk variants on the risk-ratio scale for susceptibility to MS. Cohort specific associations were combined with inverse variance meta-analysis with random effects. ≥ 1 pregnancy before symptom onset cohort specific ORs and 95% CIs estimated with logistic regression models. Parity before symptom onset cohort specific ORs and 95% CI estimated with proportional odds regression models. Adjusted models included age of MS onset and principal components for genetic ancestry.

[b] Estimates from observed data as presented in Table 2.

[c] Estimate of multiplicative interaction between pregnancy exposure defined as have at least one live or non-live birth before MS symptom onset. Data on non-live births only available for Kaiser Permanente and MS Sunshine participants, n = 968.

[d] Estimate of multiplicative interaction between pregnancy exposure defined as have at least one live pregnancy and parity within 5 years before MS symptom onset.

[e] Estimate of multiplicative interaction between pregnancy exposure with MS age of onset adjusted by -5 years.

[f] wGRS modeled as a continuous variable

**Supplement**

Tables e-1, e-2, and e-3 are provided as excel documents.

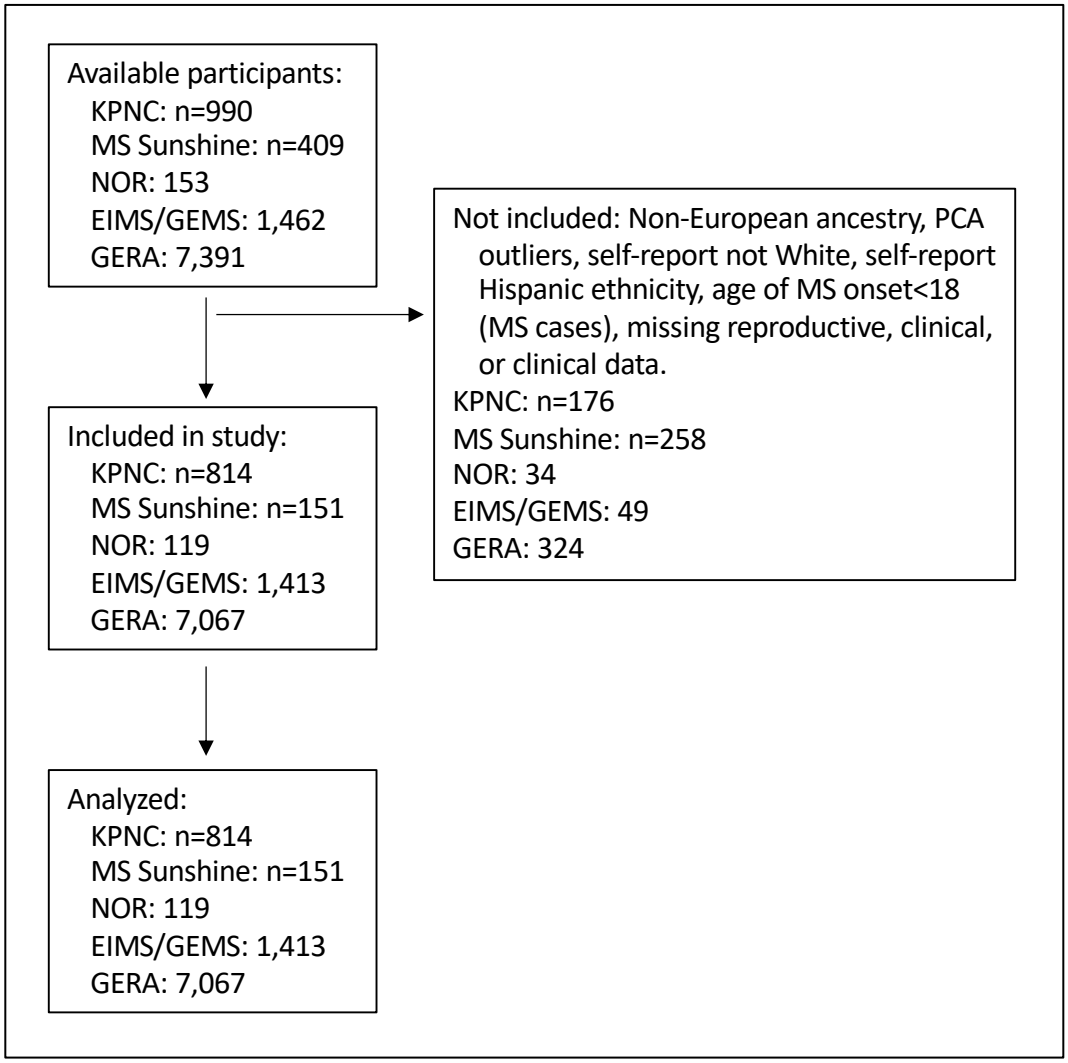



**Figure e-1.** Study flow chart for participants from the Kaiser Permanente Northern California (KPNC) MS Research Program, the Kaiser Permanente Southern California MS Sunshine Study, the Norwegian MS Registry and Biobank (NOR), the Epidemiological Investigation of Multiple Sclerosis and the Genes and Environment in Multiple Sclerosis (EIMS/GEMS) study, and the Genetic Epidemiology Research on Adult Health and Aging (GERA) study.

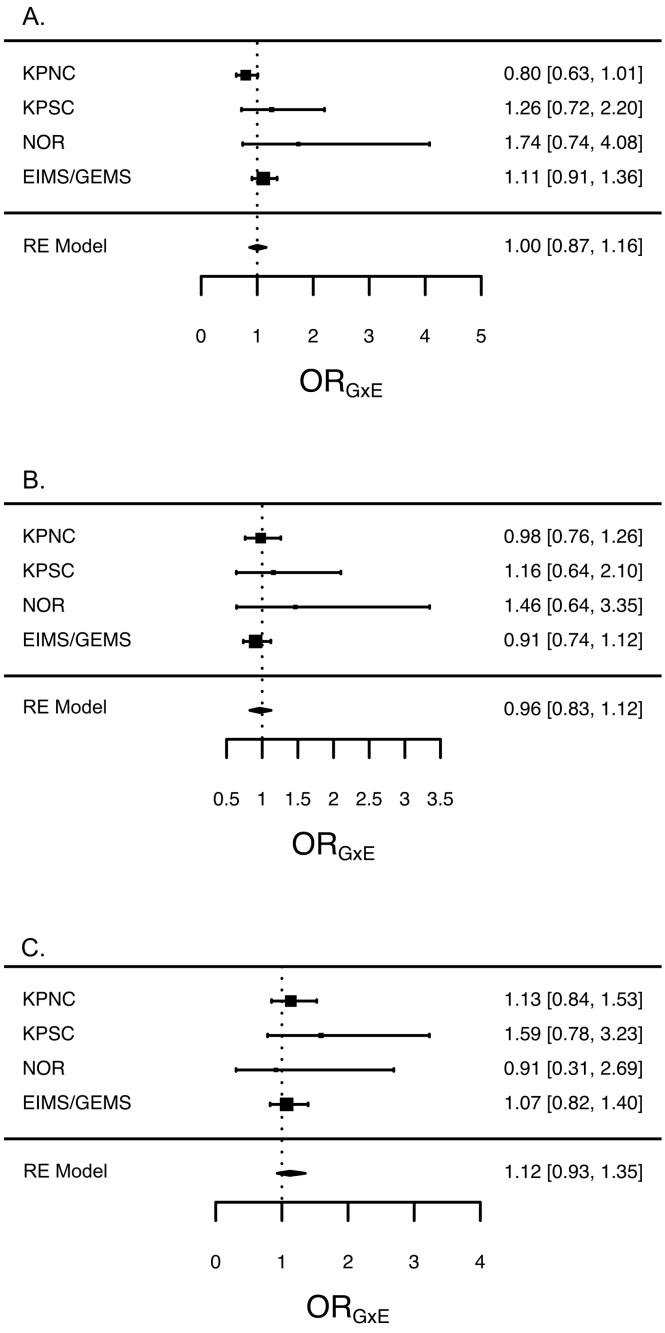



**Figure e-2.** Case-only logistic regression meta-analysis results among each study cohort: A) *HLA-DRB1*15:01*, B) *HLA-A*02:01*, and C) continuous MS weighted genetic risk score. ORs estimate the departure of the multiplicative joint effects of pregnancy and risk variants on the risk-ratio scale for susceptibility to MS.

## Chapter 3 - Evidence supports causal association between allele-specific vitamin D receptor binding and susceptibility to multiple sclerosis among Europeans


### Abstract

*Background and Objectives*: There is evidence for a causal association between 25-hydroxy vitamin D (25(OH)D) serum levels and multiple sclerosis (MS) susceptibility among Europeans, but the underlying mechanisms are unknown. Previous studies have established that 25(OH)D signals through the nuclear vitamin D receptor (VDR), a ligand-regulated transcription factor that modulates vitamin D regulated gene expression. Directly testing for associations between VDR binding and phenotypes in large-scale human studies poses many challenges. Using mendelian randomization methods, SNPs associated with VDR binding were used as instrumental variables (IVs) to test for an association between VDR binding affinity at a locus and MS susceptibility.

*Methods*: Data for 13,598 MS cases and 38,887 healthy controls from the Kaiser Permanente Northern California MS Research Program, two Swedish MS case-control studies, and the UK Biobank were included in the analyses. All participants were of European ancestry. SNPs associated with allele-specific VDR binding affinity (VDR-BVs) were previously identified using ChIP-exo data from 16 calcitriol-stimulated Lymphoblastoid Cell-Lines followed by Allele-seq. 112 VDR-BVs were present in the data after QC. The polygenic risk score (PRS) for 25(OH)D level was calculated using summary statistics from two recent GWAS. Within each study, logistic regression was used to estimate the independent association between each VDR-BV IV and MS as well as interaction between each VDR-BV IV and the 25OHD PRS for MS risk. Analyses were adjusted for HLA-DRB1*15:01, sex, age, and genetic ancestry. Meta-analysis with random effects was used to combine associations.

*Results*: We found evidence for interaction between two VDR-BV (rs2881514 and rs2531804) and MS after correcting for ($P_{FDR}<0.05$). There was also evidence of interaction between rs2881514 and a polygenic risk score for serum vitamin D, providing evidence of a causal association between rs2881514 and MS not biased on horizontal pleiotropy.

*Conclusions*: This study is the first to demonstrate that genetic variation in VDR binding affinity at a single locus contributes to MS susceptibility and our results highlight the importance of the Vitamin D pathway in MS pathogenesis.

## Introduction

The geographic distribution of multiple sclerosis (MS), which is more prevalent among people residing in latitudes farther from the equator[1] has led to the hypothesis that high vitamin D levels are associated with MS susceptibility.[2] Several studies have shown that high levels of serum 25-hydroxyvitamin D (25(OH)D)[3], greater sunlight exposure[4,5], and vitamin D supplements and diets rich in vitamin D[6–8] are associated with a decreased risk of MS. However, reverse causation and unmeasured confounding factors remain hard to assess alternative explanations. Recent studies using Mendelian Randomization (MR) have demonstrated strong evidence for a causal relationship between low serum 25(OH)D levels and MS susceptibility in three large, adult-onset MS case-control datasets[9,10] and one pediatric-onset MS study.[11] The contribution of genetic instruments suggests that more of the human genome is involved in explaining variation in serum 25(OH)D levels, and that different aspects of the vitamin D pathway, specifically transcription and expression mediated by vitamin D receptor DNA binding are involved in the relationship between vitamin D and MS.

Vitamin D has immunomodulatory properties and participates in the regulation of calcium metabolism, cellular growth, proliferation and apoptosis, and other important immunologic functions such as modulation of inflammatory processes.[12] It is well established that 25(OH)D signals through the nuclear vitamin D receptor (VDR), a ligand-regulated transcription factor which mediates all genomic actions of 25(OH)D.[13] Upon activation by vitamin D, the VDR forms a heterodimer and acts as a transcription factor that binds to specific vitamin D response elements (VDREs) located within regulatory regions of target genes. Strong experimental evidence has recently demonstrated alternative mechanisms of gene regulation by 25(OH)D stimulated VDRs, and many VDR binding sites identified through ChIP-seq experiments do not contain VDREs.[14,15] Transcriptional regulation by the VDR, similar to other nuclear receptors, has been characterized by its capacity to recognize high affinity binding sites. Bound receptors recruit coregulatory proteins, leading to transactivation of adjacent target genes. Stimulation with vitamin D also shifts the protein interaction profile of DNA-bound VDR from co-repressor proteins to co-activator proteins to further regulate transcription.[16] Gene expression profiling studies have revealed that 25(OH)D signaling through the VDR can lead to activation or repression of target gene transcription.

Previous research has found that genetic variation of tagging Single Nucleotide Polymorphisms (SNPs) within these VDR binding sites (VDR binding variants, VDR-BVs), can alter binding affinity and evidence from 1000 Genomes samples has revealed that these VDR-BVs are enriched within several autoimmune disease GWAS regions, including MS.[17] These findings indicate that genetic variation in VDR binding affinity may be one of the causal mechanisms that underlie MS GWAS findings. The effect of genetic variation in VDR binding on gene regulation and expression is partially dependent on upstream bioavailability of vitamin D. MR studies have shown that 25(OH)D genetic instrumental variables (GIVs), conceptualized as a measure of genetic variation in bioavailability of 25(OH)D, are causally associated with MS. Therefore, bioavailability of 25(OH)D should modulate the effect of genetic variation in VDR binding affinity on MS. No formal investigation of genetic variation of individual VDR-BVs near GWAS SNPs or across the genome in MS cases and controls has been reported, and genomic regions to which VDR bind are strong candidates to investigate for genetic variation relevant to MS.

The aim of this research was to identify VDR-BVs that have a causal effect on Multiple Sclerosis (MS) susceptibility via their effect on VDR DNA binding using two-sample Mendelian Randomization (MR). We used data on MS cases and healthy controls from three case-control studies and the UK Biobank. VDR-BVs were previously identified through ChIP-exo analysis.[17] Our hypothesis was that altered VDR binding disrupts downstream gene regulation by vitamin D and increases the risk of developing MS. Identification of VDR-BVs associated with MS susceptibility will improve understanding of the biological mechanisms through which vitamin D acts to affect MS and further elucidate the causes of MS.

## Methods

### *MS case control studies*

Individual level data used in this research was from one US based case control study and two Swedish based MS case control studies. From the US, MS cases and controls were from the Kaiser Permanente Northern California (KPNC) MS Research Program.[18] Additional KPNC controls were participants of the Genetic Epidemiology Research on Adult Health and Aging (GERA; dbGaP phs000674.v2. p2).[19] From Sweden, MS cases and controls were from the Epidemiological Investigation of Multiple Sclerosis (EIMS) and the Genes and Environment in Multiple Sclerosis (GEMS) study.[20,21] Methods for identification and confirmation of MS diagnosis in each study are provided within the corresponding publications. Briefly, MS cases from KPNC, EIMS, and GEMS had their disease status confirmed by an independent neurologist. All study participants provided written informed consent, and all studies obtained approval from Institutional Review Boards of KPNC, local Ethical Committees, and the University of California, Berkeley.

### *UK Biobank*

The UK Biobank (UKB, http://www.ukbiobank.ac.uk) is a prospective cohort study of approximately 500,000 individuals from the United Kingdom.[22] Recruitment took place between 2006-2010 in 22 assessment centers located across the United Kingdom. Participants were aged 40-69 at time of recruitment. MS cases were identified using the self-reported first date of MS symptoms. Non-MS cases were defined as those not reporting MS symptoms or other demyelinating disease at time of enrollment. Controls were frequency matched to cases by year of birth ($\pm$2 years) and sex at a ratio of 10 controls per MS case. Methods for identification of MS cases and controls are provided in the Supplement. Data were accessed under approval of UKB within project 69668. All participants gave prior written informed consent, and the study was conducted following the principles of the declaration of Helsinki.

### *Genotype and exposure assessment*

All participants in the MS case-control studies completed an interview or self-reported questionnaire related to MS disease events, reproductive history, and environmental exposures.[18–21] Participants provided blood or saliva samples for genotyping. Genome-wide single nucleotide polymorphism (SNP) genotyping was performed using the Illumina Infinium 660K BeadChip Array and Human Omni Express Array (KPNC), Illumina Global Screening Array (GEMS), Human Omni Expression (EIMS), and Affymetrix Axiom Array (GERA).

Details on KPNC, GEMS, and EIMS genome-wide SNP imputation are provided in supplement. UKB genotyping and imputation is described previously.[23] Participants with missing genotypes that met QC thresholds (info score>0.8, missingness per SNP<0.05, missingness per cohort<0.05, and minor allele frequency (MAF)>0.05) were imputed using the mean MAF within each study.

*Identifying valid genetic instruments for VDR binding*

The occupancy of VDR at a given genomic locus is the exposure of interest in this study. Single nucleotide polymorphisms (SNPs) associated with VDR binding affinity at a locus were previously identified by Gallone et al.[24] Briefly, occupancy of VDR using ChIP-exo data was quantified from sixteen calcitriol-stimulated Lymphoblastoid Cell-Lines (LCLs). To satisfy the first assumption of MR, at least one variant associated with variation in VDR occupancy at a locus was needed. SNPs associated with VDR binding in cis were identified by observing differential occupancy over each allele of a given heterozygous SNP. Allele-specific VDR binding was identified using AlleleSeq.[25] After applying correction for multiple testing using, 305 SNPs were identified that were associated with allele-specific VDR binding affinity (VDR-BVs). Of these, 112 VDR-BVs were present in each dataset for analysis. SNP-VDR binding effect sizes were estimated by regressing the read counts from ChIP-exo data at a VDR binding region described above against the corresponding VDR-BV alleles. For each VDR-BV the GIV was derived by multiplying the SNP-VDR binding effect size by the number alleles carried by each participant (hereafter, $GIV_{VDR}$).

*Pleiotropy and Bioavailability of Vitamin D*

The causal effect estimates from MR analysis are unbiased in the absence of pleiotropic effects. A common approach to estimating the pleiotropic bias is based on over-dispersion across multiple instrumental SNPs for a given exposure. Given that there was only one instrumental SNP for each $GIV_{VDR}$, sensitivity analyses for multi-SNP MR instruments were not applicable. However, given that VDR binding is dependent upon bioavailability of vitamin D, we used polygenic risk scores (PRSs) for serum 25(OH)D as instrumental variables for bioavailability of vitamin D to perform subgroup analyses. As the occupancy of VDR at a locus depends upon the bioavailability of 25(OH)D levels, the causal effect of VDR occupancy should be modulated by 25(OH)D Serum levels. Interaction between a $GIV_{VDR}$ and bioavailability of 25(OH)D was evidence of an association between VDR binding and MS susceptibility not biased by horizontal pleiotropy.

*25(OH)D polygenic risk scores*

We used PRSs as GIVs for the bioavailability of 25(OH)D (hereafter $GIV_{25OHD}$). SNPs and the estimated effect sizes from two recent GWAS on serum 25(OH)D were used to construct each $GIV_{25OHD}$.[26,27] GWAS summary statistics were extracted from MR-Base R platform.[28] Clumping and thresholding methods were used to identify SNPs for the PRS. Linkage disequilibrium clumping and p-value thresholding was performed using extract_instruments() function in the *TwoSampleMR* R package. Independent SNPs with GWAS $p < 1 \times 10^{-8}$ were identified within genomic windows of 10,000Kb ($R^2$<0.001) using a European reference panel from 1000 Genomes Phase 3. After identifying these SNPs, the PRS was calculated in plink.

*Statistical methods*

Logistic regression was used to estimate the associations all GIVs ($GIV_{25OHD}$ and $GIV_{VDR}$) and MS susceptibility within each study:

$$logit\big(P(MS=1)\big) \sim \beta_0 + \beta_1 GIV_{25OHD}$$

$$logit\big(P(MS=1)\big) \sim \beta_0 + \beta_1 GIV_{VDR_j}$$

Additionally, to estimate an association not biased by horizontal pleiotropy, we estimated interaction between each $GIV_{VDR}$ and $GIV_{25OHD}$ for MS susceptibility:

$$logit\big(P(MS=1)\big) \sim \beta_0 + \beta_1 GIV_{VDR_j} + \beta_2 GIV_{25OHD} + \beta_3 \left(GIV_{VDR_j} \times GIV_{25OHD}\right),$$

where $\beta_3$ is the estimate of multiplicative interaction between VDR binding at a locus and the bioavailability of 25(OH)D. A non-zero value of $\beta_3$ indicates that pleiotropy does not explain all the observed association and therefore, the association between VDR-BV and MS is causal.

All models were adjusted for sex (male or female), quintiles of year of birth, carriage of the *HLA-DRB1*15:01* allele, and six genome-wide principal components. Random effect meta-analysis was used to combine study-specific associations. All meta-analyses were performed using the metagen() function in the Meta R package.[29] The DerSimonian-Laird estimator was used to estimate the between study variance. Between study heterogeneity was assessed using Cochran's $Q$ statistic and Higgins & Thompson's $I^2$ statistic. P-values from associations between each $GIV_{25OHD}$ and MS and the interaction parameter between each $GIV_{VDR}$ and $GIV_{25OHD}$ were corrected for multiple tests using the Benjamini-Hochberg method.[30]

Annotation of VDR-BVs to the nearest transcription start sites (TSS) and GO enrichment analysis was performed using rGREAT.[31] Expression quantitative trait loci (eQTL) for tissue specific associations between VDR-BVs and gene expression are from the GTEx Project (https://gtexportal.org/). Linkage disequilibrium (LD) between VDR-BVs and MS GWAS risk variants was estimated using the ld_matrix() function from the ieugwasr R package using the European LD reference panel from 1000 Genomes Phase 3.[32]

**Results**

*Characteristics of MS cases and controls*

Characteristics of MS cases and controls included in this study are presented in Table 1. The largest source of MS cases was the EIMS study, with 6,709 MS cases and 5,881 controls, followed by the GEMS study (3,718 MS cases, 1,180 controls), the UKB study (2,087 MS cases, 20,870 controls), and KPNC (1,082 cases, 10,956 controls). All studies were primary comprised of female participants with KPNC having the highest proportion of females (80.2%) and UKB having the lowest (72.4%). The average year of birth was earlier within KPNC and UKB studies (mid 1950s) compared to the GEMS and EIMS studies (1960-1970s). The proportion of MS cases with at least one HLA-DRB1*15:01 allele was similar across all studies (~49-57%) with UKB cases having the lowest carriage rate.

*$GIV_{25(OH)D}$ associated with MS susceptibility*

The mean $GIV_{25(OH)D}$ for increased 25(OH)D serum levels was lower among cases than controls within each study (Table 2). The difference in the Jiang et al. and Revez et al. $GIV_{25(OH)D}$ $GIVs_{25(OH)D}$ between cases and controls was largest among GEMS participants. Differences in the mean was smallest across both $GIVs_{25(OH)D}$ among UKB cases and controls. Results from logistic regression and meta-analyses indicate that decreased 25(OH)D serum level is associated with an increased risk of MS susceptibility (Figure 1). The association was higher for the Jiang et al. $GIV_{25(OH)D}$ (OR: 1.85, 95% CI: 1.30-2.63) compared to the Revez et al. $GIV_{25(OH)D}$ (OR: 1.32, 95% CI: 1.07-1.61). Estimates for between study variance were low for the Jiang et al. $GIV_{25(OH)D}$ ($I^2$=0%, $P_Q$=0.56) indicating very consistent associations across all studies. There was some evidence for moderate heterogeneity for the Revez et al. $GIV_{25(OH)D}$ ($I^2$=55%, $P_Q$=0.56), however, the direction of effect was consistent across all studies. Associations within the UKB participants were the closet to the null compared to the other studies

*VDR-BVs associated with MS susceptivity in meta-analyses*

Seven $GIV_{VDR}$ were associated with MS susceptibility at p<0.05 (Table 3). Following multiple testing correction, two $GIV_{VDR}$ were significant ($P_{FDR}$<0.05). Meta-analyses an association between reduced VDR binding affinity at rs2881514_A and increase in risk of MS susceptibility across all four studies (Meta OR: 1.10, 95% CI: 1.05-1.15, $P=9.4 \times 10^{-5}$; Figure 2). rs2531804_A was associated with a decreased risk of MS susceptibility (Meta OR: 0.82, 95% CI: 0.73-0.92, $P=6.4 \times 10^{-4}$). There was evidence of interaction between 13 $GIVs_{VDR}$ and $GIVs_{25OHD}$ at $P$<0.05 (Table 4). No interactions were significant after multiple testing correction. There was evidence of interaction between the top $GIV_{VDR}$ from the independent models, rs2881514_A, and the Jiang et al. $GIV_{25OHD}$ (Meta OR: 2.17, 95% CI: 1.10-4.29; Table 4, Figure 4). The association was consistent across three of the four studies, with the KPNC estimate being protective and the others being harmful. The direction of this interaction association was evidence of synergistic interaction between decreased VDR binding affinity at rs2881514 and decreased bioavailability of 25(OH)D for increased risk of MS.

*VDR-BV with evidence of association with MS are eQTLS*

Several of the $GIVs_{VDR}$ that were independently associated at p<0.05 with MS or had evidence of interaction with $GIV_{25(OH)D}$ are eQTLs for one or more human tissues in GTEx (Figure S-1). rs2881514_A is an eQTL for *RFTN1* in esophageal tissue. The A allele, which is associated with decreased VDR binding affinity, is associated with decreased expression of *RFTN1* in esophageal tissues. Other $GIV_{VDR}$, including rs2531804, rs55811049, rs961320, rs871699, rs62200158, which had evidence of association with MS at p<0.05, are eQTLs in brain tissue and skin tissue, among others.

*LD between $GIV_{VDR}$ and MS GWAS risk variants*

Within the European reference from 1000 Genomes project, LD between VDR-BVs and MS GWAS variants was generally low. Of the 112 VDR-BVs included in this research, only 3 VDR-BV had an $R^2$ >0.5 with an MS GWAS variant (Table S-1). Two VDR-BVs with independent associations with MS (Table 3) were in moderate LD with MS GWAS variants (VDR-BV:

rs55792977, MS GWAS: rs13385171, $R^2$=0.48; VDR-BV: rs13098781, MS GWAS: rs9863496, $R^2$=0.34).

## Discussion

In this study we observed evidence for an association between variation in VDR binding at a locus and MS susceptibility. Two VDR-BVs, rs2881514 and rs2531804, were independently associated with MS susceptibility after correcting for multiple tests in meta-analyses. Models evaluating interaction between VDR-BV and MS by 25(OH)D serum levels also found evidence of interaction between several VDR-BVs and 25(OH)D. Although none were significant after multiple testing correction, there was evidence for interaction for between rs2881514 and the Jiang PRS at p<0.05. Both PRSs for 25(OH)D serum levels were significantly associated with MS susceptibility, with lower 25(OH)D serum being associated with increased risk of MS in all four studies.

The first study to report a causal association between vitamin D serum levels and MS using MR methods was by Mokry et al.[10] Another study using data from KPNC, EIMS, and GEMS participants published similar findings soon after using three GWAS SNPs identified in 2010.[9] Our findings using PRSs constructed with summary statistics from two recent GWASs on 25(OH)D serum levels further establish vitamin D insufficiency as a risk factor for MS. The direction of association between each $GIV_{25(OH)D}$ and MS was consistent across all four studies. Measures of heterogeneity between studies indicate little variation between studies for Jiang $GIV_{25(OH)D}$, although there is some evidence of moderate heterogeneity for the Revez $GIV_{25(OH)D}$ ($I^2$=55%, $p_Q$=0.09). Associations were the closest to the NULL for both $GIV_{25(OH)D}$ in UKB participants. Unlike the other studies, MS cases from UKB did not have their MS status confirmed by a neurologist. MS cases status was inferred from electronic health records in approximately half MS cases in UKB Biobank or from self-report. A likely explanation for the smaller associations in UKB participants is due to misclassification of MS case status resulting in individuals without MS being included as MS cases. Evidence of this misclassification can also be seen from the frequency of carriage of *HLA-DRB1*15:01* alleles, which was lower among UKB MS cases (49.7%) compared to cases in the other three studies (>53%).

The top VDR-BV from meta-analyses was rs2881514. The A allele of this SNP was associated with decreased VDR binding affinity in LCLs. Evidence from results within each study and meta-analyses indicate a harmful effect of decreased binding affinity at that locus on MS risk. Further, interaction between rs2881514 and $GIV_{25(OH)D}$ indicate presence of synergist interaction between decreased VDR binding affinity at rs2881514 and decreased 25(OH)D serum levels for a substantial increase in the risk of MS susceptibility (Meta OR: 2.17, 95% CI: 1.10-4.29). This association was consistent among UKB, GEMS, and EIMS MS cases and control, but was slightly protective in KPNC, though with very wide confidence intervals (KPNC OR: 0.79, 95% CI: 0.14-4.47). This SNP is located on chromosome 3, 1533 bases upstream of the transcription start site of *RFTN1. RFTN1* encodes Raftlin, a protein that enables double-stranded RNA binding. Raftlin is critical for the production of lipid rafts, which are membrane microdomains that play a critical role in B cell activation through B cell receptor signaling.[33] Raftlin has also been found to be essential in Toll-Like Receptor 3 and 4 signaling pathways, which are implicated is MS pathogenesis and symptom modulation through their role in the myeloid differentiation 88-TLR pathway.[34–36] *RFTN1* has not been directly implicated in MS risk;

however; one study found increased expression of *RFTN1* in chronic active brain lesions from MS cases compared to healthy brain tissue from controls.[37]

Several of the VDR-BV with evidence for association with MS had evidence of SNP associated tissue specific expression in GTEx (Figure S-1). rs25318104, rs9621320, rs871699, rs689384, rs62200158, rs558110449, and rs2286576, show evidence of being an eQTL in one or more brain tissues and many of these appear to be eQTLs across wide range of tissue types. Of the set of VDR-BVs tested for association with MS, those with evidence of association with MS susceptibility showed evidence for enrichment in genes involved in immune related processes (leukocyte mediate immunity, immune effector process, myeloid leukocyte activation, and phospholipase activity, among others; Table S-1). Annotations to eQTL and enrichment results provide evidence supporting a mediating role of allele specific VDR binding in the expression of nearby genes and immune pathways in multiple sclerosis risk.

This study had several strengths. This is the first study to investigate the association of VDR binding with MS among cases and controls. An investigation of variation in VDR binding among cases and controls would not be possible in large sample sizes without MR methodology. We used previously identified VDR-BVs to create GIVs for VDR binding a locus. Another strength was the large sample size and use of data from separate studies. There were some limitations in our work including the possibility of horizontal pleiotropy biasing associations between $GIV_{VDR}$ and MS. Each $GIV_{VDR}$ was constructed using one VDR-BV preventing the use of standard sensitivity analyses for assessing the validity of MR assumptions in our analyses. However, our use of a PRS for 25(OH)D serum level as a second GIV provided a method to estimate an association between $GIV_{VDR}$ and MS unbiased by horizontal pleiotropy. Another limitation that MS cases and controls in the study are of European ancestry, limiting the generalizability of our findings to non-European populations. Further, the VDR-BVs were identified in LCLs which are EBV-transformed B-cell lines. Variations in VDR binding associated with MS likely occurs in other lymphocytes, including CD4+ and CD8+ T cells. And lastly, we were only able to assess variations in VDR binding at 112 loci, which represent only a small portion of the total regions where VDR binding occurs across the genome.

This study is the first to demonstrate that genetic variation in VDR binding affinity at a single locus contributes to MS susceptibility and our results highlight the importance of the Vitamin D pathway in MS pathogenesis. Future studies of VDR binding and MS should identify VDR-BVs in lymphocytes not captured by LCLs, including CD4+ and CD8+ T cells.

## References


1. Rosati G. The prevalence of multiple sclerosis in the world: an update. *Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*. 2001 Apr;**22**(2):117–39.

2. Goldberg P. Multiple sclerosis: vitamin D and calcium as environmental determinants of prevalence. *International Journal of Environmental Studies*. Taylor & Francis Group ; 1974 Jan 25;**6**(1):19–27.

3. Munger KL, Levin LI, Hollis BW, Howard NS, Ascherio A. Serum 25-Hydroxyvitamin D Levels and Risk of Multiple Sclerosis. *JAMA*. 2006 Dec 20;**296**(23):2832.

4. Mei IAF van der, Ponsonby A-L, Dwyer T, et al. Past exposure to sun, skin phenotype, and risk of multiple sclerosis: case-control study. *BMJ*. 2003 Aug 9;**327**(7410):316–0.

5. Kampman MT, Wilsgaard T, Mellgren SI. Outdoor activities and diet in childhood and adolescence relate to MS risk above the Arctic Circle. *Journal of Neurology*. 2007 Apr 21;**254**(4):471–477.

6. Munger KL, Zhang SM, O'Reilly E, et al. Vitamin D intake and incidence of multiple sclerosis. *Neurology*. 2004 Jan 13;**62**(1):60–5.

7. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. *Annals of Internal Medicine*. American College of Physicians; 2017 Aug 15;**167**(4):268.

8. Ascherio A, Munger KL, Simon KC. Vitamin D and multiple sclerosis. *The Lancet Neurology*. 2010 Jun;**9**(6):599–612.

9. Rhead B, Bäärnhielm M, Gianfrancesco M, et al. Mendelian randomization shows a causal effect of low vitamin D on multiple sclerosis risk. *Neurology Genetics*. American Academy of Neurology; 2016 Oct;**2**(5):e97.

10. Mokry LE, Ross S, Ahmad OS, et al. Vitamin D and Risk of Multiple Sclerosis: A Mendelian Randomization Study. Muraro PA, editor. *PLOS Medicine*. Public Library of Science; 2015 Aug 25;**12**(8):e1001866.

11. Gianfrancesco MA, Stridh P, Rhead B, et al. Evidence for a causal relationship between low vitamin D, high BMI, and pediatric-onset MS. *Neurology*. American Academy of Neurology; 2017 Apr 25;**88**(17):1623–1629.

12. Wöbke TK, Sorg BL, Steinhilber D. Vitamin D in inflammatory diseases. *Frontiers in Physiology*. 2014 Jul 2;**5**:244.

13. Bikle DD. Vitamin D metabolism, mechanism of action, and clinical applications. *Chemistry & biology*. NIH Public Access; 2014 Mar 20;**21**(3):319–29.

14. Ramagopalan S V., Heger A, Berlanga AJ, et al. A ChIP-seq defined genome-wide map of vitamin D receptor binding: Associations with disease and evolution. *Genome Research*. 2010 Oct 1;**20**(10):1352–1360.

15. Tuoresmäki P, Väisänen S, Neme A, Heikkinen S, Carlberg C. Patterns of Genome-Wide VDR Locations. Slominski AT, editor. *PLoS ONE*. Public Library of Science; 2014 Apr 30;**9**(4):e96105.

16. Heikkinen S, Väisänen S, Pehkonen P, Seuter S, Benes V, Carlberg C. Nuclear hormone 1α,25-dihydroxyvitamin D3 elicits a genome-wide shift in the locations of VDR chromatin occupancy. *Nucleic Acids Research*. 2011 Nov;**39**(21):9181–9193.

17. Gallone G, Haerty W, Disanto G, Ramagopalan S V., Ponting CP, Berlanga-Taylor AJ. Identification of genetic variants affecting vitamin D receptor binding and associations with autoimmune disease. *Human Molecular Genetics*. Narnia; 2017 Jun 1;**26**(11):2164–2176.

18. Briggs FBS, Acuña BS, Shen L, et al. Adverse socioeconomic position during the life course is associated with multiple sclerosis. *J Epidemiol Community Health*. 2014 Jul;**68**(7):622–629.

19. Banda Y, Kvale MN, Hoffmann TJ, et al. Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the genetic epidemiology research on adult health and aging (GERA) cohort. *Genetics*. Genetics Society of America; 2015 Aug;**200**(4):1285–1295.

20. Hedström AK, Hillert J, Olsson T, Alfredsson L. Alcohol as a Modifiable Lifestyle Factor Affecting Multiple Sclerosis Risk. *JAMA Neurology*. 2014 Mar;**71**(3):300–300.

21. Hedström A, Åkerstedt T, Olsson T, Alfredsson L. Shift work influences multiple sclerosis risk. *Multiple Sclerosis Journal*. 2015 Aug;**21**(9):1195–1199.

22. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine*. Public Library of Science; 2015 Mar 31;**12**(3):e1001779.

23. Bycroft C, Freeman C, Petkova D, et al. Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv*. Cold Spring Harbor Laboratory; 2017;166298–166298.

24. Gallone G, Haerty W, Disanto G, Ramagopalan SV, Ponting CP, Berlanga-Taylor AJ. Identification of genetic variants affecting vitamin D receptor binding and associations with autoimmune disease. *Human Molecular Genetics*. Narnia; 2017 Jun;**26**(11):2164–2176.

25. Rozowsky J, Abyzov A, Wang J, et al. AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol Syst Biol*. 2011 Aug 2;**7**:522.

26. Jiang X, O'Reilly PF, Aschard H, et al. Genome-wide association study in 79,366 European-ancestry individuals informs the genetic architecture of 25-hydroxyvitamin D levels. *Nature Communications*. Nature Publishing Group; 2018 Dec 17;**9**(1):260.

27. Revez JA, Lin T, Qiao Z, et al. Genome-wide association study identifies 143 loci associated with 25 hydroxyvitamin D concentration. *Nature Communications*. Nature Research; 2020 Dec;**11**(1):1–12.

28. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018 May 30;**7**:e34408.

29. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. *Evid Based Ment Health*. 2019 Nov;**22**(4):153–160.

30. Yekutieli D, Benjamini Y. The control of the false discovery rate in multiple testing under dependency. *Ann Statist*. 2001 Aug;**29**(4):1165–1188.

31. Gu Z. rGREAT: Client for GREAT Analysis. 2021.

32. Hemani G. ieugwasr: R interface to the IEU GWAS database API [Internet]. 2021. Available from: https://github.com/mrcieu/ieugwasr

33. Saeki K. The B cell-specific major raft protein, Raftlin, is necessary for the integrity of lipid raft and BCR signal transduction. *The EMBO Journal*. 2003 Jun 16;**22**(12):3015–3026.

34. Watanabe A, Tatematsu M, Saeki K, et al. Raftlin Is Involved in the Nucleocapture Complex to Induce Poly(I:C)-mediated TLR3 Activation. *J Biol Chem*. 2011 Mar 25;**286**(12):10702–10711.

35. Tatematsu M, Yoshida R, Morioka Y, et al. Raftlin Controls Lipopolysaccharide-Induced TLR4 Internalization and TICAM-1 Signaling in a Cell Type–Specific Manner. *The Journal of Immunology*. American Association of Immunologists; 2016 May 1;**196**(9):3865–3876.

36. Zheng C, Chen J, Chu F, Zhu J, Jin T. Inflammatory Role of TLR-MyD88 Signaling in Multiple Sclerosis. *Frontiers in Molecular Neuroscience* [Internet]. 2020 [cited 2022 May 27];**12**. Available from: https://www.frontiersin.org/article/10.3389/fnmol.2019.00314

37. Hendrickx DAE, Scheppingen J van, Poel M van der, et al. Gene Expression Profiling of Multiple Sclerosis Pathology Identifies Early Patterns of Demyelination Surrounding Chronic Active Lesions. *Frontiers in Immunology* [Internet]. 2017 [cited 2022 May 29];**8**. Available from: https://www.frontiersin.org/article/10.3389/fimmu.2017.01810

## Figures

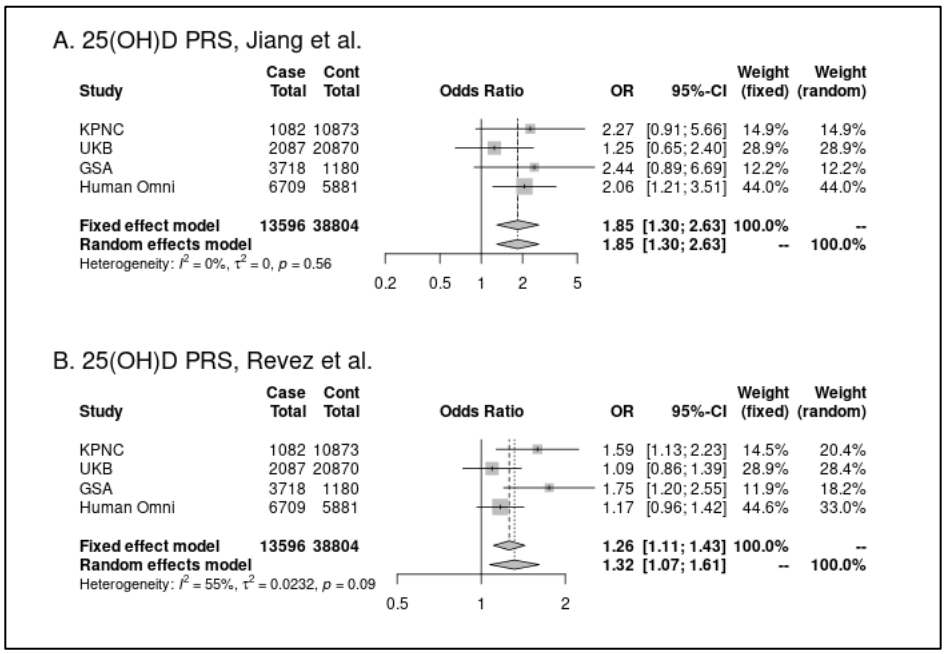



**Figure 1.** Results from meta-analysis of association between 25(OH)D polygenic risk scores (PRS) and multiple sclerosis susceptibility. A) 25(OH)D calculated using summary statistics from Jiang et al. 2019; B) 25(OH)D calculated using summary statistics from Revez et al. 2020.

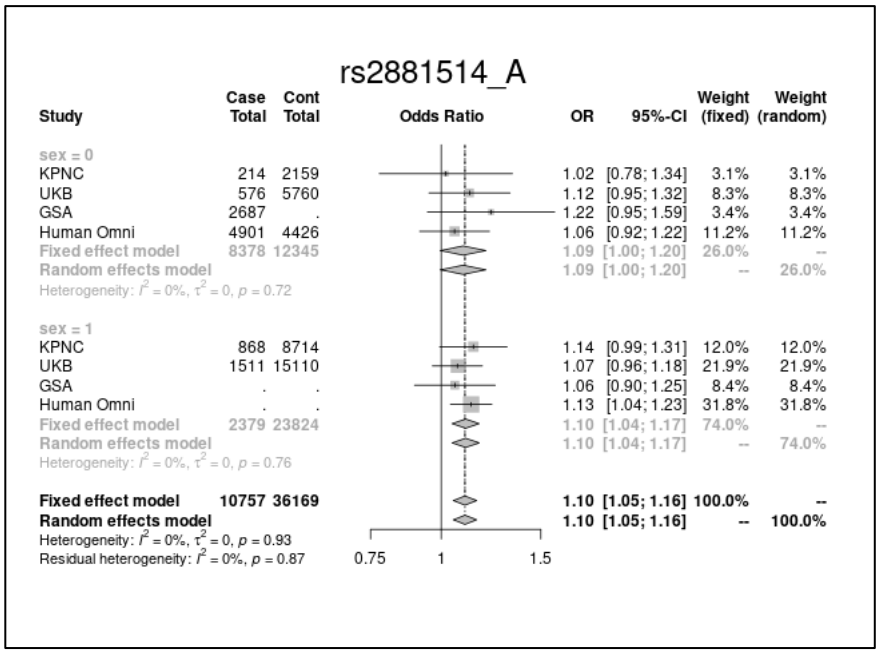



**Figure 2.** Meta-analysis results for association between VDR-BV rs2881514_A and multiple sclerosis stratified by sex.

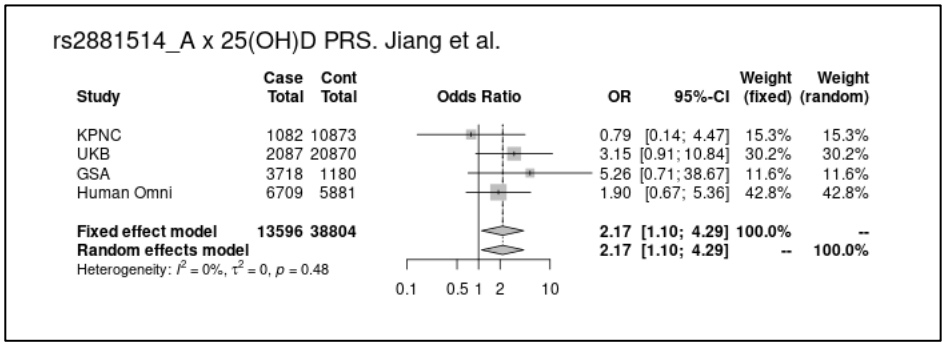



**Figure 3.** Evidence of interaction between rs2881514_A and Jiang et a. 25(OH)D polygenic risk score.

## Tables

**Table 1.** Characteristics of MS cases and healthy controls.

| | KPNC | | UKB | | GEMS | | EIMS | |
|---|---|---|---|---|---|---|---|---|
| | **Case** | **Control** | **Case** | **Control** | **Case** | **Control** | **Case** | **Control** |
| n | 1084 | 10956 | 2087 | 20870 | 3718 | 1180 | 6709 | 5881 |
| Sex, n (%) | | | | | | | | |
| Male | 215 (19.8) | 2165 (19.8) | 576 (27.6) | 5760 (27.6) | 1031 (27.7) | 342 (29.0) | 1808 (26.9) | 1455 (24.7) |
| Female | 869 (80.2) | 8791 (80.2) | 1511 (72.4) | 15110 (72.4) | 2687 (72.3) | 838 (71.0) | 4901 (73.1) | 4426 (75.3) |
| Birth year, mean (SD) | 1957.6 (8.9) | 1957.1 (8.6) | 1952.7 (7.6) | 1952.7 (7.6) | 1973.1 (12.4) | 1968.9 (13.7) | 1961.0 (13.7) | 1960.5 (13.3) |
| *DRB1*15:01* carrier, n (%) | | | | | | | | |
| 0 alleles | 506 (46.7) | 8280 (75.6) | 1050 (50.3) | 15271 (73.2) | 1621 (43.6) | 843 (71.4) | 2913 (43.4) | 4052 (68.9) |
| 1 or 2 alleles | 578 (53.3) | 2676 (24.4) | 1037 (49.7) | 5599 (26.8) | 2097 (56.4) | 336 (28.5) | 3796 (56.6) | 1829 (31.1) |

Abbreviations: EIMS, Epidemiological Investigation of Multiple Sclerosis; GEMS, Genes Environment and MS study KPNC, Kaiser Permanente Northern California MS case control study; UKB, UK Biobank

**Table 2.** Summary of vitamin D polygenic risk scores used within each study.

| | KPNC | | UKB | | GEMS | | EIMS | |
|---|---|---|---|---|---|---|---|---|
| | **Case** | **Control** | **Case** | **Control** | **Case** | **Control** | **Case** | **Control** |
| n | 1084 | 10956 | 2087 | 20870 | 3718 | 1180 | 6709 | 5881 |
| 25(OH)D PRS, Jiang et al.[a] (mean (SD)) | 0.285 (0.072) | 0.288 (0.069) | 0.292 (0.069) | 0.293 (0.069) | 0.289 (0.070) | 0.293 (0.066) | 0.286 (0.069) | 0.290 (0.069) |
| 25(OH)D PRS, Revez et al.[a] (mean (SD)) | 2.150 (0.194) | 2.163 (0.186) | 2.184 (0.189) | 2.187 (0.188) | 2.057 (0.186) | 2.074 (0.182) | 2.066 (0.187) | 2.070 (0.189) |

Abbreviations: EIMS, Epidemiological Investigation of Multiple Sclerosis GEMS, Genes Environment and MS study; KPNC, Kaiser Permanente Northern California MS case control study; PRS, polygenic risk score; UKB, UK Biobank

[a] PRSs calculated from independent GWAS variants (Linkage disequilibrium $R^2<0.001$) with $p<5x10^{-8}$.

**Table 3.** Vitamin D Receptor binding variants (VDR-BVs) instrumental variable (IV) associated with and multiple sclerosis susceptibility at p<0.05.

| VDR-BV[a] | chr:bp | GREAT annotation[b] | OR[c] | 95% CI | *P* | $P_{FDR}$[c] |
|---|---|---|---|---|---|---|
| rs2881514_A | chr3:16553679 | RFTN1 (+1533); OXNAD1 (+246974) | 1.10 | (1.05-1.15) | 9.4E-05 | 0.011 |
| rs2531804_A | chr6:28411302 | ZSCAN23 (-59) | 0.82 | (0.73-0.92) | 6.4E-04 | 0.036 |
| rs12048389_T | chr1:107538722 | PRMT6 (-60578) | 0.92 | (0.87-0.99) | 1.8E-02 | 0.660 |
| rs55792977_T | chr2:65650863 | SPRED2 (+8447); ACTR2 (+195893) | 1.05 | (1.00-1.09) | 3.2E-02 | 0.660 |
| rs7309003_C | chr12:97751789 | NEDD1 (+450546) | 0.95 | (0.91-1.00) | 3.2E-02 | 0.660 |
| rs10995246_C | chr10:64391845 | ADO (-172670); ZNF365 (+257895) | 1.06 | (1.00-1.13) | 3.5E-02 | 0.660 |
| rs10232857_C | chr7:55601335 | VOPP1 (+38882); LANCL2 (+168195) | 0.93 | (0.87-1.00) | 4.9E-02 | 0.780 |

Abbreviations: bp, base pair; chr; chromosome; CI, Confidence Interval; FDR, false discovery rate; GREAT, Genomic Regions Enrichment of Annotations Tool; OR, Odds Ratio; VDR-BV, Vitamin D Receptor Binding Variant

[a] VDR-BV using as instrumental variable for allelic specific binding. Allele including is the allele associated with decreased binding affinity.

[b] PRS calculated using summary statistics from Jiang et al 2019 and Revez et al 2020 GWAS.

[c] Distance from VDR-BV to TSS of nearest upstream and downstream gene from rGREAT.

[d] Odds ratios are from random effect meta-analyses combining estimates from the four studies. ORs are the multiplicative interaction coefficient for interaction between 25(OH)D PRS and VDR-BV.

**Table 4.** Vitamin D Receptor binding variants (VDR-BVs) with evidence of interaction with 25(OH)D vitamin D polygenic risk score.

| VDR-BV[a] | 25(OH)D PRS[b] | chr:bp | GREAT annotation[b] | OR[c] | 95% CI | *P* | $P_{FDR}$[c] |
|---|---|---|---|---|---|---|---|
| rs11733032_G | Jiang et al. | chr4:79560226 | BMP2K (-137305); ANXA3 (+87554) | 2.91 | (1.50-5.66) | 0.002 | 0.24 |
| rs2286576_T | Jiang et al. | chr12:4714007 | DYRK4 (+14764); AKAP3 (+44205) | 2.47 | (1.31-4.65) | 0.005 | 0.45 |
| rs961320_C | Jiang et al. | chr12:49297863 | CCDC65 (-68) | 3.20 | (1.35-7.62) | 0.008 | 0.46 |
| rs55811049_A | Revez et al. | chr5:118324297 | DTWD2 (-58) | 1.53 | (1.11-2.12) | 0.009 | 0.46 |
| rs11729497_G | Revez et al. | chr4:126551382 | FAT4 (+313829) | 0.72 | (0.56-0.93) | 0.01 | 0.46 |
| rs62200158_C | Jiang et al. | chr20:61557905 | TCFL5 (-64791); DIDO1 (+11368) | 0.22 | (0.07-0.72) | 0.011 | 0.47 |
| rs55811049_A | Jiang et al. | chr5:118324297 | DTWD2 (-58) | 2.76 | (1.15-6.64) | 0.024 | 0.53 |
| rs4440604_G | Jiang et al. | chr8:105290834 | DCSTAMP (-61219); RIMS2 (+777720) | 0.50 | (0.28-0.92) | 0.024 | 0.53 |
| rs2881514_A | Jiang et al. | chr3:16553679 | RFTN1 (+1533); OXNAD1 (+246974) | 2.17 | (1.10-4.29) | 0.025 | 0.53 |
| rs12144635_T | Jiang et al. | chr1:173246261 | PRDX6 (-200143); TNFSF4 (-69810) | 0.26 | (0.08-0.85) | 0.026 | 0.53 |
| rs11733032_G | Revez et al. | chr4:79560226 | BMP2K (-137305); ANXA3 (+87554) | 1.32 | (1.03-1.69) | 0.026 | 0.53 |
| rs3825776_T | Jiang et al. | chr15:58746829 | LIPC (+44062); ADAM10 (+295347) | 0.38 | (0.16-0.91) | 0.030 | 0.56 |
| rs871699_T | Jiang et al. | chr11:18127678 | SAAL1 (-41) | 0.31 | (0.10-0.94) | 0.039 | 0.60 |
| rs689384_C | Jiang et al. | chr18:54318453 | WDR7 (-162) | 2.62 | (1.02-6.75) | 0.047 | 0.63 |

Abbreviations: bp, base pair; chr; chromosome; CI, Confidence Interval; FDR, false discovery rate; GREAT, Genomic Regions Enrichment of Annotations Tool; OR, Odds Ratio; VDR-BV, Vitamin D Receptor Binding Variant

[a] VDR-BV using as instrumental variable for allelic specific binding. Allele including is the allele associated with decreased binding affinity.

[b] PRS calculated using summary statistics from Jiang et al 2019 and Revez et al 2020 GWAS.

[c] Distance from VDR-BV to TSS of nearest upstream and downstream gene from rGREAT.

[d] Odds ratios are from random effect meta-analyses combining estimates from the four studies. ORs are the multiplicative interaction coefficient for interaction between 25(OH)D PRS and VDR-BV.

## Supplement

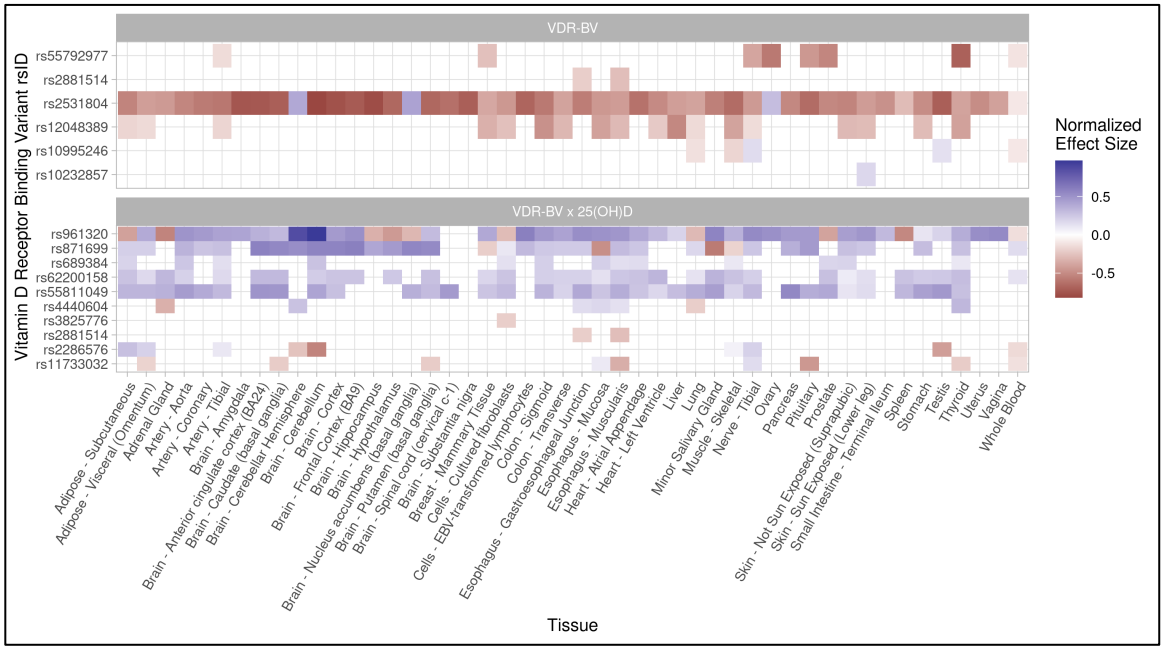
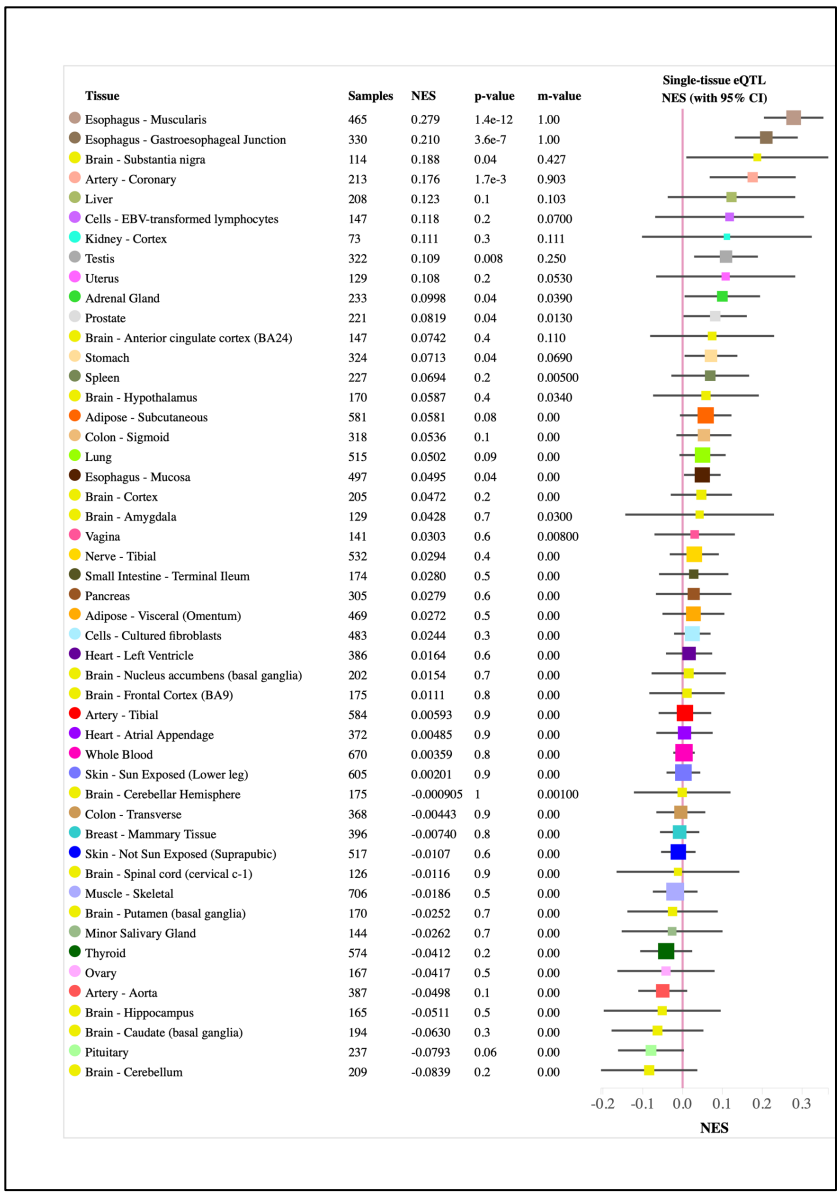



**Figure S-1.** Evidence of SNP associated tissue expression among VDR-BVs presented in Tables 3 and 4. Expression data from GTEx (https://gtexportal.org/). Blue indicates evidence of association between decreased VDR binding affinity and increased expression, red indicates evidence of association between decreased VDR binding affinity and decreased expression.




| Tissue | Samples | NES | p-value | m-value |
|---|---|---|---|---|
| Esophagus - Muscularis | 465 | 0.279 | 1.4e-12 | 1.00 |
| Esophagus - Gastroesophageal Junction | 330 | 0.210 | 3.6e-7 | 1.00 |
| Brain - Substantia nigra | 114 | 0.188 | 0.04 | 0.427 |
| Artery - Coronary | 213 | 0.176 | 1.7e-3 | 0.903 |
| Liver | 208 | 0.123 | 0.1 | 0.103 |
| Cells - EBV-transformed lymphocytes | 147 | 0.118 | 0.2 | 0.0700 |
| Kidney - Cortex | 73 | 0.111 | 0.3 | 0.111 |
| Testis | 322 | 0.109 | 0.008 | 0.250 |
| Uterus | 129 | 0.108 | 0.2 | 0.0530 |
| Adrenal Gland | 233 | 0.0998 | 0.04 | 0.0390 |
| Prostate | 221 | 0.0819 | 0.04 | 0.0130 |
| Brain - Anterior cingulate cortex (BA24) | 147 | 0.0742 | 0.4 | 0.110 |
| Stomach | 324 | 0.0713 | 0.04 | 0.0690 |
| Spleen | 227 | 0.0694 | 0.2 | 0.00500 |
| Brain - Hypothalamus | 170 | 0.0587 | 0.4 | 0.0340 |
| Adipose - Subcutaneous | 581 | 0.0581 | 0.08 | 0.00 |
| Colon - Sigmoid | 318 | 0.0536 | 0.1 | 0.00 |
| Lung | 515 | 0.0502 | 0.09 | 0.00 |
| Esophagus - Mucosa | 497 | 0.0495 | 0.04 | 0.00 |
| Brain - Cortex | 205 | 0.0472 | 0.2 | 0.00 |
| Brain - Amygdala | 129 | 0.0428 | 0.7 | 0.0300 |
| Vagina | 141 | 0.0303 | 0.6 | 0.00800 |
| Nerve - Tibial | 532 | 0.0294 | 0.4 | 0.00 |
| Small Intestine - Terminal Ileum | 174 | 0.0280 | 0.5 | 0.00 |
| Pancreas | 305 | 0.0279 | 0.6 | 0.00 |
| Adipose - Visceral (Omentum) | 469 | 0.0272 | 0.5 | 0.00 |
| Cells - Cultured fibroblasts | 483 | 0.0244 | 0.3 | 0.00 |
| Heart - Left Ventricle | 386 | 0.0164 | 0.6 | 0.00 |
| Brain - Nucleus accumbens (basal ganglia) | 202 | 0.0154 | 0.7 | 0.00 |
| Brain - Frontal Cortex (BA9) | 175 | 0.0111 | 0.8 | 0.00 |
| Artery - Tibial | 584 | 0.00593 | 0.9 | 0.00 |
| Heart - Atrial Appendage | 372 | 0.00485 | 0.9 | 0.00 |
| Whole Blood | 670 | 0.00359 | 0.8 | 0.00 |
| Skin - Sun Exposed (Lower leg) | 605 | 0.00201 | 0.9 | 0.00 |
| Brain - Cerebellar Hemisphere | 175 | -0.000905 | 1 | 0.00100 |
| Colon - Transverse | 368 | -0.00443 | 0.9 | 0.00 |
| Breast - Mammary Tissue | 396 | -0.00740 | 0.8 | 0.00 |
| Skin - Not Sun Exposed (Suprapubic) | 517 | -0.0107 | 0.6 | 0.00 |
| Brain - Spinal cord (cervical c-1) | 126 | -0.0116 | 0.9 | 0.00 |
| Muscle - Skeletal | 706 | -0.0186 | 0.5 | 0.00 |
| Brain - Putamen (basal ganglia) | 170 | -0.0252 | 0.7 | 0.00 |
| Minor Salivary Gland | 144 | -0.0262 | 0.7 | 0.00 |
| Thyroid | 574 | -0.0412 | 0.2 | 0.00 |
| Ovary | 167 | -0.0417 | 0.5 | 0.00 |
| Artery - Aorta | 387 | -0.0498 | 0.1 | 0.00 |
| Brain - Hippocampus | 165 | -0.0511 | 0.5 | 0.00 |
| Brain - Caudate (basal ganglia) | 194 | -0.0630 | 0.3 | 0.00 |
| Pituitary | 237 | -0.0793 | 0.06 | 0.00 |
| Brain - Cerebellum | 209 | -0.0839 | 0.2 | 0.00 |

**Figure S-2.** Multi-tissue eQTL Comparison for rs2881514_A. The effect allele for normalized expression is G.

**Table S-1.** Results from GREAT enrichment analyses of VDR-BVs associated with MS susceptibility.

| Ontology | ID | Description | GREAT[a] Enrichment *P*-value | VDR-BV |
|---|---|---|---|---|
| BP | GO:0002443 | leukocyte mediated immunity | 0.0044 | rs12144635; rs2881514; rs11733032; rs3825776 |
| BP | GO:0002252 | immune effector process | 0.0074 | rs12144635; rs2881514; rs11733032; rs3825776 |
| BP | GO:0002274 | myeloid leukocyte activation | 0.0074 | rs12144635; rs11733032; rs4440604; rs3825776 |
| BP | GO:0006887 | exocytosis | 0.0074 | rs12144635; rs11733032; rs4440604; rs3825776 |
| BP | GO:0032940 | secretion by cell | 0.0074 | rs12144635; rs11733032; rs4440604; rs3825776 |
| BP | GO:0045055 | regulated exocytosis | 0.0074 | rs12144635; rs11733032; rs4440604; rs3825776 |
| BP | GO:0046903 | secretion | 0.0074 | rs12144635; rs11733032; rs4440604; rs3825776 |
| CC | GO:0099503 | secretory vesicle | 0.0048 | rs12144635; rs11733032; rs2286576; rs3825776; rs689384 |
| CC | GO:0031410 | cytoplasmic vesicle | 0.0060 | rs12144635; rs2881514; rs11733032; rs4440604; rs2286576; rs3825776; rs689384 |
| CC | GO:0097708 | intracellular vesicle | 0.0060 | rs12144635; rs2881514; rs11733032; rs4440604; rs2286576; rs3825776; rs689384 |
| CC | GO:0042581 | specific granule | 0.0110 | rs11733032; rs3825776 |
| CC | GO:0030141 | secretory granule | 0.0120 | rs12144635; rs11733032; rs2286576; rs3825776 |
| MF | GO:0004620 | phospholipase activity | 0.0300 | rs12144635; rs3825776 |

Abbreviations: BP, Biological Process; CC, Cellular Component; MF, Molecular Function; GREAT, Genomic Regions Enrichment of Annotations Tool; VDR-BV, Vitamin D Receptor Binding Variant.
[a] Enrichment analyses performed using rGREAT. Background regions were locations of all 112 VDR-BV used in regression analyses, test regions were the 12 VDR-BVs with nominal associations with MS presented in Tables 3 and 4.

## S1. Study descriptions

*KPNC participants*

Study recruitment is described in detail elsewhere.[18] Briefly, study participants were recruited from KPNC membership between 2006 and 2014. Prevalent MS cases were the focus of recruitment. Participants were aged between 18 and 69 years and were KPNC members at initial contact.

*GEMS and EIMS participants*

EIMS and GEMS are Swedish population-based case-control studies.[20] At enrollment, EIMS participants were aged 18–70 years and had recently (within 2 years) confirmed MS. GEMS participants were identified from the Swedish National MS registry and recruited between 2009 and 2011. All EIMS participants were distinct from the GEMS study.

## S2. Genotype and imputation Data description and quality control protocols

*KPNC participants*

Genome-wide SNP imputation was performed using SHAPEIT2 and IMPUTE2. Participants with missing genotypes that met QC thresholds (info score > 0.8, missingness per SNP < 0.05, missingness per cohort < 0.05, and minor allele frequency (MAF) > 0.05) were imputed using the average MAF within each cohort.

*GEMS and EIMS participants*

Phasing was performed using Eagle2. Samples were merged with the Haplotype Reference Consortium. The output VCF files were then used for imputation using Minimac4 and the same reference panel using 500 Kb windows. A small fraction of chunks had to be imputed on Minimac3 (18/2,702 or 0.7%). The chunks were then merged into single chromosome VCF files. After imputation variants with MAF<1% and imputation score <0.3 were excluded.

## Chapter 4 - Identification of cell-specific DNA methylation changes associated with methotrexate treatment response in rheumatoid arthritis


### Abstract

*Background*: MTX is the recommended first treatment for rheumatoid arthritis (RA); however, only ~40% respond adequately to MTX. Significant joint damage can occur in the early phase of RA, and response to the first treatment is an indicator of long-term prognosis. Changes in DNA methylation (DNAm) associated with response to these treatments are potential biomarkers for prediction of treatment response.

*Methods*: We estimated changes in cell-specific DNAm associated with MTX response from whole-blood samples collected from RA patients before and after initiation of MTX treatment. Patients included in this study were from the Rheumatoid Arthritis Medication Study (RAMS; n=66) and University of California San Francisco Rheumatoid Arthritis study (UCSF-RA; n=11). All patients met the American College of Rheumatology RA criteria. Blood samples were collected at baseline and after treatment (UCSF-RA: 3-6 months, RAMS: 4 weeks). DAS28-CRP was collected at baseline and after 3-6 months of treatment. Genome-wide methylation profiles were generated with Illumina *450K* (RAMS) and *EPIC* BeadChips (UCSF-RA) from whole blood. Functional normalization and background subtraction with dye-bias normalization and other QC procedures were performed using *minfi*. Differences between 450k and EPIC platforms were adjusted using *Harman*. MTX response was defined using the EULAR criteria for DAS28-CRP (Responder: good/moderate response, Non-responder: no response). Differentially methylated positions (DMPs) were identified using *limma* and *tensor composition analysis* (*TCA*). *TCA* is a method for identifying cell-specific differential DNAm at the CpG level from bulk tissue. B cells, CD4 and CD8 T cells, monocytes, neutrophils, and Natural Killer (NK) cells were included in cell-specific analyses. Linear models evaluated differential DNAm between MTX response groups over time and within each time point. Sex, age, smoking history, estimated global cell-proportions, and batch were included as covariates. Differentially Methylated Regions (DMRs) were identified using Comb-p.

*Results*: We found evidence for differential global methylation between response groups after treatment. Further, we found patterns of cell-specific differential global methylation associated with MTX response. One DMP was associated at genome wide significance with differential DNAm between responders and non-responders at baseline in CD4T, CD8T, and NK cells. Additionally, we identified 39 cell-specific DMRs associated with MTX response. There were no significant findings in blood analyses.

*Conclusion*: We identified cell-specific changes in DNAm associated with MTX response in RA patients. Future studies into DNAm and MTX response should include measurements of DNAm from sorted cells in bulk tissues.

## Introduction

Rheumatoid arthritis (RA) is the most common systemic autoimmune disease and affects up to 1% of the global population.[1] Methotrexate (MTX) is a disease-modifying antirheumatic drug (DMARD) that is the recommended first treatment for RA.[2] Approximately 30-40% of patients continue with MTX treatment after two years.[3,4] Reasons for discontinuation of treatment include inefficacy and adverse events. Significant joint damage can occur in the early phase of RA, and response to the first treatment regime is an important indicator of long-term prognosis. [5–8]

MTX is a synthetic folate that has greatly increased binding affinity for *DHFR* than folic acid.[9] The specific mechanisms of the anti-inflammatory effects of MTX in RA are not fully explained, but are believed to include the accumulation of adenosine as a result of a reduction in purine metabolism, decreased proliferation and increased apoptosis of immune cells, and inhibition of cytokine production.[10] Previous research suggests that medications, including MTX, alter patterns of DNA methylation (DNAm).[9,11–13] This is of interest for several reasons, including that treatment-associated changes in the epigenome may explain, at least in part, the mechanisms of MTX in RA. Further, DNAm patterns prior to treatment and changes in DNAm associated with treatment may serve as predictors of treatment response.[14]

The most common tissue for epigenome wide studies (EWAS) using DNAm is blood or PBMCs extracted from blood.[15] Peripheral blood contains several different cell types, each of which have different methylation profiles. DNAm measurements from blood are a combination of DNAm in the constituent cell-types and adjustment for global cell-proportions is critical when performing epigenome wide association studies.[16,17] However, adjustment global cell-proportions can limit researchers ability to detect differential DNAm and cell-specific differential DNAm may be obscured if, for example, the differential DNAm at a CpG for a given phenotype is present in only one cell-type or if the direction of differential DNAm is in opposing directions between certain cell-types. Previous research has identified patterns of global methylation associated with MTX response and two CpGs associated with MTX response using DNAm from T-lymphocytes measured before treatment initation.[18,19] Two previous studies investigating the association between DNAm and MTX treatment response using DNAm from whole blood and isolated PBMCs found limited evidence of DNAm associated with MTX response.[20,21] No studies have investigated DNAm from sorted cells and MTX response.

In this study we estimated changes in cell-specific DNAm associated with treatment response from whole-blood samples collected from RA patients before and after initiation of MTX treatment. Patients included this study were from the Rheumatoid Arthritis Medication Study (RAMS) and University of California San Francisco Rheumatoid Arthritis treatment study (UCSF-RA).

## Patients and Methods

*Patient data*

A flow chart of study procedures and analyses is displayed in Figure 1. Participants included this is research were from two studies: University of California, San Francisco RA Treatment response study (UCSF-RA) and the Rheumatoid Arthritis Medication Study (RAMS). UCSF-RA

participants were recruited from rheumatology clinics in San Francisco County between 2016 and 2020. The methods of the RAMS study have been described previously.[20] Briefly, RAMS is a one-year United Kingdom based multicenter study of RA patients that observed RA patients who began first treatment with MTX. Both UCSF-RA and RAMS participants were naïve to MTX at baseline. For RAMS participants, blood samples were collected at baseline and four weeks after treatment initiation. DASs were measured at baseline before treatment initiation and at 6 months after treatment initiation. For UCSF-RA participants, blood samples were collected at baseline and at the next follow-up visit, approximately 3-6 months after treatment initiation. Genome-wide DNAm profiles were generated from the Illumina Infinium EPIC (UCSF-RA) and 450K BeadChips (RAMS) using peripheral blood.

*Disease Activity Scores and Treatment Response Criteria*

For this study, the Disease Activity Score based on 28 joints with C-reactive protein (hereafter, DAS28) was used as the primary measure of disease activity.[22] Component measurements of the DAS28 were measured at the pre-treatment baseline visit and at the follow-up visit (RAMS: 6 months after baseline; UCSF-RA: 3-6 months after baseline). CRP levels were measured from sera collected from whole blood collected at each visit. Missing DAS28 scores were imputed using Clinical Disease Activity Index scores.[23] Response to MTX was defined with the European Alliance of Associations for Rheumatology (EULAR) criteria: Good response = DAS28 $\leq$ 3.2 after treatment and DAS28 improvement $>$1.2 or Moderate response = 3.2 $<$ DAS28 after treatment $\leq$ 5.1 and 0.6 $<$ DAS28 improvement $\leq$ 1.2; and No response = DAS28 after treatment $>$ 5.1 and DAS28 improvement $\leq$ 0.6 [24].

*Methylation Data Quality Control*

DNAm data for each participant was measured using the Illumina 450K (RAMS) and EPIC (UCSF-RA) platforms from DNA extracted from whole blood samples. To prevent inducing batch effects associated with pre- and post-treatment visits stratified random sampling was used to balance out the number of pre- and post-treatment samples on each DNAm array (RAMS) or each participant's pre- and post-treatment samples was processed on the same array (UCSF-RA).

*DNA methylation data processing*

DNAm data was processed using the *minfi* R package.[25] Raw DNAm data from 450K and EPIC and chips were combined including only overlapping CpG sites (452,567 sites). Normalization and QC steps were performed on the combined DNAm data. We excluded samples with ≥5% detection p<0.01 (0 samples) and CpGs with ≥5% detection p<0.01 (834 sites excluded). CpG sites with annotated SNPs in the single base extension or CpG were excluded (16,130 sites excluded). Previously identified cross-reactive probes were also excluded (26,854 sites excluded).[26] And lastly, we excluded CpG sites on sex chromosomes, leaving 399,716 CpG sites for analyses. Background subtraction and dye-bias correction was performed using p*reprocessNoob* and between-array normalization was performed using *preprocessFunnorm.*

*Global cell proportion estimation*

Global cell type proportions for B cells, CD8+ and CD4+ T-cells, monocytes, natural killer (NK) cells, and neutrophils, were estimated separately within each platform (450K and EPIC) using *estimateCellCounts2()* with IDOL Optimized CpGs.[27]

*Correction for platform*

There is substantial overlap in CpG coverage between EPIC and 450K platforms; however, previous studies have found global DNAm differences between two platforms.[28] We used *Harman*, a method to correct DNAm measurements for batch effects constrained by the probability of overcorrection, to remove the effect of platform (EPIC vs. 450K) on global DNAm.[29] A matrix of normalized and QC'ed M-values were used as input with a confidence limit of 95% corresponding to a 95% probability that only batch variation is being removed (Figure S-1).

*Differential Position Analyses*

*Blood methylation*

EWAS on association between differential DNAm in blood and treatment response was performed with *limma*.[30] EULAR treatment response was collapsed into a binary variable, TR (TR=1: Good or moderate EULAR response, TR=0: No EULAR response). The primary parameter of interest in this study was the difference in the change in DNAm between follow-up (time=1) and baseline (time=0) between treatment response groups. A time interaction model was used to estimate this parameter:

$$CpG_j \sim \beta_0 + \beta_1 TR_i + \beta_2 Time_i + \beta_3 (TR_i \times Time_i)$$

$$\beta_3 = (\beta_{TR=1}^{T1} - \beta_{TR=1}^{T0}) - (\beta_{TR=0}^{T1} - \beta_{TR=0}^{T0})$$

where $\beta_3$is the estimate of the difference in DNAm at $CpG_j$ between treatment responders (TR=1) and non-responders (TR=0) between baseline and follow-up visits. Additionally, we also implemented the following models: (1) change in DNAm from baseline to follow-up among all participants, (2) change in DNAm from baseline to follow-up separately among treatment responders and non-responders, (3) change in DNAm at baseline between treatment responders and non-responders, and (4) change in DNAm at follow-up between treatment responders and non-responders. Covariates in all models included age, sex, tobacco smoking history (ever or never smoked), array slide, and estimated global cell proportions.

*Cell specific analyses*

*Tensor Composition Analysis* (*TCA*) was used to estimate differential DNAm between treatment response and cell-specific DNAm. *TCA* is a method for estimating cell type-specific associations between DNAm and disease phenotypes using bulk DNAm data. The linear models defined above were also implemented in *TCA*. Estimated global cell proportions were included as input. We implemented a two-step pipeline to estimate differential DNAm in *TCA*: (1) a joint model (JM) that tests for evidence of differential DNAm within any cell-type at a CpG; and (2) a

marginal conditional model (MCM), which tests for evidence of differential DNAm within a particular cell type adjusted for the other cell-types at a CpG. The JM can be thought of as an ANOVA test and did not provide evidence for differential DNAm in a specific cell type. All CpGs with $P < 1 \times 10^{-6}$ in the joint model were tested for cell specific differential DNAm with the MCM. *P*-values were adjusted for the false discovery rate using the Benjamini-Hochberg method, and adjusted MCM *P*-values <0.05 were considered significant.[31]

*Differential Region Analyses*

Exploratory Differentially Methylated Region (DMR) analyses were performed using the *Comb-P* method implemented in the ENmix R package.[32,33] *Comb-P* is a moving-averages method that uses autocorrelation between adjacent *P*-values within a genomic window to identify regions of differential DNAm. The method is agnostic to the statistical test used to generate the CpG level *P*-values. We identified DMRs using *P*-values estimated in *limma* and the MCM test in *TCA*. To generate genome-wide p-values for cell-specific DMR analyses, MCMs were applied to all CpGs in *TCA*. Parameter settings used were: bin.size=310, seed=0.001, dist.cutoff=750bp. Benjamini-Hochberg and Dunn–Šidák methods were used to correct for multiple tests[31]. DMRs with less than two CpGs were excluded from results.

*Whole blood global methylation estimates*

Global DNAm in whole blood was estimated by taking the mean methylation value across all CpGs sites that passed QC in each sample. Differences in global DNAm between response groups at baseline and follow-up were estimated using t-tests.

*Annotation of DMPs and DMRs*

DMPs and DMRs were annotated to genic features and CpG island location using Illumina and UCSF gene annotation. Additionally, the distance to the nearest upstream and downstream transcription start sites for CpGs was annotated using rGREAT.[34]

*Pathway analysis*

Analyses for enrichment of gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways was performed using gometh() from the *missMethyl* R package separately for the top 1000 nominally significant CpGs identified in *limma* ($P < 0.05$) and *TCA* ($P_{MCM} < 0.05$) analyses.

All statistical analyses were conducted using R 4.0.2[35].

**Results**

*Study participant characteristics*

Baseline characteristics of participants included in this research are reported in Table 1. Thirty-six of RAMS participants and four UCSF-RA participants were identified as treatment

responders. Age at baseline was similar between response groups among RAMS participants, however, UCSF-RA responders were approximately 6 years older than non-responders. All RAMS participants self-reported as White. Approximately half of UCSF-RA participants were White, with three participants identifying as African American, one participant identifying as Asian, and one participant identifying as Other. Baseline DAS28 was similar among the responders in both datasets. Non-responders in RAMS (mean=4.75, SD=1.19) had a higher baseline DAS28 than UCSF-RA (mean 3.44, SD=0.98) counterparts.

*Global methylation is different between responders and non-responders*

The distribution of differential DNAm estimates from *limma* and *TCA* analyses are presented in Figure 2. Among all participants, following treatment approximately half of CpGs were hypo- vs. hypomethylated (Figure 2A). The direction of global DNAm was different between EULAR responders and non-responders. Treatment responders had more hypomethylation following treatment compared to non-responders. Within time-point analyses comparing responder found reduced methylated among responders compared to non-responders at the follow-up visit.

*Global methylation changes are cell-specific*

The proportion of CpGs that were hypomethylated following treatment with MTX were different between cell-types (Figure 2B). The direction of DNAm change following MTX treatment was different between treatment response groups. For example, in neutrophils non-responders were more likely to be hypomethylated compared to responders. This contrasted with DNAm changes in monocytes, where there was more hypermethylation in non-responders than in responders. These cell-specific differences were also seen between treatment response groups within each visit.

*DMP results*

Five CpGs had $P < 1 \times 10^{-6}$ in *limma* DMP analyses (Table 2). Two CpGs (cg06336912 and cg15936718) were associated with decreased DNAm in responders compared to non-responders at the follow-up visit. The other three CpGs were from the model comparing change in DNAm over time between responders and non-responders (cg16868591), change in DNAm over time among responders (cg19506849), and change in DNAm among non-responders (cg18224793).

One CpGs reached genome-wide significance in the *TCA* JM (Table 3). cg13249593 was associated with cell-specific differential DNAm between responders and non-responders at baseline in CD4+ T-cells, CD8+ T-cell, and NK cells. cg06336912 is located approximately 21Kb upstream of *KRT19* and 23Kb and *KRT9*.

*DMR results*

No DMRs were identified from *limma* analyses. Using p-values from *TCA* MCMs, 39 cell-specific DMRs were identified in models comparing: change in DNAm over time among non-responders (3 DMRs), change in DNAm over time among responders (3 DMRs), and change in DNAm between responders and non-responders at baseline (16 DMRs) and at follow-up (17

DMRs). The top ten DMRs ranked by p-value are presented in Table 3. Top DMRs were located close to transcription start sites and CpG Islands.

*Pathway analyses*

No GO pathways reached significance after correction for multiple testing; however, several GO ontologies that reached nominal significance in cell-specific differential CpGs from *TCA* analyses are related to immune function and methotrexate response (Figure 3). At baseline, CpGs associated with differential DNAm between treatment non-responders were enriched for pathways related to regulation of immune system, adaptive immune response, and lymphocyte differentiation, among others. After treatment, CpGs associated with differential DNAm between treatment response groups were enriched for pathways possibly related to MTX response (antiporter activity and regulation of purine nucleotide metabolic process), and lymphocyte differentiation and proliferation.

In the KEGG pathway analyses Wnt signaling pathway (hsa043103, $P$: $8.9 \times 10^{-5}$, $P_{FDR}$: 0.03) was significant after correction for multiple testing in *limma* analyses estimating differential DNAm over time among all participants.

*Comparison of DNAm estimates with previous studies*

We compared our global DNAm estimates from DNAm estimates with CpGs identified by Glossop et al.[18] Glossop et al identified two CpGs (cg03018489 and cg14345882) associated with DAS28 treatment response before treatment in T lymphocytes. In *limma* analyses, cg14345882 was associated with hypermethylation in responders compared to non-responders at baseline at nominal significance ($\Delta$ Beta: 6.1%, $p$: $3.1 \times 10^{-3}$). In *TCA* analyses, the estimated direction of DNAm change was the same in CD4+ T-cells and CD8+ T-cells as in Glossop et al and *limma* analyses, but results did not reach nominal significance. cg03018489 was excluded during QC. Gosselt et al found evidence of differences in global DNAm in leukocytes between treatment response groups, with higher DNAm associated with non-response. [36] In our DNAm data from whole blood, we found evidence of differences in global DNAm between responders and non-responders after treatment ($\beta$: -0.24%, $P$=0.03), but not at baseline ($\beta$: -0.03%, $P$=0.79). Estimated cell-specific differential DNAm between responders and non-responders from *TCA* models indicates increased DNAm among non-responders at baseline in CD8+ T cells and monocytes, and increased DNAm among responders at baseline in CD4+ T cells and neutrophils (Figure 2B).

**Discussion**

In this study we investigated differential DNAm associated with EULAR treatment response to MTX among MTX native RA patients. We are the first to report results for cell-specific differential DNAm. Similar to recent studies investigating differential DNAm and response to MTX treatment, no significant DMPs or DMRs associated with treatment response were identified after applying multiple testing correction in bulk tissue analyses.[20,21] One DMP and 39 DMRs from cell-specific analyses were identified after multiple testing correction. cg13249593 was associated with differential DNAm in responders compared to responders in CD4+ (hyper) and CD8+ (hypo) T-cells and NK (hypo) cells. It is located approximately 20Kb upstream and

23Kb downstream of the transcription starts sites of *KRT9* and *KRT19*, respectively. *KRT19* produces the protein Keratin 19, which has been found in the synovial fluid of RA patients. Keratin 19 has been found to be an autoantigen in anti-CCP positive RA patients.[37] Evidence for differential DNAm at cg13249593 was found in the *limma* analyses as well (Δ Beta: -0.37, $P: 1.5 \times 10^{-3}$) but did not reach genome-wide significance. Estimated coefficients from models indicate that treatment with MTX altered DNAm at cg13249593 in non-responders and had little effect among non-responders at baseline (Figure S-2). The top DMR analyses identified several cell-specific DMRs. Of note, a DMR located 53 bases from the TSS of *ALOX12* was associated with differential DNAm between responders and non-responders at follow-up in monocytes and NK cells. Previous research has found that increased expression of *ALOX12* in monocytes is associated with juvenile rheumatoid arthritis and that MTX decreases the expression of *ALOX12*.[38,39] Estimated coefficients from *TCA* indicate that DNAm was higher among treatment responders than non-responders, suggesting decreased expression among responders.

There was evidence of differential DNAm in the top 20 genes associated with methotrexate from the Comparative Toxicogenomics Database (http://ctdbase.org/) in *limma* and *TCA* analyses at nominal significance ($P$<0.05; Figure S-4). These genes include *SLC19A1*, the gene most associated with MTX response, as well as *DHFR*, and *BLC2*. *SLC19A* is a folate transporter and is integral to the methotrexate pathway. Inhibition of *DHFR* through increased binding affinity of MTX, a synthetic folate, to *DHFR* compared to folate is the primary mechanism of methotrexate in cancer.[9] *BCL2* is an apoptosis regulator and inhibition of B-cell apoptosis has been associated with increased expression of *BCL2* in RA patients.[40]

Our results compared favorably to results from previous study by Glossop et al which identified two CpGs predictive of MTX response before treatment in T-lymphocytes.[18] While our results were not genome wide significant, the direction of DNAm differences at the one CpG that passed QC, both in blood and in CD4+ and CD8+ T-cells were the same as reported by Glossop et al, with higher baseline DNAm among MTX responders than non-responders (Figure S-3). We were not able to replicate findings of differences in global DNAm between response groups at baseline from Gosselt et al; however, we did find evidence of differential global DNAm after treatment in whole blood. We did not have cell-specific measurements of DNAm and therefore could not estimate global cell-specific DNAm differences directly. However, the distribution of hyper- vs. hypomethylated CpGs sites between response groups at baseline and follow-up from *TCA* analyses indicates suggestive evidence of differences at baseline DNAm between response groups, specifically in CD4+ and CD8+ T cell, monocytes, and neutrophils.

The lack significant DMPs and DMRs in DNAm from whole blood in this study is similar to results from two previous studies investigating response to DNA methylation which did not find evidence of differential DNAm associated with MTX treatment response. Recent studies with evidence of association between differential DNAm and MTX response used DNAm derived from T-lymphocytes.[18] This, along with our evidence of differences in the patterns of global cell-specific DNA methylation between response groups and the identification of a cell-specific DMP and several cell-specific DMRs suggest that associations with DNAm and MTX response are cell-specific.

Strengths of this research include the prospective study design used for each dataset. Patients in both RAMS and UCSF-RA studies were naïve to MTX at baseline and blood samples and disease activity scores were collected both before and after treatment initiation. Further, all participants included in this research received MTX monotherapy rather than combination therapy during the observation period. This was the first study to report on cell-specific DNAm associated with MTX using methods to estimate cell-specific DNAm from whole-blood. Patients in the RAMS study reported treatment compliance >80%.[20] We were unable to evaluate MTX treatment compliance in UCSF-RA participants.

Limitations include residual confounding from combining DNAm data from the 450K and EPIC platforms; however, PCA plots and results from analyses comparing associations between genome-wide DNAm principal components and platform indicate that we were able to remove platform effects with *Harman* (Figure S-1). Another limitation is the composition and number of participants in this research. Participants were recruited from rheumatology clinics in the UK and San Francisco Bay Area and nearly all participants self-identified as White, limiting generalizability of findings. Further, the sample size was relatively small preventing the identification of small changes in DNAm. Other limitations include differences in the study protocols between RAMS and UCSF-RA. In RAMS, blood samples for DNAm were collected at baseline and after 4 weeks of treatment, while UCSF-RA participants DNAm blood samples were collected at the baseline and follow-up visit. *TCA* has some limitations. The first is that the estimates of differential DNAm from *TCA* were not replicated in actual cell-specific DNAm measurements in our participants. Another limitation is that there is more power to detect differential DNAm in more abundant cell types compared to less abundant cell types. And lastly, the two time-point models used in the *limma* and *TCA* were slightly different. The *limma* models included a random effect for each participant. The *TCA* software cannot perform paired analyses using random effects or by including participant IDs as a covariate. Including IDs as a covariate would create an unidentifiable model (No. cell-types × No. parameters in model).

In, conclusion, we estimated changes in DNAm associated with response to MTX in RA patients using methods that deconvolute cell-specific DNAm at the CpG level. We identified evidence of cell-specific differential DNAm between responders and non-responders at baseline in one DMP at genome-wide significance. We also identified 39 cell-specific DMRs. No DMPs or DMRs were identified in whole blood analyses. Our findings of cell-specific differential DNAm associated with MTX response and the paucity of evidence of differential DNAm in this present research and similar studies using DNAm from whole blood and PBMCs indicate that future studies into DNAm and MTX response either need larger sample sizes to identify the modest effects of MTX on the methylome or DNAm from sorted cells.

## References


1. Almutairi KB, Nossent JC, Preen DB, Keen HI, Inderjeeth CA. The Prevalence of Rheumatoid Arthritis: A Systematic Review of Population-based Studies. The Journal of Rheumatology. 2021 May 1;48(5):669–76.

2. Lopez-Olivo MA, Siddhanamatha HR, Shea B, Tugwell P, Wells GA, Suarez-Almazor ME. Methotrexate for treating rheumatoid arthritis. Cochrane Database of Systematic Reviews [Internet]. 2014 [cited 2022 Apr 25];(6). Available from: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD000957.pub2/full

3. Romão VC, Canhão H, Fonseca JE. Old drugs, old problems: where do we stand in prediction of rheumatoid arthritis responsiveness to methotrexate and other synthetic DMARDs? BMC Medicine. 2013 Jan 23;11(1):17.

4. Verstappen SM, Owen SA, Hyrich KL. Prediction of response and adverse events to methotrexate treatment in patients with rheumatoid arthritis. International Journal of Clinical Rheumatology. 2012;7(5):559.

5. Lindqvist E, Jonsson K, Saxne T, Eberhardt K. Course of radiographic damage over 10 years in a cohort with early rheumatoid arthritis. Ann Rheum Dis. 2003 Jul;62(7):611–6.

6. Rezaei H, Saevarsdottir S, Forslind K, Albertsson K, Wallin H, Bratt J, et al. In early rheumatoid arthritis, patients with a good initial response to methotrexate have excellent 2-year clinical outcomes, but radiological progression is not fully prevented: data from the methotrexate responders population in the SWEFOT trial. Ann Rheum Dis. 2012 Feb;71(2):186–91.

7. Vastesaeger N, Xu S, Aletaha D, St Clair EW, Smolen JS. A pilot risk model for the prediction of rapid radiographic progression in rheumatoid arthritis. Rheumatology. 2009 Sep 1;48(9):1114–21.

8. Breedveld FC, Weisman MH, Kavanaugh AF, Cohen SB, Pavelka K, van Vollenhoven R, et al. The PREMIER study: A multicenter, randomized, double-blind clinical trial of combination therapy with adalimumab plus methotrexate versus methotrexate alone or adalimumab alone in patients with early, aggressive rheumatoid arthritis who had not had previous methotrexate treatment. Arthritis Rheum. 2006 Jan;54(1):26–37.

9. Brown PM, Pratt AG, Isaacs JD. Mechanism of action of methotrexate in rheumatoid arthritis, and the search for biomarkers. Nat Rev Rheumatol. 2016 Dec;12(12):731–42.

10. Wessels JAM, Huizinga TWJ, Guchelaar HJ. Recent insights in the pharmacological actions of methotrexate in the treatment of rheumatoid arthritis. Rheumatology. 2008 Mar 1;47(3):249–55.

11. Kim YI, Logan JW, Mason JB, Roubenoff R. DNA hypomethylation in inflammatory arthritis: reversal with methotrexate. The Journal of laboratory and clinical medicine. 1996;128:165–72.

12. Ellis JA, Munro JE, Chavez RA. Genome-scale case-control analysis of CD4+ T- cell DNA methylation in juvenile idiopathic arthritis reveals potential targets involved in disease. Clinical epigenetics. 2012;4(20).

13. Plant D, Wilson AG, Barton A. Genetic and epigenetic predictors of responsiveness to treatment in RA. Nature Reviews Rheumatology. 2014 Feb;10:329–329.

14. Nair N, Wilson AG, Barton A. DNA methylation as a marker of response in rheumatoid arthritis. Pharmacogenomics. 2017 Sep;18(14):1323–32.

15. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biol. 2014;15(2):R31.

16. Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlén SE, Greco D, et al. Differential DNA Methylation in Purified Human Blood Cells: Implications for Cell Lineage and Studies on Disease Susceptibility. PLoS One. 2012 Jul 25;7(7):e41361.

17. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012 May 8;13:86.

18. Glossop JR, Nixon NB, Emes RD, Sim J, Packham JC, Mattey DL, et al. DNA methylation at diagnosis is associated with response to disease-modifying drugs in early rheumatoid arthritis. Epigenomics. 2017 Apr;9(4):419–28.

19. Gosselt HR, van Zelst BD, de Rotte MCFJ, Hazes JMW, de Jonge R, Heil SG. Higher baseline global leukocyte DNA methylation is associated with MTX non-response in early RA patients. Arthritis Res Ther. 2019 Jun 26;21(1):157.

20. Nair N, Plant D, Plant D, Verstappen SM, Verstappen SM, Isaacs JD, et al. Differential DNA methylation correlates with response to methotrexate in rheumatoid arthritis. Rheumatology (United Kingdom). 2020 Jun;59(6):1364–71.

21. Gosselt HR, Vallerga CL, Mandaviya PR, Lubberts E, Hazes JMW, de Jonge R, et al. Epigenome wide association study of response to methotrexate in early rheumatoid arthritis patients. PLoS One. 2021 Mar 10;16(3):e0247709.

22. Tamhane A, Redden DT, McGwin G, Brown EE, Westfall AO, Reynolds RJ, et al. Comparison of the disease activity score using erythrocyte sedimentation rate and C-reactive protein in African Americans with rheumatoid arthritis. The Journal of rheumatology. 2013 Nov;40(11):1812–22.

23. Leong KP, Tan JWL, Gao X, Koh ET. Conversion among the 28-joint count activity indices for rheumatoid arthritis. Eur J Rheumatol. 2020 Jul;7(3):105–11.

24. Fransen J, van Riel PLCM. The Disease Activity Score and the EULAR response criteria. Clin Exp Rheumatol. 2005 Oct;23(5 Suppl 39):S93-99.

25. Fortin JP, Triche TJ, Hansen KD, Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. Bioinformatics (Oxford, England). 2017;33(4):558–60.

26. Chen Y an, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics. 2013 Feb;8(2):203–9.

27. Salas LA, Koestler DC. FlowSorted.Blood.EPIC: Illumina EPIC data on immunomagnetic sorted peripheral adult blood cells [Internet]. 2020. Available from: https://github.com/immunomethylomics/FlowSorted.Blood.EPIC

28. Logue MW, Smith AK, Wolf EJ, Maniates H, Stone A, Schichman SA, et al. The correlation of methylation levels measured using Illumina 450K and EPIC BeadChips in blood samples. Epigenomics. 2017 Nov;9(11):1363–71.

29. Oytam Y, Sobhanmanesh F, Duesing K, Bowden JC, Osmond-McLeod M, Ross J. Risk-conscious correction of batch effects: maximising information extraction from high-throughput genomic datasets. BMC Bioinformatics. 2016 Sep 1;17(1):332.

30. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research. 2015 Jan;43(7):e47–e47.

31. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological). 1995;57(1):289–300.

32. Pedersen BS, Schwartz DA, Yang IV, Kechris KJ. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. Bioinformatics. 2012 Nov 15;28(22):2986–8.

33. Xu Z, Xie C, Taylor JA, Niu L. ipDMR: identification of differentially methylated regions with interval P-values. Bioinformatics. 2021 Mar 1;37(5):711–3.

34. Gu Z. rGREAT: Client for GREAT Analysis. 2021.

35. R Core Team (2020). — European Environment Agency [Internet]. [cited 2021 Oct 1]. Available from: https://www.eea.europa.eu/data-and-maps/indicators/oxygen-consuming-substances-in-rivers/r-development-core-team-2006

36. Gosselt HR, van Zelst BD, de Rotte MCFJ, Hazes JMW, de Jonge R, Heil SG. Higher baseline global leukocyte DNA methylation is associated with MTX non-response in early RA patients. Arthritis Research & Therapy. 2019 Jun 26;21(1):157.

37. Poulsen TBG, Damgaard D, Jørgensen MM, Senolt L, Blackburn JM, Nielsen CH, et al. Identification of Novel Native Autoantigens in Rheumatoid Arthritis. Biomedicines. 2020 May 29;8(6):141.

38. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, Wiegers J, Wiegers TC, et al. Comparative Toxicogenomics Database (CTD): update 2021. Nucleic Acids Research. 2021 Jan 8;49(D1):D1138–43.

39. Cepika AM, Banchereau R, Segura E, Ohouo M, Cantarel B, Goller K, et al. A multidimensional blood stimulation assay reveals immune alterations underlying systemic juvenile idiopathic arthritis. J Exp Med. 2017 Nov 6;214(11):3449–66.

40. Yang J, Zhao S, Yang X, Zhang H, Zheng P, Wu H. Inhibition of B-cell apoptosis is mediated through increased expression of Bcl-2 in patients with rheumatoid arthritis. Int J Rheum Dis. 2016 Feb;19(2):134–40.

**Figures**

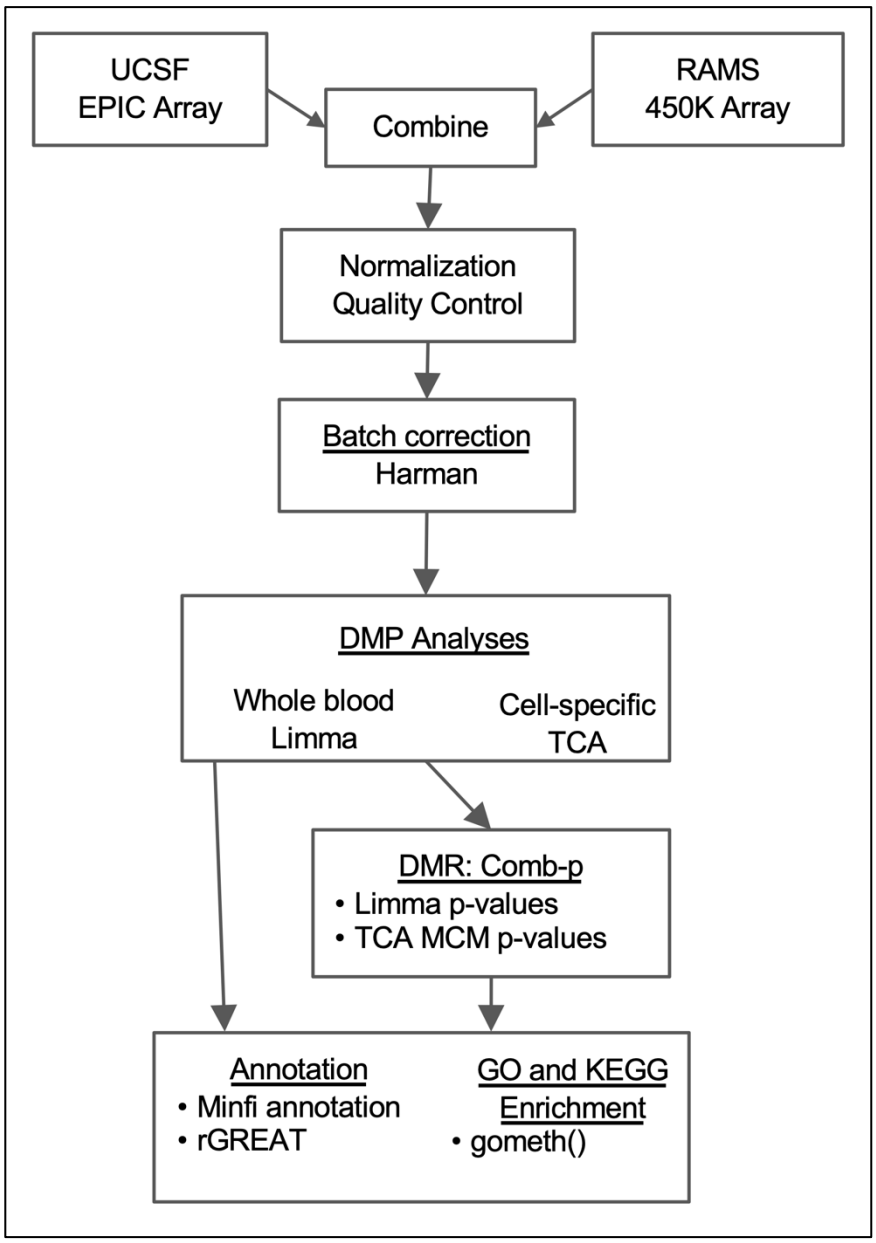



**Figure 1.** Flow chart of quality control procedures and differential DNA methylation analyses.

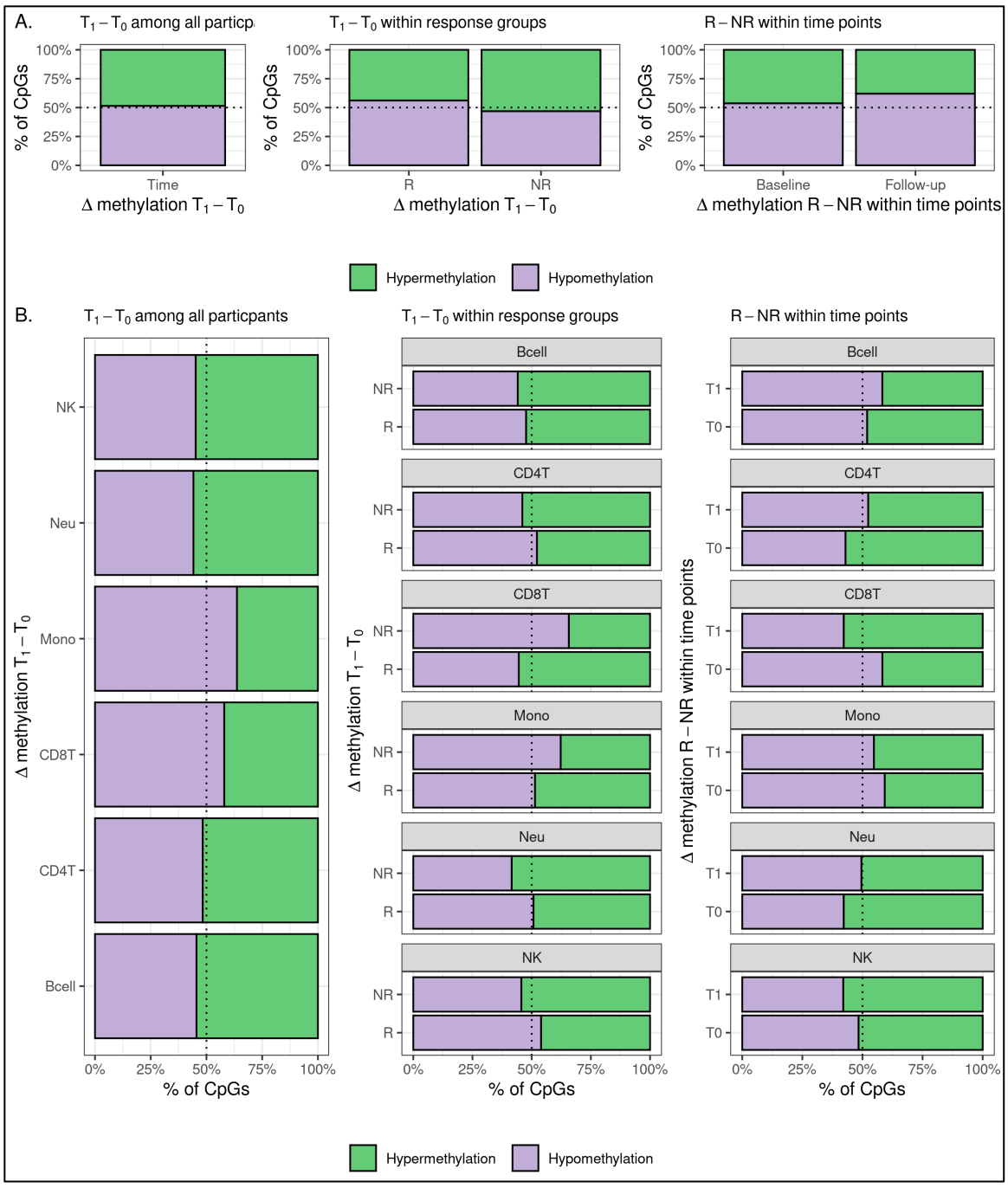



**Figure 2.** Global DNAm estimated in A) *limma* models and B) *TCA* models. Abbreviations: R, EULAR treatment responders; NR, EULAR treatment non-responders; $T_{0,}$ baseline visit DNAm; $T_1$ follow-up visit DNAm.

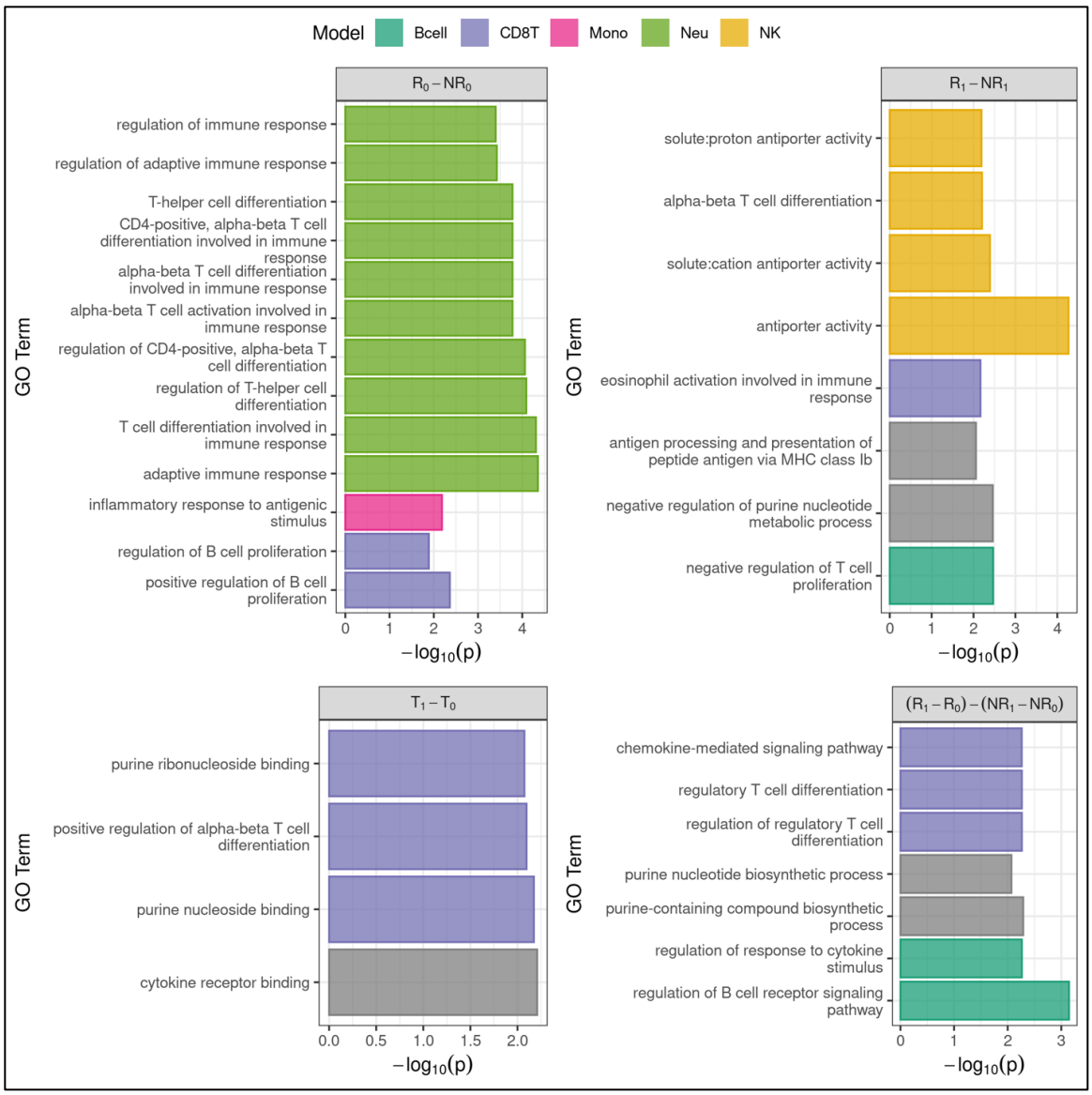



**Figure 3.** GO pathway analysis results for ontologies related to immune function and MTX response from top 1000 DMPs ($P$<0.05) for each cell-type. Abbreviations: $R_0$-$NR_0$, model for difference in DNAm between EULAR responders and non-responders at baseline; $R_1$-$NR_1$, model for difference in DNAm between EULAR responders and non-responders at follow-up; ($R_1$- $R_0$)- ($NR_1$- $NR_0$), Difference in change in DNAm over time between treatment responders and non-responders.

## Tables

**Table 1.** Baseline characteristics of participants.

| | RAMS (n=66) | | UCSF-RA (n=11) | |
|---|---|---|---|---|
| | **Response[a]** | **No Response[b]** | **Response[a]** | **No Response[b]** |
| n | 36 | 30 | 4 | 7 |
| Female, n (%) | 25 ( 69.4) | 24 ( 80.0) | 4 (100.0) | 6 (85.7) |
| Age at baseline, mean (SD), years | 60.13 (14.02) | 59.17 (15.24) | 61.85 (10.48) | 54.95 (12.05) |
| Self-reported race, n (%) | | | | |
| Asian | 0 ( 0.0) | 0 ( 0.0) | 0 ( 0.0) | 1 (14.3) |
| African American | 0 ( 0.0) | 0 ( 0.0) | 1 ( 25.0) | 2 (28.6) |
| Other | 0 ( 0.0) | 0 ( 0.0) | 0 ( 0.0) | 2 (28.6) |
| White | 36 (100.0) | 30 (100.0) | 3 ( 75.0) | 2 (28.6) |
| Baseline DAS28, mean (SD) | 4.96 (1.02) | 4.10 (1.30) | 3.83 (0.70) | 3.15 (1.04) |
| Ever smoke=Yes, n (%) | 20 ( 55.6) | 16 ( 53.3) | 1 ( 25.0) | 5 (71.4) |

Abbreviations: DAS28, Disease Activity Score 28 point with C-reactive protein; RAMS, Rheumatoid Arthritis Medication Study; SD, standard deviation UCSF-RA, University of California, San Francisco Rheumatoid Arthritis Study

[a] EULAR treatment response is “Good” or “Moderate”.

[b] EULAR treatment response is “None”.

**Table 2.** *limma* DMP results with $P < 1 \times 10^{-6}$.

| | | Annotation | | | | |
|---|---|---|---|---|---|---|
| **CpG** | **chr:bp (hg19)** | **GREAT[a]** | **UCSC** | **Model[b]** | **% ΔBeta** | ***P*** |
| cg06336912 | chr16:90173019 | *PRDM7* (-30681) | | $T_1$: R-NR | -0.17 | 6.5e-6 |
| cg15936718 | chr16:90088801 | *GAS8* (-207) | *GAS8* | $T_1$: R-NR | -0.28 | 8.9e-6 |
| cg16868591 | chr19:12803493 | *DHPS* (-10777); *FBXW9* (+3964) | *FBXW9* | ($R_1$-$R_0$)-($NR_1$-$NR_0$) | -0.30 | 3.0e-6 |
| cg19506849 | chr10:114767609 | *HABP2* (-545176); *TCF7L2* (+57600) | *TCF7L2* | $R_1$-$R_0$ | 0.92 | 9.4e-6 |
| cg18224793 | chr7:100222124 | *MOSPD3* (+11991); *TFR2* (+18220) | *TFR2* | $NR_1$-$NR_0$ | 1.50 | 4.7e-6 |

Abbreviations: bp, base pair; chr; chromosome; CI, Confidence Interval; FDR, false discovery rate; GREAT, Genomic Regions Enrichment of Annotations Tool; NR, treatment non-responder; R, treatment responder; $T_0$, baseline visit.

[a] Distance from CpG to transcription start site of nearest upstream and downstream gene from rGREAT.

**Table 3.** Top *TCA* DMP results

| **CpG** | **chr:bp (hg19)** | **GREAT annotation** | **Model** | **Cell type** | **Δ Meth.** | **Joint model** | **Marginal conditional model *P*** |
|---|---|---|---|---|---|---|---|
| | | | | CD4T | 0.226 | | 4.6E-6 |
| | chr17: | KRT19 (-20674); | $T_0$: R- | CD8T | -0.310 | *P*: 2.4E-8 | 1.8E-5 |
| cg13249593 | 39705234 | KRT9 (+23076) | NR | NK | -0.489 | $P_{FDR}$: 9.5E-3 | 2.3E-6 |

Abbreviations: bp, base pair; chr; chromosome; CI, Confidence Interval; FDR, false discovery rate; GREAT, Genomic Regions Enrichment of Annotations Tool; NR, treatment non-responder; R, treatment responder; $T_0$, baseline visit.

[a] Distance from CpG to transcription start site of nearest upstream and downstream gene from rGREAT.

[b] Model comparing difference in DNA methylation baseline before MTX treatment between those were identified as treatment responders compared to non-responders.

[c] Joint model from tensor composition analysis tests for evidence of differential methylation within any cell-type at each CpG. All CpGs with evidence of cell-specific methylation in the joint model at $p<1 \times 10^{-6}$ were tested for differential methylation within each cell-type using the marginal conditional model.

**Table 4.** Top 10 *TCA* DMRs.

| | | | | | | GREAT annotation[a] | Illumina annotation | |
|---|---|---|---|---|---|---|---|---|
| **Region chr:bp** | **Model** | **Cell type** | **No. CpG** | **ΔMeth.** | ***P*** | **Gene (Dist. to TSS)** | **Gene** | **Feature** |
| 1:161008461-61008826 | $T_0$: R-NR | CD8T | 8 | hyper | 9.5e-18 | *TSTD1* (+136) | *TSTD1* | TS200 |
| 12:9217389-9217907 | $T_0$: R-NR | Bcell | 10 | hyper | 5.8e-15 | | *LOC144571* | TS200 |
| 1:202310823-202311278 | $T_1$: R-NR | Neu | 7 | hyper | 3.1e-14 | *UBE2T* (+57) | *UBE2T* | TS200 |
| 17:6899084-6899577 | $T_1$: R-NR | Mono | 11 | hyper | 2.0e-13 | *ALOX12* (-53) | *ALOX12* | TS200 |
| 13:36871753-36872346 | R: $T_1$-$T_0$ | NK | 9 | hypo | 8.6e-13 | | *C13orf38* | TS200 |
| 1:161008461-161008826 | $T_0$: R-NR | CD4T | 8 | hypo | 1.8e-12 | *TSTD1* (+136) | *TSTD1* | TS200 |
| 17:6899084-6899577 | $T_1$: R-NR | NK | 11 | hypo | 4.9e-12 | *ALOX12* (-53) | *ALOX12* | TS200 |
| 10:77542301-77542585 | $T_0$: R-NR | CD4T | 9 | hyper | 1.5e-11 | *C10orf11* (-76) | *C10orf11* | TS200 |
| 6:32063725-32064258 | $T_1$: R-NR | NK | 16 | hypo | 5.9e-11 | *TNXB* (-50087) | *TNXB* | Body |
| 6:31148331-31148748 | $T_0$: R-NR | CD8T | 15 | hyper | 1.2e-10 | | | |

Abbreviations: bp, base pair; chr; chromosome; CI, Confidence Interval; FDR, false discovery rate; GREAT, Genomic Regions Enrichment of Annotations Tool; Hyper, Hypermethylation; Hypo, Hypomethylation; NR, treatment non-responder; R, treatment responder; $T_0$, baseline visit.

[a] Distance from CpG to transcription start site of nearest upstream and downstream gene from rGREAT.

**Dissertation Publications**

1. Pregnancy does not modify risk of multiple sclerosis in genetically susceptible women (Chapter 2)

1.

Adams C, Wu SL, Shao X, Bradshaw PT, Gonzales E, Smith JB, et al. Pregnancy does not modify the risk of MS in genetically susceptible women. Neurology - Neuroimmunology Neuroinflammation. 2020 Nov;7(6):e898–e898.