Uso de la aleatorización mendeliana en el análisis de la asociación entre niveles de vitamina D y el riesgo de cáncer de mama.

Association between Vitamin D levels and breast cancer risk using a Mendelian randomization approach

Pablo Fernández Navarro

Bioinformatics and Data Management (BIODAMA)
Cancer and Environmental Epidemiology Unit
National Center for Epidemiology
Carlos III Institute of Health

- Several studies have been done to evaluate the association of vitamin D deficiency and breast cancer risk. **There is controversy in the literature about this association**.

NIH NATIONAL CANCER INSTITUTE

- Epidemiologic studies have investigated whether people with higher vitamin D intakes or higher blood levels of vitamin D have lower risks of specific cancers.
- **The results of these studies have been inconsistent**, possibly because of the challenges in carrying out such studies.
- **Several randomized trials** of vitamin D intake have been carried out, but these were designed to assess bone health or other non-cancer outcomes.
- Although some of these trials have yielded information on **cancer incidence and mortality, the results need to be confirmed by additional research** because the trials were not designed to study cancer specifically.

https://www.cancer.gov/about-cancer/causes-prevention/risk/diet/vitamin-d-fact-sheet

Breast Cancer Res. 2011; 13(4): 217. PMCID: PMC3236325
Published online 2011 Aug 16. doi: 10.1186/bcr2846

Vitamin D and breast cancer: interpreting current evidence

Rowan T Chlebowski[1]

- Current evidence regarding vitamin D and breast cancer was reviewed to inform clinical practice and identify potential research directions.

- The evidence was inconsistent, inconclusive as to causality, and insufficient to inform nutritional requirements.

- **Current evidence is sufficient to support further study** of factors influencing 25(OH)D levels, associations between 25(OH)D levels and breast cancer in **premenopausal** and Black women, moderate dose (≤2,000 IU D3/day) supplemental vitamin D use and breast cancer incidence, and observational studies evaluating whether a threshold higher 25(OH)D level is associated with adverse clinical outcome in **women with breast cancer**.

- **Before routine clinical application of any strategies** targeting vitamin D status for breast cancer prevention or therapy are undertaken, the limitations of the current evidence should be considered.

NEJM Journal Watch

INFORMING PRACTICE

November 18, 2016

Vitamin D Levels and Breast Cancer Outcome

William J. Gradishar, MD reviewing Yao S et al. JAMA Oncol 2016 Nov 10.

Highest versus lowest serum concentrations were associated with superior survival.

Controversy has swirled around the relationship between serum vitamin D levels and both the risk for developing breast cancer and the prognosis for those who develop breast cancer. Results from various reports exploring these potential associations have been criticized for issues with methodology and conflicting results.

Now, investigators have conducted a prospective, case–cohort analysis of the Pathways Study (Cancer Causes Control 2008; 19:1065), involving 4505 women (median age, 58.7) in the Kaiser Permanente Northern California health plan who had an incident diagnosis of breast cancer and were followed for outcomes and comorbidities at baseline, 12, 24, 48, 72, and 96 months. Most patients had estrogen-receptor–positive, stage I–III disease and luminal (A, B) subtype, though all stages and subtypes were represented. Blood samples were collected at a median of 69 days from diagnosis.

At baseline, 48% of participants were vitamin D deficient, and 35% were insufficient. African Americans and Hispanics had lower vitamin D levels than whites. Current smokers had lower levels than former smokers, whereas physical activity and vitamin D supplements and dietary intake were associated with higher levels. Body-mass index was inversely associated with vitamin D levels. Across tertiles of vitamin D levels, those in the highest third had superior overall survival, breast cancer–specific survival, and invasive disease-free survival compared with those in the lowest third. The associations were strongest in premenopausal women.

COMMENT

The results of this report add to the body of evidence that vitamin D levels are linked to breast cancer recurrence events and overall mortality. However, the conclusions regarding the impact of vitamin D levels and outcomes selected by clinical subtypes or menopausal status must be interpreted with caution because of limitations in sample size and pathologic information available from the clinical record. Even with those caveats, the baseline finding of almost 85% of patients having vitamin D deficiency or insufficiency is alarming from a bone-health issue alone.

EDITOR DISCLOSURES AT TIME OF PUBLICATION

Disclosures for William J. Gradishar, MD at time of publication

Consultant / Advisory board	Eisai
Editorial boards	Clinical Breast Cancer; Oncology
Leadership positions in professional societies	National Comprehensive Cancer Network (Chair, Breast Cancer Panel)

William J. Gradishar, MD

Editor-in-Chief
NEJM JOURNAL WATCH
ONCOLOGY AND HEMATOLOGY

Biography | Disclosures | Summaries

PLOS ONE

RESEARCH ARTICLE

Randomized controlled trials of vitamin D and cancer incidence: A modeling study

William B. Grant[1]*, **Barbara J. Boucher**[2]

1 Sunlight, Nutrition, and Health Research Center, San Francisco, California, United States of America, 2 Blizard Institute, Barts & The London School of Medicine & Dentistry, Queen Mary University of London, London, United Kingdom

* wbgrant@infionline.net

OPEN ACCESS

Citation: Grant WB, Boucher BJ (2017) Randomized controlled trials of vitamin D and cancer incidence: A modeling study. PLoS ONE 12 (5): e0176448. https://doi.org/10.1371/journal.pone.0176448

Editor: Andrzej T. Slominski, University of Alabama at Birmingham, UNITED STATES

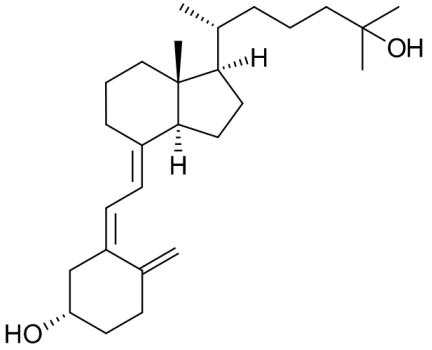
Received: March 3, 2017

Accepted: April 11, 2017

Published: May 1, 2017

Abstract

Although geographic ecological studies and observational studies find that ultraviolet B exposure and 25-hydroxyvitamin D [25(OH)D] concentrations are inversely correlated with 15–20 types of cancer, few randomized controlled trials (RCTs) of vitamin D support those findings. The poor design of some RCTs may account for that lack of support. Most vitamin D RCTs to date have considered the vitamin D dose, rather than initial, final, or changes in, serum 25(OH)D concentrations. Here a model is developed for use in designing and analyzing vitamin D RCTs with application to cancer incidence. The input variables of the model are vitamin D dose, baseline and achieved 25(OH)D concentrations, known rates of cancer for the population, and numbers of participants for the treatment and placebo arms is estimated—vitamin D dosage and numbers of participants are varied to achieve desired hazard ratio significance, using information from two vitamin D RCTs on cancer incidence conducted in Nebraska with good agreement between the model estimates and reported hazard ratios. Further improvements to the conduct of vitamin D RCTs would be to start the trial with a moderate bolus dose to achieve the desired 25(OH)D concentrations, and bloodspot 25(OH)D assay use in summer and winter annually to monitor seasonal and long-term changes in 25(OH)D concentration and compliance, and to allow dosage adjustment for achievement of desired vitamin D status.



Calcifediol, calcidiol, 25-hydroxycholecalciferol, or 25-hydroxyvitamin D (abbreviated 25(OH)D)

- Confounding, reverse causation and various biases can generate the associations, and even with careful study design and statistical adjustment, incorrect causal inference is possible

- The recognition of these problematic aspects of epidemiological investigation has led to the **application of a series of methods** aimed **at improving causal inference**

- A successful approach is to use **genetic variants as exposure indicators** that are not subject to the influences **that vitiate conventional study designs**, an approach known as **Mendelian randomization (MR)**

Mendelian Randomization: Genes as Instrumental Variables

- **Mendelian randomization (MR)** is a method of using measured variation in genes of known function to examine the causal effect of a modifiable exposure on disease in non-experimental studies.

- Mendelian randomization is a method that allows one to test for, or in certain cases to estimate, a causal effect from observational data in the presence of confounding factors.

- From a statistical point of view, MR is an application of the technique of instrumental variables with genotype acting as an instrument for the exposure of interest.

- Because genotypes are assigned randomly when passed from parents to offspring during meiosis, if we assume that choice of mate is not associated with genotype (panmixia), then the **population genotype distribution** should be **unrelated to the confounders** that typically plague observational epidemiology studies.

- In this regard, Mendelian randomization can be thought of as a **"natural"** randomized controlled trial.

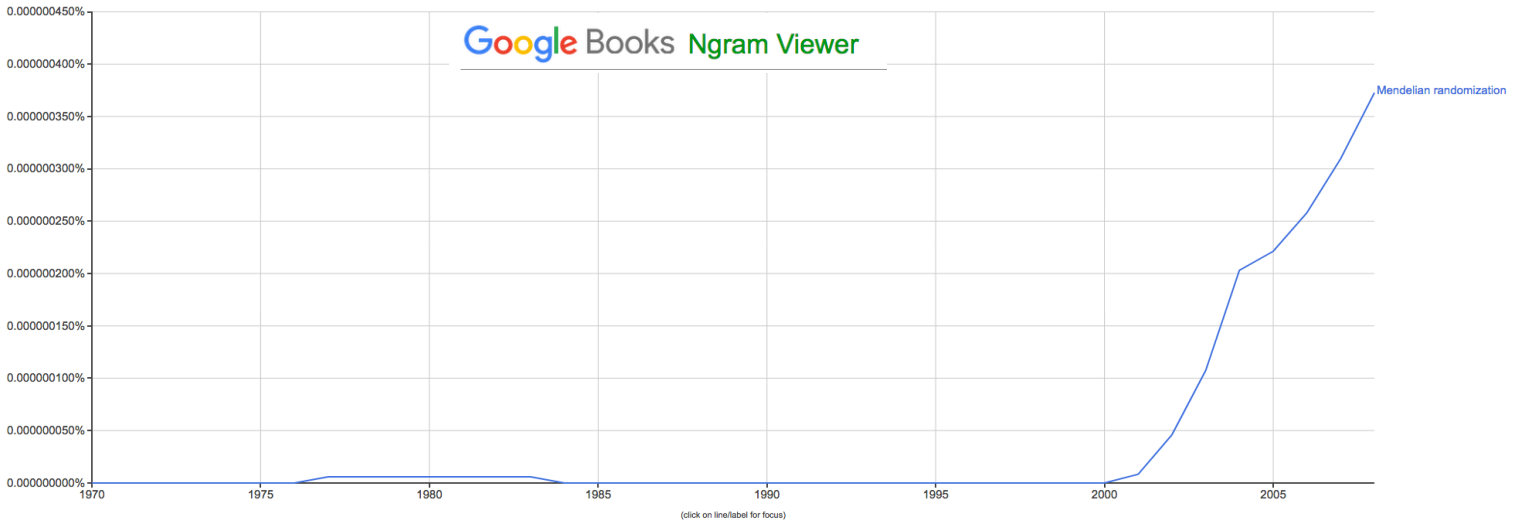
- Because the polymorphism is the instrument, Mendelian randomization is dependent on genetic association studies having provided good candidate genes for response to risk exposure.

- The design of MR was first proposed in **1986**:

  Katan MB. Apolipoprotein E isoforms, serum cholesterol, and cancer. Lancet. 1986 Mar 1;1(8479):507-8.

- Gray and Wheatley (**1991**) describe it as a method for obtaining unbiased estimates of the effects of a putative causal variable without conducting a traditional randomized trial. These authors also coined the term Mendelian randomization.

  Gray R, Wheatley K. How to avoid bias when comparing bone marrow transplantation with chemotherapy. Bone Marrow Transplant. 1991;7 Suppl 3:9-12. Review



International Journal of Epidemiology 2003;**32**:1–22
DOI: 10.1093/ije/dyg070

30TH THOMAS FRANCIS JR MEMORIAL LECTURE

'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?*

George Davey Smith and Shah Ebrahim

Associations between modifiable exposures and disease seen in observational epidemiology are sometimes confounded and thus misleading, despite our best efforts to improve the design and analysis of studies. Mendelian randomization—the random assortment of genes from parents to offspring that occurs during gamete formation and conception—provides one method for assessing the causal nature of some environmental exposures. The association between a disease and a polymorphism that mimics the biological link between a proposed exposure and disease is not generally susceptible to the reverse causation or confounding that may distort interpretations of conventional observational studies. Several examples where the phenotypic effects of polymorphisms are well documented provide encouraging evidence of the explanatory power of Mendelian randomization and are described. The limitations of the approach include confounding by polymorphisms in linkage disequilibrium with the polymorphism under study, that polymorphisms may have several phenotypic effects associated with disease, the lack of suitable polymorphisms for studying modifiable exposures of interest, and canalization—the buffering of the effects of genetic variation during development. Nevertheless, Mendelian randomization provides new opportunities to test causality and demonstrates how investment in the human genome project may contribute to understanding and preventing the adverse effects on human health of modifiable exposures.

- Associations between modifiable exposures and disease seen in observational epidemiology are sometimes confounded and thus misleading, despite our best efforts to improve the design and analysis of studies.

Statistical Methods in Medical Research 2007; **16**: 309–330

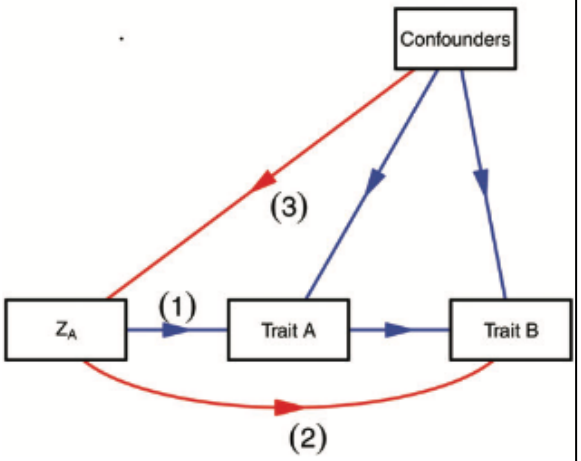
Mendelian randomization as an instrumental variable approach to causal inference

Vanessa Didelez Departments of Statistical Science, University College London, UK and
Nuala Sheehan Departments of Health Sciences and Genetics, University of Leicester, UK

In epidemiological research, the causal effect of a modifiable phenotype or exposure on a disease is often of public health interest. Randomized controlled trials to investigate this effect are not always possible and inferences based on observational data can be confounded. However, if we know of a gene closely linked to the phenotype without direct effect on the disease, it can often be reasonably assumed that the gene is not itself associated with any confounding factors – a phenomenon called *Mendelian randomization*. These properties define an instrumental variable and allow estimation of the causal effect, despite the confounding, under certain model restrictions. In this paper, we present a formal framework for causal inference based on Mendelian randomization and suggest using directed acyclic graphs to check model assumptions by visual inspection. This framework allows us to address limitations of the Mendelian randomization technique that have often been overlooked in the medical literature.

- **Frequent use of causal vocabulary** to express something that is more than association between genotypes, intermediate phenotypes and disease.

- While this is common practice in the medical literature where underlying knowledge about the biology of the problem may indeed allow one to deduce the direction of an observed association and where 'causal pathways' for disease are familiar concepts, **it is important for our purposes that we make a formal distinction between association and causation.**



Human Molecular Genetics, 2014, Vol. 23, Review Issue 1 **R89–R98**
doi:10.1093/hmg/ddu328
Advance Access published on July 4, 2014

Mendelian randomization: genetic anchors for causal inference in epidemiological studies

George Davey Smith* and Gibran Hemani

MRC Integrative Epidemiology Unit (IEU) at the University of Bristol, School of Social and Community Medicine, Bristol, UK

Received May 14, 2014; Revised June 19, 2014; Accepted June 20, 2014

Observational epidemiological studies are prone to confounding, reverse causation and various biases and have generated findings that have proved to be unreliable indicators of the causal effects of modifiable exposures on disease outcomes. Mendelian randomization (MR) is a method that utilizes genetic variants that are robustly associated with such modifiable exposures to generate more reliable evidence regarding which interventions should produce health benefits. The approach is being widely applied, and various ways to strengthen inference given the known potential limitations of MR are now available. Developments of MR, including two-sample MR, bidirectional MR, network MR, two-step MR, factorial MR and multiphenotype MR, are outlined in this review. The integration of genetic information into population-based epidemiological studies presents translational opportunities, which capitalize on the investment in genomic discovery research.

Type	Exposure/trait	Disease/outcome	Conclusion
Biomarkers	CRP	Coronary heart disease	Observational association between CRP and coronary heart disease is a result of confounding and/or reverse causation (18)
	Serum iron	Parkinson's disease	Higher serum iron levels lower the risk of Parkinson's disease (19)
	Uric acid	Coronary heart disease	Observational association between uric acid and coronary heart disease is, in part, due to confounding by BMI (20)
	Macrophage migration inhibitory factor (MIF)	Type 2 diabetes	Elevated MIF, amongst other factors, increases the risk of type 2 diabetes (21)
	Interleukin 6 (IL6)	Coronary heart disease	IL6 increases the risk of coronary heart disease (22,23)
Behaviours	Smoking	Anxiety/depression	Anxiety and depression amongst smokers does not appear to be a consequence of smoking (24,25)
	Alcohol consumption	Blood pressure	Alcohol use increases blood pressure (26)
Physiological measures	BMI	Symptomatic gallstone disease	Higher BMI increases the risk of symptomatic gall stone disease (27).
Maternal influences (corrected for genetic correlation between mother and child)	Alcohol consumption	Childhood school performance	The observational finding that moderate maternal alcohol intake is associated with more favourable school performance is due to confounding, and the casual association is in the opposite direction (28)
	Maternal BMI	Fat mass of offspring	Fat mass in children aged 9–11 is not strongly influenced by BMI of mothers during pregnancy (29)

Published in final edited form as:
Epidemiology. 2014 May ; 25(3): 427–435. doi:10.1097/EDE.0000000000000081.

Methodological challenges in Mendelian randomization

Tyler J. VanderWeele,
Departments of Epidemiology and Biostatistics, Harvard School of Public Health

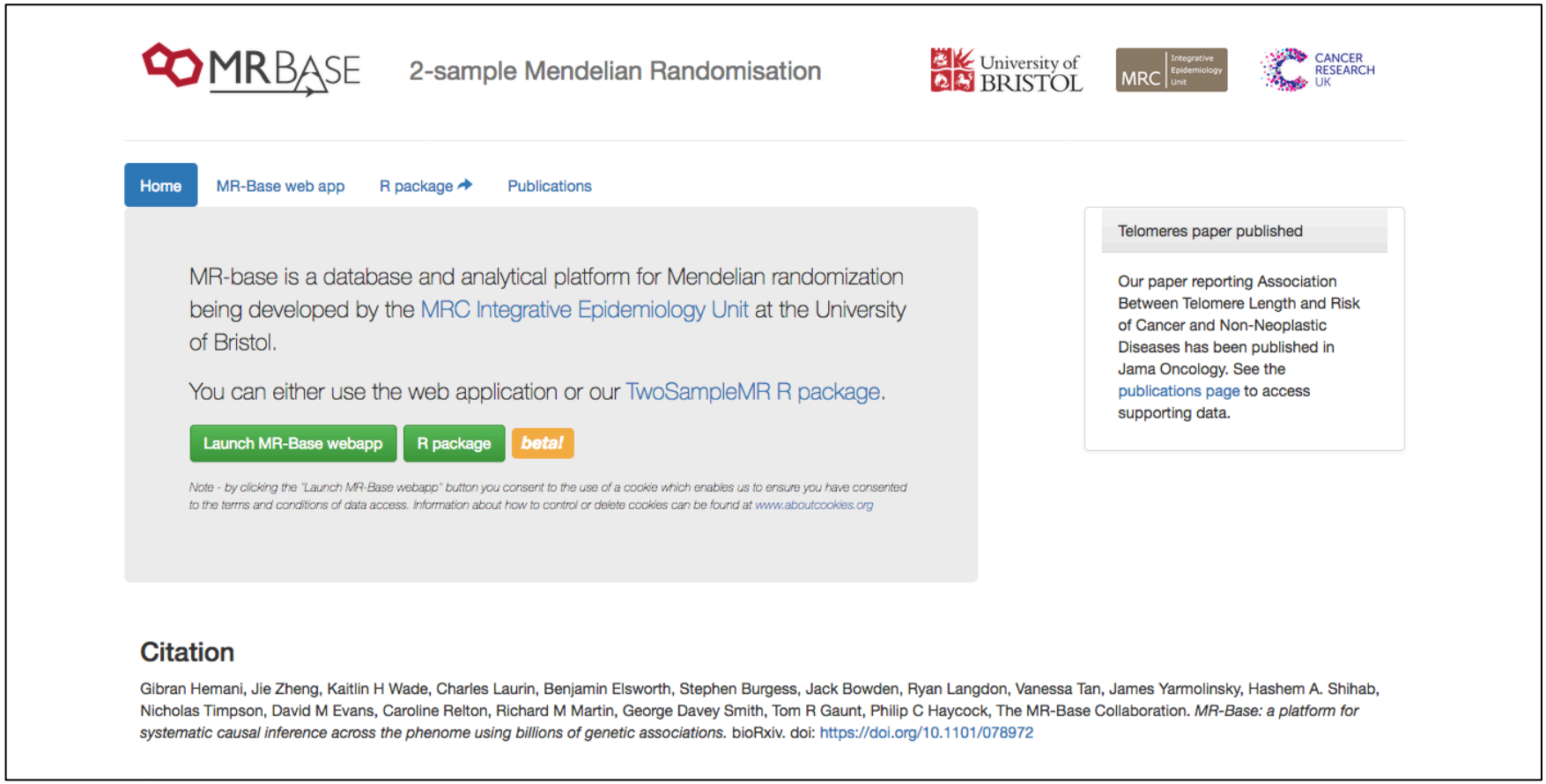
Eric J. Tchetgen Tchetgen,
Departments of Epidemiology and Biostatistics, Harvard School of Public Health

Marilyn Cornelis, and
Department of Nutrition, Harvard School of Public Health

Peter Kraft
Departments of Epidemiology and Biostatistics, Harvard School of Public Health677 Huntington Avenue

Abstract

We give critical attention to the assumptions underlying Mendelian randomization analysis and their biological plausibility. Several scenarios violating the Mendelian randomization assumptions are described, including settings with inadequate phenotype definition, the setting of time-varying exposures, the presence of gene-environment interaction, the existence of measurement error, the possibility of reverse causation, and the presence of linkage disequilibrium. Data analysis examples are given illustrating that the inappropriate use of instrumental variable techniques when the Mendelian randomization assumptions are violated can lead to biases of enormous magnitude. To help address some of the strong assumptions being made, three possible approaches are suggested. First, the original proposal of Katan (*Lancet*. 1986; 1:507-508) for Mendelian randomization was not to use instrumental variable techniques to obtain estimates, but merely to examine genotype-outcome associations to test for the presence of an effect of the exposure on the outcome. We show that this more modest goal and approach can circumvent many, though not all of, the potential biases described. Second, we discuss the use of sensitivity analysis in evaluating the consequences of violations in the assumptions and attempts to correct for those violations. Third, we suggest that a focus on negative, rather than positive, Mendelian randomization results may turn out to be more reliable.



http://www.mrbase.org/

Denis Whelan

EMORY ROLLINS SCHOOL OF PUBLIC HEALTH
Department of Biostatistics and Bioinformatics

May 25, 2017

Bioinformatics and Research Group in Genetic & Environmental Epidemiology
National Center for Epidemiology (Pab. 12)
Instituto de Salud Carlos III
C/ Monforte de Lemos 5, 28029 Madrid

To Whom It May Concern,

This summer, Denis Whelan, a second-year PhD student in Biostatistics intends to participate in a research rotation from June 19, 2017 to August 11, 2017 with the Bioinformatics and Research Group in Genetic & Environmental Epidemiology at the National Center for Epidemiology at Instituto de Salud Carlos III (ISCIII). Given that he will receive funding from the Laney Graduate School at Emory University, we understand that ISCIII is not responsible for covering his expenses and is not liable for any work-related injury. Furthermore, as Denis is covered under Emory Student Health Aetna Insurance Plan, he will not seek insurance coverage from ISCIII.

This project will expose him to causal inference methods applied to genetic data and will provide him with the opportunity to work towards a peer-reviewed publication in a scientific journal. Specifically, this project will involve using Mendelian randomization techniques to analyze polymorphisms involved in the relationship between Vitamin D and breast cancer. Using data from MCC-Spain, a large, population-based multi case-control study of five common tumors in Spain, this project will contribute to a deepened understanding of the complex causal relationship between Vitamin D and breast cancer risk.

Under the mentorship of the Drs. Pablo Fernández Navarro and Marina Pollán, Denis intends to deepen his understanding of concepts of causal inference and contribute to ISCIII's mission of using innovative scientific methods to prevent disease. This research experience provides a unique opportunity to enhance Denis's doctoral training and supports his long-term career goals of working as a biostatistician in an international context.

We appreciate your consideration.

Sincerely,

Lance A. Waller, Rollins Professor and Chair
Zhaohui (Steve) Qin, Associate Professor

Department of Biostatistics and Bioinformatics
Rollins School of Public Health, Emory University
1518 Clifton Road NE, Atlanta GA 30340

Objective: To assess the association between Vitamin D levels (blood 25-hydroxyvitamin D levels (25(OH)D) and breast cancer risk using a Mendelian randomization approach.

Design: MCC-Study (Case-Control, "Breast Cancer")

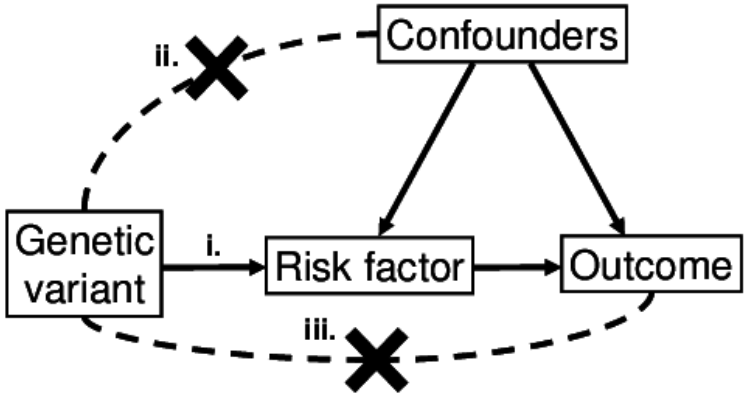
1. SNPs related with Vitamin D levels (SNPs VD)
2. Association Vitamin D levels and SNPs VD (vs Polygenic Risk Score) (Additive model) **(N=432)** (linear regression models)
3. Association SNPs VD and Breast Cancer risk (vs Polygenic Risk Score) (Additive model) **(Cases=1019,Controls=1195)** (mixed logistic regression models)
4. MR causal estimates and MR assumptions

Genotyping: MCC-Exome Array

Polygenic Risk Scores (Vitamin D) (Our final proposal of instrumental variable)

- Unweighted
- Weighted

Software: R Software



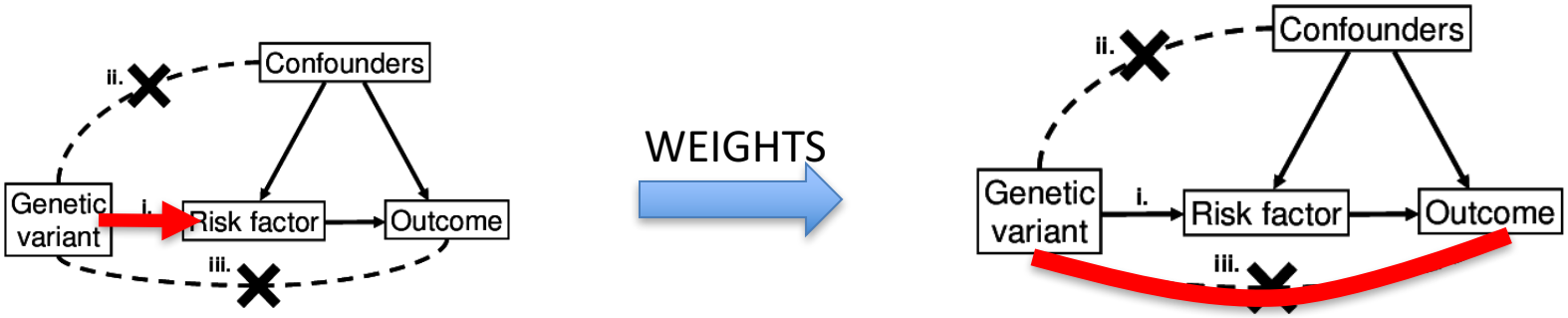
(1) The instrumental variable assumptions:

i. the genetic variant is associated with the risk factor.
ii. the genetic variant is not associated with confounders of the risk factor–outcome relationship.
iii. the genetic variant is not associated with the outcome conditional on the risk factor and confounders of the risk factor–outcome relationship.

(2) In addition to direct effects of the instrument on the disease misleading the analyst:

- Misleading conclusions may also arise in the presence of **linkage disequilibrium** with unmeasured directly-causal variants, **genetic heterogeneity, pleiotropy** (often detected as a genetic correlation), or **population stratification**.

Weighted Polygenic Risk Score



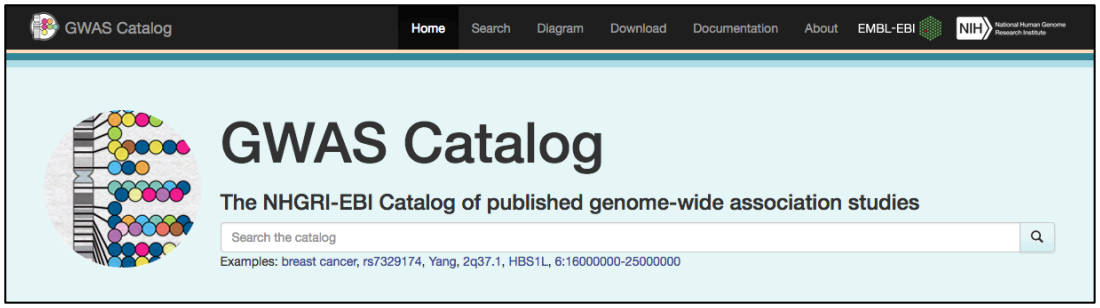

```
Call:
lm(formula = vd_25ohd3 ~ rs12512631 + rs705117 + rs2282679 +
    rs7041 + rs10485165 + rs10741657 + rs12785878 + rs3829251 +
    rs4762651 + rs10507577 + rs2547231 + rs6013897, data = datos.risk.controles)

Residuals:
    Min      1Q  Median      3Q     Max
-43.817 -13.231   0.052  12.150  92.413

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  49.5499     3.2874  15.072   <2e-16 ***
rs12512631    2.1367     1.6916   1.263   0.2072
rs705117      5.5406     8.3554   0.663   0.5076
rs2282679     2.1527     8.1598   0.264   0.7920
rs7041       -8.3473     8.0224  -1.041   0.2987
rs10485165    2.0965     1.7455   1.201   0.2304
rs10741657    3.6986     1.4011   2.640   0.0086 **
rs12785878    0.1649     2.0146   0.082   0.9348
rs3829251    -1.3967     2.2424  -0.623   0.5337
rs4762651    -0.2202     1.3635  -0.161   0.8718
rs10507577   -2.0503     2.3125  -0.887   0.3758
rs2547231     2.8891     1.8800   1.537   0.1251
rs6013897    -0.5731     1.6920  -0.339   0.7350
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 19.41 on 419 degrees of freedom
Multiple R-squared:  0.08453, Adjusted R-squared:  0.05832
F-statistic: 3.224 on 12 and 419 DF,  p-value: 0.0001892
```

GENETIC VARIANTS ASSOCIATED WITH VITAMIN D



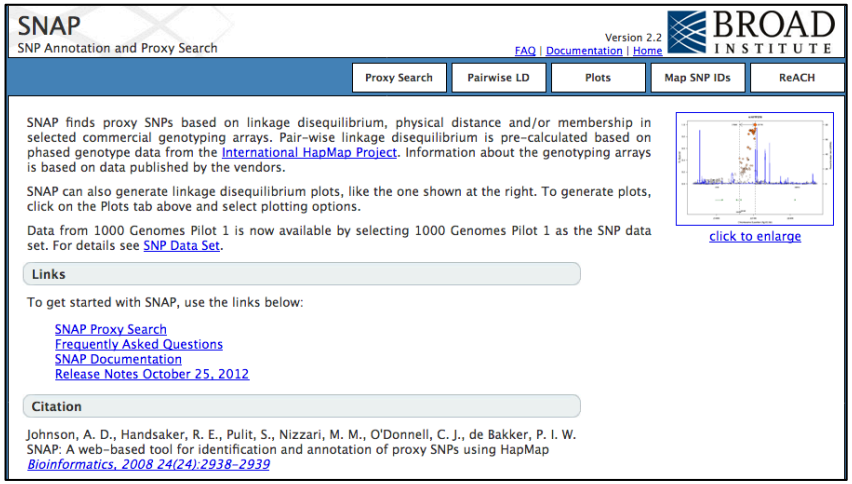
MULTI-CASO CONTROL - SPAIN

Team

illumina®

Exome Array

248264 genetic variants



Instituto de Salud Carlos III

ciberesp

MULTI-CASO CONTROL - SPAIN

Team

SNP	Gene	Database
rs10741657	CYP2R1	Wang 2010
rs10766192	CYP2R1	SUNLIGHT Consortium, validated
rs12794714	CYP2R1	SUNLIGHT Consortium, validated
rs1993116	CYP2R1	Ahn 2010
rs2060793	CYP2R1	Ahn 2010
rs1155563	GC	Ahn 2009
rs12512631	GC	Wang 2010
rs2282679	GC	Wang 2010
rs3755967	GC	SUNLIGHT Consortium, validated
rs7041	GC	Ahn 2009
rs11234027	DHCR7,NADSYN1	Ahn 2010
rs11603330	DHCR7,NADSYN1	SUNLIGHT Consortium, validated
rs12785878	DHCR7,NADSYN1	Wang 2010
rs3829251	DHCR7,NADSYN1	Ahn 2010
rs1321465	CYP24A1	Wang 2010
rs1321468	CYP24A1	Wang 2010
rs17216707	CYP24A1	SUNLIGHT Consortium, validated
rs17217119	CYP24A1	Wang 2010
rs6013897	CYP24A1	Wang 2010
rs6014011	CYP24A1	Wang 2010
rs6097931	CYP24A1	Wang 2010
rs6097944	CYP24A1	Wang 2010
rs8121940	CYP24A1	Wang 2010
rs4762651	AMDHD1	SUNLIGHT Consortium, not validated
rs182421	chr19	SUNLIGHT Consortium, not validated
rs2972516	chr19	SUNLIGHT Consortium, not validated
rs2972619	chr19	SUNLIGHT Consortium, not validated
rs1017993	SULT2A1	SUNLIGHT Consortium, not validated
rs2547231	SULT2A1	SUNLIGHT Consortium, not validated
rs2637125	SULT2A1	SUNLIGHT Consortium, not validated
rs296381	SULT2A1	SUNLIGHT Consortium, not validated
rs2463963	chr10	SUNLIGHT Consortium, not validated, not potential
rs10911580	chr1	SUNLIGHT Consortium, not validated, not potential
rs4861475	chr4	SUNLIGHT Consortium, not validated, not potential
rs11195965	chr10	SUNLIGHT Consortium, not validated, not potential
rs6008843	chr22	SUNLIGHT Consortium, not validated, not potential
rs6469626	chr8	SUNLIGHT Consortium, not validated, not potential
rs7981	chr7	SUNLIGHT Consortium, not validated, not potential
rs17609390	chr18	SUNLIGHT Consortium, not validated, not potential
rs878459	chr3	SUNLIGHT Consortium, not validated, not potential
rs17163911	chr7	SUNLIGHT Consortium, not validated, not potential

SNPS	DATE.ADDED.TO.CATALOG	PUBMEDID	FIRST.AUTHOR	DATE	JOURNAL
rs2282679	2010-07-01	20541252	Wang TJ	2010-06-09	Lancet
rs12785878	2010-07-01	20541252	Wang TJ	2010-06-09	Lancet
rs10741657	2010-07-01	20541252	Wang TJ	2010-06-09	Lancet
rs1155563	2015-05-07	25208829	Anderson D	2014-09-11	Genes Immun
rs3829251	2010-05-23	20418485	Ahn J	2010-04-23	Hum Mol Genet
rs2060793	2010-05-23	20418485	Ahn J	2010-04-23	Hum Mol Genet

SNPS	DISEASE.TRAIT
rs2282679	Vitamin D insufficiency
rs12785878	Vitamin D insufficiency
rs10741657	Vitamin D insufficiency
rs1155563	Vitamin D levels
rs3829251	Vitamin D levels
rs2060793	Vitamin D levels

SNPS	INITIAL.SAMPLE.SIZE
rs2282679	16,125 European ancestry individuals
rs12785878	16,125 European ancestry individuals
rs10741657	16,125 European ancestry individuals
rs1155563	1,073 European ancestry age 14 individuals, 67 age 14 individuals
rs3829251	4,501 European ancestry individuals
rs2060793	4,501 European ancestry individuals

SNPS	REPLICATION.SAMPLE.SIZE
rs2282679	17,871 European ancestry individuals
rs12785878	17,871 European ancestry individuals
rs10741657	17,871 European ancestry individuals
rs1155563	<NA>
rs3829251	2,221 European ancestry individuals
rs2060793	2,221 European ancestry individuals

Instituto de Salud Carlos III

ciberesp

MCC MULTI-CASO CONTROL - SPAIN

illumina®

Exome Array

248264 genetic variants

	nombre.original	ID.PABLO	chr.original	Gen.dist.original	pos.original	rs.del.nombre.original	ID
CREALExomeplus_rs10001480	4--87074032	4	94.12202	87074032	rs10001480	rs59201453;rs10001480	
CREALExomeplusrs10009993	4--87021505	4	94.03247	87021505	rs10009993	rs59515177;rs10009993	
CREALExomeplusrs10010011	4--87139714	4	94.23331	87139714	rs10010011	rs57005316;rs10010011	
CREALExomeplusrs1001190	3--183210972	3	186.28050	183210972	rs1001190	rs60053582;rs56742247;rs56498451;rs1001190	
CREALExomeplusrs10019305	4--70719533	4	80.48745	70719533	rs10019305	rs60794243;rs17684544;rs10019305	
CREALExomeplusrs1002204	7--87141497	7	99.44940	87141497	rs1002204	rs60650756;rs10358750;rs1002204	

	longitud	duplicado	marker	genesymbol	locusID	chr	chrpos	fxn_class	species	dupl_loc	current.rsid	flag	SNP_ID.QC	Strand.QC	Build.QC	Chr.QC
	2	FALSE	rs59201453	MAPK10	5602	4	87074032	intron-variant	Homo sapiens		rs10001480	0	rs10001480	+	b37	4
	2	FALSE	rs59515177	MAPK10	5602	4	87021505	intron-variant	Homo sapiens		rs10009993	0	rs10009993	+	b37	4
	2	FALSE	rs57005316	MAPK10	5602	4	87139714	intron-variant	Homo sapiens		rs10010011	0	<NA>	<NA>	<NA>	NA
	4	FALSE	rs60053582	KLHL6	89857	3	183210972	intron-variant	Homo sapiens		rs1001190	0	rs1001190	+	b37	3
	3	FALSE	rs60794243	SULT1E1	6783	4	70719533	intron-variant	Homo sapiens		rs10019305	0	rs10019305	+	b37	4
	3	FALSE	rs60650756	ABCB1	5243	7	87141497	intron-variant	Homo sapiens		rs1002204	0	rs1002204	+	b37	7

	Pos.QC	A1.QC	A2.QC	HWE.QC	MAF.QC	Type.QC	rs.QC	Blat_Flag.QC	Gene.QC	Location.QC	Amino_Acid_Change.QC	sc_damaging.QC	Trait_associationGWAS.QC	VarCat.QC
	87074032	A	G	0.8506	0.1869	custom	rs10001480	NA	MAPK10	INTRON		NA		
	87021505	A	G	0.8432	0.3915	custom	rs10009993	NA	MAPK10	INTRON		NA		
	NA	<NA>	<NA>	NA	NA	<NA>	<NA>	NA	<NA>	<NA>	<NA>	NA	<NA>	
	183210972	C	T	0.7456	0.2833	custom	rs1001190	NA	KLHL6	INTRON		NA		
	70719533	G	A	0.2382	0.3386	custom	rs10019305	NA	SULT1E1	INTRON		NA		
	87141497	A	C	0.4919	0.4420	custom	rs1002204	NA	ABCB1	INTRON		NA		

<NA>

Instituto de Salud Carlos III

ciberesp

MCC MULTI-CASO CONTROL - SPAIN

SNAP

SNP Annotation and Proxy Search

Version 2.2

FAQ | Documentation | Home

BROAD INSTITUTE

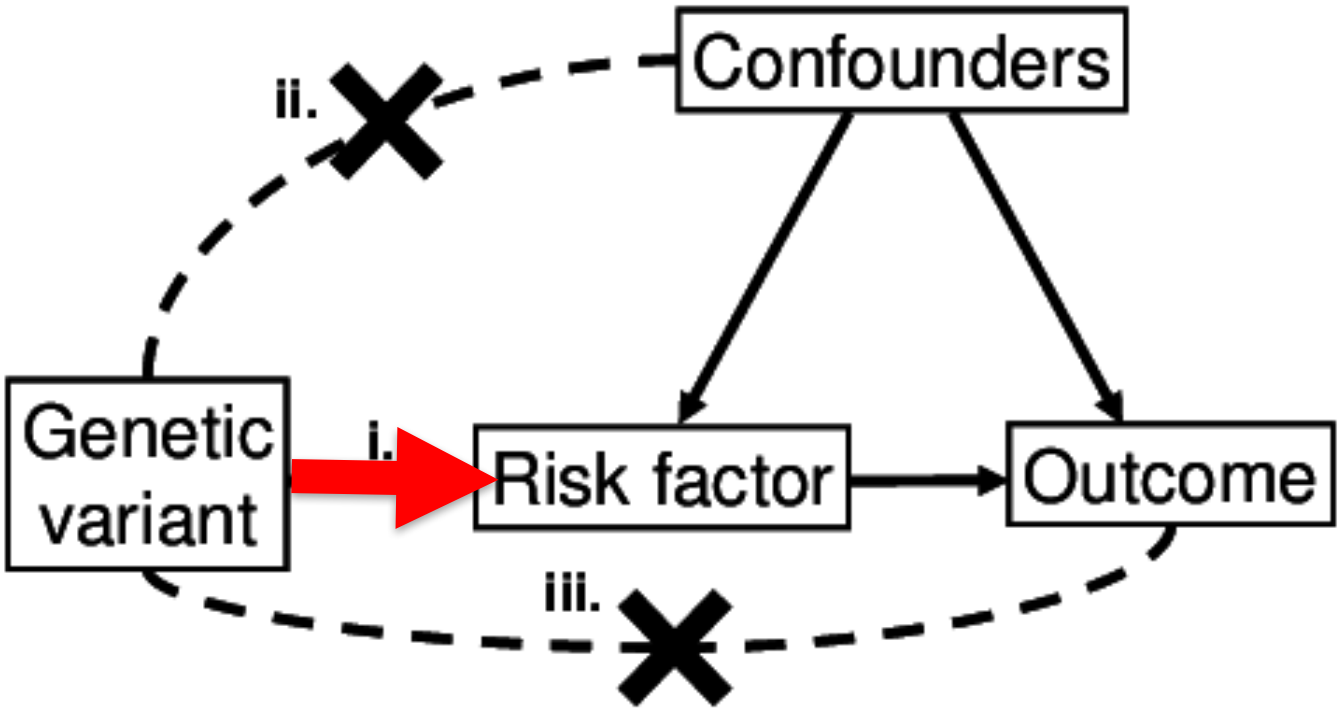
Proxy Search | Pairwise LD | Plots | Map SNP IDs | ReACH

SNP	Proxy	Distance	RSquared	DPrime	Arrays	Chromosome	Coordinate_HG18
rs12144344	rs12144344	0	1.000	1.000	AS,A5,A6,I2,I5,I6,I6Q,IM,IMD,OQ,IWQ,054,05E,AAH	chr1	76612124
rs12144344	rs12133165	484	1.000	1.000	None	chr1	76611640
rs12144344	rs1540964	1192	1.000	1.000	AN,A5,A6	chr1	76613316
rs12144344	rs1540963	1931	1.000	1.000	None	chr1	76614055
rs12144344	rs12239582	2299	1.000	1.000	A6,CYT,OQ,AxM,OE,024,028,054,05E,OEE	chr1	76609825
rs12144344	rs59741131	3798	1.000	1.000	None	chr1	76615922
rs12144344	rs1952021	6368	1.000	1.000	None	chr1	76618492
rs12144344	rs6593533	11369	1.000	1.000	CM,AAH	chr1	76623493
rs12144344	rs1359468	11599	1.000	1.000	AN,A6	chr1	76623723
rs12144344	rs9437145	12929	1.000	1.000	None	chr1	76625053
rs12144344	rs2893413	13040	1.000	1.000	None	chr1	76625164
rs12144344	rs4459157	22356	1.000	1.000	None	chr1	76634480
rs12144344	rs80136151	15657	0.967	1.000	None	chr1	76627781
rs12144344	rs2209458	23112	0.967	1.000	A6,I3,I5,I6,I6Q,IM,IMD,IC,ICQ,OQ,IWQ,OE,024,028,054,05E,OEE	chr1	76635236
rs12144344	rs2893414	23970	0.967	1.000	None	chr1	76636094
rs12144344	rs2392031	24138	0.967	1.000	None	chr1	76636262
rs12144344	rs2392034	24350	0.967	1.000	AX	chr1	76636474
rs12144344	rs9437435	7228	0.846	1.000	I2,I5,I6,I6Q,IM,IMD,IWQ,054,05E	chr1	76619352
rs10766192	rs10766192	0	1.000	1.000	IM,IMD,OQ,AxM,054,05E	chr11	14768262
rs10766192	rs3206554	13741	1.000	1.000	A6,I1	chr11	14754521
rs10766192	rs12416696	20797	1.000	1.000	IM,IMD,OQ,054,05E,AAH	chr11	14747465
rs10766192	rs1007392	37095	1.000	1.000	AG,I3,I5,I6,I6Q,IM,IMD,IC,ICQ,CYT,OQ,IWQ,OE,024,028,054,05E,OEE,AAH	chr11	14731167
rs10766192	rs10832299	42294	1.000	1.000	IM,IMD,OQ,OE,024,028,054,05E,OEE	chr11	14725968
rs10766192	rs12788030	47002	1.000	1.000	None	chr11	14721260
rs10766192	rs4144487	61235	1.000	1.000	None	chr11	14707027
rs10766192	rs10219313	61837	1.000	1.000	IM,OQ	chr11	14706425
rs10766192	rs10832294	64259	1.000	1.000	IM,IMD,IBC,OQ,OE,054,05E	chr11	14704003
rs10766192	rs7938266	102362	1.000	1.000	I3,I5,I6,I6Q,IM,IMD,IC,ICQ,CYT,OQ,CM,IWQ,OE,024,028,054,05E,OEE	chr11	14665900
rs10766192	rs10128681	110110	1.000	1.000	OQ	chr11	14658152
rs10766192	rs12290926	116640	1.000	1.000	None	chr11	14651622
rs10766192	rs11023302	117751	1.000	1.000	None	chr11	14650511
rs10766192	rs4757261	141884	1.000	1.000	I1,IM,IMD,OQ,OE,024,028,054,05E,OEE	chr11	14626378
rs10766192	rs10832297	53949	0.966	1.000	None	chr11	14714313
rs10766192	rs1451677	65674	0.966	1.000	None	chr11	14702588
rs10766192	rs10766189	77763	0.966	1.000	None	chr11	14690499
rs10766192	rs4757269	24506	0.965	1.000	IM,IMD,OQ,OE,024,028,054,05E,OEE	chr11	14792768
rs10766192	rs11023332	27576	0.933	1.000	None	chr11	14740686
rs10766192	rs10832301	22722	0.932	1.000	None	chr11	14745540
rs10766192	rs66752127	22723	0.932	1.000	None	chr11	14745539
rs10766192	rs11023350	49831	0.931	0.965	I3,I5,I6,I6Q,IM,IMD,IC,ICQ,OQ,IWQ,OE,024,028,054,05E,OEE	chr11	14818093

12 SNPs

nombre.original	SNP_ID.QC	ID
CREALExomeplusrs12512631	rs12512631	rs60927268;rs17709228;rs12512631
CREALExomeplusrs2547231	rs2547231	rs60928263;rs2547231
CREALExomeplusrs4762651	rs4762651	rs58367634;rs4762651
CREALExomeplusrs6013897	rs6013897	rs61437793;rs8116041;rs6013897
CREALExomeplusrs7041	rs7041	rs60269151;rs52808188;rs17349198;rs3737551;rs3172682;rs1047213;rs222037;rs7041
CREALExomeplusrs705117	rs705117	rs57497402;rs705117
exm-rs10485165	rs10485165	rs60623809;rs17210017;rs10485165
exm-rs10507577	rs10507577	rs57261782;rs10507577
exm-rs10741657	rs10741657	rs61548184;rs17488643;rs10741657
exm-rs12785878	rs12785878	rs56554767;rs12785878
exm-rs2282679	rs2282679	rs116893012;rs60434403;rs2282679
exm-rs3829251	rs3829251	rs60451067;rs3829251

nombre.original	SNP_ID.QC	genesymbol	chr	chrpos	fxn_class	Build.QC	HWE.QC	MAF.QC
CREALExomeplusrs12512631	rs12512631		4	72601331		b37	0.444600	0.3660
CREALExomeplusrs2547231	rs2547231	SULT2A1	19	48385057	intron-variant	b37	0.581000	0.1399
CREALExomeplusrs4762651	rs4762651	AMDHD1	12	96350254	intron-variant	b37	0.006751	0.3353
CREALExomeplusrs6013897	rs6013897		20	52742479		b37	0.817200	0.2051
CREALExomeplusrs7041	rs7041	GC	4	72618334	missense,reference	b37	1.000000	0.4486
CREALExomeplusrs705117	rs705117	GC	4	72608115	intron-variant	b37	0.467800	0.1381
exm-rs10485165	rs10485165		6	89112817		b37	0.946200	0.1672
exm-rs10507577	rs10507577		13	53968091		b37	0.082460	0.1027
exm-rs10741657	rs10741657	CYP2R1	11	14914878	upstream-variant-2KB	b37	0.603100	0.3798
exm-rs12785878	rs12785878	NADSYN1	11	71167449	intron-variant	b37	0.022070	0.3402
exm-rs2282679	rs2282679	GC	4	72608383	intron-variant	b37	0.756100	0.3096
exm-rs3829251	rs3829251	NADSYN1	11	71194559	intron-variant	b37	0.315300	0.2135



904 women with 25-hydroxyvitamin D

- 432 controls

- 472 breast cancer cases

Characteristic	(N=432)
Age	56.4 ± 12.1
Autonomous community/study site	
Madrid	106 (25)
Barcelona	32 (7)
Navarra	33 (8)
Guipúzcoa	129 (30)
León	49 (11)
Asturias	34 (8)
Huelva	2 (0)
Cantabria	39 (9)
Valencia	3 (1)
Gerona	5 (1)
Highest level of education	
Primary school or less	54 (12)
General basic education	129 (30)
Secondary school	150 (35)
University	99 (23)
Family history of breast cancer	54 (12)
None	378 (88)
Second grade only	25 (6)
First grade	29 (7)
Body Mass Index (BMI, kg/m^2)	
<18.5	4 (1)
18.5-24.9	224 (52)
25-29.9	147 (34)
≥30	57 (13)
Nulliparous	1 (0)

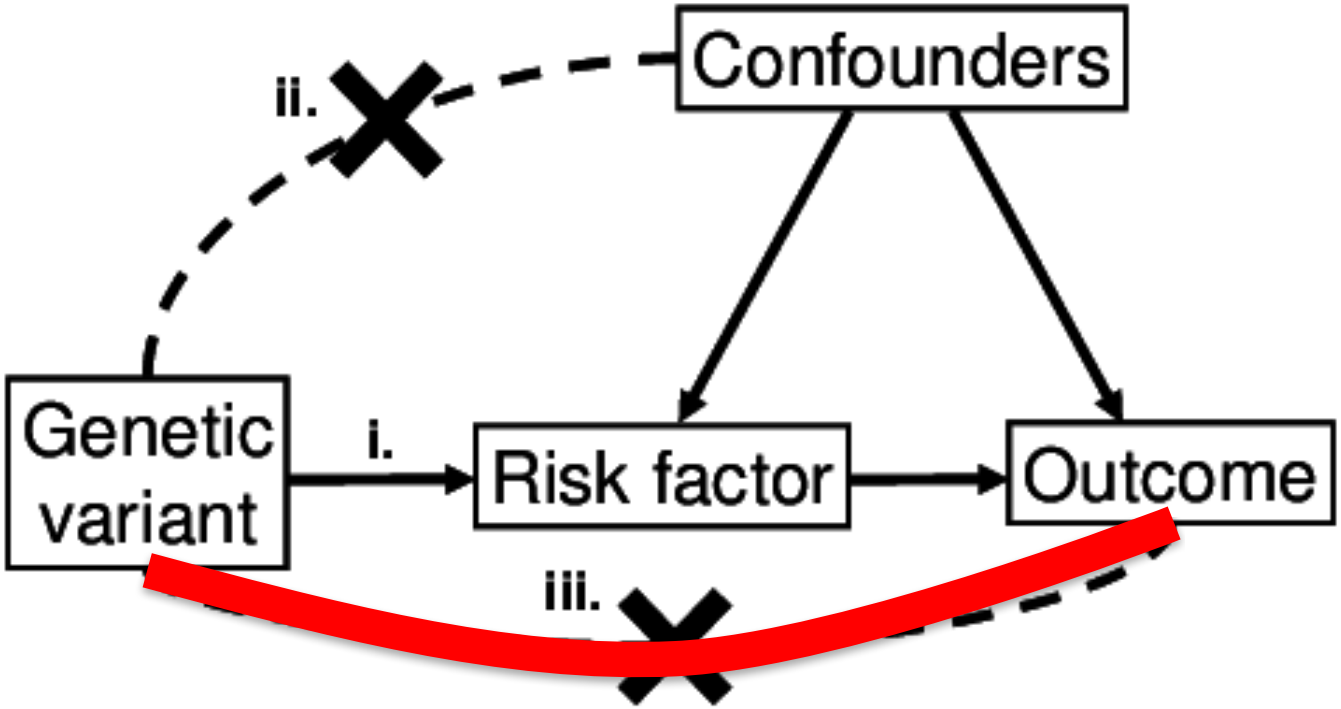
Characteristic	(N=432)
Age at first delivery	
<20 y	13 (4)
20-24 y	95 (28)
25-29 y	133 (39)
>29 y	100 (29)
Postmenopausal	278 (64)
Biopsy	8 (2)
Use of TSH	33 (8)
NA	17 (4)
Recreational Activity in metabolic equivelant* ($kcal*kg^{-1}*h^{-1}$)	13.4 ± 21.6
Vitamin D levels (25(OH)D)	49.9 ± 20.0
Skin color	
Pale white	36 (8)
White	196 (45)
Fair	148 (34)
Brown	52 (12)
Cholesterol	
Normal levels	306 (71)
High, untreated	96 (22)
High, treated	30 (7)
Season	
Spring	157 (36)
Summer	56 (13)
Fall	92 (21)
Winter	127 (29)

Confounders: Latitude, BMI, Skin color, cholesterol , sample day, recreational Activity in metabolic equivelant (kcal*kg^{-1}*h^{-1})

	Estimate	Std. Error	t value	Pr(>|t|)
rs125126311	4.468	1.345	3.322	0.001
rs7051171	-2.299	1.895	-1.213	0.226
rs22826791	-6.670	1.510	-4.417	0.000
rs70411	-6.371	1.317	-4.838	0.000
rs104851651	2.063	1.699	1.215	0.225
rs107416571	3.312	1.382	2.397	0.017
rs127858781	-0.680	1.364	-0.498	0.618
rs38292511	-0.608	1.521	-0.400	0.689
rs47626511	0.435	1.345	0.324	0.746
rs105075771	-2.124	2.264	-0.938	0.349
rs25472311	2.739	1.853	1.478	0.140
rs60138971	0.080	1.654	0.048	0.962

Weighted Risk Score	0.938	0.154	6.094	0.000
Weighted Risk Score/sd	5.457	0.896	6.094	0.000

sd= 5.815915



2214 women

- 1195 controls
- 1019 breast cancer cases

Characteristic	Cases (N=1019)	Controls (N=1195)
Age	57.2 ± 12.4	58.8 ± 13.2
Autonomous community/study site		
Madrid	223 (22)	312 (26)
Barcelona	121 (12)	182 (15)
Navarra	115 (11)	96 (8)
Guipúzcoa	202 (20)	223 (19)
León	186 (18)	108 (9)
Asturias	66 (6)	100 (8)
Cantabria	106 (10)	174 (15)
Highest level of education		
Primary school or less	150 (15)	213 (18)
General basic education	344 (34)	371 (31)
Secondary school	329 (32)	367 (31)
University	196 (19)	244 (20)
Family history of breast cancer		
None	759 (74)	1,021 (85)
Second grade only	112 (11)	79 (6)
First grade	148 (15)	104 (9)
Body Mass Index (BMI, kg/m^2)		
<18.5	15 (1)	23 (2)
18.5-24.9	459 (45)	585 (49)
25-29.9	363 (36)	397 (33)
≥30	182 (18)	190 (16)
Nulliparous	100 (21)	90 (21)
Age at first pregnancy		
<20 y	37 (5)	33 (3)
20-24 y	227 (29)	284 (29)
25-29 y	312 (39)	381 (40)
>29 y	215 (27)	266 (23)
NA	9 (1)	4 (0)

Characteristic	Cases (N=1019)	Controls (N=1195)
Postmenopausal	675 (66)	342 (70)
Biopsy	80 (8)	23 (2)
Use of TSH	78 (8)	93 (8)
NA	28 (3)	42 (4)
Recreational Activity in metabolic equivelant* (kcal*kg^{-1}*h^{-1})	13.8 ± 22.3	14.0 ± 22.3
Vitamin D levels (25(OH)D)	43.2 ± 22.5	49.9 ± 19.8
Skin color		
Pale white	71 (7)	89 (7)
White	458 (45)	521 (44)
Fair	367 (36)	448 (37)
Brown	123 (12)	137 (11)
Cholesterol		
Normal levels	780 (77)	844 (71)
High, untreated	177 (17)	279 (23)
High, treated	62 (6)	72 (6)
Season		
Spring	288 (28)	395 (33)
Summer	210 (21)	168 (14)
Fall	269 (26)	274 (23)
Winter	252 (25)	358 (30)

Confounders: Family history of breast cancer, age, BMI, education level, menopausal status, age at first delivery, use of TSH, biopsia

Random effect: CCAA

	OR	Std. Error	t value	Pr(>|t|)
rs12512631	0.8885831	0.06591437	-1.79212925	0.07311226
rs705117	0.9560836	0.09197829	-0.48826681	0.62536087
rs2282679	1.1815747	0.06963027	2.39619979	0.01656606
rs7041	1.1423412	0.06407866	2.07681975	0.03781820
rs10485165	0.9550389	0.08554330	-0.53777742	0.59073072
rs10741657	0.9961653	0.06546969	-0.05868540	0.95320269
rs12785878	0.9961062	0.06611953	-0.05900468	0.95294838
rs3829251	0.9660456	0.07615066	-0.45363069	0.65009466
rs4762651	0.9890263	0.06573253	-0.16786748	0.86668754
rs10507577	0.9254790	0.10886948	-0.71134618	0.47686974
rs2547231	1.1272692	0.09042239	1.32487183	0.18521367
rs6013897	0.9758167	0.08054846	-0.30392245	0.76118698

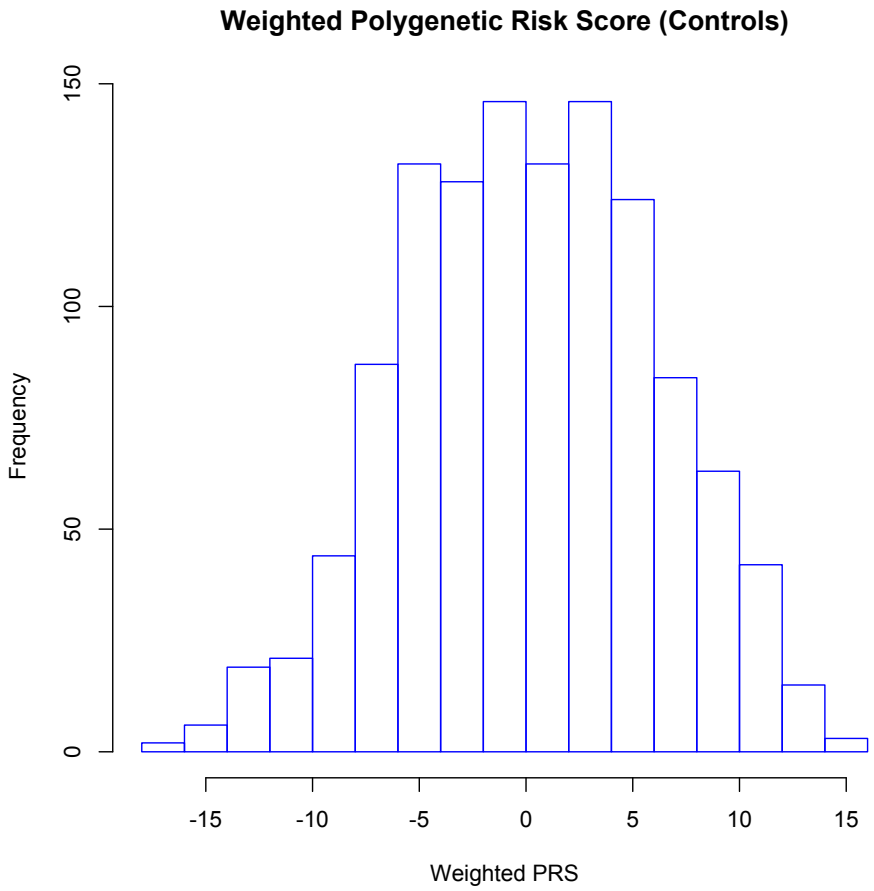
Weighted Risk Score

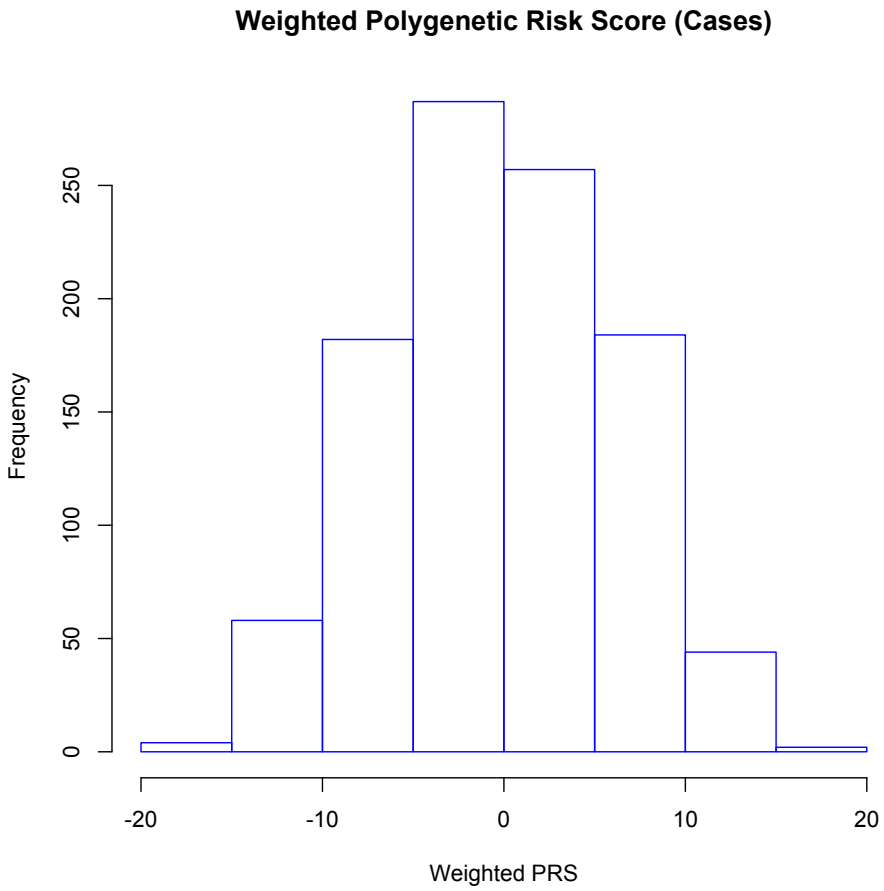
OR	Std. Error	z value	Pr(>|z|)
0.9857108	0.007334785	-1.9622	0.04973926

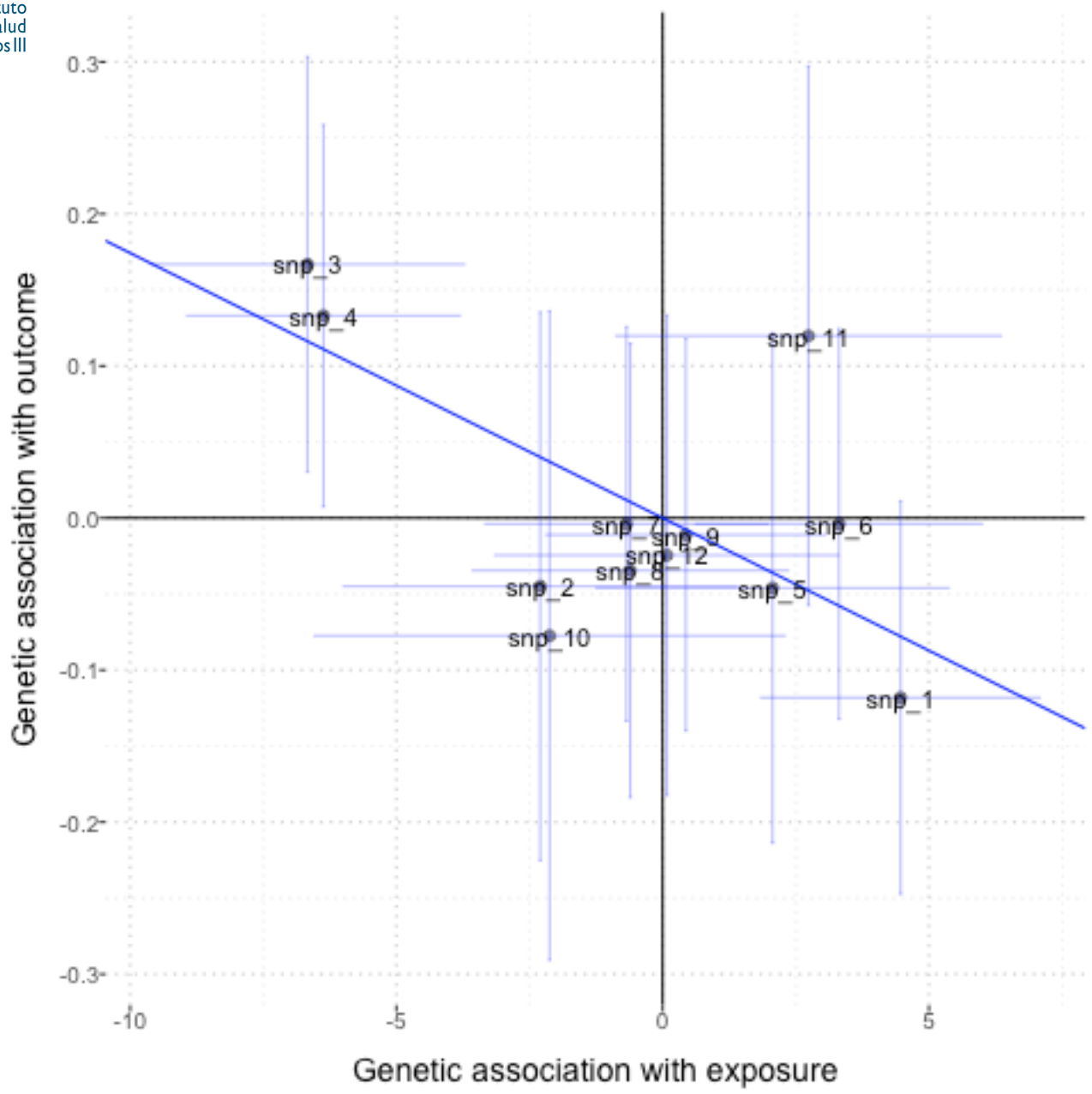
Weighted Risk Score/sd

OR	Std. Error	t value	Pr(>|t|)
0.9162039	0.04460114	-1.9622	0.0497392

sd= 6.080771


Weighted Polygenetic Risk Score (Controls)
Frequency
0
50
100
150
-15
-10
-5
0
5
10
15
Weighted PRS


Weighted Polygenetic Risk Score (Cases)
Frequency
0
50
100
150
200
250
-20
-10
0
10
20
Weighted PRS



Label	SNP
snp_1	rs12512631
snp_2	rs705117
snp_3	rs2282679
snp_4	rs7041
snp_5	rs10485165
snp_6	rs10741657
snp_7	rs12785878
snp_8	rs3829251
snp_9	rs4762651
snp_10	rs10507577
snp_11	rs2547231
snp_12	rs6013897

```
Inverse-variance weighted method
(variants uncorrelated, random-effect model)

Number of Variants : 12

------------------------------------------------------------------
 Method Estimate Std Error  95% CI        p-value
    IVW   -0.017     0.006 -0.029, -0.006   0.003
------------------------------------------------------------------
Residual standard error =  0.831
Residual standard error is set to 1 in calculation of confidence interval when its estimate is less than 1.
Heterogeneity test statistic = 7.5958 on 11 degrees of freedom, (p-value = 0.7490)
```

OR= 0.98 (0.97-0.99)

```
MR-Egger method
(variants uncorrelated, random-effect model)

Number of Variants =  12

------------------------------------------------------------------
      Method Estimate Std Error  95% CI        p-value
    MR-Egger   -0.026     0.009 -0.044, -0.007   0.006
 (intercept)    0.039     0.034 -0.028,  0.107   0.251
------------------------------------------------------------------
Residual Standard Error :  0.792
Residual standard error is set to 1 in calculation of confidence interval when its estimate is less than 1.
Heterogeneity test statistic = 6.2787 on 10 degrees of freedom, (p-value = 0.7913)
I^2_GX statistic: 63.2%
```

OR= 0.97 (0.96-0.99)

Results from Various Mendelian Randomization Methods for Adjusted Models

Method	Estimate	Std Error	95% CI		P-value
Simple median	-0.011	0.014	-0.038	0.016	0.421
Weighted median	-0.022	0.008	-0.038	-0.006	0.007
Penalized weighed median	-0.022	0.008	-0.038	-0.006	0.007
IVW	-0.017	0.006	-0.029	-0.006	0.004
Penalized IVW	-0.017	0.006	-0.029	-0.006	0.004
Robust IVW	-0.019	0.008	-0.034	-0.004	0.004
Penalized robust IVW	-0.019	0.008	-0.034	-0.004	0.004
MR-Egger	-0.026	0.009	-0.044	-0.007	0.006
(intercept)	0.039	0.027	-0.014	0.093	0.148
Penalized MR-Egger	-0.026	0.009	-0.044	-0.007	0.006
(intercept)	0.039	0.027	-0.014	0.093	0.148
Robust MR-Egger	-0.026	0.005	-0.036	-0.016	0.000
(intercept)	0.036	0.022	-0.007	0.079	0.102
Penalized robust MR-Egger	-0.026	0.005	-0.036	-0.016	0.000
(intercept)	0.036	0.022	-0.007	0.079	0.102

Type of Analysis	Heterogeneity	Pleiotropy	Max LD (Theoretical)	Max LD (Empirical)	Population Statifcation
Adjusted Individual SNPs Model	0.749	0.251	<0.8	0.54	Caucasian women only

- Subtypes of breast cancer
- Premenopausal and postmenopausal women, and other possible interactions.
- Analysis with other non-related SNPs
- Other approximations without controls included in the Vitamin D and SNPs study
- Exclude SNPs associated with breast cancer
- Analysis of association with confounders

"*A consortium of leading European research centres and pharmaceutical companies will this week announce a plan to transform epidemiology by combining it with the new techniques of high-throughput biology*".

Butler D. Epidemiology set to get fast-track treatment. Nature. 2001 Nov 8;414(6860):139.


SE HA DESCUBIERTO EL GEN QUE VINCULA A LOS MOTEROS RUIDOSOS CON LA MADRE QUE LOS PARIÓ
LOADO SEA EL CIELO


El cromosoma masculino es pequeño y muy repetitivo
CHORRADAS CHORRADAS CHORRADAS CHORRADAS CHORRADAS
FÚMBOL TOTAS


"CIENTÍFICOS ESPAÑOLES DESCUBREN EN UNA IDEOLOGÍA ANCESTRAL IBÉRICA EL GEN DEL «¿QUE INVENTEN ELLOS?»"
¿SÓLO EN UNA IDEOLOGÍA?
NO; SEGURO QUE CON EL PRESUPUESTO QUE TIENEN NO HAN PODIDO INVESTIGAR A MÁS