



# Cohort profile: the CORDELIA study (*Collaborative cOHORTs Reassembled Data to study mEchanisms and Longterm Incidence of chronic diseAses*)

Álvaro Hernáez · Anna Camps-Vilaró · Sara Polo-Alonso · Isaac Subirana · Rafel Ramos · Rafael de Cid, et al. [full author details at the end of the article]

Received: 10 July 2024 / Accepted: 2 April 2025  
© The Author(s) 2025

## Abstract

The CORDELIA Study (Collaborative Cohorts Reassembled Data to Study Mechanisms and Long-term Incidence of Chronic Diseases) combines 35 Spanish population cohorts to investigate the clinical, environmental, genetic, and omics determinants of cardiovascular disease in the Southern European population. It aims to conduct the largest genome-wide association study to date on cardiovascular disease in this population, improve predictions of cardiovascular incidence using genomic and clinical data, and identify subgroups that would benefit most from targeted pharmacological and lifestyle interventions. CORDELIA includes 196,632 individuals (ages 18–84, 54% female, 96% born in Spain, 20% with higher education, recruited from 1989 to 2020, with follow-up periods ranging from 5 to 30 years), with DNA samples available for 117,342 participants (60%). Of the participants, 24% were current smokers, 43% hypertensive, 11% diabetic, 15% medicated with lipid-lowering drugs, 44% overweight, and 27% obese. If not already available, genotyping is being performed using the Axiom™ Spain Biobank array (~750,000 variants, including 115,000 specific and 50,000 rare functional variants from the Spanish population). The cohort also includes incident events (coronary heart disease, stroke, heart failure, peripheral artery disease, hypertension, diabetes); date and cause of death; and harmonized data on risk factors (body mass index, waist circumference, lipid profile, blood pressure, glucose, creatinine), lifestyle (smoking, physical activity, diet, alcohol), and socioeconomic status. 99,019 participants (50%) also provide plasma samples. CORDELIA will significantly contribute to understanding the complex interplay of risk factors contributing to cardiovascular disease and advance the fields of precision medicine and public health in Southern European individuals.

**Keywords** Cohort · Genome-wide association study · Cardiovascular disease · Mortality · Southern Europe

## Abbreviations

ASCVD	Atherosclerotic cardiovascular disease	Control-MCC-Spain	MultiCase-Control Study-Spain (Control participants)
ACRISC	Auto-cribado del riesgo cardiovascular	CORSAIB	COR SA a les Illes Balears
AEGIS	A Estrada Glycation and Inflammation Study	DIBETES-EUSKADI	Prevalence of diabetes mellitus and impaired glucose metabolism in the adult population of the Basque Country
ARTPER	PERipheral ARTery disease study	DRECE	Dieta y Riesgo de Enfermedad Cardiovascular en España
ASTURIAS	Diabetes y prediabetes en Asturias	EFRCV	Encuesta de Factores de Riesgo Cardiovascular de Murcia
AWHS	Aragón Workers' Health Study	EMMA	Estació de Monitorització de Malalties Arterioscleròtiques
BARCOS	Cohort de dones post-climatèriques de l'àrea de Barcelona	ENRICA	Estudio sobre Nutrición y Riesgo Cardiovascular en España
CARGENCORS	CARDiovascular GENetic risk score for Risk Stratification of patients positive for SARS-CoV-2 (COvid 19) virus	EPIC	European Prospective Investigation into Cancer and Nutrition
CDC de Canarias	Cardiovascular, Diabetes, Cáncer	GCAT	Genomes for Life

HERMEX	Harmonizing Equations of Risk in Mediterranean Countries-Extremadura
HORTEGA	Evaluation of non-traditional risk factors of cardiometabolic and other chronic diseases in a general population from Spain
IBERICAN	<i>Estudio para la identificación de la población española de riesgo cardiovascular y renal</i>
ILERVAS	<i>Estudio de intervención longitudinal prospectivo sobre enfermedades vasculares y renales ocultas</i>
NAVARRA 93	<i>Riesgo Cardiovascular en Navarra</i>
NEFRONA	National Observatory of Atherosclerosis in Nephrology
PIZARRA	Study of diet and health in a population from southern Spain
PREDAPS	<i>Evolución de pacientes con PRE-Diabetes en Atención Primaria de Salud</i>
PREDIMERC	<i>PREvalencia de Diabetes Mellitus y Riesgo Cardiovascular</i>
RECCYL	<i>Riesgo de Enfermedad Cardiovascular en Castilla y León</i>
REGICOR	<i>Registre Gironí del Cor</i>
RICARTO	<i>Riesgo CARDiovascular y eventos cardiovasculares en la población general del área sanitaria de Toledo</i>
RIVANA	<i>Estudio de Riesgo Vascular en Navarra</i>
SALMANTICOR	<i>Estudio clínico-epidemiológico salmantino de las patologías del corazón</i>

## Background

The incidence and mortality rates of atherosclerotic cardiovascular disease (ASCVD) are decreasing worldwide and also in Europe, although the absolute number of cases continue to increase due to the aging of population and it remains the most relevant cause of morbidity and mortality [1]. ASCVD risk can be decreased by a healthy lifestyle (not smoking, a balanced diet, regular physical activity) and the available pharmacological treatments for ASCVD risk factors. However, these approaches are often insufficient, and it is essential to have a good understanding of all environmental, genetic, and socioeconomic factors involved in the development of ASCVD [2, 3]. These factors are locally divergent, which

justifies the creation of large population cohorts in specific regions in the world [4–6]. Very few large cohorts are available in Southern European and Mediterranean populations [7]. In Spain, the recently launched IMPaCT Precision Medicine initiative is the only cohort study of its kind, aiming to recruit approximately 200,000 participants by 2028 and complete a 10-year follow-up period by 2038 [8]. Southern European populations are scarcely represented in genome-wide association studies (GWAS) in individuals of European ancestry [9]. In addition, it remains unknown whether genetic variants related to ASCVD can improve the prediction of the disease in this population. European and American studies indicate that around two-thirds of cardiovascular events occur in individuals classified as low or moderate risk by current predictive models [10, 11], who therefore do not qualify for intensive preventive strategies. Further research is needed to determine if these genetic variants could also identify individuals who would benefit most from specific lifestyle changes and pharmacological interventions to prevent ASCVD [12].

## The cordelia study: rationale and aims

In this context, the general aim of the CORDELIA Study (*Collaborative cOHORTs Reassembled Data to study mEchanisms and Longterm Incidence of chronic diseAses*) is to investigate a comprehensive range of risk factors for ASCVD, including environmental, socioeconomic, clinical, genetic, and other omics data, using information from 35 existing Spanish cohorts. In particular, CORDELIA aims to conduct the largest GWAS on ASCVD in the Southern European population to date, aiming to identify the genetic variants linked to the occurrence of ASCVD in these individuals. It also aims to create polygenic risk scores for ASCVD adapted to the characteristics of the Spanish population, and to assess whether their inclusion in predictive functions improves their capacity to predict the 10-year incidence of ASCVD. CORDELIA also aims to identify subgroups of the population according to their genetic predisposition to disease that would benefit the most from the available pharmacological treatments and lifestyle modifications. All the previous results will be stratified by sex, highlighting the relevance of sex-specific differences in ASCVD, which can lead to more precise and effective prevention strategies.

## Methods

### Study population

The CORDELIA Study includes 196,632 individuals, aged 18 to 84 years, recruited in Spain from 1989 to 2020, who gave their informed consent in their respective cohorts, and

a completed follow-up on cardiovascular events between 5 and 30 years. DNA samples are available in 117,342 of the participants (60%). Serum or plasma samples are available in 99,019 of the participants (50%). The general characteristics of the 35 cohorts included in the CORDELIA Study are available in Table 1.

## Genotype data

Genotype information for five participating cohorts has already been provided. Specifically, AWHs, Control-MCC-Spain, and HORTEGA cohorts have utilized the Infinium™ Global Screening Array-24 BeadChip platform, with their genotypes aligned to the GRCh38/hg38 reference genome and imputed using the TOPMed imputation panel with an r2 quality metric. Additionally, the NEFRONA and CARGEN-CORS cohorts have employed the Axiom™ Spain Biobank Array-1, with their genotypes referenced to the GRCh37/hg37 genome and imputed using the TOPMed imputation panel with an r3 quality metric.

For the rest of the cohorts, obtaining genotype information is underway. DNA extraction in the CORDELIA study samples was performed using a liquid–liquid method with the FlexiGene DNA Kit reagent from QIAGEN (QIAGEN, Valencia, California, USA). The performance of the extraction, and the quality of the obtained DNA samples were assessed using an Infinite M200 spectrophotometer (Tecan Group Ltd., Männedorf, Switzerland). DNA was quantified with the NanoQuant absorbance method. DNA integrity was also evaluated using 1% agarose gels and quantified by Picogreen (fluorescence) prior to genotyping. During the genotyping process, the samples were quantified with spectrophotometry to verify appropriate amplification. The fragmentation step was monitored with 4% agarose gels. RECCYL was the only cohort that provided isolated DNA samples with verified quality control (performed in the DNA National Bank, University of Salamanca, Salamanca, Spain).

DNA samples are being genotyped at the Fundación Pública Galega de Medicina Xenómica (Santiago de Compostela, Spain), using the Axiom™ Spain Biobank Array-2 (ThermoFisher Scientific, Cornellà de Llobregat, Spain) [45]. This array contains 757,836 variants selected to provide functional genomic coverage, including 114,898 specific variants and 50,536 rare functional variants from the Spanish population. Total genomic DNA (200 ng) is amplified and fragmented to 25–125 base pair oligomers. These fragments are purified and transferred to the GeneTitan Multichannel Instrument for automated hybridization, staining, washing, and scanning. Genotyping calling and clustering is performed with the software Axiom Analysis Suite v5.3.0.45.

Regarding quality control: (1) probes showing genotyping problems (e.g., poor cluster separation) are not included; (2) markers with minor allele frequency < 1%, genotyping

rate < 98%, and deviating from Hardy–Weinberg equilibrium ( $p$ -value <  $10^{-6}$  with a mid- $p$  adjustment) are removed [46]; and (3) samples with < 98% genotyping rate and a heterozygosity rate that deviates more than 5 standard deviations from the mean heterozygosity rate of the study are also excluded. A subgroup of > 100,000 independent autosomal SNPs (minor allele frequency > 5% excluding regions with high linkage disequilibrium [47], pruned with a window size of 1000 markers, step size of 80, and  $r^2$  of 0.1) will be used to detect kinship (by identity-by-descent estimation, removing individuals at first and second degree of kinship) and population stratification. After removing related individuals and outliers, the first 10 ancestry-informative genetic principal components will be calculated, to be used as covariates in the association analyses. Unrelated individuals will be analyzed in the Admixture software to assess their ancestry using independent markers overlapping the 1000 Genomes Project superpopulations [48, 49]. Individuals will be assigned to a specific ancestry if their probability of being part of that group is  $\geq 80\%$ . Otherwise, they will be classified as “mixed” individuals. Finally, after quality control, imputation of genetic variants will be performed using the TOPMed r3 reference panel (GRCh38) [50]. The post-imputation filtering applied will be an INFO score  $\geq 0.8$  and minor allele frequency > 1%.

## Study variables: harmonization

### Administrative variables

The components cohorts’ questionnaires were based on standardized surveys by the World Health Organization [51]. The following basic administrative variables are available for all cohorts: center code (a consecutive number starting at 01 for each cohort), cohort name (the acronym recognized and published by the cohort investigators), participant number (pseudonymized identifiers used by the cohorts), date of inclusion (the date each participant signed the informed consent and was first examined), date/year of birth, age at inclusion (self-reported or calculated as the difference between date of birth and inclusion), and sex (female or male phenotype, collected by the cohort personnel). Some cohorts also collected information on country of origin (self-reported), civil status (harmonized to single, married/with a partner, divorced/separated, widow, or not available), and city/town/village of residence.

### Clinical outcomes

Regarding the clinical outcomes of the study, there is a diagnosis recorded in a medical record or a medical report indicating that the participant has developed the condition plus the diagnosis date. The cardiometabolic events identified in the follow-up of the cohorts participating in

**Table 1** Characteristics of the cohorts included in the CORDELIA Study (population aged 18–84 years)

Cohort	Region	Sample size		Recruited	Response rate (%)	Sampling	Sampling procedure	References
		Total	DNA available					
Registre Gironí del Cor (REGICOR)	Catalonia	11,632	10,980	1995–2018	72	Population-based	Two-stage cluster: age, sex	[13]
Auto-cribado del riesgo cardiovascular (ACRISC)	Catalonia	743	728	2015–2017	60	Population-based	Random	[14]
Cohort de dones post-climatèriques de l'àrea de Barcelona (BARCOS)	Catalonia	473	473	1989–2010	100	Population-based	Unselected	[15]
Aragón Workers' Health Study (AWHS)	Aragon	5,670	5,670	2009–2010	90	Workers	Unselected, consecutive inclusion of volunteers	[16]
Estudio clínico-epidemiológico salmantino de las patologías del corazón (SALMANTICOR)	Castilla y León	1,993	1,993	2015–2018	58	Population-based	Three-stage cluster: age, sex, residence	[17]
Prevalence of diabetes and impaired glucose regulation (DI@BET.ES)	Spain	5,419	4,687	2007–2010	56	Population-based	Random. Cluster: town	[18]
Study of diet and health in a population from southern Spain (PIZARRA)	Andalusia	2,089	1,422	2007–2012	75	Population-based	Random	[19]
Evolución de pacientes con PREDiabetes en Atención Primaria de Salud (PRE-DAPS)	Spain	2,022	341	2012	70	Population-based	Unselected, consecutive inclusion of patients	[20]
Dieta y Riesgo de Enfermedad Cardiovascular en España (DRECE)	Spain	3,465	-	1991	96	Population-based	Two-stage cluster: age, sex	[21]
Estudio sobre Nutrición y Riesgo Cardiovascular (ENRICA)	Spain	13,018	-	2008–2010	51	Population-based	Three-stage cluster: province, municipality,	[22]
European Prospective Investigation into Cancer and Nutrition (EPIC): EPIC-Granada, EPIC-Gipuzkoa, EPIC-Murcia, EPIC-Navarra	Andalusia Basque Country Murcia Navarra	7,864 8,417 8,515 7,953	6,448 8,300 7,920 7,806	1992–1996	88	Population-based	Healthy volunteers	[23]
Encuesta de Factores de Riesgo Cardiovascular de Murcia (EFRVCV)	Murcia	3,089	-	1992–1993	61	Population-based	Multi-stage random within clusters: residence, age, sex	[24]

**Table 1** (continued)

Cohort	Region	Sample size		Recruited	Response rate (%)	Sampling	Sampling procedure	References
		Total	DNA available					
MultiCase-Control Study-Spain (Control-MCC-Spain)	Spain	4,077	2,823	2008–2013	53	Population-based	Population-based randomly selected controls, frequency matched to cancer cases by age, sex, and region	[25]
COR SA a les Illes Balears (CORSAIB)	Balearic Islands	1,685	-	1999–2000	92	Population-based	Random in clusters: region	[26]
Estació de Monitorització de Malalties Arterioscleròtiques (EMMA)	Catalonia	35,175	-	1997–2002	100	Population-based	Random	[27]
PERIPHERAL ARTery disease study (ARTPER)	Catalonia	3,736	1,809	2006–2008	100	Population-based	Random in clusters: primary care center	[28]
National Observatory of Atherosclerosis in Nephrology (NEFRONA)	Spain	3,004	3,000	2009–2012	90	Hospital-based	Cluster: kidney disease stage	[29]
Estudio de intervención longitudinal prospectivo sobre enfermedades vasculares y renales ocultas (ILERVAS)	Catalonia	8,330	7,229	2015–2018	83	Population-based	Cluster: age, primary care center	[30]
Diabetes y prediabetes en Asturias (ASTURIAS)	Asturias	1,034	232	1998–1999	64	Population-based	Two-stage cluster: region, age	[31]
Cardiovascular, Diabetes, Cáncer (CDC de Canarias)	Canary Islands	7,160	7,081	2000–2005	70	Population-based	Two-stage cluster: island, region	[32]
Harmonizing Equations of Risk in Mediterranean Countries-Extremadura (HERMEX)	Extremadura	2,833	-	2007–2009	81	Population-based	Random, stratified by age and sex	[33]
Evaluation of non-traditional risk factors of cardiometabolic and other chronic diseases in a general population from Spain (HORTEGA)	Castilla y León	1,475	1,475	1999–2000	62	Population-based	Random, stratified by age and sex	[34]
Riesgo Cardiovascular en Navarra (NAVARRA 93)	Navarra	1,480	-	1993	81	Population-based	Three-stage cluster: age, sex, region	[35]
Estudio de Riesgo Vascular en Navarra (RIVANA)	Navarra	4,164	892	2004–2005	74	Population-based	Random: stratified by age and sex	[36]
Riesgo de Enfermedad Cardiovascular en Castilla y León (RECCYL)	Castilla y León	3,761	3,761	2004	87	Population-based	Two-stage stratified random sampling (region, rurality)	[37]

**Table 1** (continued)

Cohort	Region	Sample size		Recruited	Response rate (%)	Sampling	Sampling procedure	References
		Total	DNA available					
PREvalencia de Diabetes Mellitus y Riesgo Cardiovascular (PREDIMERC)	Madrid Autonomous Region	2,268	649	2,000	2007	56	Population-based	Two-stage stratified random sampling (socioeconomic status, primary care center) [38]
Estudio para la identificación de la población española de riesgo cardiovascular y renal (IBERICAN)	Spain	8,035	6,008	-	2014–2018	87	Population-based	Unselected, consecutive in primary care centers [39]
A Estrada Glycation and Inflammation Study (AEGIS)	Galicia	1,486	1,315	-	2012–2015	68	Population-based	Random in cluster: age, sex, neighborhood [40]
Prevalence of diabetes mellitus and impaired glucose metabolism in the adult population of the Basque Country (DI@BETES-Euskadi)	Basque Country	567	567	-	2010–2012	68	Population-based	Random in cluster: primary care center [41]
CARDiovascular GENetic risk score for Risk Stratification of patients positive for SARS-CoV-2 (Covid 19) virus (CARGEN-CORS)	Catalonia	2,838	2,575	-	2020	80	Population-based	Consecutive: hospital and primary care COVID-19 triage [42]
Genomes for Life (GCAT)	Catalonia	19,187	19,183	19,183	2014–2017	NA	Population-based	Unselected, volunteers and blood donors [43]
Riesgo CARDiovascular y eventos cardiovasculares en la población general del área sanitaria de Toledo (RICARTO)	Castilla La Mancha	1,975	1,975	-	2011–2020	32	Population-based	Random: stratified by age, sex, and region [44]
<b>All cohorts</b>		196,632	117,342	99,019	1989–2020	81		



the meta-cohort include: fatal and non-fatal acute myocardial infarction, exertional or unstable angina, fatal and non-fatal stroke, heart failure, peripheral artery disease, chronic kidney disease, diabetes and hypertension in initially non-diabetic or non-hypertensive participants. The definition of these non-fatal events in the follow-up is performed by clinical committees of the individual cohorts (following the criteria agreed by the Committee for the Harmonization of Outcomes and Variables of the CORDELIA Study). The following International Classification of Diseases (ICD) codes are being considered: coronary heart disease (ICD-9: 410–414; ICD-10: I20–I25); cerebrovascular disease (ICD-9: 430–437.0; ICD-10: I60–I64); heart failure (ICD-9: 402, 404, 428; ICD-10: I11, I13.0, I13.2, I50), and peripheral artery disease (ICD-9: 440.2; ICD-10: I70.2, I70.3, I70.92, I73.9, E10.5, E11.5). Incident cases of hypertension and type-2 diabetes are also being collected. The CORDELIA infrastructure enables researchers to propose the future collection of data on additional fatal and non-fatal chronic conditions. Fatal events, as well as the primary cause of death for deceased participants, are verified against the official databases of the Spanish National Institute of Statistics' Mortality Registry and the Spanish hospital discharges dataset ("*Conjunto Mínimo Básico de Datos*"). The most recent vital status (alive, dead, or unknown) of all cohort participants is available.

The previous history of cardiovascular morbidities of the study participants is also available. Following the standard questionnaires set by the World Health Organization, participants were asked whether their doctors had informed them that they suffered from any of the following: heart attack/myocardial infarction, stroke/brain hemorrhage/brain embolism/thrombosis, angina pectoris, arterial hypertension/high blood pressure, hypercholesterolemia/high blood cholesterol levels, or diabetes. They were also queried about any doctor-prescribed drug treatments they were taking to manage blood pressure, diabetes, or cholesterol.

Given that the definitions of hypertension and diabetes have varied since 1989 (the recruitment year in the oldest cohort), we developed a harmonized algorithm for history of hypertension and diabetes at baseline in all the participating cohorts based on the history of morbidities reported by the participants, some clinical examinations (blood pressure, fasting glucose; see following paragraph), and drug use information (see Supplemental Tables 1 and 2, respectively). By doing so, we maximize the number of participants with valid information on their history of hypertension and diabetes as a covariate for subsequent analyses and ensure these outcomes are defined according to current clinical guidelines.

## Clinical measurements and laboratory determinations

The clinical examinations at the baseline visit include measurements of weight (kg), height (cm), body mass index (BMI, kg/m<sup>2</sup>), waist circumference (cm), heart rate (beats/minute, a single measurement or the average of 2–3 measurements), systolic blood pressure (SBP, mmHg; one or two readings, retaining the lowest), and diastolic blood pressure (DBP, mmHg; one or two readings, retaining the lowest). Participants were categorized according to their BMI values as underweight (< 18.5 kg/m<sup>2</sup>), normal weight (18.5 to < 25.0 kg/m<sup>2</sup>), overweight (25.0 to < 30.0 kg/m<sup>2</sup>), obese (30.0 to < 40.0 kg/m<sup>2</sup>), and morbidly obese (≥ 40 kg/m<sup>2</sup>). Abdominal obesity was defined as a waist circumference of ≥ 89 cm (female) and ≥ 103 cm (male).

Laboratory variables were measured in blood samples taken after an overnight fast, performed in reference laboratories on fresh plasma/serum aliquots stored at −80°C until use, in samples not previously thawed. Triglycerides, glucose, total cholesterol, and high-density lipoprotein (HDL) cholesterol were measured using enzymatic methods, and creatinine using the Jaffe method. All local laboratories met external quality control standards [13, 26, 28, 32, 33, 36, 52–55]. Lipid profile and glucose measurements in 8 of the cohorts (ARTPER, CDC, CORSAIB, HERMEX, PREDIMERC, RECCYL, REGICOR, RIVANA) were validated for concordance with a central laboratory [56]. The values for triglycerides, glucose, creatinine, total cholesterol, and HDL cholesterol are expressed in mg/dL. Low-density lipoprotein (LDL) cholesterol was calculated using the Friedewald formula (LDL cholesterol = total cholesterol – HDL cholesterol – triglycerides/5) when triglycerides were < 300 mg/dL or otherwise directly determined. When lipid profile and glucose values were originally in mmol/L, they were converted to mg/dL by multiplying by 88.57 for triglycerides, 18.016 for glucose, and 38.67 for total, HDL, and LDL cholesterol. When creatinine was provided in μmol/L, it was converted to mg/dL by multiplying by 0.0113. The estimated glomerular filtration rate for participants with creatinine values (in mL/min/1.73 m<sup>2</sup>) was calculated using the Chronic Kidney Disease Epidemiology Collaboration formula [57]. After harmonization, we defined the following valid data ranges for continuous cardiovascular risk factors: 15–50 kg/m<sup>2</sup> for BMI, 60–150 cm for waist circumference, 40–250 mmHg for SBP, 20–130 mmHg for DBP, 30–1000 mg/dL for triglycerides, 30–350 mg/dL for glucose, 100–450 mg/dL for total cholesterol, 20–130 mg/dL for HDL cholesterol, and 0.2–5.0 mg/dL for creatinine.

## Socioeconomic and lifestyle variables

Data on these variables was available from the baseline visit of each cohort. Information on educational attainment

**Table 2** Population characteristics of the CORDELIA aggregated cohorts

	All participants ( <i>n</i> = 196,632)	Female ( <i>n</i> = 105,513)	Male ( <i>n</i> = 91,119)	Valid data ( <i>n</i> , %)
Age, years (mean $\pm$ SD)	52.5 $\pm$ 13.2	52.2 $\pm$ 13.3	52.8 $\pm$ 13.1	196,632 (100.0)
Female sex ( <i>n</i> , %)	105,513 (53.7%)	105,513 (100%)	0 (0.00%)	196,632 (100.0)
Born in Spain ( <i>n</i> , %)	77,291 (95.5%)	42,416 (95.0%)	34,875 (96.1%)	80,825 (41.1)
Civil status:				102,474 (52.1)
Single ( <i>n</i> , %)	13,653 (13.3%)	7,623 (12.9%)	6,029 (13.8%)	
Married/partnered ( <i>n</i> , %)	78,133 (76.2%)	43,197 (73.4%)	34,936 (80.1%)	
Divorced/separated ( <i>n</i> , %)	5,600 (5.46%)	3,728 (6.33%)	1,872 (4.29%)	
Widow ( <i>n</i> , %)	5,088 (4.97%)	4,326 (7.35%)	762 (1.75%)	
Educational attainment:				131,186 (66.7)
None/functionally illiterate ( <i>n</i> , %)	20,788 (15.8%)	13,865 (18.8%)	6,923 (12.1%)	
Elementary/primary school ( <i>n</i> , %)	47,311 (36.1%)	26,824 (36.3%)	20,487 (35.7%)	
High/secondary school ( <i>n</i> , %)	37,392 (28.5%)	19,402 (26.3%)	17,990 (31.4%)	
University studies ( <i>n</i> , %)	25,695 (19.6%)	13,755 (18.6%)	11,940 (20.8%)	
Tobacco use:				193,457 (98.4)
Never smoker ( <i>n</i> , %)	101,078 (52.2%)	68,902 (66.4%)	32,176 (35.9%)	
Current smoker/ex-smoker for < 1 year ( <i>n</i> , %)	47,285 (24.4%)	19,588 (18.9%)	27,697 (30.9%)	
Ex-smoker for $\geq$ 1 year ( <i>n</i> , %)	45,094 (23.3%)	15,242 (14.7%)	29,852 (33.3%)	
Participants with hypertension ( <i>n</i> , %)	84,085 (42.8%)	40,347 (38.3%)	43,738 (48.0%)	196,518 (99.9)
Systolic blood pressure (minimum), mmHg (mean $\pm$ SD)	129 $\pm$ 19.6	126 $\pm$ 20.4	132 $\pm$ 18.0	164,895 (83.9)
Diastolic blood pressure (minimum), mmHg (mean $\pm$ SD)	77.8 $\pm$ 10.5	76.1 $\pm$ 10.4	79.7 $\pm$ 10.3	164,895 (83.9)
Heart rate, beats/min (mean $\pm$ SD)	71.1 $\pm$ 12.0	72.5 $\pm$ 11.9	69.6 $\pm$ 12.1	96,342 (49.0)
Participants with diabetes ( <i>n</i> , %)	21,644 (11.0%)	9,654 (9.18%)	11,990 (13.2%)	196,040 (99.7)
Glucose, mg/dL (mean $\pm$ SD)	99.2 $\pm$ 27.3	95.9 $\pm$ 24.4	103 $\pm$ 29.7	138,870 (70.6)
Total cholesterol, mg/dL (mean $\pm$ SD)	210 $\pm$ 41.3	212 $\pm$ 41.2	207 $\pm$ 41.2	155,414 (79.0)
HDL cholesterol, mg/dL (mean $\pm$ SD)	54.7 $\pm$ 14.6	59.6 $\pm$ 14.7	49.5 $\pm$ 12.6	147,016 (74.8)
Participants medicated with cholesterol-lowering drugs ( <i>n</i> , %)	21,919 (15.5%)	10,600 (14.4%)	11,319 (16.5%)	141,869 (72.1)
Triglycerides, mg/dL (median [1 st-3rd quartile])	98.3 [72.0–139]	89.5 [66.0–124]	110 [80.0–155]	148,351 (75.4)
Body mass index, kg/m <sup>2</sup> (median [1 st-3rd quartile])	27.3 [24.6–30.4]	26.9 [23.8–30.5]	27.7 [25.3–30.2]	193,836 (98.6)
Categories for body mass index:				193,836 (98.6)
< 18.5 kg/m <sup>2</sup> ( <i>n</i> , %)	1,093 (0.56%)	831 (0.80%)	262 (0.29%)	
18.5–24.9 kg/m <sup>2</sup> ( <i>n</i> , %)	54,794 (28.3%)	35,179 (33.9%)	19,615 (21.8%)	
25.0–29.9 kg/m <sup>2</sup> ( <i>n</i> , %)	84,920 (43.8%)	38,686 (37.3%)	46,234 (51.4%)	
30.0–39.9 kg/m <sup>2</sup> ( <i>n</i> , %)	50,268 (25.9%)	27,075 (26.1%)	23,193 (25.8%)	
$\geq$ 40.0 kg/m <sup>2</sup> ( <i>n</i> , %)	2,761 (1.42%)	2,044 (1.97%)	717 (0.80%)	
Waist circumference, cm (mean $\pm$ SD)	96.1 $\pm$ 23.5	91.5 $\pm$ 22.5	102 $\pm$ 23.5	140,541 (71.5)
Participants with abdominal obesity ( <i>n</i> , %)	76,433 (54.4%)	36,413 (48.0%)	40,020 (61.9%)	140,541 (71.5)
Creatinine, mg/dL (mean $\pm$ SD)	0.87 $\pm$ 0.30	0.76 $\pm$ 0.25	0.98 $\pm$ 0.32	91,282 (46.4)
Glomerular filtration rate, mL/min/1.73 m <sup>2</sup> (mean $\pm$ SD)	90.8 $\pm$ 18.2	91.3 $\pm$ 18.3	90.3 $\pm$ 18.1	91,282 (46.4)

has been harmonized into the following categories: none/functionally illiterate (illiterate, no studies, or only can read and write), elementary/primary school (compulsory basic education), high/secondary school (including “*Bachillerato Unificado Polivalente*”, “*Curso de Orientación Universitaria*”, vocational training, “*Bachillerato*”, and university access courses), university studies, and not available. Age at completion of last studies is also available for some cohorts. Smoking status has been harmonized as follows:

current smoker/ex-smoker of less than one year (anyone who smoked  $\geq$  1 cigarette/day was considered a smoker), ex-smoker of one year or more, never smoker, not available. Additionally, some cohorts provided details on the number of cigarettes per day participants currently smoke or smoked in the past (for smokers and ex-smokers), the age at which participants began smoking (for smokers and ex-smokers), and the age at which ex-smokers quit smoking. Leisure-time physical activity estimates for the participants



were harmonized to average kcal/day whenever possible. When data were reported in metabolic equivalents of task per minute per day, these values were converted to kcal/day by multiplying those values by the participant's weight in kilograms. Dietary data, whenever possible, were used to construct a Mediterranean diet adherence score, ranging from 1 to 14, with higher scores indicating greater adherence. The diet score positively rated the intake of virgin olive oil, fruits, vegetables, "sofrito" (sauce or base made from sautéed vegetables such as tomatoes, onions, garlic, and peppers), legumes, mixed nuts, fish, poultry, and moderate doses of wine; and it negatively rated the intake of red and processed meats, butter/cream, sugar-sweetened beverages, and non-homemade sweets/pastries [58].

Additional details on the collected variables (whether tobacco use, weight, height, physical activity, and diet information was collected by trained personnel or self-reported; technical details regarding biological samples; and the detailed methodology and equipment used to measure blood pressure and laboratory determinations) is available in Supplemental Table 3.

All cohorts' data have been harmonized and integrated into a single, curated database hosted at the Hospital del Mar Research Institute in Barcelona. Whenever between-cohort heterogeneity is substantial, phenotype data analyses will be conducted in the pooled CORDELIA database using mixed-effects models (considering the cohort as a random effect variable), and genotype analyses will employ a meta-analysis framework. According to the consortium's by-laws, CORDELIA investigators will access the integrated dataset until the end of 2026. Thereafter, any researcher may request access by submitting a standardized application form available at [www.cordeliaproject.net](http://www.cordeliaproject.net); each request will undergo scientific and ethical review to ensure proper handling of participant data.

### GWAS sample size

Accepting an alpha risk of  $5 \times 10^{-8}$  (the standard genome-wide significance threshold), a statistical power of 80% in a two-sided test, and a 10-year incidence rate of coronary events of 3.7% in the group of participants not exposed to the risk allele, 117,342 individuals allow the identification of the following relative risks depending on the risk allele frequency (AF):  $\geq 1.84$  (AF 1%),  $\geq 1.38$  (AF 5%),  $\geq 1.28$  (AF 10%),  $\geq 1.24$  (AF 15%),  $\geq 1.21$  (AF 20%),  $\geq 1.20$  (AF 25%),  $\geq 1.19$  (AF 30%),  $\geq 1.18$  (AF 40%), and  $\geq 1.17$  (AF 50%). In the sex-stratified analyses (54% of the participants are female), under the same assumptions, our population allows the identification of the following relative risks:  $\geq 2.15$  (AF 1%),  $\geq 1.52$  (AF 5%),  $\geq 1.38$  (AF 10%),  $\geq 1.32$  (AF 15%),  $\geq 1.29$  (AF 20%),  $\geq 1.27$  (AF 25%),  $\geq 1.26$  (AF 30%),  $\geq 1.24$  (AF 40%), and  $\geq 1.24$  (AF 50%).

## Results

### Characteristics of the study population

As described in Table 2, participants in the CORDELIA study were, on average, 52 years old, with female participants comprising 54% of the group. According to cohorts that provided this information, the majority were born in Spain (96%). 20% had attained a university degree. Smoking status showed that 52% had never smoked, and 24% were current smokers. Cardiovascular and metabolic conditions were prevalent, with 43% having hypertension, 11% diagnosed with diabetes, and 15% taking cholesterol-lowering drugs. The average triglyceride level was 98 mg/dL. The percentages of participants who were overweight, obese, and morbidly obese were 44%, 26%, and 1.4%, respectively. Abdominal obesity was present in 54% of the participants. The mean participants' estimated glomerular filtration rate was 90.8 mL/min/1.73 m<sup>2</sup>. A description of the variables in each of the CORDELIA cohorts is available in Table 3. The proportion of participants with valid data for the study variables in each of the individual cohorts is described in Supplemental Table 4.

## Discussion

### Findings to date

This cohort profile is the first joint publication of the CORDELIA Study. However, as evidenced by several high-impact publications, its participating cohorts have substantially contributed to the epidemiological understanding of various health outcomes, including acute myocardial infarction [59, 60], stroke [61], peripheral artery disease [62], subclinical atherosclerosis [63], cancer [64–67], neurodegenerative disease [68, 69], diabetes [18], obesity [70, 71], disability [72], sleep-related diseases [73], and COVID-19 [42, 74].

### Strengths and limitations

The CORDELIA Study has several strengths. It is, to date, the largest meta-cohort on cardiovascular disease of South European population with almost 200,000 participants, maximizing the statistical power of the individual cohorts. To our knowledge, it is also one of the largest meta-cohorts combining a comprehensive set of clinical, socioeconomic, and lifestyle variables, DNA samples (available in 117,342 participants, 60%) and serum/plasma samples (available in 99,019 participants, 50%) in European population. CORDELIA represents an exhaustively defined population, aiming

**Table 3** Characteristics of the participants of individual cohorts in CORDELIA

Sample size	Age, years (mean $\pm$ SD)	Female (%)	Born in Spain (%)	%Univ. Education (%)	Current smokers (%)	HT (%)	SBP, mmHg (mean $\pm$ SD)	DBP, mmHg (mean $\pm$ SD)	HR, beat/min (mean $\pm$ SD)	Diabetes (%)	Glucose, mg/dL (mean $\pm$ SD)	TC, mg/dL (mean $\pm$ SD)	HDL-C, mg/dL (mean $\pm$ SD)	LDL-C, mg/dL (mean $\pm$ SD)	Trig., mg/dL median [1 st-3rd quartile]	Chol.-Lowering treat. (%)	BMI, kg/m <sup>2</sup> (mean $\pm$ SD)	Creat. mg/dL (mean $\pm$ SD)
REGICOR	11,632	54 $\pm$ 13.52	97	18	24	42	127 $\pm$ 20.78 $\pm$ 11	69 $\pm$ 11.14	101 $\pm$ 25 $\pm$ 42	215 $\pm$ 52 $\pm$ 14	141.93 $\pm$ 38	10	27 $\pm$ 5	0.9 $\pm$ 0.2				
ACRISC	743	48 $\pm$ 10.51	-	44	24	22	109 $\pm$ 17.73 $\pm$ 11	68 $\pm$ 11.6	93 $\pm$ 19 $\pm$ 34	197 $\pm$ 57 $\pm$ 15	118.96 $\pm$ 30	8	26 $\pm$ 4	-				
BARCOS	473	58 $\pm$ 8.100	76	-	14	20	129 $\pm$ 17.77 $\pm$ 10 $\pm$ 35	122.4	97 $\pm$ 17 $\pm$ 38	223 $\pm$ 63 $\pm$ 16	138.91 $\pm$ 36	3	26 $\pm$ 4	0.8 $\pm$ 0.2				
AWHS	5,670	47 $\pm$ 9.7	-	-	37	34	125 $\pm$ 15.83 $\pm$ 10	71 $\pm$ 12.5	94 $\pm$ 19 $\pm$ 37	208 $\pm$ 52 $\pm$ 11	128.120 $\pm$ 31	8	27 $\pm$ 4	1.0 $\pm$ 0.1				
SALMAN-TICOR	1,993	55 $\pm$ 16.54	96	24	20	60	140 $\pm$ 19.87 $\pm$ 11	68 $\pm$ 11.12	95 $\pm$ 27 $\pm$ 33	183 $\pm$ 54 $\pm$ 14	106.97 $\pm$ 29	25	26 $\pm$ 4	0.8 $\pm$ 0.2				
DI@BETES	5,419	50 $\pm$ 17.58	91	16	26	44	129 $\pm$ 20.75 $\pm$ 11	- 12	98 $\pm$ 26 $\pm$ 39	197 $\pm$ 52 $\pm$ 13	105.102 $\pm$ 30	12	28 $\pm$ 5	0.8 $\pm$ 0.2				
PIZARRA	2,089	51 $\pm$ 10.61	99	7	26	45	133 $\pm$ 21.78 $\pm$ 11	72 $\pm$ 11.14	103 $\pm$ 24 $\pm$ 40	207 $\pm$ 55 $\pm$ 15	128.103 $\pm$ 37	12	29 $\pm$ 5	0.8 $\pm$ 0.2				
PREDAPS	2,022	58 $\pm$ 10.51	95	17	17	81	130 $\pm$ 16.79 $\pm$ 10	73 $\pm$ 10.0	98 $\pm$ 13 $\pm$ 37	210 $\pm$ 56 $\pm$ 15	129.110 $\pm$ 33	100	29 $\pm$ 5	0.8 $\pm$ 0.2				
DRECE	3,465	37 $\pm$ 12.52	-	12	43	22	120 $\pm$ 17.75 $\pm$ 12	- 6	95 $\pm$ 22 $\pm$ 44	201 $\pm$ 54 $\pm$ 14	124.94 $\pm$ 38	5	26 $\pm$ 4	1.0 $\pm$ 0.2				
ENRICA	13,018	47 $\pm$ 17.53	94	29	27	36	127 $\pm$ 18.74 $\pm$ 11	70 $\pm$ 11.10	93 $\pm$ 20 $\pm$ 38	197 $\pm$ 53 $\pm$ 14	122.93 $\pm$ 32	44	27 $\pm$ 5	0.9 $\pm$ 0.2				
EPIC-Granada	7,864	50 $\pm$ 9.77	-	12	19	26	125 $\pm$ 19.78 $\pm$ 11	77 $\pm$ 12.8	82 $\pm$ 44 $\pm$ 41	231 $\pm$ 54 $\pm$ 14	145.142 $\pm$ 35	-	29 $\pm$ 5	0.8 $\pm$ 0.2				
EPIC-Gipuzkoa	8,417	50 $\pm$ 8.51	-	9	30	20	125 $\pm$ 19.77 $\pm$ 11	69 $\pm$ 11.6	81 $\pm$ 9 $\pm$ 43	231 $\pm$ 56 $\pm$ 15	152.97 $\pm$ 38	-	27 $\pm$ 4	0.8 $\pm$ 0.2				

Table 3 (continued)

Sample size	Age, years (mean $\pm$ SD)	Female (%)	Born in Spain (%)	%Univ. Education (%)	Current smokers (%)	HT (%)	SBP, mmHg (mean $\pm$ SD)	DBP, mmHg (mean $\pm$ SD)	HR, beat/min (mean $\pm$ SD)	Diabetes (%)	Glucose, mg/dL (mean $\pm$ SD)	TC, mg/dL (mean $\pm$ SD)	HDL-C, mg/dL (mean $\pm$ SD)	LDL-C, mg/dL (mean $\pm$ SD)	Trig., mg/dL median [1 st-3rd quartile]	Cholesterol treatment (%)	BMI, kg/m <sup>2</sup> (mean $\pm$ SD)	Creat. mg/dL (mean $\pm$ SD)
EPIC-Murcia	8,515	49 $\pm$ 8.68	-	15	25	26	124 $\pm$ 17	77 $\pm$ 10	74 $\pm$ 11.6	-	89 $\pm$ 19	222 $\pm$ 19	54 $\pm$ 14	147 $\pm$ 14	147.89 $\pm$ 37	-	28 $\pm$ 4	0.7 $\pm$ 0.2
EFRCV	3,089	40 $\pm$ 13.51	-	10	41	35	126 $\pm$ 18	76 $\pm$ 12	-8	-	-	193 $\pm$ 45	62 $\pm$ 17	-96	-96 [70-141]	5	27 $\pm$ 4	-
Control-MCC-SPAIN	4,077	63 $\pm$ 12.50	95	21	19	39	130 $\pm$ 18	75 $\pm$ 10	-16	-	89 $\pm$ 33	205 $\pm$ 36	58 $\pm$ 13	85 $\pm$ 20	118 [87-164]	25	27 $\pm$ 4	0.8 $\pm$ 0.2
CORSAIB	1,685	54 $\pm$ 11.52	-	8	30	45	130 $\pm$ 20	80 $\pm$ 10	75 $\pm$ 11.12	-	104 $\pm$ 30	217 $\pm$ 41	53 $\pm$ 13	139 $\pm$ 36	107 [77-153]	6	28 $\pm$ 5	-
EMMA	35,175	58 $\pm$ 12.52	-	-	19	67	137 $\pm$ 21	79 $\pm$ 9	-19	-	104 $\pm$ 35	224 $\pm$ 42	53 $\pm$ 14	149 $\pm$ 40	96 [72-132]	9	28 $\pm$ 4	-
ARTPER	3,736	64 $\pm$ 9.54	97	7	17	65	136 $\pm$ 18	78 $\pm$ 10	73 $\pm$ 11.20	-	107 $\pm$ 29	216 $\pm$ 39	55 $\pm$ 14	135 $\pm$ 34	109 [81-151]	31	29 $\pm$ 5	-
NEFRONA	3,004	57 $\pm$ 13.40	-	-	20	88	141 $\pm$ 22	81 $\pm$ 12	-26	-	106 $\pm$ 35	184 $\pm$ 39	50 $\pm$ 15	107 $\pm$ 35	105 [82-128]	57	28 $\pm$ 5	1.9 $\pm$ 1.0
ILERVAS	8,330	57 $\pm$ 6.51	-	-	30	56	131 $\pm$ 17	82 $\pm$ 10	-0	-	-	206 $\pm$ 38	57 $\pm$ 15	145 $\pm$ 26	137 [103-189]	18	29 $\pm$ 5	0.8 $\pm$ 0.2
ASTURIAS	1,034	53 $\pm$ 13.54	-	-	-	43	134 $\pm$ 22	84 $\pm$ 13	-13	-	100 $\pm$ 24	229 $\pm$ 41	56 $\pm$ 14	150 $\pm$ 37	99 [71-141]	-	28 $\pm$ 5	1.0 $\pm$ 0.2
CDC Canarias	7,160	43 $\pm$ 13.56	98	14	26	32	119 $\pm$ 19	75 $\pm$ 12	73 $\pm$ 11.10	-	97 $\pm$ 25	203 $\pm$ 41	51 $\pm$ 13	127 $\pm$ 37	99 [71-146]	39	27 $\pm$ 5	-
HERMEX	2,833	51 $\pm$ 15.54	98	12	32	38	124 $\pm$ 23	72 $\pm$ 11	72 $\pm$ 11.14	-	104 $\pm$ 25	208 $\pm$ 38	56 $\pm$ 14	129 $\pm$ 33	93 [68-132]	17	29 $\pm$ 5	0.8 $\pm$ 0.2
HORTEGA	1,475	52 $\pm$ 19.50	-	28	23	40	131 $\pm$ 21	79 $\pm$ 11	-4	-	-	201 $\pm$ 38	56 $\pm$ 19	109 $\pm$ 39	150 [106-212]	7	26 $\pm$ 4	0.9 $\pm$ 0.2

Table 3 (continued)

Sample size	Age, years (mean $\pm$ SD)	Female (%)	Born in Spain (%)	%Univ. Education (%)	Current smokers (%)	HT (%)	SBP, mmHg (mean $\pm$ SD)	DBP, mmHg (mean $\pm$ SD)	HR, beat/min (mean $\pm$ SD)	Diabetes (%)	Glucose, mg/dL (mean $\pm$ SD)	TC, mg/dL (mean $\pm$ SD)	HDL-C, mg/dL (mean $\pm$ SD)	LDL-C, mg/dL (mean $\pm$ SD)	Trig., mg/dL median [1 st-3rd quartile]	Chol.-Loving treat. (%)	BMI, kg/m <sup>2</sup> (mean $\pm$ SD)	Creat. mg/dL (mean $\pm$ SD)
NAVARRA 93	40 $\pm$ 14.50	99	99	15	47	24	118 $\pm$ 18	76 $\pm$ 11	69 $\pm$ 10.3	96 $\pm$ 17	213 $\pm$ 58	213 $\pm$ 58	133 $\pm$ 16	133 $\pm$ 16	23	26 $\pm$ 4	-	-
EPIC-Navarra	50 $\pm$ 8.52	-	-	9	33	20	-	-	-3	77 $\pm$ 17	228 $\pm$ 54	228 $\pm$ 54	153 $\pm$ 14	153 $\pm$ 14	153 $\pm$ 14	-	28 $\pm$ 4	0.8 $\pm$ 0.2
RIVANA	54 $\pm$ 13.55	96	96	22	27	40	128 $\pm$ 19	78 $\pm$ 10	52 $\pm$ 14.12	102 $\pm$ 22	212 $\pm$ 64	212 $\pm$ 64	126 $\pm$ 16	126 $\pm$ 16	126 $\pm$ 16	12	27 $\pm$ 4	0.9 $\pm$ 0.2
RECCyL	52 $\pm$ 18.52	98	98	-	24	41	128 $\pm$ 19	76 $\pm$ 11	-10	98 $\pm$ 26	209 $\pm$ 59	209 $\pm$ 59	131 $\pm$ 15	131 $\pm$ 15	131 $\pm$ 15	11	27 $\pm$ 5	0.9 $\pm$ 0.2
PREDIMERC	50 $\pm$ 13.52	87	87	24	26	38	127 $\pm$ 19	76 $\pm$ 10	71 $\pm$ 11.9	99 $\pm$ 22	201 $\pm$ 57	201 $\pm$ 57	123 $\pm$ 15	123 $\pm$ 15	123 $\pm$ 15	12	27 $\pm$ 5	0.8 $\pm$ 0.2
IBERICAN	58 $\pm$ 15.55	-	-	14	19	56	129 $\pm$ 16	77 $\pm$ 10	73 $\pm$ 11.21	102 $\pm$ 28	195 $\pm$ 55	195 $\pm$ 55	118 $\pm$ 10.7	118 $\pm$ 10.7	118 $\pm$ 10.7	37	28 $\pm$ 5	0.8 $\pm$ 0.3
AEGIS	52 $\pm$ 17.55	94	94	12	21	41	128 $\pm$ 16	76 $\pm$ 9	65 $\pm$ 11.12	94 $\pm$ 23	197 $\pm$ 59	197 $\pm$ 59	115 $\pm$ 17	115 $\pm$ 17	115 $\pm$ 17	19	28 $\pm$ 5	0.7 $\pm$ 0.2
DI@BETES-EUSKADI	52 $\pm$ 15.53	95	95	24	24	42	132 $\pm$ 21	80 $\pm$ 12	-10	92 $\pm$ 22	206 $\pm$ 61	206 $\pm$ 61	124 $\pm$ 17	124 $\pm$ 17	124 $\pm$ 17	11	26 $\pm$ 5	0.8 $\pm$ 0.2
CARGEN-CORS	57 $\pm$ 13.55	-	-	-	9	45	129 $\pm$ 17	77 $\pm$ 11	-15	101 $\pm$ 29	199 $\pm$ 56	199 $\pm$ 56	122 $\pm$ 10.8	122 $\pm$ 10.8	122 $\pm$ 10.8	-	28 $\pm$ 5	0.8 $\pm$ 0.3
GCAT	51 $\pm$ 7.59	96	96	38	22	27	122 $\pm$ 15	77 $\pm$ 10	74 $\pm$ 11.4	92 $\pm$ 18	209 $\pm$ 59	209 $\pm$ 59	130 $\pm$ 15	130 $\pm$ 15	130 $\pm$ 15	7	27 $\pm$ 5	0.8 $\pm$ 0.2
RICARTO	50 $\pm$ 16.56	-	-	22	23	33	123 $\pm$ 17	72 $\pm$ 11	74 $\pm$ 11.9	89 $\pm$ 19	193 $\pm$ 57	193 $\pm$ 57	114 $\pm$ 15	114 $\pm$ 15	114 $\pm$ 15	20	27 $\pm$ 5	0.8 $\pm$ 0.2
All	52 $\pm$ 13.54	96	96	20	24	43	129 $\pm$ 20	78 $\pm$ 11	71 $\pm$ 12.11	99 $\pm$ 27	210 $\pm$ 55	210 $\pm$ 55	132 $\pm$ 15	132 $\pm$ 15	132 $\pm$ 15	15	28 $\pm$ 5	0.9 $\pm$ 0.3

BMI body mass index, *Chol.-lowering treat.* cholesterol-lowering treatment, *Creat.* Creatinine, *DBP* diastolic blood pressure, *HChol* hypercholesterolemia, *HDL-C* HDL cholesterol, *HR* heart rate, *HT* hypertension, *LDL-C* LDL cholesterol, *SBP* systolic blood pressure, *TC* total cholesterol, *Trig.* Triglycerides, *Univ.* university

to be representative of the Spanish adult population thus improving the generalizability of our findings to Southern European adult populations (genetic ancestry, lifestyle, and clinical variables are expected to be comparable across cohorts). It will also conduct the largest GWAS on ASCVD in Southern European population. Due to its unique characteristics, it can assess the association of certain rare variants that are specific to Spanish or South European population with the occurrence of ASCVD. It will also investigate whether a polygenic risk score can improve the current predictive functions for the 10-year incidence of ASCVD in this same population. Genetic characteristics vary according to ethnicity and location; therefore, it is plausible that calculating a polygenic risk score that considers “local” genetic variants improves risk prediction. Additionally, it will explore whether there are relevant interactions between genetic and lifestyle information on the development of ASCVD that are specific to Southern European population. It is plausible that some interactions are only evident if both the local genetic background and a local environmental exposure are considered together (e.g., Southern European populations have greater adherence to a Mediterranean diet than other regions of the world).

However, it also has limitations. First, inter-cohort data may be heterogeneous, as the cohorts may have used different methods for recruitment, phenotype definition and data collection, which was also performed at distinct times (between 1989 and 2020). To minimize this bias, the Data Management team, in collaboration with the Committee for the Harmonization of Outcomes and Variables of the CORDELIA Study, has performed exhaustive harmonization of the available data (to the lowest common denominator for categorical variables) and the statistical analyses will undergo central validation and be adapted to the nature of the data (whenever possible, mixed models that consider the cohort as a random effect will be used). Regarding genotype heterogeneity and population stratification, these issues will be minimized by analyzing participants’ relatedness and adjusting for genetic principal components. If significant genotype heterogeneity among cohorts is detected, individual GWAS analyses in each participating cohort may be performed and subsequently meta-analyzed [75]. Second, while imputation allows obtaining results for a very high number of genetic variants (about 9–10 million), the number of directly genotyped variants will be about 750,000. To minimize potential biases, imputation will be performed using the comprehensive TOPMed r3 (GRCh38) reference panel and a stringent post-imputation filtering (INFO score  $\geq 0.8$  and a minor allele frequency  $> 1\%$ ). Third, our data may not be fully representative of the Spanish population due to different designs for population selection (two/three-stage clusters, random, and non-random selections of population-based, hospital-based, and worker studies), the characteristics of

the participants in the individual cohorts (some focused on healthy individuals, while others included populations with various pathologies; additionally, each cohort recruited individuals of diverse age ranges) and a recruitment over a period of almost 30 years. The large number of participants included in the meta-cohort may help mitigate these differences, but an exhaustive description of the study population must be provided in each publication derived from our data. Fourth, the follow-up times and the diverse age ranges in the participating cohorts may decrease the capacity of cohorts with shorter follow-ups or younger participants to detect outcomes. To minimize this limitation, all cohorts will conduct a follow-up update of fatal and non-fatal events during 2024, allowing for more homogeneous tracking of the participants across studies. Fifth, anthropometric, lifestyle, and socioeconomic variables in certain cohorts are self-reported, which may increase measurement error. Sixth, although the sample size in such a study in Southern European population is the largest to date, the statistical power to detect associations between rare genetic variants (those with an AF of  $\leq 1\%$ ) and the outcomes of interest may be limited. Seventh, since CORDELIA leverages existing data and biological samples from 35 cohorts, genotype data will be available for at most 60% of participants, and biological samples will be available for 50%. Biological samples cannot be obtained from cohorts that do not already have stored DNA or serum/plasma samples. Eighth, the definitions of hypertension and diabetes across the cohorts may vary, as they might be based on self-reported information, individual-level measurements, and treatment information. This variation could introduce heterogeneity and bias in assessing these conditions across cohorts. To minimize this limitation, we have developed a harmonized algorithm to maximize the number of participants with available information on their history of hypertension and diabetes (available in 99.9% and 99.7% of the participants, respectively). Finally, while our conclusions could be generalized to Southern European adults with similar characteristics, they may not be extended to populations of other genetic ancestry or different regions.

## Conclusions

The CORDELIA Study marks a significant advancement as the largest cohort for the study of cardiovascular disease within South European adult population to date. CORDELIA also establishes the foundation for the most extensive GWAS on ASCVD in Southern Europe using a well-characterized population with long follow-up. This infrastructure will also support future projects targeting other diseases. CORDELIA aims to enhance the representation of Southern Europeans in genomic studies and to develop tailored polygenic risk scores that refine ASCVD prediction and

potentially improve targeted prevention strategies. By integrating comprehensive clinical, environmental, genetic, and socioeconomic data, CORDELIA exemplifies how large-scale collaborative research can advance precision medicine.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10654-025-01229-6>.

**Acknowledgements** The CORDELIA team would like to thank the participants who were part of the individual cohorts that make up the project, the health personnel and field work coordinators who facilitated the collection of data and biological samples in the individual cohorts, Marta Cabañero for her work in the construction of the study database, Esmeralda Gómez and Sandra Farré for their work as project managers, the Hospital del Mar Research Institute laboratory team (Daniel Muñoz-Aguayo, Gemma Blanchart, Sònia Gaixas) for their work in the extraction and quality control of DNA samples of the study, and the Centro Nacional de Genotipado (CEGEN) and the Centro Singular de Investigación en Medicina Molecular y Enfermedades Crónicas (CIMUS) team (Silvia Diz) for performing the genotyping and for their support with genotyped data quality control. CIBER de Enfermedades Cardiovasculares (CIBERCV), CIBER de Epidemiología y Salud Pública (CIBERESP), CIBER de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), CIBER de Enfermedades Raras (CIBERER), CIBER de Fragilidad y Envejecimiento Saludable (CIBERFES) y CIBER de Fisiopatología de la Obesidad y Nutrición (CIBEROBN) are initiatives of Instituto de Salud Carlos III (Madrid, Spain) and are financed by the European Regional Development Fund. The CORDELIA researchers belong to the following groups of the CIBER areas: CIBERCV (CB16/11/00229, CB16/11/00246, CB16/11/00226, CB16/11/00451, CB16/11/00374), CIBERESP (CB06/02/0001, CB06/02/0042, CB06/02/0072, CB06/02/0074, CB06/02/0023, CB06/02/0084, CB06/02/0076), CIBERDEM (CB07/08/0019, CB15/00071), CIBERER (CB19/07/00057), CIBERFES (CB16/10/00245) and CIBEROBN (CB12/03/30017). A full list of the investigators of the cohorts included in the CORDELIA Study is available in the Appendix I and at the project website (<https://corde.liaproject.net/component-cohorts/>).

**Author contributions** Conceptualization and Methodology: J.M.; Project administration: J.M., Á.H., R.R., R.D.C., F.R.-A., R.E., M.D.C., P.A., M.B.-L., M.G., M.-J.S., A.C.D.L., S.C.-S., M.L., G.R.-M., M.J.G.-S., B.P.-G., T.V.-A., P.T.-M., D.L.-P., J.M.H., J.M.V., A.C.-V., F.J.F.-R., A.M.G., S.V., X.M.-T., P.L.S., V.M.-S., F.R., C.M.-I., M.A.-S., J.C.M.-E., E.D., M.G., L.A.C.-G., D.O.; Formal analysis and investigation: Á.H., A.C.-V., S.P.-A., I.S., I.Q., R.C.; Writing—original draft preparation: Á.H.; Writing—review and editing: A.C.-V., S.P.-A., I.S., R.R., R.D.C., F.R.-A., R.E., M.D.C., P.A., M.B.-L., M.G., M.-J.S., A.C.D.L., S.C.-S., M.L., G.R.-M., M.J.G.-S., B.P.-G., T.V.-A., P.T.-M., D.L.-P., J.M.H., J.M.V., I.R.D., F.J.F.-R., A.M.G., S.V., X.M.-T., P.L.S., V.M.-S., F.R., C.M.-I., M.A.-S., J.C.M.-E., E.D., M.G., L.A.C.-G., D.O., I.Q., R.M.-L., N.B., J.R.B., H.T., J.H.G., A.A., E.C.-B., J.D., M.R.-B., D.A.-G., M.A.P.-D., B.M.-F., W.O.-B., C.S.-O., E.P.-B., J.E.L., V.M.L.-L., P.C.-N., N.C.-C., S.C., L.Z.-P., D.F.-B., E.D.-N., C.M.-A., J.F.-N., P.I.D.-D., A.V.-C., G.F.-J., N.A., F.G., M.T.-P., J.A.-B., R.C., M.R.-A., J.B., P.G.-C., M.C.-A., S.M.C.-Y., A.L., J.M.D.-T., E.A., N.F.F.-M., M.D.C.-R.-P., R.M.M.-P., A.C., A.L.C.-P., J.M. All authors read and approved the final manuscript.

**Funding** The CORDELIA Study variable harmonization and GWAS were funded by Instituto de Salud Carlos III and co-funded by the European Union (Proyecto de Investigación de Medicina Personalizada de Precisión PMP22/00033, and projects PI21/00040 and PI21/00163). A detailed description of the funding sources of the individual cohorts is available in Appendix II.

## Declarations

**Competing interests** The authors have no relevant financial or non-financial interests to disclose.

**Ethics approval** This study was performed in line with the principles of the Declaration of Helsinki. Approval for the CORDELIA Study was granted by the Ethics Committee of the Parc de Salut Mar (2023/10785/I, date: March 29, 2023). All cohorts have obtained informed consent by their participants and were approved by the corresponding ethics committees. The original informed consents have been expanded for this study by applying the Additional Disposition 17.2.C of the Spanish Organic Law 3/2018 on the Protection of Personal Data and the guarantee of digital rights, which allow for participant follow-up, aggregation of anonymized data, and the analyses described in this project. Extended information about ethical approval in the individual cohorts is available in Appendix III.

**Consent to participate** All participants in the CORDELIA cohorts agreed to participate after providing written informed consent for each cohort. Participants were given the option to decline their participation in the CORDELIA project through a document made available online in the websites of the individual cohorts, in accordance with Spanish law.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. GBD 2021 Diseases and Injuries Collaborators. Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet*. 2024;403:2133–61.
2. O'Sullivan JW, Raghavan S, Marquez-Luna C, Luzum JA, Damrauer SM, Ashley EA, et al. Polygenic Risk Scores for Cardiovascular Disease: A Scientific Statement From the American Heart Association. *Circulation*. 2022;146:e93–118.
3. Visseren FLJ, Mach F, Smulders YM, Carballo D, Koskinas KC, Bäck M, et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice. *Eur Heart J*. 2021;42:3227–337.
4. Chen Z, Chen J, Collins R, Guo Y, Peto R, Wu F, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol*. 2011;40:1652–66.
5. All of Us Research Program Investigators, Denny JC, Rutter JL, Goldstein DB, Philippakis A, Smoller JW, et al. The “All of Us” Research Program. *N Engl J Med*. 2019;381:668–76.
6. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, 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 Med.* 2015;12: e1001779.
7. Zins M, Goldberg M, CONSTANCES team. The French CONSTANCES population-based cohort: design, inclusion and follow-up. *Eur J Epidemiol.* 2015;30:1317–28.
  8. Instituto de Salud Carlos III. Infraestructura de Medicina de Precisión asociada a la Ciencia y la Tecnología - IMPaCT [Internet]. [cited 2024 Apr 10]. Available from: <https://www.isciii.es/QueHacemos/Financiacion/IMPACT/Paginas/Plan.aspx>
  9. Mills MC, Rahal C. A scientometric review of genome-wide association studies. *Commun Biol.* 2019;2:9.
  10. Marrugat J, Vila J, Baena-Díez JM, Grau M, Sala J, Ramos R, et al. [Relative validity of the 10-year cardiovascular risk estimate in a population cohort of the REGICOR study]. *Rev Esp Cardiol.* 2011;64:385–94.
  11. Emerging Risk Factors Collaboration, Kaptoge S, Di Angelantonio E, Pennells L, Wood AM, White IR, et al. C-reactive protein, fibrinogen, and cardiovascular disease prediction. *N Engl J Med.* 2012;367:1310–20.
  12. Khera AV, Kathiresan S. Genetics of coronary artery disease: discovery, biology and clinical translation. *Nat Rev Genet.* 2017;18:331–44.
  13. Grau M, Subirana I, Elosua R, Solanas P, Ramos R, Masiá R, et al. Trends in cardiovascular risk factor prevalence (1995–2000–2005) in northeastern Spain. *Eur J Cardiovasc Prev Rehabil.* 2007;14:653–9.
  14. Barroso M, Pérez-Fernández S, Vila MM, Zomeño MD, Martí-Lluch R, Cordon F, et al. Validity of a method for the self-screening of cardiovascular risk. *Clin Epidemiol.* 2018;10:549–60.
  15. Bustamante M, Nogués X, Mellibovsky L, Agueda L, Jurado S, Cáceres E, et al. Polymorphisms in the interleukin-6 receptor gene are associated with bone mineral density and body mass index in Spanish postmenopausal women. *Eur J Endocrinol.* 2007;157:677–84.
  16. Casasnovas JA, Alcaide V, Civeira F, Guallar E, Ibañez B, Borreguero JJ, et al. Aragon workers' health study—design and cohort description. *BMC Cardiovasc Disord.* 2012;12:45.
  17. Melero-Alegria JJ, Cascon M, Romero A, Vara PP, Barreiro-Perez M, Vicente-Palacios V, et al. SALMANTICOR study. Rationale and design of a population-based study to identify structural heart disease abnormalities: a spatial and machine learning analysis. *BMJ Open.* 2019;9:e024605.
  18. Soriguer F, Goday A, Bosch-Comas A, Bordiú E, Calle-Pascual A, Carmena R, et al. Prevalence of diabetes mellitus and impaired glucose regulation in Spain: the Di@bet.es Study. *Diabetologia.* 2012;55:88–93.
  19. Soriguer F, Gutiérrez-Repiso C, Rubio-Martín E, García-Fuentes E, Almaraz MC, Colomo N, et al. Metabolically healthy but obese, a matter of time? Findings from the prospective Pizarra study. *J Clin Endocrinol Metab.* 2013;98:2318–25.
  20. Díaz-Redondo A, Giraldez-García C, Carrillo L, Serrano R, García-Soidán FJ, Artola S, et al. Modifiable risk factors associated with prediabetes in men and women: a cross-sectional analysis of the cohort study in primary health care on the evolution of patients with prediabetes (PREDAPS-Study). *BMC Fam Pract.* 2015;16:5.
  21. Gómez Gerique JA, Ferreira Gonzalez I, Rubio Herrera MA, Lora Pablos D, Martín Ballesteros B, García Sardina R, et al. Improvement of serum lipids concentration in a general population historical cohort. Why? *Clin Investig Arterioscler.* 2017;29:239–47.
  22. Rodríguez-Artalejo F, Graciani A, Guallar-Castillón P, León-Muñoz LM, Zuluaga MC, López-García E, et al. [Rationale and methods of the study on nutrition and cardiovascular risk in Spain (ENRICA)]. *Rev Esp Cardiol.* 2011;64:876–82.
  23. Riboli E, Kaaks R. The EPIC Project: rationale and study design. *European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol.* 1997;26 Suppl 1:S6–14.
  24. Huerta JM, Tormo M-J, Gavrilá D, Navarro C. Cardiovascular risk estimated after 13 years of follow-up in a low-incidence Mediterranean region with high-prevalence of cardiovascular risk factors. *BMC Public Health.* 2010;10:640.
  25. Castaño-Vinyals G, Aragonés N, Pérez-Gómez B, Martín V, Llorca J, Moreno V, et al. Population-based multicase-control study in common tumors in Spain (MCC-Spain): rationale and study design. *Gac Sanit.* 2015;29:308–15.
  26. Rigo Carratalá F, Frontera Juan G, Llobera Cànoves J, Rodríguez Ruiz T, Borrás Bosch I, Fuentespiná Vidal E [Prevalence of cardiovascular risk factors in the Balearic Islands (CORSIAIB Study)]. *Rev Esp Cardiol.* 2005;58:1411–9.
  27. Ramos R, Balló E, Marrugat J, Elosua R, Sala J, Grau M, et al. Validity for use in research on vascular diseases of the SID-IAP (Information System for the Development of Research in Primary Care): the EMMA study. *Rev Esp Cardiol (Engl Ed).* 2012;65:29–37.
  28. Alzamora MT, Forés R, Baena-Díez JM, Pera G, Toran P, Sorribes M, et al. The peripheral arterial disease study (PERART/ARTPER): prevalence and risk factors in the general population. *BMC Public Health.* 2010;10:38.
  29. Betriu A, Martínez-Alonso M, Arcidiacono MV, Cannata-Andía J, Pascual J, Valdivielso JM, et al. Prevalence of subclinical atheromatosis and associated risk factors in chronic kidney disease: the NEFRONA study. *Nephrol Dial Transplant.* 2014;29:1415–22.
  30. Bermúdez-López M, Martínez-Alonso M, Castro-Boqué E, Betriu À, Cambray S, Farràs C, et al. Subclinical atheromatosis localization and burden in a low-to-moderate cardiovascular risk population: the ILERVAS study. *Rev Esp Cardiol (Engl Ed).* 2021;74:1042–53.
  31. Botas Cervero P, Delgado Alvarez E, Castaño Fernández G, Díaz De Greñu C, Prieto Santiago J, Díaz Cadórniga FJ [Prevalence of diabetes mellitus and glucose intolerance in the population aged 30 to 75 years in Asturias, Spain]. *Rev Clin Esp.* 2002;202:421–9.
  32. Cabrera de León A, Rodríguez Pérez M del C, Almeida González D, Domínguez Coello S, Aguirre Jaime A, Brito Díaz B, et al. [Presentation of the “CDC de Canarias” cohort: objectives, design and preliminary results]. *Rev Esp Salud Publica.* 2008;82:519–34.
  33. Félix-Redondo FJ, Fernández-Bergés D, Fernando Pérez J, Zaro MJ, García A, Lozano L, et al. [Prevalence, awareness, treatment and control of cardiovascular risk factors in the Extremadura population (Spain). HERMEX study]. *Aten Primaria.* 2011;43:426–34.
  34. Tellez-Plaza M, Briongos-Figuero L, Pichler G, Dominguez-Lucas A, Simal-Blanco F, Mena-Martin FJ, et al. Cohort profile: the Horteiga Study for the evaluation of non-traditional risk factors of cardiometabolic and other chronic diseases in a general population from Spain. *BMJ Open.* 2019;9: e024073.
  35. Guembe Suescun, M. J. Factores de riesgo cardiovascular en población navarra de 18 a 64 años. *Boletín de Salud Pública de Navarra.* 1994;2–8.
  36. Viñes JJ, Díez J, Guembe MJ, González P, Amézqueta C, Barba J, et al. [Study of vascular risk in Navarre: objectives and design. Prevalence of metabolic syndrome and of vascular risk factors]. *An Sist Sanit Navar.* 2007;30:113–24.
  37. Junta de Castilla y León. Riesgo de Enfermedad Cardiovascular en Castilla y León [Internet]. Portal de Salud de la Junta de Castilla y León. [cited 2024 Apr 11]. Available from: <https://www.saludcastillayleon.es/profesionales/es/cardiovascular-4eea>
  38. Gil Montalbán E, Ortiz Marrón H, López-Gay Lucio-Villegas D, Zorrilla Torrás B, Arrieta Blanco F, Nogales Aguado P [Validity and concordance of electronic health records in primary care

- (AP-Madrid) for surveillance of diabetes mellitus. PREDIMERC study]. *Gac Sanit.* 2014;28:393–6.
39. Cinza Sanjurjo S, Llisterrí Caro JL, Barquilla García A, Polo García J, Velilla Zancada S, Rodríguez Roca GC, et al. [Description of the sample, design and methods of the study for the identification of the Spanish population at cardiovascular and renal risk (IBERICAN)]. *Semergen.* 2020;46:4–15.
40. Gude F, Díaz-Vidal P, Rúa-Pérez C, Alonso-Sampedro M, Fernández-Merino C, Rey-García J, et al. Glycemic Variability and Its Association With Demographics and Lifestyles in a General Adult Population. *J Diabetes Sci Technol.* 2017;11:780–90.
41. Aguayo A, Urrutia I, González-Frutos T, Martínez R, Martínez-Indart L, Castaño L, et al. Prevalence of diabetes mellitus and impaired glucose metabolism in the adult population of the Basque Country. *Spain Diabet Med.* 2017;34:662–6.
42. Camps-Vilaró A, Pinsach-Abuin ML, Degano IR, Ramos R, Martí-Lluch R, Elosua R, et al. Genetic characteristics involved in COVID-19 severity. The CARGENCORS case-control study and meta-analysis. *J Med Virol.* 2024;96:e29404.
43. Obón-Santacana M, Vilardell M, Carreras A, Duran X, Velasco J, Galván-Femenía I, et al. GCAT|Genomes for life: a prospective cohort study of the genomes of Catalonia. *BMJ Open.* 2018;8:e018324.
44. Rodríguez-Roca GC, Segura-Fragoso A, Villarín-Castro A, Alonso-Moreno FJ, Rodríguez-Padial L, Rodríguez-García ML, et al. [Characteristics and cardiovascular events in a general population included in the RICARTO (Riesgo CARDiovascular TOledo) study: Data from the first 1,500 individuals included in the study]. *Semergen.* 2018;44:180–91.
45. de Coö A, Cruz R, Quintela I, Herrera D, Sanz M, Diz P, et al. Genome-wide association study of stage III/IV grade C periodontitis (former aggressive periodontitis) in a Spanish population. *J Clin Periodontol.* 2021;48:896–906.
46. Graffelman J, Moreno V. The mid p-value in exact tests for Hardy-Weinberg equilibrium. *Stat Appl Genet Mol Biol.* 2013;12:433–48.
47. Price AL, Weale ME, Patterson N, Myers SR, Need AC, Shianna KV, et al. Long-range LD can confound genome scans in admixed populations. *Am J Hum Genet.* 2008;83:132–5; author reply 135–139.
48. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19:1655–64.
49. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature.* 2015;526:68–74.
50. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature.* 2021;590:290–9.
51. The World Health Organization MONICA Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. WHO MONICA Project Principal Investigators. *J Clin Epidemiol.* 1988;41:105–14.
52. Valverde JC, Tormo M-J, Navarro C, Rodríguez-Barranco M, Marco R, Egea J-M, et al. Prevalence of diabetes in Murcia (Spain): a Mediterranean area characterised by obesity. *Diabetes Res Clin Pract.* 2006;71:202–9.
53. Gil Montalbán E, Zorrilla Torras B, Ortiz Marrón H, Martínez Cortés M, Donoso Navarro E, Nogales Aguado P, et al. [Prevalence of diabetes mellitus and cardiovascular risk factors in the adult population of the autonomous region of Madrid (Spain): the PREDIMERC study]. *Gac Sanit.* 2010;24:233–40.
54. Vega Alonso AT, Lozano Alonso JE, Alamo Sanz R, Lleras Muñoz S, Escribano Hernández A, De la Iglesia RP, et al. [Design of a population-based study of cardiovascular risk in Castile and Leon [Spain] through primary care teams]. *Gac Sanit.* 2007;21:84–7.
55. Segura Fragoso A, Rius Mery G [Cardiovascular risk factors in a rural population of Castilla-La Mancha]. *Rev Esp Cardiol.* 1999;52:577–88.
56. Grau M, et al. [Cardiovascular risk factors in Spain in the first decade of the 21st Century, a pooled analysis with individual data from 11 population-based studies: the DARIOS study]. *Rev Esp Cardiol.* 2011;64:295–304.
57. Inker LA, Schmid CH, Tighiouart H, Eckfeldt JH, Feldman HI, Greene T, et al. Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med.* 2012;367:20–9.
58. Schröder H, Fitó M, Estruch R, Martínez-González MA, Corella D, Salas-Salvadó J, et al. A short screener is valid for assessing Mediterranean diet adherence among older Spanish men and women. *J Nutr.* 2011;141:1140–5.
59. Sala J, Marrugat J, Masiá R, Porta M. Improvement in survival after myocardial infarction between 1978–85 and 1986–88 in the REGICOR study. (Registre Gironí del COR) registry. *Eur Heart J.* 1995;16:779–84.
60. Gil M, Marrugat J, Sala J, Masiá R, Elosua R, Albert X, et al. Relationship of therapeutic improvements and 28-day case fatality in patients hospitalized with acute myocardial infarction between 1978 and 1993 in the REGICOR study, Gerona. Spain The REGICOR Investigators. *Circulation.* 1999;99:1767–73.
61. Ramírez-Moreno JM, Rebollo B, Macías-Sedas P, Valverde N, Parejo A, Felix-Redondo FJ, et al. Strength of association of classical vascular risk factors in young patients with ischaemic stroke: a case-control study. *Neurologia (Engl Ed).* 2022;S2173–5808(22):00142.
62. Alzamora MT, Forés R, Pera G, Baena-Díez JM, Heras A, Sorribes M, et al. Incidence of peripheral arterial disease in the ART-PER population cohort after 5 years of follow-up. *BMC Cardiovasc Disord.* 2016;16:8.
63. Laclaustra M, Casasnovas JA, Fernández-Ortiz A, Fuster V, León-Latre M, Jiménez-Borreguero LJ, et al. Femoral and Carotid Subclinical Atherosclerosis Association With Risk Factors and Coronary Calcium: The AWHs Study. *J Am Coll Cardiol.* 2016;67:1263–74.
64. Gutiérrez-González E, Pastor-Barriuso R, Castelló A, Castaño-Vinyals G, Fernández de Larrea-Baz N, Dierssen-Sotos T, et al. Toenail zinc and risk of prostate cancer in the MCC-Spain case-control study. *Environ Res.* 2024;245:118065.
65. Botteri E, Peveri G, Berstad P, Bagnardi V, Hoff G, Heath AK, et al. Lifestyle changes in middle age and risk of cancer: evidence from the European Prospective Investigation into Cancer and Nutrition. *Eur J Epidemiol.* 2024;39:147–59.
66. Bonet C, Crous-Bou M, Tsilidis KK, Gunter MJ, Kaaks R, Schulze MB, et al. The association between body fatness and mortality among breast cancer survivors: results from a prospective cohort study. *Eur J Epidemiol.* 2023;38:545–57.
67. Papantoniou K, Castaño-Vinyals G, Espinosa A, Aragonés N, Pérez-Gómez B, Ardanaz E, et al. Breast cancer risk and night shift work in a case-control study in a Spanish population. *Eur J Epidemiol.* 2016;31:867–78.
68. Andreu-Reinón ME, Huerta JM, Gavrilá D, Amiano P, Mar J, Tainta M, et al. Incidence of Dementia and Associated Factors in the EPIC-Spain Dementia Cohort. *J Alzheimers Dis.* 2020;78:543–55.
69. Gallo V, Vanacore N, Bueno-de-Mesquita HB, Vermeulen R, Brayne C, Pearce N, et al. Physical activity and risk of Amyotrophic Lateral Sclerosis in a prospective cohort study. *Eur J Epidemiol.* 2016;31:255–66.
70. Martínez-García F, Mansego ML, Rojo-Martínez G, De Marco-Solar G, Morcillo S, Soriguer F, et al. Impact of obesity-related genes in Spanish population. *BMC Genet.* 2013;14:111.

71. Soriguer F, Rojo-Martínez G, Esteva de Antonio I, Ruiz de Adana MS, Catalá M, Merelo MJ, et al. Prevalence of obesity in south-east Spain and its relation with social and health factors. *Eur J Epidemiol*. 2004;19:33–40.
72. Struijk EA, Banegas JR, Rodríguez-Artalejo F, Lopez-Garcia E. Consumption of meat in relation to physical functioning in the Seniors-ENRICA cohort. *BMC Med*. 2018;16:50.
73. Salinas-Roca B, Sánchez E, Bermúdez-López M, Valdivielso JM, Farràs-Sallés C, Pamplona R, et al. Association between adherence to the mediterranean diet, physical activity, and sleep apnea-hypopnea syndrome (SAHS) in a middle-aged population with cardiovascular risk: Insights from the ILERVAS cohort. *Sleep Med*. 2024;116:19–26.
74. Kogevinas M, Karachaliou M, Espinosa A, Aguilar R, Castaño-Vinyals G, Garcia-Aymerich J, et al. Long-Term Exposure to Air Pollution and COVID-19 Vaccine Antibody Response in a General Population Cohort (COVICAT Study, Catalonia). *Environ Health Perspect*. 2023;131:47001.
75. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26:2190–1.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

Álvaro Hernández<sup>1,2,3</sup>  · Anna Camps-Vilaró<sup>1,2,4</sup> · Sara Polo-Alonso<sup>1,2,5</sup> · Isaac Subirana<sup>2,6</sup> · Rafel Ramos<sup>7,8,9,10</sup> · Rafael de Cid<sup>11,12</sup> · Fernando Rodríguez-Artalejo<sup>13,14,15</sup> · Roberto Elosua<sup>2,4,6</sup> · M. Dolores Chirlaque<sup>15,16,17</sup> · Pilar Amiano<sup>15,18,19</sup> · Marcelino Bermúdez-López<sup>20,21</sup> · Marcela Guevara<sup>15,22,23</sup> · Sergio Cinza-Sanjurjo<sup>2,24,25,26</sup> · María-José Sánchez<sup>15,27,28</sup> · Antonio Cabrera de León<sup>29,30</sup> · Martín Laclaustra<sup>2,31,32,33</sup> · Gemma Rojo-Martínez<sup>34,35,36</sup> · María J. Guembe-Suescun<sup>37,38</sup> · Beatriz Pérez-Gómez<sup>15,39</sup> · Tomás Vega-Alonso<sup>40</sup> · Pere Torán-Monserrat<sup>7,41,42,43</sup> · David Lora-Pablos<sup>44,45,46</sup> · José María Huerta<sup>15,16</sup> · José M. Valdivielso<sup>20,47</sup> · Irene R. Dégano<sup>1,2,4,48</sup> · Francisco J. Félix-Redondo<sup>49,50,51</sup> · Ana María Gandarillas<sup>52</sup> · Sergio Valdés<sup>34,35,36</sup> · Xavier Mundet-Tuduri<sup>53</sup> · Pedro L. Sánchez<sup>2,54,55</sup> · Vicente Martín-Sánchez<sup>14,56</sup> · Fernando Rigo<sup>57</sup> · Manuela Alonso-Sampedro<sup>10,25</sup> · Conchi Moreno-Iribas<sup>15,22,23</sup> · Juan Carlos Martín-Escudero<sup>58</sup> · Elías Delgado<sup>59,60,61,62</sup> · Maria Grau<sup>15,63,64</sup> · Inés Urrutia<sup>36,61,65,66,67</sup> · Diana Ovejero<sup>68,69</sup> · Inés Quintela<sup>70</sup> · Ruth Martí-Lluch<sup>7,8,9,10</sup> · Natalia Blay<sup>11,12</sup> · José R. Banegas<sup>13,15</sup> · Helena Tizón-Marcos<sup>1,2,71</sup> · Jesús Humberto Gómez<sup>15,16</sup> · Amaia Aizpurua<sup>18,19</sup> · Eva Castro-Boqué<sup>20</sup> · Josu Delfrade<sup>15,22,23</sup> · Miguel Ángel Prieto-Díaz<sup>72</sup> · Miguel Rodríguez-Barranco<sup>15,27,28</sup> · Delia Almeida-González<sup>29</sup> · Belén Moreno-Franco<sup>2,31,32,33</sup> · Wasima Oualla-Bachiri<sup>34,35,36</sup> · Carmen Sayón-Orea<sup>22,23,73,74</sup> · Elena Plans-Beriso<sup>15,39</sup> · José Eugenio Lozano<sup>40</sup> · Víctor M. López-Lifante<sup>41,75</sup> · Pilar Cancelas-Navia<sup>44,45</sup> · Natalia Cabrera-Castro<sup>15,16</sup> · Serafí Cambray<sup>20,76</sup> · Lluís Zacarías-Pons<sup>7,10</sup> · Daniel Fernández-Bergés<sup>49,50</sup> · Encarnación Donoso-Navarro<sup>77</sup> · Cristina Maldonado-Araque<sup>34,35,36</sup> · Josep Franch-Nadal<sup>36,53</sup> · Pedro Ignacio Dorado-Díaz<sup>2,54,55</sup> · Alejandro Villarín-Castro<sup>78</sup> · Guillem Frontera-Juan<sup>57</sup> · Francisco Gude<sup>10,25,79</sup> · Naroa Andueza<sup>22,23</sup> · María Téllez-Plaza<sup>13,39</sup> · Jessica Ares-Blanco<sup>59,60,61</sup> · Raquel Cruz<sup>62,80</sup> · Marc Ribas-Aulinas<sup>7</sup> · Jordi Barretina<sup>11</sup> · Pilar Guallar-Castillón<sup>13,14,15</sup> · Miguel Caínzos-Achirica<sup>1,2,71</sup> · Sandra Milena Colorado-Yohar<sup>15,16,81</sup> · Adrián Llorente<sup>18,19</sup> · Juan Miguel Diaz-Tocados<sup>20</sup> · Eva Ardanaz<sup>15,22,23</sup> · Rafael Manuel Micó-Pérez<sup>82</sup> · Nicolás Francisco Fernandez-Martinez<sup>15,27,28</sup> · María del Cristo Rodríguez-Pérez<sup>29</sup> · Ana Cenarro<sup>2,31,32,83</sup> · Alfonso L. Calle-Pascual<sup>36,84,85</sup> · Jaume Marrugat<sup>1,2</sup>

✉ Álvaro Hernández  
ahernaez@researchmar.net

✉ Jaume Marrugat  
jmarrugat@researchmar.net

<sup>1</sup> REGICOR Study Group, Hospital del Mar Research Institute (IMIM), Carrer Doctor Aiguader 88, 08003 Barcelona, Spain

<sup>2</sup> CIBER de Enfermedades Cardiovasculares (CIBERCV), Instituto de Salud Carlos III, Madrid, Spain

<sup>3</sup> Blanquerna School of Health Sciences, University Ramon Llull, Barcelona, Spain

<sup>4</sup> School of Medicine, University of Vic-Central University of Catalonia (UVic-UCC), Vic, Spain

<sup>5</sup> PhD Program, Barcelona, Spain

<sup>6</sup> Cardiovascular Epidemiology and Genetics Research Group, Hospital del Mar Research Institute (IMIM), Barcelona, Spain

<sup>7</sup> Department of Medical Science, Faculty of Medicine, University of Girona (UdG), Girona, Spain

<sup>8</sup> Vascular Health Research Group, Institut Universitari Per a La Recerca en Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Girona, Spain

<sup>9</sup> Girona Biomedical Research Institute (IDIBGI), Doctor Trueta University Hospital, Girona, Spain

- 10 Network for Research On Chronicity, Primary Care, and Health Promotion (RICAPPS), Instituto de Salud Carlos III, Madrid, Spain
- 11 Genomes For Life-GCAT Lab, CORE Program, Germans Trias I Pujol Research Institute (IGTP), Badalona, Spain
- 12 Grup de Recerca en Impacte de les Malalties Cròniques I les Seves Trajectòries (GRIMTRA), Germans Trias I Pujol Research Institute (IGTP), Badalona, Spain
- 13 Department of Preventive Medicine and Public Health, School of Medicine, Universidad Autónoma de Madrid, Madrid, Spain
- 14 IMDEA/Food Institute. CEI UAM+CSIC, Madrid, Spain
- 15 CIBER de Epidemiología y Salud Pública (CIBERESP), Instituto de Salud Carlos III, Madrid, Spain
- 16 Department of Epidemiology, Regional Health Council, IMIB-Arrixaca, Murcia, Spain
- 17 Department of Health and Social Sciences, University of Murcia, Murcia, Spain
- 18 Sub Directorate for Public Health and Addictions of Gipuzkoa, Ministry of Health of the Basque Government, San Sebastian, Spain
- 19 Epidemiology of Chronic and Communicable Diseases Group, BioGipuzkoa Health Research Institute, San Sebastián, Spain
- 20 Vascular and Renal Translational Research Group, Institut de Recerca Biomèdica de Lleida (IRBLleida), Lleida, Spain
- 21 Department of Experimental Medicine, University of Lleida, Lleida, Spain
- 22 Instituto de Salud Pública y Laboral de Navarra, Pamplona, Spain
- 23 Navarra Institute for Health Research (IdiSNA), Pamplona, Spain
- 24 Milladoiro Health Centre, Health Area of Santiago de Compostela, Santiago de Compostela, Spain
- 25 Research Methods Group (RESMET), Health Research Institute of Santiago de Compostela (IDIS), Santiago de Compostela, Spain
- 26 Department of Medicine, University Hospital of Santiago de Compostela, Santiago de Compostela, Spain
- 27 Escuela Andaluza de Salud Pública, Granada, Spain
- 28 Instituto de Investigación Biosanitaria Ibs.GRANADA, Granada, Spain
- 29 University Hospital Nuestra Señora de Candelaria, Santa Cruz de Tenerife, Spain
- 30 Department of Preventive Medicine, Faculty of Medicine, University of la Laguna, San Cristóbal de La Laguna, Spain
- 31 Instituto de Investigación Sanitaria de Aragón (IIS Aragón), Saragossa, Spain
- 32 Translational Research Unit, Hospital Universitario Miguel Servet, Saragossa, Spain
- 33 Faculty of Medicine, Universidad de Zaragoza, Saragossa, Spain
- 34 Department of Endocrinology and Nutrition, Hospital Regional Universitario de Málaga, Universidad de Málaga, Málaga, Spain
- 35 Instituto de Investigación Biomedica de Málaga (IBIMA-Plataforma Bionand), Málaga, Spain
- 36 CIBER de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), Instituto de Salud Carlos III, Madrid, Spain
- 37 Vascular Risk in Navarre Investigation Group, Department of Health, Government of Navarre, Pamplona, Spain
- 38 Servicio de Apoyo a la Gestión Clínica y Continuidad Asistencial, Hospital Universitario de Navarra, Pamplona, Spain
- 39 Department of Chronic Diseases, National Center of Epidemiology, Instituto de Salud Carlos III, Madrid, Spain
- 40 Dirección General de Salud Pública, Consejería de Sanidad, Junta de Castilla y León, Valladolid, Spain
- 41 Unitat de Suport a la Recerca Metropolitana Nord, Institut Universitari Per a la Recerca a L'Atenció Primària de Salut Jordi Gol I Gurina (IDIAPJGol), Mataró, Spain
- 42 Multidisciplinary Research Group in Health and Society (GREMAS), Institut Universitari Per a la Recerca en Atenció Primària de Salut Jordi Gol I Gurina (IDIAPJGol), Barcelona, Spain
- 43 Germans Trias I Pujol Research Institute, Can Ruti Campus, Badalona, Spain
- 44 Instituto de Investigación Sanitaria del Hospital Universitario, 12 de Octubre (imas12), Madrid, Spain
- 45 Spanish Clinical Research Network (SCReN), Madrid, Spain
- 46 Facultad de Estudios Estadísticos, Universidad Complutense de Madrid, Madrid, Spain
- 47 Department of Medicine and Surgery, University of Lleida, Lleida, Spain
- 48 Institute for Research and Innovation in Life Sciences and Health in Central Catalonia (IRIS-CC), Vic, Spain
- 49 Dirección General de Asistencia Sanitaria. Servicio Extremeño de Salud, Vic, Spain
- 50 Research Unit of Don Benito-Villanueva de La Serena Health Area, Servicio Extremeño de Salud. Villanueva de La Serena, Badajoz, Spain
- 51 University Institute for Biosanitary Research of Extremadura (INUBE), Badajoz, Spain
- 52 Subdirección General de Vigilancia en Salud Pública, Dirección General de Salud Pública, Consejería de Sanidad, Madrid, Spain
- 53 Unitat de Suport a la Recerca Barcelona, Institut Universitari Per a la Recerca a L'Atenció Primària de Salut Jordi Gol I Gurina (IDIAPJGol), Barcelona, Spain
- 54 Complejo Asistencial Universitario de Salamanca, Salamanca, Spain
- 55 Instituto de Investigación Biomédica de Salamanca (IBSAL), Universidad de Salamanca, Salamanca, Spain
- 56 Instituto de Biomedicina (IBIOMED), Universidad de León, León, Spain

- 57 Hospital Universitari Son Espases Atención Primaria Mallorca, Institut d'Investigació Sanitària Illes Balears (IdISBa), Palma, Spain
- 58 Department of Internal Medicine, Hospital Universitario Río Hortega, University of Valladolid, Valladolid, Spain
- 59 Endocrinology, Nutrition, Diabetes and Obesity Group, Health Research Institute of the Principality of Asturias (ISPA), Oviedo, Spain
- 60 Asturias Central University Hospital, Oviedo, Spain
- 61 Department of Medicine, University of Oviedo, Oviedo, Spain
- 62 CIBER de Enfermedades Raras (CIBERER), Instituto de Salud Carlos III, Madrid, Spain
- 63 Department of Medicine, School of Medicine and Health Sciences, Serra-Hunter Fellow, University of Barcelona, Barcelona, Spain
- 64 August Pi I Sunyer Biomedical Research Institute (IDIBAPS), Barcelona, Spain
- 65 Instituto de Investigación Sanitaria Biocruces, Hospital Universitario de Cruces, Barakaldo, Spain
- 66 European Reference Network On Rare Endocrine Conditions (Endo-ERN), Leiden, The Netherlands
- 67 Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU), Leioa, Spain
- 68 Musculo-Skeletal Research Unit, Hospital del Mar Research Institute (IMIM), Barcelona, Spain
- 69 CIBER de Fragilidad y Envejecimiento Saludable (CIBERFES), Instituto de Salud Carlos III, Madrid, Spain
- 70 Fundación Pública Galega de Medicina Xenómica, Centro Nacional de Genotipado (CEGEN), Santiago de Compostela, Spain
- 71 Cardiology Service, Hospital del Mar, Barcelona, Spain
- 72 Vallobín-La Florida Health Center, Principality of Asturias Health Service, Oviedo, Spain
- 73 CIBER de Fisiopatología de La Obesidad y Nutrición (CIBEROBN), Instituto de Salud Carlos III, Madrid, Spain
- 74 Department of Preventive Medicine and Public Health, University of Navarra, Pamplona, Spain
- 75 Primary Healthcare Centre Palau Solità i Plegamans, Gerència d'Atenció Primària Metropolitana Nord, Institut Català de la Salut, Barcelona, Spain
- 76 Department of Basic Medical Sciences, Serra Hunter Lecturer, University of Lleida, Lleida, Spain
- 77 Department of Biochemistry-Clinical Analysis, Puerta de Hierro Majadahonda University Hospital, Madrid, Spain
- 78 Unidad Docente Multiprofesional de Atención Familiar y Comunitaria de Toledo, Gerencia de Atención Primaria de Toledo, Servicio de Salud de Castilla-La Mancha, Toledo, Spain
- 79 Faculty of Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain
- 80 Centro Singular de Investigación en Medicina Molecular y Enfermedades Crónicas (CIMUS), Universidade de Santiago de Compostela, Santiago de Compostela, Spain
- 81 Research Group On Demography and Health, National Faculty of Public Health, University of Antioquia, Medellín, Colombia
- 82 Xàtiva-Ontinyent Department of Health, Fontanars dels Alforins Health Centre, Valencia, Spain
- 83 Instituto Aragonés de Ciencias de la Salud (IACS), Saragossa, Spain
- 84 Endocrinology and Nutrition Department, Hospital Clínico Universitario San Carlos, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos (IdISSC), Madrid, Spain
- 85 Facultad de Medicina, Medicina II Department, Universidad Complutense de Madrid, Madrid, Spain