



## OPEN ACCESS

## EDITED BY

Hui-Qi Qu,  
Children's Hospital of Philadelphia,  
United States

## REVIEWED BY

Paola Giusti-Rodriguez,  
University of Florida, United States  
Sarah S. Kalia,  
National Cancer Institute (NIH), United States

## \*CORRESPONDENCE

Melisa Naranjo Vanegas,  
✉ msnaranjo@sura.com.co

## †PRESENT ADDRESS

Melisa Naranjo Vanegas,  
Ayudas Diagnósticas SURA, Bioscience Center,  
Protección Building, Suramericana

RECEIVED 07 January 2025

ACCEPTED 04 June 2025

PUBLISHED 02 July 2025

## CITATION

Velasco Parra HM, Cardona DS, Buitrago CA,  
Vanegas MN, Hincapié SG, Jaramillo CJ,  
Cock-Rada AM, Benavides Duque C,  
Piedrahita CP, Bustamante C,  
González Niño LA, Kintle J, Kulm S, Bolli A,  
Dominico PD, Botta G, Busby GB and  
Valencia-Arango JP (2025) Clinical validation of  
an integrated risk assessment test incorporating  
genomic and non-genomic data for sporadic  
breast cancer in Colombia.  
*Front. Genet.* 16:1556907.  
doi: 10.3389/fgene.2025.1556907

## COPYRIGHT

© 2025 Velasco Parra, Cardona, Buitrago,  
Vanegas, Hincapié, Jaramillo, Cock-Rada,  
Benavides Duque, Piedrahita, Bustamante,  
González Niño, Kintle, Kulm, Bolli, Dominico,  
Botta, Busby and Valencia-Arango. This is an  
open-access article distributed under the terms  
of the [Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in  
other forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in this  
journal is cited, in accordance with accepted  
academic practice. No use, distribution or  
reproduction is permitted which does not  
comply with these terms.

# Clinical validation of an integrated risk assessment test incorporating genomic and non-genomic data for sporadic breast cancer in Colombia

Harvy Mauricio Velasco Parra<sup>1</sup>, Danny Styvens Cardona<sup>1,2</sup>,  
Cesar Augusto Buitrago<sup>1</sup>, Melisa Naranjo Vanegas <sup>1\*†</sup>,  
Sebastián Gutiérrez Hincapié<sup>3</sup>, Carolina Jaramillo Jaramillo<sup>3</sup>,  
Alicia Maria Cock-Rada<sup>1,2,3,4</sup>, Carolina Benavides Duque<sup>5</sup>,  
Clara Patricia Piedrahita<sup>6</sup>, Catalina Bustamante<sup>6</sup>,  
Leonel Andrés González Niño<sup>1,2,3</sup>, Jen Kintle<sup>7</sup>, Scott Kulm<sup>7</sup>,  
Alessandro Bolli<sup>7</sup>, Paolo Di Dominico<sup>7</sup>, Giordano Botta<sup>7</sup>,  
George B. Busby<sup>7</sup> and Juan Pablo Valencia-Arango<sup>1,2</sup>

<sup>1</sup>Personalized Medicine Group, Unidad de Bioentendimiento, Bioscience Center- Ayudas Diagnósticas SURA, Medellín, Colombia, <sup>2</sup>Data Science Department, Bioscience Center - Ayudas Diagnósticas SURA, Medellín, Colombia, <sup>3</sup>Omics Science Center, Unidad de Bioentendimiento, Bioscience Center – Ayudas Diagnósticas SURA, Medellín, Colombia, <sup>4</sup>Ayudas Diagnósticas Sura, Bioscience Center, Medellín, Colombia, <sup>5</sup>Clinical Research Group, Bioscience Center – Ayudas Diagnósticas SURA, Medellín, Colombia, <sup>6</sup>Medical imaging & AI in health SURA, Bioscience Center – Ayudas Diagnósticas SURA, Medellín, Colombia, <sup>7</sup>Allelica Inc., New York, United States

**Introduction:** Breast cancer risk arises from a complex interaction of genetic, environmental, and physiological factors. Integrating Polygenic Risk Scores (PRS) with clinical risk factors can enhance personalized risk prediction, especially in diverse populations like Colombia.

**Objective:** To evaluate the predictive performance of ancestry-specific PRS combined with clinical and imaging risk factors for breast cancer in Colombian women.

**Methods:** We developed and validated ancestry-specific PRS using diverse genetic datasets. A cohort of 1,997 Colombian women, including 510 breast cancer cases (25.5%) and 1,487 controls (74.5%), were recruited. Clinical data, such as breast density and family history, were analyzed for predictive ability using the area under the receiver operating characteristic curve (AUC). Participants were categorized into genetic ancestry groups: Admixed American, African, and European. PRS were applied to the cohort and adjusted for clinical factors to assess risk prediction.

**Results:** Breast density and family history were the strongest individual predictors, with AUCs of 0.66 and 0.64, respectively. Most participants were of Admixed American ancestry (70% of cases, 73% of controls). The combined PRS showed an Odds Ratio per Standard Deviation of 1.56 (95% CI 1.40–1.75) and an AUC of 0.72 (95% CI 0.69–0.74) when adjusted for family history. Incorporating PRS with clinical and imaging data improved the AUC to 0.79 (95% CI 0.76–0.81), significantly enhancing predictive accuracy.

**Conclusion:** Combining ancestry-specific PRS with clinical risk factors provides a more accurate approach for breast cancer risk stratification in Colombian women. These findings support the development of precise, population-specific risk assessment models.

#### KEYWORDS

polygenic risk score, breast cancer, risk stratification, prediction, epidemiology, Colombia

## Introduction

Cancer, including breast cancer, remains a leading cause of global mortality, driven by both inherited genetic predispositions and accumulated environmental exposures (Kocarnik et al., 2022; Bray et al., 2024). Polygenic inheritance describes the collective influence of numerous genetic variants and their interaction with environmental factors on traits, risk, or diseases, deviating from traditional Mendelian genetics (Crouch and Bodmer, 2020). Polygenic Risk Scores (PRS) estimate the cumulative effects of these common genetic variants on disease risk, offering a single metric that represents an individual's genetic predisposition to a particular condition. PRS are derived from Genome-Wide Association Studies (GWAS), which examine the associations of millions of genetic variants with various diseases and traits (Visscher et al., 2017; Buniello et al., 2019). While individual Single Nucleotide Variants (SNVs) generally have modest effects, their combined influence can account for a significant proportion of heritability in many disorders (Wray et al., 2007). PRS have garnered attention as valuable tools for stratifying individuals into risk categories, thereby enabling targeted medical interventions.

Notably, the risk associated with PRS in the upper percentiles of a population distribution often mirrors well-established clinical risk factors, such as elevated LDL cholesterol for cardiovascular disease or high breast density for breast cancer (Busby et al., 2020; Mars et al., 2022; Mars et al., 2020; Levin and Rader, 2020; Ma and Zhou, 2021; Vachon et al., 2015; Cupido et al., 2021). Despite their promise, most PRS models are developed using datasets from individuals of European ancestry, creating challenges when applied to populations with different genetic backgrounds, such as Latinos (Lambert et al., 2019; Martin et al., 2019). Differences in allele frequencies, linkage disequilibrium patterns, and effect size may impact the predictive performance of PRS in non-European populations (Fatumo et al., 2022). Therefore, it is essential to validate PRS within the target populations for which they are intended, particularly those with diverse and admixed genetic ancestries (Gri et al., 2019; Park et al., 2017).

Breast cancer is the most prevalent malignancy among women worldwide, representing a quarter of all malignant tumors and the leading cause of cancer-related mortality in developing countries (Harbeck and Gnant, 2017). In Colombia, 17,018 new cases of breast cancer and 4,752 related deaths were reported in 2023, accounting for 14.5% of cancer incidence and 8.4% of cancer mortality, respectively (World Health Organization, 2023). While reductions in breast cancer mortality have been observed in some countries due to improved screening and rapid therapeutic interventions, these strategies primarily focus on early detection and treatment rather than prevention (Viaña González, 2020).

Mammography remains a cornerstone for breast cancer screening, yet its role as a standalone risk assessment tool is limited. Incorporating genetic testing and other innovative methodologies can enhance the accuracy of risk prediction, aiding both prevention and early diagnosis. This underscores the need for integrated approaches combining genetic, clinical and imaging data to refine risk stratification and improve outcomes.

This study evaluates the clinical utility of combining five ancestry-specific PRS with clinical and mammographic data in a cohort of Colombian women with sporadic breast cancer. By focusing on a genetically diverse population, this research contributes to the growing evidence supporting the integration of PRS into clinical risk models, particularly in underrepresented populations such as Latin Americans. The findings aim to bridge the gap in precision medicine applications from non-European populations, with implications for improving breast cancer prevention and care in Colombia and beyond.

## Methods

### Study design and setting

This study was an observational case-control study, called “Soy Generación”, which included data collected between July and December 2022 in 5 cities in Colombia (Bogotá, Medellín, Barranquilla, Bucaramanga and Cali). The study included women (ages 40–80 years) with and without confirmed breast cancer diagnosis.

### Study population

The case group consisted of women (aged 40–65 years) who had been diagnosed with breast cancer during the study period and who were part of “Tiempo para ti” (Time for you—in English) program. This program offers integrated care for breast cancer patients through risk factor evaluation and clinical monitoring, provided by SURA Colombia.

The control group was composed of women (aged over 65 years old) who were insured by SURA Colombia, and who had no history of breast cancer. These patients were identified through negative mammography results. They were included to provide a comparison for assessing risk factors in a population without breast cancer, serving as a baseline for evaluating genetic and environmental influences on the development of the disease. Informed consent was obtained for all participants.

## Ancestry estimation and principal component analysis (PCA)

We applied a consistent ancestry estimation and PCA pipeline across all individual-level datasets used in this study, including PRS Training, Testing, and Validation cohorts. Genetic ancestry of individuals was estimated using 99,561 single nucleotide variants (SNVs) and the iAdmix method (Bansal et al., 2015), leveraging the 1000 Genomes Project reference dataset. Individuals were classified into five Super-populations: AFR (African), AMR (American), EAS (East Asian), EUR (European), and SAS (South Asian). Principal component analysis (PCA) was then performed by projecting individual-level genotype data onto a pre-computed PCA space derived from the 1000 Genomes Project dataset. This analysis yielded ancestry proportions from the five 1,000 Genomes Superpopulations and generated up to 10 principal components (PCs), which were used as covariates in GWAS analyses and downstream PRS modeling.

## Description of cohort used in PRS development

We employed a multi-stage study design integrating diverse population cohorts to develop, train, test, and validate ancestry-specific polygenic risk scores (PRS) for breast cancer. The foundation of PRS construction relied on three ancestry-specific genome-wide association studies (GWAS), which provided effect size estimates for score development. These included datasets representing East Asian, European, and African ancestries, derived from Biobank Japan (Sakaue et al., 2021; Michailidou et al., 2017), and the internally conducted Ghana Breast Health Study (Brinton et al., 2017), respectively.

Individual-level genotype data from five additional multi-ancestry cohorts were used for training and testing the PRS: the UK Biobank (UKB) (Bycroft et al., 2018), the Multi-Ethnic Study of Atherosclerosis (MESA, dbgap accession: phs000209. v2. p1) (Bild et al., 2002), the San Francisco Bay Area Latina Breast Cancer Study (SFBALCS, dbgap accession: phs000912. v1. p) (Michailidou et al., 2017), the Estrogen Receptor Negative Breast Cancer in African American Women study (CIDR, dbgap accession: phs000669. v1. p1) and High-Risk Breast Cancer GWAS (HRBC, dbgap accession: phs000929. v1. p1). These cohorts were selected for their diverse representation of ancestral backgrounds and phenotypic richness.

For external validation, we used an independent cohort of 1,997 admixed American women from SURA Colombia, selected based on clinical and demographic inclusion criteria to reflect the target population for clinical implementation.

## GWAS discovery datasets

GWAS summary statistics used for PRS construction were sourced from three datasets. Two were obtained from previously published studies: one from Biobank Japan, which included individuals of East Asian ancestry, and one from Michailidou et al. (2017), which included primarily European ancestry individuals. A third GWAS was conducted internally on the

Ghana Breast Health Study cohort using the fastGWA-GLMM framework (Jiang et al., 2021; Nyante et al., 2019), applied to individual-level genotype data. For this African ancestry GWAS, a genetic relationship matrix (GRM) was computed based on approximately 100,000 linkage disequilibrium (LD)-independent variants, with age and the first four principal components of ancestry included as covariates.

Summary statistics underwent quality control and filtering steps prior to PRS development. Variants with minor allele frequency (MAF)  $\leq 1\%$  were excluded based on ancestry-specific allele frequencies from the 1000 Genomes Project. In the case of the largest European GWAS, fine-mapping was conducted using PolyFun (Weissbrod et al., 2020), with default parameters, with either one or ten causal variants selected per LD block and three different ancestry-specific LD maps (AFR, EUR, EAS) sourced from the PolyFun repository. Ancestry-specific LD maps were sourced from the PolyFun repository to improve the precision of effect size estimates. These processing steps yielded three ancestry-specific sets of GWAS summary statistics and six fine-mapped datasets, which were used in subsequent PRS construction. These filtering and fine-mapping procedures resulted in three sets of ancestry-specific summary statistics and six ancestry-specific fine-mapped datasets, which were subsequently used in the generation of PRS (Supplementary Table S1).

## Framework for PRS development

In this study, we used Allelica's DISCOVER v1.3 software (Busby et al., 2023), a flexible PRS development platform that enables the integrated implementation of a range of existing polygenic score construction algorithms. For this analysis, we applied two distinct strategies: a single-ancestry PRS, developed using the stacked clumping and thresholding (SCT) method (Privé et al., 2019) method optimized in a European-ancestry Training dataset, and a trans-ancestry PRS, built using PRS-CSx (Ruan et al., 2022), a Bayesian approach that integrates GWAS summary statistics across multiple ancestral groups while accounting for linkage disequilibrium. DISCOVER generated a range of candidate scores, selected the best-performing PRS based on independent ancestry-specific Training datasets, and confirmed its predictive performance in separate Testing dataset. This dual framework, allowing for both population-specific optimization and improved cross-ancestry generalizability, is described in detail in the following sections.

## Single-ancestry PRS development

Due to its large sample size and increased statistical power, the European GWAS summary statistics were used to construct the single-ancestry polygenic risk score. We applied the SCT method (Privé et al., 2019; Ruan et al., 2022) within Allelica's DISCOVER PRS development platform to generate a final panel of variants, risk alleles and effect sizes for downstream assessment. SCT generates ~124,000 PRS panels by filtering the variants from the original summary statistics using a variety of different parameters related to Linkage Disequilibrium (LD) and significance thresholds. The

TABLE 1 Datasets used to develop novel ancestry specific PRSs for breast cancer.

Cohort	Name	Country	Cases	Controls	Ethnicity/ Ancestry	Model development stage
BBJ	Biobank Japan	Japan	5,551	89,371	EAS	PRS Discovery
BCAC	Breast Cancer Association Consortium	Global	122,977	105,974	EUR	PRS Discovery
GHANA	Ghana Breast Cancer study	Ghana	869	1,618	AFR	PRS Discovery
CIDR	Centre for Inherited Disease Research	United States, Barbados, Nigeria	1,681	2,085	AFR	PRS Training <sup>a</sup>
MESA	Multi-Ethnic Study of Atherosclerosis	United States	76	2,947	AFR, EAS, EUR	PRS Training <sup>a</sup>
SFBALCS	San Francisco Bay Area Latinas Cancer study	United States	622	60	AMR	PRS Testing
HRBC	High Risk Breast Clinic	United States	2,343	2,059	EUR	PRS Training <sup>a</sup>
UKBB	United Kingdom Biobank	United Kingdom	143	2,607	AFR, EAS, SAS	PRS Training <sup>a</sup>
UKBB	United Kingdom Biobank	United Kingdom	6,269	71,725	EUR	PRS Training <sup>b</sup>
UKBB	United Kingdom Biobank	United Kingdom	14,641	164,309	AFR, EAS, EUR, SAS	PRS Testing <sup>t</sup>
Total			155,172	444,755		

<sup>a</sup>Training dataset for single- and trans-ancestry PRSs.

<sup>b</sup>Training dataset for single-ancestry PRSs.

resultant 124,000 different panels were then used in a penalized logistic regression to generate the optimal lineal combination of scores which maximizes outcome discrimination. The single-ancestry PRS training dataset, comprising both genotype and phenotypic data, was used to optimize penalization hyperparameters.

## Trans-ancestry PRS development

To develop polygenic risk score (PRS) panels, we combined nine sets of prepared genome-wide association study (GWAS) summary statistics with the PRS-CSx (Ruan et al., 2022), algorithm (Supplementary Table S2), which was implemented using Allelica's DISCOVER software. PRS-CSx utilizes a high-dimensional Bayesian regression framework that continuously shrinks the effect of each variant across multiple trans-ancestry GWAS. This allows for more accurate estimation of genetic risk by integrating information from various ancestry groups.

Given that PRS-CSx can simultaneously consider summary statistics from multiple ancestry groups to develop a consensus posterior effect, we employed several strategies to combine our available summary statistics. After combining the datasets, we ran the analyses with PRS-CSx using a range of hyperparameter values. In total, we generated 15 distinct PRS panels (Supplementary Table S2). For each GWAS combination, PRS-CSx was run 4 times varying the value of the global shrinkage parameter PHI ( $10^0$ ,  $10^{-2}$ ,  $10^{-4}$ ,  $10^{-6}$ ), as suggested within PRS-CSx documentation. All the other parameters were set to default values.

## PRS training

We utilized individual-level genotype data from five cohorts to train and test the PRS: UKB, MESA, HRBC, and SFBALCS, and CIDR.

Details on the size and sources of these cohorts are provided in Table 1. Training of the single-ancestry PRS was performed in 77,994 European women from the first release of UKB. This allowed us to identify the best linear combination of 124,000 different panels, generated through a range of clumping/thresholding hyperparameters. The best performing ancestry-specific PRS among single- and trans-ancestry PRS was identified within the following cohorts: HRBC as European, MESA as multi-ancestry, and CIDR as African. Cohorts were combined together and best performing PRS were identified in each ancestry-specific subgroup. This approach allowed us to identify the most robust PRS for each population group. PRS performance were quantified by means of the PRS Odd ratio per Standard deviation (ORxSD) in logistic regression models with PRS as predictive variable, Breast Cancer as dependent variable, and age, the first four principal components of ancestry (PC1-4) and family history of Breast Cancer (when available in the dataset) as control variables.

## PRS testing

To assess the generalizability and predictive performance of the trained PRS, we used independent samples from the second release of UKB and the SFBALCS cohorts. Testing was conducted in ancestry-stratified subgroups corresponding to the populations included in the training step. PRS performances were quantified in each ancestry-specific group according to the same methodology described above. Age, PC1-4 and family history were used as control variables in the logistic regression models.

## PRS validation

For the Colombian validation dataset, we preselected 1997 patients from a total of 20,666 SURA Colombia affiliates

based on adherence to specific inclusion criteria, which were validated using the company's TERADATA SQL Assistant. Inclusion criteria included: females aged 40–80 years, living status, health conditions (including breast cancer), and mammography reports performed between 2017 and 2021. Additionally, breast cancer patients who met genetic testing criteria according to NCCN guidelines were tested with a Next-Generation Sequencing panel of 30 hereditary cancer genes. Patients with confirmed pathogenic/likely pathogenic variants were excluded. Of the 2,618 patients who met these criteria, 2055 consented to participate in the study. After validation of inclusion and exclusion criteria, 58 patients were excluded (see flowchart in [Supplementary Figure S1](#)). The final clinical validation cohort included 1997 patients: 1,487 controls and 510 cases. Baseline demographics and clinical data were collected from the electronic medical record (EMR) database.

DNA was extracted from blood samples collected from individuals in the validation cohort using the Applied Biosystems MagMAX DNA Multi-Sample Ultra 2.0 kit system, operated on the KingFisher Flex Magnetic Particle Processor 96DW platform. DNA quantity was determined using the Qubit 3 machine. Genotyping was performed using the Illumina Infinium HTS (High-Throughput Screening) Global Screening Array v3.0 (GSAv3.0) platform on the Illumina iScan Platform, capable of examining approximately 750,000 single-nucleotide variants (SNVs) and copy number variants (CNVs) per sample. The end-to-end analysis workflow was handled by the Biodiscovery NxClinical SW platform. Raw genotype data were imputed into a joint callset of 82,579,889 genome-wide variants using the BEAGLEv5.4 imputation tool and the 1000 Genomes Project v3 Imputation Reference Panel.

## Statistical analysis

Descriptive statistics (mean, median, percentiles, proportions) were used, with associations assessed using Fisher's exact test or Chi-square ( $p < 0.05$  considered significant). Differences between groups were tested using Student's t-test, Mann-Whitney U, and Kruskal–Wallis tests, with normality assessed by the Shapiro–Wilk test. Polygenic score differences across ancestries were evaluated using Student's t-test or Mann-Whitney U, as appropriate. Model applicability was assessed by calculating Odds Ratios (OR), standard deviation (SD), p-values, and the Area Under the Receiver Operating Characteristic (AUC) curve.

## Estimating risk

Polygenic risk scores (PRS) were calculated as the additive sum of effect sizes for risk alleles:

$$PRS = \sum_{i=1}^n \beta_i x_i$$

Where  $\beta_i$  is the effect size (logarithm of the OR) for SNP $_i$  and  $x_i$  is the allele dose (0, 1, or 2). PRS values were adjusted using principal components from the 1000 Genomes Project. Relative risk was

estimated using the z-score of the PRS value in an ancestry-matched reference distribution:

$$Relative\ Risk = \exp(Z * \log(OR \times SD))$$

Relative risks were averaged across five PRS, weighted by ancestry proportions. After obtaining the relative risk values for each patient, we estimated the absolute 10-year risk of breast cancer using obtained epidemiological data for specific localities in Colombia (Medellin, Bogota, Cali, Barranquilla) collected for the purposes of this study ([Supplementary Figure S2](#)) and generated recalibrated risk curves for each using the approach of [Darst et al. \(2021\)](#). The approach constrains the PRS-specific absolute risks for a given age to be equivalent to the age-specific incidence for the entire population. Therefore age-specific incidence rates were calculated to increase or decrease based on the PRS category estimated risk and the proportion of the population within the PRS category. Note that both incidence and survival are a function of time (age) and are calculated for a given percentile of the PRS ( $k$ ).

Absolute 10-year breast cancer risk was estimated using Colombian epidemiological data, recalibrated with risk curves based on [Darst et al. \(2021\)](#). Cumulative 10-year risks were calculated using the following formula (34):

$$AR_k(t) = \sum_0^t P_{ND}(t) S_k(t) I_k(t)$$

Where  $P_{ND}(t)$  is the probability of survival from other causes,  $S_k(t)$  is survival from breast cancer, and  $I_k(t)$  is breast cancer incidence for risk category  $k$ .

## Reclassification methods

Reclassification was performed to evaluate the impact of integrating Polygenic Risk Scores (PRS) with clinical risk factors on breast cancer risk stratification. The reclassification process involved comparing the predicted risk categories before and after the inclusion of PRS. We used the Net Reclassification Improvement (NRI) and Integrated Discrimination Improvement (IDI) metrics to quantify the improvement in risk prediction. Participants were categorized into risk strata based on their predicted probabilities, and changes in these categories were analyzed to assess the effectiveness of the PRS integration. To further assess the impact of PRS on screening recommendations, we estimated the age at which each woman's risk was equal to that of a 50-year-old woman, which is the current recommended age to start mammography screening in Colombia. This estimation was based on the PRS percentile assigned to each woman. We then compared this estimated age to the actual age of diagnosis for women diagnosed under the age of 50.

## Incorporating breast cancer PRS into risk models

Logistic regression was used to estimate effect sizes of risk factors, adjusting for ancestry and family history components.

TABLE 2 Clinical and Phenotypic characteristics of the patients in "Soy Generación" Study.

Characteristic	Case, N = 510 <sup>a</sup>	Control, N = 1,487 <sup>a</sup>	p-value <sup>b</sup>
Age	55 (49, 61)	69 (67, 71)	<0.001
BMI when diagnosed	26 (24, 30)	0 (0, 0)	NA
Study entry weight	66 (59, 74)	63 (56, 70)	<0.001
Age of menarche	13 (12, 14)	13 (12, 15)	<0.001
History of at least one pregnancy	433 (85%)	1,251 (84%)	0.7
Age at first deliver	24.0 (20.0, 28.0)	23.0 (20.0, 27.0)	0.3
Age at Menopause	48 (45, 51)	49 (45, 52)	0.2
Hormone Replacement Therapy			0.7
Estrogen treatment only	22 (81%)	19 (76%)	
Estrogen and Progestogen treatment	5 (19%)	6 (24%)	
Hormone replacement therapy			<0.001
2–5 years (%)	15 (56%)	3 (12%)	
Unknown	1 (3.7%)	3 (12%)	
More than 5 years	3 (11%)	16 (64%)	
Less than 2 years	8 (30%)	3 (12%)	
Breast Density Residual			<0.001
A	41 (8.0%)	302 (20%)	
B	218 (43%)	866 (58%)	
C	204 (40%)	288 (19%)	
D	47 (9.2%)	31 (2.1%)	

<sup>a</sup>Median (IQR); Mean (SD); n (%).

<sup>b</sup>Wilcoxon rank sum test; Welch Two Sample t-test; Pearson's Chi-squared test; Fisher's exact test.

<sup>c</sup>BMI: body mass index.

Discriminatory capacity was assessed using the ROC curve. The PRS model included:

$$Ca_{breast} = \alpha_0 + \alpha_1 PGS + \alpha_2 edad + \alpha_3 F_1 + \dots + \alpha_k PC_1 + \dots + \alpha_{k+l} PC_l + \epsilon$$

In the given context,  $Ca_{breast}$  indicates a dichotomous variable (1: case, 0: control),  $F_i$  represents the literature-reported factors associated with cancer, and  $PC_l$  corresponds to the principal component  $l$  associated with ancestry.

In addition to evaluating single-ancestry and trans-ancestry PRS separately, we developed a combined PRS model by integrating the five ancestry-specific PRS into a weighted linear model. Weights were derived from the ancestry proportions estimated for each individual through iAdmix, as previously described. The combined PRS was calculated as a weighted average of each ancestry-specific PRS, where the contribution of each score was proportional to the individual's inferred ancestry composition. This approach aimed to maximize predictive performance in admixed populations by leveraging information from multiple ancestries. The combined PRS was subsequently included as a predictive variable in logistic regression models, alongside age, principal components of ancestry (PC1-PC4), and family history of breast cancer.

Discriminatory performance was assessed using the area under the received operating characteristic curve (AUC) and compared with single-ancestry and trans-ancestry models.

## Results

### Clinical and phenotypic characteristics

A total of 1,997 women were recruited for the study, of whom 510 (25.5%) were sporadic breast cancer cases, and 1,487 (74.5%) were assigned to the control group. Following laboratory-based revalidation of the data and variables collected during recruitment and project phases, clinical validation led to the exclusion of 21 individuals from the control group and 37 from the case group (Supplementary Figure S1). The mean age of the cases at the time of study entry was 55 years, whereas the age of the controls was 69 years. Clinical and phenotypic characteristics, known risk factors for breast cancer, were compared between cases and controls.

Significant differences were observed for several variables. Study entry weight was higher in cases than controls (median: 66 kg vs. 63 kg,  $p < 0.001$ ), and age at menarche was slightly lower among cases (median: 12 vs. 13,  $p < 0.001$ ). Duration of hormone

TABLE 3 OR\*SD of the SURA Population and other populations using TARGET DATA.

Genetic ancestry	PRS	ORxSD	Total samples in testing dataset	Algorithm used and panel
AMR	Allelica_BC_AMR_2022	1.50 (1.17–1.93)	368	PRS-CSx - AFR-F10_AFR-G_EUR-F10_EUR-G (PHI 10 <sup>-4</sup> )
AFR	Allelica_BC_AFR_2022	1.45 (1.31–1.60)	3,379	PRS-CSx - EAS-F1_EUR-F1 (PHI 10 <sup>-4</sup> )
EUR	Allelica_BC_EUR_2020	1.75 (1.73–1.77)	172,588	SCT - single-ancestry PRS
SAS	Allelica_BC_SAS_2022	1.61 (1.43–1.84)	1,764	PRS-CSx -AFR-F1_EAS-F1 (PHI 10 <sup>-4</sup> )
EAS	Allelica_BC_EUR_2020	1.71 (1.49–1.95)	1,219	SCT - single-ancestry PRS
SURA/Allelica	multiple	<u>1.56 (1.40–1.75)</u>	1,997	multiple

Details of polygenic risk scores derived from DISCOVER. The odds ratio per standard deviation (ORxSD) and total sample size of the Testing cohort are listed. The final row describes the performance of the above five ancestry-specific scores, combined using an ancestry-specific weighted average, across all individuals from the Colombia PRS, Validation cohort. For each ancestry-specific, the algorithm used and hyperparameter value is shown. See [Supplementary Tables S1,S2](#) for further details.

replacement therapy (HRT) use also differed between groups ( $p < 0.0001$ ), with higher proportion of controls having used HRT for more than 5 years. Conversely, no statistically significant differences were found for several reproductive factors, including history of at least one pregnancy, age at first delivery, age at menopause, or type of hormone replacement therapy regimen ([Table 2](#)).

Mammographic breast density exhibited a statistically significant relationship with breast cancer ( $p < 0.001$ ) (see also [Supplementary Table S3](#)). The breast density distribution according to the ACR BI-RADS classification, showed a higher proportion of B, C and D types in the cases compared to the controls: 43% of cases were classified as type B compared to 19% of controls, 40% as type C of cases compared to 19% of controls, and 9.2% as type D of cases compared to 2.1% in controls ([Table 2](#)). When analyzing the impact of the clinical variables detected on their ability to discriminate for breast cancer, we observed areas under the curve ranging from 0.51 to 0.66, and for the multivariate model it was equal to 0.76. The variables of breast density and family history are the ones with the highest predictive power on their own with an AUC of 0.66 and 0.64 respectively.

## Development and testing of 5 ancestry-specific PRS for breast cancer

We evaluated 15 trans-ancestry PRS panels developed with PRS-CSx alongside a single-ancestry PRS across five genetic ancestry-specific cohorts. The best-performing PRS in each cohort was identified based on the highest Odds Ratio per Standard Deviation (ORxSD) ([Table 3](#)). The ORxSD ranged from 1.45 (1.31–1.77) for Allelica\_BC\_AFR\_2022 in the African ancestry group to 1.75 (1.73–1.77) in the European ancestry group. The Allelica\_BC\_EUR\_2020 PRS, previously developed using the Stacked Clumping and Thresholding algorithm ([Patel and Khera, 2022](#)), was the best-performing PRS in the European and East Asian ancestry groups. An ancestry adjustment and recalibration model revealed that 85% of study participants self-identified as admixed.

Ancestry classification using iAdmix and 1000 Genomes Project superpopulations categorized participants into major ancestry groups ([Figure 1A](#)). Most individuals were classified as Admixed American (AMR) (84% of cases and 91% of controls), followed by

European ancestry (14% of cases and 6% of controls). African ancestry was the least represented (1.5% of cases and 2.1% of controls). Significant differences in PRS distributions between cases and controls were observed across ancestry groups ([Figure 1B](#); [Supplementary Tables S4,S5](#)).

## Predictive performance of the multi-ancestry PRS in the colombian population

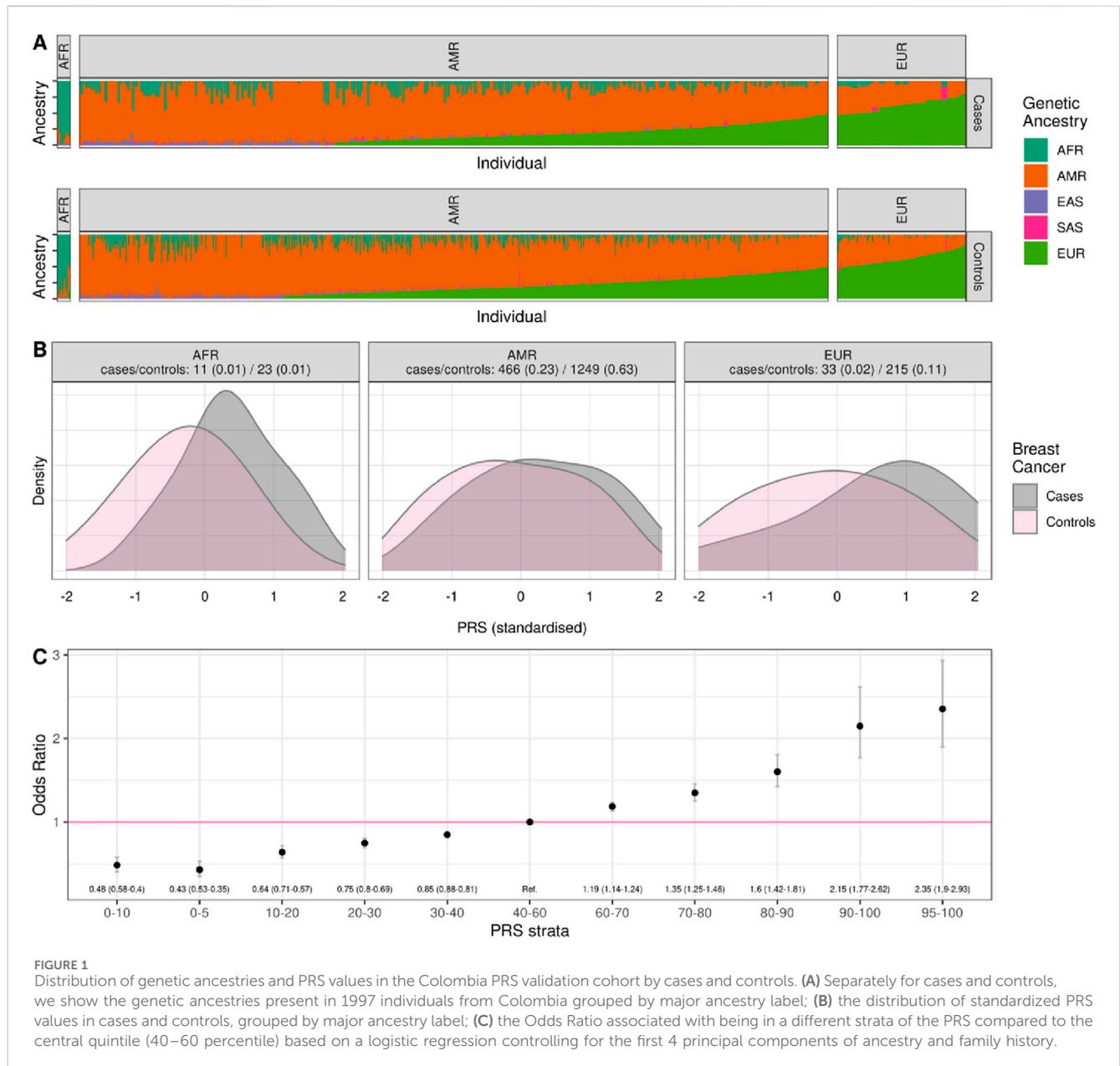
The ORxSD of the multi-ancestry PRS applied to the Colombian population was 1.56 (95% CI: 1.39–1.74), indicating a 1.56-fold increased risk of breast cancer per standard deviation in the PRS. This result was consistent with values reported for European (1.75; 95% CI: 1.73–1.77) and African (1.45; 95% CI: 1.31–1.60) populations but exceeded those from smaller American cohorts (1.50; 95% CI: 1.17–1.93) ([Table 3](#)).

PRS values were used to estimate relative and absolute genetic risk. Women in the top decile of the PRS distribution had a 2.15-fold (95% CI: 1.77–2.62) increased risk compared to the reference quintile, while those in the highest 5% exhibited a 2.35-fold (95% CI: 1.90–2.93) increased genetic risk. These findings underscore the efficacy of PRS in stratifying individuals by genetic susceptibility to breast cancer ([Figure 1C](#); [Supplementary Table S6](#)).

Adjusted for family history and ancestry, the AUC of the PRS was 0.72 (95% CI: 0.69–0.74). When combining clinical, imaging, and genomic data, the AUC increased to 0.79 (95% CI: 0.76–0.81), with genetic factors contributing approximately 8% to the prediction when family history alone was considered, and 3% when all factors were included ( $p < 1.4e-12$ ) ([Figure 2](#)). Comparisons with datasets from other studies revealed similarities in PRS metrics ([Supplementary Table S7](#)). Despite a relatively modest AUC, PRS-based patient reclassification demonstrated clinical relevance, particularly when integrated with modalities like BI-RADS.

## Reclassification results

The reclassification analysis demonstrated significant improvements in risk prediction when PRS was integrated with clinical factors. The NRI indicated that 15% of participants were



correctly reclassified into higher risk categories, while 10% were correctly reclassified into lower risk categories. The IDI showed an overall improvement in the discrimination ability of the model, with an increase in the AUC from 0.72 to 0.79. These results underscore the utility of PRS in enhancing breast cancer risk stratification.

PRS stratification categorized participants into four risk groups based on the percentile distribution in the control population: low-risk (<30th percentile), standard-risk (30th–69th), moderate-risk (70th–89th), and high-risk (≥90th percentile). Among breast cancer cases, 8.6% were classified as high-risk, 17.8% as moderate-risk, 33.7% as standard-risk, and 39.8% as low-risk. In the control group, 9.7% were high-risk, 12.8% moderate-risk, 31.3% standard-risk and 46.3% low-risk.

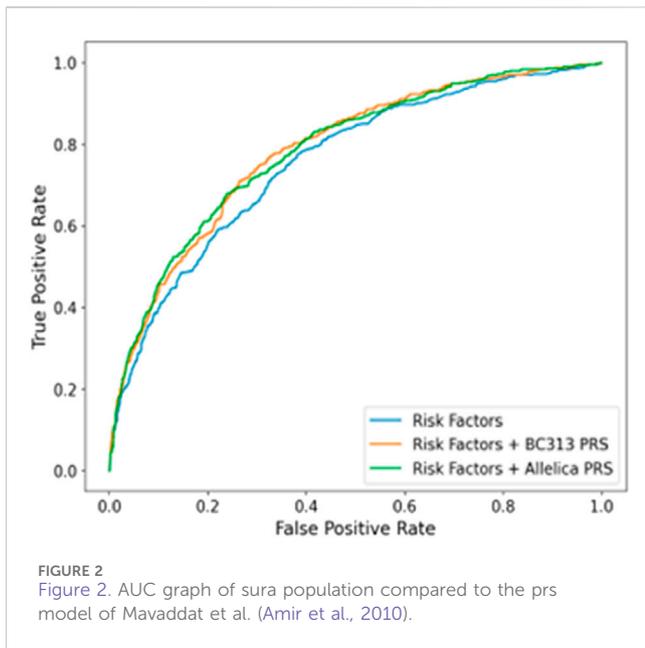
Among the 231 breast cancer cases diagnosed before age 50, 104 (45.0%) had an estimated screening eligibility age lower than their diagnosis age. On average, high-risk patients would have qualified

for screening 15.2 years earlier, moderate-risk patients 8.4 years earlier, and low-risk individuals approximately 5.3 years later than the standard screening age of 50.

Cumulative breast cancer risk was estimated based on observed incidence and mortality data across five Colombian regions (Figure 3). Risk was highest in Barranquilla and Medellín, with individuals in the highest PRS decile showing more than double the lifetime risk compared to those with average PRS values, and a sevenfold increased risk compared to the lowest PRS group (Figure 4). Ten-year risk peaked around age 60 in all regions.

## Discussion

Few studies have examined the effect of polygenic risk scores (PRS) in Latin American or Hispanic populations (Patel and Khera,



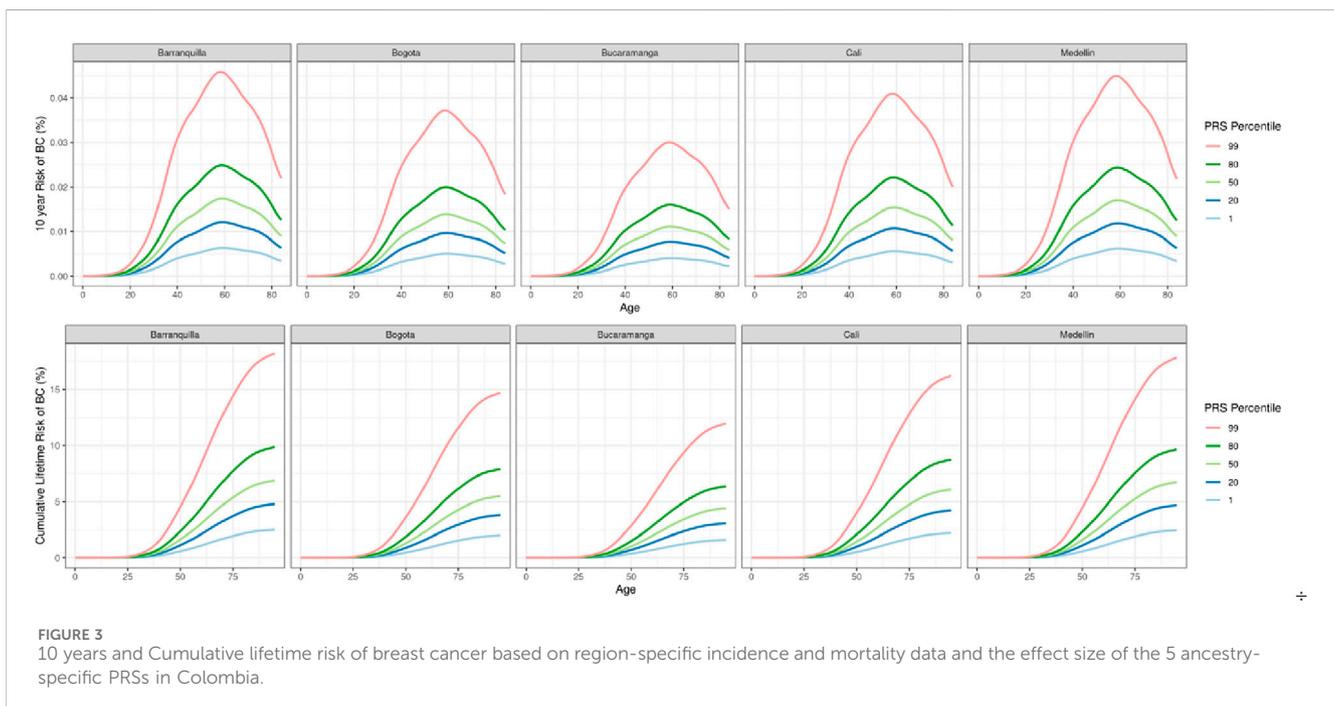
2022; Gaziano et al., 2016; Duncan et al., 2019; McArdle et al., 2021; Félix et al., 2023). The limited genomic characterization of these admixed populations poses challenges in validating PRS-based strategies, particularly for complex diseases such as breast cancer (Félix et al., 2023). This study provides clinical validation of PRS in a highly admixed Colombian cohort. Despite the diverse ancestry composition, the observed OR per standard deviation and AUC values suggest comparable predictive performance to European datasets, reinforcing the potential to extend genomic risk prediction to underrepresented populations (Sweeney et al., 2008; Martínez et al., 2017). Our findings were additionally supported by

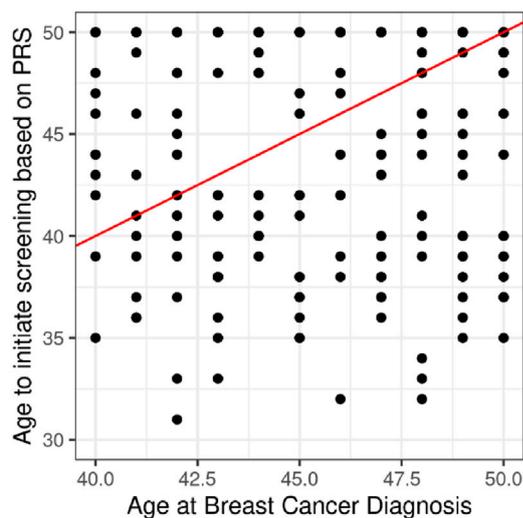
comparisons with external biobanks, showing consistent risk discrimination.

Multiple PRS studies have focused on breast cancer risk, largely in European and North American cohorts. Mavaddat et al. developed a PRS with 313 SNPs optimized for estrogen receptor (ER)-specific breast cancer, reporting an ORxSD of 1.61 (95% CI: 1.57–1.65) and an AUC of 0.630 (95% CI: 0.628–0.651) in a European cohort (Mavaddat et al., 2019). Despite including only 2.2% Hispanic participants, these results are comparable to our Colombian population. Similarly, Allman et al. observed improved ORxSD values when integrating 71 SNPs with the Gail or IBIS models in Hispanic women, findings corroborated in their later work (Allman et al., 2021; Allman et al., 2015). Shieh et al. demonstrated that integrating classical risk factors with PRS improved AUC values in Latin American women, aligning closely with our findings (Shieh et al., 2016; Shieh et al., 2020).

Triviño et al. reported PRS performance in a Spanish cohort, yielding an ORxSD of 1.41 (95% CI: 1.24–1.61) and an AUC of 0.8 (95% CI: 0.77–0.83), emphasizing ancestry-related nuances (Trivino et al., 2020). Ray et al. confirmed adequate discrimination of PRS-based models in a Colombian cohort, although without reporting overall ORxSD (Ray et al., 2017). Collectively, these studies highlight the robust predictive power of our PRS, which appears well-calibrated for the Colombian population.

Breast cancer prevalence differs across populations due to variations in reproductive behavior (Sweeney et al., 2008), socioeconomic factors (Martínez et al., 2017), and ancestry (Hines et al., 2017). The ancestry composition of our cohort aligns with meta-analyses of Latin Americans but differs from the higher European ancestry proportions reported by Ruiz et al. and Rodrigues de Moura et al. in the central regions of Colombia and





**FIGURE 4**  
Projection of cases under 50 years of age. We applied the PRS to each woman to assign them a PRS percentile and from this we estimated the age at which her risk was equal to that of a 50-year-old woman and define this as the age to start screening. Of the 231 cases diagnosed under the age of 50, 104 (45.0%) have an estimated age of screening that is younger than the age of diagnosis.

Latin America (Ruiz et al., 2022; Rodrigues de Moura et al., 2015). Interestingly, Rey et al. found that Colombian women with higher American and lower European ancestry proportions had elevated odds of developing breast cancer (Rey-Vargas et al., 2022).

A significant risk of breast cancer is associated with germline pathogenic/likely pathogenic (P/LP) variants in high- or moderate-penetrance cancer-predisposing genes (Rojas and Stuckey, 2016). Although carrier status of P/LP variants was not systematically assessed across the entire cohort, individuals identified through NCCN clinical criteria were excluded from the analysis.

Other known risk factors for breast cancer include family history, particularly in first-degree relatives (Collaborative Group on Hormonal Factors in Breast Cancer, 2001), hormonal factors, lifestyle (Mcperson et al., 2001), and breast density (Amir et al., 2010). All these variables, except the age of first birth, showed significant differences in women with breast cancer in our cohort. Alongside traditional risk prediction models such as BRCAPRO, Gail, BCSC, or Tyrer-Cuzick, there is growing interest in incorporating genomic risk factors like PRS to refine clinical risk stratification strategies.

The debate on optimal breast cancer screening strategies has shifted toward more personalized approaches. Wolfson et al. argued that PRS-based stratification may be more impactful than focusing solely on rare deleterious variants or family history (Wolfson et al., 2021). Ho et al. demonstrated the superiority of combinatorial models that include PRS, family history, and deleterious variant detection, achieving a two-year AUC of 0.622 (0.608–0.636) (Ho et al., 2023; Morabia et al., 2003; He et al., 2006; Sims et al., 2008; Boora et al., 2016; Versmissen et al., 2015; Li et al., 2019; Kryukov et al., 2009).

Notably, PRS stratification identified significant proportions of women under 50 years of age who would benefit from screening—an improvement over Colombia's current guidelines, which recommend mammography starting at age 50. High-risk patients in our cohort would have benefited from screening an average of 15 years earlier, with moderate-risk patients beginning screening approximately 8 years earlier. This underscores the value of PRS in personalizing screening strategies to better address early-onset cases. Our team is developing follow-up protocols tailored to patient risk levels, incorporating mammography, ultrasound, and MRI with annual monitoring starting at age 40. While these protocols align with international debates, further evaluation of clinical utility is underway (Ray et al., 2017; Eby, 2017).

The PRS presented here offers a significant advancement in breast cancer risk stratification for Colombian women. By integrating clinical, imaging, and genomic data, this model provides a comprehensive framework for personalized secondary prevention and anticipatory health management across diverse populations. Future work should prioritize the inclusion of underrepresented ancestries—such as Afro-Colombian and Indigenous populations—to improve the calibration and equity of PRS models in Latin America. Although most PRS are trained on European ancestry data, local validation is essential across all ancestral backgrounds presented in admixed populations such as Colombia (Bryc et al., 2010). These findings provide a foundation for incorporating genomic medicine into regional screening programs while paving the way for novel population management strategies and will help reduce disparities in predictive accuracy and ensure equitable implementation of genomic tools (Clarke and Thirlaway, 2011; Sweet et al., 2017; Roux et al., 2022; Brooks et al., 2021; French et al., 2020).

## Conclusion

This study provides a PRS and data consistent with findings from European populations, with clinical validation values demonstrating greater precision than several reports analyzing Latin populations, including those with Iberian or Colombian ancestry. At the time of this publication, it represents the most comprehensive Colombian case-control association study, with the best representation of the country's multi-territoriality and multi-ancestry, focused on the clinical validation of a risk stratification test that integrates clinical, imaging, and genomic data. This unique model offers accurate risk stratification for sporadic breast cancer and holds promise for clinical applications in personalized secondary prevention, aiming to anticipate health management needs for patients and populations across Colombia.

## Limitations

Although this study focuses on analyzing the risk of sporadic breast cancer, there remains a possibility that some cases—clinically not meeting NCCN criteria for hereditary breast cancer—may carry pathogenic or likely pathogenic variants. Our PRS, based on an array platform, was not

designed to detect such variants, representing a potential limitation in the scope of genetic information captured.

One limitation of our study is the variability in the completeness of family history information among participants. Approximately 30% of the participants had limited family history data, which may have impacted their classification as sporadic cases. This limitation should be considered when interpreting the results, as some individuals may have undetected genetic predispositions due to incomplete family history information. However, we now have comprehensive electronic medical records that include detailed family history information for first and second-degree relatives, with an average follow-up period of 5–8 years.

On the other hand, it is important to acknowledge that 55% of women diagnosed before age 50 would not have met the PRS-based threshold for early screening. This highlights a key limitation of current polygenic risk models and underscores the need for more comprehensive approaches. To address this, our team is actively working on the development of integrated models that combine PRS with clinical, imaging, and lifestyle factors to improve early detection, particularly in younger women. These efforts will be supported by future prospective studies aimed at validating the clinical utility and equity of such models in diverse Latin American populations.

While this is a clinical validation study, future research should evaluate qualitative and quantitative aspects related to the psychosocial impact on participants. This includes anxiety stemming from cancer risk stratification, concerns about perceived cancer risk, efforts to seek support and information, and behaviors associated with understanding and accepting risk information. Such evaluations are crucial to recognizing the broader consequences of these interventions at both personal and population levels. Additionally, expanding the sample size and including individuals with diverse ancestries, particularly Afro-Colombian and European populations, is necessary to enhance the predictive quality and accuracy of risk states. This is especially important given the known lower performance of predictive parameters in populations with African ancestry and Caribbean women compared to other groups.

## Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request. Restrictions apply to the availability of these data due to ethical and privacy considerations.

## Ethics statement

The studies involving humans were approved by Comité de Ética y BPC en Investigación Sura. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

HV: Conceptualization, Supervision, Writing – original draft, Writing – review and editing. DC: Formal Analysis, Methodology, Resources, Visualization, Writing – original draft, Writing – review and editing. CeB: Investigation, Project administration, Writing – original draft, Writing – review and editing. MV: Conceptualization, Investigation, Methodology, Resources, Supervision, Writing – original draft, Writing – review and editing. SH: Data curation, Investigation, Writing – original draft, Writing – review and editing. CJ: Funding acquisition, Supervision, Writing – original draft, Writing – review and editing. AC-R: Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review and editing. CrB: Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review and editing. CP: Conceptualization, Methodology, Supervision, Validation, Writing – original draft, Writing – review and editing. CtB: Conceptualization, Investigation, Validation, Writing – original draft, Writing – review and editing. LG: Conceptualization, Funding acquisition, Writing – original draft, Writing – review and editing. JK: Investigation, Validation, Writing – original draft, Writing – review and editing. SK: Investigation, Validation, Writing – original draft, Writing – review and editing. AB: Investigation, Writing – original draft, Writing – review and editing. PD: Investigation, Validation, Writing – original draft, Writing – review and editing. GiB: Investigation, Validation, Writing – original draft, Writing – review and editing. GeB: Conceptualization, Methodology, Supervision, Validation, Writing – original draft, Writing – review and editing. JV-A: Data curation, Formal Analysis, Investigation, Methodology, Supervision, Writing – original draft, Writing – review and editing.

## Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

## Acknowledgments

We thank the entire work of the SURA Bioscience Center working group, EPS SURA and Ayudas Diagnósticas SURA. Also, we thank all Allelica team members that made this study possible.

## Conflict of interest

Authors JK, SK, AB, PD, GiB, and GeB were employed by Allelica Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations,

## References

- Allman, R., Dite, G. S., Hopper, J. L., Gordon, O., Starlard-Davenport, A., Chlebowski, R., et al. (2015). SNPs and breast cancer risk prediction for African American and Hispanic women. *Breast Cancer Res. Treat.* 154 (3), 583–589. doi:10.1007/s10549-015-3641-7
- Allman, R., Spaeth, E., Lai, J., Gross, S. J., and Hopper, J. L. (2021). A streamlined model for use in clinical breast cancer risk assessment maintains predictive power and is further improved with inclusion of a polygenic risk score. *PLoS One* 16 (1 January), e0245375. doi:10.1371/journal.pone.0245375
- Amir, E., Freedman, O. C., Seruga, B., and Evans, D. G. (2010). Assessing women at high risk of breast cancer: a review of risk assessment models. *J. Natl. Cancer Inst.* 102 680–691. doi:10.1093/jnci/djq088
- Bansal, N., Katz, R., De Boer, I. H., Peralta, C. A., Fried, L. F., Siscovick, D. S., et al. (2015). Development and validation of a model to predict 5-year risk of death without ESRD among older adults with CKD. *Clin. J. Am. Soc. Nephrol.* 10 (3), 363–371. doi:10.2215/CJN.04650514
- Bild, D. E., Bluemke, D. A., Burke, G. L., Detrano, R., Diez Roux, A. V., Folsom, A. R., et al. (2002). Multi-ethnic study of Atherosclerosis: objectives and design. *Am. J. Epidemiol.* 156 (9), 871–881. doi:10.1093/aje/kwf113
- Boora, G. K., Kanwar, R., Kulkarni, A. A., Abyzov, A., Sloan, J., Ruddy, K. J., et al. (2016). Testing of candidate single nucleotide variants associated with paclitaxel neuropathy in the trial NCCTG N08C1 (Alliance). *Cancer Med.* 5 (4), 631–639. doi:10.1002/cam4.625
- Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., et al. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 74 (3), 229–263. doi:10.3322/caac.21834
- Brinton, L. A., Awuah, B., Nat Clegg-Lamprey, J., Wiafe-Addai, B., Ansong, D., Nyarko, K. M., et al. (2017). Design considerations for identifying breast cancer risk factors in a population-based study in Africa. *Int. J. Cancer* 140 (12), 2667–2677. doi:10.1002/ijc.30688
- Brooks, J. D., Nabi, H. H., Andrusis, I. L., Antoniou, A. C., Chiquette, J., Després, P., et al. (2021). Personalized risk assessment for prevention and early detection of breast cancer: integration and implementation (PERSPECTIVE I&I). *J. Pers. Med.* 11 (6), 511. doi:10.3390/jpm11060511
- Bryc, K., Velez, C., Karafet, T., Moreno-Estrada, A., Reynolds, A., Auton, A., et al. (2010). Colloquium paper: genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proc. Natl. Acad. Sci. U. S. A.* 107 (Suppl. 2), 8954–8961. doi:10.1073/pnas.0914618107
- Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 47 (D1), D1005–D1012. doi:10.1093/nar/gky1120
- Busby, G. B., Craig, P., Yousfi, N., Hebbalkar, S., Domenico, P. Di, and Bottà, G. (2020). Genetic assessments of breast cancer risk that do not account for polygenic background are incomplete and lead to incorrect preventative strategies. doi:10.1101/2021.08.13.21262050
- Busby, G. B., Kulm, S., Bolli, A., Kintzle, J., Domenico, P. D., and Bottà, G. (2023). Ancestry-specific polygenic risk scores are risk enhancers for clinical cardiovascular disease assessments. *Nat. Commun.* 14 (1), 7105. doi:10.1038/s41467-023-42897-w
- Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562 (7726), 203–209. doi:10.1038/s41586-018-0579-z
- Clarke, A., and Thirlaway, K. (2011). “Genomic counseling”? Genetic counseling in the genomic era. *Genome Med.* 3, 7. doi:10.1186/gm221
- Collaborative Group on Hormonal Factors in Breast Cancer (2001). Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58 209 women with breast cancer and 101 986 women without the disease. *Lancet.* 358 (9291), 1389–1399. doi:10.1016/S0140-6736(01)06524-2
- Crouch, D. J. M., and Bodmer, W. F. (2020). Polygenic inheritance, GWAS, polygenic risk scores, and the search for functional variants. *Proc. Natl. Acad. Sci. U. S. A.* 117 (32), 18924–18933. doi:10.1073/pnas.2005634117
- Cupido, A. J., Tromp, T. R., and Hovingh, G. K. (2021). The clinical applicability of polygenic risk scores for LDL-cholesterol: considerations, current evidence and future perspectives. *Curr. Opin. Lipidol.* 32, 112–116. doi:10.1097/MOL.0000000000000741
- Darst, B. F., Sheng, X., Eeles, R. A., Kote-Jarai, Z., Conti, D. V., and Haiman, C. A. (2021). Combined effect of a polygenic risk score and rare genetic variants on prostate cancer risk. *Eur. Urol.* 80 (2), 134–138. doi:10.1016/j.eururo.2021.04.013
- Duncan, L., Shen, H., Gelaye, B., Meijssen, J., Ressler, K., Feldman, M., et al. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. *Nat. Commun.* 10 (1), 3328. doi:10.1038/s41467-019-11112-0
- Eby, P. R. (2017). Evidence to support screening women annually. *Radiologic Clin. North Am.* 55, 441–456. doi:10.1016/j.rcl.2016.12.003
- Fatumo, S., Chikowore, T., Choudhury, A., Ayub, M., Martin, A. R., and Kuchenbaecker, K. (2022). A roadmap to increase diversity in genomic studies. *Nat. Med.* 28, 243–250. doi:10.1038/s41591-021-01672-4
- Félix, T. M., Fischinger Moura de Souza, C., Oliveira, J. B., Rico-Restrepo, M., Zanoteli, E., Zatz, M., et al. (2023). Challenges and recommendations to increasing the use of exome sequencing and whole genome sequencing for diagnosing rare diseases in Brazil: an expert perspective. *Int. J. Equity Health* 22, 11. doi:10.1186/s12939-022-01809-y
- French, D. P., Astley, S., Astley, S., Brentnall, A. R., Cuzick, J., Dobrashian, R., et al. (2020). What are the benefits and harms of risk stratified screening as part of the NHS breast screening Programme? Study protocol for a multi-site non-randomised comparison of BC-predict versus usual screening (NCT04359420). *BMC Cancer* 20 (1), 570. doi:10.1186/s12885-020-07054-2
- Gaziano, J. M., Concato, J., Brophy, M., Fiore, L., Pyarajan, S., Breeling, J., et al. (2016). Million Veteran Program: a mega-biobank to study genetic influences on health and disease. *J. Clin. Epidemiol.* 70, 214–223. doi:10.1016/j.jclinepi.2015.09.016
- Grinde, K. E., Qi, Q., Thornton, T. A., Liu, S., Shadyab, A. H., Chan, K. H. K., et al. (2019). Generalizing polygenic risk scores from Europeans to Hispanics/Latinos. *Genet. Epidemiol.* 43 (1), 50–62. doi:10.1002/gepi.22166
- Harbeck, N., and Gnant, M. (2017). Breast cancer. *Lancet* 389, 1134–1150. doi:10.1016/S0140-6736(16)31891-8
- He, J. Q., Burkett, K., Connett, J. E., Anthonisen, N. R., Paré, P. D., and Sandford, A. J. (2006). Interferon gamma polymorphisms and their interaction with smoking are associated with lung function. *Hum. Genet.* 119 (4), 365–375. doi:10.1007/s00439-006-0143-z
- Hines, L. M., Sedjo, R. L., Byers, T., John, E. M., Fejerman, L., Stern, M. C., et al. (2017). The interaction between genetic ancestry and breast cancer risk factors among hispanic women: the breast cancer health disparities study. *Cancer Epidemiol. Biomarkers Prev.* 26 (5), 692–701. doi:10.1158/1055-9965.EPI-16-0721
- Ho, P. J., Lim, E. H., Hartman, M., Wong, F. Y., and Li, J. (2023). Breast cancer risk stratification using genetic and non-genetic risk assessment tools for 246,142 women in the UK Biobank. *Genet. Med.* 25 (10), 100917. doi:10.1016/j.gim.2023.100917
- Jiang, L., Zheng, Z., Fang, H., and Yang, J. (2021). A generalized linear mixed model association tool for biobank-scale data. *Nat. Genet.* 53, 1616–1621. doi:10.1038/s41588-021-00954-4
- Kocarnik, J. M., Compton, K., Dean, F. E., Fu, W., Gaw, B. L., Harvey, J. D., et al. (2022). Cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life years for 29 cancer groups from 2010 to 2019: A systematic analysis for the global burden of disease study 2019. *JAMA Oncol.* 8 (3), 420–444. doi:10.1001/jamaoncol.2021.6987
- Kryukov, G. V., Shpunt, A., Stamatoyannopoulos, J. A., and Sunyaev, S. R. (2009). Power of deep, all-exon resequencing for discovery of human trait genes. *Proc. Natl. Acad. Sci. U. S. A.* 106 (10), 3871–3876. doi:10.1073/pnas.0812824106

or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2025.1556907/full#supplementary-material>

- Lambert, S. A., Abraham, G., and Inouye, M. (2019). Towards clinical utility of polygenic risk scores. *Hum. Mol. Genet.* 28, R133–R142. doi:10.1093/hmg/ddz187
- Levin, M. G., and Rader, D. J. (2020). Polygenic risk scores and coronary artery disease: ready for prime time? *Circulation* 141, 637–640. doi:10.1161/CIRCULATIONAHA.119.044770
- Li, Y., Levran, O., Kim, J. J., Zhang, T., Chen, X., and Suo, C. (2019). Extreme sampling design in genetic association mapping of quantitative trait loci using balanced and unbalanced case-control samples. *Sci. Rep.* 9 (1), 15504. doi:10.1038/s41598-019-51790-w
- Ma, Y., and Zhou, X. (2021). Genetic prediction of complex traits with polygenic scores: a statistical review. *Trends Genet.* 37, 995–1011. doi:10.1016/j.tig.2021.06.004
- Mars, N., Kerminen, S., Feng, Y. C. A., Kanai, M., Läll, K., Thomas, L. F., et al. (2022). Genome-wide risk prediction of common diseases across ancestries in one million people. *Cell Genomics* 2 (4), 100118. doi:10.1016/j.xgen.2022.100118
- Mars, N., Widén, E., Kerminen, S., Meretoja, T., Pirinen, M., della Briotta Parolo, P., et al. (2020). The role of polygenic risk and susceptibility genes in breast cancer over the course of life. *Nat. Commun.* 11 (1), 6383. doi:10.1038/s41467-020-19966-5
- Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., and Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* 51 (4), 584–591. doi:10.1038/s41588-019-0379-x
- Martínez, M. E., Gomez, S. L., Tao, L., Cress, R., Rodriguez, D., Unkart, J., et al. (2017). Contribution of clinical and socioeconomic factors to differences in breast cancer subtype and mortality between Hispanic and non-Hispanic white women. *Breast Cancer Res. Treat.* 166 (1), 185–193. doi:10.1007/s10549-017-4389-z
- Mavaddat, N., Michailidou, K., Dennis, J., Lush, M., Fachal, L., Lee, A., et al. (2019). Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am. J. Hum. Genet.* 104 (1), 21–34. doi:10.1016/j.ajhg.2018.11.002
- McArdle, C. E., Bokhari, H., Rodell, C. C., Buchanan, V., Preudhomme, L. K., Isasi, C. R., et al. (2021). Findings from the hispanic community health study/study of Latinos on the importance of sociocultural environmental interactors: polygenic risk score-by-immigration and dietary interactions. *Front. Genet.* 12, 720750. doi:10.3389/fgene.2021.720750
- McPherson, K., Steel, C. M., and Dixon, J. M. (2001). ABC of Breast Diseases Breast cancer-epidemiology, risk factors, and genetics. *BMJ*. 321 (7261), 624–628. doi:10.1136/bmj.321.7261.624
- Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., et al. (2017). Association analysis identifies 65 new breast cancer risk loci. *Nature* 551 (7678), 92–94. doi:10.1038/nature24284
- Morabia, A., Cayanis, E., Costanza, M. C., Ross, B. M., Flaherty, M. S., Alvin, G. B., et al. (2003). Association of extreme blood lipid profile phenotypic variation with 11 reverse cholesterol transport genes and 10 non-genetic cardiovascular disease risk factors. *Hum. Mol. Genet.* 12 (21), 2733–2743. doi:10.1093/hmg/ddg314
- Nyante, S. J., Biritwum, R., Figueroa, J., Graubard, B., Awuah, B., Addai, B. W., et al. (2019). Recruiting population controls for case-control studies in sub-Saharan Africa: the Ghana breast health Study. *PLoS One* 14 (4), e0215347. doi:10.1371/journal.pone.0215347
- Park, A., Lee, J. H., Park, A., Jung, Y. H., Chu, H. J., Bae, S. S., et al. (2017). Prevalence rate and clinical characteristics of esophageal ectopic sebaceous glands in asymptomatic health screen examinees. *Dis. Esophagus* 30 (1), 1–5. doi:10.1111/dote.12453
- Patel, A. P., and Khera, A. V. (2022). Advances and Applications of Polygenic Scores for Coronary Artery Disease. *Annu Rev Med.* 74, 141–154. doi:10.1146/annurev-med-042921-112629
- Privé, F., Vilhjálmsson, B. J., Aschard, H., and Blum, M. G. B. (2019). Making the most of clumping and thresholding for polygenic scores. *Am. J. Hum. Genet.* 105 (6), 1213–1221. doi:10.1016/j.ajhg.2019.11.001
- Ray, K. M., Price, E. R., and Joe, B. N. (2017). Evidence to support screening women in their 40s. *Radiologic Clin. N. Am.* 55, 429–439. doi:10.1016/j.rcl.2016.12.002
- Rey-Vargas, L., Bejarano-Rivera, L. M., Mejia-Henao, J. C., Sua, L. F., Bastidas-Andrade, J. F., Ossa, C. A., et al. (2022). Association of genetic ancestry with HER2, GRB7 AND estrogen receptor expression among Colombian women with breast cancer. *Front. Oncol.* 12, 989761. doi:10.3389/fgene.2022.989761
- Rodrigues de Moura, R., Coelho, A. V. C., de Queiroz Balbino, V., Crovella, S., and Brandão, L. A. C. (2015). Meta-analysis of Brazilian genetic admixture and comparison with other Latin America countries. *Am. J. Hum. Biol.* 27 (5), 674–680. doi:10.1002/ajhb.22714
- Rojas, K., and Stuckey, A. (2016). Breast cancer epidemiology and risk factors. *Clin. Obstet. Gynecol.* 59 (4), 651–672. doi:10.1097/GRF.0000000000000239
- Roux, A., Cholerton, R., Sicsic, J., Moumjid, N., French, D. P., Giorgi Rossi, P., et al. (2022). Study protocol comparing the ethical, psychological and socio-economic impact of personalised breast cancer screening to that of standard screening in the “My Personal Breast Screening” (MyPeBS) randomised clinical trial. *BMC Cancer* 22 (1), 507. doi:10.1186/s12885-022-09484-6
- Ruan, Y., Lin, Y. F., Feng, Y. C. A., Genet, N., Guo, Z., Sawa, A., et al. (2022). Improving polygenic prediction in ancestrally diverse populations. *Nat. Genet.* 54 (1), 573–580. doi:10.1038/s41588-022-01054-7
- Ruiz, L. K. A., Sánchez, P. V., Barbudo, D. C., Hernández, L. G., and Hernández, J. M. R. (2022). Sistematización de Experiencias: una reflexión sobre sus potencialidades para la salud pública. *Salud Unimorte* 38 (1), 299–326. doi:10.14482/sun.38.1.614.41
- Sakaue, S., Kanai, M., Tanigawa, Y., Karjalainen, J., Kurki, M., Koshiba, S., et al. (2021). A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* 53, 1415–1424. doi:10.1038/s41588-021-00931-x
- Shieh, Y., Fejerman, L., Lott, P. C., Marker, K., Sawyer, S. D., Hu, D., et al. (2020). A polygenic risk score for breast cancer in US latinas and Latin American women. *J. Natl. Cancer Inst.* 112 (6), 590–598. doi:10.1093/jnci/djz174
- Shieh, Y., Hu, D., Ma, L., Huntsman, S., Gard, C. C., Leung, J. W. T., et al. (2016). Breast cancer risk prediction using a clinical risk model and polygenic risk score. *Breast Cancer Res. Treat.* 159 (3), 513–525. doi:10.1007/s10549-016-3953-2
- Sims, A. M., Shephard, N., Carter, K., Doan, T., Dowling, A., Duncan, E. L., et al. (2008). Genetic analyses in a sample of individuals with high or low BMD shows association with multiple Wnt pathway genes. *J. Bone Mineral Res.* 23 (4), 499–506. doi:10.1359/jbmr.071113
- Sweeney, C., Baumgartner, K. B., Byers, T., Giuliano, A. R., Herrick, J. S., Murtaugh, M. A., et al. (2008). Reproductive history in relation to breast cancer risk among Hispanic and non-Hispanic white women. *Cancer Causes Control* 19, 391–401. doi:10.1007/s10552-007-9098-1
- Sweet, K., Sturm, A. C., Schmidlen, T., McElroy, J., Scheinfeldt, L., Manickam, K., et al. (2017). Outcomes of a randomized controlled trial of genomic counseling for patients receiving personalized and actionable complex disease reports. *J. Genet. Couns.* 26 (5), 980–998. doi:10.1007/s10897-017-0073-z
- Trivino, J. C., Ceba, A., Rubio-Solsona, E., Serra, D., Sanchez-Guiu, I., Ribas, G., et al. (2020). Combination of phenotype and polygenic risk score in breast cancer risk evaluation in the Spanish population: a case – control study. *BMC Cancer* 20, 1079. doi:10.1186/s12885-020-07584-9
- Vachon, C. M., Pankratz, V. S., Scott, C. G., Haeberle, L., Ziv, E., Jensen, M. R., et al. (2015). The contributions of breast density and common genetic variation to breast cancer risk. *J. Natl. Cancer Inst.* 107 (5), dju397. doi:10.1093/jnci/dju397
- Versmissen, J., Oosterveer, D. M., Yazdanpanah, M., Dehghan, A., Hölm, H., Erdman, J., et al. (2015). Identifying genetic risk variants for coronary heart disease in familial hypercholesterolemia: an extreme genetics approach. *Eur. J. Hum. Genet.* 23 (3), 381–387. doi:10.1038/ejhg.2014.101
- Viaña González, L. F. (2020). Retos para el control del cáncer en Colombia: ante todo más acción. *Rev. Colomb. Cancerol.* 24 (3), 96–97. doi:10.35509/01239015.712
- Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., et al. (2017). 10 Years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* 101, 5–22. doi:10.1016/j.ajhg.2017.06.005
- Weissbrod, O., Hormozdiari, F., Benner, C., Cui, R., Ulirsch, J., Gazal, S., et al. (2020). Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nat. Genet.* 52 (12), 1355–1363. doi:10.1038/s41588-020-00735-5
- Wolfson, M., Gribble, S., Pashayan, N., Easton, D. F., Antoniou, A. C., Lee, A., et al. (2021). Potential of polygenic risk scores for improving population estimates of women’s breast cancer genetic risks. *Genet. Med.* 23 (11), 2114–2121. doi:10.1038/s41436-021-01258-y
- World health organization (2023). Cancer today Available online at: <https://gco.iarc.fr/today/home>
- Wray, N. R., Goddard, M. E., and Visscher, P. M. (2007). Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res.* 17 (10), 1520–1528. doi:10.1101/gr.6665407