scientific reports

Check for updates

OPEN

Metabolomic-driven prediction of the mutational status of healthy individuals with a family history of hereditary breast and ovarian cancer syndrome: the HRRmet study

Bàrbara Roig[1,2,3], Sara Fernández-Castillejo[1,2]✉, Josep Gumà[1,2,3], Joan Badia[1,2], Mireia Melé[1,2], Mònica Salvat[1,2], Montserrat Querol[1,2], Raquel Cumeras[1,2,3] & Marta Rodríguez-Balada[1,2]

Pathogenic variants (PVs) identified in genes involved in the DNA homologous recombination repair (HRR) mechanism are the main cause of hereditary breast and ovarian cancer syndrome (HBOC). The main objective of this study was to identify differential plasma metabolomic profiles associated with the HRR genotype in healthy individuals. Cascade testing was performed by Sanger sequencing in healthy carrier and noncarrier individuals with a familial history of HBOC. PVs associated with HRR genes (*BRCA1*, *BRCA2*, *PALB2*, *ATM*, *CHEK2* and *RAD51*) were identified. Untargeted metabolomics of plasma samples was performed by liquid chromatography coupled with mass spectrometry. Predictive models were developed using a machine learning approach. Thirty-one metabolites were selected to create the global predictive model (accuracy 61.9%), whereas the gene-specific models had a better performance (accuracy > 80%) and were constructed with fewer metabolites. The present study is the first to characterize the phenotype associated with the HRR-deficient genotype in healthy individuals with a familial history of HBOC. Metabolomic profiles may be useful for differentiating carriers from noncarriers of PVs in the HRR genes, and therefore, with potential predictive capacity of the HRR germline mutational status.

Keywords Metabolome, Metabolomic profile, Mutational status, Homologous recombination repair, Hereditary breast and ovarian cancer

DNA repair involves a collection of normal physiological processes whose objective is to maintain the integrity and fidelity of the genome to guarantee the correct transmission of genetic material after cell duplication[1]. Different DNA repair mechanisms exist depending on whether a single-strand or double-strand DNA break has occurred. Following double-strand breaks, the cell loses the undamaged template to repair the damaged strand and must use the homologous sequence of the sister chromatid as a template. This DNA repair mechanism only occurs during S and G2 phases, when the sister chromatids are closest to each other. This process is known as the homologous recombination repair (HRR) mechanism. The main genes involved in this double-strand break repair system are *ATM*, *BRCA1*, *BRCA2*, *CHEK2*, *RAD50*, *RAD51C*, *RAD51D*, *BRIP1* and *PALB2*. When these tumour suppressor genes are inactivated, mutations in other cell cycle regulatory genes can accumulate, leading to genomic instability and, in turn, increasing the likelihood of abnormal cell proliferation and cancer development[2,3]. The tumours that show a deficit in the HRR mechanism present a tumour phenotype known as homologous recombination deficiency (HRD) or the BRCAness phenotype. Although this phenotype is caused

[1]Institute of Oncology of Southern Catalonia (IOCS), Sant Joan University Hospital of Reus (HUSJR), Reus, Spain. [2]Translational, Epidemiological, Clinical, and Oncological Research Group (GIOTEC), Pere Virgili Health Research Institute (IISPV), Reus, Spain. [3]Rovira i Virgili University (URV), Reus, Spain. ✉email: sara.fernandez@iispv.cat

mainly by inactivation of the *BRCA1* and *BRCA2* genes, the presence of pathogenic or likely pathogenic variants (PVs) in the remaining genes responsible for HRR may also be involved[4].

Hereditary breast and ovarian cancer syndrome (HBOC) is one of the most common hereditary cancer syndromes. It is characterized by germline alterations in genes involved in the HRR pathway. The main genes known to be related to HBOC are the high-penetrance genes *BRCA1*, *BRCA2* and *PALB2*, although additional genes with moderate penetrance, such as *ATM*, *CHEK2* and *BRIP1*, have also associated with HBOC[5]. In approximately 10–15% of HBOC cases, a PV in the *BRCA1* and *BRCA2* (*BRCAs)* genes is identified. Similarly, the presence of alterations in *BRCAs* increases the risk of breast cancer 10–20-fold, and alterations in *BRCA1* increase the risk of colorectal cancer 4-fold[2,6–9].

According to the Leiden Open Variation Database (LOVD) of the global variome, approximately 6071 unique *BRCA1* variants have been described, 2628 of which are considered PVs: 8417 variants in *BRCA2* (2875 PV), 1407 variants in *PALB2* (170 PV), 3559 in *ATM* (473 PV), 893 in *BRIP1* (46 PV), 708 in *CHEK2* (76 PV), 314 in *RAD51C* (36 PV), and 278 in *RAD51D* (14 PV) (data retrieved on July 8, 2025)[10]. Notably, the *BRCA* genes were the first genes to be cloned, identified, and related to hereditary cancer syndromes[11,12]. Most of the variants described in these genes (≈ 60%) are family-specific (described only once) and are randomly distributed throughout the gene. However, some recurrent PVs at specific geographical or population levels have also been reported[13].

In recent years, a paradigm shift has occurred in the field of oncology, with clinical oncology evolving into precision oncology (also known as personalized oncology or genomic medicine). This shift has promoted the exponential appearance of targeted treatments directed at specific molecular alterations of each tumour, giving way to the appearance of the concept of molecular markers or biomarkers. In this sense, the role of omic sciences is highly relevant. Omic sciences allow the identification and quantification of a wide range of molecules, yielding a holistic view of an organism and integrating data obtained from genomics, transcriptomics, proteomics and metabolomics, among other methods. In particular, metabolomics involves the analysis of a set of small metabolites (called the metabolome) involved in various biological pathways and thus present in biofluids and tissue samples. An individual's genome is translated into the transcriptome, then the proteome, and eventually the metabolome, which is the result of the individual's metabolism. Hence, alterations present in a gene result in modifications to the metabolic profile, which might ultimately promote the growth of cancer cells[14]. Metabolomics is a state-of-the-art and high-throughput technology that offers the unprecedented possibility of conducting untargeted analyses to characterize differences in the metabolomes of different clusters of individuals. In this sense, metabolomics is commonly applied as a tool for biomarker discovery for early diagnosis, prognosis, treatment, and prevention. In addition, due to the inherent sensitivity of the analytical technology applied for metabolomics, subtle alterations in biological pathways can be detected to provide insights into the biochemical mechanisms underlying various physiological, pathological, and disease-predictive conditions[15,16].

The application of metabolomics in oncology, and more specifically in hereditary cancer syndromes, is currently thriving. Some authors have explored the possibility of using metabolomic profiles as potential early detection biomarkers in asymptomatic patients, for tumour characterization, or as prognostic factors[17,18]. However, evidence for biomarkers associated with the mutational status in the absence of an oncological disease and, more specifically, in healthy individuals with a familial history of cancer is lacking. Moreover, most of the published studies have focused on an analysis of tumour tissues or cell lines, and very few have used the plasma or serum of subjects[14,19–21].

However, evidence on the usefulness of metabolomics in characterizing the mutational status of healthy individuals with a family history of hereditary cancer is scarce. In a previous metabolomic study, we demonstrated the existence of free methylated nucleotides that could be considered biomarkers of the mutational status of *BRCA1* in HBOC patients[22]. These results suggest that mutational status could be inferred through a metabolic approach.

In the present study, we wanted to further evaluate the existence of a differential and characteristic metabolomic profile comprising molecular biomarkers of the mutational status for the first time, which allows us to infer the HRD genotype in healthy individuals with a family history of HBOC.

Results

Demographic characteristics and mutational status of the participants

A total of 260 healthy participants with a familial history of hereditary cancer were enrolled in this study (130 carriers of PVs in genes involved in HRR and 130 noncarriers). As shown in (Table 1), 56.2% of the participants were women (49.2% carriers and 63.1% noncarriers), and the mean age was 48 ± 20 years (50 ± 22 years for carriers and 46 ± 17 years for noncarriers). Most of the genes studied in both carriers and noncarriers were *BRCA2* (40.0 and 44.6%, respectively) and *BRCA1* (24.6 and 30.0%), followed by *ATM* (11.5 and 8.5%), *PALB2* (10.0 and 7.7%), *CHEK2* (6.9 and 6.2%) and *RAD51* (6.9 and 3.1%).

In the present study, *RAD51C* and *RAD51D* carriers were clustered together (i.e., *RAD51*) because few individuals exhibited PVs in these two genes. Noncarrier subjects were selected from our database by matching age and sex with carriers so that no significant differences existed between the two subpopulations (Table 1).

Metabolites identified by metabolomics

Liquid chromatography coupled with mass spectrometry (LC–MS) allowed us to identify 285 plasma metabolites. After data curation, 213 (74.7%) metabolites were selected, and 169 (59.3%) were present in at least 80% of the samples. These metabolites were further subjected to imputation, normalization, and analysis as detailed in the Methods section.

	Carriers		Non-carriers		Total		p-value
	(n = 130)		(n = 130)		(n = 260)		
Gender; females (%)	64	(49.2%)	82	(63.1%)	146	(56.2%)	0.136
Age; yo (mean ± SD)	50	± 22	46	± 17	48	± 20	0.102
Gene studied; n (%)							
ATM	15	(11.5%)	11	(8.5%)	26	(10.0%)	0.433
BRCA1	32	(24.6%)	39	(30.0%)	71	(27.3%)	0.406
BRCA2	52	(40.0%)	58	(44.6%)	110	(42.3%)	0.567
CHEK2	9	(6.9%)	8	(6.2%)	17	(6.5%)	0.808
PALB2	13	(10.0%)	10	(7.7%)	23	(8.8%)	0.532
RAD51	9	(6.9%)	4	(3.1%)	13	(4.2%)	0.166

Table 1. Demographic characteristics and mutational status of the participants. A total of 260 healthy participants with a Familial history of hereditary cancer were enrolled in this study (130 carriers of PV in genes involved in the homologous recombination repair pathway and 130 noncarriers). Abbreviations: SD, standard deviation; yo, years old.

	Fold-change		Univariate analysis	Volcano Plots	Machine learning
	< 0.75	> 1.25	(p-value < 0.05)	(FC and p-value criteria)	(VI metabolites)
Global	2	17	18	7	31
Gene-segregated					
BRCA1	8	68	26	25	8
BRCA2	5	26	7	5	21
PALB2	39	41	3	3	3
ATM	40	47	6	6	20
CHEK2	29	48	4	4	9
RAD51	27	71	6	6	-
Penetrance-segregated					
High penetrance	3	28	25	17	14
Moderate penetrance	15	38	13	12	7

Table 2. Summary of the number of differentially abundant metabolites between the carrier and noncarrier subpopulations. The fold change (FC) cut-off values (FC > 1.25 and FC < 0.75) were selected arbitrarily; since participants were healthy individuals, no higher FCs were expected. Univariate analyses were performed with the Mann–Whitney–Wilcoxon test. Data in the volcano plots refer to those metabolites meeting both the p-value and FC criteria. The machine learning approach was performed by a linear support vector machine (SVM) with recursive feature elimination (RFE), for which the variant influence (VI) metabolites were included.

Metabolite profiling of the carrier and noncarrier subpopulations

The univariate analysis revealed that of the 169 metabolites profiled, 57 metabolites were significantly altered ($p < 0.05$) in at least one of the performed comparisons: data analysed globally; data segregated by gene (*BRCA1*, *BRCA2*, *PALB2*, *ATM*, *CHEK2*, and *RAD51*); and data segregated by gene penetrance (high penetrance genes: *BRCA1*, *BRCA2*, and *PALB2*; and moderate penetrance genes: *ATM*, *CHEK2*, and *RAD51*). Some metabolites were significantly altered in only one comparison (e.g., CAR 2:0), whereas others were significantly altered in several comparisons (e.g., CAR 3:0) (Table 2 and Supplementary Table S1).

Metabolites for predicting the global HRR mutational status

Metabolomic data were used to compare carrier and noncarrier subpopulations regardless of the gene affected or gene penetrance. Eighteen metabolites were significantly altered ($p < 0.05$) (Table 2 and details in Supplementary Table S1). These significant differences were selected according to cut-off values of fold change (FC) of > 1.25 and FC < 0.75. As shown in Table 2, 17 metabolites had a FC > 1.25, whereas 2 metabolites had a FC < 0.75 (details are given in Supplementary Table S2). Finally, volcano plots were constructed to examine the metabolites that met both the *p*-value and FC criteria. As observed in the volcano plots (see Supplementary Fig. S1 online), no significant ($p < 0.05$) metabolite had a FC < 0.75, but seven had a FC > 1.25when globally comparing carriers and noncarriers.

Further analysis using a machine learning approach based on a linear support vector machine (SVM) algorithm allowed us to select the most important metabolites (i.e., the metabolomic profile) to construct a model that could predict the mutational status of HRR-associated genes in healthy individuals with a familial history

of HBOC. Table 2 summarizes the results obtained from the machine learning process. When the global dataset was used, the cross-validated maximum area under the receiver operating characteristic curve (maxAUC) of the receiver operating characteristic (ROC) curve was achieved with 31 specific metabolites (maxAUC of 0.694) (Fig. 1). Among these 31 metabolites, 17 were found to be important for predicting the carrier mutational status, and 14 were important for predicting the noncarrier mutational status. Using only these selected 31 metabolites, the SVM algorithm was applied to obtain the final predictive model (Table 3), as well as the ROC curves of the model (see Supplementary Fig. S2 online). As shown in Table 3, the specificity and sensitivity of this model were greater than 60%, while the maxAUC was 0.671. Using this model constructed with the entire dataset, the assignment of the mutational status was correct in 61.9% of the subjects (accuracy).

The importance of the metabolites in predicting the mutational status was quantified using the variable influence (VI) score. The VI scores of the 31 metabolites selected for the global model ranged from 9.00 to 30.00. Some of these metabolites were significant ($p < 0.05$; $n = 6$), and most had a FC greater than 1.25 ($n = 4$) (Table 3; Fig. 1).

Metabolites for predicting the HRR gene-specific mutational status

The data segregated by gene revealed 26 differentially abundant metabolites for *BRCA1*, 7 for *BRCA2*, 6 for *ATM* and *RAD51*, 4 for *CHEK2*, and 3 for *PALB2* ($p < 0.05$) (Table 2 and details in Supplementary Table S1). Moreover, 68 metabolites had a FC > 1.25 for *BRCA1*, 26 for *BRCA2*, 41 for *PALB2*, 47 for *ATM*, 48 for *CHEK2*, and 71 for *RAD51*. Fewer metabolites had a FC < 0.75: 8 for *BRCA1*, 5 for *BRCA2*, 39 for *PALB2*, 40 for *ATM*, 29 for *CHEK2*, and 27 for *RAD51*. When both *p*-values and FC values were considered, only 25 metabolites met both criteria in the *BRCA1* subgroup, 5 in the *BRCA2* subgroup, 3 in the *PALB2* subgroup, 6 in the *ATM* subgroup, 4 in the *CHEK2* subgroup, and 6 in the *RAD51* subgroup, as shown in the corresponding volcano plots (see Supplementary Fig. S1 online).

The gene-specific models constructed with machine learning algorithms enhanced the predictive capacity in our study and decreased the number of metabolites (range 3–21) compared with those models constructed with the whole dataset (i.e., the global model; Table 3). The specificity of the 5 models was > 83%, and the sensitivity was > 75%. The predictive accuracy of these 5 gene-specific models was > 82%, and the maxAUC values were far greater than 0.750, the threshold of a good predictive model (see Supplementary Fig. S2 online for details). Notably, the *BRCA1* and *BRCA2* predictive models were shown to be consistent, with maxAUC values indicative of good models (0.881 and 0.894, respectively). As shown in (Fig. 2), only 8 metabolites were selected by the algorithm to construct the *BRCA1* predictive model, while 21 metabolites were needed for *BRCA2* model construction. In both cases, the specificity, sensitivity, and accuracy were greater than 82%.

As detailed in Table 3, the VI scores of the gene-specific models ranged from 1.67 (phenylalanine in the *PALB2* model) to 18.67 (2,2,2-trifluoroacetophenone in the *BRCA2* model). Interestingly, the latter metabolite was significant and had a FC > 1.25 in the *BRCA2* model. The VI scores of additional metabolites were significant and met the FC criteria (i.e., FC < 0.75 or > 1.25; $n = 5$ in the *BRCA1* model, $n = 1$ in the *BRCA2* model, none in the *PALB2* model, $n = 5$ in the *ATM* model, and $n = 3$ in the *CHEK2* model) (Table 3). The predictive model for *RAD51* genes was not performed because of a lack of sufficient subjects to generate a solid and reliable model.

Metabolites for predicting HRR gene penetrance-specific mutational status

Metabolomic data were also segregated by gene penetrance. This analysis revealed that 25 metabolites were differentially abundant among the high-penetrance genes and that 13 were differentially abundant among the moderate-penetrance genes ($p < 0.05$) (Table 2 and Supplementary Table S1). In addition, 28 metabolites for high-penetrance genes had a FC > 1.25, and only 3 had a FC < 0.75, whereas 38 metabolites for moderate-penetrance genes had a FC > 1.25, and 15 had a FC < 0.75. As shown in the volcano plots (see Supplementary Fig. S1 online), of the 31 metabolites that met the FC criteria, only 17 were significant ($p < 0.05$) in the high-penetrance gene category, and 12 of the 53 meeting the FC criteria were in the moderate-penetrance gene category.

Machine learning models constructed when data were segregated by gene penetrance had different predictive capacities. The high-penetrance model included 14 metabolites and had a modest maxAUC of 0.691 (specificity, sensitivity, and accuracy all > 60%). In contrast, the moderate-penetrance model included only 7 metabolites and correctly assigned the mutational status in 80.4% of the subjects (specificity and sensitivity both > 75%), with an maxAUC close to optimal (0.921). As shown in Table 3, the VI scores ranged from 2.33 to 7.00 (moderate-penetrance model) and from 4.00 to 12.33 (high-penetrance model), the latter corresponding to the metabolite 4,4'-dihydrazino-biphenyl in the high-penetrance model. This VI metabolite was not significant ($p < 0.05$) in the univariate analyses and did not meet the FC criteria. In contrast, 3 VI metabolites were significant and met the FC criteria in the high-penetrance model, and there were 5 such metabolites in the moderate-penetrance model. Notably, no metabolite resulted in a VI score in both models (Table 3; Fig. 1, and Supplementary Fig. S2 online).

Carriers vs. noncarriers heatmap

A heatmap of the metabolites whose abundances differed between carrier and noncarrier participants was constructed for each set of comparisons (Fig. 3). As shown in this heatmap, most of the significant metabolites were increased in carriers vs. noncarriers (i.e., FC > 1.25), and most also met the *p*-value criteria (indicated with a #). Notably, few significant differentially abundant metabolites had a different FC direction, including CAR14:0;O and trans-4-aminocyclohexane carboxylic acid.

Mutation heatmap

To investigate whether the metabolite abundance is related to the gene affected, a mutation heatmap was constructed with data obtained from the 130 carrier individuals included in our study. As shown in Fig. 4,



Fig. 1. Metabolites selected by a machine learning approach based on linear support vector machine (SVM) algorithms. (**a**) Global model. (**b**) *BRCA1* model. (**c**) *BRCA2* model. (**d**) *PALB2* model. (**e**) *ATM* model. (**f**) *CHEK2* model. (**g**) High-penetrance model. (**h**) Moderate-penetrance model. The variant importance (VI) score for each selected metabolite was quantified using the coefficients (weights) calculated by the SVM model. The orange bars indicate metabolites important for carrier prediction, whereas the purple bars indicate those important for noncarrier prediction. Statistically significant comparisons are indicated with an * symbol, and comparisons with FC > 1.25 or FC < 0.75 are indicated with a # symbol. In both cases, p values and FC values are also indicated.

	Global model	Gene-segregated models					Penetrance-segregated models	
		BRCA1	BRCA2	PALB2	ATM	CHEK2	High penetrance	Moderate penetrance
Feature Selection step								
MaxAUC (AU)	0.694	0.906	0.904	0.951	1.000	1.000	0.739	0.960
Selected metabolites (n)	31	8	21	3	20	9	14	7
Metabolites important for predicting carrier (n)	17	5	10	1	7	4	6	5
Metabolites important for predicting non-carrier (n)	14	3	11	2	13	5	8	2
Testing Step (ROC analysis and confusion matrix)								
Specificity (%)	62.6	83.7	84.5	100	90.9	88.9	67.3	77.3
Sensitivity (%)	61.3	89.3	82.7	76.5	93.3	100	64.9	82.4
Accuracy (%)	61.9	85.9	83.6	82.6	92.3	94.1	66.2	80.4
maxAUC (AU)	0.671	0.881	0.894	0.885	0.982	0.986	0.691	0.921
Minimal VI Score	9.00	2.00	2.67	1.67	7.00	3.33	4.00	2.33
Maximal VI Score	30.00	7.33	18.67	2.33	16.67	8.67	12.33	7.00
VI scores of the selected metabolites								
(1R,9 S)-10-(Cyclopropylmethyl)-12-ethyl-4-hydroxy-13-methyl-10-azatricyclo[7.3.1.02,7] trideca-2(7),3,5-trien-8-one					10.67 [1,2]			7.00 [1,2]
(9Z,12Z)-15-Hydroxyoctadeca-9,12-dienoic acid					7.67 [3]			
(R)-3-Hydroxyhexanoic acid		2.67						
[3-[2-Aminoethoxy(hydroxy)phosphoryl]oxy-2-hydroxypropyl] hexadecanoate			15.67					
1-(2-Chloroethyl)-1-nitrosourea	13.67						6.33	
1,2,3,4-Tetrahydroquinoline	17.33							
1-[(5-Methoxy-2,3-dihydro-1 H-indol-3-yl)methylideneamino]-2-pentylguanidine	9.00							
17-Ethynyl-16-fluoroestradiol					8.33 [1,3]			
1-Methyl-5-(4-benzoyl)pyrrole-2-acetic acid 2-(theophylline-7-yl)ethyl ester			11.00 [1]			3.67 [2]		
1-Phenylethylamine			15.33			3.33 [3]		
2-(1-Phenylpropan-2-yl)-1,3,2-dioxazetidine	22.00							
2-(2-Chlorophenyl)ethylbiguanide		5.00 [1,3]						
2,2,2-Trifluoroacetophenone	17.33		18.67 [1,3]				6.67 [1,3]	
2,6-Ditert-butyl-4-[3-(3,5-ditert-butyl-4-hydroxyphenoxy)propoxy]phenol						5.33 [1,2]		
2-Nitrophenyl octyl ether								2.67
3-Butylphthalic acid						4.67 [1,3]		
3-Pyridinemethanol, 6-amino-alpha-(((1-methyl-4-phenylbutyl)amino)methyl)	21.33							
4,4'-Dihydrazino-biphenyl			18.00		7.67 [3]		12.33	
4-Aminoantipyrine	22.33							
7-Hexadecynoic acid	15.33		10.67					
9(10)-Epoxy-12Z-octadecenoic acid			4.33					
Acetyl-N-formyl-5-methoxykynurenamine		5.00[1,3]	7.67					
alpha-(4-Fluorophenyl)-4-(5-fluoro-2-pyrimidinyl)-1-piperazine butanol			9.00					
Anhydromethylecgonine			11.33					
Benzamide, 4-chloro-N-(2-(4-morpholinyl)ethyl)-, N-oxide				2.00 [3]				
Bergamotin		3.67			13.33 [1,3]			
Butanedioic acid, octenyl					7.33			
CAR 12:0					10.67 [2]			
CAR 14:0;O	14.33			2.33 [3]			8.00 [1,3]	
CAR 14:2					14.67 [1,2]			
CAR 16:1			2.67					
CAR 2:0			10.00 [1]					
CAR 3:0	11.00 [1,3]							
CAR 4:0	15.33 [1,3]							
CAR 5:0						5.00 [3]	5.67 [1]	
CAR 6:0					15.00 [2]			
CAR14:1;OH								4.33 [1,2]
Carnitine	15.33 [1,3]				7.00 [3]			

Continued

	Global model	Gene-segregated models					Penetrance-segregated models	
		BRCA1	BRCA2	PALB2	ATM	CHEK2	High penetrance	Moderate penetrance
Cytosine-5-carboxylic acid			13.33		10.33 [3]			
Diisopropylethylamine			9.00 [3]					
DL-Valine-4-antipyrineamide	10.00						8.33	
DMAP-ethyl-PAF					16.67 [3]			
Fluoro-beta-alanine								3.33 [1,2]
Guanidinosuccinic acid	19.00 [1]							
Indole	18.67							
L-Isopulegol		5.33 [1,2]			7.67		9.33 [1]	
LPC 15:0	19.33							
LPC 17:0	16.00							
LPC 18:0 A	21.67						7.00	
LPC 18:1 A	9.33	7.33 [1,3]					7.33	
LPC 18:1 B	22.67		10.33					
LPC 18:1 C	14.00							
LPC 18:2 B	20.00							
LPC 20:5							6.00	
LPC 22:6 B	15.67		15.67					
LPC O-16:0 B	11.00							
LPC P-16:0			9.33					
LPE 18:0			8.33		13.67 [2]			
L-Phenylalanine, methyl ester	10.33 [1]							
Metaldehyde	12.00						10.00	
Methyl 6,7-dimethoxy-4-ethyl-beta-carboline-3-carboxylate					9.33 [3]	4.33 [3]		
Myristoleic acid							4.00	
N6-Methyladenosine	9.33							
PC O-18:0 (PC O-16:0/2:0) - PAF			12.33		12.00 [3]	5.67 [2]		
PG 34:1 (PG 16:1_18:0)					8.67 [1,3]			2.33 [1,3]
Phenyl-Alanine	9.00 [3]			1.67 [3]				
Pro-hydroxyPro					11.33 [3]			
Theobromine						8.67 [1,2]		5.67 [1,2]
Trans-4-Aminocyclohexanecarboxylic acid					11.00 [3]			
Trichloroethyl	14.67					4.33 [2]		2.67
Triethylene glycol dimethacrylate	30.00 [1,3]				7.00 [3]		5.67 [1,3]	
Trimethylolmelamine			7.67					
Tryptophan			1.67					
Tryptophan betaine		2.00 [1,3]						
Tyrosine	19.00							
Valine		5.00					8.33	

Table 3. **Data from the predictive models constructed and the variable influence scores of the selected metabolites for each model.** Those metabolites that were not selected for any model are not included in this table. Model MaxAUC values: 0.5–06, bad model; 0.6–0.75, regular model; 0.75–0.9, good model; 0.9–0.97, very good model; 0.97–1, excellent model. [1]*p*-value < 0.05; [2] FC < 0.75; [3] FC > 1.25. Abbreviations: MaxAUC, maximal area under the curve; AU, arbitrary units; FC, fold change.

metabolites did not cluster according to the gene affected but rather according to the nature of the metabolite (lipids, amino acids, etc.).

Discussion

In the present study, we characterized the metabolomic phenotype associated with the HRD genotype (i.e., mutational status) in healthy individuals with a familial history of HBOC for the first time. In particular, our results have allowed us to identify the metabolomic profile characteristics of healthy individuals who carry a PV in HRR genes. To our knowledge, this study is the first in which a metabolomic profile with potential predictive capacity of the HRR germline mutational status in healthy individuals has been described.

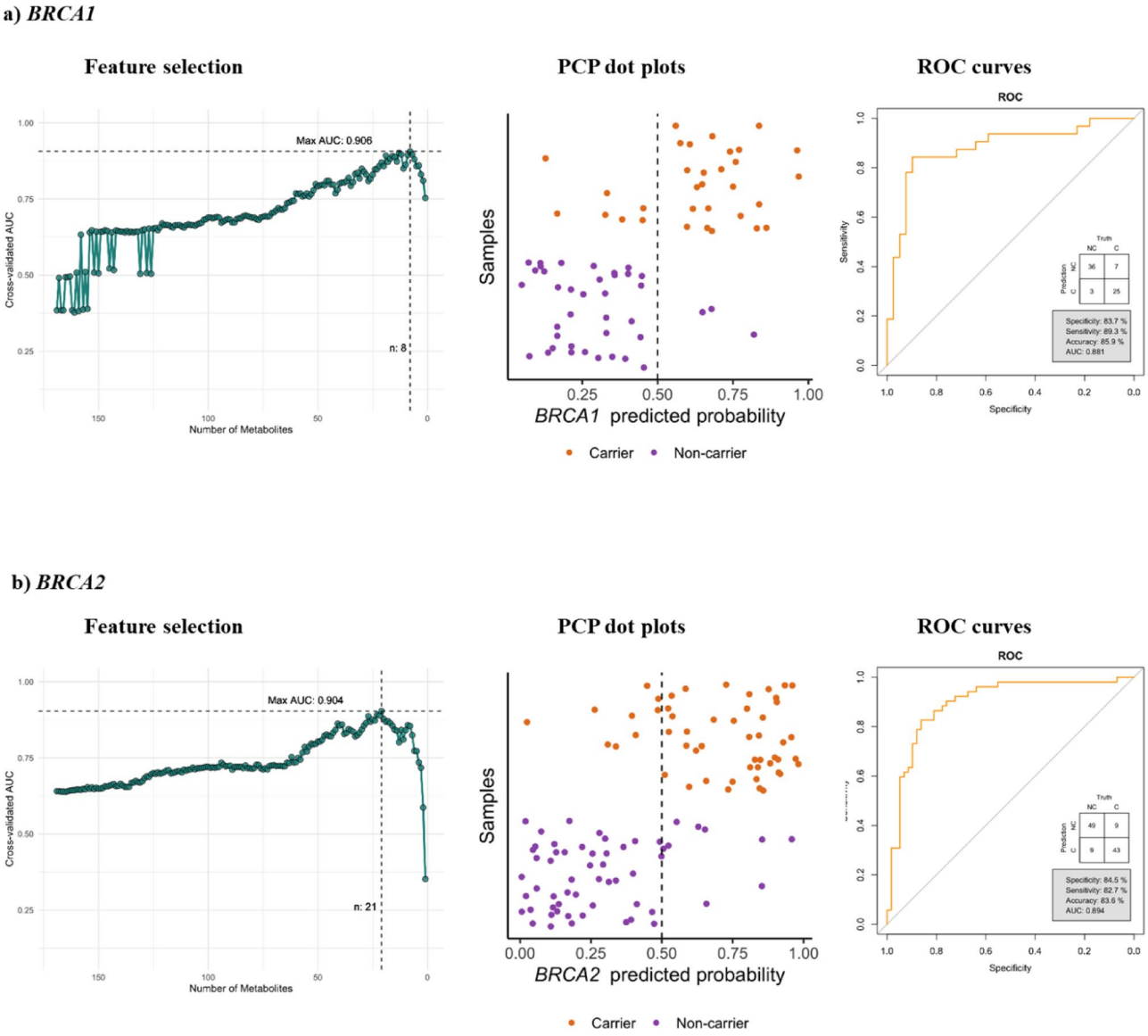


Fig. 2. Machine learning approach based on a linear support vector machine (SVM). (**a**) *BRCA1* model. (**b**) *BRCA2* model. A linear SVM was used to compute the predicted class probability for carriers vs. noncarriers for each subset of samples. First, in the feature selection step, the recursive feature elimination (RFE) method was chosen with 3-fold cross-validation for the iterative removal of the least important features (metabolites) to increase model simplicity. The maximum predictive performance, measured by the maximum area under the receiver operating characteristic curve (maxAUC) with the minimal number of metabolites, was used to select the final metabolite set. The VI score for each selected metabolite was calculated using the coefficients (weights) calculated by the SVM model. The individual maxAUC for each selected metabolite was also calculated from its own abundance. Second, the final predictive model was built with the VI metabolites selected in the feature selection step using a 3-fold cross-validation SVM. The receiver operating characteristic (ROC) curve and the predicted carrier probability (PCP) dot plot were constructed using the predicted SVM scores. The confusion matrix, specificity, sensitivity and accuracy were calculated using a 0.5 probability threshold.

A large body of knowledge exists that provides evidence of the possibility of using metabolomic profiles for tumour characterization or the identification of prognostic factors in breast cancer patients[14,17,18,23–25]. However, scarce evidence exists on the metabolome in healthy individuals or on the possibility of using metabolomic profiles as potential early detection biomarkers in asymptomatic patients. Moreover, most published metabolomic studies have focused on the analysis of tumour tissues or cell lines[14,19–21,26], and very few have used the plasma or serum of subjects. These studies reported the associations among certain metabolites and prognosis and disease progression, among other factors, of patients with breast cancer[17,18,27–31], as comprehensively described below.

In the present study, a total of 260 healthy participants with a familial history of hereditary cancer were enrolled. In this population, we profiled a total of 285 plasma metabolites, 169 of which were subjected to a

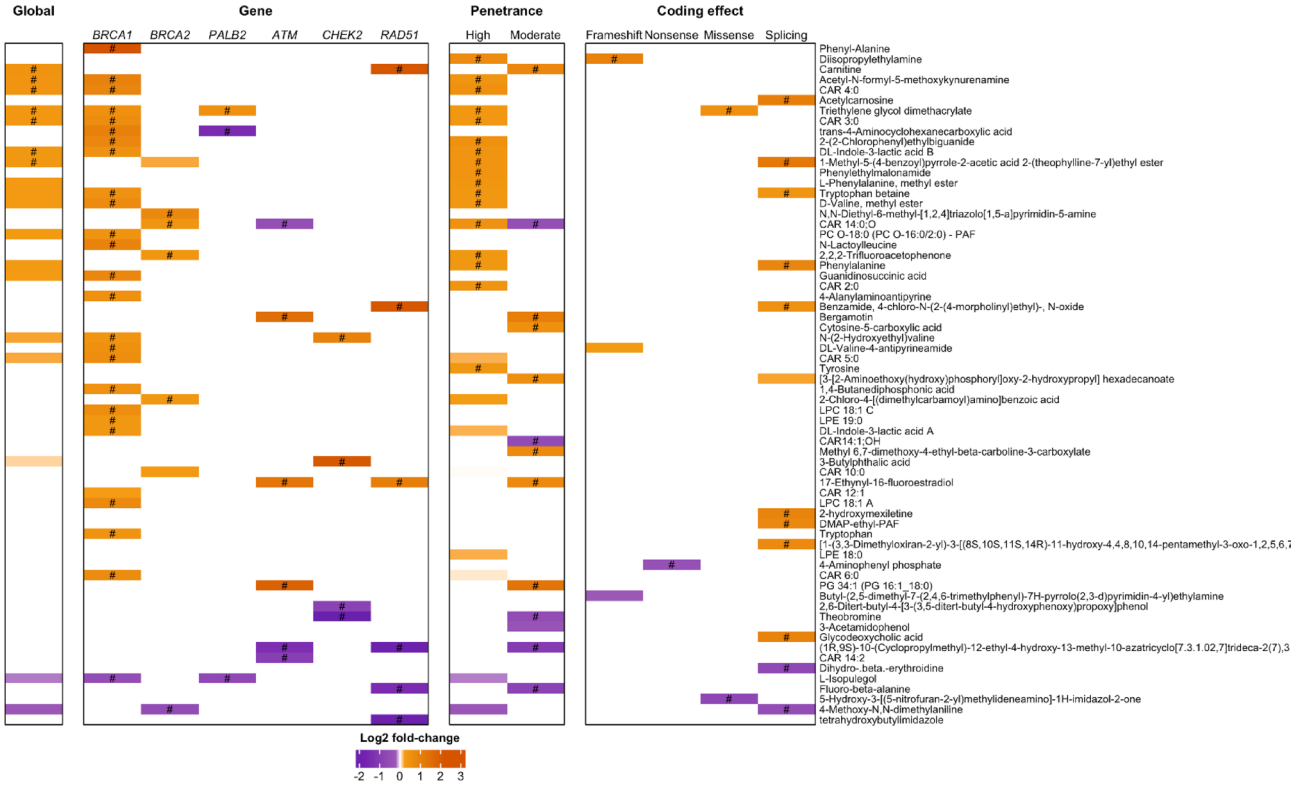


Fig. 3. Carrier heatmap of the profiled metabolites in this study that met the FC criteria (FC > 1.25 or FC < 0.75) in any of the comparisons performed. The colour gradient ranges from purple (FC greater in noncarriers) to orange (FC greater in carriers). Comparisons with a *p*-value < 0.05 are indicated with the # symbol.

further comprehensive investigation. Considering that our population was composed only of healthy individuals, the exploratory analysis surprisingly identified 7 metabolites that differentiated the subpopulations (carriers and noncarriers) regardless of the gene affected. Similar results were found when the data were segregated by gene or gene penetrance, with different numbers of differentiating metabolites. For example, a considerable number of metabolites (18) allowed us to differentiate *BRCA1* carriers but only 2 metabolites were identified in *BRCA2* carriers, with both genes traditionally linked to HBOC.

On the basis of our preliminary results, we aimed to further evaluate the existence of a differential and characteristic metabolomic profile that could infer the HRD genotype in healthy individuals with a family history of HBOC for the first time. With these data, predictive models were constructed using a machine learning approach. Our data revealed that while the model constructed with the whole dataset showed a moderate predictive capacity, those models constructed when the data were segregated by gene or gene penetrance demonstrated greater predictive performance. In general, models with maxAUCs > 0.70 are considered predictive with clinical relevance. In particular, maxAUC values of 0.75–0.90 are indicative of good models, maxAUC values of 0.91–0.97 indicate very good models, and maxAUC values > 0.97 are specific to excellent predictive models[32]. Therefore, the constructed models for predicting *BRCA1*, *BRCA2*, and *PALB2* mutational status were good, whereas those constructed for the *ATM* and *CHEK2* mutational status were excellent. Interestingly, models predicting the mutational status of high- or moderate-penetrance genes resulted in regular and very good models, respectively. In addition, up to 31 metabolites were needed to construct the global predictive model (with a low accuracy of 61.9%), whereas fewer metabolites were needed to construct the models with the best performance, mainly the *CHEK2* (9 metabolites; accuracy = 94.1%) and *ATM* (20 metabolites; accuracy = 92.3%) models.

These findings have several considerations. Metabolic heterogeneity exists in healthy individuals irrespective of their mutational status (carrier or noncarrier). On the other hand, such heterogeneity in turn subsides when individuals are grouped according to the gene affected. An emerging hallmark of cancer is the ability of cells to modify or reprogram cellular metabolism to most effectively support the uncontrolled neoplastic proliferation that defines it[33]. In particular, breast cancer prognosis and metastasis involve a complex network in which glucose, amino acid, fatty acid, and cholesterol metabolism are affected[14]. Although only healthy individuals were included in the present study, metabolic changes can precede cancer onset[15,16]. Namely, in asymptomatic individuals, subtle metabolomic changes emerge that can be detected only using a metabolomic approach because of the inherent sensitivity of metabolomics. These mild alterations in biological pathways might provide insights into the biochemical mechanisms underlying the preclinical steps of cancer development.

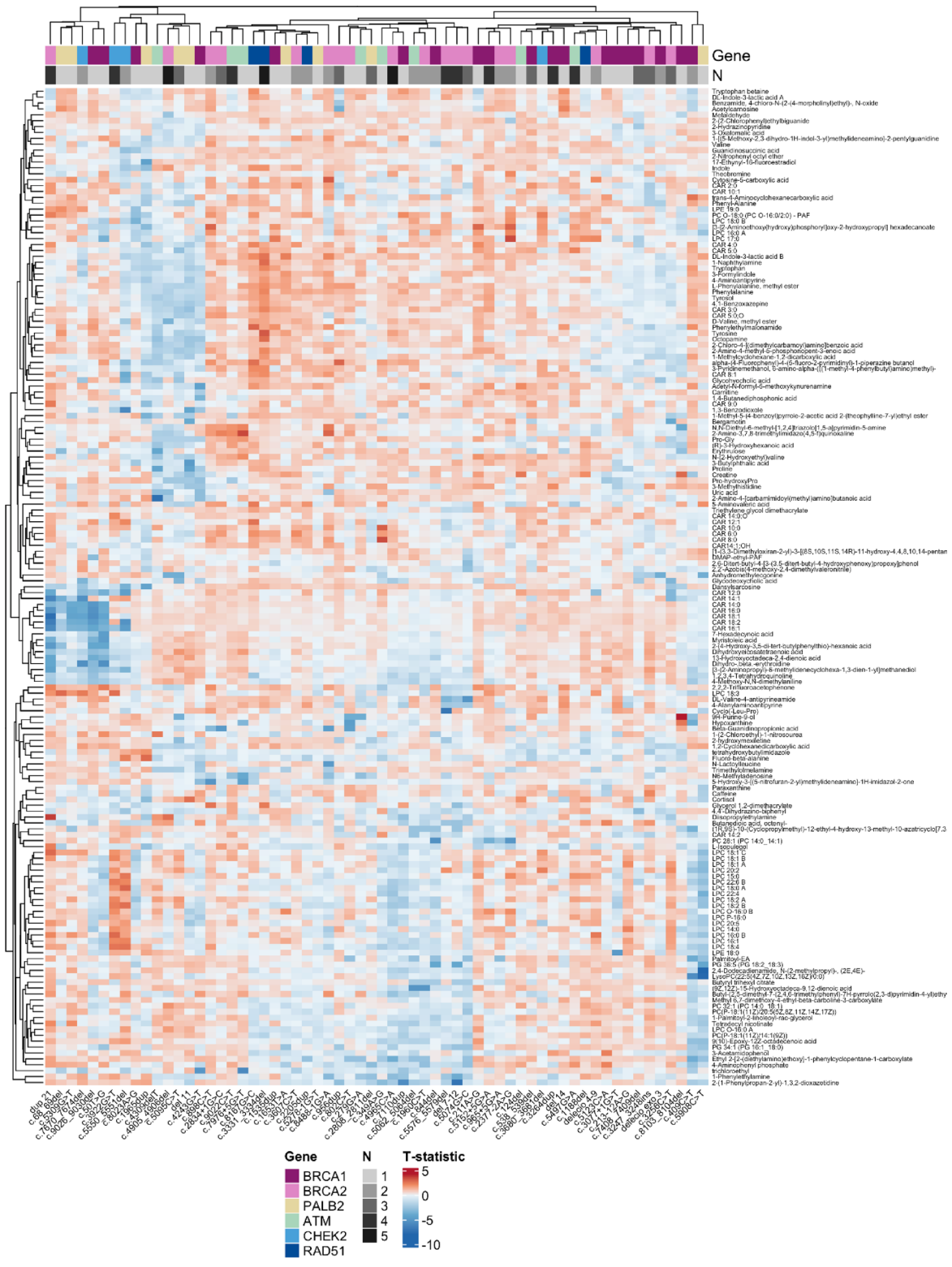


Fig. 4. Mutation heatmap constructed with data obtained from the 130 carrier individuals included in our study according to their identified PVs. The t statistic value from the logistic regression analysis is graphically represented as a hierarchically clustered heatmap. The colour gradient indicates the t statistic value ranging from red (higher t statistic value) to blue (lower t statistic value). N refers to the number of participants harbouring a particular PV.

In our study, metabolomic heterogeneity became evident in the mutational heatmap, where clusters were not arranged according to the gene affected but rather according to the nature of the metabolites. In addition, metabolic heterogeneity in our population became more evident when we examined the VI metabolites obtained in each predictive model in depth. In this sense, of the thirty-one VI metabolites selected by the global model, only 14 were shared with the other predictive models constructed here. More interestingly, 17 metabolites were not shared with other predictive models and were thus specific for predicting the global mutational status. The same condition applied to all the predictive models constructed since only a few metabolites, if any, were jointly selected by these models. This selection occurred for several lysophosphatidylcholines (LPCs) and carnitines (CARs). In our study, although some LPCs were considered VI metabolites in the global model only (for example, LPC 15:0), other LPCs were also considered VI metabolites in other predictive models (as in the case of LPC 18:1 A, which was also present in the *BRCA1* and high-penetrance models). LPCs are a class of chemical compounds produced by the enzyme phospholipase A2, which removes one of the fatty acids from phosphatidylcholine to produce LPCs. Among other functions, LPCs serve as substrates to generate lysophosphatidic acid, which, in turn, promotes tumour progression via cell proliferation, invasion, metastasis, tumorigenicity, and angiogenesis, as has been recently reviewed[34]. CAR3:0 and CAR 4:0 were specific metabolites for the global prediction model, whereas CAR 14:0;O and L-carnitine were also present in the *PALB2* and *ATM* models, respectively. Additional CARs were not considered VI metabolites in the global model, but they were included in other predictive models, such as CAR 6:0, CAR 12:0, and CAR 14:2 (*ATM* model), CAR 2:0 and CAR16:1 (*BRCA2* model), CAR 5:0 (*CHEK2*), and CAR14:1;OH (moderate-penetrance model). CARs are a group of metabolites that, among other functions, transport long-chain fatty acids across the mitochondrial membrane for energy production. The increase in the abundance of metabolites involved in fatty acid transport and synthesis has been observed in breast cancer patients and they have thus been proposed as potential biomarkers of breast cancer diagnosis and treatment[35]. In addition, CARs regulate cellular metabolism by participating in the conversion and utilization of different fuel sources, enabling cells to switch between carbohydrate and fatty acid metabolism as needed[36]. Thus, CARs regulate ketone body production, a process intimately linked to the metabolic adaptation of cancer cells, as we recently reported[37]. Moreover, as recently reviewed, carnitine levels are higher in HER2-positive patients than in HER2-negative patients[14]. Therefore, in our study, we reported lipid metabolites that can both predict the mutational status in healthy individuals and are known to be closely associated with cellular metabolism reprogramming in cancer cells.

Accumulating evidence has identified the metabolomic changes associated with cancer onset and development in breast cancer patients, especially in tumour tissues or cell lines. In this context, Privat et al. reported that glycolysis is increased in *BRCA1*-mutated breast cancer cell lines, allowing tumour transformation[38]. Shen et al. observed a differential metabolomic plasma profile of patients and controls conditioned by race, status, and hormone receptor status[21]. Asiago et al. developed a predictive model for early breast cancer detection based on 11 metabolites that represent some of the changes in the metabolic activity of several pathways associated with cancer. Among these 11 metabolites, two involved in amino acid metabolism were decreased (proline) and increased (tyrosine) in patients with recurrent breast cancer[20]. Moreover, proline and valine have been reported to be major amino acids involved in the prediction of molecular subtypes of breast cancer. In this context, proline and valine levels are increased in HER2-positive breast cancer patients, whereas valine levels are decreased in ER-positive patients compared with ER-negative patients[14]. Consistent with these results, in our study, proline levels were increased (FC > 1.25) in carriers when all datasets were considered; in *BRCA1*, *BRCA2*, and *PALB2* carriers when the data were segregated by gene; and in high-penetrance gene comparisons when the data were segregated by penetrance. In addition, we found that tyrosine levels were also increased in *BRCA1* and in high-penetrance comparisons, with the latter being statistically significant. More interestingly, tyrosine was one of the 31 VI metabolites selected for the global predictive model, and valine was a vital VI metabolite in the *BRCA1* and high-penetrance prediction models. Substantial evidence has demonstrated that amino acids contribute to tumorigenesis and tumour immunity by acting as nutrients, signalling molecules, gene transcription regulators, and epigenetic modifiers[39].

Yang et al. reported a differential lipidomic profile in plasma capable of classifying patients with benign breast tumours and pathogenic cancer. These authors reported that PG 34:1 levels, among other factors, were significantly increased in groups with benign breast tumours and breast cancer compared with healthy controls, with FCs ranging from 1.79 to 2.07, respectively[19]. In our study, the PG 34:1 (16:1; 18:0) level was significantly increased in both *ATM* carriers (FC = 3.47; p value = 0.02) and moderate-penetrance carriers (FC = 2.59; p value = 0.01). More interestingly, this lipid was considered a VI metabolite in both predictive models (VI scores of 8.67 and 2.33, respectively). In general, PGs act as precursors for the synthesis of cardiolipin, and they are essential for the optimal function of several enzymes involved in mitochondrial energy metabolism. Lipids seem to be among the major metabolites involved in breast cancer onset and development. As reviewed by Subramani et al., LPC levels are correlated with a lower risk of breast cancer; therefore, the cellular level of LPC could serve as a good predictor of breast cancer risk[35]. Moreover, as reviewed by Alvarez-Frutos et al., lower plasma levels of LPCs were observed in cancer patients than in healthy controls. Interestingly, LPCs have been associated with cancer onset[25] and breast cancer subtypes, since LPC16:1 levels are decreased in HER-2-positive patients, whereas the opposite trend is observed for the LPC 20:4 levels in these patients[14]. In our study, LPC 16:1 levels were increased in carriers of PVs in *PALB2* and *CHEK2* (FC = 1.276 and FC = 1.363, respectively), although these increases did not reach statistical significance.

Although several authors have reported that these metabolomic changes are associated with cancer onset and development, there is limited evidence on the metabolomic changes associated with the mutational status of healthy individuals. Penkert et al.. reported the presence of different metabolites in healthy women who were carriers of a PV in *BRCA1* compared with noncarriers, revealing the role of *BRCA1* in cellular metabolism beyond its known role in HRR[40]. In particular, Penkert et al. found that pyruvate levels were increased in

carrier individuals, whereas the RI1984/pyruvic acid and lactic acid/pyruvic acid ratios were significantly decreased in these individuals. None of these metabolites were identified in our study, surely due to the dissimilar methodologies used in both studies (gas chromatography–mass spectrometry (GC–MS) and liquid chromatography–mass spectrometry (LC–MS)). Our group previously performed a nontargeted metabolomic study on breast cancer cell lines from carriers and noncarriers of a PV in *BRCA1*. We reported for the first time that these cells have different metabolomic profiles based on *BRCA1* functionality. Based on the results of the in vitro study, we subsequently performed a targeted metabolomic study of patients with HBOC who were carriers or noncarriers of a PV in *BRCA1* to determine their metabolomic signature and to assess the contribution of the identified metabolites to the genetic diagnosis of breast cancer. In these studies, we demonstrated the existence of free methylated nucleotides (N6-methyladenosine and 1-methylguanine) capable of distinguishing the plasma samples of carrier patients with HBOC from those of noncarrier patients[22]. Interestingly, in the present study, N6-methyladenosine was present in the global predictive model. In our previous study, we found that the plasma levels of these free methylated nucleotides were significantly lower in patients who were carriers of PVs in *BRCA1* than in noncarrier patients. In the present study, the plasma N6-methyladenosine levels were lower in noncarriers than in carriers in all the comparisons except for the *BRCA1*- and *RAD51*-specific comparisons, although these changes were not statistically significant. The differences in the N6-methyladenosine levels between the two studies may be due to the different populations studied. While Roig et al. enrolled breast cancer patients, in the present study, we included only healthy subjects.

The main limitation of our study is the dissimilar number of individuals harbouring a PV in different HRR genes, especially *RAD51*, as the limited availability of individuals in this subgroup hindered predictive model construction. However, such versatility in our set of individuals represents the natural variety found in the general population. A major strength of our study is the fact that the population included was composed of healthy individuals without cancer. Therefore, the metabolic changes observed here are of paramount importance since these changes are associated with the genotype of the individuals rather than with cancer onset and development. More studies are warranted to identify those individuals included in this study who may develop breast or ovarian cancer in the future.

In conclusion, our study is the first to establish a link between the metabolome (phenotype) and genetic alterations (i.e., mutational status) in healthy carriers of PVs in genes involved in HRR with a familial history of HBOC. In particular, we reported amino acids, lipid metabolites, and methylated nucleotides that can both predict the mutational status of these healthy individuals and are known to be closely associated with cellular metabolic reprogramming in cancer cells. This characteristic and distinctive metabolomic profile could infer the HRD genotype in these healthy individuals. Overall, a seemingly safe assumption is that the metabolomic heterogeneity observed here may become a cornerstone in inferring the mutational status in healthy subpopulations with a familial history of HBOC syndrome.

Methods

Participant selection

A total of 260 individuals were selected from the *Institut d'Oncologia de la Catalunya Sud (IOCS)*. The inclusion criteria included the following: (1) healthy individuals; (2) carriers and noncarriers of PVs in genes involved in the HRR DNA repair pathway; (3) a familial history of HBOC; (4) fulfilling the clinical criteria to undergo cascade genetic testing according to the Catalan Health Service guidelines regarding the determination of genetic profiles of hereditary cancer syndromes in adults and paediatrics[41]; and (5) signed informed consent forms. Individuals were allocated to one of the following groups (n = 130 each): (a) individuals carrying PVs in genes responsible for HRR or (b) individuals who did not carry PVs in these genes.

The study was performed in accordance with the Declaration of Helsinki and was approved by the Clinical Research Ethics Committee of Sant Joan University Hospital (Reus, Spain) (ref: 105/2022). Written informed consent was obtained from all participants.

Sample collection

A venous blood sample (10 mL) was collected from each patient by venepuncture in Vacutainer™ collection tubes with K_2EDTA anticoagulant (Becton, Dickinson, and Company, Franklin Lakes, NJ, USA). The plasma fraction was isolated by centrifugation. DNA was extracted from peripheral blood lymphocytes using the Gentra® PureGene DNA Isolation Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. The DNA concentration and purity were measured with a Nanodrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

Mutational status

The mutational status was analysed using automatic capillary (Sanger) sequencing with a BigDye™ Terminator v3.1 kit (Life Technologies) in a SeqStudio sequencer (Applied Biosystems) and analysed using Sequencher v5.0 software (Gene Codes Corporation). The clinical significance of the PVs included in this study was examined according to a 5-tier classification system[42]. This classification was performed according to the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP) standards and guidelines[43], the Cancer Variant Interpretation Group UK guidelines (CanVIG-UK), and the ClinGen Variant Expert Curation Panel specifications for *BRCA*s[44] and *ATM*[45]. In silico predictive studies were performed with tools recommended elsewhere[43,46]. The predicted consequences of splice variants were determined mainly by Splice AI. Variants were examined in databases such as ClinVar, OMIM, the Leiden Open Variation Database, and BRCA Share and by reviewing updated bibliographies. Variant frequencies were obtained from the Genome Aggregation Database (gnomAD) browser and the 1000 Genomes Project. The consequences of missense variants were examined with REVEL[47].

Metabolite extraction, analysis, and identification

The untargeted metabolomic analysis of the blood plasma of the individuals was assessed using LC–MS. Sample preparation, acquisition and compound identification were performed by oloBion® (Barcelona, Spain) using proprietary software (oloMAP® v.1.0). Each plasma sample (20 µL) was mixed with 188 µL of cold acetonitrile/H_2O/isopropanol (2:2:3) containing a mixture of internal standards, and the resulting sample was shaken for 6 min at 4 °C and 2000 rpm. The samples were subsequently centrifuged for 30 s at 14,000 rpm and evaporated to dryness. The residue was reconstituted in acetonitrile/H_2O (4:1), stored at 20 °C for 1 h, sonicated for 5 min, centrifuged for 10 min at 14,000 rpm, filtered through 0.2 µm RC filters (Phenomenex, Spain), and analysed by high-resolution MS using a 1290 Infinity II ultrahigh-performance liquid chromatograph (Agilent Technologies, CA, USA) coupled with a 6560 Ion Mobility Q-TOF mass spectrometer (Agilent Technologies, CA, USA). The metabolites were separated at 45 °C on a Waters Acquity UPLC BEH C18 column (100 mm length × 2.1 mm id; 1.7 µm particle size) equipped with a Waters Acquity VanGuard BEH C18 precolumn (5 mm × 2.1 mm id; 1.7 µm particle size) using mobile phases A (water with 0.1% formic acid) and B (acetonitrile with 0.1% formic acid) and gradient elution as follows: initial conditions of 0.5% B for 0–0.1 min; increase to 80% B from 0.1 to 10 min; increase to 99.5% B from 10 to 10.1 min; hold at 99.5% B from 10.1 to 12 min; and hold at 0.5% B from 12 to 14.5 min. The mobile phase flow rate was set to 0.3 mL/min, and the injection volume was 15 µL. Sample temperature was maintained at 4 °C. Following separation, the eluate was introduced into the mass spectrometer for electrospray ionization (ESI) in positive mode with the following parameters: capillary voltage, ± 3 kV; gas temperature, 250 °C; drying gas (nitrogen), 13 L/min; nebulizer gas (nitrogen), 50 psi; sheath gas temperature, 315 °C; sheath gas flow (nitrogen), 12 L/min; and acquisition rate, 1 spectrum/s. For metabolite identification, MS/MS spectra were collected at collision energies of 10, 20 and 50 eV with an MS1 acquisition rate of 4 spectra/s (100 ms) and an MS/MS acquisition rate of 3 spectra/s (77 ms) with 4 precursor ions per cycle. Quality control samples were used every 10 samples to ensure that the equipment was working properly and to be able to correct any instrumental drift. The compounds were identified by comparison with available databases, such as BinBase and NIST20.

Metabolomic data processing

After database curation, features with a percentage of missing values greater than 20% were excluded from further analysis. The remaining metabolites were imputed with the GBM method and normalized to their total abundance (metabolic total ion count (mTIC) method). For the multivariate analysis, log2 transformation and autoscaling were also performed.

Statistical analysis

Univariate statistical analyses were performed with the Mann–Whitney–Wilcoxon test. Metabolites were compared between each PV sample and the pooled ($N = 130$) noncarrier samples by performing one-hot encoding of the PV samples for the logistic regression analysis. The t statistic value from the logistic regression analysis was then graphically represented in a hierarchically clustered heatmap. FCs were calculated as the ratio of the normalized metabolite peak intensities (abundances) of carriers/noncarriers. $FC > 1.25$ and $FC < 0.75$ values were selected arbitrarily; since participants were healthy individuals, no higher FCs were expected. All analyses were performed with the entire dataset, with data segregated by gene (*BRCA1*, *BRCA2*, *PALB2*, *ATM*, *CHEK2*, and *RAD51*), and with data segregated by penetrance (high-penetrance genes: *BRCA1*, *BRCA2*, and *PALB2*; moderate-penetrance genes: *ATM*, *CHEK2*, *RAD51*). R software (v.4.4.2) with the R packages stats (4.4.2), ggpubr (v.6.0), and ComplexHeatmap (v.2.20.0) were used for data analysis.

Predicted probability of the germline mutational status

A machine learning approach based on a linear SVM was used to compute the predicted class probability for carriers vs. noncarriers for each subset of samples. First, in the feature selection step, recursive feature elimination (RFE) was chosen with 3-fold cross-validation for the iterative removal of the least important features (metabolites) to improve the simplicity of the model. The maximum predictive performance, measured by the maxAUC with the minimal number of metabolites, was used to select the final metabolite set. The VI score for each selected metabolite was calculated using the coefficients (weights) calculated by the SVM model. The individual maxAUC for each selected metabolite was also calculated from its own abundance. Second, the final predictive model was built with the VI metabolites selected in the feature selection step by a 3-fold cross-validation SVM (Supplementary Fig. S3). The ROC curve and the predicted carrier probability (PCP) dot plot were constructed using the predicted SVM scores. The confusion matrix, specificity, sensitivity and accuracy were calculated using a 0.5 probability threshold. As in the univariate analyses, the machine learning approach was performed with the entire dataset, with the data segregated by gene and by penetrance. R software was used (v.4.4.2), with the R packages mlr3verse (v.0.3.1), mlr3learners (v.0.9.0), mlr3extralearners (v.1.0.0), mlr3fselect (v.1.3.0), mlr3pipelines (v.0.7.1) and pROC (v.1.18.5). The number of seeds used for reproducibility purposes was 3 (set.seed).

Data availability

The datasets used in the current study are available from the corresponding author upon reasonable request.

Received: 5 August 2025; Accepted: 8 January 2026
Published online: 31 January 2026

References

1. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715 (1993).
2. Toss, A. et al. Hereditary ovarian cancer: Not only BRCA 1 and 2 Genes. *Biomed Res. Int.* (2015).
3. Goyal, G., Fan, T. & Silberstein, P. T. Hereditary cancer syndromes: utilizing DNA repair deficiency as therapeutic target. *Fam Cancer*. **15**, 359–366 (2016).
4. Lord, C. J. & Ashworth, A. BRCAness revisited. *Nat. Rev. Cancer*. **16**, 110–120 (2016).
5. Sociedad Española de Oncología Médica (SEOM). Cáncer de Mama y Ovario Hereditario. https://seom.org/informacion-sobre-el-cancer/consejo-genetico?start=4 (2021).
6. Fackenthal, J. D. et al. Naturally occurring BRCA2 alternative mRNA splicing events in clinically relevant samples. *J. Med. Genet.* **53**, 548–558 (2016).
7. Kuchenbaecker, K. B. et al. Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. *JAMA - J. Am. Med. Assoc.* **317**, 2402–2416 (2017).
8. Díez, O. Síndromes hereditarios en oncología. *Hosp. la St. Creu i St. Pau. Barcelona Divers.* III, 664 (2016).
9. Matanes, E. et al. Endometrial cancer in germline BRCA mutation carriers: A systematic review and Meta-analysis. *J. Minim. Invasive Gynecol.* **28**, 947–956 (2021).
10. LOVD. LOVD Variants Database. https://www.lovd.nl/
11. Miki, Y. et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**, 66–71 (1994).
12. Wooster, R. et al. Identification of the breast cancer susceptibility gene BRCA2. *Nature* **378**, 789–792 (1995).
13. National Human Genome Research Institute. Breast cancer information core. https://research.nhgri.nih.gov/bic/
14. Kaur, R., Gupta, S., Kulshrestha, S., Khandelwal, V. & Pandey, S. Metabolomics-driven biomarker discovery for breast cancer prognosis and metabolomics-driven biomarker discovery for breast cancer prognosis and diagnosis. *Cells* 14, (2025).
15. Johnson, C. H., Ivanisevic, J., Siuzdak, G. & Metabolomics beyond biomarkers and towards mechanisms. *Nat. Rev. Mol. Cell Biol.* **17** 451–459 https://doi.org/10.1038/nrm.2016.25 (2016).
16. Ni, Y., Xie, G. & Jia, W. Metabonomics of human colorectal cancer: new approaches for early diagnosis and biomarker discovery. *J. Proteome Res.* **13**, 3857–3870 (2014).
17. Sitter, B. et al. Quantification of metabolites in breast cancer patients with different clinical prognosis using HR MAS MR spectroscopy. *NMR Biomed.* **23**, 424–431 (2010).
18. Shaw, P. G. et al. Integrated proteomic and metabolic analysis of breast cancer progression. *PLoS One*. **8**, 1–12 (2013).
19. Yang, L. et al. Comprehensive lipid profiling of plasma in patients with benign breast tumor and breast cancer reveals novel biomarkers. *Anal. Bioanal Chem.* **407**, 5065–5077 (2015).
20. Asiago, V. M. et al. Early detection of recurrent breast cancer using metabolite profiling. *Cancer Res.* **70**, 8309–8318 (2010).
21. Shen, J., Yan, L., Liu, S., Ambrosone, C. B. & Zhao, H. Plasma metabolomic profiles in breast cancer patients and healthy controls: by race and tumor receptor subtypes. *Transl Oncol.* **6**, 757–765 (2013).
22. Roig, B. et al. Metabolomics reveals novel blood plasma biomarkers associated to the BRCA1-mutated phenotype of human breast cancer. *Sci. Rep.* **7**, 17831 (2017).
23. Wu, H. C. et al. Plasma metabolomics profiles and breast cancer risk. *Breast Cancer Res.* **26**, 141 (2024).
24. Neagu, A. N. et al. Omics-based investigations breast cancer molecules **28**, (2023).
25. Alvarez-Frutos, L. et al. Multiomics insights on the onset, progression, and metastatic evolution of breast cancer. *Front. Oncol.* **13**, 1–19 (2023).
26. Li, H. et al. The landscape of cancer cell line metabolism. *Nat. Med.* **25**, 850–860 (2019).
27. März, J. et al. Plasma metabolome profiling for the diagnosis of catecholamine producing tumors. *Front. Endocrinol. (Lausanne).* **12**, 1–11 (2021).
28. Drew, D. A. Metabolomics as a tool for biomarker discovery in gastric cancer. *Cancer Epidemiol. Biomarkers Prev.* **9**, 1601–1603 (2021).
29. Adams, C. D. et al. Circulating metabolic biomarkers of screen-detected prostate cancer in the protect study. *Cancer Epidemiol. Biomarkers Prev.* **28**, 208–216 (2019).
30. Lahiguera, Á. et al. Tumors defective in homologous recombination rely on oxidative metabolism: relevance to treatments with PARP inhibitors. *EMBO Mol. Med.* **12**, 1–23 (2020).
31. Bliziotis, N. G. et al. Pre- versus post-operative untargeted plasma nuclear magnetic resonance spectroscopy metabolomics of pheochromocytoma and paraganglioma. *Endocrine* **75**, 254–265 (2022).
32. Davey, G. K. et al. EPIC–Oxford:lifestyle characteristics and nutrient intakes in a cohort of 33 883 meat-eaters and 31 546 Non meat-eaters in the UK. *Public. Health Nutr.* **6**, 259–268 (2003).
33. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
34. Takai, M., Mori, S., Honoki, K. & Tsujiuchi, T. Roles of lysophosphatidic acid (LPA) receptor-mediated signaling in cancer cell biology. *J. Bioenerg Biomembr.* **56**, 475–482 (2024).
35. Subramani, R., Poudel, S., Smith, K. D., Estrada, A. & Lakshmanaswamy, R. *Metab. Breast Cancer Rev. Metab.* **12**, (2022).
36. Longo, N., Frigeni, M. & Pasquali, M. Carnitine transport and fatty acid oxidation. *Biochim. Biophys. Acta - Mol. Cell. Res.* **1863**, 2422–2435 (2016).
37. Fernández-Castillejo, S. et al. Article ketone bodies are potential prognostic biomarkers in Relapsed/ refractory diffuse large B-Cell lymphoma: results from the R2-GDP-GOTEL trial. *Cancers (Basel).* **17**, 532 (2025).
38. Privat, M. et al. BRCA1 induces major energetic metabolism reprogramming in breast cancer cells. *PLoS One* **9**, (2014).
39. Chen, J., Cui, L., Lu, S. & Xu, S. Amino acid metabolism in tumor biology and therapy. *Cell Death Dis.* **15**, (2024).
40. Penkert, J. et al. Plasma metabolome signature indicative of BRCA1 germline status independent of cancer incidence. *Front. Oncol.* **11**, 1 (2021).
41. Servei Català de la Salut. Determinacions del perfil genètic de les síndromes hereditàries de càncer en l'adult i pediatria. https://catsalut.gencat.cat/web/.content/minisite/catsalut/proveidors_professionals/documentacio-professionals/oncologia-precisio-panel-germinals.pdf (2021).
42. Plon, S. E. et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum. Mutat.* **29**, 1282–1291 (2021).
43. Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American college of medical genetics and genomics and the association for molecular pathology. *Genet. Med.* **17**, 405–424 (2015).
44. Parsons, M. T. et al. Evidence-based recommendations for gene-specific ACMG/AMP variant classification from the ClinGen ENIGMA BRCA1 and BRCA2 variant curation expert panel. *Am. J. Hum. Genet.* **111**, 2044–2058 (2024).
45. Richardson, M. E. et al. Specifications of the ACMG / AMP variant curation guidelines for the analysis of germline ATM sequence variants authors ARTICLE specifications of the ACMG / AMP variant curation guidelines for the analysis of germline ATM sequence variants. *Am J. Hum. Genet* **111**, 2411–2426 .
46. Rehm, H. L. et al. *ClinGen-The Clin. Genome Resource N engl. j. med.* **23** 372 (2015).
47. Ioannidis, N. M. et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am. J. Hum. Genet.* **99**, 877–885 (2016).

Acknowledgements
The authors would like to thank the technicians from the Biobanc-IISPV in Reus (http://www.iispv.cat) for sample management, oloBion SL (Barcelona, Spain) for their support with the metabolomic analysis, and Marta Pulido, MD, PhD, for editing the manuscript and editorial assistance supported by Institut d'Investigació Sanitària Pere Virgili (IISPV), Reus, Spain. We are grateful to all the individuals and families who kindly participated in this study.

Author contributions
BR, SFC and MRB formulated the research goals and aims of the present work and curated the data included. JG and MRB managed and coordinated the planning and execution of the research activities. JG was responsible for acquiring financial support for the project and was the supervisor of all the research activities. BR, SFC, JG, JB, MM, MS, MQ, RC, and MRB were involved in conducting the investigation by both collecting and analysing the data. SFC, JB, and RC performed the comprehensive statistical analyses. BR, SFC and MRB validated and reproduced the outcomes presented here and were responsible for the preparation and creation of the paper. SFC wrote the initial draft, and BR and MRB comprehensively reviewed it. Second and final revisions were performed by JG, JB, MM, MS, MQ, and RC.

Funding
This work was supported by *Lliga contra el Càncer*, Comarques de Tarragona i Terres de l'Ebre (Ref PR-010-2022).

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-35789-8.

Correspondence and requests for materials should be addressed to S.F.-C.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/.

© The Author(s) 2026