

Original Article

Open Access



Identification of molecular signatures and pathways of obese breast cancer gene expression data by a machine learning algorithm

Betul Comertpay, Esra Gov

Department of Bioengineering, Adana Alparslan Türkeş Science and Technology University, Adana 01250, Turkey.

Correspondence to: Assoc. Prof. Esra Gov, Department of Bioengineering, Adana Alparslan Türkeş Science and Technology University, Adana 01250, Turkey. E-mail: egov@atu.edu.tr

How to cite this article: Comertpay B, Gov E. Identification of molecular signatures and pathways of obese breast cancer gene expression data by a machine learning algorithm. *J Transl Genet Genom* 2022;6:84-94. <https://dx.doi.org/10.20517/jtgg.2021.44>

Received: 30 Aug 2021 **First Decision:** 9 Nov 2021 **Revised:** 29 Nov 2021 **Accepted:** 27 Dec 2021 **Published:** 20 Jan 2022

Academic Editors: Sanjay Gupta, Nathan A. Berger **Copy Editor:** Yue-Yue Zhang **Production Editor:** Yue-Yue Zhang

Abstract

Aim: Currently, the obesity epidemic is one of the biggest problems for human health. Obesity is impacted on survival in patients with breast cancer. However, key biomarkers of obesity-related breast cancer risk are still not well known. Thus, using machine learning to identify the most appropriate features in obesity-associated breast cancer patients may improve the predictive accuracy and interpretability of regression models.

Methods: In the present study, we identified 23 differentially expressed genes (DEGs) from the GSE24185 transcriptome dataset. Seed genes were identified from DEGs, the co-expression network genes and hub genes of the protein-protein interaction network. Pathway enrichment analysis was performed for DEGs. The Ridge penalty regression model was executed by using *P*-values of enriched pathways and seed gene pathway association score to obtain the most relevant molecular signatures. The model was performed using 10-fold cross-validation to fit the penalized models.

Results: Angiotensin II receptor type 1 (AGTR1), cyclin D1 (CCND1), glutamate ionotropic receptor AMPA type subunit 2 (GRIA2), interleukin-6 cytokine family signal transducer (IL6ST), matrix metalloproteinase 9 (MMP9), and protein kinase CAMP-dependent type II regulatory subunit beta (PRKAR2B) were considered as candidate molecular signatures of obese patients with breast cancer. In addition, RAF-independent MAPK1/3 activation, collagen degradation, bladder cancer, drug metabolism-cytochrome P450, and signaling by Hedgehog pathways in cancer were primarily associated with obesity-associated breast cancer.



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



Conclusion: These genes may be used for risk analysis of the disease progression of obese patients with breast cancer. Corresponding genes and pathways should be validated via experimental studies.

Keywords: Obesity, breast cancer, machine learning, penalty regression models

INTRODUCTION

Breast cancer is the second largest cause of mortality from cancer among women; however, detection at an early stage and treatment could significantly improve outcomes^[1]. The World Health Organization (WHO) stated that last year breast cancer was diagnosed in 2.3 million women worldwide and resulted in 685,000 deaths^[2]. It has a complex etiology that involves various genetic, physiological, and lifestyle-related risk factors (alcohol/smoking, excessive body weight, *etc.*)^[3,4]. Particularly, several studies have demonstrated the association between obesity status and breast cancer, highlighting the potential of an increase in personal health behaviors to reduce the burden of disease^[4]. In the WHO report, overweight and obesity are determined as a surplus fat aggregation that may harm to health. Body mass index (BMI) is a basic height-weight index mostly used to categorize overweight and obesity in adults (BMI > 30 kg/m²). According to the most recent WHO case report, currently, more than 1.9 billion adults and 650 million people worldwide can be categorized as overweight or obese, respectively, and these rates are predicted to increase more rapidly in the coming decades^[2].

Numerous studies have examined the association between obesity and cancer development in various cancer types, such as esophagus, pancreas, prostate, colorectal, and breast cancer^[5]. Although there is substantial evidence indicating a high BMI is linked to a growing risk of breast cancer in postmenopausal women and poorer clinical outcomes in people of all ages, the specific nature of the exposure is unknown.

This uncertainty is mirrored in the variety of methodologies used in the research to characterize or define body composition: BMI, body weight, body composition, metabolic state, and nutritional condition^[6].

Obesity is linked to a higher incidence of postmenopausal estrogen receptor-positive breast cancer and poorer cancer-associated results across the board^[7]. The obesity-cancer relationship is thought to be influenced by significant quantities of circulating and local estrogens, changed concentrations of adipokines^[8] (adiponectin and leptin), disrupted insulin/IGF signaling, changes in the microbiome, and local and systemic inflammatory effects (e.g., WAT)^[7]. The latest studies indicate that obesity-associated insulin/insulin-like growth factor-1 axis, adipokines, inflammatory cytokines and leptin, sex hormones^[9], adiponectin^[8], ORPS^[10], and HER2^[11] proteins play a significant role in breast cancer-related pathways. On the other hand, CD68 immunohistochemistry (CD68 + CLS-B) expression has been related to insulin resistance and negative prognosis in obesity-associated breast cancer^[12]. According to another study, vitamin D supplementation may have varied impacts on gene expression in breast and adipose tissue during weight loss^[13].

Obesity affects various aspects of breast cancer treatment, including surgery, chemotherapy, endocrine therapy, and radiotherapy. In addition, breast cancer risk and recurrence are affected by anti-inflammatory drugs, metformin, diet, and physical activity^[7,14]. Surgery, radiation, and chemotherapy problems are more common in obesity-associated breast cancer patients. Furthermore, obese patients have a higher chance of local recurrence than women of normal weight. Mechanistically driven approaches, involving biomarker development, are essential for the prevention and treatment of obesity-related malignancies, much as they are for tumor-directed pharmacologic therapy in oncology^[14].

Although various studies are being conducted to gain a better understanding of the association between obesity and breast cancer, integrative analysis is needed to detect novel molecular signatures and pathways to determine the obesity related breast cancer risk biomarkers.

In the present study, a gene expression dataset was analyzed to compare obesity-associated breast cancer samples and non-obesity-associated with breast cancer samples. The co-expression network and protein-protein interaction (PPI) network of differentially expressed genes (DEGs) were determined. Seed genes, common DEGs, were then identified from the co-expression gene network and hub genes of the PPI network. Next, to examine the molecular mechanisms of obesity-associated breast cancer, statistically significant pathways were determined. The Ridge penalty regression model was executed by using p-values of enriched pathways and seed gene pathway association score to determine the potential to be a molecular signature of seed genes in obese patients with breast cancer to obtain the most relevant molecular signatures. Finally, we identified several candidate genes and pathways in obese patients with breast cancer.

METHODS

Gene expression datasets and identification of differentially expressed genes

To characterize gene expression profiles of obesity in breast cancer, raw data of the obesity-related high-throughput gene expression dataset GSE24185^[15] in breast cancer were obtained from the Gene Expression Omnibus^[16]. In total, 74 samples were analyzed, including those from 36 historically normal (BMI \leq 24.9) breast cancer patients as a control sample and 38 obese patients with breast cancer (BMI \geq 30). The affy package of the R/Bioconductor platform (version 3.6) was used. Normalization for each dataset was performed with robust multiarray^[17] techniques. Normalized log-expression values, which were calculated using multiple test options of linear models for microarray data^[18] to define DEGs, were used in the statistical analysis of each dataset to contrast obese vs. non-obese breast cancer patients. For DEGs identification, they were selected according to computed *P*-values greater than the significance level (*P* value $<$ 0.05) with the fold change of 1.5 used as statistical threshold parameters.

Construction of co-expression networks in breast cancer and obese states

By separating the expression profiles of non-obesity-associated and obesity-associated breast cancer samples, two new data subsets were generated using the expression profiles of resultant DEGs. The co-expression network of DEGs was reconstructed by calculating the Pearson correlation coefficients of the mean expression values of DEGs in samples from obese patients with breast cancer and non-obese patients with breast cancer. To specify the statistical meaning of binary gene correlations, the obtained correlation coefficients were normally distributed (*P*-value $<$ 0.05), and positive and negative correlation cutoff significance levels (cutoffs $>$ 0.47 and \leq 0.47) were selected, respectively. An obesity-associated breast cancer-specific co-expression network was reconstructed, including 15 nodes and 17 edges, by using significant pairwise gene correlations.

PPI network reconstruction and identification of seed genes

The physical protein-protein interaction information was obtained from the BioGRID^[19] database, which includes 43,219 physical interactions associated with proteins. Resultant DEGs of PPI networks were reconstructed using Cytoscape^[20]. Seed genes were obtained from the intersection of DEGs, co-expressed genes, and hub genes of the PPI network.

Gene set overrepresentation analyses

Overrepresentation analyses were built using the ConsensusPathDB^[21] bioinformatics tool to determine biological processes, molecular functions, metabolic pathways, and signaling information crucially associated with DEGs of obese patients with breast cancer and seed genes. The Kyoto Encyclopedia of

Genes and Genomes^[22] (KEGG) and Reactome^[23] were used as a pathway database for the analyses. Statistically significant values ($P < 0.05$) representing the significance of enrichment analysis were obtained by Fisher's exact test.

Performance evaluation of the seed genes with a classification algorithm

The Ridge regression approach was used to understand the importance of seed genes in obese patients with breast cancer. This method is modeled as a linear weighted sum of biomarkers, performing a regularization punishment to limit the enormity of the regression coefficients. This gives rise to a sparse set of genes (i.e., biomarkers) that predict disease. This method limits the estimates of the regression coefficients towards no correlation to the maximum likelihood estimates. Ridge regression employs a penalization term to reduce overfitting. However, instead of using the sum of the absolute values, it uses the sum of the squares of the coefficients. As a result, under Ridge regression, the coefficients are not zero. The Ridge function is:

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 - \lambda \sum_{j=1}^P \beta_j^2$$

The machine learning algorithm was used to check the path validity of the identified common seed genes. To execute the regression algorithm, the NumPy^[24] and Pandas^[25] packages of the Python^[26] platform were used. Furthermore, to overcome the difficulty of insufficient data on obese patients with breast cancer with genotypes to train a high-performance model of risk prediction for obese patients with breast cancer, we interpret our recommended method at 10 replicates of five-fold cross-validation. Mathematically, Ridge regression can be defined by using a single penalty function " α "^[27]. A penalty parameter $\alpha = 0.1077$ was used in the Ridge algorithm. A high value for the penalty parameter (α) will result in a heavy penalty, leading to the selection of fewer variables. In addition, test size and random state were taken as 0.25 and 42, respectively.

RESULTS

Transcriptome profiling of obese patients with breast cancer

The statistical analyses of the gene expression dataset resulted in the identification of up- and downregulated DEGs with $P < 0.05$ and $FC > 1.5$ or $FC < 0.67$. Nineteen downregulated and four upregulated genes were identified. 4-Aminobutyrate aminotransferase (ABAT), beta polypeptide (ADH1B), angiotensin II receptor type 1 (AGTR1), cyclin D1 (CCND1), dual specificity phosphatase 4 (DUSP4), flavin containing dimethylaniline monooxygenase 2 (FMO2), FRY microtubule binding protein (FRY), polypeptide n-acetylgalactosaminyltransferase 7 (GALNT7), glutamate ionotropic receptor AMPA type subunit 2 (GRIA2), glycogenin 2 (GYG2), interleukin-6 cytokine family signal transducer (IL6ST), keratin 6B (KRT6B), mesoderm specific transcript (MEST), matrix metalloproteinase 12 (MMP12), matrix metalloproteinase 9 (MMP9), phospholamban (PLN), protein kinase CAMP-dependent type II regulatory subunit beta (PRKAR2B), ribonuclease a family member 4 (RNASE4), S100 calcium binding protein A2 (S100A2), signal peptide, CUB domain and EGF-like domain containing 2 (SCUBE2), semaphorin 3C (SEMA3C), tissue factor pathway inhibitor (TFPI), and transforming growth factor beta receptor 3 (TGFB3) were identified as DEGs. A small number of DEGs may have been obtained due to the study of the effect of obesity on tumor tissues.

Biological and clinical features of seed genes

Co-expression network analyses were performed, which identified 23 DEGs of obese patients with breast cancer samples obtained from the GSE24185 dataset [Figure 1]. Co-expressed genes were identified as

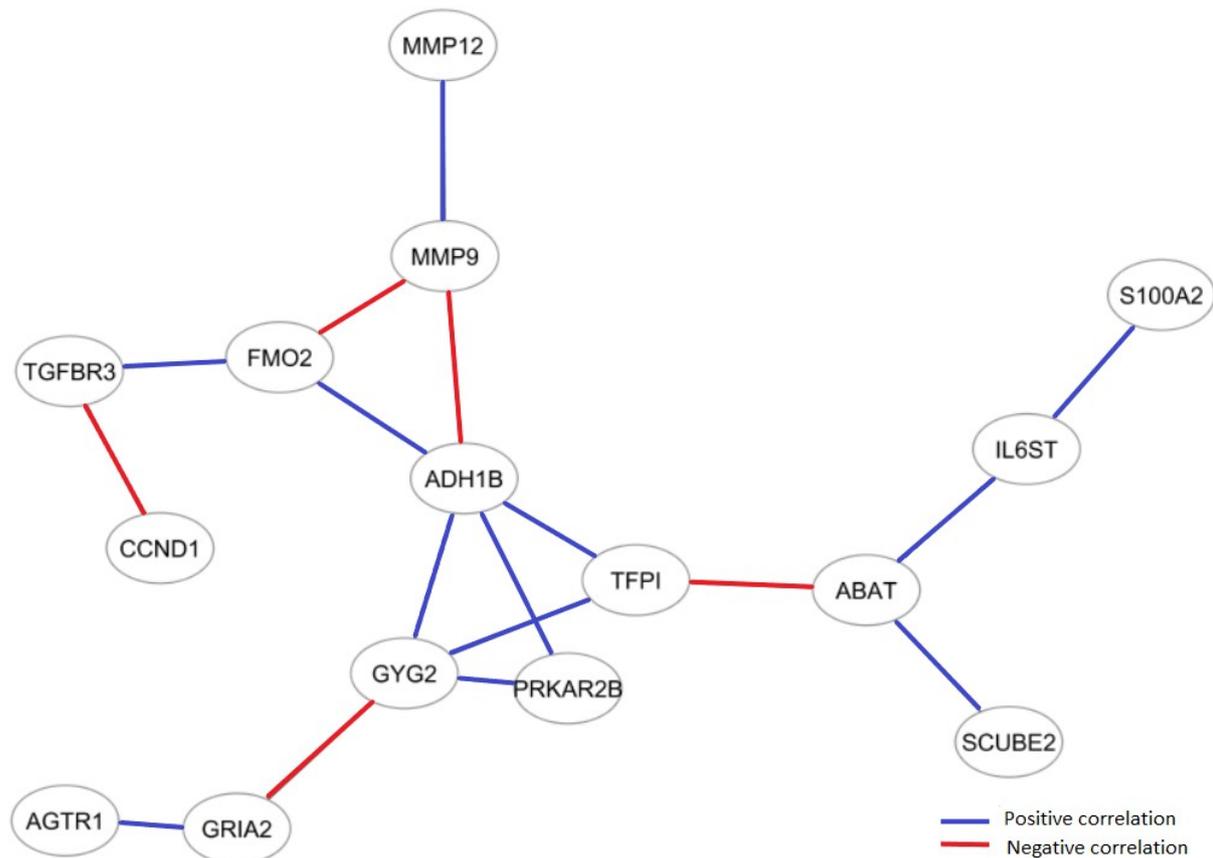


Figure 1. Co-expression network of the differentially expressed genes. Mutual differentially expressed genes (DEGs) that were significantly correlated are depicted as nodes, and statistically significant correlations between DEGs are represented as edges. Blue and red color edges represent positive and negative correlations, respectively.

ABAT, alcohol dehydrogenase 1B (Class I), ADH1B, AGTR1, CCND1, FMO2, GRIA2, GYG2, IL6ST, MMP12, MMP9, PRKAR2B, S100A2, SCUBE2, TFPI, and TGFBR3.

The first neighbor enriched PPI network was constructed by using DEGs [Figure 2]. Hub proteins with degree score ≥ 2 were determined as CCND1, PRKAR2B, IL6ST, PLN, GRIA2, S100A2, DUSP4, KRT6B, MMP9, AGTR1, and GYG2. Seed genes were also identified as common DEGs among co-expressed genes and hub genes of the PPI network [Figure 3A]. AGTR1, CCND1, GRIA2, GYG2, IL6ST, and PRKAR2B were downregulated while S100A2 and MMP9 were upregulated seed genes [Figure 3B]. The biological importance of the seed genes was described according to GeneCard^[28] in Table 1. The protein product of AGTR1 is a vasopressor hormone that affects the narrowing of blood arteries. CCND1 functions as a regulator of CDK kinases. Another seed gene, *IL6ST*, is a signal transducer and part of the cytokine receptor complex. GRIA2 and S100A2 are related to physiological processes, while *GYG2* and *PRKAR2B* are metabolism-related genes. It has been reported that *MMP9* is a metastasis-associated gene.

To identify important signaling pathways to the obesity-related carcinogenesis mechanism, pathway enrichment analysis was performed via KEGG and Reactome databases [Table 2]. Especially tumor-associated signaling pathways were obtained. RAF-independent MAPK1/3 activation, collagen degradation, bladder cancer, drug metabolism-cytochrome P450, and signaling by Hedgehog pathways in cancer were determined as significant pathways (P value < 0.01).

Table 1. Biological features of seed genes

AGTR1	It is a considerable effector in the cardiovascular system that controls blood pressure and volume
CCND1	It is associated with the cyclin family and their regulatory CDK kinases
GRIA2	It is activated in a variety of ordered neurophysiological processes
GYG2	The gene is associated with initiation reactions of glycogen biosynthesis and involved in blood glucose homeostasis
IL6ST	The activation of this protein is dependent on the binding of cytokines to their receptors
MMP9	Its family is concerned with the breakdown of the extracellular matrix in regular physiological processes (embryonic development, tissue remodeling, etc.) and disease processes such as metastasis
PRKAR2B	PKA is related to the organizing of lipid and glucose metabolism
S100A2	It may act as a calcium sensor and modulator, indirectly playing a role in various physiological processes

Table 2. Significant enrichment analyses results

Pathways	P-value	DEGs
RAF-independent MAPK1/3 activation	0.0006	DUSP4, IL6ST
Collagen degradation	0.0013	MMP12, MMP9
Bladder cancer	0.0018	CCND1, MMP9
Drug metabolism-cytochrome P450	0.0051	FMO2, ADH1B
Signaling by Hedgehog	0.0068	SCUBE2, PRKAR2B
Pathways in cancer	0.0077	CCND1, AGTR1, IL6ST, MMP9
Prostate cancer	0.0096	CCND1, MMP9
AGE-RAGE signaling pathway in diabetic complications	0.0100	CCND1, AGTR1
Degradation of the extracellular matrix	0.0112	MMP12, MMP9
Phase I functionalization of compounds	0.0116	FMO2, ADH1B
Thyroid hormone signaling pathway	0.0135	CCND1, PLN
Apelin signaling pathway	0.0185	CCND1, AGTR1
Adrenergic signaling in cardiomyocytes	0.0203	PLN, AGTR1
Hepatitis B	0.0203	CCND1, MMP9
Cushing syndrome	0.0231	CCND1, AGTR1
JAK-STAT signaling pathway	0.0253	CCND1, IL6ST
cGMP-PKG signaling pathway	0.0256	PLN, AGTR1
Calcium signaling pathway	0.0324	PLN, AGTR1
Kaposi sarcoma-associated herpesvirus infection	0.0327	CCND1, IL6ST
cAMP signaling pathway	0.0370	GRIA2, PLN
Proteoglycans in cancer	0.0377	CCND1, MMP9
Viral carcinogenesis	0.0377	CCND1, IL6ST
MAPK1/MAPK3 signaling	0.0380	DUSP4, IL6ST
Transmission across chemical synapses	0.0455	GRIA2, ABAT
Biological oxidations	0.0466	FMO2, ADH1B

DEGs: Differentially expressed genes.

Gene and pathway relationships were established, and Ridge regression machine learning analysis was performed [Figure 4]. CCND1, GRIA2, IL6ST, MMP9, and PRKAR2B were determined as molecular signatures of obese breast cancer patients according to Ridge regression results.

DISCUSSION

The obesity epidemic is recognized as one of the most serious health issues affecting public health worldwide today. Numerous observational studies have shown that it is associated with obesity and poor

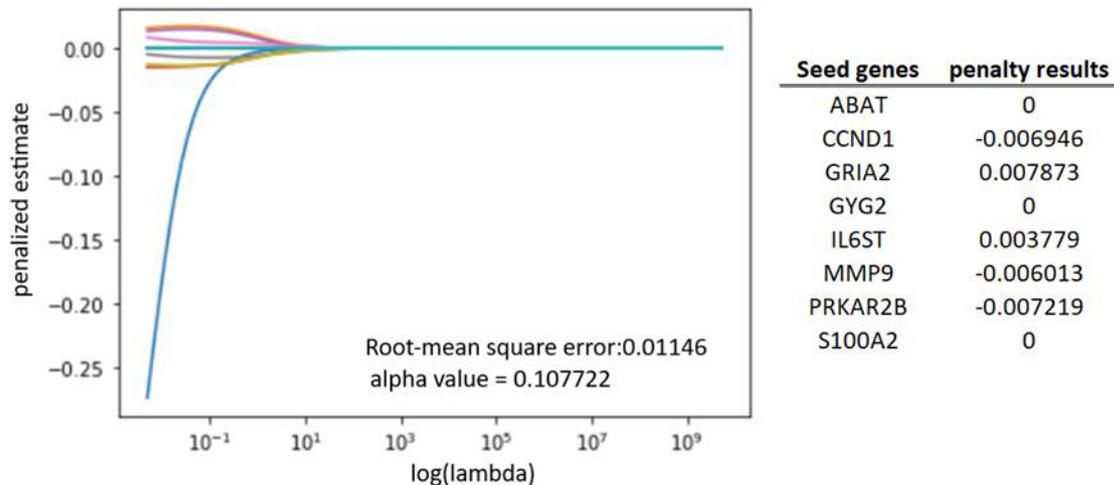


Figure 4. Ridge machine learning analysis results. Each curve represents a penalty estimate score in the model.

survival in patients with breast cancer. On the other hand, key biomarkers of obesity-associated breast cancer risk are still lacking. The present study employed a gene co-expression network analysis to decipher the crucial genes and pathways of obese patients with breast cancer. We identified 23 DEGs from the GSE24185 transcriptome dataset. The seed genes were identified from common DEGs in the co-expression network genes and hub genes of the PPI network. The pathway enrichment analysis was conducted for the seed genes and DEGs. The validity of the identified seed genes was checked by ridge regression.

The putative molecular markers of obese women with breast cancer were identified as CCND1, GRIA2, IL6ST, MMP9, and PRKAR2B. In the literature, recent studies supported the analysis results. It was reported that CCND1 deficiency has a crucial impact on obesity/diabetes-associated liver tumors^[29]. In another study, it was concluded that obesity may enhance asthma and associated mechanisms via CCND1 gene activity^[30]. IL6ST appears to be a positive prognostic factor that is linked to estrogen receptor status in breast cancer^[31]. In addition, interleukin-6 actions in the hypothalamus protect against obesity and play a role in the regulation of neurogenesis^[32]. It was found that upregulated gene expression of MMP9 is linked to visceral obesity in esophageal adenocarcinoma tumor biopsies^[33]. In addition, MMP9 could be regulated by DNA methylation in breast cancer^[34]. According to single-nucleotide polymorphisms analysis results, PRKAR2B may play a role in antipsychotic-induced weight increase in schizophrenia patients^[35]. There is limited literature on GRIA2, and no obesity-related research was found. It is important to conduct more experimental studies to evaluate all these results together.

Mutual risk factors were examined for similar subtypes of obesity and basal subtypes in breast cancer and bladder cancer^[36]. Obesity has been linked to the development of advanced prostate cancer. In the presence of obesity, tumor-associated neutrophils and B cells may promote prostate cancer^[37].

Cancer-associated pathways including RAF-independent MAPK1/3 activation, collagen degradation, bladder cancer, drug metabolism-cytochrome P450, and signaling by Hedgehog were determined as significant pathways. Cytochrome P450 is a hemoprotein that plays a role in drug metabolism. Drug-metabolizing enzyme activity body composition, dietary consumption, and nutritional status all affect cytochrome P450. This link could lead to drug toxicity or reduced therapeutic efficacy, as well as a change in the cost-effectiveness of medical care^[38]. The Hedgehog signaling pathway is critical for breast cancer growth and metastasis^[39], and inhibiting Hedgehog signaling reprograms the breast cancer immune

microenvironment^[40]. Moreover, the Indian Hedgehog signaling system has been linked to the development of hepatocellular cancer in obese mice^[41], and downregulation of Sonic Hedgehog signaling in the hippocampus leads to neuronal death in mice fed a high-fat diet^[42].

In conclusion, this unique approach provides a generic paradigm for mapping complex genetic networks underlying human disease from gene expression data, and the understanding of the reciprocal interplay between obesity and cancer elucidated can begin to affect clinical practice. Therefore, response to conventional and targeted therapies is an essential issue to investigate in experimental and computational studies. As with the development of personalized oncology approaches, there is a need to evaluate new diagnostic and therapeutic strategies to understand the obesity and cancer interplay. In the present study, it was represented that *CCND1*, *GRIA2*, *IL6ST*, *MMP9*, and *PRKAR2B*, as well as pathways associated with these genes, may be molecular signatures in obese patients with breast cancer. These genes may be used for risk analysis of the disease progression of obese patients with breast cancer. Further experimental studies should be conducted and large sample studies should be carried out.

DECLARATIONS

Authors' contributions

Conceptualization, data curation, formal analysis, investigation, methodology, visualization, and writing - original draft: Comertpay B

Supervision, validation, writing review and editing: Gov E

All authors have read and agreed to the published version of the manuscript.

Availability of data and materials

Not applicable.

Financial support and sponsorship

None.

Conflicts of interest

Both authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2022.

REFERENCES

1. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. *Nature* 2020;577:89-94. DOI PubMed
2. World Cancer Report: Cancer Research for Cancer Prevention (PDF). Available from: <https://shop.iarc.fr/products/world-cancer-report-cancer-research-for-cancer-prevention-pdf> [Last accessed on 7 Jan 2022].
3. McPherson K, Steel CM, Dixon JM. ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. *BMJ* 2000;321:624-8. DOI PubMed PMC
4. Andò S, Gelsomino L, Panza S, et al. Obesity, leptin and breast cancer: epidemiological evidence and proposed mechanisms. *Cancers (Basel)* 2019;11:62. DOI PubMed PMC
5. Afshin A, Forouzanfar MH, Reitsma MB, et al; GBD 2015 Obesity Collaborators. Health effects of overweight and obesity in 195 countries over 25 years. *N Engl J Med* 2017;377:13-27. DOI PubMed PMC

6. James FR, Wootton S, Jackson A, Wiseman M, Copson ER, Cutress RI. Obesity in breast cancer--what is the risk factor? *Eur J Cancer* 2015;51:705-20. DOI PubMed
7. Argolo DF, Hudis CA, Iyengar NM. The impact of obesity on breast cancer. *Curr Oncol Rep* 2018;20:47. DOI PubMed
8. Gui Y, Pan Q, Chen X, Xu S, Luo X, Chen L. The association between obesity related adipokines and risk of breast cancer: a meta-analysis. *Oncotarget* 2017;8:75389-99. DOI PubMed PMC
9. Wu X, Zhang X, Hao Y, Li J. Obesity-related protein biomarkers for predicting breast cancer risk: an overview of systematic reviews. *Breast Cancer* 2021;28:25-39. DOI PubMed
10. Diao S, Wu X, Zhang X, et al. Obesity-related proteins score as a potential marker of breast cancer risk. *Sci Rep* 2021;11:8230. DOI PubMed PMC
11. Modi ND, Tan JQE, Rowland A, et al. The obesity paradox in early and advanced HER2 positive breast cancer: pooled analysis of clinical trial data. *NPJ Breast Cancer* 2021;7:30. DOI PubMed PMC
12. Chang MC, Eslami Z, Ennis M, Goodwin PJ. Crown-like structures in breast adipose tissue of breast cancer patients: associations with CD68 expression, obesity, metabolic factors and prognosis. *NPJ Breast Cancer* 2021;7:97. DOI PubMed PMC
13. Mason C, Wang L, Duggan C, et al. Gene expression in breast and adipose tissue after 12 months of weight loss and vitamin D supplementation in postmenopausal women. *NPJ Breast Cancer* 2017;3:15. DOI PubMed PMC
14. Lee K, Kruper L, Dieli-Conwright CM, Mortimer JE. The impact of obesity on breast cancer diagnosis and treatment. *Curr Oncol Rep* 2019;21:41. DOI PubMed PMC
15. Creighton CJ, Sada YH, Zhang Y, et al. A gene transcription signature of obesity in breast cancer. *Breast Cancer Res Treat* 2012;132:993-1000. DOI PubMed
16. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* 2013;41:D991-5. DOI PubMed PMC
17. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003;4:249-64. DOI PubMed
18. Smyth GK. Limma: linear models for microarray data. In: Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S, editors. *Bioinformatics and computational biology solutions using R and bioconductor*. New York: Springer-Verlag; 2005. p. 397-420. DOI
19. Chatr-Aryamontri A, Oughtred R, Boucher L, et al. The BioGRID interaction database: 2017 update. *Nucleic Acids Res* 2017;45:D369-79. DOI PubMed PMC
20. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498-504. DOI PubMed PMC
21. Kamburov A, Wierling C, Lehrach H, Herwig R. ConsensusPathDB--a database for integrating human functional interaction networks. *Nucleic Acids Res* 2009;37:D623-8. DOI PubMed PMC
22. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* 2014;42:D199-205. DOI PubMed PMC
23. Croft D, O'Kelly G, Wu G, et al. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res* 2011;39:D691-7. DOI PubMed PMC
24. Harris CR, Millman KJ, van der Walt SJ, et al. Array programming with NumPy. *Nature* 2020;585:357-62. DOI PubMed PMC
25. Betancourt R, Chen S. Pandas library. Python for SAS users. Berkeley: Apress; 2019. p. 65-109. DOI
26. Perez F, Granger BE. IPython: a system for interactive scientific computing. *Comput Sci Eng* 2007;9:21-9. DOI
27. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010;33:1. PubMed PMC
28. Safran M, Dalah I, Alexander J, et al. GeneCards version 3: the human gene integrator. *Database (Oxford)* 2010;2010:baq020. DOI PubMed PMC
29. Luo C, Liang J, Sharabi K, et al. Obesity/Type 2 diabetes-associated liver tumors are sensitive to cyclin D1 deficiency. *Cancer Res* 2020;80:3215-21. DOI PubMed PMC
30. Thun GA, Imboden M, Berger W, Rochat T, Probst-Hensch NM. The association of a variant in the cell cycle control gene CCND1 and obesity on the development of asthma in the Swiss SAPALDIA study. *J Asthma* 2013;50:147-54. DOI PubMed
31. Martínez-Pérez C, Leung J, Kay C, et al. The signal transducer IL6ST (gp130) as a predictive and prognostic biomarker in breast cancer. *J Pers Med* 2021;11:618. DOI PubMed PMC
32. Bobbo VC, Engel DF, Jara CP, et al. Interleukin-6 actions in the hypothalamus protects against obesity and is involved in the regulation of neurogenesis. *J Neuroinflammation* 2021;18:192. DOI PubMed PMC
33. Allott EH, Lysaght J, Cathcart MC, et al. MMP9 expression in oesophageal adenocarcinoma is upregulated with visceral obesity and is associated with poor tumour differentiation. *Mol Carcinog* 2013;52:144-54. DOI PubMed
34. Klassen LMB, Chequin A, Manica GCM, et al. MMP9 gene expression regulation by intragenic epigenetic modifications in breast cancer. *Gene* 2018;642:461-6. DOI PubMed
35. Gagliano SA, Tiwari AK, Freeman N, et al. Protein kinase cAMP-dependent regulatory type II beta (PRKAR2B) gene variants in antipsychotic-induced weight gain. *Hum Psychopharmacol* 2014;29:330-5. DOI PubMed
36. Sun X, Hoadley KA, Kim WY, Furberg H, Olshan AF, Troester MA. Age at diagnosis, obesity, smoking, and molecular subtypes in muscle-invasive bladder cancer. *Cancer Causes Control* 2017;28:539-44. DOI PubMed PMC
37. Fujita K, Hayashi T, Matsushita M, Uemura M, Nonomura N. Obesity, inflammation, and prostate cancer. *J Clin Med* 2019;8:201. DOI PubMed PMC
38. Zarezadeh M, Saedisomeolia A, Shekarabi M, Khorshidi M, Emami MR, Müller DJ. The effect of obesity, macronutrients, fasting and

- nutritional status on drug-metabolizing cytochrome P450s: a systematic review of current evidence on human studies. *Eur J Nutr* 2021;60:2905-21. DOI PubMed
39. Riobo-Del Galdo NA, Lara Montero Á, Wertheimer EV. Role of Hedgehog signaling in breast cancer: pathogenesis and therapeutics. *Cells* 2019;8:375. DOI PubMed PMC
 40. Hanna A, Metge BJ, Bailey SK, et al. Inhibition of Hedgehog signaling reprograms the dysfunctional immune microenvironment in breast cancer. *Oncoimmunology* 2019;8:1548241. DOI PubMed PMC
 41. Chong YC, Lim TE, Fu Y, Shin EM, Tergaonkar V, Han W. Indian Hedgehog links obesity to development of hepatocellular carcinoma. *Oncogene* 2019;38:2206-22. DOI PubMed
 42. Qin S, Sun D, Zhang C, et al. Downregulation of sonic hedgehog signaling in the hippocampus leads to neuronal apoptosis in high-fat diet-fed mice. *Behav Brain Res* 2019;367:91-100. DOI PubMed