Breast and Colorectal Cancers in Women: a Meta-Analysis Driven by BioOptimatics

Xavier J. Aguilar-García, MS*; Alibeth E. Luna Alvear, MS†; Isis Narváez-Bandera, PhD†¶; Deiver Suárez-Gómez, PhD†; Clara E. Isaza, PhD†‡; Mauricio Cabrera-Ríos, PhD*†

Objective: This meta-analysis explored genes in common between breast cancer (BC) and colorectal cancer (CRC) in women. Breast cancer and CRC are causes of significant morbidity and mortality in women worldwide. Research has shown that women are underrepresented in clinical trials, especially in oncology; studying sex differences in cancer addresses this lack.

Methods: Ten GEO (Gene Expression Omnibus) dataset (5 BC and 5 CRC) were used to identify genes in common. Correlated networks were constructed and analyzed using BioOptimatics methodologies, including multiple criteria optimization and minimum spanning tree.

Results: Eighteen differentially expressed genes were identified, with such core genes as B3GNT3, CALU, CD46, DCN, DLX4, and others showing high frequencies. The study also identified 289 diseases related to core genes, further narrowed down to 37, including BC and CRC. Direct associations with BC and CRC were found for 5 genes, while 7 were linked to other cancer types.

Conclusion: The results of this study underscore the importance of sex differences in cancer biology and indicate that this methodology, BioOptimatics, can help in the discovery of new pathways and biomarkers for BC and CRC in women.

[P R Health Sci J 2024;43(4):186-195]

Key words: Breast cancer, Colorectal cancer, Meta-analysis, Multiple criteria optimization, Minimum spanning tree

Globally, breast cancer (BC) is the most diagnosed cancer in women, with an estimated 2.3 million new cases (1,2) each year. In men, prostate cancer is the most diagnosed one, with an estimated 1.4 million new cases, annually. For both sexes, colorectal cancer (CRC) is the third most common cancer, having an estimated 1.9 million new cases in 2020. In the US, the mortality rate due to BC has been declining, but it is still the leading cause of cancer death in women, accounting for 6.9% of all cancer deaths (in women) in 2020. Accounting for 9.2% of all cancer deaths in that same year, CRC is also a significant cause of cancer death. The burdens of both BC and CRC are expected to rise, with a projected 41% and 68.5% increase in cases of BC and CRC, respectively, by 2040 (1–3). These statistics emphasize the importance of continuing efforts in cancer prevention, early detection, and effective treatment to address the growing burdens of BC and CRC.

There are notable differences in cancer incidence, survival, and mortality rates between men and women. In BC, women have a higher incidence rate than men do, but men have higher mortality rates (1,2,4). In contrast, CRC is more common in men than women (1). These differences are due to the role played by sex in cancer biology. Research on sex differences in cancer mechanisms shows that sex chromosomes and hormones play significant roles in gene expression, immune system response, and cancer progression (5,6). Moreover, female patients have more robust innate and adaptive immune responses than do male patients, reducing cancer mortality risk in female patients because of differences in hormone levels and epigenetic, genetic, and psychosocial factors. Women often show different symptoms, respond differently to treatments, and experience other side effects than men do (5). However, sex-specific differences are often overlooked in clinical trials, with women being underrepresented in the sample populations (with the exception, of course, of those trials in which women-specific cancers are the focus). Several reports have highlighted this underrepresentation, calling for greater attention being paid to women's health in clinical trials (excluding BC) (7,8). In addition, identifying the genes and mechanisms underlying these differences is essential to developing more targeted and effective cancer treatments for this population—female cancer patients.

Microarray experiments have significantly advanced the field of oncology by enabling the simultaneous measurement of thousands of gene expression levels across various samples. The technology used in these experiments, namely microarrays, provides a comprehensive view of gene expression patterns in different types of cancer. However, the analysis of microarray experiments typically requires normalization and shows

*Industrial Engineering Department, The Applied Optimization Group, University of Puerto Rico at Mayagüez, Mayagüez, Puerto Rico; †Graduate Program in Bioengineering, The Applied Optimization Group, University of Puerto Rico at Mayagüez, Mayagüez, Puerto Rico; ‡Biology Department, University of Puerto Rico at Mayagüez, Mayagüez, Puerto Rico; ¶Department of Biostatistics and Bioinformatics, Moffitt Cancer Center

The authors have no conflict of interest to disclose.

Address correspondence to: Prof. Mauricio Cabrera-Ríos, Department of Industrial Engineering and Graduate Program in Bioengineering, University of Puerto Rico at Mayagüez, Mayagüez, PR 00681-9000. Email: mauricio.cabrera1@upr.edu

redundancy, and their results depend on managing parameters, which can affect meta-analysis (MA) outcomes. Despite these challenges, MAs of gene expression data can identify common signaling pathways in cancer, leading to valuable insights into cancer biology (9–11).

A significant MA of gene expression data from microarray and next-generation sequencing experiments across multiple cancers, including BC and CRC, revealed correlations in gene expression patterns. Notably, this analysis pinpointed 2 genes, ABCA8 and S100P, as being overexpressed in both BC and CRC during the tumoral process. Moreover, gene expression–based signature approaches have identified 3 differentially expressed genes (DEGs), CEACAM5, GATA3, and TSPAN8, that aid in the identification of tumor tissue origins (10,12–14). Subsequent studies have delved into broader multi-cancer biomarker signatures, spotlighting key genes and signaling pathways linked to BC and CRC. This body of research has led to a deeper understanding of cancer biology and the potential for developing targeted therapies based on the insights derived from it (15–26).

This work aimed to investigate commonalities between BC and CRC by evaluating gene expression in samples from women that were taken from 10 GEO (Gene Expression Omnibus) dataset, 5 of BC (27–31) and 5 of CRC (31–34). The objective was to identify shared DEGs and patterns. The study used BioOptimatics tools, which are based on mathematical optimization. The BioOptimatics tools can be accessed at the address https://github.com/DeiverSuarez/OBAMA. The results revealed novel correlations in common signaling pathways, presenting new research possibilities for BC and CRC in the female population.

Materials and Methods

The present study followed the recommendations specified in a 2022 article from Diaz-Uriarte et al. (11). These recommendations were related to choosing a suitable study design that would effectively integrate different data types, assessing the value of clinical versus omics data, and choosing suitable preprocessing and filtering approaches. In addition, the adoption of BioOptimatics methods helps preserve objectivity and repeatability in the results. Figure 1 shows the 4 stages of the study.

Stage 1: GEO Dataset Selection & Preprocessing

Experimental data for BC and CRC were obtained from public datasets (27–34). The first step in selecting the data was establishing the inclusion and exclusion criteria. One of the inclusion criteria was that the microarray dataset must contain samples without cancer (controls) and samples with cancer (disease). Another criterion was that each microarray dataset had to have samples from women in both groups (control and disease). Finally, at least 1 microarray dataset per disease needed to be included.

The exclusion criteria were as follows: Samples from men were not included in the study; samples consisting of tissue containing any disease or diseases other than the diseases of interest were not included in the study. Samples from men were not included in this study because it was aligned with the objectives and efforts required to address female health and avoid sex-biased outcome results. In addition, samples consisting of tissue containing any disease or diseases other than the diseases of interest and those that had been treated for conditions other than BC and CRC were not included because this study intended to identify potential biomarkers by analyzing the relative gene expression through the absolute differences between control and disease states.

Posterior data extraction and preprocessing were conducted using R code developed by our research group and described in Sánchez-Peña et al., Camacho-Cáceres et al., Lorenzo et al., Isaza et al, and Narváez-Bandera et al. (35–39). This step calculated the median of the relative expression to consolidate repeated genes and remove the samples from men, those consisting of tissue containing any disease or diseases other than the diseases of interest, and those that had been treated for conditions other than BC and CRC and that had been included in some of the 10 microarray datasets. The data extraction and synthesis tasks enabled the subsequent BioOptimatics analyses (39).

Stage 2: Preliminary Multiple Criteria Optimization Analysis

This study used multiple criteria optimization (MCO) to identify DEGs. The performance measures selected were the absolute values of the differences between the medians and means. These values were calculated using both control samples and disease samples from each BC and CRC microarray dataset. Ten consecutive Pareto-efficient frontiers were consistently identified for each MCO run. Two approaches were used in this step: MCO individual (MCO Ind) and MCO MA. The combination of MCO Ind and MCO MA results for each platform were tallied to compute the frequency per gene.

To organize the MAs, an R-driven clustering procedure was used in conjunction with the generation of heatmaps. Clustering is a technique that requires scaling to analyze multiple studies under the same dimension and group a set of objects based on similarity between them (40). The data used in this case included the results of the preliminary 15 MCOs with a coverage of 2 or more in all the GEO datasets. The scale, then, was associated with the number of Pareto-efficient frontiers (from 1 to 10) in which the gene was identified.

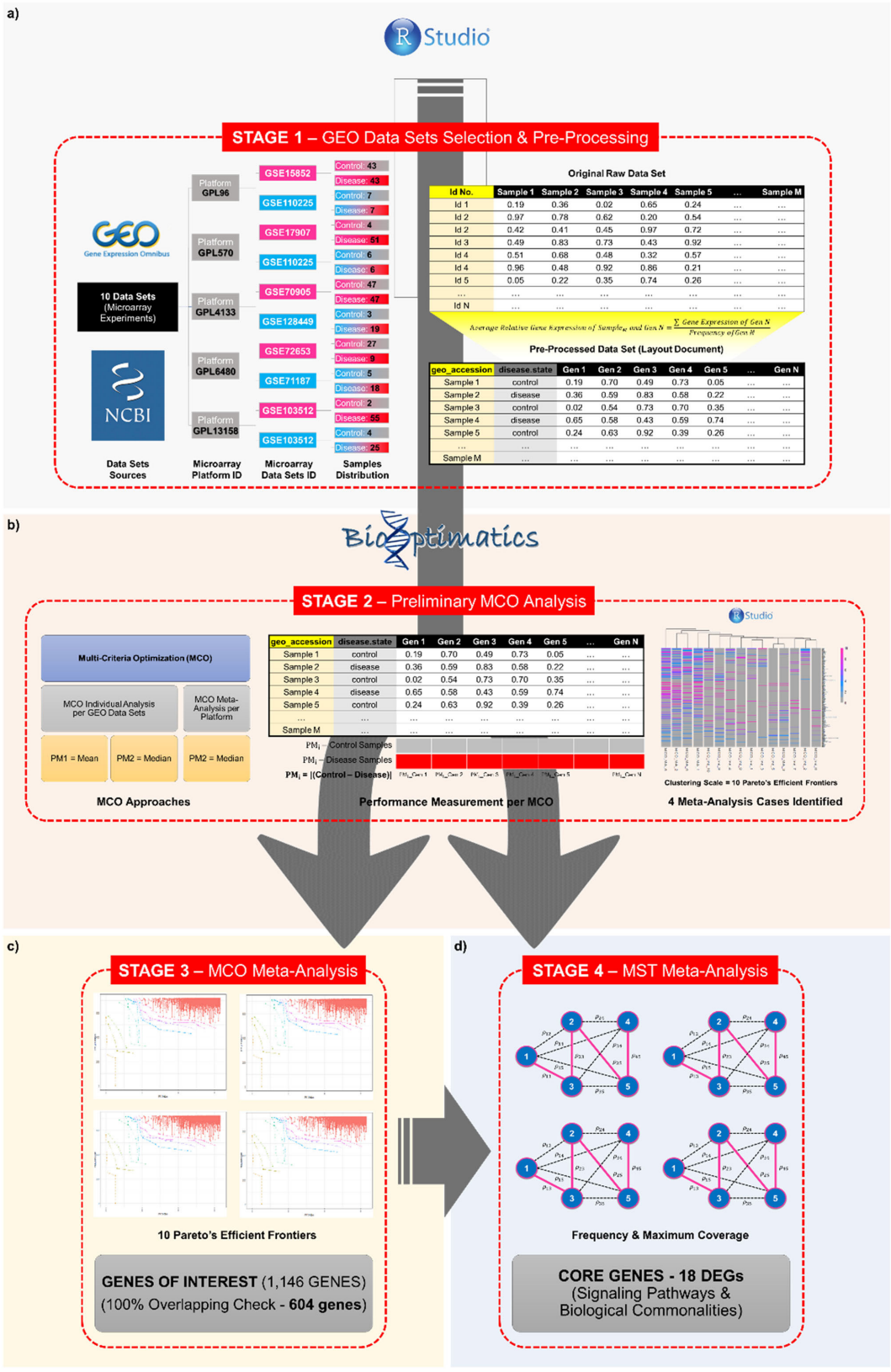
Stage 3: Multiple Criteria Optimization Meta-analysis

In this stage, the heatmap generated previously was used to identify all the study cases to be included in the subsequent steps. Only studies integrating at least 1 GEO dataset of BC and CRC were analyzed. To identify the genes of interest, an MCO MA was performed for each study case. Then, a frequency analysis and a maximum coverage analysis were performed on the results obtained. Due to the aggregated evidence, these genes were included as potential genes of interest in this study.

Stage 4: Minimum Spanning Tree Meta-analysis

In stage 3, an overlap check between the previously identified potential genes of interest was required to perform the MST analysis described in the steps that follow. The resulting genes were considered to be of further interest. Then, an MST MA was

Figure 1. Material and Methodology



a) STAGE 1 – GEO Data Sets Selection & Pre-Processing. In this study, a total of 10 Microarray Data Sets were selected: 5 of BC and 5 of CRC. These GEO Data Sets are configured under five (5) platforms. Then, with a code in RStudio, the pre-processing step was performed per each microarray data set selected. This step generates a layout document that facilitates the analysis of them in next stages using the BioOptimatics tool.

b) STAGE 2 – Preliminary MCO Analysis. In this stage, three (3) steps were performed: MCO Individual (MCO Ind.), MCO Meta-Analysis per Platform (MCO MA per Platform), and Clustering – Heatmap. For MCO Ind., each of all GEO Data Set was analyzed and for MCO MA per Platform step, with the meta-analysis approach, two GEO Data Set per platform were analyzed simultaneously. Both analyses were performed through MCO methodology to obtain the 10 Pareto's Efficient Frontiers. A total of 15 MCO were analyzed. Finally, in the Clustering – Heatmap step, the clustering technique was implemented to identify the Meta-Analysis (MA) cases to analyze in this study. The heatmap used as scale the 10 Pareto's Efficient Frontiers. As result, a total of 4 MA cases of studies were identified.

c) STAGE 3 – MCO Meta-Analysis. The MCO MA for each case previously identified were performed. It was to identify the genes of interest. Then, a quality assessment was performed to identify that of these genes of interest are overlapping between the GEO Data Sets.

d) STAGE 4 – MST Meta-Analysis. In this stage, the MST MA methodology was performed to identify the core genes of this study on 4 MA cases of studies. It was to establish a network correlated between genes of interest. To identify the core genes of this study, the frequency per case and the summatory of correlations between pairwise of genes were considered. Then, MST MA were performed considering the genes linked on signaling pathways associated on BC and CRC and overlapped on 10 GEO data sets. Finally, the potential correlations between core genes and genes of signaling pathways of BC and CRC were reviewed considering the insight on BC and CRC.

Gray Arrows (➡) Indicate the Flow Process of Methodology

performed for each study case. A complete pairwise correlation was computed. A correlation analysis of frequency and an analysis of maximum coverage were performed to identify the core genes of this study.

Furthermore, an overlap check was performed on genes linked to signaling pathways associated with BC and CRC (according to the Kyoto Encyclopedia of Genes and Genomes) among the 10 GEO datasets selected. An MST MA was performed for each study case, taking into account the core genes and the genes of the signaling pathways selected. Finally, a literature reviewed was performed with 2 objectives: to validate the results using GeneMANIA and to establish associations between the core genes and gain new insights into BC and CRC in female patients.

Results

The results from each stage of analysis are discussed next.

Stage 1: GEO Dataset Selection & Preprocessing

In this stage, 10 microarray datasets were selected and downloaded from the GEO public repository, 5 of BC (28–32) and 5 of CRC (32–35). Table 1 summarizes the initial quantity, the final number of samples per microarray dataset, and the distribution per type of cancer after the preprocessing step.

Each BC microarray dataset had only women's samples; the men's samples were removed from the CRC microarray datasets. Regarding the disease samples, only those samples with cancer, carcinoma, and/or adenocarcinoma was accepted; those with polyps and adenomas were removed, resulting in a total of 749 samples. Finally, 432 samples were included: 150 control samples and 282 disease samples, the latter being distributed between BC and CRC, as seen in Table 1.

Stage 2: Preliminary Multiple Criteria Optimization Analysis

A total of 686 genes were identified across 10 Pareto frontiers from 15 individual MCO analyses. Only 246 of them were used to generate the heatmap owing to the frequency of their appearances in the result. Each gene selected was a solution to the MCO problem for 2 or more of the 15 MCOs analyzed following the principles of MA. The heatmap that was generated is presented in Figure 2.

Six MA cases were identified using the heatmap: 4 consisted of BC with CRC, 1 was BC, and another was CRC. These cases were identified through clustering. Four of them were finally included in this study and are named MCO & MST MA #1 through MCO & MST MA #4. Figure 2 summarizes the microarray datasets used in the MCO and MST MA study cases.

Stage 3: Multiple Criteria Optimization Meta-analysis

A total of 4,729 genes from all the MCO MA study cases were analyzed. However, only 1,146 genes following the MA principle were included. In other words, the resulting genes were relatively differentially expressed in at least 2 MCO MA study cases.

This step required a quality check to move to the next stage. That check verified 100% overlap of the 1,146 genes of interest across 10 selected GEO datasets. As a result, a total of 604 of them had 100% overlap and were then analyzed using MST in the next stage.

Stage 4: Minimum Spanning Tree Meta-analysis

In this stage, the MST MA analysis was performed using the BioOptimatics capabilities of our group's code. The 4 study cases previously mentioned (refer to Figure 2) were included. For this analysis, we calculated all pairwise linear correlations among 604 genes of interest and identified the most correlated structure for each MST MA case. To further streamline the results following MA principles, 9 pairs of genes (18 DEGs) were considered. These were correlated in at least 2 MST MA cases. Refer to Figure 3 for the most correlated path for each MST MA case. From the optimal MST solution, 39 correlated paths were identified. From them, the CD46-SORD, PIP5K1B, B3GNT3, and VCAN-CALU pairs had the highest frequency of appearance across 4 MST MA cases.

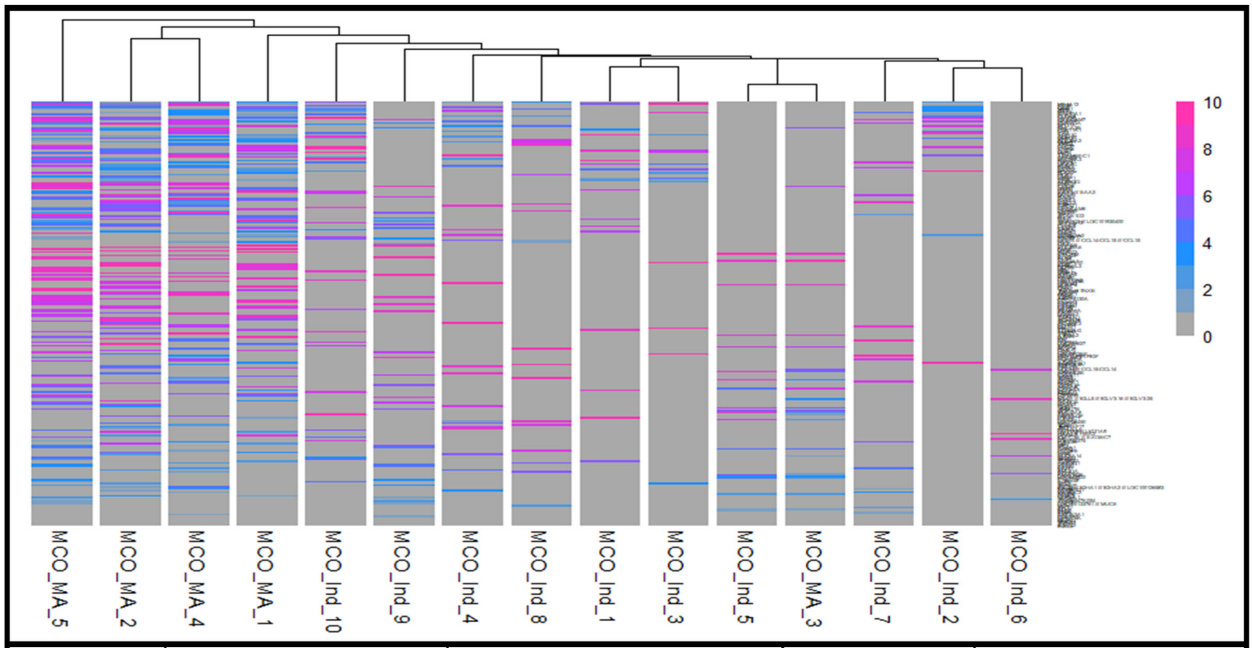
According to our results, these gene relationships, composed of 18 DEGs, have important roles in BC and CRC in women.

Table 1. Quantity of Genes and Samples per GEO Dataset

Type of Cancer	GPL ID	GSE ID	Qty Genes	Initial Qty Samples	Qty Samples Removed	Final Qty Samples	Qty Samples (Control)	Qty Samples (Disease)
BC	GPL96	GSE15852	13515	86	0	86	43	43
CRC	GPL96	GSE110225	13515	26	12	14	7	7
BC	GPL570	GSE17907	13362	55	0	55	4	51
CRC	GPL570	GSE110225	23520	34	18	16	8	8
BC	GPL4133	GSE70905	4217	137	43	94	47	47
CRC	GPL4133	GSE128449	19712	58	36	22	3	19
BC	GPL6480	GSE72653	19595	36	0	36	27	9
CRC	GPL6480	GSE71187	19595	189	166	23	5	18
BC	GPL13158	GSE103512	20741	280	223	57	2	55
CRC	GPL13158				251	29	4	25
		Total Samples		1,181	749	432	150	282

Abbreviations: BC, breast cancer; CRC, colorectal cancer; GEO, gene expression omnibus; GPL, GEO platform; GSE, GEO series

Figure 2. Heat Map – MCO per Dataset & MCO per Platform Distribution: Case-Study Design



Type of Cancer	Platform	GEO Accession	MCO/MST Id.	MCO Id. Description
BC	GPL96	GSE15852	MCO Ind #1	MCO per Data Set
CRC		GSE110225	MCO Ind #2	
BC	GPL570	GSE17907	MCO Ind #3	
CRC		GSE110225	MCO Ind #4	
BC	GPL4133	GSE70905	MCO Ind #5	
CRC		GSE128449	MCO Ind #6	
BC	GPL6480	GSE72653	MCO Ind #7	
CRC		GSE71187	MCO Ind #8	
BC	GPL13158	GSE103512 (BC)	MCO Ind #9	
CRC		GSE103512 (CRC)	MCO Ind #10	
BCRC	GPL 96	GSE15852 + GSE110225	MCO MA #1	MCO per Platform
	GPL570	GSE17907 + GSE110225	MCO MA #2	
	GPL4133	GSE70905 + GSE128449	MCO MA #3	
	GPL6480	GSE72653 + GSE71187	MCO MA #4	
	GPL13158	GSE103512 (BC) + GSE103512 (CRC)	MCO MA #5	
CRC	GPL4133 + GPL96	GSE128449 + GSE110225	N/A[1]	MCO Inds.
BCRC	GPL4133 + GPL96 + GLP6480	GSE128449 + GSE110225 + GSE72653	MCO & MST MA #1	MCO Inds.
BCRC	GPL4133	GSE70905 + GSE128449	MCO & MST MA #2[2]	MCO per Platform + MCO Ind.
BC	GPL570 + GPL96	GSE17907 + GSE15852	N/A[1]	MCO Inds.
BCRC	GPL4133 + GPL570 + GPL96	GSE70905 + GSE128449 + GSE17907 + GSE15852	MCO & MST MA #3[3]	MCO per Platform + MCO Inds.
BCRC	GLP6480 + GPL570	GSE72653 + GSE71187 + GSE17907 + GSE110225	MCO & MST MA #4	MCO per Platforms

[1] These MCO & MST cases were not considered due to apply only for BC or CRC cased, instead BCRC as required for this study.
[2] MCO & MST MA #2 = MCO MA #3 + MCO Ind. #5. Due to repeatability of analysis and avoid redundance of analysis, only consider the MCO MA #3 results are used for genes of interest selection.
[3] MCO & MST MA #3 = MCO MA #3 + MCO Ind. #5 + MCO Ind. #3 + MCO Ind. #1. Due to repeatability of analysis and avoid redundance of analysis, only consider the MCO MA #3 + MCO Ind. #3 + MCO Ind. #1 results are used for genes of interest selection.

Discussion

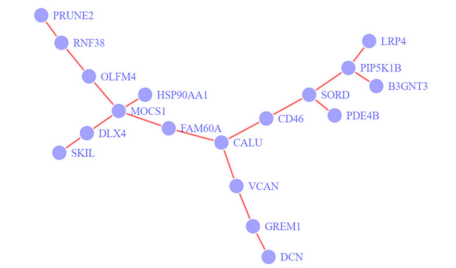
Microarray technology and next-generation sequencing have dramatically transformed oncology by enabling the simultaneous analysis of thousands of gene expression patterns across different cancer types. This advancement has facilitated a deeper understanding of cancer biology, allowing researchers to explore global gene expression trends and identify commonalities across various tumors (9,10). However, the analysis of the resulting data typically requires the user to select normalization procedures and manipulate parameters that affect the results of an MA (10,11).

Table 2. Diseases Related to Core Genes by GeneCards

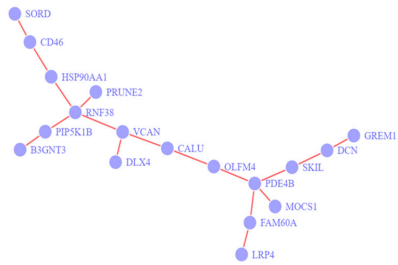
Disorder	B3GNT3	CALU	CD46	DCN	DLX4	FAM60A	GREM1	HSP90AA1	LRP4	MOCS1	OLFM4	PDE4B	PIP5K1B	PRUNE2	RNF38	SKIL	SORD	VCAN	Qty. DEGs
Colorectal Cancer				✓			✓	✓											3
Colon Mucinous Adenocarcinoma															✓				1
Colonic Benign Neoplasm											✓								1
Colorectal Adenoma							✓												1
Familial Colorectal Cancer							✓												1
Lipoma of the Colon							✓												1
Breast Cancer								✓											1
Hereditary Breast Ovarian Cancer Syndrome							✓												1
Ovarian Cancer			✓	✓				✓											3
Pancreatic Cancer								✓			✓							✓	3
Prostate Cancer								✓						✓					2
Gastric Cancer							✓	✓											2
Lung Cancer								✓											1
Thymus Cancer									✓										1
Bladder Cancer								✓											1
Testicular Cancer								✓											1
Corneal Cancer																✓			1
Renal Fibrosis							✓									✓			2
Muscular Dystrophy				✓				✓											2
Methylmalonic Aciduria and Homocystinuria, Cblc Type			✓							✓									2
Cataract																	✓	✓	2
Axonal Neuropathy								✓									✓		2
Charcot–Marie–Tooth Disease					✓												✓		2
Synostosis							✓		✓										2
Chromosome 2q35 Duplication Syndrome							✓		✓										2
Multiple Sclerosis			✓					✓											2
Aortic Aneurysm, Familial Thoracic 1				✓														✓	2
Orofacial Cleft					✓		✓												2
Asthma								✓				✓							2
Peripheral Nervous System Disease								✓									✓		2
Parkinson's Disease, Late-Onset								✓	✓										2
Ehlers–Danlos Syndrome				✓														✓	2
Sclerosteosis							✓		✓										2
Glomerulonephritis			✓	✓															2
Systemic Lupus Erythematosus			✓					✓											2
Interstitial Lung Disease 2				✓			✓												2
Orthostatic Intolerance				✓														✓	2

Abbreviation: DEGs, differentially expressed genes

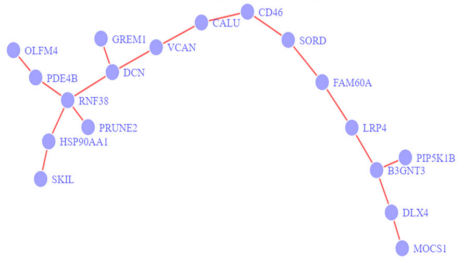
Figure 3. Summary of MST MA study cases: Pathway Structures from Pairwise Correlated Genes



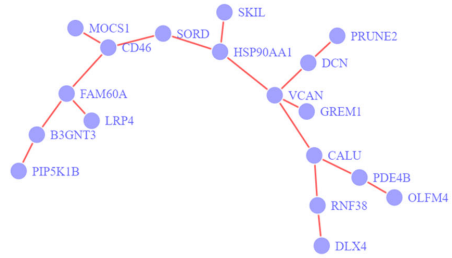
MST MA Case Study Pathway Structure #1



MST MA Case Study Pathway Structure #2



MST MA Case Study Pathway Structure #3

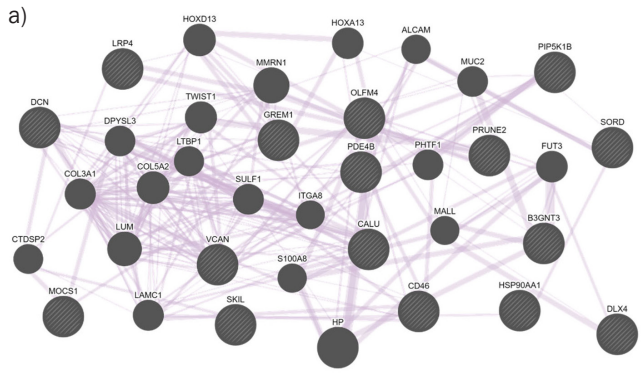


MST MA Case Study Pathway Structure #4

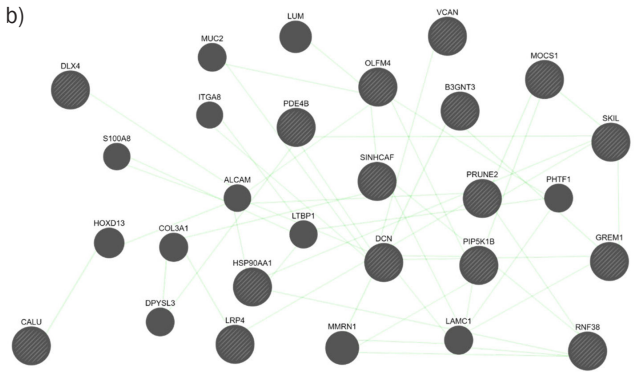
Pairwise Correlated Genes	MST MA #1	MST MA #2	MST MA #3	MST MA #4	Frequency
CD46[1] - SORD[2]	✓	✓	✓	✓	**4**
PIP5K1B[3] - B3GNT3[4]	✓	✓	✓	✓	**4**
VCAN[5] - CALU[6]	✓	✓	✓	✓	**4**
RNF38[7] - PRUNE2[8]	✓	✓	✓		**3**
OLFM4[9] - PDE4B[10]		✓	✓	✓	**3**
DCN[11] - GREM1[12]	✓	✓	✓		**3**
FAM60A[13] - LRP4[14]		✓	✓	✓	**3**
VCAN - GREM1	✓			✓	2
HSP90AA1[15] – SKIL[16]			✓	✓	2
CALU - CD46	✓		✓		2
DCN - VCAN			✓	✓	2
DLX4[17] - MOCS1[18]	✓		✓		2
HSP90AA1-RNF38		✓	✓		2

[1] CD46 is CD46 molecule, a protein coding.
[2] SORD Sorbitol Dehydrogenase
[3] PIP5K1B is Phosphatidylinositol-4-Phosphate 5-Kinase Type 1 Beta, a protein coding.
[4] B3GNT3 is UDP-GlcNAc:BetaGal Beta-1,3-N Acetylglucosaminyltransferase 3, a protein coding.
[5] VCAN is Versican, a protein coding.
[6] CALU is Calumenin, a protein coding.
[7] RNF38 IS RING FINGER PROTEIN 38, a protein coding.
[8] PRUNE2 is Prune Homolog 2 With BCH Domain, a protein coding.
[9] OLFM4 is Olfactomedin 4, a protein coding.
[10] PDE4B is Phosphodiesterase 4b, a protein coding.
[11] DCN is Decorin, a protein coding.
[12] GREM1 is Gremlin 1, DAN Family Bmp Antagonist, a protein coding.
[13] FAM60A (SINHCAF) id SIN3-HDAC Complex Associated Factor, a protein coding.
[14] LRP4 is LDL Receptor Related Protein 4, a protein coding.
[15] HSP90AA1 is Heat Shock Protein 90 Alpha Family Class A Member 1, as protein codng.
[16] SKIL is SKI Like Proto-Oncogene, as protein coding.
[17] DLX4 is Distal-Less Homeobox 4, as protein coding.
[18] MOCS1 is Molybdenum Cofactor Synthesis 1, as protein coding.

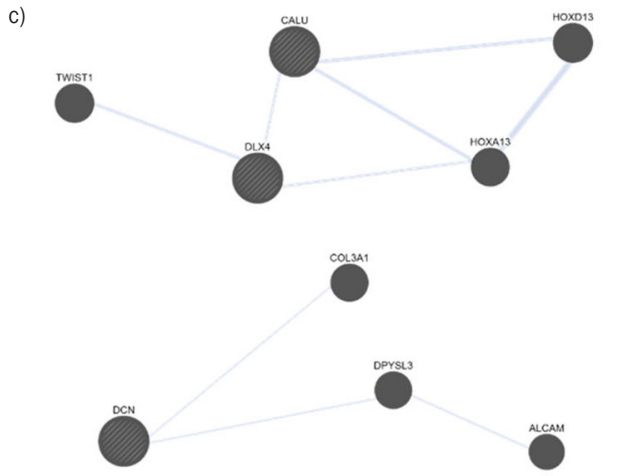
Figure 4. GeneMANIA Networks



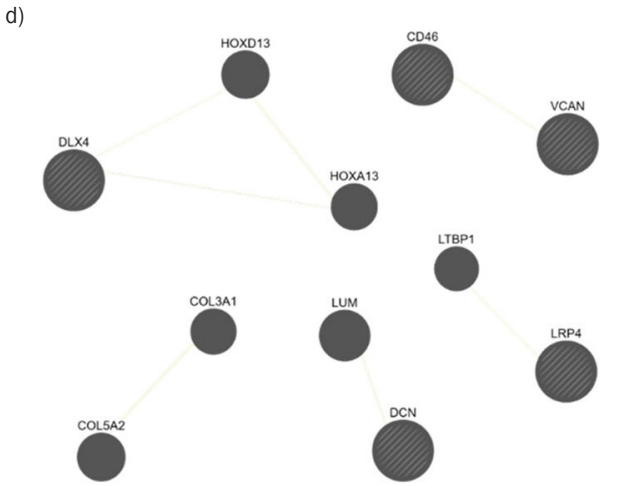
a) Co-expression, GeneMANIA Network. With respect to the co-expression network, 84.88% of 38 genes are linked. Of 18 DEGs, RNF38 and FAM60A (SINHCAF) were not connected to the rest of the genes.



b) Genetic Interactions, GeneMANIA Network. For the genetic interactions network, 7.41% of 38 genes are linked. Of the 18 DEGs, CD46 and SORD were not connected to the rest of genes.



c) Co-localization, GeneMANIA Network. In the co-localization network, 6.42% of 38 genes are linked. Of the 18 DEGs, B3GNT3, SINHCAF, SKIL, PDE4B, GREM1, MOCS1, RNF38, CD46, SORD, LRP4, PRUNE2, HSP90AA1, VCAN, PIP5K1B, and OLFM4 were not connected to the rest of the genes.



d) Shared Protein Domains, GeneMANIA Network. Of the shared protein domains network, 1.29% of 38 genes are linked. Of the 18 DEGs, HSP90AA1, SORD, GREM1, RNF38, PDE4B, B3GNT3, CALU, PIP5K1B, OLFM4, PRUNE2, SKIL, MOCS1, and SINHCAF were not connected to the rest of the genes.

Our study focused on identifying shared genetic and biological pathways between BC and CRC in women. We used GEO datasets and employed GeneCards (41), the Kyoto Encyclopedia of Genes and Genomes (42), and GeneMANIA (43) to conduct gene ontology enrichment analyses, pathway analyses, and network constructions. Our analyses revealed that the DEGs were involved in various biological processes, including the negative regulation of transcription by RNA polymerase II and signal transduction. Cellular components primarily consist of the membrane, nucleus, and cytosol, while molecular functions involve protein binding and metal ion binding.

In comparison, Vogelstein et al. (45) and Poliakov et al. (46) provided broader genomic landscapes and metabolic pathways in various cancers; their respective focuses were not explicitly on sex-specific differences. Vogelstein's work mapped the genomic landscape of common human cancers through comprehensive sequencing and identifying driver mutations and classifying them into 12 signaling pathways regulating key cellular processes (44). Poliakov's study examined overexpressed genes across various cancer types, focusing on metabolic shifts such as glycolysis and oxidative phosphorylation (45). Our methodology, which employed MCO and MST, led to the identification of 18 DEGs associated with BC and CRC in women. These core genes were also tied to 37 other diseases, demonstrating potential associations between specific gene expression patterns and cancer types.

Another study, this one by Lee et al. (46), conducted an MA on BRCA genes and those of related cancers but did not consider genes beyond the BRCA spectrum. Vogelstein et al.'s work on driver mutations and Poliakov et al.'s research on overexpressed genes showed no overlap with our findings. It is important to note

that the divergence between our results and those from said studies is not solely due to our focus on sex-specific differences. The lack of overlap may also be attributed to methodological differences, dataset variations, or distinct study goals. While our study's unique focus on women allowed for insights into sex-based gene expression, other factors such as study design, scope, and analysis techniques could also play a role in these differing outcomes.

Of the 18 DEGs shared between BC and CRC in women and that were identified in our work, 4—B3GNT3, GREM1, OLFM4, and PDE4B—overlapped in breast, sigmoid colon, and transverse colon tissues (12). This overlap highlights the potential for discovering new biomarkers and signaling pathways that could lead to personalized diagnostic tools and targeted therapies for female cancer patients.

A multi-cancer biomarker signature developed in a different study included 41 genes differentially expressed in multiple cancers, although only PDC and RPS6KA1 were specifically tied to BC and CRC (13). Chen et al. focused on gene signatures associated with immune-suppressive tumor microenvironments, suggesting a link between high expression levels of RAB31, IRAK3, and TNPO2 and dysfunctional T-cell phenotypes across 6 different cancers (14).

Except for the study by Lopes-Ramos et al., which revealed significant regulatory differences between the sexes across multiple human tissues, highlighting the fact that sex-biased regulation is systemic and not isolated to a few tissues (15), most studies overlook sex-specific variations. Recent research has shown that signaling pathways can differ between men and women in various cancers (15–17). The Wnt/beta-catenin pathway, MAPK/ERK pathway, and PI3K/Akt/mTOR pathway are examples in which sex-specific variations have been observed, emphasizing the need for a more nuanced approach in studying these pathways and their therapeutic potential (5,22–26).

Given the emphasis on sex as a critical biological variable, our study contributes to cancer research by highlighting the sex-specific differences between men and women in the context of BC and/or CRC. Future research should continue exploring such differences, potentially expanding such studies to include male samples for comparative analysis. This approach would help validate and refine the current findings and identify additional DEGs specific to men. We suggest that BioOptimatics could play a pivotal role in exploring these relationships, providing a framework for further investigation into cancer-related genetic pathways and aiding in the development of targeted therapies.

Resumen

Objetivo: Este metaanálisis exploró los genes comunes entre el cáncer de mama (CM) y el cáncer colorrectal (CCR) en mujeres. El cáncer de mama y el cáncer colorrectal son causas significativas de morbilidad y mortalidad entre mujeres a nivel mundial. La subrepresentación de mujeres en ensayos clínicos, especialmente en oncología, y la necesidad de estudiar las diferencias de género en el cáncer hacen que esta exploración sea necesaria. Métodos: Se utilizaron 10 conjuntos de datos obtenidos del repositorio GEO (Gene Expression Omnibus) (5 CM y 5 CCR) para identificar genes en común. Se construyeron redes correlacionadas y se analizaron con metodologías de BioOptimatics, como la optimización de criterios múltiples y el árbol de expansión mínima. Resultados: Se identificaron 18 genes diferencialmente expresados, con genes principales como B3GNT3, CALU, CD46, DCN, DLX4, entre otros, que mostraron una alta frecuencia dentro de los resultados. El estudio también identificó 289 enfermedades relacionadas con los genes principales, posteriormente reducidas a 37, donde se incluyen CM y CCR. Se encontraron asociaciones directas con CM y CCR para cinco genes, mientras que siete se relacionaron con otros tipos de cáncer. Conclusión: Los resultados de este estudio enfatiza la importancia de las diferencias de género en la biología del cáncer y sugiere que esta metodología, BioOptimatics, puede ayudar a descubrir nuevas vías de señalización y biomarcadores para CM y CCR en mujeres.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-249. doi:10.3322/caac.21660
2. Arnold M, Morgan E, Rumgay H, et al. Current and future burden of breast cancer: Global statistics for 2020 and 2040. Breast. 2022; 66:15-23. doi: 10.1016/j.breast.2022.08.010
3. Xi Y, Xu P. Global colorectal cancer burden in 2020 and projections to 2040. Transl Oncol. 2021;14(10):101174. doi: 10.1016/j.tranon.2021.101174
4. Gao Y, Heller SL, Moy L. Male Breast Cancer in the Age of Genetic Testing: An Opportunity for Early Detection, Tailored Therapy, and Surveillance. Radiographics. 2018;38(5):1289-1311. doi:10.1148/rg.2018180013
5. Rubin JB, Lagas JS, Broestl L, et al. Sex differences in cancer mechanisms. Biol Sex Differ. 2020;11(1):17. Published 2020 Apr 15. doi:10.1186/s13293-020-00291-x
6. Rubin JB. The spectrum of sex differences in cancer. Trends Cancer. 2022;8(4):303-315. doi: 10.1016/j.trecan.2022.01.013
7. Ludmir EB, Fuller CD, Moningi S, et al. Sex-Based Disparities Among Cancer Clinical Trial Participants. J Natl Cancer Inst. 2020; 112(2):211-213. doi:10.1093/jnci/djz154
8. Steinberg JR, Turner BE, Weeks BT, et al. Analysis of Female Enrollment and Participant Sex by Burden of Disease in US Clinical Trials Between 2000 and 2020. JAMA Netw Open. 2021;4(6):e2113749. Published 2021 Jun 1. doi: 10.1001/jamanetworkopen.2021.13749
9. Hambali MA, Oladele TO, Adewole KS. Microarray cancer feature selection: Review, challenges and research directions. Int J Cogn Comput Eng. 2020; 1:78-97. doi: 10.1016/j.ijcce.2020.11.001
10. Feltes BC, Poloni JF, Nunes IJG, Faria SS, Dorn M. Multi-Approach Bioinformatics Analysis of Curated Omics Data Provides a Gene Expression Panorama for Multiple Cancer Types. Front Genet. 2020; 11:586602. Published 2020 Nov 23. doi:10.3389/fgene.2020.586602
11. Diaz-Uriarte R, Gómez de Lope E, Giugno R, et al. Ten quick tips for biomarker discovery and validation analyses using machine learning. PLoS Comput Biol. 2022;18(8): e1010357. Published 2022 Aug 11. doi: 10.1371/journal.pcbi.1010357
12. Feltes BC, Poloni JF, Nunes IJG, Faria SS, Dorn M. Multi-Approach Bioinformatics Analysis of Curated Omics Data Provides a Gene Expression Panorama for Multiple Cancer Types. Front Genet. 2020; 11:586602. Published 2020 Nov 23. doi:10.3389/fgene.2020.586602
13. Martinez-Ledesma E, Verhaak RG, Treviño V. Identification of a multi-cancer gene expression biomarker for cancer clinical outcomes using a network-based algorithm. Sci Rep. 2015; 5:11966. Published 2015 Jul 23. doi:10.1038/srep11966
14. Chen JH, Wu ATH, Lawal B, et al. Identification of Cancer Hub Gene Signatures Associated with Immune-Suppressive Tumor Microenvironment and Ovatodiolide as a Potential Cancer Immunotherapeutic Agent. Cancers (Basel). 2021;13(15):3847. Published 2021 Jul 30. doi:10.3390/cancers13153847

15. Lopes-Ramos CM, Chen CY, Kuijjer ML, et al. Sex Differences in Gene Expression and Regulatory Networks across 29 Human Tissues. Cell Rep. 2020;31(12):107795. doi: 10.1016/j.celrep.2020.107795
16. Khodursky S, Jiang CS, Zheng EB, Vaughan R, Schrider DR, Zhao L. Sex differences in interindividual gene expression variability across human tissues. PNAS Nexus. 2022;1(5): pgac243. Published 2022 Oct 26. doi:10.1093/pnasnexus/pgac243
17. Sanchez-Vega F, Mina M, Armenia J, et al. Oncogenic Signaling Pathways in The Cancer Genome Atlas. Cell. 2018;173(2):321-337.e10. doi: 10.1016/j.cell.2018.03.035
18. Yip HYK, Papa A. Signaling Pathways in Cancer: Therapeutic Targets, Combinatorial Treatments, and New Developments. Cells. 2021;10(3):659. Published 2021 Mar 16. doi:10.3390/cells10030659
19. He H, Shao X, Li Y, et al. Targeting Signaling Pathway Networks in Several Malignant Tumors: Progresses and Challenges. Front Pharmacol. 2021; 12:675675. Published 2021 May 31. doi:10.3389/fphar.2021.675675
20. Sheng KL, Kang L, Pridham KJ, Dunkenberger LE, Sheng Z, Varghese RT. An integrated approach to biomarker discovery reveals gene signatures highly predictive of cancer progression. Sci Rep. 2020;10(1):21246. Published 2020 Dec 4. doi:10.1038/s41598-020-78126-3
21. Sahasrabuddhe AA, Rolland DC, Banerjee PP. Signaling Pathways as Biomarkers. Biomark Cancer. 2016;7(Suppl 2):33-35. Published 2016 Feb 14. doi:10.4137/BIC.S37778
22. Zougros A, Michelli M, Chatziandreou I, et al. mRNA coexpression patterns of Wnt pathway components and their clinicopathological associations in breast and colorectal cancer. Pathol Res Pract. 2021; 227:153649. doi: 10.1016/j.prp.2021.153649
23. Harmston N, Lim JYS, Arqués O, et al. Widespread Repression of Gene Expression in Cancer by a Wnt/β-Catenin/MAPK Pathway. Cancer Res. 2021;81(2):464-475. doi: 10.1158/0008-5472.CAN-20-2129
24. Zhan XJ, Wang R, Kuang XR, Zhou JY, Hu XL. Elevated expression of myosin VI contributes to breast cancer progression via MAPK/ERK signaling pathway. Cell Signal. 2023; 106:110633. doi: 10.1016/j.cellsig.2023.110633
25. Stefani C, Miricescu D, Stanescu-Spinu II, et al. Growth Factors, PI3K/AKT/mTOR and MAPK Signaling Pathways in Colorectal Cancer Pathogenesis: Where Are We Now? Int J Mol Sci. 2021;22(19):10260. Published 2021 Sep 23. doi:10.3390/ijms221910260
26. Miricescu D, Totan A, Stanescu-Spinu II, Badoiu SC, Stefani C, Greabu M. PI3K/AKT/mTOR Signaling Pathway in Breast Cancer: From Molecular Landscape to Clinical Aspects. Int J Mol Sci. 2020;22(1):173. Published 2020 Dec 26. doi:10.3390/ijms22010173
27. Pau Ni IB, Zakaria Z, Muhammad R, et al. Gene expression patterns distinguish breast carcinomas from normal breast tissues: the Malaysian context. Pathol Res Pract. 2010;206(4):223-228. doi: 10.1016/j.prp.2009.11.006
28. Sircoulomb F, Bekhouche I, Finetti P, et al. Genome profiling of ERBB2-amplified breast cancers. BMC Cancer. 2010; 10:539. Published 2010 Oct 8. doi:10.1186/1471-2407-10-539
29. Quigley DA, Tahiri A, Lüders T, et al. Age, estrogen, and immune response in breast adenocarcinoma and adjacent normal tissue. Oncoimmunology. 2017;6(11): e1356142. Published 2017 Aug 10. doi:10.1080/2162402X.2017.1356142
30. Abdalla M, Tran-Thanh D, Moreno J, et al. Mapping genomic and transcriptomic alterations spatially in epithelial cells adjacent to human breast carcinoma. Nat Commun. 2017;8(1):1245. Published 2017 Nov 1. doi:10.1038/s41467-017-01357-y
31. Brouwer-Visser J, Cheng WY, Bauer-Mehren A, et al. Regulatory T-cell Genes Drive Altered Immune Microenvironment in Adult Solid Cancers and Allow for Immune Contextual Patient Subtyping. Cancer Epidemiol Biomarkers Prev. 2018;27(1):103-112. doi: 10.1158/1055-9965.EPI-17-0461
32. An N, Shi X, Zhang Y, et al. Discovery of a Novel Immune Gene Signature with Profound Prognostic Value in Colorectal Cancer: A Model of Cooperativity Disorientation Created in the Process from Development to Cancer. PLoS One. 2015;10(9): e0137171. Published 2015 Sep 1. doi: 10.1371/journal.pone.0137171
33. Esteban-Gil A, Pérez-Sanz F, García-Solano J, et al. Differences in expression profiling between tumoral colorectal carcinomas subsets and normal tissue and polyps. Mar 19, 2019. Accessed December 15, 2019. https://https.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE128446
34. Vlachavas EI, Pilalis E, Papadodima O, et al. Radiogenomic Analysis of F-18-Fluorodeoxyglucose Positron Emission Tomography and Gene Expression Data Elucidates the Epidemiological Complexity of Colorectal Cancer Landscape. Comput Struct Biotechnol J. 2019; 17:177-185. Published 2019 Jan 25. doi: 10.1016/j.csbj.2019.01.007
35. Sánchez-Peña ML, Isaza CE, Pérez-Morales J, Rodríguez-Padilla C, Castro JM, Cabrera-Ríos M. Identification of potential biomarkers from microarray experiments using multiple criteria optimization. Cancer Med. 2013;2(2):253-265. doi:10.1002/cam4.69
36. Camacho-Cáceres KI, Acevedo-Díaz JC, Pérez-Marty LM, et al. Multiple criteria optimization joint analyses of microarray experiments in lung cancer: from existing microarray data to new knowledge. Cancer Med. 2015;4(12):1884-1900. doi:10.1002/cam4.540
37. Lorenzo E, Camacho-Caceres K, Ropelewski AJ, et al. An Optimization-Driven Analysis Pipeline to Uncover Biomarkers and Signaling Paths: Cervix Cancer. Microarrays (Basel). 2015;4(2):287-310. doi:10.3390/microarrays4020287
38. Isaza C, Rosas JF, Lorenzo E, et al. Biological signaling pathways and potential mathematical network representations: biological discovery through optimization. Cancer Med. 2018;7(5):1875-1895. doi:10.1002/cam4.1301
39. Narváez-Bandera I, Suárez-Gómez D, Isaza CE, Cabrera-Ríos M. Multiple Criteria Optimization (MCO): A gene selection deterministic tool in RStudio. PLoS One. 2022;17(1): e0262890. Published 2022 Jan 27. doi: 10.1371/journal.pone.0262890
40. Menyhárt O, Győrffy B. Multi-omics approaches in cancer research with applications in tumor subtyping, prognosis, and diagnosis. Comput Struct Biotechnol J. 2021; 19:949-960. Published 2021 Jan 22. doi: 10.1016/j.csbj.2021.01.009
41. Stelzer G, Rosen N, Plaschkes I, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Curr Protoc Bioinformatics. 2016; 54:1.30.1-1.30.33. Published 2016 Jun 20. doi:10.1002/cpbi.5
42. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000;28(1):27-30. doi:10.1093/nar/28.1.27
43. Warde-Farley D, Donaldson SL, Comes O, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010;38(Web Server issue): W214-W220. doi:10.1093/nar/gkq537
44. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. Science. 2013;339(6127):1546-1558. doi:10.1126/science.1235122
45. Poliakov E, Managadze D, Rogozin IB. Generalized portrait of cancer metabolic pathways inferred from a list of genes overexpressed in cancer. Genet Res Int. 2014; 2014:646193. doi: 10.1155/2014/646193
46. Lee YC, Lee YL, Li CY. BRCA Genes and Related Cancers: A Meta-Analysis from Epidemiological Cohort Studies. Medicina (Kaunas). 2021;57(9):905. Published 2021 Aug 30. doi:10.3390/medicina57090905