2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI) | 978-1-6654-0358-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/BHI50953.2021.9508518

Computational insights on the molecular mechanisms across breast cancer progression combining gene differential expression and co-expression

1st Emmanouil K. Ikonomakis
Computational Intelligence Laboratory
Department of Mathematics
University of Patras
Patras, Greece
eki@math.upatras.gr

2nd Marilena M. Bourdakou
Laboratory of Pharmacology
Faculty of Medicine
Democritus University of Thrace
Alexandroupolis, Greece
kmarlen1988@gmail.com

3rd George Kolios
Laboratory of Pharmacology,
Faculty of Medicine
Democritus University of Thrace
Alexandroupolis, Greece
gkolios@med.duth.gr

4th Michael N. Vrahatis
Computational Intelligence Laboratory
Department of Mathematics
University of Patras
Patras, Greece
vrahatis@math.upatras.gr

5th George M. Spyrou
The Cyprus Institute of Neurology and Genetics
The Cyprus School of Molecular Medicine
Ayios Dometios, Nicosia, Cyprus
georges@cing.ac.cy

***Abstract*—As a complex disease with mechanisms that are not fully understood, breast cancer pathology and progression is expected to be determined not only by individual genes, but also by their coordinated effects though a more systemic framework. In this contribution, we combine gene differential analysis (single gene view) with gene co-expression analysis (systemic view) to provide insights on the implicated molecular mechanisms across breast cancer progression. Important gene-gene links in a gene co-expression network are identified and clustered with Density and Distance Content driven (DeDiCo) algorithm, which has been recently published by members of our groups. This algorithm presents very good performance regarding its accuracy on both synthetic and real-life datasets and ability to determine arbitrarily shaped clusters without any assumptions on the shape and the number of the clusters. This work provides a pipeline, exploiting a novel clustering algorithm, in order to analyse gene-gene links based on their co-expression relationships and gene differential analysis across breast cancer progression stages. Existing bibliography verifies the validity of the resulted clusters while exclusive pathways per breast cancer stage as well as a common pathway signature across all breast cancer stages are reported and proper discussion takes place.**

***Index Terms*—Gene-gene links clustering, hybrid space of co-expression and differential expression, density and distance content driven clustering algorithm, pathway analysis per breast cancer stage, gene co-expression networks**

I. Introduction

Breast cancer is the most frequent malignancy in women and one of the most common forms of cancer worldwide. Unfortunately, it is also a complex disease whose mechanisms are not yet fully understood [1]. In this work we aim to further analyze genes related to each breast cancer stage, by performing clustering on their gene-gene links characterized by a synthetic descriptor that combines both differential expression and co-expression. This has allowed for cluster detection of similarly related genes as opposed to solely relying on the proximity of two genes based on a single measure and diversifies from other approaches which cluster genes together based on their distance on a (weighted) graph.

The application of clustering algorithms in order to detect clusters of genes has already received a lot of attention [2]. A common choice for this problem is the usage of the hierarchical clustering algorithms [3]–[5]. Various other clustering algorithms have also been utilized [6], [7] or proposed their own clustering schemes [8]. The concept of identifying gene-gene relationships has also received significant attention as by Wu et al. [9].

In the current study, we applied a recently proposed Density and Distance Content driven (DeDiCo) clustering algorithm

George M. Spyrou holds the Bioinformatics ERA Chair, The Cyprus Institute of Neurology and Genetics, 6 International Airport Avenue, 2370 Ayios Dometios, Nicosia, Cyprus, funded by the European Commission Research Executive Agency (REA) Grant BIORISE (Num. 669026), under the Spreading Excellence, Widening Participation, Science with and for Society Framework. Marilena M. Bourdakou's postdoctoral research is being funded by the IKY (State Scholarships Foundation) scholarship, co-financed by Greece and the European Union (European Social Fund- ESF) through the Operational Programme "Human Resources Development, Education and Lifelong Learning" in the context of the project "Reinforcement of Post-doctoral Researchers - 2nd Cycle" (MIS5033021), implemented by the State Scholarships Foundation (IKY). This work was supported by computational time granted from the Greek Research & Technology Network (GRNET) in the National HPC facility – ARIS – under project ID PA070211-DeDiCoBC.

called DeDiCo [10], in a space defined by gene differential expression and co-expression relationships in order to cluster gene-to-gene links for each breast cancer stage. The choice of DeDiCo is justified since it makes no assumption in the number of clusters. Hence it addresses the problem of clustering gene-gene links without the restriction of providing the cluster number at hand or following a specific distribution, e.g. Gaussian. Our goal is to highlight inter-related genes and subsequently identify the corresponding molecular mechanisms per breast cancer stage.

II. Methods

A. From genes to networks and beyond

Microarray gene expression data (mRNA) for Breast Invasive Carcinoma were obtained from The Cancer Genome Atlas through Firehose (http://gdac.broadinstitute.org/). The total number of samples was 587 from which 526 are primary solid tumor and 61 normal samples. A subset of tumor samples which contain clinical information for the breast cancer staging (Stages I, II, III and IV) was selected. Each subset was statistically analyzed with the R package LIMMA (Linear Models for Microarray and RNA-Seq Data) [11] in order to find the differentially expressed genes (DEGs) in breast cancer samples for each stage compared with the corresponding normal ones. To start with a manageable gene set for each Stage, we filtered for the most statistically significant DEGs (with p-value< 0.01 and q-value< 0.01) and then we kept the top 1000 of them sorted in descending order based on their absolute log Fold Change.

Based on our previous work [12], we applied the correlation-based network inference method, GeneNet which achieved the best score in the analysis of breast cancer stages. A gene co-expression network for each breast cancer stage was constructed for the selected DEGs. For the construction of the co-expression networks we used the R package ENA (Ensemble Network Aggregation) and more specifically, the GeneNet function [13], [14]. Since co-expression networks are fully connected but usually with a great number of edges with extremely low weights, we kept a manageable number of the edges with the highest weight, namely the top 30000 gene-gene links that correspond to the top 6% of the network edges. For each link two features have been calculated: its corresponding weight and the maximum absolute log Fold Change of the two linked genes. All the feature measurements have been normalized to the unit interval.

B. The Density and Distance Content driven (DeDiCo) clustering algorithm

The notion of density is used in clustering as an expression of larger point groupings in datasets. On the other hand, points may be grouped together based on their distance, interpreting their similarity as their proximity. The DeDiCo clustering algorithm was recently proposed with the aim of combining those notions [10]. It identifies regions containing the majority of points by applying the notion of density and exploits the distance on those subsets of the dataset in order to refine the clustering. Those high density regions are called windows and are basically a hyper-rectangle [15] which is moved and enlarged in order to capture the majority of an existing cluster. The Affinity Propagation algorithm [16] is used so that potential cluster overlaps will not lead to merging adjacent clusters and allow the algorithm to capture clusters of arbitrary shape. Additionally, it has the ability to choose whether to use or not an evolutionary optimization algorithm. For this utilization of DeDiCo, Differential Evolution (DE) was used in an attempt to reduce the computational cost. That is, at each of its iterations, DeDiCo decides whether to use DE or simply evaluates the density around each point of the dataset at hand. Its choice is based on achieving the lowest number of density evaluations. The resulting clusters are merged together based on their distance. In the benchmarking presented in [10], DeDiCo was consistently in the top-two clustering algorithms when tested against six real-life datasets, in parallel with Affinity Propagation, k-Means, DBSCAN and FSDP (fast search and find of density peaks) clustering algorithms. Also it has presented the ability to determine arbitrarily shaped clusters without any assumptions on the shape and number of the clusters. Nevertheless, DeDiCo has a significant number of parameters, which allows it to adapt to a large range of clustering problems. For this reason, we have exploited the advantage of the High-Performance Computer Center ARIS of the Greek Research and Technology Network. For the purposes of this work, approximately 80K core hours where used. This involved initial experimentation as well as the parameters combination exploration. The clustering with the highest Silhouette [17] value has been selected for further analysis.

III. Results

The DeDiCo algorithm has been applied to the filtered co-expression network of each breast cancer (BC) stage. In order to find the underlying significant biological pathways derived from gene-gene links of each cluster for all breast cancer stages, KEGG pathway enrichment analysis was performed using the PathwayConnector (http://bioinformatics.cing.ac.cy/PathwayConnector/index.php?app=PathwayConnector) [18] – a web-tool that provides an easy way for rapidly relating human pathways together, by creating complementary networks of pathways related to a specific biological status. We have used the p-value< 0.05 in order to select the significantly enriched pathways for each cluster and stage. We have investigated the common and exclusive mechanisms of each stage and cluster and we have found four common cluster and stage related pathways (Table I). From these pathways, Complement and coagulation cascades has been reported to be a key regulatory mechanism in cancer development and growth [19] and cytochromes P450 (CYPs) are key enzymes in cancer formation and cancer treatment [20].

We have also investigated the Stage and cluster exclusive molecular mechanisms. As it is presented in Table I, there are three cluster and stage exclusive molecular pathways for Stage I. Examining the biological relevance of those pathways,

Helicobacter pylori was categorized as a carcinogen and it has been correlated with the tumor grade of gastric cancer [21]. For the case of Stage II, three common cluster and stage exclusive pathways were found and six cluster exclusive pathways (Table I). From these mechanisms, it has been reported that choline metabolism is associated with oncogenesis and tumor growth [22] and tight junction with the poor prognosis of breast cancer [23].

Following pathway analysis of our findings for the case of Stage III, six pathways have been found as cluster and stage exclusive (Table I). From these pathways, Adrenergic signaling in cardiomyocytes has been associated with breast cancer progression [24] and fructose and mannose metabolism has been found deregulated in breast cancer patients [25].

For the case of Stage IV, 18 stage exclusive pathways were found, from which five (Mismatch repair, Pancreatic cancer, Cysteine and methionine metabolism, Oxytocin signaling pathway and Leukocyte transendothelial migration) are also exclusive between clusters. It has been reported that Mismatch Repair Polymorphisms may be possible markers of breast cancer [26] and with carcinogenesis, as it is involved in the development and cancer progression [27]. From the remaining 13 molecular pathways that were found as exclusive for Stage IV melanomas and lymphomas are the most commonly reported tumors metastasizing to the breast [28] and AMPK and Chemokine signaling pathways have been proposed as possible targets for the prevention, development and metastasis of breast cancer [29], [30].

DeDiCo clusters gene-gene links by taking into account the fold change between two genes, as well as their co-expression. Both differential expression and differential co-expression analyses, are broadly considered to be useful tools in understanding gene regulation related to complex diseases, such as breast cancer. As it is presented in Fig. 1, each cluster consists of gene-gene relationships with similar co-expressions and differential expressions. More specifically, Fig. 1 represents the gene-gene relationships of Stage I. The remaining figures from the other Stages can be found in our GitHub (https://github.com/CING-BIG/Gene-Gene-Links-Clustering). Setting a threshold of 0.25 in both the normalized co-expression and differential expression axes to find the top gene-gene links from all clusters, we have resulted to 77 gene-gene links and 91 unique genes for Stage I, 24 significant gene-gene links and 43 unique genes for Stage II, 48 gene-gene links and 68 genes for Stage III and 24 gene associations as well as 39 unique genes for Stage IV respectively.

It is worth mentioning that the sub-networks from all stages consist of significant genes for breast cancer such as KRT19, HOTAIR, KRT5, SCGB2A2, CXCR4, MUCL1, MMP1, CEACAM5, MMP9, TYMS, FOS, PTGS2, S100A7, SPP1, IL6 and MMP13 that are strongly related to breast cancer based on the Malacards database (http://www.malacards.org/card/).

TABLE I
THE LETTER C MARKS THE PATHWAYS THAT ARE CONSERVED ACROSS ALL CLUSTERS IN EACH BC STAGE WHILE E MARKS THE PATHWAYS THAT ARE EXCLUSIVE IN EACH BC STAGE

Pathway Name	Stages			
	I	II	III	IV
Metabolic pathways	**E**			
Epithelial cell signaling in Helicobacter pylori infection	**E**			
Pentose and glucuronate interconversions	**E**			
Amoebiasis		**E/C**		
Choline metabolism in cancer		**E/C**		
Tight junction		**E/C**		
Fc gamma R-mediated phagocytosis		**E**		
Cocaine addiction		**E**		
Malaria		**E**		
Insulin resistance		**E**		
ABC transporters		**E**		
Glycolysis / Gluconeogenesis		**E**		
Legionellosis			**E**	
Salivary secretion			**E**	
Prion diseases			**E**	
Adrenergic signaling in cardiomyocytes			**E**	
Fructose and mannose metabolism			**E**	
Thyroid cancer			**E**	
Mismatch repair				**E**
Pancreatic cancer				**E**
Cysteine and methionine metabolism				**E**
Oxytocin signaling pathway				**E**
Leukocyte transendothelial migration				**E**
Neuroactive ligand-receptor interaction				**E/C**
Viral carcinogenesis				**E/C**
Melanoma				**E/C**
Systemic lupus erythematosus				**E/C**
Glycerolipid metabolism				**E/C**
Fatty acid degradation				**E/C**
Cell cycle				**E/C**
AMPK signaling pathway				**E/C**
Serotonergic synapse				**E/C**
Chemokine signaling pathway				**E/C**
Nitrogen metabolism				**E/C**
Regulation of lipolysis in adipocytes				**E/C**
Retinol metabolism				**E/C**
Tyrosine metabolism	**C**	**C**	**C**	**C**
Phenylalanine metabolism	**C**	**C**	**C**	**C**
Drug metabolism - cytochrome P450	**C**	**C**	**C**	**C**
Complement and coagulation cascades	**C**	**C**	**C**	**C**
AGE-RAGE signaling pathway in diabetic complications				**C**

IV. DISCUSSION

Clinical predictors are the most important features for breast cancer prognosis but they are not enough for the accurate prediction of the disease [31]. Furthermore, a single gene is deficient as a biomarker for the prediction of patient survival. In the current study, we have investigating the cluster formation of gene-to-gene links for each breast cancer stage. We have applied a recently proposed density and distance content driven clustering algorithm in a space defined by gene differential expression and co-expression relationships. The motivation of clustering gene-to-gene links per breast cancer stage is to highlight inter-related genes and subsequently identify the corresponding molecular mechanisms per breast cancer stage.

Our results reveal four common cluster and stage related pathways including Complement and coagulation cascades among others. The functions of the complement system protect the organism from pathogens. In the breast cancer staging, these functions are regulated to change the environment in

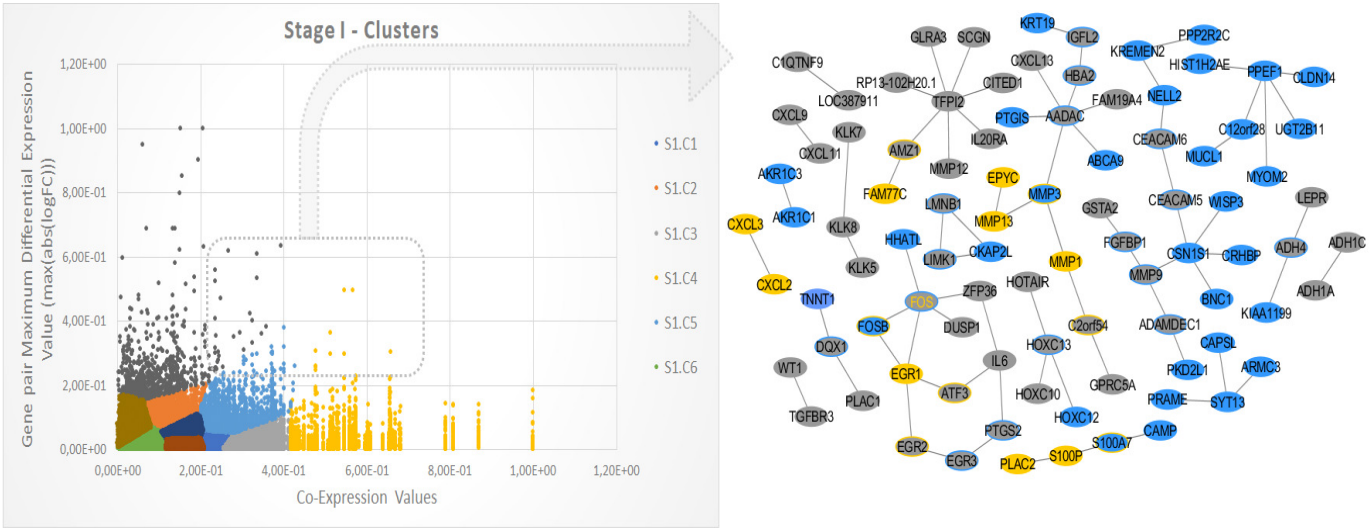

Fig. 1. Stage I: Scatterplot regarding the gene-pair maximum differential expression value vs gene-pair co-expression value. A rectangular region in this scatterplot with down-left corner at (Dif.Expr. Threshold, Coexpr.Threshold) = (0.25, 0.25) includes the best gene-gene links of simultaneous high co-expression and dif. expression values. The small network pieces that correspond to these links are embedded in a picture-in-picture mode.

order to suit for cancer progression. Furthermore, proteins involved in complement and metabolism pathways were altered in several tumor stages [32], [33].

For the case of exclusive pathways, Helicobacter pylori infection is one of the exclusive pathways that was found for Stage I. In addition it has been associated with gastric cancer but there is no evidences that it is associated with breast cancer [34], [35]. Moreover, three pathways were found to be common in all clusters of Stage II and six cluster and stage exclusive. From these pathways, alterations in fructose and mannose metabolism have also been found in human breast cancer [25]. Moreover, evidences suggest Adrenergic signaling in cardiomyocytes pathway may be related with breast cancer development [24]. Finally, the increased expression levels of prion proteins in breast cancer are associated with poor prognosis [36], [37].

Following pathway analysis of our findings for the case of Stage IV, 18 stage exclusive pathways were found, from which five of them were also exclusive between clusters: Namely, Mismatch repair, Pancreatic cancer, Cysteine and methionine metabolism, Oxytocin signaling pathway as well as Leukocyte transendothelial migration. Oxytocin signaling pathway has been indicated as possible biomarker for breast cancer [27]. From the other 13 exclusive pathways for Stage IV, AMPK signaling pathway is related with breast cancer growth and metastasis [29].

In this study we have also examined the conserved subnetworks for each breast cancer stage, to highlight the significant genes. PBK that was found exclusively in Stage II subnetwork, has significantly higher expression levels in patients at Stages II and III compared to Stage I [38]. CXCL9 (Stages I and II) has a central role in tumor progression and may be a possible target for cancer therapy [39]. Moreover, the expression levels of IL6 (found in Stages I and III) were correlated with Stages II-III breast carcinoma [40]. Finally, CXCR4 that was found exclusively in Stage IV, plays a pivotal role in metastatic breast cancer [41].

A limitation in the investigation of gene-to-gene links and their corresponding molecular pathways is the fact that there is often no ground truth. For this reason it is difficult to validate the highlighted molecular pathways apart from comparing results from the literature. Although our findings are computational, they give insights for experimental validation. The novelty of our approach lies in the exploitation of the complex gene-to-gene relationships via a clustering algorithm that allows researchers to obtain results that are more closely tied to the biological mechanisms per breast cancer stage.

V. Conclusions

The analysis in this work with DeDiCo has been concluded to significant breast cancer stages-related mechanisms, genes and gene-gene links. Further investigation of the suggested and unexplored so far mechanisms and gene-gene links that have been proposed by DeDiCo clustering will provide a more systematic understanding of the complex breast cancer staging mechanism, which in turn yields useful insights in the development of new therapeutic strategies.

Competing Interests

The authors have no competing financial interests.

References

[1] N. Harbeck, F. Penault-Llorca, J. Cortes, M. Gnant, N. Houssami, P. Poortmans, K. Ruddy, J. Tsang, and F. Cardoso, "Breast cancer," *Nature Reviews Disease Primers*, vol. 5, no. 1, p. 66, 2019.

[2] J. Oyelade, I. Isewon, F. Oladipupo, O. Aromolaran, E. Uwoghiren, F. Ameh, M. Achas, and E. Adebiyi, "Clustering algorithms: Their application to gene expression data," *Bioinformatics and Biology Insights*, vol. 10, pp. 237–253, 2016.

[3] A. El Guerrab, A. Cayre, F. Kwiatkowski, M. Privat, J.-M. Rossignol, F. Rossignol, F. Penault-Llorca, and Y.-J. Bignon, "Quantification of hypoxia-related gene expression as a potential approach for clinical outcome prediction in breast cancer," *PLoS ONE*, vol. 12, no. 4, p. e0175960, 2017.

[4] M. Verbanck, S. Lê, and J. Pagès, "A new unsupervised gene clustering algorithm based on the integration of biological knowledge into expression data," *BMC Bioinformatics*, vol. 14, no. 1, p. 42, 2013.

[5] K. Sirinukunwattana, R. S. Savage, M. F. Bari, and D. R. J. Snead, "Bayesian hierarchical clustering for studying cancer gene expression data with unknown statistics," *PLoS ONE*, p. e75748, 2013.

[6] A. E. Brown, E. I. Yemini, L. J. Grundy, T. Jucikas, and W. R. Schafer, "A dictionary of behavioral motifs reveals clusters of genes affecting caenorhabditis elegans locomotion," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 2, pp. 791–796, 2013.

[7] Z. Liu, Y.-q. Song, C.-h. Xie, and Z. Tang, "A new clustering method of gene expression data based on multivariate gaussian mixture models," *Signal, Image and Video Processing*, vol. 10, no. 2, pp. 359–368, 2016.

[8] X. Yu, G. Yu, and J. Wang, "Clustering cancer gene expression data by projective clustering ensemble," *PLoS ONE*, vol. 12, no. 2, p. e0171429, 2017.

[9] J. Wu, S. Hu, Y. Chen, Z. Li, J. Zhang, and H. Yuan, "BCIP: a gene-centered platform for identifying potential regulatory genes in breast cancer," *Scientific Reports*, vol. 7, p. 45235, 2017.

[10] E. K. Ikonomakis, G. M. Spyrou, and M. N. Vrahatis, "Content driven clustering algorithm combining density and distance functions," *Pattern Recognition*, vol. 87, pp. 190–202, 2019.

[11] G. K. Smyth, "Linear models and empirical Bayes methods for assessing differential expression in microarray experiments," *Statistical Applications in Genetics and Molecular Biology*, vol. 3, no. 1, p. 3, 2004.

[12] M. M. Bourdakou, E. I. Athanasiadis, and G. M. Spyrou, "Discovering gene re-ranking efficiency and conserved gene-gene relationships derived from gene co-expression network analysis on breast cancer data," *Scientific Reports*, vol. 6, p. 20518, 2016.

[13] M. K. Lau, S. R. Borrett, D. E. Hines, and P. Singh, *enaR: Tools for Ecological Network Analysis*, 2017, r package version 2.10.0. [Online]. Available: https://CRAN.R-project.org/package=enaR

[14] R. Zhong, J. D. Allen, G. Xiao, and Y. Xie, "Ensemble-based network aggregation improves the accuracy of gene network reconstruction," *PLoS ONE*, vol. 9, no. 1, p. e106319, 2014.

[15] M. N. Vrahatis, B. Boutsinas, P. Alevizos, and G. Pavlides, "The new k-windows algorithm for improving the k-means clustering algorithm," *Journal of Complexity*, vol. 18, no. 1, pp. 375–391, 2002.

[16] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, pp. 972–976, 2007.

[17] P. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.

[18] G. Minadakis, M. Zachariou, A. Oulas, and G. M. Spyrou, "Pathwayconnector: finding complementary pathways to enhance functional analysis," *Bioinformatics*, vol. 35, pp. 889–891, 2019.

[19] M. Krupp, T. Maass, J. U. Marquardt, F. Staib, T. Bauer, R. König, S. Biesterfeld, P. R. Galle, A. Tresch, and A. Teufel, "The functional cancer map: A systems-level synopsis of genetic deregulation in cancer," *BMC Medical Genomics*, vol. 4, no. 1, p. 53, 2011.

[20] C. Rodriguez-Antona and M. Ingelman-Sundberg, "Cytochrome P450 pharmacogenetics and cancer," *Oncogene*, vol. 25, no. 11, pp. 1679–1691, 2006.

[21] S. Mohammadi Saravle, M. Ahmadi Hedayati, E. Mohammadi, F. Sheikhesmaeili, and B. Nikkhoo, "Sirt1 gene expression and gastric epithelial cells tumor stage in patients with helicobacter pylori infection," *Asian Pacific journal of cancer prevention*, vol. 19, no. 4, pp. 913–916, 2018.

[22] K. Glunde, Z. M. Bhujwalla, and S. M. Ronen, "Choline metabolism in malignant transformation," *Nature reviews. Cancer*, vol. 11, no. 12, pp. 835–848, 2011.

[23] T. A. Martin, G. Watkins, R. E. Mansel, and W. G. Jiang, "Loss of tight junction plaque molecules in breast cancer tissues is associated with a poor prognosis in patients with breast cancer," *European Journal of Cancer*, vol. 40, no. 18, pp. 2717–2725, 2004.

[24] E. I. Obeid and S. D. Conzen, "The role of adrenergic signaling in breast cancer biology," *Cancer Biomarkers*, vol. 13, no. 3, pp. 161–169, 2013.

[25] G. Schramm, E.-M. Surmann, S. Wiesberg, M. Oswald, G. Reinelt, R. Eils, and R. König, "Analyzing the regulation of metabolic pathways in human breast cancer," *BMC Medical Genomics*, vol. 3, no. 1, p. 39, 2010.

[26] M. Kappil, M. B. Terry, L. Delgado-Cruzata, Y. Liao, and R. M. Santella, "Mismatch repair polymorphisms as markers of breast cancer prevalence in the breast cancer family registry," *Anticancer research*, vol. 36, no. 9, pp. 4437–4441, 2016.

[27] B. Lerman, T. Harricharran, and O. O. Ogunwobi, "Oxytocin and cancer: An emerging link," *World Journal of Clinical Oncology*, vol. 9, pp. 74–82, 2018.

[28] M. Moschetta, M. Telegrafo, N. M. Lucarelli, G. Martino, L. Rella, A. A. Stabile Ianora, and G. Angelelli, "Metastatic breast disease from cutaneous malignant melanoma," *International Journal of Surgery Case Reports*, vol. 5, no. 1, pp. 34–36, 2014.

[29] H. Zhao, Y. C. Orhan, X. Zha, E. Esencan, R. T. Chatterton, and S. E. Bulun, "AMP-activated protein kinase and energy balance in breast cancer," *American Journal of Translational Research*, vol. 9, no. 2, pp. 197–213, 2017. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5340661/

[30] G. Xiao, X. Wang, J. Wang, L. Zu, G. Cheng, M. Hao, X. Sun, Y. Xue, J. Lu, and J. Wang, "CXCL16/CXCR6 chemokine signaling mediates breast cancer progression by pERK1/2-dependent mechanisms," *Oncotarget*, vol. 6, no. 16, pp. 14 165–14 178, 2015.

[31] E. S. Papadakis, T. Reeves, N. H. Robson, T. Maishman, G. Packham, and R. I. Cutress, "BAG-1 as a biomarker in early breast cancer prognosis: a systematic review with meta-analyses," *British Journal of Cancer*, vol. 116, pp. 1585–1594, 2017.

[32] S. J. Pitteri, K. S. Kelly-Spratt, K. E. Gurley, J. Kennedy, T. B. Buson, A. Chin, H. Wang, Q. Zhang, C.-H. Wong, L. A. Chodosh, P. S. Nelson, S. M. Hanash, and C. J. Kemp, "Tumor microenvironment-derived proteins dominate the plasma proteome response during breast cancer induction and progression," *Cancer Research*, vol. 71, no. 15, pp. 5090–5100, 2011.

[33] M.-N. Song, P.-G. Moon, J.-E. Lee, M. Na, W. Kang, Y. S. Chae, J.-Y. Park, H. Park, and M.-C. Baek, "Proteomic analysis of breast cancer tissues to identify biomarker candidates by gel-assisted digestion and label-free quantification methods using LC-MS/MS," *Archives of Pharmacal Research*, vol. 35, no. 10, pp. 1839–1847, 2012.

[34] S. Alzahrani, T. T. Lina, J. Gonzalez, I. V. Pinchuk, E. J. Beswick, and V. E. Reyes, "Effect of helicobacter pylori on gastric epithelial cells," *World journal of gastroenterology*, vol. 20, no. 36, pp. 12 767–12 780, 2014.

[35] M. Naumann and J. E. Crabtree, "Helicobacter pylori-induced epithelial cell signalling in gastric carcinogenesis," *Trends in Microbiology*, vol. 12, no. 1, pp. 29–36, 2004.

[36] C.-H. Yang, Y.-D. Lin, L.-Y. Chuang, and H.-W. Chang, "Double-bottom chaotic map particle swarm optimization based on chi-square test to determine gene-gene interactions," *BioMed Research International*, vol. 2014, p. 10, 2014.

[37] M.-A. Déry, J. Jodoin, J. Ursini-Siegel, O. Aleynikova, C. Ferrario, S. Hassan, M. Basik, and A. C. LeBlanc, "Endoplasmic reticulum stress induces PRNP prion protein gene expression in breast cancer," *Breast Cancer Research*, vol. 15, no. 2, p. R22, 2013.

[38] Q. Luo, B. Lei, S. Liu, Y. Chen, W. Sheng, P. Lin, W. Li, H. Zhu, and H. Shen, "Expression of PBK/TOPK in cervical cancer and cervical intraepithelial neoplasia," *International Journal of Clinical and Experimental Pathology*, vol. 7, no. 11, pp. 8059–8064, 2014. [Online]. Available: https://europepmc.org/articles/PMC4270522

[39] Q. Ding, P. Lu, Y. Xia, S. Ding, Y. Fan, X. Li, P. Han, J. Liu, D. Tian, and M. Liu, "CXCL9: evidence and contradictions for its role in tumor progression," *Cancer Medicine*, vol. 5, no. 11, pp. 3246–3259, 2016.

[40] Y. Ma, Y. Ren, Z.-J. Dai, C.-J. Wu, Y.-H. Ji, and J. Xu, "IL-6, IL-8 and TNF-α levels correlate with disease stage in breast cancer patients," *Advances in Clinical and Experimental Medicine*, vol. 26, no. 3, pp. 421–426, 2017.

[41] D. Mukherjee and J. Zhao, "The role of chemokine receptor CXCR4 in breast cancer metastasis," *American Journal of Cancer Research*, vol. 3, no. 1, pp. 46–57, 2013.