Understanding Heterogeneity in Pregnancy-Associated Breast Cancer

T. Murlidharan Nair
Department of Biological Sciences and Computer Science/Informatics
Indiana University South Bend
1700 Mishawaka Ave, South Bend, IN-46634 USA

***Abstract*- Breast cancers diagnosed during pregnancy are generally in their advanced stages. These cancers occurring during pregnancy and up to one year postpartum are termed as pregnancy associated breast cancer. Recent genomic studies by Harvell *et. al.* was the first large scale study that attempted to identify molecular signatures associated with pregnancy associated breast cancer. In this study, we have rigorously analyzed the data with a view to identify features involved in within-class and between-class separation. The results reveal features that are unique between classes, as well as point to the importance of understanding heterogeneity within the same class.**

Keywords: Pregnancy associated breast cancer; multiple comparison

I. INTRODUCTION

Breast cancer is a heterogeneous disease and the second most common cancer in women after skin cancer. The number of new cases of breast cancer that are detected is 124.8 per 100,000 women each year [1]. Based on the SEER stat fact sheet, using the 2010-2012 data, approximately 12.3 percent women will be diagnosed with breast cancer sometime in their lifetime. Major risk factors for the disease include age, family history, which doubles, if a first-degree relative (mother, sister or daughter) had the disease and high risk mutations in the BRCA1 and BRCA2 genes. There is data to suggest an increase in breast cancer after pregnancy [2]. The term pregnancy-associated breast cancer (PABC) is used for cases that are diagnosed during gestation, lactation and up to one year postpartum [2-4].

PABC cases diagnosed are typically in advanced stages and have poor prognosis compared to women who are not pregnant [5]. Women with PABC have higher grade tumors and are often estrogen and progesterone negative [6, 7]. It is thus important to understand the relationship between pregnancy and breast cancer. Duration of exposure to endogenous hormones viz. estrogen and progesterone produced in the ovaries have been related to breast cancer risk. There an inverse relationship between a women's lifetime number of menstrual cycles, and both pregnancy and breastfeeding [8]. Nursing for an extended period also decreases the risk for estrogen positive and negative breast cancers [9]. Further, there is evidence to suggest that early full term pregnancy and multiparity have a decreased risk in women of all ethnic groups for developing breast cancer [10, 11]. However, the risk of PABC is greatest in older first time mothers. Metastasis is common in PABC, which then results in higher mortality rates [2]. Breast cancer metastasis associated with pregnancy may be attributed to the promotional effects of pregnancy associated hormones as well as the delay in diagnosis [2].

Pathologically it has been noted that invasive ductal and lobular carcinomas is the most prevalent type, while the inflammatory types are relatively rare [12]. As mentioned earlier, estrogen and progesterone receptors have been found to be mostly negative in PABC [12], but this high negativity could also be attributed to the downregulation of these receptors [13].

The relationship between breast cancer and pregnancy is complex. Several studies have shown that prognosis of PABC is invariant to non-pregnancy associated breast cancer (non-PABC), if tumor size, nodal status and other established prognostic markers are comparable[14, 15]. However, normal gestational changes may mask masses in breast during pregnancy, and result in a delay in diagnosis, and contribute to poor prognosis [2, 7, 16].

More recently, the role of gestational hormones on breast cancer have been evaluated using gene expression patterns [17]. This is one of the largest high throughput study undertaken to understand the molecular signatures associated with PABC. The study analyzed expression patterns from epithelial and stromal cells in breast. The epithelial cells line the ducts and milk producing lobules, while the stromal cells contribute to the makeup of vasculature, basement membrane and extracellular matrix. This allowed the study to capture differential expression in malignant epithelial and tumor associated stromal cells. The analysis revealed a subset of genes associated with cell cycle to be enriched in PABC and immune-related genes enriched in non-PABC [17]. In order to understand the complexities associated with PABC, we have rigorously analyzed this data for changes in expression between PABC and non-PABC (epithelial and stromal) using multiple comparison. Further, we have also studied the changes in expression within the same class with a view to illustrate the heterogeneity associated with PABC. Our analysis revealed common and disparate features in expression patterns, as well as the importance of understanding differential expression within the same class, when studying complex systems.

II. MATERIALS AND METHODS

A. Data

The expression data used in this study were obtained from the earlier work by Harvel *et al* (2013) [17]. The data consisted of gene expression data profiled using HG-U133 Plus 2.0 (Affymetrix) gene chips that contained over 54,000 probe sets. (Accession number GSE31192).

978-1-4673-6799-8/15/$31.00 ©2015 IEEE

The data set consisted of a total of 33 samples, out of which 13 were normal and 20 were tumor. Out of the 13 normal samples, 8 were PABC and 5 were non-PABC. Out of the 20 tumor samples, 12 were PABC and 8 were non-PABC. Further, within each of the two classes (normal and tumor), the samples were further categorized as Er-Pos or Er-Neg and of epithelial or stromal origin.

B. Statistical analysis

A scoring scheme using multiple comparisons has been used to extract features between classes. We carried out comparisons within and between all possible categories. In each case we established contrasts, which we tested using Tukey's HSD method [18-20]. All the statistical computation were done using the multcomp package from the R statistical computing environment (http://www.r-project.org) [21]. Exhaustive comparison of expression levels was carried out within and between each class. Each gene or probe that showed a significant difference in either a within-class or between-class comparison was given a unit score. The sum of the scores was used to rank the genes in the respective comparisons. This can be clearly understood from Figure 1, which depicts four genes, NHP2L1, MTHFD2, EGR1, and FAM66D, with scores, one, two three and four respectively. The scores reflect the total number of comparisons that reveled a significant difference. Significant comparisons are those that do not intersect the zero line. The top ranking genes were then selected as features for class separation using a clustering algorithm [22].

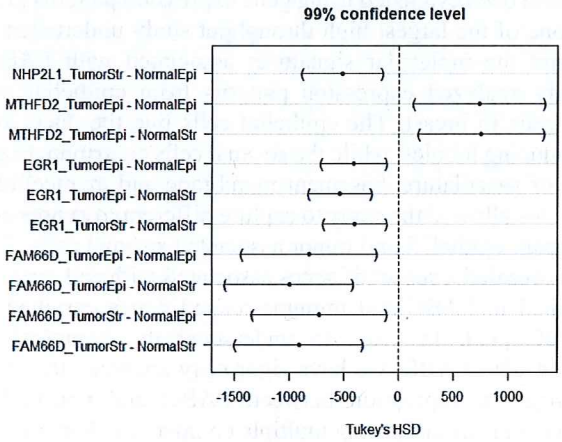

Figure 1: Multiple comparison of four genes viz NHPL1, MTHFD2, EGR1 and FAM66D

III. RESULTS AND DISCUSSION

The expression levels that were captured by Harvell *et. al.* [17] were categorized in several ways, each time ensuring that statistical power is not compromised. They are discussed in the subsections below.

A. Epithelial vs Stromal cells

The dataset was categorized as normal or tumor cells associated with epithelium or stroma. All of the samples were from invasive ductal carcinomas and majority of patients did not have nodal involvement and had stage III disease [17]. There were a total of 13 normal samples and 20 tumor samples. Among the 13 normal samples, 7 were stromal, and out of the 20 tumor samples, 10 were stromal. The categorization here did not differentiate the samples as PABC and non-PABC. The goal of this comparison was to delineate the similarities and differences between stromal tumors, epithelial tumors as well as between normal stroma and epithelium.

The comparison between normal and tumor cells revealed a total of 147 features that were significantly different. All of the 147 features had a score of four, which meant that they were significantly different in all possible between-class comparisons. These features were then used to hierarchically cluster the two classes under study.

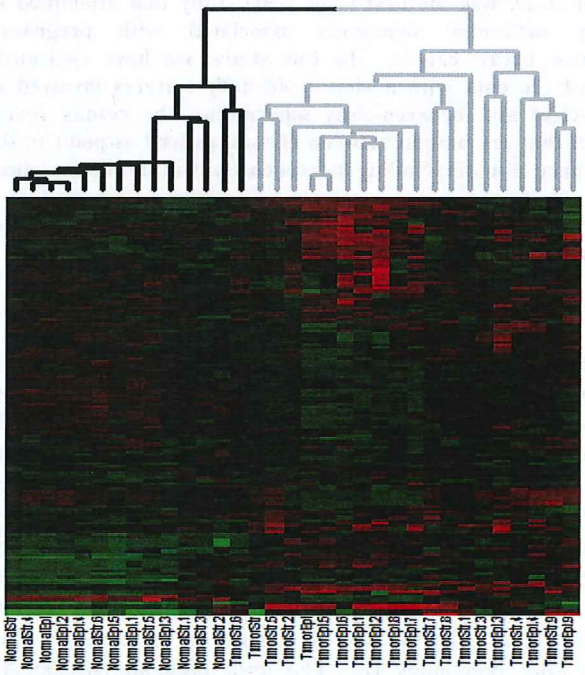

Figure 2: Hierarchical clustering using 147 features extracted by multiple comparison. Black lines represents normal and gray represents tumor.

The hierarchical clustering in Figure 2 clearly shows that the features that were extracted using multiple comparison are capable of good class separation, with only one of the tumor samples being clustered with the normal sample. Functional analysis using the selected features reveal that disparate sets of genes are up or down regulated between the two classes.

We also analyzed the data to understand the difference between tumor in the epithelium and stroma as well as the difference between the corresponding normal samples. There were close to 2000 feature that were differentially expressed between the tumors in the epithelium and stroma. The features were capable of clearly separating the two types of tumors robustly. Figure 3 shows the hierarchical clustering of the two tumor types based on the features selected. It is interesting to note that there is a clear difference in the expression levels of the genes associated with stroma and epithelium. The ones that are upregulated in one type is down regulated in the other. This

difference points to the fact that different pathways are disrupted in these two tumor types, and either of these disruptions are capable of inducing tumorigenesis in the cell.

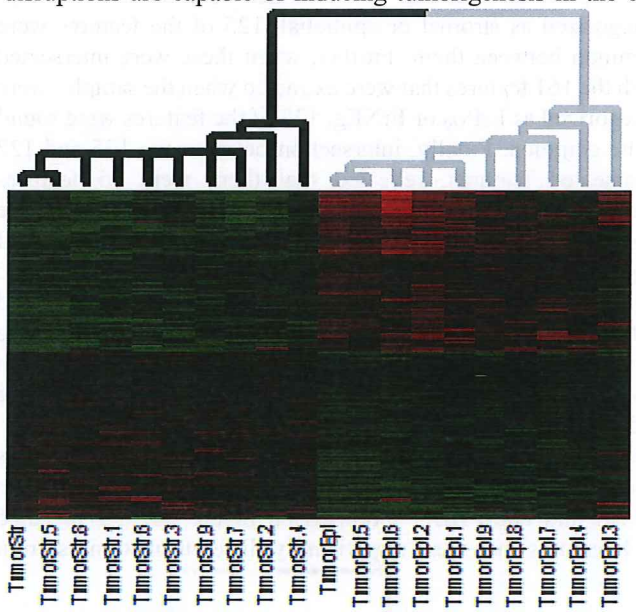

Figure 3: Hierarchical clustering of tumor in the stroma (black) and epithelium (gray)

We were also interested in understanding the difference within the normal samples. In order to achieve this we extracted features that were differentially expressed in normal stoma and normal epithelium. There were over a thousand features that were differentially expressed.

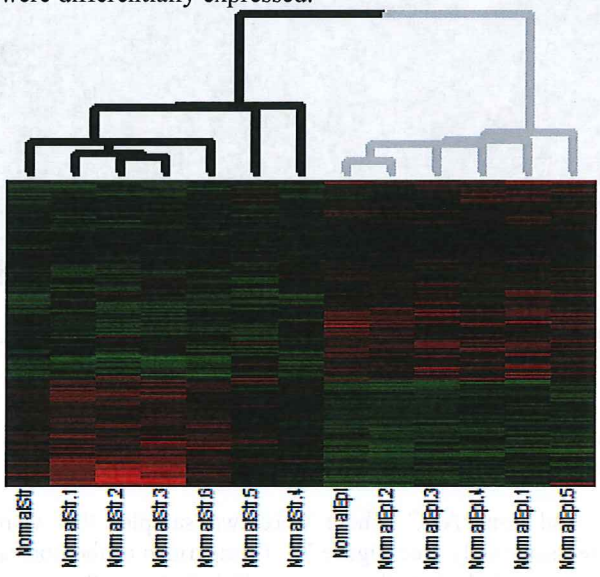

Figure 4: Hierarchical clustering of normal stromal (black) and epithelium (gray) using the extracted features.

Clustering the data using these features is shown in Figure 4, very clearly distinguishes the two types of samples. It is noteworthy that the features that were upregulated in stroma are down regulated in epithelium. The differences would be expected, since stroma and epithelium are functionally different. Stroma is part of a tissue that has a connective or structural role, while epithelial tissues lines the cavities and surfaces of the organ.

Interestingly enough, the intersection of features that differentiated the within-class sample of tumor and normal revealed that there were 224 features that were common. These genes could have tissue specific expression and are likely to be unrelated to tumorigenesis. This still leaves over 1700 features that could have tumor specific expression, probably indicating that different pathways that have been disrupted in stromal vs epithelial tumors [23].

B. Estrogen Positive vs Estrogen negative

In this categorization the data was classified based on hormonal receptor signatures. The distribution of the samples based on the hormonal receptor signature was the same as that in stroma vs epithelium. The normal samples had 6 Er-Pos samples while in the tumor there were 12 Er-Pos samples. Here again the categorization did not differentiate the sample as PABC or non-PABC. We also ignored their origin, i.e. whether they were stromal or epithelial. The goal here was to delineate the differences and similarities based on hormonal receptor signatures.

The comparison between normal tumor samples revealed about 167 signatures that were capable of class separation with one misclassification (data not shown). The classification was similar to that, when the samples were categorized as stromal and epithelial (Figure 2). The intersection of these features revealed that 109 features were common between the two categories. This also brings to light that there were about 50 features that were different. We hypothesize that these features are specific to hormonal function. Detail functional analysis is still work in progress.

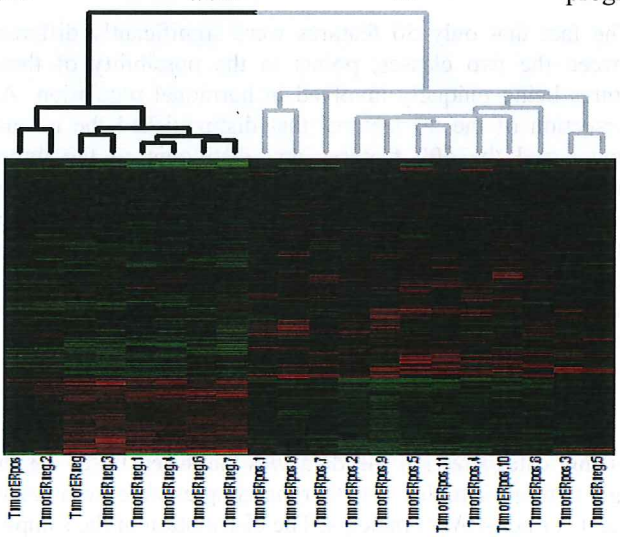

Figure 5: Hierarchical clustering within-class tumor ErPos ErNeg

Analysis of the feature within-class were done to understand the unique signatures in tumor and normal based on hormonal regulations. Analysis of the tumor data revealed that there were

over 500 features that were different in Er-Pos and Er-Neg samples. These features were sufficient for class separation (see Figure 5). These results point to the likelihood of different underlying biological phenomena associated with Er-Pos and Er-Neg tumors.

Features were also extracted in a within-class comparison of the normal samples. This was done with the view to identify the differences between the normal samples based on hormone receptor status. The comparison revealed over 50 features that were significantly different within the normal samples. These features were capable of grouping the samples as Er-Pos or Er-Neg (See Figure 6).

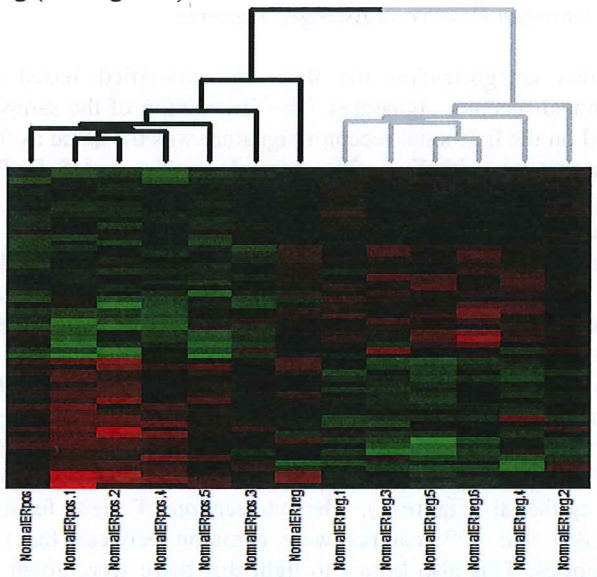

Figure 6: Hierarchical clustering of the normal samples showing class separation based on hormone receptor status.

The fact that only 50 features were significantly different between the two classes, points to the possibility of these features being uniquely involved in hormonal regulation. An intersection of the 50 features that distinguished the normal samples and the 500 features that distinguished the tumor sample revealed only two common features. These results reveal the possibility that tumors with differing hormonal status, likely have completely different pathways for transformation. The underlying pathways involved and disrupted is still under study as part of the detailed functional analysis.

C. PABC vs non-PABC

In this categorization the data was classified based on the origin of tissue samples. Whether the samples were from PABC patients or non-PABC patients. The distribution of the samples were slightly different. Out of the 13 normal samples, 8 of them were from PABC patients. Further, out of the 20 tumor samples, 12 were from PABC patients. The goal here was to capture features unique to PABC. Multiple comparison were done as described in the methods section and the features extracted. Analysis of the features revealed 185 features that were significantly different between the tumor and normal samples. When the features obtained in this comparison were intersected with the 148 features that were obtained when the samples were categorized as stromal or epithelial, 125 of the features were common between them. Further, when these were intersected with the 161 features that were extracted when the samples were categorized as ErPos or ErNEg, 122 of the features were found to be common. Finally, intersection between the 125 and 122 intersected features, revealed that there were 95 features common in all. Based on the number of features that are significantly different between the classes compared, we could hypothesize that there are distinct features that are unique to PABC and non-PABC. The class separation that we obtained between the normal and tumor samples were similar to those obtained when they were categorized as being from stromal/epithelial origin or Er-Pos/Er-Neg. As in those case we did have one sample misclassified.

Within-class analysis was done in this case as well. This was done to understand the unique features distinguishing tumors in PABC and non-PABC. Comparison of the samples revealed 274 features that were significantly different in tumors from

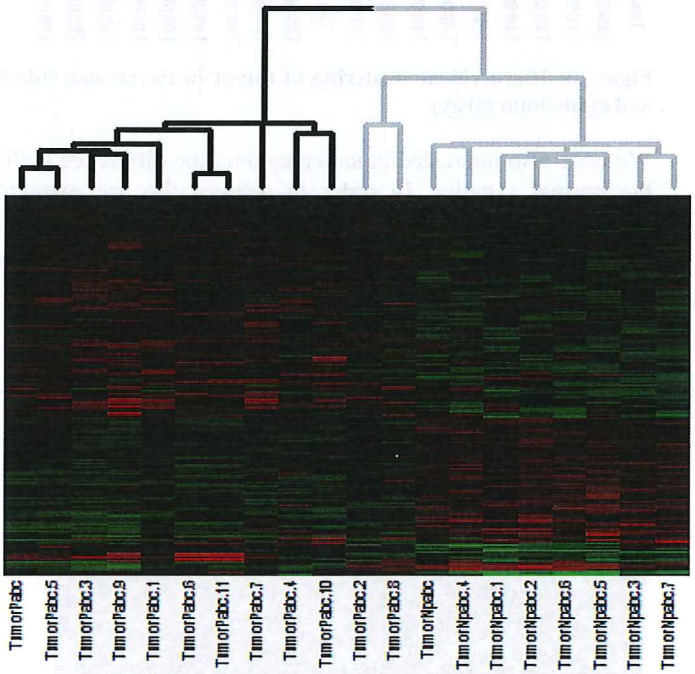

Figure 7: Hierarchical clustering of the tumor samples showing separation between PABC and non-PABC

PABC and non-PABC. There were two samples that were clustered separately (see Figure 7). Comparison of the normal samples revealed that there were 327 features that were significantly different between PABC and non-PABC. These features were able to provide good class separation when clustered (data not shown). Intersection of features that were significantly different between tumor and normal, revealed that there were only 8 features that were common. This reveals that there is a significant difference between the molecular signatures in PABC and non-PABC. We are currently working on rigorously analyzing these features to mine the underlying

biology that distinguishes PABC and non-PABC. It would be worthwhile to unravel the complex interplay between hormones during pregnancy, and the aggressiveness of cancers, when compared to cancers from non-pregnant women.

IV. Conclusion

Understanding precisely the molecular signatures associated with cancer is a complex and challenging problem. Even when features are identified that are capable of class separation, they may not necessarily be biologically relevant. The true features may be masked because biological samples are inherently heterogeneous. The rigorous comparisons that were undertaken here point to importance of understanding within-class differences. Such comparisons provide a view of the complexity in the data and the results could be used to enhance the true difference that exist between the classes. These are preliminary results and we are still working on the functional analysis of the identified features.

Acknowledgment

I would like to thank IUSB for supporting my research. The paper is dedicated to the memory of my father (Col. T. S. S. Nair, VSM) who recently passed away after a long battle with cancer.

References

[1] N. Howlader, S. F. Altekruse, C. I. Li, V. W. Chen, C. A. Clarke, L. A. Ries, and K. A. Cronin, "US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status," *J Natl Cancer Inst,* vol. 106, no. 5, May, 2014.

[2] P. Schedin, "Pregnancy-associated breast cancer and metastasis," *Nat Rev Cancer,* vol. 6, no. 4, pp. 281-91, Apr, 2006.

[3] K. H. Dow, "Pregnancy and breast cancer," *J Obstet Gynecol Neonatal Nurs,* vol. 29, no. 6, pp. 634-40, Nov-Dec, 2000.

[4] T. R. Lyons, P. J. Schedin, and V. F. Borges, "Pregnancy and breast cancer: when they collide," *J Mammary Gland Biol Neoplasia,* vol. 14, no. 2, pp. 87-98, Jun, 2009.

[5] D. M. Barnes, and L. A. Newman, "Pregnancy-associated breast cancer: a literature review," *Surg Clin North Am,* vol. 87, no. 2, pp. 417-30, x, Apr, 2007.

[6] P. B. Gupta, D. Proia, O. Cingoz, J. Weremowicz, S. P. Naber, R. A. Weinberg, and C. Kuperwasser, "Systemic stromal effects of estrogen promote the growth of estrogen receptor-negative cancers," *Cancer Res,* vol. 67, no. 5, pp. 2062-71, Mar 1, 2007.

[7] M. Ulery, L. Carter, B. L. McFarlin, and C. Giurgescu, "Pregnancy-associated breast cancer: significance of early detection," *J Midwifery Womens Health,* vol. 54, no. 5, pp. 357-63, Sep-Oct, 2009.

[8] D. Schottenfeld, and J. F. Fraumeni, "Cancer epidemiology and prevention," Oxford University Press, 2006, pp. xviii, 1392 p.

[9] H. Ma, L. Bernstein, M. C. Pike, and G. Ursin, "Reproductive factors and breast cancer risk according to joint estrogen and progesterone receptor status: a meta-analysis of epidemiological studies," *Breast Cancer Res,* vol. 8, no. 4, pp. R43, 2006.

[10] L. Bernstein, "Epidemiology of endocrine-related risk factors for breast cancer," *J Mammary Gland Biol Neoplasia,* vol. 7, no. 1, pp. 3-15, Jan, 2002.

[11] K. Polyak, "Pregnancy and breast cancer: the other side of the coin," *Cancer Cell,* vol. 9, no. 3, pp. 151-3, Mar, 2006.

[12] E. Vinatier, B. Merlot, E. Poncelet, P. Collinet, and D. Vinatier, "Breast cancer during pregnancy," *Eur J Obstet Gynecol Reprod Biol,* vol. 147, no. 1, pp. 9-14, Nov, 2009.

[13] A. J. Nixon, D. Neuberg, D. F. Hayes, R. Gelman, J. L. Connolly, S. Schnitt, A. Abner, A. Recht, F. Vicini, and J. R. Harris, "Relationship of patient age to pathologic features of the tumor and prognosis for patients with stage I or II breast cancer," *J Clin Oncol,* vol. 12, no. 5, pp. 888-94, May, 1994.

[14] S. Loibl, G. von Minckwitz, K. Gwyn, P. Ellis, J. U. Blohmer, B. Schlegelberger, M. Keller, S. Harder, R. L. Theriault, D. Crivellari, T. Klingebiel, F. Louwen, and M. Kaufmann, "Breast carcinoma during pregnancy. International recommendations from an expert meeting," *Cancer,* vol. 106, no. 2, pp. 237-46, Jan 15, 2006.

[15] H. B. Kal, and H. Struikmans, "Breast carcinoma during pregnancy. International recommendations from an expert meeting," *Cancer,* vol. 107, no. 4, pp. 882-3; author reply 883, Aug 15, 2006.

[16] A. L. Johansson, T. M. Andersson, C. C. Hsieh, K. Jirstrom, P. Dickman, S. Cnattingius, and M. Lambe, "Stage at diagnosis and mortality in women with pregnancy-associated breast cancer (PABC)," *Breast Cancer Res Treat,* vol. 139, no. 1, pp. 183-92, May, 2013.

[17] D. M. Harvell, J. Kim, J. O'Brien, A. C. Tan, V. F. Borges, P. Schedin, B. M. Jacobsen, and K. B. Horwitz, "Genomic signatures of pregnancy-associated breast cancer epithelia and stroma and their regulation by estrogens and progesterone," *Horm Cancer,* vol. 4, no. 3, pp. 140-53, Jun, 2013.

[18] P. H. Westfall, "Multiple Testing of General Contrasts Using Logical Constraints and Correlations," *Journal of the American Statistical Association,* vol. 92, no. 437, pp. 299-306, 1997.

[19] F. Bretz, T. Hothorn, and P. H. Westfall, *Multiple comparisons using R,* Boca Raton, FL: CRC Press, 2011.

[20] P. H. Westfall, R. D. Tobias, R. D. Wolfinger, and SAS Institute., *Multiple comparisons and multiple tests using SAS,* 2nd ed., Cary, N.C.: SAS Institute, 2011.

[21] M. J. Crawley, "Statistics An Introduction Using R," Wiley,, 2014, p. 1 online resource (357 p.).

[22] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, "Cluster analysis and display of genome-wide expression patterns," *Proc Natl Acad Sci U S A,* vol. 95, no. 25, pp. 14863-8, Dec 8, 1998.

[23] W. Huang da, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nat Protoc,* vol. 4, no. 1, pp. 44-57, 2009.