Intl. J. Bioinformatics and Biological Sci.: (V. 6 n.1, p. 01-06): June 2018

DOI: 10.30954/2319-5169.01.2018.1

Identification of Therapeutic Targets and Biomarker for Breast Cancer Using Microarray Datamining

Joe Arun Raja P[1]. and C. Nelson Kennedy Babu[2]

[1]Research Scholar, Manonmaniam Sundaranar Univerisity, Tirunelveli, Tamil Nadu, India
[2]Department of Computer Science & Engineering, Dhanalakshmi Srinivasan College of Engineering, Coimbatore, India

Corresponding author: joearunraja@gmail.com / cnkbabu@gmail.com

Abstract

Neuropilin2 is a family of receptor protein involved in endocrine development in Breast and ducts. The bHLH regulates transcription of Neuropilin2 (NRP2) of progenitor ductal and endocrine cells to a neurogenesis and neurogen specific transcriptional mechanisms that mediates different signaling pathways. The differentiated genes such as RAB40A, FGFR1, TPM1, NRP2, and CLMN genes regulating the development of islet of endocrine and ductal cells, but the molecular mechanism and classification of gene expression is remain unknown. There are several transcriptional gene mutations may regulate transcription of islet cells of ductal and endocrine regions of the pancreas and intestine that may lead to cancer. However, our knowledge of microarray data analysis methods helps to classify the genes associated with differential and undifferential endocrine lineage, ductal cell and exocrine regions determine neurogenesis and neuron specific signaling pathways. Using Meta analysis of statistical rank correlation algorithm to rank the genes based on gene signatures. The reveal predicts 154 (38%) genes that were consistently and significantly up regulated 247 (62%) were down regulated in Breast cancer. Using functional annotation of gene clusters reveals only 47 genes is presumably associated with neuronal development and cell differentiation. Furthermore, this experiment helps to understand candidate genes for novel biomarkers identification, diagnosis and therapeutic approaches to Breast cancer.

Keywords: Breast cancer, NRP2, gene expression, meta analysis, microarray, neuropilin, neuronal development, biomarker

Breast cancer (BC) is the fourth leading cause of cancer deaths[1] among women, being responsible for 6% of all cancer-related deaths and it is very difficult to diagnose in its early stages. At the time of diagnosis, 26% have survival rate, past 5 years only 21.5% of survival rate has been recorded[2]. With non-specific symptoms 90% of patients are diagnosed for surgery in advanced stage. Prior to this study, nearly 95% of BC is associated with hereditary factors that may greater risk of BC[3]. The deadly nature of BC stems from its tendency to rapidly spread to the lymphatic system and distant organs. There are various types of Breast cancer include 78% are ductal carcinomas of the ductal epithelium. Only 5% of tumors of the breast are benign.

In most patients, the BC is observed in both exocrine and endocrine regions, however the biological phenomenon shows the exocrine region release insulin-producing beta cells during embryogensis and endocrine region regulates glucose tolerance[4]. The pathological association of diabetes in exocrine region and Breast cancer is in the endocrine region[5,6]. During the development of progenitor cells, which express specific genes *Foxa1, Foxa2, Pdx1, Pbx1, Hes1, Ptf1a, Ngn3, HNF6, Pax4, NeuroD1, Nkx2.2* and *Sox9*, give escalate to express in both the ductal, exocrine and endocrine Breast cell development[7]. In cell division, the inactivation of genes that may transcriptionally regulate embryogenesis of neuronal cells it may leads to cancer[8]. The Neuropilin2 (NRP2) controls a complex of gene regulatory transcriptional networks in endocrine progenitor cells which regulate insulin secreting islets of Langerhans[9]. The relationship of exocrine and endocrine development of multi potent intracellular genes that may limit the expression during precursors of ductal polypeptide producing cells.

The experimental results suggest that there are other transcriptional signaling pathways play important regulatory roles during BC development and the molecular mechanism are unknown. There are various techniques to comprehend the mechanism of molecular and genes resoponsibility for intra cellular signaling pathways. Using gene expression program determines the characterization of NRP2-mediated duct and endocrine cell mediated signaling pathways. The studies generated large set of adenoviruses expressing NRP1 and NRP2 of genome wide mRNa profiling data is revealed the expression analysis to reprogram ngn3 expression on Breast cancer gene analysis[11].

The objective of the study is to understand the disease etiology of novel candidate gene expression in disease progression, cell cycle, neuronal development and cell differentiation. We used gene expression dataset to understand the characteristic NRP-2 mediated mRNA profiling within duct, exocrine and endocrine cell functional genes that may involved in BC and diabetes mellitus. We evaluate the transcriptional regulation of genes that may involved in expression with β-islets within endoderm, that may helps to synthesize insulin, neuronal development mediated gene signaling networks that may helps to synthesize neuronal cells to form interconnection between mammary alveoli and ducts.

In these we temporarily targets NRP-2 induced Breast with β-islets within endocrine cells, the genes that differentially expressed in both drug targets and disease conditions are typically predicts for molecular biomarkers. The biomarkers helps to identify the genes associated with neuronal cell development of pancreas along with other types of diseases also.

MATERIALS AND METHODS

Disease dataset selection

The data used in this study is publicly available from Gene expression Ominibus (GEO) database (GEO dataset: GSE67301)[11]. The dataset contains 37 breast tissue samples were collected under ultrasound guidance from patients with stage three breast cancer before four cycles of whole breast hyperthermia. Gene expression analysis was done using Affymetrix U133 Plus 2.0 GeneChip arrays.

The advantage of using this data is to understand the expression of NRP2 mediated human duct cells is recombinant with adenovirus and the induction of NRP2 expressions with different time intervals. The different time intervals are helps to predict the differential expression of NRP2 in duct, endocrine and exocrine cell differentiation in breast.

Normalization of gene expression dataset

The dataset is from a gene expression levels of Breast cancer samples from homosapiens, the details of study is given in Cao *et al.* (2012). The dataset contains 7 controls of human ductal cells is transfected with AdGFP control vector biological samples and 7 treated human duct cell after transduction with Ad-NRP2-GFP vectors in 3, 14 and 20 days time series. The samples is labeled with Cy3 (green dye) of control and Cy5 (red dye) of treated samples is hybridized with Affymetrix HG133A samples. However, the preprocessing and normalization procedures is implemented using Bioconductor packages [12] to remove systematic variance and prepare datasets for further analysis.

The preprocessing of raw data using bioconductor tools to make over statistical metrics to predict quality of outlier data. The GCRMA (Gene-Chip Robust Multiarray Average)[13] algorithm include noise and non-specific binding (NSB) data calculation to optimize background intensities that adjust probe intensities to expression measurements and the normalized data is summarized using Robust Microarray Analysis (RMA) algorithm[14]. The raw signal intensities of randomly retained datasets are in position of the PM and MM for every probe pairs, the MM probes introduces more background noise on a raw intensity scale to correct for NSB. The log2 transformation of two-color arrays convert data to a linear scale value of each background corrected PM probe is obtained and these values are normalized using Q function. The RMA and GCRMA packages helps for only PM signals of background corrections. Thus the normalization techniques used only for conjunction with (*pmonly)* PM correction method. We used Li-Wong procedure to normalize arrays using invariant set of genes and then fit probe set data to parametric model[15].

Differential gene expression of NRP2 in Breast data

After normalization and preprocessing the resultant data is used for statistical analysis to predict the differential gene expression. Clustering algorithms such as *hierarchical*[16] and *K-means* clustering[17] or *self-organizing maps (SOM)*[18] used to generate partial solutions of single factors. Hiearchical clustering specifically describes the differential gene expression based on distance matrix are connected by a series of branches (clustering tree or dendrogram). This method helps to classify gene expression analysis of complete linkage outforms. K-means clustering algorithm predicts the best fitting between clusters and their representation using a predefined number of clusters. The prototype of randomly selected datasets of lowest dissimilarity and represents smallest cluster variance.

Evaluation of functional Annotation

The clustering data of functional data is used for GO annotation of the genes extracted from Affymetrix HG133A annotation file. The Database for Annotation, Visualization and Integrated Discovery (DAVID) Gene Functional Classification Tool (http://david.abcc.ncifcrf.gov)[19] used to identify list of genes associated with biological terms into organized classes. The organizes of significant genes shared specific terms by <5% of the genes paired with common GO terms that functionally related. Genome enrichment analysis is predicted using Gorilla (http://cbl-gorilla.cs.technion.ac.il/)[20] to get gene rank list.

RESULTS AND DISCUSSION

1. Experimental analysis of NRP2 over expression in Breast cell lines

How NRP2 is expressed differentially in both exocrine, endocrine and ductal cells. We performed an expression analysis of 14 datasets of NRP2 mediated reprogrammed datasets is annotated with Affymetrix Human Genome U133A of adult Breast duct cells of expression datasets. The isolates of NRP2 deficient Breast duct cells that are abundant expression in islet endocrine and neural tissues. There are 14 experimental datasets of control vectors of mRNA transcriptional gene factors data of 3, 14 and 20 days is compared with NRP2 expression of NRP2 cell cultures of datasets from GEO database. The transcripts of differentially regulated NRP2 are unpaired with P<0.05 of up and down regulations is calculated using R and BioConductor.

The differentially expressed dataset contains 22285 genes, the preprocessing and normalization is carried out using simple affy statistical software package to calculate thresholds of raw data. The normalization of expression values is calculated using RMA is close with GCRMA of standard processing of expression values, the differential gene expression of dataset is treated with NRP2 expression array of treated datasets is presented in principal component analysis (PCA) of 3, 14 and 20 days datasets from both early and late stage of gene expression (Fig. 1). The incremental analysis of repeated genes across independent datasets is biologically active that filtering out differentially.

We determine the differential expressions of Breast cancer genes were independently classified. Using fold change (FC) cutoff of 14 independent datasets and a FDR-corrected P-values (P<0.05) of 15410 genes were filtered from background corrections. Using functional enrichment and gene ontology to classified 11721 genes were differentially expressed in overall dataset. The Pearson's Correlation coefficient of hierarchical clustering shows the differentially expressed genes is aligned based on Euclidean distance. The highly significant Euclidean distance with poor correlations did not show statistically significance with correlations of series of outcomes (Fig. 2). A set of 400 differentially expressed genes of significant (P <0.05) thresholds of well defined clusters is used for linkage analysis.

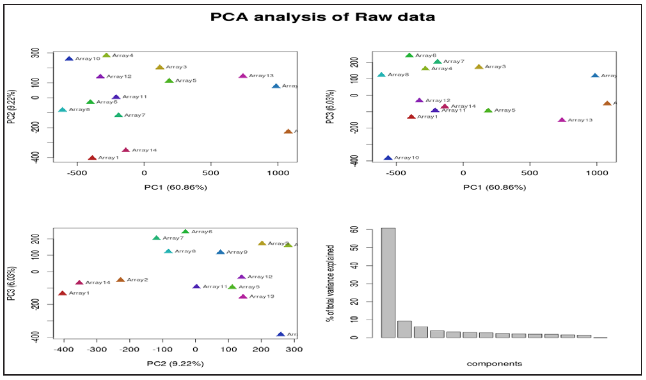

Fig. 1: Principle component analysis of 14 datasets including NRP2 data analysis

Identification of upregulated genes

We used geNetClassifier to classify significant genes of differential expression at <0.95 threshold shows 182 up regulated, 229 down regulated. The initial screening is carried out with all pairwise z-vlues associated with significant correlation of up regulated and down regulated datasets. Using false positive prediction of 46 gene ranks were up regulated at two folds, 65 genes are also significantly expressed in precisely explained Breast cancer (Fig: 3a, 3b). Supplementary Table 1 provides the up regulated and down regulated genes have well established and significantly associated with lack of neuronal cell development in Breast cell development. Some well known genes include RAB40A,FGFR1,TPM1,NRP2, and CLMN genes is down regulated in Breast duct, endocrine and exocrine cells and is less effective to synthesize neuronal cell development and differentiation into insulin production. Hence, these genes is potential targets to Breast cancer. RAB40A,FGFR1,TPM1,NRP2, and CLMN genes is significantly down regulated genes in Breast cancer on neuronal cell development, but these genes directly associated with other types of cancer such as pancreatic cancer, ovarian cancer and other cancer types.

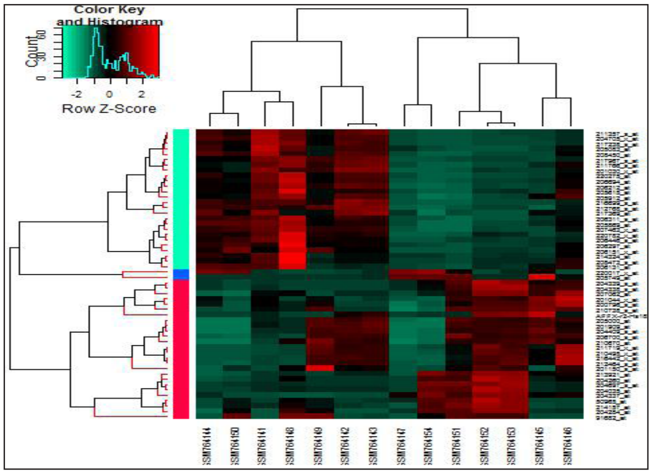

Fig. 2: Hierchical clustering of 400 differentially expressed significant genes predicted with p<0.05 threshold

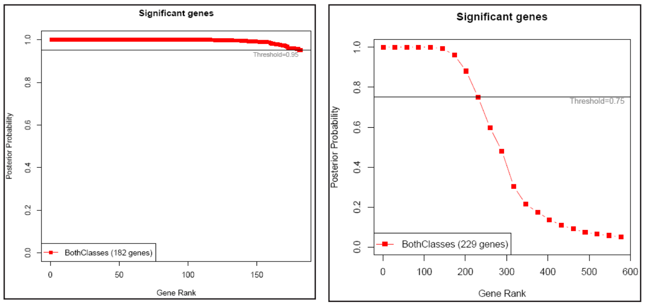

Fig. 3a: Differential expression of case and control dataset genes were predicted with significance of <0.05 threshold

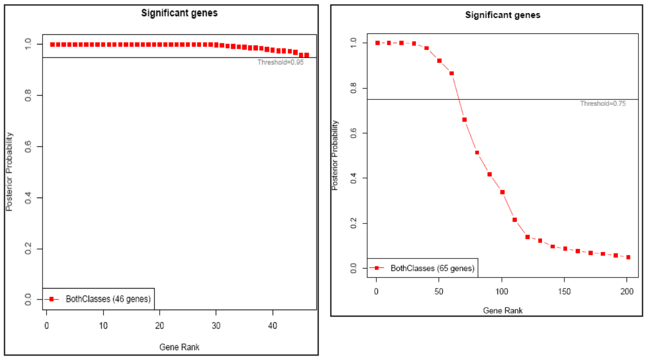

Fig. 3b: Differential expression of both commonly associated biomarker genes is predicted with significance of <0.05 threshold

Functional analysis of NRP2 associated significant genes

We identified the functions of 46 significant genes of both up regulated and down regulated genes were analyzed using GOrilla online database. The twelve up regulated genes are functionally involved in cell-differentiation, cell development and disease progression with in Breast cells. The dysregulation of RAB40A, FGFR1, TPM1, NRP2, CLMN of p-value 2.46E-5 is significantly involved in different pathways with in cells of breast.

CONCLUSION

We have predicted the gene expression of NRP2 gene regulation in Breast neuronal cell development that leads to cancer. The ductal, exocrine and endocrine cells of neurogenesis of expressed datasets predicts 46 genes of up regulated and down regulated that potentially signified in Breast cancer and also associated with other significant cancer types. Based on this analysis the genes RAB40A, FGFR1, TPM1, NRP2, and CLMN are down regulated in Breast cells and also with different cancer types is potentially predicted best biomarkers to understand Breast cancer and associated cell types. Our results are helps to predict the gene regulation networks of other significant gene regulatory networks is helps to predict best drug targets and is used for drug discovery.

REFERENCES

1. Ulirsch J, Fan C, Knafl G, Wu MJ *et al.* Vimentin DNA methylation predicts survival in breast cancer. *Breast Cancer Res Treat*, **137**(2): 383-96.

2. Howlader N, Noone AM, Krapcho M, Neyman N, Aminou R, *et al.* 2011. SEER Cancer Statistics Review, 1975–2009.

3. Esserman LJ, Berry DA, Cheang MC, Yau C *et al.* Chemotherapy response and recurrence-free survival in neoadjuvant breast cancer depends on biomarker profiles: results from the I-SPY 1TRIAL (CALGB 150007/150012; ACRIN 6657). *Breast Cancer Res Treat*, **132**(3): 1049-62.

4. Solar, M., Cardalda, C., Houbracken, I., Martin, M., Maestro, M.A., De Medts, N., Xu, X., Grau, V., Heimberg, H., Bouwens, L., Ferrer, J., (2009). Pancreatic exocrine duct cells give rise to insulin-producing beta cells during embryogenesis but not after birth. *Dev. Cell*, **17**: 849–860.

5. D'Amour, K.A., Agulnick, A.D., Eliazer, S., Kelly, O.G., Kroon, E., Baetge, E.E. 2005. Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nat. Biotechnol.*, **23**: 1534–1541.

6. Duvillie, B., Attali, M., Bounacer, A., Ravassard, P., Basmaciogullari, A., Scharfmann, R. 2006. The mesenchyme controls the timing of pancreatic beta-cell differentiation. *Diabetes*, **55**: 582–589.

7. Prashantha Nagaraja, Kavya Parashivamurthy, Nandini Sidnal, Siddappa Mali, Dakshyani Nagaraja, and Sivarami Reddy 2013. Analysis of gene expression on ngn3 gene signaling pathway in endocrine pancreatic cancer, *Bioinformation*, **9**(14): 739–747.

8. Ramiya, V. K., Maraist, M., Arfors, K. E., Schatz, D. A., Peck, A. B. and Cornelius, J. G. 2000. Reversal of insulin-dependent diabetes using islets generated *in vitro* from pancreatic stem cells. *Nat. Med.*, **6**: 278-282.

9. Gradwohl, G., Dierich, A., LeMeur, M. and Guillemot, F. 2000. neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. *Proc. Natl. Acad. Sci. USA* **97**: 1607-1611.

10. O. Karlsson, S. Thor, T. Norberg, H. Ohlsson, T. Edlund 1990. Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and a Cys-His domain Nature, 344 (1990), pp. 879–882.

11. Farrell AS, Pelz C, Wang X, Daniel CJ *et al. Pin1* regulates the dynamics of c-Myc DNA binding to facilitate target gene regulation and oncogenesis. *Mol. Cell Biol.*, **33**(15): 2930-49.

12. Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J. Y., and Zhang, J. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**: R80.

13. Jean(ZHIJIN) Wu, Rafael Irizarry James MacDonald 2014. Background Adjustment Using Sequence Information.

14. Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., Speed, T. P. 2003. Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data. Accepted for publication in Biostatistics.

15. Li C, Wong WH. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc. Natl Acad. Sci.*, USA **98**: 31-36.

16 Hartigan JA. Clustering algorithms. New York: John Wiley & Sons, 1975: 351pp.

17. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM. 1999. Systematic determination of genetic network architecture. *Nat Genet.*, **22**: 281–5.

18. Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, et al. Interpreting patterns of gene expression with selforganizing maps: methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci., USA* **96**: 2907–12.

19. Da Wei Huang, Brad T Sherman, Qina Tan,1 Jack R Collins, W Gregory Alvord, Jean Roayaei, Robert Stephens, Michael W Baseler, H Clifford Lane, and Richard A Lempicki 2007. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Research, Vol. 35, Web Server issue W169–W175.

20. Eran Eden*, Roy Navon*, Israel Steinfeld, Doron Lipson and Zohar Yakhini. 2009. “GOrilla: A Tool For Discovery And Visualization of Enriched GO Terms in Ranked Gene Lists”, *BMC Bioinformatics*, **10**: 48.