CTCBC – Recognition of Tumor Cells in Breast Cancer by using Clustering Methods

S.Mythili,
Research Scholar,
Department of Computer Applications
Hindusthan College of Arts & Science,
Coimbatore

A.V.Senthil Kumar, Ph.D.
Director,
Deparment of Computer Applications
Hindusthan College of Arts & Science,
Coimbatore

ABSTRACT:
Breast cancer is a highly disease which the women's are mostly affected. The main cause of womens death is not only by the tumor cells that affects the Breast Cancer but its metastases at different sites, such as lumph nodes and other organs (i.e lung, liver and bones). Identifying the circulating tumor cells in the blood that results from tumor cell invasion and intravascular filtration highlights its role which concerns the tumor cells aggressiveness and metastasis. Biological research regarding CTC's monitoring for Breast Cancer is limited due to the cause of indicative genes for their detection and isolation of genes . By using the direct CTC detection , we focus on the identification of factors in peripheral blood that can be indirectly reveals the presence of tumor cells. By selecting the publicly available Breast Cancer tissues and peripheral blood microarray datasets. In this 2 steps are to be followed by eliminating the procedures for the identification of several discrimant factors. The new algorithm CTCBC is proposed to identify the BC in an earlier stage. This procedure provides the facilities of identifying the major genes which involved in the causes of Breast cancer.

Keywords
Breast cancer (BC) signature, Circulating tumor cells (CTC), Peripheral Blood (PB), Biological processes, gene Elimination.

1. INTRODUCTION:
Breast cancer (BC) affects the western side women's in large ratio. Tumor cells are the main cause of cancer which affects both the men and women in lungs, liver or bones which causes the death for the patients, so there is a need to improve the diagnostic tools in clinical procedure[4]. CTC in PB are the important factor which diagnosis the arise of the cancer in tumor cells. In this the tumor aggressiveness and metastasis DNA microarrays techniques are used for high- dimensional gene expression datasets [3]. This gene expression with CTC will provide the opportunity to identify the markers and diagnosis the BC patients for the treatment of the BC.

Several microarray studies on BC tissues samples are given in the process of gene deformations [7]. These similar gene deformations appear in the analyzed portion of PB. BC specific alterations can be identified in affected tissues cells in the blood some specific alternations in the blood samples will indicate the cancer in the cells with the existence of CTC's [8].

2. EXISTING WORK:
The hypothesis supports the identification of cancer tissues, the cancer blood indicates the ability of tumor to diffuse and can be used as factors for CTC estimation without the proper detection [16]. In this two –stage are used several procedure applied on available DNA microarray datasets from the different tissues and blood samples[2]. The first stage aims in extracting gene signatures associated with pairwise differentiation between cell types and the disease states[1]. The comparison of cancer and tissues provides the information about the disease. Next we compare the cancer blood and the peripheral blood which can derive the alterations that relates the CTC content at the beginning stage[15]. There are 3 stages which we drive the BC signature. In all these stages, the cancer blood is compared with the genes and the final comparison result is obtained. C1, C2 & C3 specifies the comparison of the cancer blood with the PB cells.

C1$\rightarrow$ it compares the cancer tissues with the normal tissue to identify whether it is a primary cancer.

C2$\rightarrow$ It compares the BC Patients PB to the normal tissues and results the level of BC in the patient blood.

C3$\rightarrow$ It compares normal PB with the cancer tissues and identifies genes in the PB cells.

Here, the cancer patients PB information is passed out tumor cells status is found.

3. PROPOSED WORK:
The hypothesis of the PB from the cancer patients is retrieve and it process whether the tumor cells are in primary or secondary stage. While comparing the C1, C2, and C3 results the closer BC signature is not to be denoted at the certain level of cancer aspects, which does not leads to the closer aspects of CTC's.

For this, we proposed a procedure, which consider the intersection of the BC signature with the tumor cells. We consider the following intersections.

1) C1 n C2 : it derives the genes over expressed in normal blood with the cancer tissues.

2) C1 n C2 n C3 : it denotes the genes in blood which has more cancer blood tissues.

The intersection of C1, C2 & C3 will denote the presence of cancer tissues. When the tumor cells are found then it can be directly compared with the factors used for isolation of the PB cells. Here C3 is based on the specific differences in the expression of genes which are used in the biological studies. The C3 aims to educates the cross tissues for the control of BC. C1 performs the action which is similar to C3 but the new information about the BC cannot be provided.C1 performs the level of cancer tissues in the blood and c3 compares the level of cancer cells in the blood. The first step of C1 n c2 identifies the level of cancer tissues in the blood control. The second

step of C1 n c2 n c3 shows the overall cancer tissues in the blood cells. We have used 2 models to study about the tumor cells in the breast cancer. We have used 2 models to study about the tumor cells in the breast cancer. Progression model and metastatic predestination mode, this provides the clear state of CTC's. In this we applied the two stages methodological which is used to drive a panel of genes used in tumor cells in the BC.

The analysis of the intersection signature might be used in terms of new process and a technical pathway, which may be useful for diagnosing the BC in the clinical lab. We have designed a data integration methods that combines different datasets is a single platform which is known as meta-analysis. Datasets and their data integration is designed by the heterogenetic datasets in the means of gene transcripts used and also the biological representation. Here, the dataset Integration approach performs 3 comparisons in a step by step process.

1. The dataset is downloaded as GEO Format. The raw data is not taken for the process, the data is pre-processed and it performs K-Nearest Neighbour algorithm, then the data is transformed for the processing.
2. The dataset is identified by the platform called EGIC enterz gene identifiers. This platform is used for mapping specific annotations in the Bioconductors libraries.
3. The mapping can be resulted in different probes that are same in Entrez gone id which keeps the expression values in a large set of data's .
4. Then it performs the study of each dataset separately by performing the PB tissues of the cancer patients.
5. Finally all the merged tissues are compared with the C1,C2,& C3 which shows the result of the sample datasets. The final step is the selection of those that exhibit one-direction over expression.

4. DATASETS:

We have used 7 different datasets that are publicly available from the Gene expression Omnibus (GEO) database which is taken with their specific characters [12]. The samples may the normal tissues or the cancer tissues. In the process of C1 the healthy samples are compared with the cancer patient's tissues [9]. So that we may have the data integration method which combines the different database in a single platform with the multiple number of analysis. It proposed the genes that show significant regulation in the comparison of C1, C2 & C3. These genes are already used in the dataset by focusing on the discrimination of CTC's.
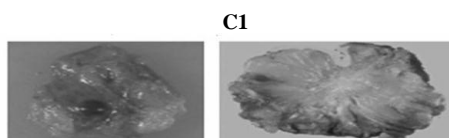
C1

Fig:1 (a)

Fig: 1 (a) represents the Cluster1, the first picture is the Normal Breast Tissue and the Second picture is Breast Cancer Tissue.
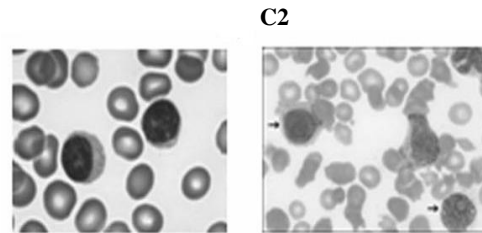
C2

Fig:1 (b)

Fig: 1 (b) Represents the Cluster2, the first picture is Non- cancerous Peripheral Blood and the second picture is Cancerous Peripheral Blood.
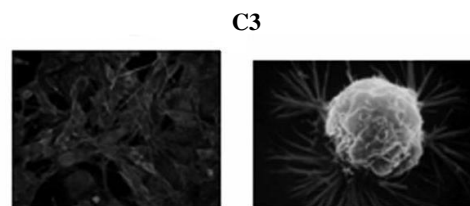
C3

Fig:1 (c)

Fig: 1(c) represents the Cluster3, the first picture is Normal Peripheral Blood and the Breast Caner Tissues.

The comparisons of C1,C2,& C3 are made by study of single datasets, in the case of C1 the healthy tissue samples are compared with the cancer patients tissues where the number of healthy tissues are large. Here we have proposed a data integrity method which combines the different data sets in a single platform or a multiplatform. We have approached 4 steps for the performance of comparing C1, C2 & C3.

1) The dataset related to the BC tissues and healthy tissues are collected, the raw data is not taken it is performed with the K-nearest Neighbor algorithm and it is log-transform the BC tissues.

2) A specific platform is allotted for a each dataset, which has the specific Bio conductor libraries used for GEO platform. We provide the expression values for the largest variation as given by the in range value.

3) We proposed an algorithm CTCBC, in which the method proceeds in each comparing by performing a cross study of datasets, batch correction of sample tissues and cross platform normalization.

4) The final step is the merging of datasets, for each comparison the subsequently used tissues are used for finding the different genes by selecting the samples from the original datasets. This datasets are taken for the cluster comparison.

Table 1 represents the number of tissues used for the Cluster comparison

Origin	Number of Cancer Tissues	Number of Healthy Tissues	Comparison
Blood Samples	158	14	C1,C2
Tissues	57	27	C1,C3
Tissues	68	11	C2,C3
Tissues	135	32	C1,C2,C3
Blood Samples	39	7	C2,C3
Tissues	82	17	C1,C2,C3

5. RESULT:

By this proposed work, a new algorithm CTCBC is obtained by comparing the specific sets of genes which exhibit differentially over expressed behaviour. It is used to find the number and reveal the biological relevance of genes belonging to the different sets, we find 35 common genes in the C1 n c2 n c3 and 24 genes in C1 n c2 . To validate our methodology, we use hierarchical clustering on the first datasets and we find that 35 genes are identical. We consider the difference between the metastatic populations to access the group of samples.

We identify the factor in PB that can indirectly reveal the CTC's , instead of detecting the CTC's this shows in the intersections of C1, C2, C3 in which the genes are over expressed in normal blood are eliminated and the remaining genes are over expressed in blood due to the cancer associated factors. Classification of 3 genes are conducted (i.e C1 n c2, C2 n c3, C1 n c3), Finally all the 3 genes comparison is made (C1 n c2 n c3). The Biological result shows the intersections of the 3 cluster, in all the 3 cluster the genes are taken with BC Tissues. The intersection of C1 n c2 = 35% , C2 n c3 = 28% , C1 n c3 = 34% according to the dataset used with BC tissues. Final result of all the 3 cluster intersection is made , it shows 24% of the genes are common so that the Breast Cancer tissues are identified easily.
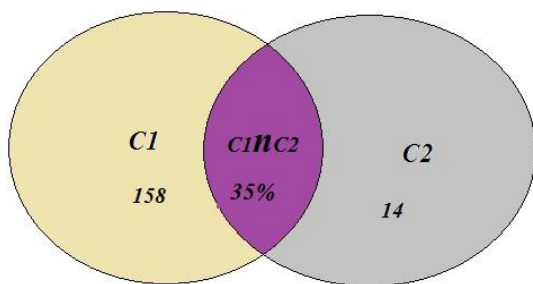

Fig: 2(a)

Fig: 2(a) represents the comparison of cluster C1 & C2. The C1 & C2 values are given with the count of number of genes. The intersection of C1 n c2 result is obtained as 35%.
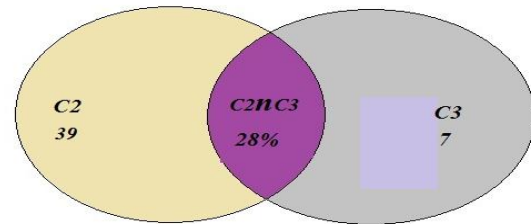

Fig: 2(b)

Fig: 2(b) represents the comparison of cluster C2 & C3. The C2 & C3 Values are given with the count of the number of genes. The intersection of C2 n c3 result is obtained as 28%.
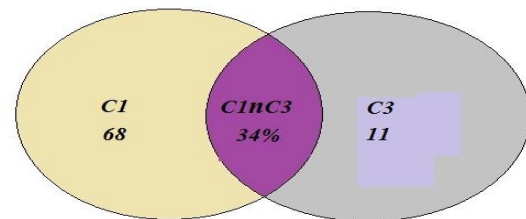

Fig: 2(c)

Fig: 2(c) represents the comparison of cluster C1 & C3. The C1 & C3 Values are given with the count of number of genes. The intersection of C1 n c3 result is obtained as 34%.
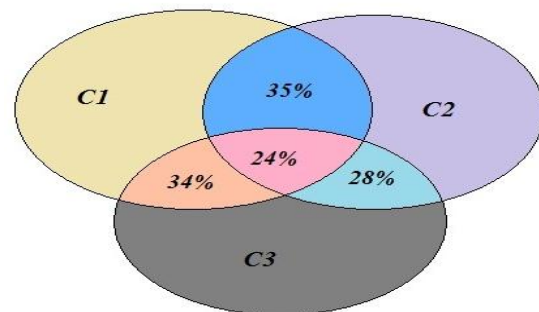

Fig : 2(d)

Fig : 2(d) represents the comparison of cluster C1,C2 & C3. The intersections of C1 n c2 n c3 final results are obtained. This result produces the common BC Tissues in the provided datasets.

6. CONCLUSION:

The identification of cancer tissues is most important for the treatment of cancer patients. The PB tissues are calculated in small number of disseminating cells which causes the BC in CTC's. Here we have given an approach for the multistudy to explore the tissues in the combination of microarray gene expression data. The intersection of C1,C2 & C3 results are used for identifying the CTC earlier. Finally the comparison we perform for the derivation of signatures C1, C2 &C3 are more or less similar to another study. The final dataset can be possibly contain a less measured gene transcripts. Then the cross study normalization of the expression values can be merged datasets by eliminating the different variants. The result produces have some of the important genes that can be commonly used for CTC recognition. Further we can examine the number of genes in terms of the cross study and their normalization. In this Hierarchical Clustering is also used to cluster the number of genes for each cluster.

7. REFERENCES:

[1] D.C.danila, K.Pantel, M.fleisher, and H.I.scher " circulating tumor cells as biomarkers, " Cancer J.vol.17, no 6 pp .438-450, 2011.

[2] Stellins Sfakianakis, Ekaterini S,Bei, Michalis Zervakis , Member, IEEE, despoina Vassou and dimitrios Kafetzopulos , "On the identification of Circulating Tumor cells in Breast cancer " IEEe Journal of Biomedical and Halth informatics Vol18, No.3,May 2014.

[3] T.J.Molloy, P.roepman, B.Naume, and L.J.V.Veer, " A prognostic gene expression profile that predicts circulating tumor cell presence in breast cancer patients , " plos One, vol.7, no.2, p.e32426, Feb 2012.

[4] M.Kandula, Ch.K.Kumar, k.Ravi Kanth, V.V.Laxmi Addala, S.Murthy, and Y.S.

[5] G.Shin T.W. Kang,S.Yang,S.J. Back,Y,-S jeong, and S.Y,Kim."GENT: Gene expression database of normal and tumor tissues,"Cancer Inf, vol,10,pp.149-157,2011.

[6] R-Z. Liu, K.Graham,D.D.Glubreeht,D.R.German,J.R.Mackey, and B.Godbout," Association of FAB P% expression with poor survival in triple- negative breast cancer," Amer .J.Pathology, vol.178.no.3.pp.997-1008,mar.2011.

[7] E.Enerly,I.Steinfed. K.Kleivi, S-k Leivonen,M.R, Aure, H.G. Russnes. J.A.Ronnenberg, h.G.Russnes,J.A.Ronneberg,H.Johnsen,R.Navon,E. Rodland,R.Makela,B.Naume,M.Perala,O.Kallioneie mi.V.N.Kristensen,Z.Yakhini,And A.L.Berresen-Dale."miRNA-mRNA integrated analysis reveals roles for miKNAs in primary breast tumors ," plos One,Vol 6no.2, P.e16915, Feb 2011.

[8] H.G.LaBreche, J.R.Nevins,and E.Huang, " Integrating factor analysis and a transgenic mouse to reveal a peripheral blood predictor of breast tumors, " BMC Med.Genomics, Vol.4.p.61,2011.

[9] I.Shabo,H.Olsson, O.Stal, and J.Svanvik, " Breast cancer expression of DAP12 is associated with skeletal and liver metastases and poor survival", Clin.Breast Cancer, vol.13,pp 371-377, Oct 2013.

[10] V.Ouellet, K.Tiedemann, A.Mourskaia, J.e.fong, D.Tran-Thanh, e.Amir, M.Clemons,B.Perbal, S.V.Komarova, and P.M.Siegel, " CNN Impairs osteoblast and stimulates oseteoclast differentiation to favour breast cancer metastasis to bone," Amer.J.Pathol, Vol.178,no,5,Pp.2377-2388, May 2011.

[11] G.Tsiliki, M.Zervakis, M.Ioannou, E.Sanidas, E.Stathopoulus, G.Potamias, M.Tsiknakis, and D.Kafetzopoulus, " Multi-platform data intergration in microarray analysis, " IEEE Transactions.Ing.Technol.Biomed, Vol.15, no.6,pp.806-812, Nov 2011.

[12] A.Turnbull, R.Kitchen, A.Larionov, L.Renshaw, J.Dixon, and A.Sims, " Direct integration of intensity level data from Affymetrix and ILlumina microarray improves statistical power for robust reanalyis, " BMC Med. Genomics, Vol.5, no.1, p.35, 2012.

[13] D.Hanahan and R. Weinberg, " The hallmarks of cancer, " Cells , Vol.100, no.1, pp.57-7, 2000.

[14] F.Lozy and V.Karantza, " Autophagy and cancer cell metabolism, " Semin cell Dev.Biol, Vol.23, no4, pp 395- 401, Jun 2012.

[15] J. H. Choi, N.R. Shin, H.J. Moon, C.H.. Kwon, G. H. Kim, G. A. Song, T.Y. Jeon, D. H. Kim, D.Y.. Park, " Identification of S100A8 and S100a9 as negative regulators for lymph node metastasis of gastric adenocarcinoma, " Histol. Histopathol, vol.27, no. 11, pp 1439 – 1448, Nov . 2012.

[16] J. H. Hung, " Gene Set/ Pathway enrichment anlaysis, " Mehtods Mol. Biol. Vol.939, pp. 201-213, 2013.

[17] Y.Park, T.Kitahara T.Urita. Y.Yoshida, and R.Kato, " Expected Clinical applications of Circulating tumor cells in Breast cancer, " Worls J.Clin Oncol, vol 2 , no.8, pp. 303 – 310, 2011.

[18] S. Riethdorf and K.Pantel, " Advancing personalized cancer therapy by detection and characterization of circulating carcinoma cells , " Ann.New york Acad. Sci, Vol.1210, no 1, pp 66-77, Oct 2010.

[19] J. Barbaz'an L. Alonso – Alconada L. Muinelo – Romay, M. Vieito , A.Abalo, M. Alonso – Nocelo, S. Candamio, E. Gallardo, B. Fern'andez, I. Abdulkader, M. de Los Angeles Casares, A. Go'mez-Tato, R. Lo'pez- L'opez, and M.Abal, " Molecular characterizaton of circulating tumor cells I human metastatic colorectal cancer, " PloS One, Vol.7, 2012.

[20] F.J. Lopez, M. Cuadros, C. Cano, A. Concha, and A. Blanco, " Biomedical application of fuzzy association rules for identifying breast cancer biomarkers, " Med.Biol. Eng.Comput, Vol.50, no.9, pp. 981- 990, sep 2012.

[21] C.Clarke, S. F.Madden, P.Doolan, S.T. Aherne, H.Joyce, L.O'Discoll, W.M. Gallagher, B.T. Hennessy, M.Moriarty, J.Crown, S.Kennedy, and M.Clynes, " Correlating transcriptional networks to breast cancer survival : A large- scale coexpression analysis," Carcinogenesis, vol. 34, pp 2300- 2308, Jul. 2013.