

## CRITICAL ANALYSIS ON THE USE OF COMPUTATIONAL TOOLS FOR THE GENOMIC ANALYSIS OF ORAL CARCINOMA

Shafana A.R.F., Uwanthika G.A.I.

<sup>1,2</sup> Department of Computer Science and Technology, Uva Wellassa University of Sri Lanka, Badulla, Sri Lanka

*shafana.cst@gmail.com* , *isuriuwanthika@gmail.com*

**ABSTRACT:** *Betel quid chewing, being the main cause of oral cancer in Sri Lanka, has been driving its mortality rate rapidly over the years. The low five-year survival rate has been witnessed since oral cancer is typically detected in advanced stages. With the advancement of medicine and technology, the issues must have been addressed to date, but have not. Thus, there is a paramount need to revise the therapeutic strategies of oral squamous cell carcinoma in order to identify the cancer in its early stages and also to identify the molecular sub-classes so as a mean to individualize the treatment. This research paper mainly reports on several methods and tools that have been used to analyze the gene expression dataset of oral carcinoma in order to identify the differentially expressed genes. Further, several other clustering methods that are used to extract molecular sub-types has also been discussed. An array of literature has been critically reviewed to find the current insights on methods used for the computational genomic analysis in order to identify the potential gap in cancer diagnosis, prognosis or drug response as to support oral cancer mitigation through the identification of molecular sub-types and the stage-specific genes of Oral Cancer with the support of computational biology and bioinformatics.*

**Keywords:** Oral Squamous Cell Carcinoma, Genomic Analysis, Microarray, Biomarkers, Differentially Expressed Genes

### 1. INTRODUCTION

Cancer is a leading genetic disease across the world that result from both inherited and acquired changes in DNA. There are about 100 types of cancer which can affect any part of the body. According to the statistics of 2008, cancers in the breast, oral cavity, lungs, oesophagus, colon & rectum, cervix, thyroid, lymphoma, ovary and leukemia are among the ten most common cancers in Sri Lanka in the years 2001- 2008. Currently, Oral pharyngeal cancer is one which carries highest mortality and morbidity among different types of cancers especially in males.

Oral Pharyngeal Cancer is the malignant growth of body cells in the oral cavity. From the updates of World Health Organization as of 2014, it is stated that oral cancer has occupied eighth place in the leading cause of death across the world. Moreover, it has occupied second place in the cancers that cause death in Sri Lanka. According to National Cancer Institute of Sri Lanka (2013), there are 900 deaths per year due to oral cancer which is approximately 3 deaths per day.

The five-year survival rate (62%) of oral cancer is among the lowest of all the major cancers in humans (Cheng et al., 2014). The main cause of this lower



survival rate is that nearly 90% of the oral pharyngeal cancers are diagnosed only at the advanced stages. Therefore, this creates an urgent need for the detection of oral cancer at an early curable stage.

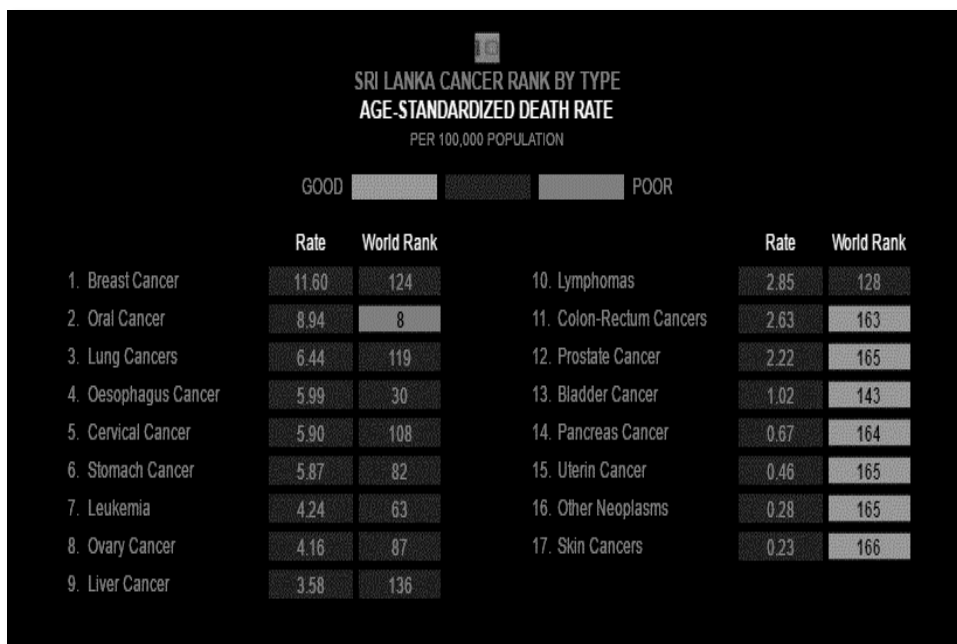


Figure 1. Updates of World Health Organization, 2014

The field that is highly promising for such prognosis is molecular biology (Jurel et al., 2014). Although there are several other technologies available in support of molecular biology, DNA microarray technology is one of the best omics, as it permits profiling of the expression of thousands of genes at once. (Gevaert & Moore, 2009).

However, due to the high dimensionality of the data from DNA microarray technology, it is hard that a clinical practitioner interprets these data directly and identify the unique gene associated with each stages of oral cancers. Thus, this causes a vital need for the intervention of bioinformatics and computational biology in order to analyze the DNA microarray efficiently, such that it can be used to detect Oral cancer at the earliest stage possible in a way to contribute to the medical science as well as to the entire community. This particular paper provides a wide variety of computational techniques which have been utilized in the recent years for the clinical management of cancer and presents a methodology to increase the survival rate and the individualization of the treatment.

## 2. CANCER RELATED RESEARCHES

Shillitoe et al. (2000) developed libraries of cDNA and deposited in GenBank. The comparison was done for pairs of genes in libraries in Cancer Genome Anatomy Project (CGAP) that was listed through a literature review.

According to the researchers, though the data were quite incomplete at the time, they have represented a large quantity of new information and clones of potential utility to the oral cancer community. This gave rise to the development of new methods of analysis.

Vijver et.al (2002) have used complementary DNA (cDNA) microarray tissues of breast cancer patients as a mean to improve the selection of cancer patients for adjuvant therapy. 70 gene-expression prognosis profiles were evaluated through microarray analysis. The predictive power of these profiles were evaluated through uni-variable and multi-variable statistical analysis. Through this research, the researchers concluded that gene-expression profiles were strong predictors of outcome related information of cancer than the conventional histologic and clinical data.

Researchers (Darby, Nettimi, Kodali & Shih, 2005) have used artificial neural network for the prediction of head and neck cancer metastasis by generalizing a pattern for it. The particular research was targeted to predict the secondary regions where cancer could metastasize. Data were collected from seven primary sites for tumors and ten secondary regions where cancer metastasize were identified. Two Artificial Neural Network methods such as supervised back-prop method and unsupervised self-organizing map (SOM) method were used for the analysis of data. Although back-prop method was highly successful, a further work was required by SOM method.

Microarray analysis has been playing an important role in tumor diagnosis and classification, prediction of prognoses and treatment and developing an understanding of molecular mechanisms, gene networks and bio-chemical pathways (Fan & Ren, 2006). According to the researchers, it has been proved that a proper and systematic statistical analysis on microarray data can pave ways for the individualized treatment for cancer patients in future as well. With the development in technologies and innovative statistical tools, genomic and cancer related researches can be undertaken for many scientific investigations in a better way than before.

Mount et al. (2014), have conducted a research based on squamous cell carcinoma of lungs and breast with the objective of increasing the survival rate of cancer patients through early detection of their gene expressions. In the particular research, refined set of genes were tested by means of logistic regression. Two-sliding window analysis and leave-one-out approaches were utilized for the validation of the gene lists in both lung and breast carcinoma and they were successful in identifying 24 prognostic genes. Researchers concluded that the Logistic Regression was more accurate in predicting clinical outcome of cancers rather than the traditional Kaplan Meier (KM) approach.

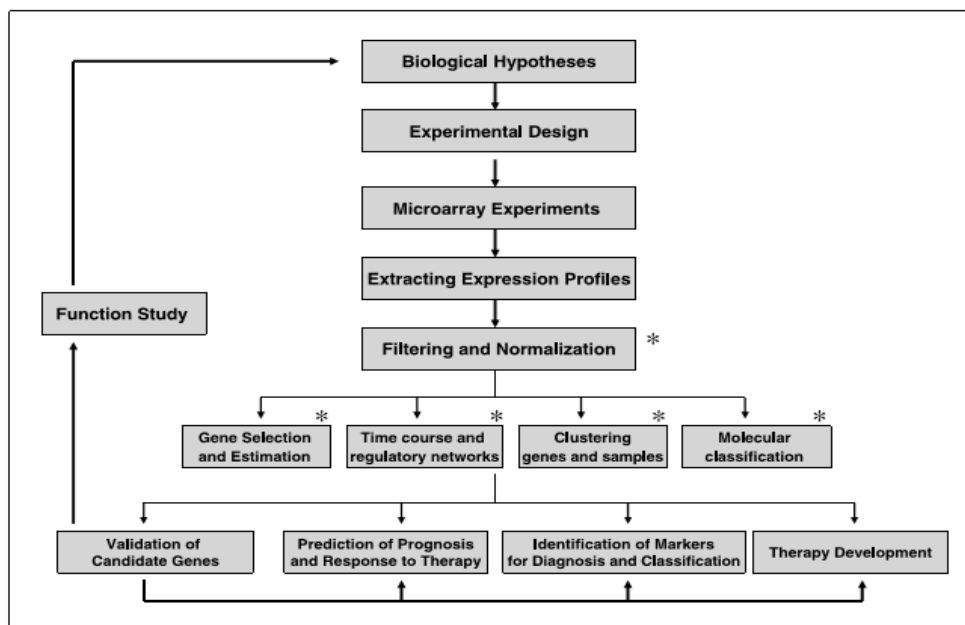


Figure 2. Schematic Representation of microarray analysis in association with the steps for statistical analysis (Steps for statistical analysis are marked with \*).

Another research (Statnikov, Aliferis, Tsamardinos, Hardin & Levy, 2005) has also developed a software system called GEMS (Gene Expression Model Selector) with the goal of automating high-quality model construction while enforcing sound optimization and performance estimation procedures. The researchers were successful in developing optimal classification models for the domain of cancer diagnosis with microarray gene expression data which could estimate their performance in future patients.

The particular software was able to investigate which one among the many powerful classifiers available presently for the diagnosis of gene expression, could perform at the best across many cancer types; how classifiers interact with existing gene selection methods in datasets with varying sample size, number of genes and cancer types; Whether it was possible to use meta-learning in the form of ensemble classification as a way to increase diagnostic performance; and how to parameterize the classifiers and gene selection procedures in order to avoid over-fitting.

### 3. BIOMARKERS AND ORAL CANCER

From the research of Li et al. (2004), salivary transcriptome can be used to detect the presence of oral cancer. Unstimulated salivary samples were obtained from ten OSCC patients and from ten normal persons. To differentiate salivary RNA profiles between cancer patients and matched normal persons, the HG U133A microarrays were used. Statistical analysis

was done for those exhibited at least 2-fold change. The research was successful in using salivary transcriptome and its microarrays to detect cancer in advance.

Differential display analysis was conducted to compare normal epithelium tissues with Oral Squamous Cell Carcinoma (Chang et.al, 2005) to be used as a predictive molecular markers of oral cancer. The researchers identified a total of 7 genes such as NPM, CDK1, NDRG1, HMGCR, EF1A, NAC and CHES1. The identified genes were classified as either over-expressed or under-expressed based on comparative real time RT-PCR.

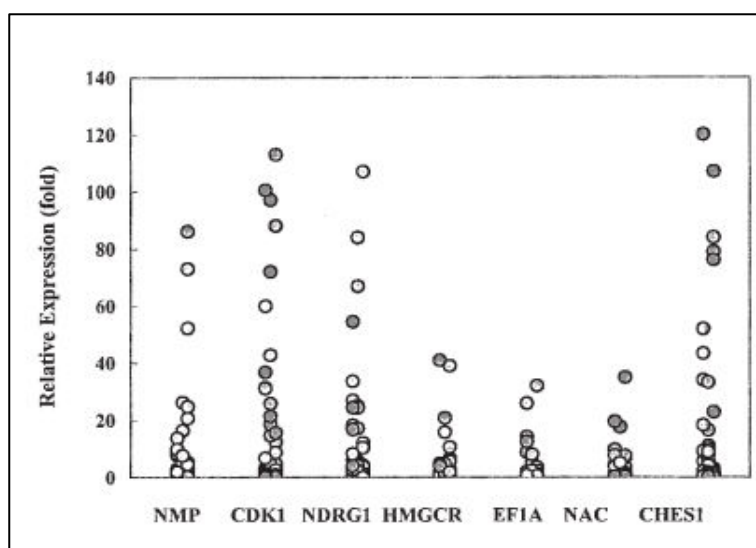


Figure 3. Relative expression levels of the 7 genes that were either overexpressed or under-expressed in tumor samples as compared to their normal counterparts. (Chang et.al, 2005)

#### 4. COMPUTATIONAL APPROACHES TO ORAL CANCER

As per Ford and Farah (2013), there is an immense need of a theory based studies with respect to oral cancer, such that the five-year survival rate of cancer detected patients can be increased. The primary prevention and early detection of oral cancer is the principal solution to reduce the mortality rate as well as to control the inflation. The survival rate of Oral cancer can be increased, if the disease condition can be diagnosed at an early stage. The field that is highly promising for such prognosis is molecular biology (Jurel, Gupta, Singh D., Singh M., Srivastava, 2014). Although there are several other technologies available in support of molecular biology, DNA microarray technology is one of the best omics, as it permits profiling of the expression of thousands of genes at once (Gevaert & Moore, 2009).

According to Kuo (2003), Bioinformatics is an emerging area which has gained wide acceptance in many of biomedical researches. And specifically, oral genomics can help researchers to get a clear understanding of genes

and their interaction. The availability of several microarray platforms for measuring gene expression allows makes consistency and reproducibility across high throughput technologies. Kuo (2003) states that when the clinical data is incorporated with bioinformatics, it can create a great understanding of genomic analysis and this in turn increase the understanding of many biological challenges in dental science.

In another research (Siow-wee & Kareem, 2010), Adaptive Network-based Fuzzy Inference Systems (ANFIS) have been used for the prediction of oral cancer survival using small medical dataset. In order to increase the accuracy of the prognosis of oral cancer, genomic markers and clinical markers were tested separately. However, it was concluded that the combination of both markers would give rise to accuracy in early prediction of oral cancer.

Salvi et al. (2013) have used a set of biomarkers in order to predict the probability of recurrence of oral cancer as a mean to optimize the treatments. They collected huge amount of datasets from genomic, imaging and clinical evidences as there was uncertainty in progression. The researchers were successful in developing a novel ICT-enabled cancer recurrence prediction method for Oral Squamous Cell Carcinoma (OSCC).

Researches (Randhawa & Acharya, 2015) have used logistic regression model and integrated network analysis to identify the key genes which correlate with the advancement of the oral squamous cell carcinoma stages. Here, they have identified thirteen key hub genes which were associated with the progression of the OSCC stages and by using classifier model they have divided these genes among cancer stages. Therefore, they emphasize that these selected gene set may yield to identify and ratify the biomarkers of the OSCC.

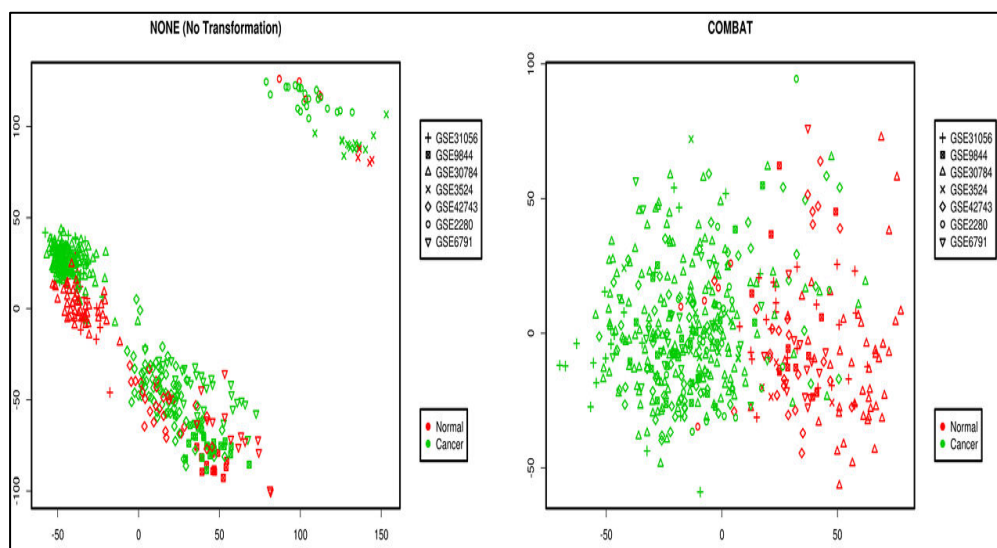


Figure 4. Multidimensional plot of merged gene expression data before and after the removal of batch effect (Randhawa & Acharya, 2015)



## 5. CONCLUSION

A series of researches has been undertaken in terms of predicting cancers and especially for oral cancer as mentioned above. However, the five year survival rate of it, remains low over the past two decades. The main reason for this could be that Oral Squamous Cell Carcinoma is identified in its advanced stages (Cheng, Rees & Wright, 2014). Further, the compendium of literature suggests that there is a variation of gene expression among subtypes of oral cancer and the stages of it. Thus, it could be derived that the effective solution to increase the survival rate is to detect it at an early stage and to individualize the treatment for specific molecular sub types of it. This paper concludes that the identification of the differentially expressed gene expressions associated with each of the molecular subtypes of oral cancer and also with the stage-specific gene expression through a comprehensive computational study on microarrays could be a timely solution.

## 6. REFERENCES

- Chang, J.T., Wang, H.M, Chang, K.W., Chen, W.H., Wen, M.C., Hsu, Y.M., Yung, B.Y.M., Chen, I.H., Liao, C.T., Hsieh, L.L. & Cheng, A.J. (2005). Identification of differently expressed genes in oral squamous cell carcinoma (OSCC): Overexpression of NPM, CDK1 and NDRG1 and under-expression of CHES1, *Publication of the International Union against Cancer*, 949(114)
- Cheng, Y.L., Rees, T. & Wright, J. (2014). A review of research on salivary biomarkers for oral cancer detection, *Clinical and Translational Medicine*, 3(3)
- Darby, E., Netti, T., Kodali, S. & Shih, L. (2005). Head and Neck Cancer Metastasis Prediction via Artificial Neural Networks, *IEEE publications*
- Fan, J. & Ren, Y. (2006). Statistical analysis of DNA microarray data in cancer research, *Clin Cancer Res* 2006, 12(15), 4469-4473
- Gevaert, O. & De Moore, B. (2009). Prediction of cancer outcome using DNA microarray technology: past, present and future, *Expert Opin. Med. Diagn*, 3(2), 157-165.
- Jurel, S.K., Gupta, D.S., Singh, R.D., Singh, M. & Srivastava, S. (2014). Genes and oral cancer, *Indian J Hum Genet*, 20(1), 4–9.
- Kuo, W.P. (2003). Overview of bioinformatics and its application to oral genomics, *Adv Dent Res*, 17, 89-94
- Kuo, W.P., Whipple, M.E., Jenssen, T.K., Todd, R., Epstein, J.B., & Ohno-Machado, L., et al. (2003b). 'Microarrays and clinical dentistry', *J Am Dent Assoc*, 134, 456-462
- Kuo, W.P., Whipple, M.E., Sonis, S.T., Ohno-Machado, L. & Jenssen, T.K. (2002c). Gene expression profiling by DNA microarrays and its application to dental research, *Oral Oncol*, 38, 650-656.

Li, Y., Maie, A. R., John, S., Zhou, X., Kim, Y., Sinha, U., Richard, C. K., Eisele, J.D., Abemayor, E., Elashoff, E., Park, N.H. & Wong, D.T. (2004). Salivary Transcriptome Diagnostics for Oral Cancer Detection, *American Association for Cancer Research*, 10, 8442–8450

Mount, D.W., Putnam, C.W., Centouri, S.M., Manziello A.M., Pandey, R., Garland, L.L. & Martinez, J.D. (2014). Using logistic regression to improve the prognostic value of microarray gene expression data sets: application to early-stage squamous cell carcinoma of the lung and triple negative breast carcinoma, *BMC medical genomics*, 33(7)

Randhawa, V. & Acharya, V. (2015). Integrated network analysis and logistic regression modeling identify stage-specific genes in Oral Squamous Cell Carcinoma, *BMC Medical Genomics*, 39 (8)

Salvi, D., Picone, M., Arredondo, M.T. & Cabrera-Umpi'Errez, M.F. (2013). Merging Person-Specific Bio-Markers for Predicting Oral Cancer Recurrence through an Ontology, *IEEE transactions on biomedical engineering*, 60(1)

Shillitoe, E.J., May, M., Patel, V., Lethanakul, C., Ensley, J.F., Strausberg, R.L., & Gutkind, J.S. (2000). Genome-wide analysis of oral cancer early results from the Cancer Genome Anatomy Project, *Oral Oncology*, 36, 8-16

Statnikov, A., Aliferis, C.F., Tsamardinos, L., Hardin, D., Levy, S. (2005). A comprehensive evaluation of multiclassification methods for microarray gene expression cancer diagnosis, *Oxford University Press*, 21(5), 621-643

Vijver, V.D., Yudong, D.H., Veer, L.J.V., Dai, H., Hart, A.A.M., Voskuil, D.W., & Bernards, R. et. al. (2002). A gene-expression signature as a predictor of survival in breast cancer, *The New English Journal of Medicine*, 347

Wee, C.S. & Kareem S.A. (2010). A Bootstrap-ANFIS Framework for Oral Cancer Prognosis Based on Clinical and Genomic Markers, IEEE publications