Identification and Functional Assessment of Novel Gene Sets towards Better Understanding of Dysplasia Associated Oral Carcinogenesis

Satarupa Banerjee [1, *], Anji Anura [1], Jitamanyu Chakrabarty [2], Sanghamitra Sengupta [3], Jyotirmoy Chatterjee [1]

[1] School of Medical Science and Technology, Indian Institute of Technology, Kharagpur 721302, India
[2] Department of Chemistry, National Institute of Technology Durgapur, India
[3] Department of Biochemistry, University of Calcutta, Kolkata, India

* Corresponding Author. E-mail: satarupa@smst.iitkgp.ernet.in

Abstract. Oral epithelial dysplasia (OED) often precedes oral cancer. Understanding the underlying complex biological aspects of dysplasia associated oral carcinogenesis using important gene sets is thus important. Computation assisted gene set identification through different feature ranking and visualization techniques was therefore attempted in this study. Result suggested that, weighted support vector machine (SVM) could be useful for feature ranking and SVM for attribute selection. Alteration in keratinization, cell–cell communication and peptidase activity was the major affected phenomena, while extracellular matrix dynamics was also found to be hampered. During best gene subset identification, set of six genes could classify normal (NOM) and oral squamous cell carcinoma (OSCC) conditions and two sets comprising four genes in each could classify NOM and dysplastic (DYS) conditions with 100% sensitivity and specificity. A gene set, comprising 32 genes showed best efficacy of 94.12% sensitivity, 99.40% specificity and 98.89% accuracy during classification of DYS and OSCC.

Keywords: OED, oral epithelial dysplasia (OED); SVM, support vector machine; NOM, normal; OSCC, oral squamous cell carcinoma; DYS, dysplastic.

1. Introduction

Oral epithelial dysplasia (OED) is often a step that precedes development of squamous cell carcinoma. It can either convert to oral squamous cell carcinoma (OSCC) or revert back to normal condition, if treated early. Till date there are no specific biomarkers which may be precisely utilized to assess malignant potentiality of oral precancers including OED. Histopathological evaluation of biopsy specimens still serves as gold standard for critical detection of grades of dysplasia and for predicting its malignant potentiality. However, the procedure lacks specificity and suffers from inter and/or intra-observer variability because of the paucity of unequivocal features of dysplasia that may be regarded as cardinal markers for accurate prediction of progression risks in oral pre-malignant disorders A recent review suggested that combination of selected biomarkers may be effective to address such problem (Banerjee and Chatterjee, 2015).

OED is a histopathological condition, where cytological and architectural characteristics of oral mucosa are altered. The role of OED in oral carcinogenesis is quite controversial. Some literature suggests that likelihood of malignant transformation of OED is significant (Al-Dakkak, 2010), while other studies have shown that there is no correlation between malignant potentiality and grade of dysplasia (Dost et al., 2014). In such circumstances, understanding the molecular progression of OED to OSCC is important and can no longer be avoided (Pitiyage et al., 2009). Semi-quantitative analysis of immunohistochemically stained tissue sections has been attempted to grade OED in precancers, (Anura et al., 2014) however, the procedures are still immature and have not yet been utilized in routine clinical practices. Comparative and quantitative assessments of histological grading and immunohistochemical expression of few key molecules to study the association between OED and OSCC were reported in few studies (Anura et al., 2014 and Tabor et al., 2003). Molecular dissection

of oral carcinogenesis has also been attempted through the analysis of proteome and deregulation of molecular network (Molinolo et al., 2009), but understanding the progression of OED to OSCC remains in its infancy. In silico analysis of microarray gene expression data is recently gaining interest for selection of candidate gene which may be subjected to gene ontology (GO) and functional enrichment analysis for understanding underlying molecular, biological and cellular activities of given gene sets and prioritizing candidate diagnostic indicators (Hindumathi et al., 2014).

In this study, an in-depth bioinformatic and statistical analyses of the microarray transcriptome were attempted to throw light on the process. Differentially expressed (DE) genes were primarily selected to dissect progression of OSCC through OED. Weighted support vector machine (SVM) was employed to select precise gene subset towards optimal classification of oral lesions, OED and OSCC. Venn diagram was implemented in visualization of complex association of different gene sets, to unearth their possible functional association (Kestler et al., 2005). The major aim of this cost-minimized strategy exercise is to select a novel gene sub-set which can modulate specificity and sensitivity of the classification task.

The main challenge of microarray data analysis includes high number of variables against a small sample size, from which meaningful gene sets have to be chosen which should classify the disease with maximum efficiency at optimum computational burden and diagnostic cost (Liu et al., 2011). Supervised machine learning classifiers such as Naïve Bayes (NB) (Wu et al., 2012) and k nearest neighbor (KNN) (Zhang and Deng, 2007) are commonly used for cancer microarray data classification in addition to support vector machine (SVM). In this study efficiency of these three classifiers were evaluated. Feature ranking and feature selection are routinely used to reduce data dimensionality and improve learning and predictive efficiencies. A recent study showed feature ranking utilizing weights from linear SVM yields better result even with non-identically distributed training and testing data [13]. ReleifF is a feature selection algorithm, which acts through filtering and is popularly used in cancer microarray data analysis. It randomly draws instances and after computing the nearest neighbors, weighs the feature. It comparatively provides higher weightage to the attribute which have higher differentiating ability of the instance from neighbors of other class (Wang and Makedon, 2004). Efficacy of feature selection algorithms such as weighted SVM was also evaluated in this study during gene selection. Several data visualization techniques are used in cancer microarray data towards knowledge discovery and class labeled data analysis [15]. Among them, VizRank is a simple gene set ranking technique, which works through utilizing visual projections of class labeled data. Here, we employed Radviz (Radial Coordinate visualization)(Novakova and Stepankova, 2009 and Mramor et al., 2005) based gene identification with minimal gene numbers (three), to reduce computation cost, as well as to identify a subset of molecular criteria showing maximum efficacy which may potentially be implemented in routine diagnostics.

2. Materials and method

GSE30784 dataset was downloaded from Gene Expression Omnibus and used in this study, which consisted of 167 OSCC, 17 OED and 45 NOM samples. DE genes for each two class conditions were obtained using GEO2R (Barrett et al., 2013). The cut-off for gene selection was p value < 0.05 and log FC value ± 2. During 3 class disease classification, cut-off value was p value < 0.05 and F score more than 100.

Initially, all DE genes, both upregulated and downregulated gene sets were identified separately and then gene ontology (GO) analysis and pathway analysis for each gene set was performed using EnrichR (Chen et al., 2013) where common pathways as well as important biological process, cellular component and, molecular function were identified. In gene ontology (GO) analysis, when minimum of 5 genes were found to be present in any condition, was considered significant. When too many processes or functions were obtained, a threshold of combined score was considered and mentioned accordingly in the “Result and discussion” section. Pathway analysis was done using KEGG 2015 pathway. Common pathway and gene ontology analyses were performed with cut-off of combined score 25. The concept of combined score in EnrichR is to integrate both p value and z score with the formula $c = \log(p) \cdot z$ where c is the combined score, represented by p, p-value computed using the

Fisher exact test, and z the z-score computed by assessing the deviation from the expected rank. Since Enrichr provides all three options for sorting enriched terms, combined score of 25, and p value < 0.005 were only considered (Chen et al., 2013). Venn diagram was prepared using three different gene subsets, as well as six sub-sets of up- and down-regulated genes to identify common and exclusive genes in each process using InteractiVenn (Heberle et al., 2015). Utilization of this method aided understanding of the complexity of association of both upregulated and downregulated gene sets. GO analysis and pathway analysis for each gene set were also again performed using EnrichR (Chen et al., 2013).

During specific gene subset selection for optimal disease classification, efficiency of different supervised classifiers namely SVM, KNN and Naïve Bayes was assessed using best features obtained through weighted SVM feature ranking method. Efficiency of another feature ranking techniques namely ReliefF was compared with weighted SVM and plotted accordingly with the best classifier obtained in the previous step, SVM. For selection of best feature subset, manual sequential feature reduction was carried out and optimal classification efficiency was evaluated at 10 fold cross-validation. The gene set obtained for NOM and OSCC was cross-validated in GSE9844 data set, which comprised of 12 NOM and 26 tongue OSCC samples. These analyses were performed in Orange 2.7 (Demšar et al., 2004). Visualization based classification by Radviz with minimal gene numbers (three) was also performed. Plots have been provided in the supplementary figure. Biological functions of the genes obtained in this study have also been mined from Genecard (Safran et al., 2010) and presented in supplementary Tables 1, 2 and 3. The schematics of the entire process have been provided in Fig. 1.

3. Result and discussion

This study was performed towards utilization of gene ranking and visualization based precise gene set selection for comprehensive biological and bioinformatic knowledge fusion. The integrated approach of analysis was performed for cost minimization of specific gene signature selection from genomics information, which can be further validated by molecular pathology techniques or other relevant datasets.

For GO and pathway analysis, when DE genes were extracted, the result suggested that out of 54,675 genes, 829 genes were initially expressed in this study. The common pathway analysis suggested that cell communication was found to be affected in all conditions; while extra-cellular matrix (ECM) receptor interaction was commonly hampered in both NOM to OED and OSCC transition. The role of alteration of cell–cell adhesion and other intercellular communications, both in junction based and non-junctional modes, especially during epithelial mesenchymal transition during carcinogenesis is an established phenomenon, and thus was supported by the results (Kandouz, 2015 and Loewenstein and Kanno, 1966). Cytokine–cytokine receptor was also found to be affected in both NOM and OED to OSCC transition. A recent study validated the findings, since it confirmed that early cytokine–cytokine receptor induction is triggering factor of oral carcinogenesis (Liu et al., 2012). It is also evident from existing literature that deregulation in cell proliferation and cellular invasion, which hampered cellular differentiation, is associated with ECM dynamics which is found to be affected in fibrosis and cancer (Lu et al., 2011 and Sainio and Järveläinen, 2014). During biological activity analysis epidermis development (GO: 0008544) was found to be hampered in both NOM-DYS and DYS-CA process, while inflammatory response (GO: 0006954), taxis (GO: 0042330) and chemotaxis (GO: 0006935) were commonly hampered in NOM-CA and DYS-CA conditions. Eight GO terms extracellular matrix organization (GO: 0030198), extracellular structure organization (GO: 0043062), collagen metabolic process (GO: 0032963), multicellular organismal macromolecule metabolic process (GO: 0044259), multicellular organismal metabolic process (GO: 0044236), collagen catabolic process (GO: 0,030,574), multicellular organismal catabolic process (GO: 0044243) and extracellular matrix disassembly (GO: 0022617) was found to be altered in all processes. Recent studies suggested that intracellular collagen degradation is also associated with ECM turnover during malignancy due to altered μPARAP/Endo180 expression in mammary gland (Curino et al., 2005), which might be also the case here. A recent study also showed expression of

inflammation and ECM components in oral carcinogenesis and validated the notion obtained in this study (Tanis et al., 2014). The important affected biological process in OED associated oral carcinogenesis has been shown in Table 1, while Table 2 presented important affected cellular components in OED associated oral carcinogenesis obtained from GO analysis using of DE genes. Fig. 1 showed Venn diagram showing association of DE genes obtained in each two class conditions, NOM-OSCC, NOM-DYS and DYS-OSCC.

When common 28 cells common in all genes were subjected to GO analysis, it was found that in epithelial cell differentiation (GO: 0030855), epidermis development (GO: 0008544) and epithelium development (GO: 0060429), more or equal to 5 genes were involved. Extracellular matrix components were also involved and were in synergy with previous results (Banerjee and Chatterjee, 2015). Peptidase regulator activity (GO: 0061134), endopeptidase activity (GO: 0004175) and calcium ion binding (GO: 0005509) were the affected molecular function. Previously other network based studies showed activity of calcium binding proteins in oral carcinogenesis too (Nomura et al., 2007).

In all conditions, ECM related areas (extracellular region (GO: 0005576), extracellular space (GO: 0005615), extracellular vesicular exosome (GO: 0070062), extracellular matrix (GO: 0031012) and proteinaceous extracellular matrix (GO: 0005578)) were found to be affected. During NOM to OSCC transition, basement membrane (GO: 0005604) was also found to be involved. This concept supports a well-established fact that malignant potentiality is correlated with enzymatic degradation of basement membrane collagen (Liotta et al., 1980). Interestingly collagen related components were found to be affected in both NOM to DYS and OSCC transition, but not during DYS to OSCC conversion. It can be hence implied that collagen related alterations are prominent in early stages of carcinogenesis, while ECM related alterations are evident in all stages of. This result supports the preconceived result that in severe dysplasia and OSCC ECM is disintegrated, while collagen III and laminin play role in neo-angiogenesis in lung cancer (Fisseler-Eckhoff et al., 1990). The same mechanism is also likely to happen in oral carcinogenesis.

When the gene sets were segregated using Venn diagram with union by list specification and both upregulated and downregulated genes were utilized (depicted in Fig. 2 and Fig. 3), total 16 classes were identified and provided in Supplementary Table 4. For understanding the process of OED, role of the exclusive genes when evaluated, the gene ontology with combined score more than 5, it was found that negative regulation of proteolysis (GO: 0045861), negative regulation of protein processing (GO: 0010955), negative regulation of protein maturation (GO: 1903318), negative regulation of endopeptidase activity (GO: 0010951), negative regulation of peptidase activity (GO: 0010466), regulation of endopeptidase activity (GO: 0052548) and regulation of peptidase activity (GO: 0052547) was found to be hampered when biological activity was evaluated. Mostly endopeptidase activity is assaulted (Serine-type endopeptidase inhibitor activity (GO: 0004867) peptidase regulator activity (GO: 0061134), endopeptidase inhibitor activity (GO: 0004866), peptidase inhibitor activity (GO: 0030414), endopeptidase regulator activity (GO: 0061135) and enzyme inhibitor activity (GO: 0004857)) during molecular function assessment. In this regard a recent study suggested that expression of ADAMTS2 is important in craniofacial fibrous dysplasia (Zhou et al., 2014), while another study showed that ADAMTS2 is associated to regulation of procollagen amino-propeptide processing and affect collagen biosynthesis (Le Goff et al., 2006). Hence this information is in synergy with our result.

From the affected genes involved in NOM-DYS-UP and DYS-CA-UP, it was found that keratinization (GO: 0031424) was the most involved biological process, and mainly cell communication is affected in KEGG pathway. A recent study also revealed such associated alteration in oral carcinogenesis (Kandouz, 2015 and Banerjee et al., 2015).

Result shown in Fig. 4 depicted that NOM and DYS as well as NOM and OSCC can be clearly differentiated on the basis of principle components, but there were significant overlapping in DYS and OSCC conditions. Again in clinical theranostics, single genes would have better implication than principal components. So further single gene based analysis was initiated, but it could be understood that complete diagnostic segregation of DYS and OSCC is quite difficult.

When performances of the classifiers were tested, SVM showed maximum efficacy and thus was chosen for further analysis (Fig. 5a). When the role of two feature ranking method, ReliefF and weighted SVM were evaluated, weighted SVM showed better potential and thus was used for feature selection (Fig. 5b). Gene selection result showed that when NOM and OSCC were classified, best 23 features according to weighted SVM method could classify with 100% sensitivity and specificity. Further manual sequential feature reduction aided selection of six genes, which also showed similar efficacy and thus helped towards computational cost reduction. The gene sets with their corresponding weighted values in first bracket are PEG3 (0.037), UPK1A (0.018), LAMB1 (0.010), GREM1 (0.008), TYRP1 (0.007) and COMP (0.007). When the efficacy of the same gene set was validated using a different dataset, GSE9844 sensitivity was found to be 66.67%, specificity 88.46% and 81.67% accuracy, but further manual sequential feature reduction and elimination of LAMB1 and UPK1A resulted in significant betterment of the classification efficacy (sensitivity 91.67%, specificity 92.31% and accuracy 92.50%). The differences in the results are might be site specific variation in gene expression, since in that study only tongue cancers were considered (Ye et al., 2008). When Radviz score based projection with maximum three genes were performed, five gene sets were found to have score more than 99. The gene sets in the third bracket with their scores in the first bracket were [PEG3, MYBPC1, COL1A1] (99.26), [LOC100506098, MYBPC1, COL1A1] (99.19), [COL3A1, CHRDL1, PEG3] (99.14), [COL3A1, UPK1A, PEG3] (99.02), [LPIN1, COL1A1, PEG3] (99.00). Both results are far better than previous results, where PCA based weightage linked gene selection was performed in the same dataset to classify NOM and OSCC, while the best classification was found for gene set [MMP1, RUNX2, MTERFD2] with 98.80% sensitivity and 95.60% specificity (Kim et al., 2014).

During NOM and DYS classification, best 20 features could classify the conditions with 100% sensitivity and specificity too. Sequential feature reduction showed that two sets of 4 genes in each also provide the same results. The first gene set comprised of PRR9 (0.002), LMO7 (0.001), CAPN14 (0.001) and LOC344887 (0.001) while KRT10 (0.002), CRYM (0.002), LMO7 (0.001) and ATP6V0A4 (0.001) were present in the second list. Interestingly during Radviz scoring based classification with maximum three genes that were performed, ten gene sets were obtained with score of 100 and CDSN was common in all sets. The gene sets shown in the third brackets are [CDSN, CRYM, HYAL1], [CDSN, SLC8A1-AS1, ANKRD20A5P], [CDSN, F2RL2, FAM3D], [CDSN, PRR9, CAPN14], [CDSN, COL3A1, CEACAM1], [CDSN, FAM3D, LRRC15], [CDSN, FAM3D, HYAL1], [CDSN, FAM3D, ANKRD20A5P], [CDSN, ANKRD20A5P, CAPN14] and [CDSN, COL3A1, FAM3D]. CDSN, gene associated with epidermal barrier integrity was found to be most interesting in this set, which was also found to be present in all sets.

Diagnostic classification of DYS and OSCC using the gene expression was found to shown comparatively lesser efficiency than other two classes and more computation cost had to be exploited, since the number of genes in the subset was quite high. It was found that, the best 45 genes in weighted SVM could classify the lesions with 88.24% sensitivity, 99.40% specificity and 98.33% accuracy. Then a gene set of 32 genes was obtained which showed better efficacy by manual sequential feature selection (94.12% sensitivity, 99.40% specificity and 98.89% accuracy). Further feature reduction was also tried, but since specificity was found to be reduced nearly up to 1% (98.20%) and sensitivity of nearly 6% (88.24%) with 22 selected genes, in spite of the large number of genes, the former gene set was considered to be optimal. Although in previous studies the number of genes in the gene set was lesser, the sensitivity and specificity are higher in this study (Kim et al., 2014 and Chen et al., 2008). The result has been shown in Table 3. When the efficacy of Further then Radviz projection based classification was performed with maximum three number of genes, the gene set comprising BPIFC, PLEK2 and TNFAIP3 was found to be the only gene saving score greater than 94 (94.29).

Finally when three classes were tried to be segregated using linear SVM the result suggested that the best ranked 23 genes of weighted SVM feature ranking method could classify NOM, DYS and OSCC with 82.35% sensitivity and 99.53% sensitivity 96.53% accuracy. When manual sequential feature reduction was performed, it was found that a gene set containing 13, could classify the

conditions with 82.35% sensitivity and 100% specificity. The gene set with their corresponding weighted value in first bracket are LOX (0.823), 2NF519 (0.531), MTERF4 (0.498), F2RL2 (0.444), CLEC3B (0.418), AADAC (0.359), HSBP1 (0.345), NDNF (0.313), MMP1 (0.285), CRISP3 (0.269), ENAH (0.258), ATP2C1 (0.229) and PAQR8 (0.192). From the selected 4 gene sets, FAM3D were found to be common in NOM-OSCC and DYS-OSCC classes, COL3A1 were common in NOM-OSCC and NOM-DYS condition, which was found to be associated with cytokine activity and in collagen III expression respectively. Previously 5 gene sub-sets were identified in colorectal cancer stage specific classification (Berdiel-Acer et al., 2014), while this study is one of the detailed endeavor to identify gene sub-sets in oral cancer and their biological pathway and GO analysis.

4. Conclusion

It can be concluded that, knowledge discovery through integration of state-of-the-art data mining followed by meaningful biological interpretation of the result has been implemented in this study for understanding OED associated carcinogenesis. Utilization of both feature ranking and visualization technique aided identification of precise gene sets with minimum number of genes for optimal classification of two or three different conditions. Gene set selection was also performed towards minimization of arbitrary selection of gene sets in this respect. In turn, the selected gene sets in this study are expected to be used in routine clinical practice towards cost minimization of molecular pathology based oral lesion diagnostics..

References

1. Al-Dakkak, I., 2010. Oral dysplasia and risk of progression to cancer. Evid. Based Dent. 11, 91–92.
2. Anura, A., Conjeti, S., Das, R.K., Pal, M., Bag, S., Paul, R.R., Ray, A.K., Chatterjee, J., 2014. Computer-aided molecular pathology interpretation in exploring prospective markers for oral sub-mucous fibrosis progression. Head Neck.
3. Banerjee, S., Chatterjee, J., 2015. Molecular pathology signatures in predicting malignant potentiality of dysplastic oral pre-cancers. Springer Sci. Rev. 3, 127–136.
4. Banerjee, S., Pal, M., Chakrabarty, J., Petibois, C., Paul, R.R., Giri, A., Chatterjee, J., 2015. Fourier- transform-infrared-spectroscopy based spectral-biomarker selection towards optimum diagnostic differentiation of oral leukoplakia and cancer. Anal. Bioanal. Chem. 407, 7935–7943.
5. Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall,
6. K.A., Phillippy, K.H., Sherman, P.M., Holko, M., Yefanov, A., Lee, H., Zhang, N.,
7. Robertson, C.L., Serova, N., Davis, S., Soboleva, A., 2013. NCBI GEO: archive for functional
8. genomics data sets—update. Nucleic Acids Res. 41, D991–D995.
9. Berdiel-Acer, M., Berenguer, A., Sanz-Pamplona, R., Cuadras, D., Sanjuan, X., Paules, M.J., Santos, C., Salazar, R., Moreno, V., Capella, G., Villanueva, A., Mollevíi, D.G., 2014. A 5-gene classifier from the carcinoma-associated fibroblast transcriptomic profile and clinical outcome in colorectal cancer. Oncotarget 5, 6437–6452.
10. Chen, C., Méndez, E., Houck, J., Fan, W., Lohavanichbutr, P., Doody, D., Yueh, B., Futran, N.D., Upton,M., Farwell, D.G., 2008. Gene expression profiling identifies genes predictive of oral squamous cell carcinoma. Cancer Epidemiol. Biomark. Prev. 17, 2152–2162.
11. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., Ma'ayan, A., 2013. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinf. 14, 1–14.
12. Curino, A.C., Engelholm, L.H., Yamada, S.S., Holmbeck, K., Lund, L.R., Molinolo, A.A.,Behrendt, N., Nielsen, B.S., Bugge, T.H., 2005. Intracellular collagen degradation mediated by uPARAP/Endo180 is a major pathway of extracellular matrix turnover during malignancy. J. Cell Biol. 169, 977–985.
13. Demšar, J., Zupan, B., Leban, G., Curk, T., 2004. Orange: FromExperimentalMachine Learning to Interactive Data Mining. Springer.
14. Dost, F., Lê Cao, K., Ford, P.J., Ades, C., Farah, C.S., 2014. Malignant transformation of oral epithelial dysplasia: a real-world evaluation of histopathologic grading. Oral Surg. Oral Med. Oral Pathol. Oral Radiol. 117, 343–352.
15. Fisseler-Eckhoff, A., Prebeg, M., Voss, B., Muller, K.M., 1990. Extracellular matrix in preneoplastic lesions and early cancer of the lung. Pathol. Res. Pract. 186, 95–101.
16. Heberle, H., Meirelles, G.V., da Silva, F.R., Telles, G.P., Minghim, R., 2015. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. BMC Bioinf. 16, 169.
17. Hindumathi, V., Kranthi, T., Rao, S.B., Manimaran, P., 2014. The prediction of candidate genes for cervix related cancer through gene ontology and graph theoretical approach. Mol. BioSyst. 10, 1450–1460.
18. Kandouz, M., 2015. Intercellular Communication in Cancer. Springer.
19. Kestler, H.A.,Müller, A., Gress, T.M., Buchholz,M., 2005. Generalized Venn diagrams: a new method of visualizing complex genetic set relations. Bioinformatics 21, 1592–1595.
20. Kim, K.-Y., Zhang, X., Cha, I.-H., 2014. Combined genomic expressions as a diagnostic factor for oral squamous cell carcinoma. Genomics 103, 317–322.
21. Le Goff, C., Somerville, R.P., Kesteloot, F., Powell, K., Birk, D.E., Colige, A.C., Apte, S.S., 2006. Regulation of procollagen amino-propeptide processing during mouse embryogenesis by specialization of homologous ADAMTS proteases: insights on collagen biosynthesis and dermatosparaxis. Development 133, 1587–1596.
22. Liotta, L., Tryggvason, K., Garbisa, S., Hart, I., Foltz, C., Shafie, S., 1980. Metastatic potential correlates with enzymatic degradation of basement membrane collagen. Nature 284, 67–68.
23. Liu, Q., Sung, A.H., Chen, Z., Liu, J., Chen, L., Qiao, M., Wang, Z., Huang, X., Deng, Y., 2011. Gene selection and classification for cancer microarray data based on machine learning and similarity measures. BMC Genomics 12, S1.
24. Liu, Y.C., Ho, H.C., Lee, M.R., Lai, K.C., Yeh, C.M., Lin, Y.M., Ho, T.Y., Hsiang, C.Y., Chung, J.G., 2012. Early induction of cytokines/cytokine receptors and Cox2, and activation of NF-kappaB in 4-nitroquinoline 1-oxide-induced murine oral cancer model. Toxicol. Appl. Pharmacol. 262, 107–116.

25. Loewenstein, W.R., Kanno, Y., 1966. Intercellular Communication and the Control of Tissue Growth: Lack of Communication between Cancer Cells.
26. Lu, P., Takai, K., Weaver, V.M., Werb, Z., 2011. Extracellular matrix degradation and remodeling in development and disease. Cold Spring Harb. Perspect. Biol. 3, a005058.
27. Molinolo, A.A., Amornphimoltham, P., Squarize, C.H., Castilho, R.M., Patel, V., Gutkind, J.S., 2009. Dysregulated molecular networks in head and neck carcinogenesis. Oral Oncol. 45, 324–334.
28. Mramor, M., Leban, G., Demšar, J., Zupan, B., 2005. Conquering the curse of dimensionality in gene expression cancer diagnosis: tough problem, simple models. In: Miksch, S., Hunter, J., Keravnou, E. (Eds.), Artificial Intelligence in Medicine. Springer, Berlin Heidelberg, pp. 514–523.
29. Nomura, H., Uzawa, K., Yamano, Y., Fushimi, K., Ishigami, T., Kato, Y., Saito, K., Nakashima, D., Higo, M., Kouzu, Y., 2007. Network-based analysis of calcium-binding protein genes identifies Grp94 as a target in human oral carcinogenesis. Br. J. Cancer 97, 792–801.
30. Novakova, L., Stepankova, O., 2009. Radviz and identification of clusters in multidimensional data, in: Information Visualisation. 2009 13th International Conference. IEEE, pp. 104–109.
31. Pitiyage, G., Tilakaratne, W.M., Tavassoli, M., Warnakulasuriya, S., 2009. Molecular markers in oral epithelial dysplasia: review. J. Oral Pathol. Med. 38, 737–752.
32. Safran, M., Dalah, I., Alexander, J., Rosen, N., Iny Stein, T., Shmoish, M., Nativ, N., Bahir, I., Doniger, T., Krug, H., Sirota-Madi, A., Olender, T., Golan, Y., Stelzer, G., Harel, A., Lancet, D., 2010. GeneCards Version 3: the Human Gene Integrator, Database, 2010.
33. Sainio, A., Järveläinen, H., 2014. Extracellular matrix macromolecules: potential tools and targets in cancer gene therapy. Mol. Cell. Ther. 2, 14.
34. Tabor, M.P., Braakhuis, B.J., van derWal, J.E., van Diest, P.J., Leemans, C.R., Brakenhoff, R.H., Kummer, J.A., 2003. Comparative molecular and histological grading of epithelial dysplasia of the oral cavity and the oropharynx. J. Pathol. 199, 354–360.
35. Tanis, T., Cincin, Z.B., Gokcen-Rohlig, B., Bireller, E.S., Ulusan, M., Tanyel, C.R., Cakmakoglu, B., 2014. The role of components of the extracellularmatrix and inflammation on oral squamous cell carcinoma metastasis. Arch. Oral Biol. 59, 1155–1163.
36. Wang, Y., Makedon, F., 2004. Application of Relief-F feature filtering algorithm to selecting informative genes for cancer classification using microarray ata. Computational Systems Bioinformatics Conference, 2004. CSB 2004. Proceedings. 2004 IEEE. IEEE, pp. 497–498.
37. Wu, M.-Y., Dai, D.-Q., Shi, Y., Yan, H., Zhang, X.-F., 2012. Biomarker identification and cancer classification based on microarray data using laplace naive bayes model with mean shrinkage. IEEE/ACM Trans. Comput. Biol. Bioinform. 9, 1649–1662.
38. Ye, H., Yu, T., Temam, S., Ziober, B.L.,Wang, J., Schwartz, J.L., Mao, L.,Wong, D.T., Zhou, X., 2008. Transcriptomic dissection of tongue squamous cell carcinoma. BMC Genomics 9, 69.
39. Zhang, J.-G., Deng, H.-W., 2007. Gene selection for classification of microarray data based on the Bayes error. BMC Bioinf. 8, 370.
40. Zhou, S.H., Yang, W.J., Liu, S.W., Li, J., Zhang, C.Y., Zhu, Y., Zhang, C.P., 2014. Gene expression profiling of craniofacial fibrous dysplasia reveals ADAMTS2 overexpression as a potential marker. Int. J. Clin. Exp. Pathol. 7, 8532–8541.