ABSTRACT

Dental informatics/dent informatics is the sub-disciplinary shoot of biomedical informatics. Bioinformatics is a relatively interdisciplinary field of life sciences. Successful completion of the Human Genome Project it became in the frame, bioinformatics made it easy to organize and access the exponentially generated large data. One discipline of the dentistry is benefitted by the bioinformatics is oral pathology. Data mining techniques and its implementation in the disease diagnosis, microarray analysis technique, genomic and proteomics used as a tool in detection of oral precancerous stages to malignant stage are discussed in this review article. Genomics belongs to the study of genome of the organisms; microarray technique is helpful in detection of genetic alteration in the oral cancer. Besides this, proteomics provides the information of total protein profiles of a cell. Proteomics allows the detection of many diseases or causative pathogens of head and neck region. Oral cancer related databases are helpful to access all the information of oral cancer related gene, carcinogens, gene function, cell signaling, pathway and mechanism. Here, in this review article attempts the computational applications which are beneficial in the development of the dent informatics, specialty and its application in oral pathology.

KEYWORDS: dent informatics, bioinformatics, genomics and proteomics, microarray, data mining technique.
INTRODUCTION

Bioinformatics has dynamic application over the biology, inter-disciplinary field of life sciences. Arrival of the genome epoch, it involved in sequence analysis, genome annotation, computational evolutionary biology, comparative genomics, gene and protein expression, network and system biology. Bioinformatics shared [fig: 1] other research fields that distinguish informatics and computational biology. It is also contributing to the medical research field and health care by their application like storing, retrieving, manipulating, excess information from databases.[1]

Dental informatics/dentoinformatics is the sub-disciplinary shoot of biomedical informatics. Biomedical informatics contains information of dentistry, medicine, pharmacy, nursing and health science. Bioinformatics plays an important role to manage and organize large data easily. Ultimate goal of the bioinformatics is to provide conventional methods and tools to understand and develop a perspective for the dentistry. In dentistry, oral pathology is benefitted from the bioinformatics. Oral pathology is the shoot of the dentistry which deals with the oral cancers.[2] Oral sub mucous fibrosis[3] has reported as precancerous stage of oral cancer.[4] It is thoroughly recognized that oral squamous cancer cell (OSCC) happened due to many molecular and biochemical and cellular alterations in the oral mucosal epithelial cells.[5] Previous studies paid attention on loss of heterozygosity (LOH) in OSCC, presently focused on LOH in premalignant lesions.[6-11] Tumor suppressor genes lost their function in carcinogenesis.[12] Inactivation of both transcripts of tumor suppressor genes must be lost (LOH) for loss of function. Rb (retinoblastoma), p53, and p16INK4A are TSG involved in HNSCC.[7, 13, 14] In this review article, comprehensive studies have done of research progress or application of bioinformatics tools, techniques and their approaches in detection of gene alteration, gene expression analysis and copy number variations of carcinogenic cells. Another aspect for bioinformatics in genome-based therapies is to assemble all the information related to genes and proteins of the human. Comparative genomics is the analysis of various genes, gene families, and the position of genes on the chromosomes of different organisms. On the basis these studies one can predict the evolutionary rearrangements and gene duplications that are the source of genetic variation between species. Variations among these genes can be used to predict gene and protein relationships and new biochemical pathways. A recent analysis of the human genome has revealed the presence of tandem duplicated regions that represent regions of genetic instability. These regions are often associated with human disease.[15]
DATABASES FOR ORAL CANCER GENES

OrCDB (oral cancer gene database) provides gene list, gene name gives detailed information on aliases, function, chromosomal location, mutations and SNPs, mRNA expression, Protein information, pathways involved and interacting proteins, expression of genes in different tissues, and clinical correlates, and interaction networks (http://www.tumor-gene.org/Oral/oral.html).

HNOCDB (http://gyanxet.com/hno.html) Head and Neck and oral cancer database. On the basis of text mining it resides information of gene and miRNA, study done on the basis of experimental evidence. Oral cancer genes are listed (http://gyanxet.com/oragen.html), through this database can browse altered genes. A total 133 altered genes of oral cancer are listed in (table no. 1). It resides details of 63 miRNA with their function and alteration in cancer. Among these 63 miR-200 and miR-99a can act as tumor suppressor.

List of databases and tools for oral cancer (table-1):

<table>
<thead>
<tr>
<th>Database/search tool</th>
<th>Type</th>
<th>Features</th>
<th>url</th>
</tr>
</thead>
<tbody>
<tr>
<td>PubMed</td>
<td>Database</td>
<td>Contains biomedical literature with 39 other literature Including information of DNA, protein sequence and gene related all information</td>
<td><a href="http://www.ncbi.nlm.nih.gov/pubmed">http://www.ncbi.nlm.nih.gov/pubmed</a></td>
</tr>
<tr>
<td>dbGaP</td>
<td>database</td>
<td>Database of genotypes and phenotypes Also includes information of genome-wide association and medical sequencing</td>
<td><a href="http://www.ncbi.nlm.nih.gov/gap">http://www.ncbi.nlm.nih.gov/gap</a></td>
</tr>
<tr>
<td>MeSH (Medical Subject Heading)</td>
<td>database</td>
<td>National library of Medicine’s record</td>
<td><a href="http://www.ncbi.nlm.nih.gov/mesh">http://www.ncbi.nlm.nih.gov/mesh</a></td>
</tr>
</tbody>
</table>
Data analysis and data mining techniques in oral pathology

Data mining is the automated techniques for data analysis to bare previously unrevealed connection among data items. Data mining involves the analysis of data stored in a data repository. Three major data mining techniques are regression, classification and clustering. Clustering is the most important process, when database is large. Data mining techniques have been known from many years, but as increasing data it became more prevalent. Data mining technique is not all about the database software. One can perform data mining technique with comparatively humble database systems and tools, including, creating or using off the shelf software packages. Complex data mining importance from the previous results and algorithms defined with existing software, with particular tools gaining a more affinity with different tools and techniques.\textsuperscript{[16]} This method is routinely use in large number data of industries such as medicine, crime analysis, web mining, and mobile computing, besides others utilize data mining.\textsuperscript{[17]}

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CanGEM (Cancer Genome mine)</td>
<td>Database</td>
<td>Contains array comparative genomic hybridization Data mining of copy number variation</td>
<td><a href="http://www.cangem.org/">http://www.cangem.org/</a></td>
</tr>
<tr>
<td>BLAST (Basic local Alignment Search Tool)</td>
<td>Tool</td>
<td>Ability to search rarely and more important sequence</td>
<td><a href="http://blast.ncbi.nlm.nih.gov/Blast.cgi">http://blast.ncbi.nlm.nih.gov/Blast.cgi</a></td>
</tr>
<tr>
<td>Pfam and SMART (Simple Modular Architecture Research Tool)</td>
<td>Tool</td>
<td>Protein domain family analysis</td>
<td><a href="http://pfam.xfam.org/">http://pfam.xfam.org/</a> and <a href="http://smart.embl-heidelberg.de/">http://smart.embl-heidelberg.de/</a></td>
</tr>
<tr>
<td>OMIM (online Mendelian inheritance in man)</td>
<td>Database</td>
<td>Gene and genetic phenotype that is freely available</td>
<td><a href="http://www.ncbi.nlm.nih.gov/omim">http://www.ncbi.nlm.nih.gov/omim</a></td>
</tr>
<tr>
<td>KEGG (Kyoto encyclopedia gene and genome)</td>
<td>Database</td>
<td>Understanding high level functions</td>
<td><a href="http://www.genome.jp/kegg/">http://www.genome.jp/kegg/</a></td>
</tr>
</tbody>
</table>
Data mining is a multistep process:\cite{16}

- Information association
- Information classification
- Information clustering
- Pattern sequencing
- Preparation of decision tree
- Implementation

PREVENTION AND DIAGNOSIS

Data mining technique prediction system plays a role in strategy preparation for prevention of communicable and non communicable diseases in a particular area. These techniques are also useful in disease diagnosis. It can provide an answer to complex queries for diagnosing of disease and helps healthcare practitioners to make strategically clinical decisions which traditional decision support systems cannot. A providing effective treatment as well as it helps to reduce treatment costs. To make increment in visualization and ease of interpretation of both tabular and graphical forms.\cite{18}

Some data mining tools and techniques are listed below (table no-2):

<table>
<thead>
<tr>
<th>Data mining tools</th>
<th>Function</th>
<th>url</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene XproTools</td>
<td>Predictive modeling</td>
<td><a href="http://www.gepsoft.com/">http://www.gepsoft.com/</a></td>
</tr>
<tr>
<td>SPAD</td>
<td>Data management, data mining, predictive analysis clustering</td>
<td><a href="http://www.spad.eu">www.spad.eu</a></td>
</tr>
<tr>
<td>DataDetective</td>
<td>A visual data mining tool with focus on ease of use</td>
<td><a href="http://www.sentient.nl/?dden">http://www.sentient.nl/?dden</a></td>
</tr>
<tr>
<td>Tree Net</td>
<td>Powerful data mining tool, generate accurate models</td>
<td><a href="http://www.salford-systems.com">http://www.salford-systems.com</a></td>
</tr>
</tbody>
</table>

APPLICATION OF MICROARRAY FOR GENE EXPRESSION ANALYSIS

Microarray technique was developed to investigate the differentially expressed genes using complex populations of RNA, DNA and protein.\cite{19} Clarification of these methods now allows the analysis of copy number variations and gene amplification of DNA.\cite{20} Recently microarray method applied to analysis of expression at protein level.\cite{21}

Microarray basically implemented for the following aim

- Analysis of multiple gene expression assay of the tissue or cells
- Pathogen detection
- Alteration and mutation diagnosis of genotyping polymorphisms

Due to some external conditions genetic alteration occurs in the oral mucosa, microarray may be able to detect these changes. It might be helpful to determine the progress of premalignant stages which is turning to malignant stage.\cite{22, 23} Commonly microarray used to detect pathogens and viruses from blood samples. Microarray, newly have been consider to detect inheritable markers and genotyping tools.\cite{24}

**Schematic representation of microarray application**

```
Normal tissue cell
+ Flourescent labeling
Tumor cells
```

Combine both in equal → Hybridize probe → Analyze by bioinformatics tools

Apply data mining technique → Scan
Gene expression analysis can be done by following bioinformatics tools and techniques (table no-3).

<table>
<thead>
<tr>
<th>Microarray techniques</th>
<th>Key features</th>
<th>Application note</th>
<th>url</th>
</tr>
</thead>
</table>

GENOMICS AND PROTEOMICS POTENTIAL TOOLS IN ORAL CANCER

With the arrival of array methods, provided a high throughput approach for the global study of RNA for mRNA expression and DNA for genomic polymorphisms. Additionally, epigenetic alterations such as global hypomethylation and promoter hypermethylation of CpG rich fields have also been detected. Proteomic technologies and mass spectrometry have provided the tools for analysis of protein profiles of OPLs and OSCCs. Biological fluids (serum, plasma, saliva and urine) can be use for the detection of biomarkers. By applying multiple analysis methods these biomarkers can use, not only for early detection of oral cancer, but also for monitoring prognosis of the disease and response to therapy.[25] In a study, used array CGH to diagnose genome-wide changes in microdissected premalignant stage and metastatic OSCCs. Genomic alterations (TGFβ2, cellular retinal binding protein 1 gene, PIK3CA, HTR1B, HRAS, ERBB3 and STK6) differed between premalignant stages to
metastatic cells. Selecting saliva as a diagnostic tool is that it contains the cells in oral cavity which allow saliva to be first choice of screening and identification for potentials of biomarkers for oral cancer. Salivary tools for detection changes in particular salivary molecules such as proteins or nucleic acids. The fact have analyzed genomic and proteomic targets such as DNA aberrations, mRNA, enzymes, cytokines, growth factors etc in oral cancer.

CONCLUSION

Previous few years, a drastic increase in our understanding of basic mechanisms of controlling the normal and aberrant cell growth. Particularly of those molecular processes or alterations that may be responsible for the acquisition of the transformed and metastatic phenotype. Recent advances in DNA sequencing and microarray technologies among other have enabled a unique opportunity for thousands of genes to monitor simultaneously. Data mining techniques are used because of data is increasing drastically. Bioinformatics provided various tools and techniques to analyze the gene expression, gene alteration detection, network pathway. Through these techniques it makes analysis easy not only for the detection of oral cancer but also for the prognosis of disease and response to the therapy. Genomic and proteomics are contributing to change the practices of dentistry and oral pathology. Bioinformatics tools and techniques made analysis of genetic alteration and detection of new altered gene in the oral cancer.

REFERENCES


