International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 41 No. 4 Feb. 2024, pp. 1098-1108

http://www.ijias.issr-journals.org/

Role of Computational Biology in Oral Science

Deepak T.A.[1], Anulekha C.K.[2], Suchindra Suchindra[3], Avinash Tejasvi[4], and Mariyam Nadhira[5]

[1]Department of Oral Medicine and Radiology, V.S Dental College and Hospital, Bangalore, India

[2]Department of Prosthodontics, Crown and Bridge, Kameneni Institute of Dental Science, Narketpally, India

[3]Department of Engineering, National Institute of Mental Health and Neurosciences, Karnataka State Govt, Bangalore, India

[4]Department of Oral Medicine and Radiology, Kameneni Institute of Dental Science, Narketpally, India

[5]Department of Computer Science, The Maldives National University, Male, Maldives

ABSTRACT: DNA sequence Cigarette Smoking, Betel leaf chewing, and alcohol consumption are major cause of oral cancer in Asia. The difficulty in quitting, coupled with patients' economic conditions affects the inability to get diagnosed early, driving death rate higher. There has been major advancement in molecular sciences, computational biology, and other fields today, but we are not still able to pinpoint the causes of oral cancer, also known as Squamous Cell Carcinoma (OSCC). Early detection leads to better survival rate, therefore, education on yearly check-ups plays a vital role. Computational analysis at the genomic (DNA sequence) can help patients with targeted cellular treatment and hopefully a cure. In this paper, we would look at computation tools used in detecting OSCC and various analysis. Analysis includes detecting abnormality in the cell and other molecular reactions which later morph into a cancerous cell. Later, we investigate all computational tools or techniques from local and global sequence alignment, protein structure, gene, functional structure analysis which help medical staff detect cancer, which in turn can help with oral cancer treatment, prognosis and hopefully a cure.

KEYWORDS: Computational biology, Sequence alignment, Genomic analysis, Gene.

1 INTRODUCTION

Oral squamous cell carcinoma (OSCC) is one of the common types of cancer of head and neck squamous cell carcinoma (HNSCC) today accounting to 500,000 new cases diagnosed and approximately every 250000 deaths every year [1]. Its prevalence is different in different continents or places. the oral cancer is most prevalent in the Asian continent than in the industrialized, more informed western countries where it accounts for 1- 2% of the overall cancer [2]. Oral cancer mortality is highest in Asia unfortunately [3]. OSCC forms almost80 - 90% of all head and neck cancer in the Asian continent [3]. There are many factors which lead to the oral cancer. in the western world, its smoking and alcohol [4].

In Asia, however, the consumption of local liquor, smokeless tobacco and cigarettes with no filters are the leading cause of oral cancer [5]. Gene mutation and the activation of certain genes in the cell are seen as early stage of oral cancer but it is not yet proved categorically [6], [7]. OSCC is commonly seen in elderly males in the larynx and lips area, accounting to almost 95% of the region [8]. With a 50% survival rate after detection [7], the need to develop newer tools and technologies for early faster detection is greater today [3]. In the next section, we will be talking about the role of computational biology or bioinformatics in diagnosis, and treatment of OSCC.

Corresponding Author: Suchindra Suchindra

2 COMPUTATIONAL BIOLOGY

Computational biology or Bioinformatics is a field which encompasses both biology and computer science. This field provides various tools in storing big data sometimes running into petabytes of data stored in databases, and efficiently retrieving the data from the database to establish a relation, provide a 3d structural analysis of proteins and today throws light on the cellular functionality in a much deeper level than what it was prior. This is truer after the genome project was started in the early 1990's which exploded the amount of data stored [9]. The field can also be viewed as a field which brings more analytical power, which in turn would make biological data analysis easier production more biological knowledge [10], [11].

In other words, Computational biology enables scientists to go through huge data easily, finding useful encoded gene segments (for example, a human genome is 3 billion characters in length, a gene could be 100k character in length) [12]. This information is then used in targeted therapy and pharmaceuticals to develop targeted drugs. One of the most pervasive uses of genome sequencing is used to find hereditary relationships and what ethnicity make is one made up of, also known as DNA test / DNA paternity test.

Bioinformatics should not be mistaken to only confined to DNA or genome research augmented study, but it is field which touches a medical science including Molecular Sciences, Drug development, nursing, oral medicine, and other associated fields augmented with pure computer science, biostatistics, and information sciences like State health Care systems or Hospital management systems [9], [13].

Computational biology involves developing new data structures, algorithms using either newer and efficient data structures or underpinning hardware coupled with new algorithms to efficiently store new data in the databases, or efficiently retrieve biological relevant data from a huge database. These involve and can vary depending on the algorithm, data mining techniques and visualization or image processing algorithms based off the known biological data from the biologist and lab technicians. Some of the biological niches today are in sequence alignment tools (local and global), multiple sequence alignments, genome development, Gene alignment, protein structure (2D and 3D image development), phylogenetic tree development, and prediction like protein – protein interactions and gene expression. Most of these predictions are based off existing prediction models which are modified based on the known biological knowledge from researchers over the years [14].

3 COMPUTATIONAL BIOLOGY IN ORAL SCIENCE

Oral Science computational biology is a niche field inside Bioinformatics as it is now catering to only head and neck although it still touches upon other molecular sciences like Genome, Gene expression, protein expression, Protein structure and Sequencing methods. Sequencing and Gene sequencing is probably more useful to dentist today than any other bioinformatic tool. Oral science tools will be discussed in detail in the subsequent sections.

Today MEDLINE (Medical Literature Analysis and Retrieval System Online) houses large number of published articles on health. MEDLINE also houses bibliographic information on articles ranging from oral science, general medicine, nursing, general health care, and veterinary science. It is an open source, and most articles can be searched through PubMed. Like MEDLINE [9], there is Latin American one, called The Literature Latino-Americana e do Caribe em Ciências da Saúde (in Portuguese), or LILACS which was called previously called Latin American Index Medicus. There is one other website which is called 'Medscape' which provides health science information for clinicians and medical scientists. Medscape provides education to health care practitioners including dentists and nurses (other health care professionals) [15]. The primary goal of any oral computational biologists is to improve oral care, which includes better and efficient diagnosis, treatment and prevention of disease and maintain overall good oral health. In regions like Asia, where one needs to see the economical angle, Oral computational biology should also strive to delivery cost effective oral care at the same time [16]. Today, we see that, oral science practitioners and scientists are using new technologies which have come out of this fusion between oral / health science and computer science [17].

Oral science pathology is one of the streams of many including orthodontics, periodontics, and others. Oral Pathology is a specialty mostly dealing with the diagnosis and understanding the causes of diseases related to head and necks, jaws, and face overall. Oral pathology is not limited to above said areas but also includes supporting bone, joints gums, tongue, and surrounding tissues. Knowledge of the above is acquired by learnt methods and disciplines which are basic to oral science practice, through microscopic anatomy, microbiology, and physiology, either through literature or knowledge acquired over a period while practicing dentistry on patients. Oral Pathology uses human science best practices, tools, and computational algorithms to diagnose and treat any diseases or abnormalities in the region [18]. Oral pathology especially in Asian, more so in South Asia, takes into consideration current and pass history coupled with diet, hygiene practice, economic condition, and lifestyle risks like smoking, areca nut chewing, alcohol consumption and recently fad called 'vaping' [9].

The abnormalities seen today are mouth sore that fails to heal or bleeds, to reddish and whitish patches inside the mouth, a lump which is growing quickly and thickening, sore throat leading to difficulty in chewing or swallowing. All the said abnormalities could be seen inside the mouth, face, or neck. Such abnormalities could be diagnosed as oral cancer, cysts, tumour (both benign or malignant), salivary gland abnormalities or cancer (benign or malignant) or simple diseases or lesions. Over the years, large data and literature have been published with varying sites, varieties of above said abnormalities. Today with the advent of projects like the Genome project [9], oral pathology and computational biology are converging ever so closely and working towards automated, efficient, and better way to diagnose pathology at the early stages before they turn life threatening thereby increasing the chances of survival or in other cases, improving the lifestyle of the patient.

4 COMPUTATIONAL TOOLS IN ORAL SCIENCE

In this section, we would look at the computational tools available at the oral professional disposal.

4.1 ORAL CANCER DIAGNOSIS

Among head of neck squamous cell carcinoma, oral cancer is most common. It is most prevalent in the males, and we see oral cancer making up to 7% of all cancers in Asia especially in south Asia. The cause for this high rate has been discussed earlier. Oral cancer if detected at an early stage improves the chances of the patient returning back to normal life quickly perhaps with a permanent cure [19] Unfortunately, when the patient presents themself for an examination, more than likely, the cancer would have spread to other regions and lymph nodes and there is very little a practitioner could do to the patient [20]. Hence yearly check-ups are advised. Sometimes, the lesions found in the mouth are not easily discernible to experienced practitioner leading to false diagnosis. Although some lesions are detected, the similarities in benign and malignant is very high leading to poor and false diagnosis [21].

OralCDx which is a brush biopsy method developed in early 1950s was successful in detecting oral cancer using computer assisted sample analysis. The biopsy collects 3 layers of cell epithelium of the oral mucus for better results. The biopsy sample is stained with Papanicolaou tint [22], it is then scanned and analysed microscopically with the help of computer which contains an image database containing different stages of abnormal cell morphology [23]. This computer program is powerful enough to detect abnormal cells among thousands of normal cells [22]. Once the computer prints the data of its analysis, a cytopathologist interprets this analysed data and classifies the sample to atypical (uncertain diagnosis), positive for oral cancer, normal cells, and incomplete sample [24].

To reduce the false negative finding, many attempts were made to the improve the analysis of the application. However, it was tuned to present day success by incorporating neural network computers initially developed for the missile defence [9]. This improved the application's ability to search for abnormal cell from its database and printed out a detailed data for the cytopathologist to interpret this detailed data in an efficient way there by reducing the chances of overlooking any abnormal cells today.

4.2 GENOME RESEARCH

Genome is defined as the entire genetic markup of an organism or sum of all genes excluding the part of the DNA which we have still not understood [25]. It can also be looked upon as a blueprint for all cellular activities including the protein structure, the region where the gene is expressed inside each cell. This blueprint contains entire set of instructions for new cell creation, construction, operation, maintenance, and repair in an organism [26]. In 2003, human genome project was initiated with a n objective to obtain and study genes, its structure and interactions amongst the genes [9]. Other genome projects were started much earlier like the Drosophila genome, which is comparatively shorter than the human genome which runs to 3 billion characters in length. All genome projects were initiated with a goal to diagnose and prevent diseases. Genome should not be confused with genetics which is a study of genes in an organism.

With new genomic discoveries, researchers and clinicians are throwing more light into our understanding of oral science. These new discoveries coupled with faster computer which are assembled to a supercomputer are helping us understand the oral and dental diseases like never before. Precision diagnostic tools, procedures, tests with new practical based research are throwing new research vectors providing novel avenues to diagnose, treatment (drug discovery) and improve the oral health.

Head and Neck Squamous Cell Carcinoma (HNSCC) is a disease with complex gene alterations, where some genes are mutated with an additional protein character or removed from the position in the gene (An indent '-' is associated when a protein is missing from its known location in a mutated gene). HNSS Tumors associated with human papillomavirus infection have a genetic profile different from the tumors associated with excessive tobacco usage [27. CCND1 gene is associated with

oral Squamous cell carcinoma [32]. Overexpression of this CCND1 is a sure sign of aggressive nature, and early recurrence [32]. Over expression has also indicator of the overall effectiveness of radio [28,29] Erb B complex is a cell cycle signaling gene and its overexpression is an indicator for OSCC [30].

Human cells contain tumor suppressor gene and one such gene is PTEN, if absent is the indicator to the SCC. Others include the growth factor MET oncogene. High expression of epithelial proactive cell nuclear antigen (PCNA) is an indicator for poor survival rate and prognosis [30]. Other genes like UGT1A7, GSTMI, GSTT1 and CYP1A1 over expression is most seen in the smoking and tobacco users [31]. Ever since the genome projects started in the early 1990s coupled with the human genome project in 2003, huge amount of data is collected in either full genome sequences which in a human can run up to 3 billion characters in length to smaller sequences of few hundred thousand like drosophila.

Walter Goad and his colleagues were the first to assemble a DNA Sequence Database in their lab at Los Alamos National Library (LANL) which later came to be known as GenBank [58], [9]. European equivalent was later developed, and it is housed in Germany, called EMBL (European Molecular Biology Laboratory) [58]. Japanese who were the first pioneers for genome research always had their own database called the DNA Databank of Japan at Mishima. Today, all these three have collaborated and formed a global International Nucleotide Sequence Database which is where every technician would go from Genome, protein, or genes [9].

4.3 PROTEOMICS AND COMPUTATIONAL SCIENCE

Proteomics is a new word which is commonly used today, but it was first used I the Therapeutic field. The world Proteome was coined by Marc Wilkins in 1995 [33]. Ever since then, the word 'omics' is attached to anything related to protein and its study. Proteomics is little different from genomics in that, a genome study is about genetic markup or a blueprint for all genes, proteins, and all cellular activity, while proteomics is a study of protein, its structure, expression, and its interaction [9]. All the three, are complex because they all depend on the organism need, time, and environmental conditions [34]. A human genome contains code to about 26000 - 32000 proteins [35], but the number of proteins expressed inside our body is about 1 million.

Today there are many procedures inside proteomics ranging from one and two -dimensional gel electrophoresis [36]. gel-free screening technique called multidimensional protein identification. methods [37], isotope labelling with amino acids [38], isotope-coded tags, isobaric tags, and their quantitation [39]. There are other methods like shotgun-proteomics [40], gel electrophoresis [41] and microarray for proteins [42], all these techniques or procedures can be used on cells, tissues, and organs [9]. Today, Large-scale western blot assays [43] and multiple reaction monitoring assays [44], are used for high-throughput processing and it is a throwing new light in our understanding of protein interaction, protein synthesis and protein culture.

Table 1. Protein Structure Databases [58]

Tools	Tool Used for
InterProScan	Used for Protein functional analysis
PfamScan	FASTA sequence searching against Pfam library
HMMER3	Used for searching one or more sequence against a sequence database
Hmmscan	Search other sequences against collection of profiles
Phobius	Transmembrane topology and signal peptides prediction [58]
RADAR	Protein sequence alignment

Computational biology or algorithms are used in the proteomics domain mainly to manage, store, retrieve markers [45]. These markers could be from DNA sequences or genes or proteins [45]. The algorithms collect and hold massive amount of data sometimes in petabytes and are effective to bridging the gaps between other streams like genomics, metabolism, and protein synthesis. This entire process of providing an association between various niche fields which are closely related (genome, genes, proteins, metabolism, and synthesis) is a computational costly process.

By costly, we mean, computer intensive process, lately proteomics has moved on to heuristics rather than dynamic programming algorithms. We will talk about these in detail in the sequence alignment algorithm section. The algorithms effectiveness in terms of sensitivity and making good biological sense is exacerbated by different parameters, the definition of quality and sensitivity definition and the standard data formats. In other words, there is no standard yet when it comes what makes an algorithm effective and what is the standard data on which algorithms are compared with [46]. A collection of reactions inside the cell that have a biological impact are called protein pathways [47], there are many tools to determine these

pathways today [47]. Some of the databases are in place today such are Kyoto Encyclopedia on Genes and Genomes, BioCarta, pathway Reactome and Ingenuity are some of the databases which have knowledge on cell reactions, metabolism, and cell - cell interactions, protein synthesis and protein - protein interaction [48, 49]. PANTHER is another database, which concentrates more on protein analysis based off the evolutionary relationship. This database also works on genomes to extract these relationships [50, 51]. Cancer related pathways and their detection are stored in a database called Netpath [52]. other protein interaction information is stored in BIOGRID and MINT [53, 53, 54, 55]. Most of this information are extracted using sequence format [56, 57], which we will talk later in this paper. Table 1 shows a collection of databases concentrating on structure analysis.

Table 2. Protein Structure Tools [58]

Tools	Description
PyMOL	Viewer and modelling
ERRAT	Verification and evaluation
PROCHECK	A structure validation web server
Swiss model	Homology modeling of a protein whose structure was unknown
RAMPAGE	Examine the quality of protein structure

Protein structure prediction and their classification is used as a tool for predicting OSCC today. To classify and prediction protein structure, well know knowledge base in the form all know protein structures for an organism is developed and stored [45]. Protein data bank (PDB) is one database which houses proteins, DNA and RNAs in a 3D format. Once OSCC is detected, the conforming proteins can then be examined, analysed and either targeted treatment or drug discovery could then be made. Typically, these methods are either NMR Spectroscopy, X-Ray Spectroscopy or Cryoelectronic microscopy [51]. Structural tools are shown in Table 2.

4.4 MICRO ARRAYS

Microarray is a tool which provides a platform to measure the expression of large set of genes simultaneously. It helps in identification of Single-nucleotide polymorphisms (SNPs), Mutations, tumor classification, target genes to suppress tumor, biomarkers, and identification of genes with chemoresistance. All these help in customized drug treatment and drug discovery [9]. A Microarray can be seen as a circuit made up of thousands of microchips, with each chip containing a short, DNA sequences which together would add up to a normal gene and its variants in human population [9], [59]. These microchips are not as advanced as to create their own memory operations like mutual exclusion or locking [94]

When the Microarray were first developed, there were primarily used for research to identify anomalies in Individuals. As time progressed and enough data collected using manual method, Microarray were used later to find how often a mutation in a particular gene would end up with a oral cancer [59]. Regions where the mutation were seen in a gene were later correlated with oral diseases, Tumor (Benign or Malignant) classification or other oral diseases. As computer became powerful both in terms processing power and memory, large features and large portion of the genome were placed in each individual chip thus enabling many types of genes mutations [3]. Microarrays are also used today to study which gene is turned on or off inside each cell in an organism. They are also used to isolate RNA from the samples for measurement [59].

Microarrays find their usefulness in the labs for clinical testing of diseases such as COVID [59]. As a result of their usefulness in determining the gene mutations, they are used to develop customized drug for patients, as genes determine how organisms can react to certain chemical drugs developed [59]. Microarrays are quickly fading away because of faster, efficient sequencing algorithms which are still costlier than Microarray technology [9].

4.5 SEQUENCE ALIGNMENT

A sequence in molecular or genome research could be either a DNA, RNA, or protein sequence [60]. These sequences are made up of amino acid representing characters. A protein sequence is made up of strings of character (A, R, N, D, C, Q, E, G, H, I, L, K, M, F, P, S, T, W, Y, V). Similarly, DNA (A, C, G, T) and RNA (A, C, G, U) [61]. Sequence alignment is a very old paradigm in the microbiology, where the objective is to find regions of similarity between either 2 pair of sequences or multiple sequences. These similar regions can be analysed later, to determine the similar genes, protein structure and other information leading to drug discovery and treatment [60], [62].

Early research depended on manual efforts to align or find regions of similarity between two sequences. Further research classified pairwise sequencing into local and global sequencing. Local sequence alignment is when regions in short burst are aligned together whereas global sequence alignment is interested in finding all regions of similarity in the entire sequence (DNA, RNA, or protein). Two early algorithms developed in 1970s and 1980s using a technique called dynamic programming, they are Smith-Waterman and Needleman (local alignment) and Wunsch algorithms (global alignment) [63], [64]. The drawback of Dynamic Programming algorithms was that they were very slow. To overcome tis shortfall, Heuristic algorithms were developed. Some of the local heuristic-based algorithms are FASTA [65], BLAST [66], Gapped BLAST [67], BLAT [68], BLASTZ [69], and PatternHunter [70]. Similarly, popular heuristic based global sequence alignment algorithms are MUMmer [71], GLASS [72], AVID [73], and LAGAN [74].

To quantify the alignment in heuristic algorithms, a scoring function is used. When characters in each sequence match or mismatches, a score is given to that character alignment. A scoring function was developed for aligned of pairs of characters, then scores of aligned pairs of characters together were added to get the final sequence alignment score [60]. A scoring matric was later developed, which is a n * n matrix in which each corresponding base pair has a score [75]. PAM (Percentage of Acceptable point Mutations per 108 years) series of matrices [76], [77] and BLOSUM (BLOcks SUbstitution Matrix) series of matrices [78] are popular scoring matrices. All heuristic algorithms use three techniques to find the final alignment, they are look-up table, short alignment of length from 8 – 15, called seeds and finally maximum match subsequence. The look up tables are used by FASTA and BLAST family. Seeds and their variety are used by BLAT, PatternHunter, Glass, AVID [60-63]. Maximal match subsequence which are a new technique was introduced in LAGAN, and MASAA family [61, 62, 63, 92, 93].

BLAST, LAGAN and MASAA family are currently the most effective of the alignment algorithms as they are fast (BLAST) or relatively and sensitive (LAGAN and MASAA family). LAGAN and MASAA variants use different techniques, data structures and stitching techniques to arrive at the final alignment. All the above 3 algorithms are used in our labs to find local and global alignments. Pairwise alignment can be used to find multiple sequence alignment. Initial multiple sequence alignment started with pairwise alignment and progressive moved to other sequences building a guide tree [79] and creating an unwanted pair group with arithmetic mean [80]. Guide tree is used to dictate how multiple sequence alignment would progress. Closely related sequences are aligned first and once there are all done, divergent sequences are then aligned on top of them. This process is called progressive multiple alignment method. ClustalW [81], MAFFT [82], Kalign [83], Probalign [84], MUSCLE [85], DIALIGN [86], PRANK [87], FSA [88], T – Coffee [89, 90], and Probcons [91] are use progressive methods. The other method is iterative method, as the name implies, the algorithm repeatedly applies Dynamic programming on an already sequenced pair to eliminate errors [96, 97].

Pairwise sequence alignment is used to detect chronic diseases like COVID - 9, oral diseases from benign to Malignant [60] and multiple sequence alignment together are used for protein structure prediction, protein-protein interaction, predicting cellular regions where genes would be expressed, Phylogenetics trees (DNA make up, paternity tests). Today artificial intelligence (AI) is also helping sequencing algorithms to predict tumour growth, or mutations in the head and neck regions [95].

5 CONCLUSION

Computational biology has played a pivotal role in our understanding and treatment of oral science. The potential to quickly diagnosis a benign tumour to malignant is not far from way with the way computation biology is progressing today. There are certain fields which may go extinct as time progresses, Microarrays is one of them. Currently the only advantage it has versus sequencing algorithms is the cost. When more commercial sequencing algorithms come into the field, the cost would dramatically come down. Genomics and proteomics will be the future as genes, and their expression (depends on the region, need and environment) is not fully understood. To understand more in detail, we need to convert huge amount of data into knowledge, this means computational ideas, computer hardware and techniques. The future is bright for drug discovery, customized treatment (gene therapy [98]) and a combination of new cellular discovery knowledge and computational prediction model would unlock better results which perhaps might lead to oral cancer / cancer cure.

ACKNOWLEDGMENTS

The authors would like to thank NIMHANS and their respective universities.

REFERENCES

[1] Capparuccia L, Tamagnone L: Semaphorin signaling in cancer cells and in cells of the tumor microenvironment--two sides of a coin. J Cell Sci. 2009; 122 (Pt 11): 1723–36. 10.1242/jcs.030197.

[2] Joshi P, Dutta S, Chaturvedi P, et al.: Head and neck cancers in developing countries. Rambam Maimonides Med J. 2014; 5 (2): e0009. 10.5041/RMMJ.10143.

[3] Al-Jaber A, Al-Nasser L, El-Metwally A: Epidemiology of oral cancer in Arab countries. Saudi Med J. 2016; 37 (3): 249–55. 10.15537/smj.2016.3.11388.

[4] Raham S, Dayal H, Rohrer T, et al.: Dentition, diet, tobacco, and alcohol in the epidemiology of oral cancer. J Natl Cancer Inst. 1977; 59 (6): 1611–8. 10.1093/jnci/59.6.1611.

[5] Muttagi SS, Chaturvedi P, Gaikwad R, et al.: Head and neck squamous cell carcinoma in chronic areca nut chewing Indian women: Case series and review of literature. Indian J Med Paediatr Oncol. 2012; 33 (1): 32–5. 10.4103/0971-5851.96966.

[6] Krishna A, Singh S, Kumar V, et al.: Molecular concept in human oral cancer. Natl J Maxillofac Surg. 2015; 6 (1): 9–15. 10.4103/0975-5950.168235.

[7] Neville BW, Damm DD, Allen CM, et al.: Oral and maxillofacial pathology. St. Louis, Mo: Saunders/Elsevier.2009.

[8] Weatherspoon DJ, Chattopadhyay A, Boroumand S, et al.: Oral cavity and oropharyngeal cancer incidence trends and disparities in the United States: 2000-2010. Cancer Epidemiol. 2015; 39 (4): 497–504. 10.1016/j.canep.2015.04.007

[9] Singaraju S, Prasad H, Singaraju M. Evolution of dental informatics as a major research tool in oral pathology. J Oral Maxillofac Pathol. 2012 Jan; 16 (1): 83-7. doi: 10.4103/0973-029X.92979. PMID: 22434944; PMCID: PMC3303529.

[10] Westhead DR, Parish JH, Twyman RM. Bioinformatics. 1st ed. UK: BIOS Scientific Publishers Limited; 2003.

[11] Hwa A Lim. Genetically yours: Bioinforming, biopharming, and biofarming. New Jersey: Word Scientific Publishing Co; 2002. p. 273-95.

[12] Srivastava. Introduction to bioinformatics. 1st ed. India: Shree Publications; 2007.

[13] Bansal M. Medical informatics: A primer. 1st ed. India: Tata Mcgraw-Hill; 2003.

[14] Rao VS, Das SK, Rao VJ, Srinubabu G. Recent developments in life sciences research: Role of bioinformatics. Afr J Biotechnol 2008: 7; 495-503.

[15] White SC. Decision-support systems in dentistry. J Dent Educ 1996; 60: 47-63.

[16] Kalkwarf KL. How the licensure process will evolve. J Am Dent Assoc 1999; 130: 1737-42.

[17] Schleyer TK, Forrest JL, Kenney R, Dodell DS, Dovgy NA. Is the Internet useful for clinical practice? J Am Dent Assoc 1999; 130: 1501-11.

[18] Rajendran R, Sivapathasundaram B. Shafer's Textbook of Oral Pathology. 6th ed. India; Elsevier; 2009.

[19] National Cancer Institute. Cancer statistics review, 1973-1990. Bethesda, Md.: U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health; DHHS publication (NIH) 93-2789; 1993.

[20] Silverman S. American Cancer Society. Oral cancer. Hamilton, Ontario, Canada: B.C. Decker; 1998: xvi, 174.

[21] Silverman S Jr. Early diagnosis of oral cancer. Cancer 1988; 62 (8 Suppl): 1796-9.

[22] Sciubba JJ. Improving detection of precancerous and cancerous oral lesions. Computer-assisted analysis of the oral brush biopsy. U.S. Collaborative OralCDx Study Group. J Am Dent Assoc. 1999; 130: 1445–1457.

[23] Patton LL, Epstein JB, Kerr AR. Adjunctive techniques for oral cancer examination and lesion diagnosis: a systematic review of the literature. J Am Dent Assoc. 2008; 139: 896–905.

[24] Scheifele C, Schmidt-Westhausen AM, Dietrich T, Reichart PA. The sensitivity and specificity of the OralCDx technique: evaluation of 103 cases. Oral Oncol. 2004; 40: 824–828.

[25] Little PF. Structure and function of the human genome. Genome Res 2005; 15: 1759-66.5.

[26] Yeager AL. Where will the genome lead us? Dentistry in the21st century. J Am Dent Assoc 2001; 132: 801-7.

[27] Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, et al. The mutational landscape of head and necksquamous cell carcinoma. Science 2011; 333: 1157-60.29.

[28] Huang M, Spitz MR, Gu J, Lee JJ, Lin J, Lippman SM, et al. CyclinD1 gene polymorphism as a risk factor for oral premalignant lesions. Carcinogenesis 2006; 27: 2034-7.30.

[29] Shintani S, Mihara M, Ueyama Y, Matsumura T, Wong DT.Cyclin D1 overexpression associates with radiosensitivity in oralsquamous cell carcinoma. Int J Cancer 2001; 96: 159-65.31.

[30] Aswini YB. The genomics of oral cancer and wound healing.J Indian Soc Pedod Prev Dent 2009; 27: 2-5.32.

[31] Pavanello S, Clonfero E. Biological indicators ofgenotoxic risk and metabolic polymorphisms. Mutat Res2000; 463: 285-308.

[32] Govindraju, Poornima & Chandra, Poornima. (2015). Implications of genomics in oral health. Journal of Advanced Clinical & Research Insights. 2. 147-150. 10.15713/ins.jcri.65.

[33] Agrawal GK, Sarkar A, Righetti PG, Pedreschi R, Carpentier S, Wang T, Barkla BJ, Kohli A, Ndimba BK, Bykova NV, Rampitsch C, Zolla L, Rafudeen MS, Cramer R, Bindschedler LV, Tsakirpaloglou N, Ndimba RJ, Farrant JM, Renaut J, Job D, Kikuchi S,

Rakwal R. A decade of plant proteomics and mass spectrometry: translation of technical advancements to food security and safety issues. Mass Spectrom Rev. 2013; 32: 335–365.
[34] Holman JD, Dasari S, Tabb DL. Informatics of protein and posttranslational modification detection via shotgun proteomics. Methods Mol Biol. 2013; 1002: 167–179.
[35] Chandramouli K, Qian PY. Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. Hum Genomics Proteomics. 2009; 2009.
[36] Vercauteren FG, Bergeron JJ, Vandesande F, Arckens L, Quirion R. Proteomic approaches in brain research and neuropharmacology. Eur J Pharmacol. 2004; 500: 385–398.
[37] Florens L, Washburn MP. Proteomic analysis by multidimensional protein identification technology. Methods Mol Biol. 2006; 328: 159–175.
[38] Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics. 2002; 1: 376–386.
[39] Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A, Pappin DJ. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics. 2004; 3: 1154–1169.
[40] Wolters DA, Washburn MP, Yates JR 3rd. An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem. 2001; 73: 5683–5690.
[41] Klose J, Nock C, Herrmann M, Stühler K, Marcus K, Blüggel M, Krause E, Schalkwyk LC, Rastan S, Brown SD, Büssow K, Himmelbauer H, Lehrach H. Genetic analysis of the mouse brain proteome. Nat Genet. 2002; 30: 385–393.
[42] Cutler P. Protein arrays: the current state-of-the-art. Proteomics. 2003; 3: 3–18.
[43] Schulz TC, Swistowska AM, Liu Y, Swistowski A, Palmarini G, Brimble SN, Sherrer E, Robins AJ, Rao MS, Zeng X. A large-scale proteomic analysis of human embryonic stem cells. BMC Genomics. 2007; 8: 478.
[44] Stahl-Zeng J, Lange V, Ossola R, Eckhardt K, Krek W, Aebersold R, Domon B. High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. Mol Cell Proteomics. 2007; 6: 1809–1817.
[45] Vihinen M. Bioinformatics in proteomics. Biomol Eng. 2001; 18: 241–248.
[46] Domon B, Aebersold R. Challenges and opportunities in proteomics data analysis. Mol Cell Proteomics. 2006; 5: 1921–1926.
[47] Aslam B, Basit M, Nisar MA, Khurshid M, Rasool MH. Proteomics: Technologies and Their Applications. J Chromatogr Sci. 2017; 55: 182–196.
[48] Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, Caudy M, Garapati P, Gopinath G, Jassal B, Jupe S, Kalatskaya I, Mahajan S, May B, Ndegwa N, Schmidt E, Shamovsky V, Yung C, Birney E, Hermjakob H, D'Eustachio P, Stein L. Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res. 2011; 39: D691–D697.
[49] Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 2012; 40: D109–D114.
[50] Mi H, Guo N, Kejariwal A, Thomas PD. PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways. Nucleic Acids Res. 2007; 35: D247–D252.
[51] Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, Buetow KH. PID: the Pathway Interaction Database. Nucleic Acids Res. 2009; 37: D674–D679.
[52] Kandasamy K, Mohan SS, Raju R, Keerthikumar S, Kumar GS, Venugopal AK, Telikicherla D, Navarro JD, Mathivanan S, Pecquet C, Gollapudi SK, Tattikota SG, Mohan S, Padhukasahasram H, Subbannayya Y, Goel R, Jacob HK, Zhong J, Sekhar R, Nanjappa V, Balakrishnan L, Subbaiah R, Ramachandra YL, Rahiman BA, Prasad TS, Lin JX, Houtman JC, Desiderio S, Renauld JC, Constantinescu SN, Ohara O, Hirano T, Kubo M, Singh S, Khatri P, Draghici S, Bader GD, Sander C, Leonard WJ, Pandey A. NetPath: a public resource of curated signal transduction pathways. Genome Biol. 2010; 11: R3.
[53] Schmidt A, Forne I, Imhof A. Bioinformatic analysis of proteomics data. BMC Syst Biol. 2014; 8 Suppl 2: S3.
[54] Chatr-aryamontri A, Ceol A, Palazzi LM, Nardelli G, Schneider MV, Castagnoli L, Cesareni G. MINT: the Molecular INTeraction database. Nucleic Acids Res. 2007; 35: D572–D574.
[55] Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, Duesbury M, Dumousseau M, Feuermann M, Hinz U, Jandrasits C, Jimenez RC, Khadake J, Mahadevan U, Masson P, Pedruzzi I, Pfeiffenberger E, Porras P, Raghunath A, Roechert B, Orchard S, Hermjakob H. The IntAct molecular interaction database in 2012. Nucleic Acids Res. 2012; 40: D841–D846.
[56] Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013; 41: D808–D815.
[57] Glaab E, Baudot A, Krasnogor N, Schneider R, Valencia A. EnrichNet: network-based gene set enrichment analysis. Bioinformatics. 2012; 28: i451–i457.

[58] Rath, Sonali & Das, Moumita & Kar, Dattatreya & Goel, Ruchi. (2021). Computational Genomic Analysis of Oral Squamous Cell Carcinoma - A New Diagnostic Approach. 13. 85-90. 10.31782/IJCRR.2021.SP247.
[59] https://www.genome.gov/about-genomics/fact-sheets/DNA-Microarray-Technology.
[60] Reddy, Bharath Govinda. Multiple Anchor Staged Local Sequence Alignment Algorithm-MASAA. Diss. The University of Northern British Columbia, 2009.
[61] Waqar Haque, Alex Aravind, and Bharath Reddy. 2009. Pairwise sequence alignment algorithms: a survey. In Proceedings of the 2009 conference on Information Science, Technology and Applications (ISTA '09). ACM, New York, NY, USA, 96-103. DOI=http://dx.doi.org/10.1145/1551950.1551980.
[62] Haque, W., Aravind, A., & Reddy, B. (2008, August). An efficient algorithm for local sequence alignment. In 2008 30th annual international conference of the IEEE engineering in medicine and biology society (pp. 1367-1372). IEEE.
[63] Smith, T.F. and Waterman, M.S. 1981. Identification of common molecular subsequences. J. Mol. Biol. 147: 195-197.
[64] Needleman, S.B. and Wunsch, C.D. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48: 443-453.
[65] D. J. Lipman and W. R. Pearson, «Rapid and Sensitive Protein Similarity Searches,» Science, vol. 227, pp. 1435-1441, 1985.
[66] S. F. Itschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, «Basic Local Alignment Search Tool,» J. Molecular Biology, vol. 215, pp. 403-410, 1990.
[67] S. F. Altschul, et al., «Gapped BLAST and PSI-BLAST: a new generation of protein database search programs,» Nucleic Acids Res., vol. 25, pp. 3389-3402, Sep. 1997.
[68] W. J. Kent, «BLAT—The BLAST-Like Alignment Tool,» Genome Research, vol. 12, pp. 656-664, 2002.
[69] S. Schwartz, et al., «Human–Mouse Alignments with BLASTZ,» Genome Research, vol. 13, pp. 103-107, 2003.
[70] B. Ma, J. Tromp, and M. Li, «Pattern-hunter: faster and more sensitive homology search,» Bioinformatics, vol. 18, pp. 440-445, 2002.
[71] A. L. Delcher, et al., «Alignment of whole genomes,» Nucl. Acids Research, vol. 27, pp. 2369-2376, 1999.
[72] L. Batzoglou, J. Pachter, B. Mesirov, B. Berger, and E. S. Lander, «Human and mouse gene structure: comparative analysis and application to exon prediction,» in RECOMB '00: Proc of the 4th Int'l Conference on Computational Molecular Biology, 2000, pp. 46-53.
[73] N. Bray, I. Dubchak, and L. Pachter, «AVID: A Global Alignment Program,» Genome Research, vol. 13, pp. 97-102, 2003.
[74] M. Brudno and B. Morgenstern, «Fast and sensitive alignment of large genomic sequences,» in Proc of IEEE Computer Science Bioinformatics Conference, 2002, pp. 138-147.
[75] Pittsburgh Supercomputing Center. (2007)
[Online]. http://www.psc.edu/research/biomed/homologous/ scoring_primer.html.
[76] D. J. States, W. Gish, and S. F. Altschul, «Improved Sensitivity of Nucleic Acid Database Search Using Application-Specific Scoring Matrices,» METHODS: A Companion to Methods in Enzymology, vol. 3, no. 1, pp. 66-70, 1991.
[77] M. O. Dayhoff, R. M. Schwartz, and B. C. Orcutt, «A Model of Evolutionary Change in Proteins,» in Atlas of Protein Sequence Structure Nat'l Biomedical Res., M. O. Dayhoff, Ed. 1978, ch. 5, pp. 345-352.
[78] S. Henikof and J. G. Henikof, «Amino acid substitution matrices from protein blocks,» in Proc Natl Acad Sci, 1992, pp. 10915-9.
[79] N. Saitou and M. Nei, «The neighbor-joining method: a new method for reconstructing phylogenetic trees,» Molecular Biology and Evolution, vol.4, no.4, pp.406–425,1987.
[80] I. Gronauand S. Moran, «Optimal implementations of UPGMA and other common clustering algorithms,» Information Processing Letters, vol.104, no.6, pp.205–210,2007.
[81] J. D. Thompson, D. G. Higgins, and T. J. Gibson, «CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice,» Nucleic Acids Research, vol. 22, no.22, pp.4673–4680,1994.
[82] K. Katoh and D. M. Standley, «MAFFT multiple sequence alignment software version 7: improvements in performance and usability,» Molecular Biology and Evolution, vol. 30, no. 4, pp.772–780,2013.
[83] T. Lassmann and E. L. L. Sonnhammer, «Kalign—an accurate and fast multiple sequence alignment algorithm,» BMC Bioinformatics, vol.6, article298,2005.
[84] U. Roshan and D. R. Livesay, «Probalign: multiple sequence alignment using partition function posterior probabilities,» Bioinformatics, vol.22, no.22, pp.2715–2721,2006.
[85] R.C. Edgar, «MUSCLE: a multiple sequence alignment method with reduced time and space complexity,» BMC Bioinformatics, vol.5, article113,2004.
[86] B. Morgenstern, «DIALIGN: multiple DNA and protein sequence alignment at BiBiServ,» Nucleic Acids Research, vol.32, supplement2, pp. W33–W36,2004.
[87] A. L¨oytynoja and N. Goldman, «Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis,» Science, vol.320, no.5883, pp.1632–1635,2008.

[88] R. K. Bradley, A. Roberts, M. Smoot et al., «Fast statistical alignment,» PLoS Computational Biology, vol. 5, no. 5, Article IDe1000392,2009.

[89] P. Di Tommaso, S. Moretti, I. Xenarios et al., «T-Coffee: a webserver for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension,» Nucleic Acids Research, vol. 39, supplement 2, pp. W13–W17,2011.

[90] C. Notredame, D.G. Higgins, and J. Heringa, «T-coffee: a novel method for fast and accurate multiple sequence alignment,» Journal of Molecular Biology, vol.302, no.1, pp.205–217,2000.

[91] C.B.Do, M.S.P. Mahabhashyam, M. Brudno, and S. Batzoglou, «ProbCons: probabilistic consistency-based multiple sequence alignment,» Genome Research, vol.15, no.2, pp.330–340,2005.

[92] B. Reddy and R. Fields, «Multiple Anchor Staged Alignment Algorithm – Sensitive (MASAA –S),» 2020 3rd International Conference on Information and Computer Technologies (ICICT), San Jose, CA, USA, 2020, pp. 361-365, doi: 10.1109/ICICT50521.2020.00064.

[93] Reddy, B., & Fields, R. (2023). Maximum Match Subsequence Alignment Algorithm Finely Grained (MMSAA FG). arXiv e-prints, arXiv-2305.

[94] Bharath Reddy and Richard Fields, «Techniques for Reader-Writer Lock Synchronization,» International Journal of Electronics and Electrical Engineering, Vol. 8, No. 4, pp. 63-73, December 2020. doi: 10.18178/ijeee.8.4.63-73.

[95] Bharath Reddy and Richard Fields. 2022. From past to present: a comprehensive technical review of rule-based expert systems from 1980 -- 2021. In Proceedings of the 2022 ACM Southeast Conference (ACM SE '22). Association for Computing Machinery, New York, NY, USA, 167–172. https://doi.org/10.1145/3476883.3520211.

[96] Reddy, Bharath. (2020). BIOINFORMATICS & PAIRWISE SEQUENCE ALIGNMENT: Local and Global Sequence Alignment Algorithms. Amazon publications, May, 2020, https://www.amazon.com/BIOINFORMATICS-PAIRWISE-SEQUENCE-ALIGNMENT-Algorithms-ebook/dp/B0879F5B5T/ref=sr_1_1?crid=2WEBAOL36MKNC&keywords=sequence+alignment+bharath&qid=1696737104&sprefix=sequence+alignment+bharath+%2Caps%2C295&sr=8-1.

[97] eddy, B., Fields, R. (2022). Multiple Sequence Alignment Algorithms in Bioinformatics. In: Zhang, YD., Senjyu, T., So-In, C., Joshi, A. (eds) Smart Trends in Computing and Communications. Lecture Notes in Networks and Systems, vol 286. Springer, Singapore. https://doi.org/10.1007/978-981-16-4016-2_9.

[98] Gonçalves GAR, Paiva RMA. Gene therapy: advances, challenges and perspectives. Einstein (Sao Paulo). 2017 Jul-Sep; 15 (3): 369-375. doi: 10.1590/S1679-45082017RB4024. PMID: 29091160; PMCID: PMC5823056.

AUTHORS

	Dr. Deepak T.A is professor and head of the department of Oral medicine and Radiology at V. Dental College and Hospital. He has been a full tenured professor for 10 years and his research areas are in oral medicine and oral cancer. Lately, he is exploring and developing computational tools to detect oral cancer.
	Dr. Anulekha C.K is a professor and head of the department of Prosthodontics, also known as dental prosthetics or prosthetic dentistry. She is a tenure track professor at Kameneni Institute of Dental Science. Her research lately is into oral cancer, prosthedontics and its impact on oral cancer, computational tools development and dentals sciences.
	Suchindra is an Engineering from Bangalore University. He is currently pursuing neuroscience and genome research at National Institute of Mental Health and Neurosciences. His research areas are Brain haemorrhages, Autism and Genome research. He has been active researcher for 10 years. He is working for Karnataka State Govt.
	Dr. Avinash is a professor and head of the department of Oral Medicine and Radiology at the Kameneni insitiute of Dental Science. He has been a full tenured professor for 7 seven years and focus most of his research on oral medicine and oral cancer.
	Ms Mariyam Nadhira is senior instructor of Computer science at The Maldives National University in the department of computer science. Her research is in Biotechnology and Biosciences, and Data analysis.