Review Article
Journal of Pierre Fauchard Academy (India Section), *Vol 37(1-2), DOI: 10.18311/jpfa/2023/32255, January-June 2023 p. 01-08*
ISSN (Print) : 0970-2199
ISSN (Online) : 2405-772X

Bioinformatics and Experimental Databases as Significant Research Paraphernalia in Oral Cancer - A Narrative Review

Deepti Sharma[1*], George Koshy[1] and Vishal Kumar Sharma[2]

[1]Department of Oral and Maxillofacial Pathology, Christian Dental College and Hospital, Ludhiana -141008, Punjab, India; deepti.sharma@cmcludhiana.in
[2]Department of Orthodontics, Desh Bhagat Dental College, Mandi Gobindgarh - 147301, Punjab, India

Abstract

Among head and neck squamous cell carcinomas the incidence of Oral Squamous Cell Carcinoma (OSCC) is highest (80%). Bioinformatics provides a wide-ranging outlook to understand the significant genes, proteins, pathways, and molecular mechanisms fundamental to oral carcinogenesis. It is a promising branch in the field of biomedical research with emerging popularity in the field of dental research too, targeting immune-mediated, developmental, genetic, and malignant diseases. High-throughput experimental technologies have created large and complex genomic data, which has revolutionized genomic research. Advanced technologies including microarray, next-generation sequencing, and bioinformatics together play a decisive role in unraveling the molecular mechanism in the process of Head and Neck cancer. This article is an attempt to give an overview of the significant role of bioinformatics in the field of genomic research.

Keywords: Bioinformatics, Data Mining, Genome, Insilco, Microarrays, Next Generation Sequencing, Oral Cancer

(**Article chronicle: Received:** 28.12.2022; **Revised:** 30.05.2023; **Accepted:** 08.07.2023)

1. Introduction

Past decades have seen a stupendous rise in the literature documenting studies using gene expression signatures based on genomic, transcriptomic, and proteomic profiling, as most genetic variations are viewed at molecular levels[1]. The recent completion of the human genome project has been a decisive factor in the molecular analysis of events that underpin the initiation and progression of the disease process[2]. The altered nucleic acid sequences can potentially be labeled as disease biomarkers that have to be identified, correlated with clinical and biological behavior and targeted for improved patient survival and therapeutics[3].

In today's era of modern technology, huge sets of gene expression data are created and gathered called transcription profiles. Gene expression profiling is important for characterizing molecular signatures, comprehending genetic functions, and identifying drug targets. Rather than focus on the expression of a limited number of genes, broad genomic expression profiles provide an opportunity for researchers to analyze genetic networks and pathways involved in oncogenesis, metabolism of carcinogens, and response to radiations, etc[4].

Head and Neck Squamous Cell Carcinoma (HNSCC) accounts for about 5% of all malignancies and is the sixth most common type of cancer worldwide. Recent studies are aimed to reveal the molecular sub-classifications/ characterization of HNSCC[5]. A large amount of gene expression information is now openly available via public databases as most biological distinctions are now observed at a genomic level. The last two decades have seen a rapid surge in biologically significant datasets; paving the way for a new era in biomedical research[6]. Gene expression profiling by DNA microarrays and Next Generation Sequencing techniques (NGS) has become a significant tool for studying the molecular intricacies of malignancies and other diseases[7].

*Author for correspondence

1.1 Bioinformatics: The New Arena

"**Bioinformatics"** involves deciphering the molecular data generated by high-throughput experimental technologies and incorporates the principles of biology and computer science with the primary objective of making biological processes more comprehensible[8]. The discipline is mainly concerned with retrieving, gathering, and analyzing data from large databases using algorithms and computationally intensive techniques[9]. It involves gene and protein expression analysis, genomic annotation, comparative genomics, network and system biology (Figure 1)[10].

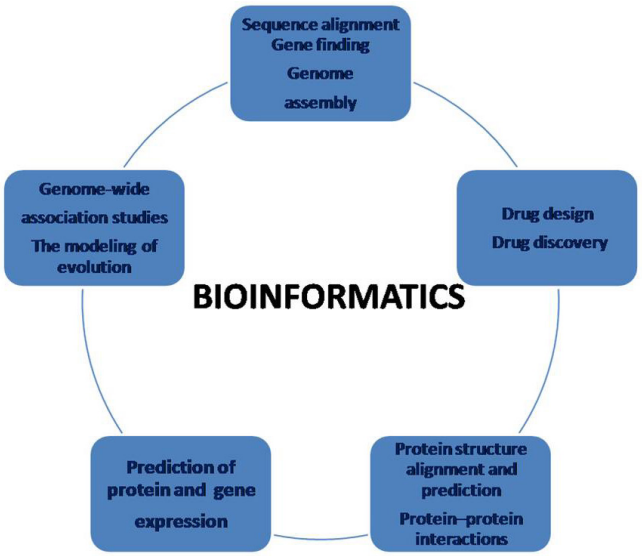


Figure 1. Scope of bioinformatics.

The application of computer and information science to improve dental practice, research, and education is "Dental informatics". In addition, to formulate new hypotheses and plan targeted experimentations, data mining of existing databases could be done[11]. Various objectives that could be targeted through microarray and next-generation sequencing-based technologies via bioinformatics could be an exploration of differentially expressed genes (mRNA and miRNA) in oral cancer and other disorders, hypothetic formulation of a statistical model for a disease process, planning experimentation, predicting prognosis and survival. Bioinformatics enables to manipulation of the data for useful encoded gene segments and utilizing this information for Precision medicine and targeted care[12].

Approximately 30,000 genes constitute the human genome encoding for at least three times a greater number of proteins. The in-depth experimental analysis of genes, proteins, and molecules involved in physiological and pathological processes is made convenient and less complicated by bioinformatics[13].

1.2 The Pillars: Microarray

In silico, analysis means "performed on computer or via computer simulation". In biomedical research in silico studies have the potential to accelerate the rate of discovery, minimizes the need for expensive lab work or clinical trials, and unravel the molecular insights of malignant cell[14]. The successful interpretation of microarray and Next Generation Sequencing (NGS) data require good experimental design and careful statistical analysis[3]. Reverse Transcription Polymerase Chain Reaction, Expressed Sequence Tags, Northern Blot, and Serial Analysis of Gene Expression (SAGE) are the traditional tools for gene expression profiling. The introduction of high throughput microarray, exome, and transcriptome sequencing (RNA-seq) facilitates rapid, comprehensive, and cost-effective quantification of mRNA abundance for multiple genes concurrently[15].

The whole genome can be monitored on a single chip containing thousands of synthetic DNA sequences (probes) via DNA microarrays which are based on the principle of hybridization. The robotically spotted cDNA arrays and short photolithographic oligonucleotide arrays are the two most commonly used microarray platforms with different protocols, designs, and analysis patterns[16]. Microarray types and functions have been explained in (Figure 2 and Table 1).

In Dentistry, DNA- microarrays provide novel insights into oral health and biology. Studies have been carried out to determine microbial signatures in oral diseases (periodontal diseases and caries)[17]. The identification of single nucleotide polymorphisms or chromosomal aberrations that cause the development of many syndromes and diseases is possible through this technology. Genomic fingerprints will also enable us to distinguish innocuous white lesions from potentially malignant lesions or early cancer. Recent studies have shown the effectiveness of microarrays in oral cancer[18]. Analysis of the salivary genome and microbiome transcripts of oral cancer patients thus enhances the diagnosis, prevention, and therapeutic interventions like

Table 1. Applications and differences between microarray and next generation sequencing

S.NO	Microarray	Next Generation Sequencing (NGS)
1.	Microarrays contain a specific set of known targets.	Next-generation sequencing is capable of interrogating the entire genome or transcriptome, not depending on pre-chosen targets.
2.	Microarrays are based on hybridization	NGS is based on synthesis, i.e., the incorporation of nucleotides by a DNA polymerase.
3.	Allows simultaneous detection and analysis of the expression of thousands of genes in a small sample. Exploration of genome-wide transcriptional profiles.	Allows sequencing of big quantity of different DNA sequences at once. Denovo-sequencing (virus, bacteria, fungi, mammalian). Includes whole-exome sequencing, whole-genome sequencing, "hotspot" sequencing, and targeted exome sequencing.
4.	Measurement of similarities or differences in genetic contents among different microbes.	Without genome sequence, deciphering a cell's transcripts.
5.	Study and identification of Single Nucleotide Polymorphism (SNP) in different populations	Structural variation (Identification of all mutations in an organism), metagenomics, methylation analysis
6.	Estimating mutations burden in germ cells. Genotyping for risk assessment	Discovery of Differentially expressed genes, transcripts, exons, and noncoding RNAs from RNA-Seq data
7.	To determine the content of the genome of living organisms via comparative genomic hybridization	Cross hybridization background, and signal saturation-induced detection dynamic range limitation.
9.	Identify drug targets by locating the genes in the pathology	Targeted resequencing (single gene, gene panels, and chromosomal regions)
10.	Identify inheritable markers, and therefore have been used as a genotyping tool. Toxicology studies.	It provides a wider expansion of the knowledge of the relationship between disease pathogenesis and genetic variations.
11.	Differential gene expression in normal versus pathological tissues	To detect early relapse or residual cancer by evaluating circulating tumor cells or free plasma DNA. Other advantages are freedom from predesigned probes.

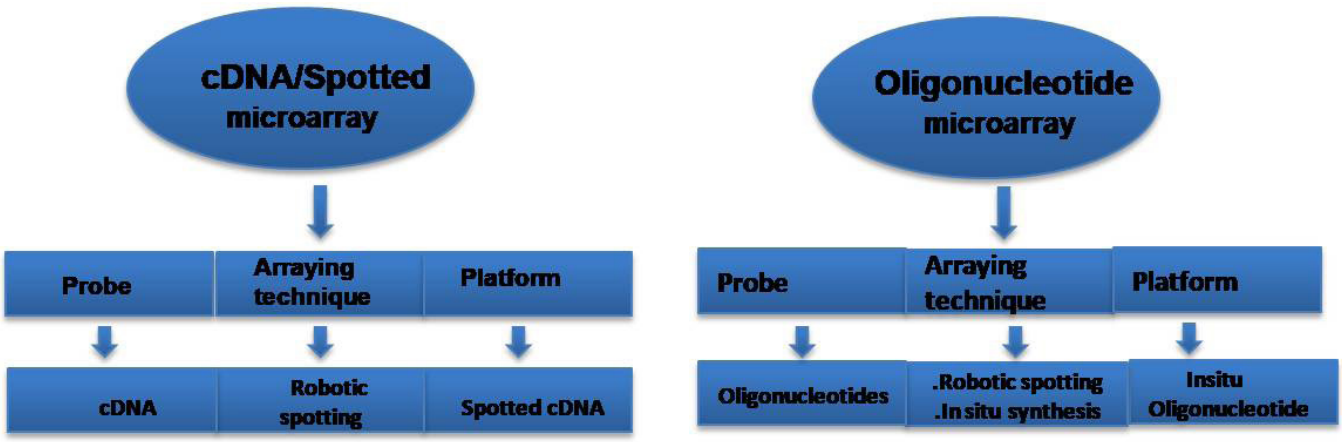


Figure 2. Microarray types.

exploring potential gene therapy[19]. Microarray analysis provides valuable information on disease pathogenesis, genetic predisposition, progression, treatment response, and resistance. Data interpretation may improve diagnosis and therapeutic approaches for cancer. Although DNA microarrays are a useful tool to study the expression

profiles of multiple genes together, but the lack of uniform data mining strategies is a major drawback[20].

1.3 The Future: Next-Generation Sequencing

In the past decades, deep sequencing Next-Generation Sequencing (NGS) technology has emerged as a powerful alternative to microarrays for the high-throughput profiling of gene expression. It has revolutionized a range of biological systems from cancer to immunology with the potential to alter medical practice, research, and genetically targeted therapeutics[21]. NGS is a method of choice for complex large-scale genetic analyses as it combines qualitative and quantitative sequence information generated[22]. The term sequencing refers to elucidating or figuring out the sequence of base pairs in DNA and RNA and the coding that makes up the genetic inheritance is determined by the order of the bases along DNA. Thus, sequencing the DNA will unveil the cause of the various diseases which could then be cured with targeted gene therapy. "The first DNA sequences were obtained by using laboratories methods based on 2- dimensional chromatography in the early 1970s"[23]. The massive sequencing at low cost is enabled by NGS or post-Sanger sequencing technologies. This has encouraged the rapid enlistment of cancer genome variations, chromosomal lesions, and copy-number alterations as well as small-scale aberrations depicted by small insertions, deletions, and point mutations[24]. The unbiased analysis of the transcriptome is allowed by RNA-seq, with dynamic detection range (8,000-fold), single base pair resolution, and minimal background signaling. This technology registers long non-coding RNA, siRNA, miRNA, and other small RNA classes (eg, snRNA and piRNA) which regulates protein translation, RNA stability, or the modulation of chromatin states[22]. The gap between genetic information encoded in DNA and phenotypes is filled by transcriptome representing the entire repertoire of gene transcripts. The systematic analysis of orthologous transcripts across species ensures to bring out important information in animal models of human diseases. However clinical trials are needed to expose the strengths and weaknesses of current NGS techniques and approaches for its implementation as a clinical diagnostic, prognostics, and theragnostic oncology tool[25].

2. Databases, Software, and Algorithms used for High-Throughput Experimental Technologies

The formulation of new hypotheses and exploratory research about biological processes have been instigated through the wide availability of microarray and NGS datasets and a variety of statistical and computational methods[26]. There have been numerous publications stating and interpreting the systematic challenges presented by big data. In the present scenario, molecular investigations of head and neck cancer including oral cancer are aimed at procuring proteomic and genomic profiles and associating the information with clinical and immune status[2]. The bulk of data generated through high throughput experimental technologies needs to be structured efficiently in the form of databases/repositories, so as to exploit the information about diagnostic, prognostic and predictive behavior[27]. These databases provide descriptions of gene lists, gene names, functions, chromosomal location, mutations, Single Nucleotide Polymorphisms (SNPs), mRNA/miRNA expression, protein interactions networks, differential gene expression in different tissues and clinical correlates. Bioinformatics includes the documentation of the research work done, dealing with the databases, software, program versions, algorithms applied and other parameters[28]. Another aspect would be the use of these published databases and methods for future research, evaluation, formulation of new hypotheses and then validating these through experimentation. Further evaluation of the associations and conflicts of the results with other available data sources urges the formulation of different propositions. The Genome Database (GDB) for the human genome is an official central repository for genomic mapping data in Homo sapiens since 1990, now it contains sequence information too[29].

The knowledge about the structure of assembled genomes, assembly names, statistical information, and links to genomic sequence data is provided by the database. The most basic databases include the repositories like National Center for Biotechnology Information (NCBI) DNA Data Bank of Japan (DDBJ), European Bioinformatics Institute (EBI), Stanford database, or Oral Cancer Database (CGDB). Different tools (packages and software) are used for searching and

retrieving information from these repositories[30]. The key NCBI Databases include PubMed/Medline (journal publication search), Genbank/Entrez (nucleotide/protein search), Taxbrowser (full species' taxonomy), GEO (Gene Expression Omnibus), and many other smaller databases. Ensembl is a joint project between GenBank, EMBL, EMBL–EBI and DDBJ have established the International Nucleotide Sequence Database Collaboration (http://www.ncbi.nlm.nih.gov/collab), to facilitate data exchange on a daily basis[31]. For protein information also a number of databases have similarly been established; including Database of Interacting Proteins (DIP) (http://dip.doembi. ucla.edu), Protein Information Resource (PIR) (http://pir.georgetown.edu/) Genomes (KEGG) (http://www.genome.jp/kegg/), Kyoto Encyclopedia of Genes and Biomolecular Interaction Network Database (BIND) (http://www.bind.ca). Data on proteins and nucleic acid structure is stored in the structural databases that include Nucleic Acid Data Base (NADB) (http://ndbserver.rutgers.edu/) and Protein Data Bank (PDB) (http://www.pdb.org/)[8,10].

Bioinformatics and data mining will aid us in comprehending and evaluating the depth of genomics, epigenomics, and proteomics and could play a crucial role in solving the mystery of disease processes at a molecular level. The need of the hour is a proper combination of experimental and theoretical results, obtained with different methods which would provide a pathway to study oral diseases, more significantly Head and Neck cancer[32,33].

The different protocols are followed for Microarray and NGS data analysis by practicing different workflows and utilizing many available packages and software. Although limitations are there for both technologies but can be overcome with appropriate solutions (Table 2).

3. Database Analysis

In cancer research and other pathologies, Microarray data analysis has opened new channels for diagnosis and treatment interventions. To evaluate gene expression various microarray platforms have been used, such as GeneChip, Agilent, NimbleGen, etc. The characteristic feature of microarray image analysis is to identify the discrete spot locations of gene expressions and quantify the spot intensities[39]. The microarray data recognizes multigene "expression profiles" for a given tumor that may be used to stratify histologically similar tumors into subtypes. Microarray analysis is a complex procedure and is performed using R statistical environment and the associated Bioconductor packages. The total analysis channel requires much software to precisely deal with the huge data[40]. The workflow for this has been explained (Figure 3).

As cancer progresses there are alterations in its signaling, methylation levels, and control networks. There has been a rapid progression in NGS technologies over recent years, which has revolutionized genomics and molecular biology including the search for structural

Table 2. Limitations of microarray and NGS technologies[34-38]

S No.	Microarray	NGS
1	It can detect sequences that are predetermined for available probes.	Expensive equipment, low yield of high-quality sequences and sequencing errors are high.
2	Difficulty in detecting low abundance transcripts. Minimal information for genes with minimal expression in the control sample.	Huge challenge to store high-throughput data.
3	The accuracy of expression measurements is limited by background hybridization. High noise level due to cross-hybridization, non-linearity, etc	Repeat sequences, Amplification bias, Homopolymers/InDels
4	No representation for the Non-coding RNAs on an array that is not yet recognized as expressed.	Contamination or new insertions
5	Relative concentration is represented by an indirect measure.	The computational challenge for reference independent de novo assembly for individual samples.
6	Difficult to design arrays in which the same probe will not bind multiple related DNA/RNA sequences.	Sequence properties and algorithmic challenges
7	No information on protein levels	Missing and fragmented gene

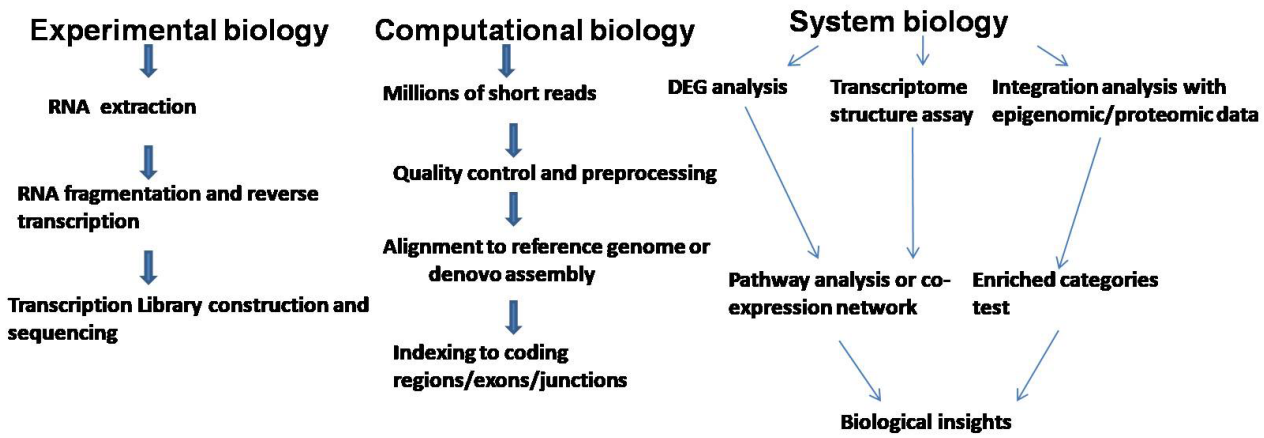


Figure 3. Microarray algorithm.

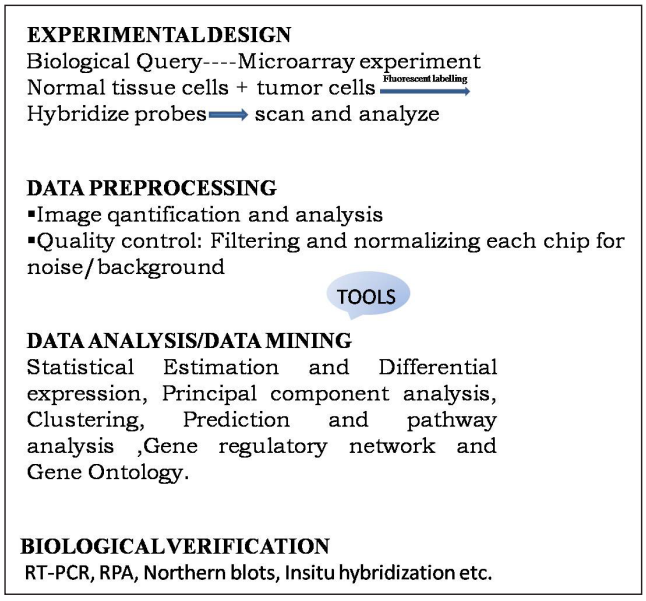


Figure 4. Work Flow for RNA sequencing.

variations in patient DNA. The NGS dataset interpretation also requires advanced and powerful computational programs. A few of the high throughput DNA sequencing methods involve Pyrosequencing, Illumina (Solexa) sequencing, SOLiD sequencing, Helioscope (TM) single molecule sequencing, etc[41].

The basic principle of NGS technology is that during sequential cycles of DNA synthesis, DNA polymerase catalyzes the inclusion of fluorescently labeled Deoxyribonucleotides Triphosphates (dNTPs) into a DNA template strand and nucleotides are identified by fluorophore excitation. This process is extended across millions of fragments in a massively parallel fashion. The Illumina NGS workflow includes library preparation, cluster generation/amplification, sequencing, and data analysis. The transcriptome under investigation is comprised of short nucleic acid sequences and the algorithm for RNA-seq includes isolation of RNA, target enrichment, fragmentation, cDNA synthesis, adaptor ligation, and amplification. The raw reads obtained are used for computational biology. An overview of a typical RNA-seq workflow is outlined in (Figure 4)[42]. A variety of freely available and commercial software tools are utilized for analyzing next-generation sequencing data to manageable sets of differentially expressed genes, functions, and pathways[43].

4. Conclusion

The future would be dedicated towards the maximum utilization of microarray and NGS data by implementing specific bioinformatics pipelines to extract meaningful biological insights. This could decipher human genomic information, elucidate mutations to increase survival rates, reduce induced co-morbidities and to better monitor the therapeutic efficacy. As this genomic asset of microarray and NGS-cancer data increase, helpful clinical-research information will be seeded to the oncologist and cancer patients as personalised therapeutic strategies.

5. References

1. Saintigny P, Zhang L, Fan YH, El-Naggar AK, Papadimitrakopoulou V, Feng L *et al.* Gene expression profiling predicts the development of oral cancer. Cancer

Prev Res (Phila). 2011; 4(2):218–29. https://doi.org/10.1158/1940-6207.CAPR-10-0155 PMid:21292635 PMCid:PMC3074595
2. Kuo WP. Overview of bioinformatics and its application to oral genomics. Adv Dent Res. 2003; 17:89-94. https://doi.org/10.1177/154407370301700121 PMid:15126216
3. Bennett NC, Farah CS. Next-generation sequencing in clinical oncology: Next steps towards clinical validation. Cancers (Basel). 2014; 6:2296-312. https://doi.org/10.3390/cancers6042296 PMid:25412366 PMCid:PMC4276967
4. Mendez E, Cheng C, Farwell G, Ricks S, Agoff N, Futran ND, Weymulle EA. Transcriptional expression profiles of oral squamous cell carcinomas. Cancer. 2002; 95:1482–94. https://doi.org/10.1002/cncr.10875 PMid:12237917
5. Kim KY, Cha IH. Possibility of the use of public microarray database for identifying significant genes associated with oral squamous cell carcinoma. Genomics Inf. 2012; 10(1):23-32. https://doi.org/10.5808/GI.2012.10.1.23 PMid:23105925 PMCid:PMC3475481
6. Murray D, Doran P, MacMathuna P, Moss AC. In silico gene expression analysis – an overview. Molecular Cancer. 2007; 6:50-60. https://doi.org/10.1186/1476-4598-6-50 PMid:17683638 PMCid:PMC1964762
7. Singh S, Lall R, Jain P. *In-silico* analysis of micro array data for prostate cancer. Int J Pharm Sci Rev Res. 2010; 3(1):46-8. ISSN 0976 – 044X
8. Singaraju S, Prasad H, Singaraju M. Evolution of dental informatics as a major research tool in oral pathology. J Oral Maxillofac Path. 2012; 16(1):83-7. https://doi.org/10.4103/0973-029X.92979 PMid:22434944 PMCid:PMC3303529
9. Rao VS, Das SK, Rao VJ, Srinubabu G. Recent developments in life sciences research: Role of bioinformatics. Afr J Biotechnol. 2008; 7:495-503. eISSN: 1684-5315
10. Verma C, Suhail S, Gupta OP, Gupta S. Recent developments in dent informatics: Role of bioinformatics. World J Pharm Res. 2015; 4(3):1299-1309
11. Zhang Y. Dental informatics and applications have an impact on dentistry. Eur J Biomed Inform. 2022; 18(4):32-3. https://doi.org/10.1353/fair.2022.0007
12. Giacomelli L, Covani U. Bioinformatics and data mining studies in oral genomics and proteomics: New trends and challenges. Open Dent J. 2010; 4:67-71. https://doi.org/10.2174/1874210601004020067 PMid:20871759 PMCid:PMC2945006
13. Wright JT, Hart TC. The genome projects: Implications for dental practice and education. J Dent Educ. 2002; 66:659-71. https://doi.org/10.1002/j.0022-0337.2002.66.5.tb03545.x PMid:12056771
14. Strausberg RL, Simpson AJG, Wooster R. Sequence-based cancer genomics: Progress, lessons and opportunities. Nat Rev Genet. 2003; 4:409-18. https://doi.org/10.1038/nrg1085 PMid:12776211
15. Malone JH, Oliver B. Microarrays, deep sequencing and the true measure of the transcriptome. BMC Biology. 2011; 9:34-43. https://doi.org/10.1186/1741-7007-9-34 PMid:21627854 PMCid:PMC3104486
16. Koschmieder K, Zimmermann K, Tri S, Stoltmannand T, Leser U. Tools for managing and analyzing microarray data. Brief Bioinform. 2011; 13(1):46-60. https://doi.org/10.1093/bib/bbr010 PMid:21422065
17. Nazmul-Hossain, ANM, Patel KJ, Rhodus NL, Moser KL. Microarrays: Applications in dental research. Oral Dis. 2008; 14(1):25-29. https://doi.org/10.1111/j.1601-0825.2006.01342.x PMid:18173445
18. Parolin C, Giordani B, Palomino RAN, Biagi E, Severgnini M, Consolandi C, *et al*. Design and validation of a DNA microarray for phylogenetic analysis of bacterial communities in different oral samples and dental implants. Sci reports. 2017; 7:6280. https://doi.org/10.1038/s41598-017-06743-6 PMid:28740183 PMCid:PMC5524749
19. Kuo WP, Whipple ME, Jenssen TK, Todd R, Epstein JB, Ohno-Machado L, Sonis ST, Park PJ. Microarrays and clinical dentistry. J Am Dent Assoc. 2003; 134(4):456-62. https://doi.org/10.14219/jada.archive.2003.0195 PMid:12733779
20. Macgregor PF, Squire JA. Application of microarrays to the analysis of gene expression in cancer. Clin Chem. 2002; 48(8):1170–7. https://doi.org/10.1093/clinchem/48.8.1170 PMid:12142369
21. Monger C, Kelly PS, Gallagher C, Clynes M, Barron N, Clarke C. Towards next generation CHO cell biology: Bioinformatics methods for RNA-seq-based expression profiling. Biotechnol J. 2015; 10(7):950-66. https://doi.org/10.1002/biot.201500107 PMid:26058739
22. Van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends Genet. 2014; 30:418–26. https://doi.org/10.1016/j.tig.2014.07.001 PMid:25108476
23. Han Y, Gao S, Muegge K, Zhang W, Zhou B. Advanced applications of RNA sequencing and challenges. Bioinform Biol Insights. 2015; 9(1):29–46. https://doi.org/10.4137/BBI.S28991 PMid:26609224 PMCid:PMC4648566
24. Upadhyay P, Dwivedi R, Dutt A. Applications of next-generation sequencing in cancer. Current Science. 2014; 107(5):795-802. https://www.jstor.org/stable/24105584
25. Bradford JR, Hey Y, Yates T, Li Y, Pepper SD, Miller CJ. A comparison of massively parallel nucleotide sequencing with oligonucleotide microarrays for global transcription profiling. BMC Genomics. 2010; 11:282-94.

https://doi.org/10.1186/1471-2164-11-282 PMid:20444259 PMCid:PMC2877694
26. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. Science. 1999; 286:531–7. https://doi.org/10.1126/science.286.5439.531 PMid:10521349
27. Gadewal N, Zingde SM. Database and interaction network of genes involved in oral cancer: Version II. Bioinformation. 2011; 6(4):169-170. https://doi.org/10.6026/97320630006169 PMid:21572887 PMCid:PMC3092954
28. Villalba GC, Ursula Matte U. Fantastic databases and where to find them: Web applications for researchers in a rush., Genet Mol Biol. 2021; 44(2):e20200203. https://doi.org/10.1590/1678-4685-gmb-2020-0203 PMid:33821874 PMCid:PMC8022358
29. Hide W. Genome Databases. Ashburner M, Ball CA, Blake JA, *et al*. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nature Genetics. 2000; 25(1):25–9 https://doi.org/10.1038/75556 PMid:10802651 PMCid:PMC3037419
30. Apweiler R, Bairoch A, Wu CH. Protein sequence databases. Current Opinion in Chemical Biology. 2004, 8:76–80 https://doi.org/10.1016/j.cbpa.2003.12.004 PMid:15036160
31. Levine AE, Steffen DL. OrCGDB: A database of genes involved in oral cancer. Nucleic Acids Res. 2001; 29(1):300-302. https://doi.org/10.1093/nar/29.1.300 PMid:11125119 PMCid:PMC29839
32. Kuo WP, Overview of bioinformatics and its application to oral genomics, Adv Dent Res. 2003; 17:89-94. https://doi.org/10.1177/154407370301700121 PMid:15126216
33. Kuo WP, Whipple ME, Epstein JB, Jenssen TK, Santos GS, Machado OL. Deciphering gene expression profiles generated from DNA microarrays and their applications in oral medicine. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2004; 97:584-91. https://doi.org/10.1016/j.tripleo.2003.11.016 PMid:15153870
34. Mehta JP, Rani S. Software and tools for microarray data analysis. Methods Mol Biol. 2011; 784:41-53. https://doi.org/10.1007/978-1-61779-289-2_4 PMid:21898212
35. Jaluria P, Konstantopoulos K, Betenbaugh M, Shiloach J. A perspective on microarrays: Current applications, pitfalls, and potential uses. Microb. Cell Factories. 2007, 6:4-18. https://doi.org/10.1186/1475-2859-6-4 PMid:17254338 PMCid:PMC1796898
36. Russo G, Zegar C,Giordano A. Advantages and limitations of microarray technology in human cancer. Oncogene 2003; 22:6497–650. https://doi.org/10.1038/sj.onc.1206865 PMid:14528274
37. Alkan C, Sajjadian S, Eichler EE. Limitations of next-generation genome sequence assembly. Nat Methods. 2011; 8(1):61-65. https://doi.org/10.1038/nmeth.1527 PMid:21102452 PMCid:PMC3115693
38. Behjati S, Tarpey PS. What is next generation sequencing? Arch Dis Child Educ Pract Ed. 2013; 98(6):236-38. https://doi.org/10.1136/archdischild-2013-304340 PMid:23986538 PMCid:PMC3841808
39. Fan J, Ren Y. Statistical analysis of DNA microarray data in cancer research. Clin Cancer Res 2006; 12(15):4469-73. https://doi.org/10.1158/1078-0432.CCR-06-1033 PMid:16899590
40. Aris VM, Cody MJ, Cheng J, Dermody JJ, Soteropoulos P, Michael R, Tolias PP. Noise filtering and nonparametric analysis of microarray data underscores discriminating markers of oral, prostate, lung, ovarian and breast cancer. BMC Bioinformatics 2004, 5:185 https://doi.org/10.1186/1471-2105-5-185 PMid:15569388 PMCid:PMC538261
41. Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol. 2008; 26:1135-45. https://doi.org/10.1038/nbt1486 PMid:18846087
42. Wolf B, Kuonen P, Dandekar T, Atlan D. DNAseq workflow in a diagnostic context and an example of a user friendly implementation. BioMed Res Int. 2015; 2015:1-10. https://doi.org/10.1155/2015/403497 PMid:26137478 PMCid:PMC4475531
43. Jiang Q, Yu YC, Ding XN, Luo Y, Ruan H. Bioinformatics analysis reveals significant genes and pathways to target for oral squamous cell carcinoma. Asian Pac J Cancer Prev. 15(5):2273-78. https://doi.org/10.7314/APJCP.2014.15.5.2273 PMid:24716969