NEXT GENERATION SEQUENCING IN DETECTING ORAL CANCER DUE TO TOBACCO CONSUMPTION

Ngoc Hoi Nguyen¹, Quang-Trung Nguyen², Deepak T.A³, Suchindra Suchindra⁴ and Avinash Tejasvi⁵

¹Institute of Chemical Technology, Vietnam Academy of Science and Technology, Ho Chi Minh City, Vietnam
²Center for Research and Technology Transfer, Vietnam Academy of Science and Technology, 18 HoangQuocViet, CauGiaY, HaNoi, Vietnam,
³Department of Oral Medicine and Radiology, V.S Dental College and Hospital, Bangalore, India
⁴Department of Engineering, National Institute of Mental Health and Neurosciences, Karnataka State Govt, Bangalore, India
⁵Department of Oral Medicine and Radiology, Kamreneni Institute of Dental Science, Narketpally, India

ABSTRACT

DNA sequence DNA Sequencing is the first step in establishing phylogenetic trees, protein structure prediction, diagnosis of cancer, discovery of drugs and hence its importance cannot be underestimated. DNA sequencing finds its use in the diagnosis of oral squamous cell carcinoma (OSCC). Oral Cancer is the most common occurring malignancies in the world, especially in India where the prevalence for smoking, Areca nut chewing coupled with a lifestyle that encourages these two activities as fashion are left many people diagnosed with OSCC. Patients with this OSCC are more likely unaware of its side effects and over time might suffer from facial deformity. The importance to understanding the symptoms, prevention and treatment of oral cancer is very much essential today. In this paper, we looked at over 2000 odd papers published and look at the correlation between the next Generation DNA sequencing algorithms (NGS) play an important role in diagnosis of OSCC. This is a further study on some of the papers which have highlighted NGS role in OSCC Diagnosis. We did like to see a comprehensive review on the papers published so far. In the discussion, we will see frequently mutated genes in the OSCC, recent discoveries and OSCC treatment based on the findings.

KEYWORDS

Sequence, phylogenetic Tree, Carcinoma, Gene.

1. INTRODUCTION

Tobacco is a consumed as a recreational substance either orally (chewed, sniffed, inhaled, and sucked) or through a third a party substance which contains tobacco, eg: Cigarette. Sometime, Tobacco is impersonalized as a cigarette in India when a tobacco is rolled like a small cigarette, eg Beedi [25,26]. In Latina America, thick tobacco leaves are fermented, and tobacco leaves of different grades are then layered to make a thick blunt which is popularly known as a cigar [27]. The consumption of tobacco increases the chances of oral cancer [11]. In India, although tobacco is still consumed orally, it is quickly fading away and is replaced by tobacco consumed through cigarette, cigar and beedis [27]. It is estimated that tobacco is the biggest killer, about 10 million deaths every year, of which tobacco consumption contributes to about 5 million [28,29]. From
the literature, we observe that tobacco smokers have about 10 - 15% chance of developing oral cancer than nonsmokers and this percentage increases with the smoking frequencies [30-33].

Tobacco smoke and its byproducts mixed with saliva after chewing produces many carcinogenic materials which are commonly called polycyclic aromatic hydrocarbons (PAHs). These PAHs are more pronounced in smokers than orally consumed tobacco patients [34]. Apart from these PAHs, Nitrosamines, aromatic amines, and other agents which are present in elevated levels in a smoker’s urine at any given point of time. The PAHs are the main causes for cellular transformation [35, 36]. These PAHs would later alter, mutate the DNA, RNA, Protein and Genes in individuals. A RNA, DNA or a protein sequence is represented by a character set. DNA (A, C, G, T), RNA (A, C, G, U) and protein molecules (A, R, N, D, C, Q, E, G, H, I, L, K, M, F, P, S, T, W, Y, V) [1 - 9]. Tobacco smoking is quite addictive, and once addicted, there is now constant exposure to these carcinogenic substances. This exposure would now lead to bondage of carcinogens to DNA genome (molecular changes or DNA changes). This bondage would now lead to DNA mutation where a protein character at one place in entire genome is replaced by another protein, there by mutating into other DNA form. In some cases, the DNA character goes missing or a new protein is added when it should not be [34].

2. Literature

Oral cancer can be defined as the presence of lesions or tumors in the lip or [10]. Oral Cancer is most common form of cancer seen across the world and its ranking within each continent varies from year to year. It occupies the highest ranking in India and surrounding south-east Asia [11, 12, 13]. From our literature review we could conclude that oral cancer diagnosis cases have more than doubled from 1990 - present. It’s about 3 out of 10000 and the adjusted life expectancy is about 66 years from the data. In Southeast Asia, the prevalence and fashion driven cigarette smoking and areca nut chewing amounts, amounts to 98% of the cases. Some cases where the oral cancer is due to passive smoking, oral Hygiene and others are reported [15, 16].

Oral Squamous Cell Carcinoma (OSCC) is the most common malignant epithelial neoplasm affect the oral cancer. Oral cancer can affect any region including oral cavity, pharyngeal regions, and salivary glands. Most of the lesions are prevalent in the anterior part of the tongue, Labial, gingiva, alveolar muscle, the palate and some cases in the parotid glands or salivary glands [11][18 – 20]. Despite many technological advancements in the recent times both in detection, prognosis and treatment, common treatment for oral cancer is still surgery or radiotherapy. If the cancer is in its initial stage, then the survival rate is considerably higher with either surgery or radiotherapy [21]. But if the cancer is detected in the 3rd or 4th stage then the survival rate falls to 30% [21] and most of the patients die within 12 - 24 months if the cancer is the advanced stage [22 - 24].

Today, efforts are made to increase the efficacy of radiotherapy and chemotherapy, either through the advent of precision targeted therapy like laser, and targeted medicine to expedite the research and development of targeted therapies and screening strategies in the oral cancer [18][19]. Targeted molecular therapy, like monoclonal antibodies and gene therapy, have been applied to oral cancer patients. These molecules are associated with the proliferation of OSCC [20].

These protein displacement or replacement in the DNA chain play the pivotal role in the advert mutation and progression of the OC. DNA genome project is perhaps the biggest project ever taken in the bioinformatics area in the early 1990s [5]. Ever since then, the appetite for DNA knowledge acquisition has grown tremendously. With this curiosity, the research produced DNA sequencing algorithms ranging from local sequence alignment algorithms such as Smith-Waterman [37], FASTA [38], BLAST [39], GappedBLAST[40], BLASTZ[41],
PatternHunter[42], YASS[43], LAMBDA[44], USearch [45], LAST [46], and ALLAlign [47] and MASAA [1-4]. These algorithms are concentrating on aligning locally similar sequences.

There are other algorithms which concentrate on aligning the entire sequences (Pairwise), these are Needleman and Wunsch [48], GLASS [49], WABA [50], and GLASS [51]. MASAA, GLASS and BLASTZ are very important in this research, in that, these algorithms a novel method of first finding regions of similarity quickly and then employ another novelty called stitching algorithm to finally find the final alignment. The regions of similarity size, and the stitching algorithm vary in these algorithms, which brings to novelty surge in multiple sequence alignment. When the need to align multiple sequences, where more than 2 sequences are aligned to find regions, came about, more researchers now had to look at the above algorithms to find regions of similarity quickly yet wanted alignment that made biological sense.

The popular multiple sequence alignment algorithms were ClustalW [52], MAFFT [53], Kalign [54], Proalign [55], MUSCLE [56], DIALIGN [57], PRANK [58], FSA [59], T − Coffee [60, 61], and Probcons [62]. Other approaches focused on Genetic and motifs algorithms [5]. Genetic algorithms got traction, while motifs-based algorithm was not successful [5].

3. NEXT GENERATION SEQUENCING

Next-generation sequencing (NGS) is a technique where powerful massive computer clusters are run to parallelly sequence millions of small segments of DNA. NGS enables researchers to study genomes amongst many sequence types systematically and make insights which in turn leading into disease understanding and drug discovery. In NGS, the entire genome or targeted segments are sequenced quickly [10]. As discussed in the previous section, researchers were prodding new algorithms ever since DNA sequence are discovered, for more insights [5]. Frederick Sanger in 1970s popularised a sequencing method using a 'plus and minus' method. It involved the usage of phix174 and chain-terminating inhibitors [11, 12]. There was Maxam-Gilbert which gained popularity for a brief period and quickly the method was abandoned, as this method was cumbersome and involved the use of many chemicals [14]. Sanger method is still considered to be one of the classical methods and is still in use today.

NGS is used in two platforms Illumina/Solexa and Ion Torrent. Of these two, the Illumina/Solexa is a low cost and high yield method, which has propelled it to be the preferred platform today. The Ion Torrent, method is rather quick and produces longer read lengths than the Illumina [15], however due to the method being expensive, it is not used widely today [14, 15]. Other platforms like the Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio) are quite new and they too produce longer lengths, however, from the literature, the price to use is not yet clear [16]. Asia especially China has poured enormous amount of money in developing new platforms, one of method coming out of ASIA is Beijing Genomic Institutes BGISEQ, standing for Beijing Genomic Institute Sequencing. BGISEQ used a new nanotechnology to generate terabytes of sequenced data [63-65]. This method is also less expensive but produces longer length compared to Illumina. It is to be seen in the future if this method would be adopted widely as illumina/Solexa is today [64,66].

One of the fields that has immensely benefitted from NGS is microbiology and dentistry. NGS has played an important role in identifying oral microbiota in humans. These newly identified microorganisms which live in human oral cavity is thus enlarging the dentistry knowledge base to support good overall good health. Two methods are used to evaluate the kind of organisms present in the oral cavity using NGS is called 16 rRNA sequencing and the other shotgun sequencing [67]. 16 rRNA is present in all bacteria and therefore this knowledge can be used to know whether the bacteria are present in one’s oral cavity or not [68]. Shot sequencing is used to
gain knowledge on microbes’ genome and resistance of genes to certain conditions [69]. Using these two methods, NGS can now classify different microbial genomes, this knowledge is now used for disease diagnosis, germ or bacterial resistance know-how and drug discovery.

NGS has overcome some of the limitations of PCR based sequencing, Sanger sequencing, and other micro array and SNP array methods, either through better sensitivity or better genomic coverage or both. This better sensitivity using NGS has through more light on some of the ambiguities in oral cancer classification. In other words, Sanger sequencing uses a minimum mutant percentage of about 25%, this is quite large, in fact, some cancer could now be diagnosed well below 25% of mutant allele [70,71]. The PCR method is quite limited to only few know sequences in detecting the tumours or oral cancer [72]. NGS expands this horizon as it can sequence the entire genome, targeted sequencing, RNA, and mRNA too [73]. Today, research has shown the tumour after biopsies are heterogenous in nature, this now presents a difficult proposition if PCR method is used, as it would not be able to detect the mutation (gene) less of 25%. NGS would otherwise be able to detect such kind of solid heterogenous tumours [74,75]. In this paper, we will review what are the genes which get modified in patients with smoking habits resulting in oral cancer using NGS technology. We do this by going through some 1800 plus papers who used NGS technology and what genes their research show cased, and we draw our conclusion from this exhaustive research.

4. RESEARCH METHODOLOGY

Our literature search was conducted according to PRISMA guidelines, where in all studies which employed NGS technology to study the impact of smoking and its direct correlation to oral cancer was studied [77-79]. We expanded the already existing study made by [76] to be broader and more exhaustive as the previous study conducted sample size was skewed towards records which did not have any NGS technology and only a 10% out of 1200 odd papers had any mention of Oral cancer and NGS simultaneously. We felt this was not exhaustive study and hence brought upon ourselves to expand appropriately to reflect the real picture of the NGS and Oral cancer correlation. We searched through freewheeling, along with Pubmed and Embase records and we want to emphasize that new expanded records all have NGS and Oral Cancer. As we searched, many duplicates, derived research papers were removed, then a consensus number were taken into the final pool was finalized.

5. METHOD

The Records were researched from Jan 2023 till September 2023 on PubMed, Embase, Researchgate and ScienceDirect and freewheeling. A total of 1921 records were researched, out of which any mention of NGS and oral cancer was 1505 records. The authenticity of the records was examined by 2 independent reviewers. Both primary and secondary and data extraction, from the papers were done by both the researchers. There were approximately 200 records bring the combined examined papers to 2021 papers. 1: There were records on Basic Analysis of NSG, Smoking and Oral Cancer (n = 29), 2: records on Diagnosis based and Analysis (n = 1), 3: records on just model (n = 1), 4: records on NSG, Smoking and oral cancer (n = 33), 5: Records on NGS but no data ( n =3), 6: Records not on either NGS or Oral Cancer or Smoking or any combination of these 2 (n = 127), 7: Records on reviews ( n = 52), 8 : Record not on oral cancer (n = 1249).
Figure 1: Flow chart of workflow

Oral cancer or OSCC or oral or squamous cell carcinoma and next generation sequencing or NGS or Smoking or DNA sequencing and Genes was used in Embase and fields like (“oral”) OR (“Carcinom”) OR (“Oral Cancer”) or (“Mouth Cancer”) Or Its combinations OR ("carcinoma"[All Fields] AND "squamous"[All Fields] AND "cell"[All Fields]) OR "squamous cell carcinoma"[All Fields] OR ("squamous"[All Fields] AND "cell"[All Fields] AND "carcinoma"[All Fields])) AND ("mouth"[MeSH Terms] OR "mouth"[All Fields] OR AND "cavity"[All Fields]))) OR (("carcinoma, squamous cell"[MeSH Terms] OR ("carcinoma"[All Fields] AND "squamous"[All Fields] AND "cell"[All Fields]) OR "squamous cell carcinoma"[All Fields] OR ("squamous"[All Fields] AND "carcinoma"[All Fields])) OR "tongue"[All Fields])) was used in pubmed and oral cancer, smoking on freewheeling [76]. Records that utilized NGS, Smoking, and genes described in OC patients with smoking were 33 with genes considered majorly is 6. The entire workflow is shown in Figure 1.

6. RESULTS AND ANALYSIS

There were 33 included in this research, with full text with 34 genes found in OC patients who smoke (Figure 2 and Table S1). Of these there are 6 genes which were found in more than 4 records. These were 11p15.5 (HRAS), 17p13.1 (TP53), 9p21.3 (CDKN2A), 9q34.3 (NOTCH1), and 4q35.2 (FAT1) and 4q31.3 (FBXW7) and 9p24.1 (PD-L1) were found to be in 3 records and the rest of the genes were found to be 1 record. This is represented in the figure 2.
From the literature, we can confidently say that smoking is directly correlated with oral cancer and the chances for OC diagnosis raises if the patient smoking frequency of smoking is high (chain smokers). It is also found out the smoking somehow is responsible for genome mutation, transcriptomic changes, some of these changes are found in some genes portion of the genome, which has catastrophic effect on the changes inside the cell (cell reaction and protein developed, and expressions). All these changes are quite complex and leads to the initiation of the oral cancer [118]. Tp53 is a gene responsible for cell health and its reproduction (growth) [118, 119]. Since it is an important gene in cell health, it also plays a pivotal role in suppressing any changes to the cell health, structure or multiplication [118, 119]. Alternation to this gene could be seen, as the cancer origination and spread of cancer [119]. From the literature, we can say with certain surety that mutations in the Tp53 gene were present in various human cancer [118]. However, when it comes to oral cancer, we have found from our research, that mutation in Tp53 were found in 78% of the 33 records, which is 16 in number [80-91]. This is quite an observation, we have found various authors [81-89, 91] and [118], believe and observe that this gene mutation is the primary initiator for the Oral Cancer proliferation. These authors [81-89, 91] also say that missense mutations are the most common. It is also observed that Tp53 is located near DNA binding area of the protein [81-89, 91] and it is associated with insufficient or poor prognosis [81-85]. Authors in [120] have noticed that there is common change or mutation is from G -> A or G -> T, and the authors observe that this mutation scenarios have more to do with smoking and consumption of tobacco in one form or the other. Authors in [121] have observed that such mutations have increases spread of OC, increased penetration, and increased resistance to drugs and eventually leading to death.

NOTCH1 is another gene mutated in OC associated with smoking. NOTCH1 is also a tumour suppressor gene very similar to TP53 and similarly to Tp53, any changes to this gene leads to initiation and multiplication of cancerous cells (tumour) [80-91]. From our research, we found out that, 63% of the records analysed or full text record had NOTCH1 gene mutated, which is a substantial percentage than the percentage found on other review [76]. We also found out from the records we looked at: all these studies were predominantly Asians (Including South - Asia). This tells us that NOTCH1 mutation is a likely event in all OC cancerous patients [118] in samples from Asia. Authors [118] evident that this observation is predominantly because of the lifestyle of population, with many consuming locally made cigarettes (Bedis in India, Secondary smoking (smoking from discarded cigarettes)) and alcohol consumption. In our research, we saw one NOTCH gene variation and this NOTCH variation was a study made in Europe, and we are now certain that this variation is also in other races but perhaps not in scale or volume found in
Asian population. The mutation in NOTCH1 genes were found in different location all along the length of the exon [118, 119] and such mutations in all studies clearly point to decreased confidence in the survival of the patient [91].

The third gene which was found to be highly mutated in records was PIK3CA, this gene accounted for 48% which is quite less than the other 2 genes found above. The exon 20 is the common area mutated of PIK3CA [86-89] and this affected area has been seen to be associated with lower lip region of the patient [86]. Authors of [82] have observed that mutation in the PIK3CA has a strong correlation with the later stage of OC, clearly showing that this mutation is clear signal that patient prognosis would be bad when this gene is mutated. Hence, a different targeted therapy to rectify this situation is highly recommended in the literature. Authors in [83] have observed that the activation of PI, suggest that Signalling pathway could be improved either by targeted treatment of other means. From the research, we conclude that a mutation of PIK3CA is a clear signal that OC is the later stage, prognosis could be hampered, a clear sign that alternative therapy should be looked and provided to the patient immediately, if not, then the patient is looking at grim and sad prognosis leading to fatality.

CDKN2A gene is associated with cell cycle inhibitor [122]. Authors have found that a downregulated expression of this gene in the life cycle of the cell correlates with OC and suggest that this gene could now be used a marker for OD diagnosis. In our research, we found 47% of the full text records to have this gene mutated with the OC occurrence and unlike the other genes, we could not correlate this gene mutation with any region of the world [123]. Most of the reported mutation in the records suggest that there is mutation from G -> T protein in the exon 2 and in position 322 of this exon, and this correlated with primary head and neck cancers along with oral cavity [124]. This gene mutation is also correlated with OC recurrence after the OC was partially cured in the patient, meaning, that the cell cycle inhibitor played an important role in recurrence of OC [125]. We also observed that majority of the tongue related OC had CDKN2A gene mutated. The prognosis was poor after this CDKN2A was mutated, leading the poor prognosis, indicating the current therapy must be stopped and alternative therapy be initiated to curb OC growth [126].

H-RAS gene is also an important regulate cell health, differentiation, migration, and apoptosis [80]. Mutation of this associated with occurrence of OC [80 - 91]. Incidentally, HRAS genes is quite less often seen in OC than other cancers [ 121 - 125], however, in neck and head and oral cancers, we observe that this gene is seen [ 80 - 91] [119 - 123, 125]. In our study we found H-RAS gene studies to be around 25%, this gene is also associated from the research to be more prevalent in patients who chew tobacco and areca nuts with alcohol consumption. From our study, we conclude that H-RAS gene to be indicator of OC in the later stage and an observation of this gene mutation correlated to patient living with poor condition, consumption of alcohol, smoking, tobacco chewing or areca nut chewing along with betel leaves.

FAT-1 gene role in cancer has been controversial in nature. FAT1 acts as a cancer suppression role especially in HNSC and oral cancers. The expression of this gene is mostly seen in patients with later stages of OC. In our study, we saw 18% of the records mentioning FAT1 gene [83]. We observed from the records, any mention of FAT1, was seen the stage 4 compared to other cancers [89 – 91]. Also, the samples included in the research records showed that OC was lip and oral cavity (Roof) than other regions in OC. FAT1 gene was also not associated with any world region or sample lifestyle. Prognosis was observed to be poor leading to fatality in 2 of the records.
7. CONCLUSION

In this study, we included 33 records with full article reads on OC, NGS and smoking correlation. We found some interesting observation, several of them, revealed that certain gene is mutated in some ethnicities than the rest, genes were mutated less than others in various regions. The correlation of certain gene mutation to lifestyle was also observed. The study has revealed that we need more research on OC to accurately diagnose and treat early-stage OC. This study was also on the NGS sequencing technique used to find OC with samples or records on smoking or areca nut consumption. This study revealed certain gene, most of the times, the first 2 genes mentioned in the study to be mutated as these are OC inhibitors. Smoking is a killer, and it is incumbent on all people in all regions around of the world to educate people on repercussions of tobacco addiction either through smoking or chewing.

APPENDIX

All data on the records is in the excel file ResearchTableWithTabs_S1S2S3.

CONFLICT OF INTEREST

The authors have no conflicts of interest in this research.

ACKNOWLEDGEMENTS

The authors would like to thank their respective universities and families for their support.

REFERENCES


WHO Tobacco Fact Sheet. 2020. https://www.who.int/newsroom/fact-sheets/detail/tobacco

The Tobacco Connection. n.d. https://oralcancerfoundation.org/understanding/tobacco/


23


Bioscience & Engineering: An International Journal (BIOEJ), January 2024, Volume 11, Number 1


[76] Sharma, Vishwas & Kumar, Dinesh & Kumar, Shravan & Singh, Harpreet & Sharma, Naveen & Gupta, Sanjay. (2022). Impact of tobacco smoking on oral cancer genetics—A next-generation sequencing perspective. iMeta. 1. 10.1002/imt2.44.

Bioscience & Engineering: An International Journal (BIOEJ), January 2024, Volume 11, Number 1


Bioscience & Engineering: An International Journal (BIOEJ), January 2024, Volume 11, Number 1


[110] Chung Ji Liu, Shu-Chun Lin, Li-Han Lin; Abstract 1234: Somatic mutations of FAT1 in oral cancer are associated with tumor progression and survival. Cancer Res 1 July 2018; 78 (13_Supplement): 1234. https://doi.org/10.1158/1538-7445.AM2018-1234


Authors

Dr Nguyen works as a professor at the Institute of Chemical technology, Vietnam Academy of Science and Technology. Her interests are in Biomedicine materials, nanomaterials, Biotechnology and Food.

Mr Nguyen received his masters of Information Technology from Thuongmai University, Vietnam. His research interest is in computational biology and software engineering.

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.

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 institute 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.