Gene mutation analysis of oral squamous cell carcinoma in the background of oral submucous fibrosis in Hainan Island (#86806)

First submission

Guidance from your Editor

Please submit by **25 Jul 2023** for the benefit of the authors (and your token reward) .

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Custom checks
Make sure you include the custom checks shown below, in your review.

Raw data check
Review the raw data.

Image check
Check that figures and images have not been inappropriately manipulated.

If this article is published your review will be made public. You can choose whether to sign your review. If uploading a PDF please remove any identifiable information (if you want to remain anonymous).

Files

Download and review all files from the materials page.

4 Figure file(s)
3 Table file(s)
1 Raw data file(s)

Custom checks

Human participant/human tissue checks

- Have you checked the authors ethical approval statement?
- Does the study meet our article requirements?
- Has identifiable info been removed from all files?
- Were the experiments necessary and ethical?

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- All underlying data have been provided; they are robust, statistically sound, & controlled.
- Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult. I suggest you have a colleague who is proficient in English and familiar with the subject matter review your manuscript, or contact a professional editing service.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

Gene mutation analysis of oral squamous cell carcinoma in the background of oral submucous fibrosis in Hainan Island

Bingxia Li [Equal first author, 1, 2], **Xinyu Chen** [Corresp., Equal first author, 3], **Haiyu Xian** [1, 2], **Qitao Wen** [1, 2], **Tao Wang** [Corresp. 1, 2]

[1] Department of stomatology, Hainan General Hospital, Haikou, China

[2] College of Stomatology, Hainan Medical College, Haikou, China

[3] College of Stomatology, Guangxi Medical University, Nanning, China

Corresponding Authors: Xinyu Chen, Tao Wang
Email address: chenxinyu814@sr.gxmu.edu.cn, 18608917377@163.com

Objective. The sequencing panel composed of 61 target genes was used to explore the related mutation genes of oral squamous cell carcinoma (OSCC) and OSCC in the background of oral submucous fibrosis (OSF), which would provide a theoretical basis for the early diagnosis of OSCC in the background of OSF. **Methods.** A total of 74 clinically diagnosed samples were included, including 36 cases of OSCC and 38 cases of OSCC in the background of OSF. DNA was extracted, and targeted gene Panel sequencing technology was used to analyze the gene frequency of pathogenic mutation sites in clinical samples. **Results.** Gene panel sequencing analysis showed that there were 69 mutations in 18 genes in OSCC and OSCC in the background of OSF. The results of gene panel sequencing were screened, and 18 mutant genes were finally screened out and their mutation frequencies in the samples were analyzed. According to the frequency of gene mutations from high to low, they were TP53, FLT4, PIK3CA, CDKN2A, FGFR4, HRAS, BRCA1, PTPN11, NF1, KMT2A, RB1, PTEN, MSH2, MLH1, KMT2D, FLCN, BRCA2, APC. Mutation of FLT4 gene was found to be unique to OSCC in the background of OSF group and its mutation frequency was significantly higher than that of OSCC group ($P<0.05$). **Conclusion.** FLT4 gene may be related to OSCC in the background of OSF and is expected to be an early diagnostic biomarker for it.

Gene mutation analysis of oral submucous fibrosis carcinogenesis in Hainan Island

Bingxia Li[1,2,a], Xinyu Chen[3,a], Haiyu Xian[1,2], Qitao Wen[1,2], Tao Wang[1,2,*]

[1]Department of stomatology, Hainan General Hospital, Haikou, Hainan, China
[2]The Affiliated Hainan Hospital of Hainan Medical University, Haikou, Hainan, China
[3]College of Stomatology, Hospital of Stomatology, Guangxi Medical University, Nanning, Guangxi, China

*** Corresponding author.**
Tao Wang
Dental Medical Center, Hainan Affiliated Hospital of Hainan medical University (Hainan General Hospital), 19 Xiuhua Road, Haikou, Hainan, 570311, China.
E-mail address: 18608917377@163.com
[a] Authors contributed equally.

Abstract

Objective. The sequencing panel composed of 61 target genes was used to explore the related mutation genes of oral squamous cell carcinoma and oral submucous fibrosis, which provided a theoretical basis for the early diagnosis of oral submucous fibrosis.

Methods. A total of 74 clinically diagnosed samples were included, including 36 cases of OSCC and 38 cases of OSF cancer patients. DNA was extracted, and targeted gene Panel sequencing technology was used to analyze the gene frequency of pathogenic mutation sites in clinical samples.

Results. Gene panel sequencing analysis showed that there were 69 mutations in 18 genes in OSCC and OSF cancerous specimens. The results of gene panel sequencing were screened, and 18 mutant genes were finally screened out and their mutation frequencies in the samples were analyzed. According to the frequency of gene mutations from high to low, they were TP53, FLT4, PIK3CA, CDKN2A, FGFR4, HRAS, BRCA1, PTPN11, NF1, KMT2A, RB1, PTEN, MSH2, MLH1, KMT2D, FLCN, BRCA2, APC. The mutation frequency of FLT4 gene was significantly higher than that of OSCC group ($P < 0.05$).

Conclusion. FLT4 gene may be related to OSF carcinogenesis and is expected to be an early diagnostic biomarker for OSF carcinogenesis.

Keywords. Oral squamous cell carcinoma; oral submucous fibrosis; gene Panel sequencing analysis; Gene mutation

Introduction

oral carcinoma is traditionally defined as oral squamous cell carcinoma (OSCC), which is the most common malignancy of squamous cell carcinoma of the head and neck. It is the sixth most common cancer in the world. It seriously affects people's chewing, swallowing and speech functions, and brings huge personal and social costs. In our country, there is a lot of cases in Hainan, Hunan, and Taiwan(Lee et al., 2011). Residents in these areas have the habit of chewing betel nut, which is the main factor that leads to the oral submucous fibrosis (OSF) and its cancerous transformation into OSCC. The International Agency for Research on Cancer (IARC) has classified areca nut as a Group I carcinogen and has identified it as an independent risk factor for OSCC. Despite the development of comprehensive and multidisciplinary therapies, the

prognosis and 5-year survival rate for OSCC remain unsatisfactory, and studies have shown a low rate of lymph node metastasis in patients with areca chewing OSCC. In the past, Some scholars defined OSF canceration as "oral cancer complicated with or accompanied by OSF", which included cases of leucoplakia canceration of oral mucosa. After years of research, our team deemed that cases developed from leucoplakia should be excluded and OSF canceration should be defined as oral squamous cell carcinoma directly transformed by OSF. Because oral cancer is often detected and diagnosed at a later stage, it is usually fatal for those affected, with a death rate of about 80 percent for advanced oral cancer, which can be reduced to about 50 percent with early screening and diagnosis(Warnakulasuriya, 2009, 2010). Therefore, it is essential to identify early diagnostic targets for OSF canceration.

Common types of gene sequencing include whole genome sequencing, whole exon sequencing, transcriptome sequencing, and targeted capture sequencing. Targeted sequencing is a research strategy in which the genome region of interest is enriched by capture kits for sequencing. According to different applications, ultra-high sensitivity and accuracy can be obtained with less data, and rapid screening of variation sites can be realized. Targeted gene Panel sequencing is a sequencing strategy and method designed on the basis of whole exome sequencing and whole genome sequencing. Compared with targeted sequencing, targeted sequencing can focus on the region of concern, remove the interference of redundant data, and maximize the use of sequencing reads, which has the advantages of deeper sequencing, lower cost and more sensitive. It is the preferred method to identify complex disease susceptibility genes at present(Kamps et al., 2017). No targeted sequencing studies have been found on OSF cancerosis. Therefore, here we used a gene sequencing panel composed of oncogenes and tumor suppressor genes to identify the most common and unique mutations in the OSF cancerosis genome.

Materials & Methods

Research object and ethical statement

This study included the primary focus tissues of 74 patients with OSCC and OSF canceration admitted to Hainan Provincial People's Hospital from June 2020 to September 2021. This study was approved by the Clinical Research Ethics Committee of Hainan Provincial People's Hospital(Approval No. [2019] 37). All samples used were informed by the patient, and the consent form for sample use was signed before the operation. The research complies with the World Medical Association's Code of Ethics (Helsinki Declaration) for experiments involving human beings. Subjects were recruited according to the defined inclusion and exclusion criteria. For cases: (1) subjects over 18 years old; (2) Diagnosed as OSCC or OSF canceration; (3) All patients with oral cancer were first-time patients without radiotherapy and chemotherapy; (4) It belongs to Hainan Province. All tissue samples were stored in the tissue sample storage tube of Kangwei Century within half an hour after the operation, without the influence of formalin fixation or paraffin embedding. The clinical diagnosis of all specimens was confirmed by histopathological examination.

DNA extraction

Use Magen tissue DNA extraction kit (Guangzhou Meiji Biotechnology Co., Ltd. D3126) to extract DNA from oral cancer tissue samples, and strictly follow the steps described in the extraction kit. The DNA concentration was determined using Qubit 3.0 (Thermofly).

Library construction and gene panel sequencing

HieffNGSOnePotlDNA library kit (Yisheng Biotechnology, 13321ES96) was used to construct the library by enzyme digestion. According to the instructions of the kit, DNA fragmentation/end repair/dA tail addition, connector connection, product purification, DNA fragment length sorting, library amplification and other operations were performed on all samples to complete the construction of the library, and quality inspection was performed using Qsep100 (Guangding Biotechnology).

For qualified libraries, the hybridized capture of the library is carried out by using the hybridized capture system (product number: P10006A) of Heyin Biological and the panel probe of oral squamous cell carcinoma independently designed by Hefei Novel Gene Technology

Service Co., Ltd., including library hybridization, elution, amplification, purification and quantification. Among them, the panel probe covers 61 genes related to oral squamous cell carcinoma, including ABCB1, APC, BRAF, BRCA1, BRCA2, CDA, CDKN2A, CYP19A1, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP3A4, CYP3A5, DHFR, DPYD, EGFR, ERBB2, ERCC1, ERCC2, FGFR4, FLCN, FLT1, FLT4, GSTM1, GSTP1, GSTT1, HRAS, IDH1, IDH2, KDR, KIT, KMT2A, KMT2D, KRAS, MET, MLH1, MSH2, MTHFR, NF1, NQO1 NRAS、NTRK1、NTRK2、NTRK3、PDGFRA、PDGFRB、PIK3CA、PTEN、PTPN11、RB1、RRM1、SLC19A1、SMAD4、SULT1A1、TP53、TPMT、TYMS、UGT1A1、XRCC1。 Finally, Hefei Novel Gene Technology Service Co., Ltd. was entrusted to use the MGISEQ-200RS high-throughput sequencing platform for sequencing and analysis of sequencing data.

Experimental grouping and statistical treatment

The subjects were divided into OSCC group and OSF canceration group for comparison. SPSS26.0 software was used for statistical analysis. The measurement data were described by ($\bar{x}\pm s$) and t-test was used; The counting data is expressed by passing (%), using 2 test; Fisher's exact test is used to detect the association between genes and OSCC; Results The difference was statistically significant with<0.05.

Results

Results of tumor specimen collection

All patients with OSCC were selected for surgical resection, and the postoperative pathological diagnosis was OSCC, and the diagnosis was clear. The following are the characteristics of 74 cases (Table 1)

Quality assessment of hybridization capture sequencing data

We performed next-generation sequencing analysis on 36 OSCC samples and 61 genes from 38 OSF cancerous samples classified by histopathology (Table S1). All sample sequencing reads are processed in the following steps: 1. Use trim-galore software to filter low-quality reads and remove adapters to obtain clean date; 2. Compare Clean Date with the reference genome

hg19 using the BWA MEN program, and remove PCR repeats during sequencing using samtools and GATK toolkits; 3. Analysis of variant sites using the GATK toolkit; 4. Annotate mutation sites using VEP, snpEFF, ANNOVAR annotation software, and ClinVar database. All samples have a read length of 100bp. The summary statistics of the sequencing results show that the total data volume of sample sequencing is 2.97G, the average median target rate is 81.12% (range: 54.52%~98.30%), the average sequencing depth range is 697.41X~9861X, the one-time average target coverage is 98.90% (range: 92.13%~100%), and the average Q30 measurement value is 92.60% (range: 88.37%~95.73%) (Table S2).

Mutation analysis of study subjects

Results by panel sequencing of targeted genes，Further bioinformatics analysis revealed 69 mutations（Table S3）.Based on the type of mutation, variants are divided into missense mutations (35), nonsense mutations (20), frameshift deletion (13), and frameshift insertion (1) (Figure 1), The largest number of missense mutations is among them. It can be seen from the gene waterfall plot that FLT4 are all frameshift deletion mutations, and they are all distributed in OSCC and OSF samples. Single nucleotide variation (SNV) is when one base is replaced by another, also known as single-nucleotide polymorphism (SNP). SNV is generally superior to SNP because only variants found in a single sample are called SNV, whereas SNP is a population concept, and this difference accounts for more than 1% of the population. Classified by transition/swap level, a total of 55 mutations were identified, of which 3 8 were converted and 17 were switched, with a higher number of transitions than reversals among all study subjects, most of which were distributed in mutation rates of 50% or more (Figure 2). Most mutations were C-to-T transitions (33, including 17 for G>A and 16 for C>T (Fig 2).Mia(Petljak et al., 2022) believes that this is due to cytosine mutations produced by overexpression of cytosine deaminase APOBEC3, and the replication of uracil may lead to C to T mutations leading to single-base substitution 2 (SBS2). We also found a G to A transition in a large sample size.

Analysis of sample gene mutation

Gene frequency statistics of pathogenic mutation sites in 74 samples. The mutation

screening conditions in each sample were satisfied: Pathogenic on the ClinVar database comment, mutation frequency AF higher than 3%, and GATK software identified the mutation as PASS. Eighteen mutant genes were screened out, and the mutation frequency of these 18 genes in 74 samples was counted, and the mutation frequency of these 18 genes in 74 samples was calculated as TP53, FLT4, PIK3CA, CDKN2A, FGFR4, HRAS, BRCA1, PTPN11, NF1, KMT2A, RB1, PTEN, MSH2, MLH1, KMT2D, FLCN, BRCA2, APC. (Figure 3). In addition to TP53, the mutant gene of FLT4 gene was the most frequent in 74 oral squamous cell carcinoma samples, The number of mutations was 27 and 7, respectively (Figure 4).

Seventy-four samples were divided into OSCC group and OSF carcinogenesis group, and the mutation ratio of mutant genes in the two groups was analyzed and compared. Six mutant genes were found to be common to both groups of samples: TP53, PIK3CA, CDKN2A, FGFR4, BRCA1, PTPN11; Five genes were unique to OSCC samples: NF1, PTEN, MSH2, BRCA2, APC; Seven genes were specific to OSF carcinogenomic samples: FLT4, HRAS, KMT2A, RB1, MLH1, KMT2D, FLCN (Table 2).

Mutations were detected in 50% (18/36) of OSCC cases and 65.8% (25/38) of OSF carcinogenesis cases, and the proportion of mutations in the OSF carcinogenesis group was higher than in the OSCC group alone, with 1.444 mutations/tumors and 1.6 mutations/tumors, respectively. The incidence of FLT4 gene mutations in OSF cancerous tissues was 18.4% (7/38), which was significantly higher than that in OSCC tissues alone (0/36) ($P<0.05$), and the difference was statistically significant (Table 3).

Discussion

Most patients with oral squamous cell carcinoma are diagnosed in the late stage, resulting in a 5-year survival rate of less than 50%(Bugshan & Farooq, 2020). The prognosis is still very poor, so early diagnosis and treatment are still the key to improve the survival rate of patients(Rivera, 2015). For many years, people have been looking for early tumor biomarkers to find prevention and better treatment options for oral cancer patients. In recent years, various sequencing technologies have found several mutations in oral cancer, including NOTCH1, FAT1,

CASP810, etc(Jayaprakash et al., 2020), and also identified mutations in the composition of tumor suppressor genes, oncogenes and mitochondrial genes, but no relevant reports have been reported on OSF oncogenic gene mutations. Compared with whole genome and whole exome analysis, there are few articles on gene Panel targeted sequencing in oral cancer genome sequencing, which has been used to analyze and detect mutations in breast cancer, ovarian cancer(Kraus et al., 2017) and non-small cell lung cancer(Mosele et al., 2020). Gene Panel targeted sequencing enriches and sequenced the genome region of interest through the capture kit, which enables rapid and cost-effective analysis of mutations in clinical samples and identification at the early stage of cancer occurrence, so as to achieve "early detection, early diagnosis and early treatment", providing an excellent opportunity for rapid clinical treatment decision-making. Panel targeting sequencing plays a unique role in the new generation of high-throughput sequencing, which has produced many exciting new discoveries and has been applied more and more widely. Therefore, we used Panel targeted sequencing to detect the most common and unique mutations in the OSF cancer group and OSCC group.

In our study, 7 specific gene mutations in the OSF cancer group were identified by gene Panel targeted sequencing technology. 7 mutations were identified in FLT4, which was the largest number of mutations, followed by 3 mutations in HRAS and 2 mutations in KMT2A. Combining the mutation frequency and distribution of OSCC and OSF cancer groups, we found that the mutation frequency of FLT4 was significantly higher in OSF cancer group, while the OSCC group had no mutation of FLT4 gene, and the mutation of FLT4 gene was unique to OSF cancer group. The analysis of the relationship between OSCC and FLT4 gene mutation with or without OSF showed that OSCC with or without OSF was correlated with FLT4 gene mutation, and FLT4 gene mutation in OSF cancer group was significantly higher than that in OSCC group alone.

FLT4, also known as vascular endothelial growth factor receptor 3 (VEGFR3), is a highly glycoylated single-chain transmembrane protein. Together with VEGF-C, it is the only group of regulatory factors for lymphangiogenesis in embryonic tissue and the physiological function of

lymphangiogenesis in mature individuals(Kaipainen et al., 1995). At the same time, it is also associated with cancer and its role in promoting new angiogenesis(Gore et al., 2011). Several studies have reported that the expression level of FLT4 in tumors is significantly positively correlated with the development of cancer cell metastasis and poor prognosis(Garouniatis et al., 2013; Martins et al., 2013). Xiao(Xiao et al., 2015) found that the increased expression of FLT4 was significantly positively correlated with the invasive tumor phenotype, resulting in poor survival rate, and was associated with lymph node metastasis of colorectal cancer and early and late death of patients.

Panel targeting sequencing found that the mutation frequency of FLT4 in cancer cells of OSCC patients with OSF was extremely high, while that of patients with simple OSCC was almost non-mutated. Chewing areca nuts was the main factor leading to OSF. Therefore, we speculated that, FLT4 gene may be related to OSCC caused by OSF canceration. At present, there are few studies on the influence and correlation of FLT4 on oral cancer cells. Therefore, in the later stage, we plan to observe whether the physiological function of cancer cells changes and the changes of its related pathways by knocking down the expression of FLT4 in oral cancer cells. In the current study, which used bioinformatics methods to identify candidate biomarkers, we anticipate that the identified FLT4 gene will contribute to our understanding of the molecular mechanisms behind the carcinogenesis of OSF into OSCC and to the identification of new targeted therapies.

Conclusions

In the present work, we found that the mutation frequency of FLT4 in cancer cells of OSCC patients with OSF was extremely high, while that of patients with simple OSCC was almost non-mutated by panel targeting sequencing. Therefore, we deem that FLT4 gene may be related to OSF carcinogenesis and is expected to be an early diagnostic biomarker for OSF carcinogenesis.

Acknowledgements

The authors gratefully acknowledge all the study participants and study staff for their help and cooperation during this study.

Funding Statement

This study was supported by National Natural Science Foundation of China (No.81960199) and Hainan Province Science and Technology special Fund (No.ZDYF2021SHFZ114).

Additional Information and Declarations

Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contributions

Bingxia Li conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
Xinyu Chen conceived and designed the experiments, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
Haiyu Xian performed the experiments, authored or reviewed drafts of the article, and approved the final draft.
Qitao Wen performed the experiments, authored or reviewed drafts of the article, and approved the final draft.
Tao Wang conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.

References

Bugshan, A., & Farooq, I. (2020). Oral squamous cell carcinoma: metastasis, potentially associated malignant disorders, etiology and recent advancements in diagnosis. *F1000Res*, *9*, 229. https://doi.org/10.12688/f1000research.22941.1

Garouniatis, A., Zizi-Sermpetzoglou, A., Rizos, S., Kostakis, A., Nikiteas, N., & Papavassiliou, A.

297 G. (2013). Vascular endothelial growth factor receptors 1,3 and caveolin-1 are implicated
298 in colorectal cancer aggressiveness and prognosis--correlations with epidermal growth
299 factor receptor, CD44v6, focal adhesion kinase, and c-Met. *Tumour Biol*, *34*(4), 2109-
300 2117. https://doi.org/10.1007/s13277-013-0776-1
301 Gore, A. V., Swift, M. R., Cha, Y. R., Lo, B., McKinney, M. C., Li, W., . . . Weinstein, B. M. (2011).
302 Rspo1/Wnt signaling promotes angiogenesis via Vegfc/Vegfr3. *Development*, *138*(22),
303 4875-4886. https://doi.org/10.1242/dev.068460
304 Jayaprakash, C., Varghese, V. K., Jayaram, P., Chakrabarty, S., Kudva, A., Ray, S., &
305 Satyamoorthy, K. (2020). Relevance and actionable mutational spectrum in oral
306 squamous cell carcinoma. *J Oral Pathol Med*, *49*(5), 427-434.
307 https://doi.org/10.1111/jop.12985
308 Kaipainen, A., Korhonen, J., Mustonen, T., van Hinsbergh, V. W., Fang, G. H., Dumont, D., . . .
309 Alitalo, K. (1995). Expression of the fms-like tyrosine kinase 4 gene becomes restricted to
310 lymphatic endothelium during development. *Proc Natl Acad Sci U S A*, *92*(8), 3566-3570.
311 https://doi.org/10.1073/pnas.92.8.3566
312 Kamps, R., Brandão, R. D., Bosch, B. J., Paulussen, A. D., Xanthoulea, S., Blok, M. J., &
313 Romano, A. (2017). Next-Generation Sequencing in Oncology: Genetic Diagnosis, Risk
314 Prediction and Cancer Classification. *Int J Mol Sci*, *18*(2).
315 https://doi.org/10.3390/ijms18020308
316 Kraus, C., Hoyer, J., Vasileiou, G., Wunderle, M., Lux, M. P., Fasching, P. A., . . . Reis, A. (2017).
317 Gene panel sequencing in familial breast/ovarian cancer patients identifies multiple novel
318 mutations also in genes others than BRCA1/2. *Int J Cancer*, *140*(1), 95-102.
319 https://doi.org/10.1002/ijc.30428
320 Lee, C. H., Ko, A. M., Warnakulasuriya, S., Yin, B. L., Sunarjo, Zain, R. B., . . . Ko, Y. C. (2011).
321 Intercountry prevalences and practices of betel-quid use in south, southeast and eastern
322 Asia regions and associated oral preneoplastic disorders: an international collaborative
323 study by Asian betel-quid consortium of south and east Asia. *Int J Cancer*, *129*(7), 1741-
324 1751. https://doi.org/10.1002/ijc.25809
325 Martins, S. F., Garcia, E. A., Luz, M. A., Pardal, F., Rodrigues, M., & Filho, A. L. (2013).
326 Clinicopathological correlation and prognostic significance of VEGF-A, VEGF-C, VEGFR-
327 2 and VEGFR-3 expression in colorectal cancer. *Cancer Genomics Proteomics*, *10*(2),
328 55-67.
329 Mosele, F., Remon, J., Mateo, J., Westphalen, C. B., Barlesi, F., Lolkema, M. P., . . . André, F.
330 (2020). Recommendations for the use of next-generation sequencing (NGS) for patients
331 with metastatic cancers: a report from the ESMO Precision Medicine Working Group. *Ann*
332 *Oncol*, *31*(11), 1491-1505. https://doi.org/10.1016/j.annonc.2020.07.014
333 Petljak, M., Dananberg, A., Chu, K., Bergstrom, E. N., Striepen, J., von Morgen, P., . . .
334 Maciejowski, J. (2022). Mechanisms of APOBEC3 mutagenesis in human cancer cells.
335 *Nature*, *607*(7920), 799-807. https://doi.org/10.1038/s41586-022-04972-y
336 Rivera, C. (2015). Essentials of oral cancer. *Int J Clin Exp Pathol*, *8*(9), 11884-11894.
337 Warnakulasuriya, S. (2009). Global epidemiology of oral and oropharyngeal cancer. *Oral Oncol*,

45(4-5), 309-316. https://doi.org/10.1016/j.oraloncology.2008.06.002

Warnakulasuriya, S. (2010). Living with oral cancer: epidemiology with particular reference to prevalence and life-style changes that influence survival. *Oral Oncol*, *46*(6), 407-410. https://doi.org/10.1016/j.oraloncology.2010.02.015

Xiao, X., Liu, Z., Wang, R., Wang, J., Zhang, S., Cai, X., . . . Fan, D. (2015). Genistein suppresses FLT4 and inhibits human colorectal cancer metastasis. *Oncotarget*, *6*(5), 3225-3239. https://doi.org/10.18632/oncotarget.3064

Table 1(on next page)

Patient characteristics of OSCC (n=74)

1 **Table 1 Patient characteristics of OSCC (n=74)**

variable	figure
Age (mean)	55.72±10.728（34-82）
Gender	
Male	52（70.3%）
Female	22（29.7%）
Primary site	
Tongue	32
Buccal mucosa	17
Gingiva	11
Other	14
TNM stage	
Ⅰ	7
Ⅱ	22
Ⅲ	24
Ⅳ	
Tumor size（cm）	
≤4（T I/2）	51
＞4（T 3/4）	23
lymph node metastasis	
Negative	40
Positive	34

2

Table 2(on next page)

Distribution of mutated genes in two groups

1 **Table 2 Distribution of mutated genes in two groups**

gene	OSCC Mutated genes	OSF Cancerous mutated genes
TP53	√	√
FLT4		√
PIK3CA	√	√
CDKN2A	√	√
FGFR4	√	√
HRAS		√
BRCA1	√	√
PTPN11	√	√
NF1	√	
KMT2A		√
RB1		√
PTEN	√	
MSH2	√	
MLH1		√
KMT2D		√
FLCN		√
BRCA2	√	
APC	√	

2

3

Table 3(on next page)

Clinicopathological features based on whether OSCC is associated with OSF

Note: Number of mutations: the number of mutated genes in each sample, for example: OSCC-mutation number≥ 1 is 18: it means that in the group [OSCC], the number of [number of mutated genes per sample≥1] is 18; Except for the t-test used for [mutation average], the rest used the chi-square test; [FLT4 mutation] was accurately tested using Fisher.

Table 3 Clinicopathological features based on whether OSCC is associated with OSF

	OSCC	OSF canceration	P value
Total	36（48.6%）	38（51.4%）	
Gender			
Female	15	7	0.05333
Male	21	31	
TNM stage			
Ⅰ、Ⅱ	13	16	0.772
Ⅲ、Ⅳ	23	22	
Lymph node metastasis			
Negative	21	19	0.4722
Positive	15	19	
Mutation number			
0	18	13	0.2542
≥1	18	25	
Mutation number			
average	1.444	1.6	0.4646
FLT4 mutations			
Negative	36	31	0.02094
Positive	0	7	

Note: Number of mutations: the number of mutated genes in each sample, for example: OSCC-mutation number≥ 1 is 18: it means that in the group [OSCC], the number of [number of mutated genes per sample≥1] is 18; Except for the t-test used for [mutation average], the rest used the chi-square test; [FLT4 mutation] was accurately tested using Fisher.

Figure 1

Mutation summary and gene waterfall plot for SCC-targeted NGS sequencing

A. Mutation aggregation. Mutation identification was performed using the "maftools" R program, which summarizes the data generated by NGS and shows the type of mutation and SNV classification, as well as the number of mutations, through a box plot；B. Gene waterfall chart. The figure, generated by the "maftools" R program, represents all types of mutations in highly mutated genes. Comment Multi_Hit: Refers to a gene that has mutated multiple times in the same sample. The figure shows a large number of nonsense and missense mutations in the TP53 gene.

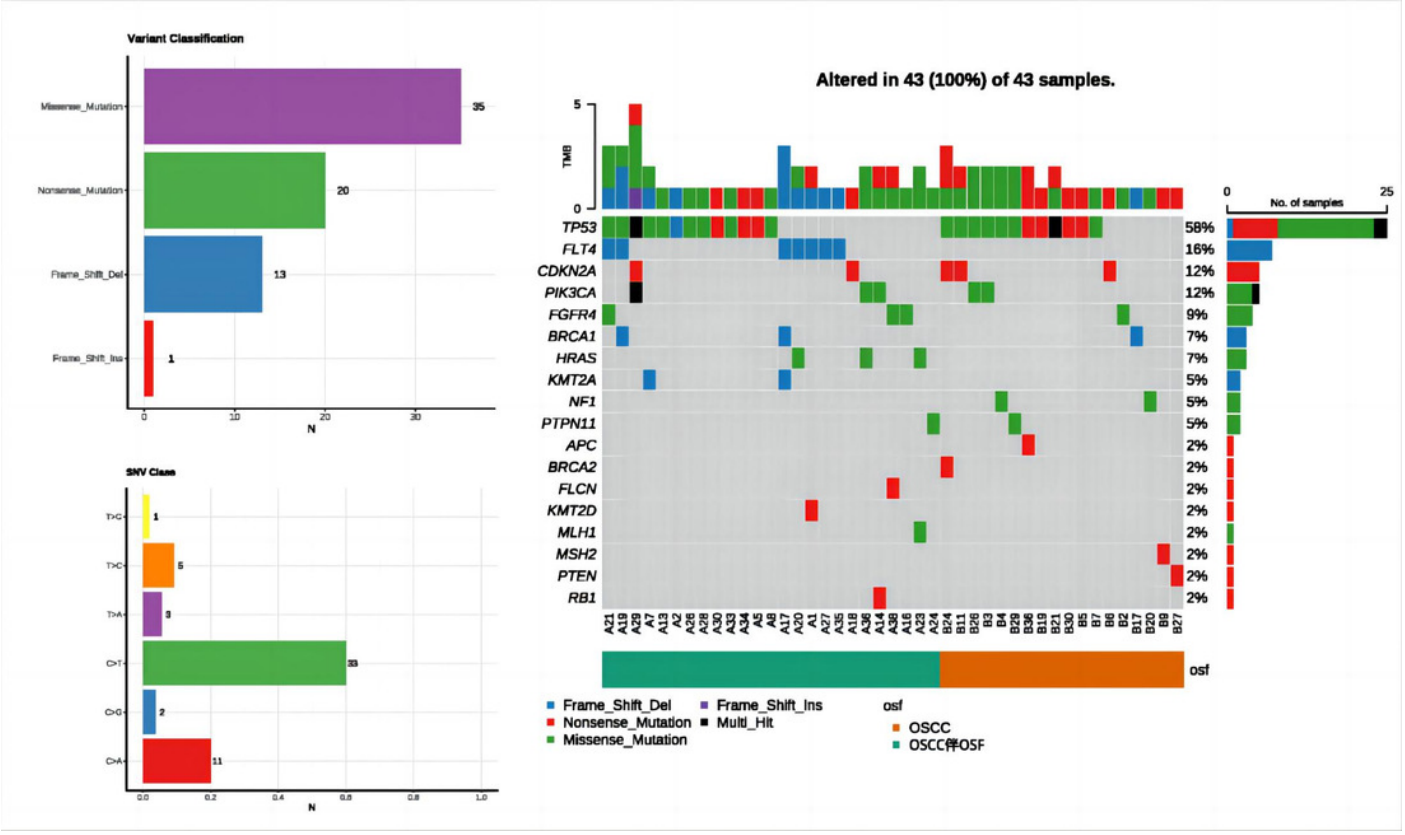


Figure 2

Mutation classification and proportion of OSCC-targeted NGS sequencing data

Mutation classification: The "maftools" R procedure is used to identify and classify mutations. The data was visualized to show six different base conversions a boxplot of the population distribution of types, a boxplot of the population distribution of transformations and reversals, and a plot showing the individual base conversion types in each sample stacked column chart of percentages

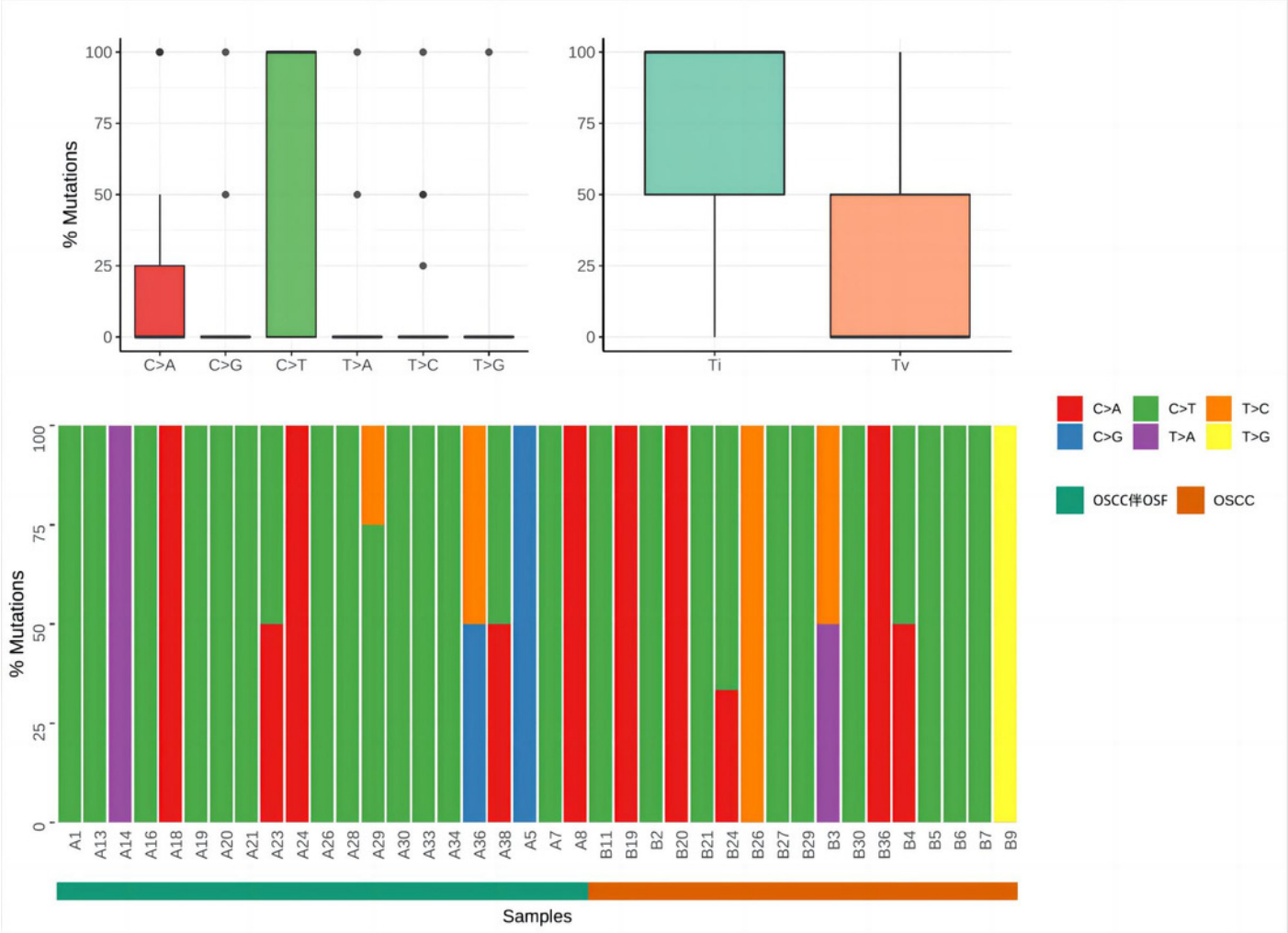


Figure 3

Mutation frequency bar plot, the abscissa is the mutation frequency in 74 samples, and the ordinate is the gene that has been mutated

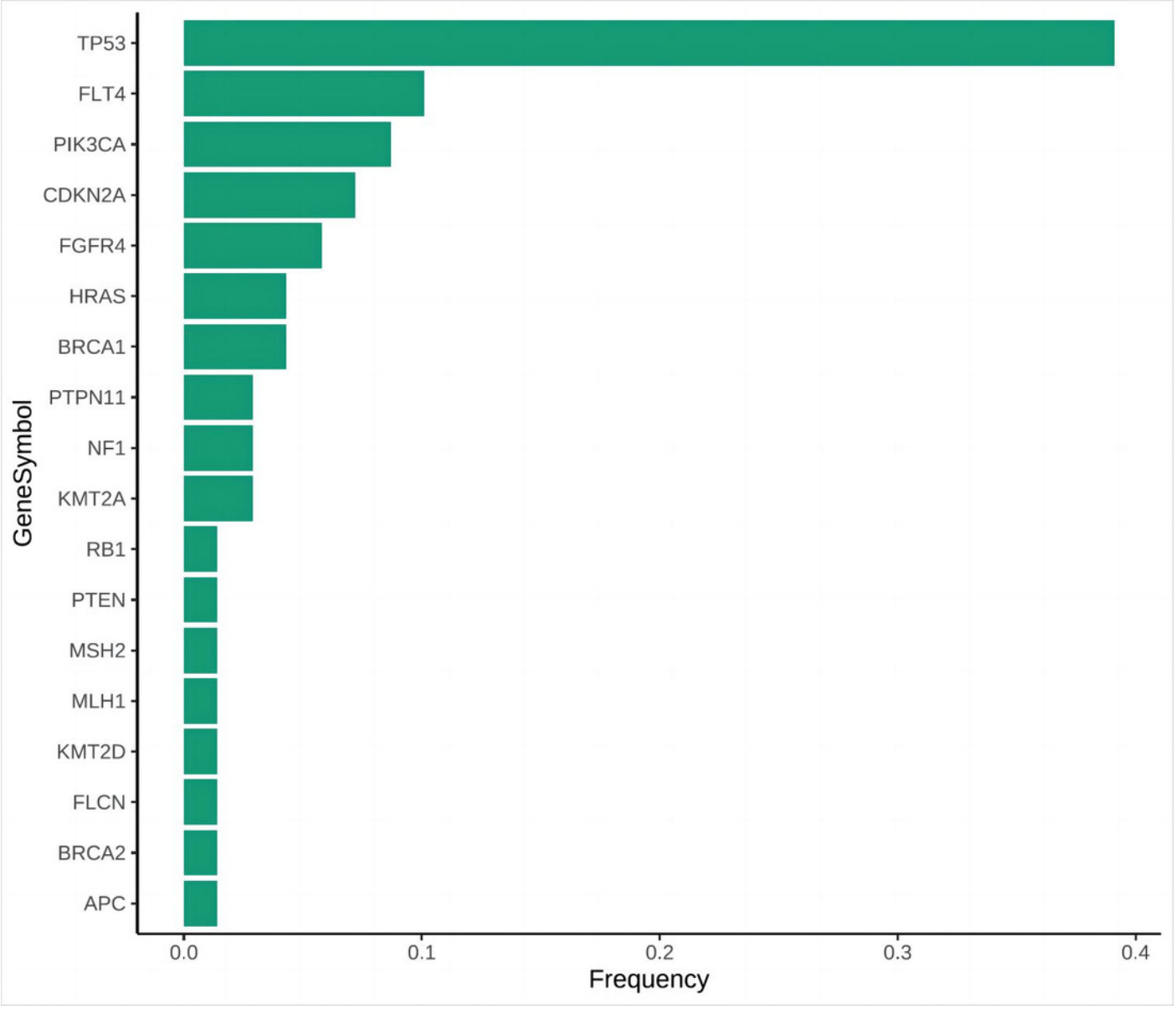
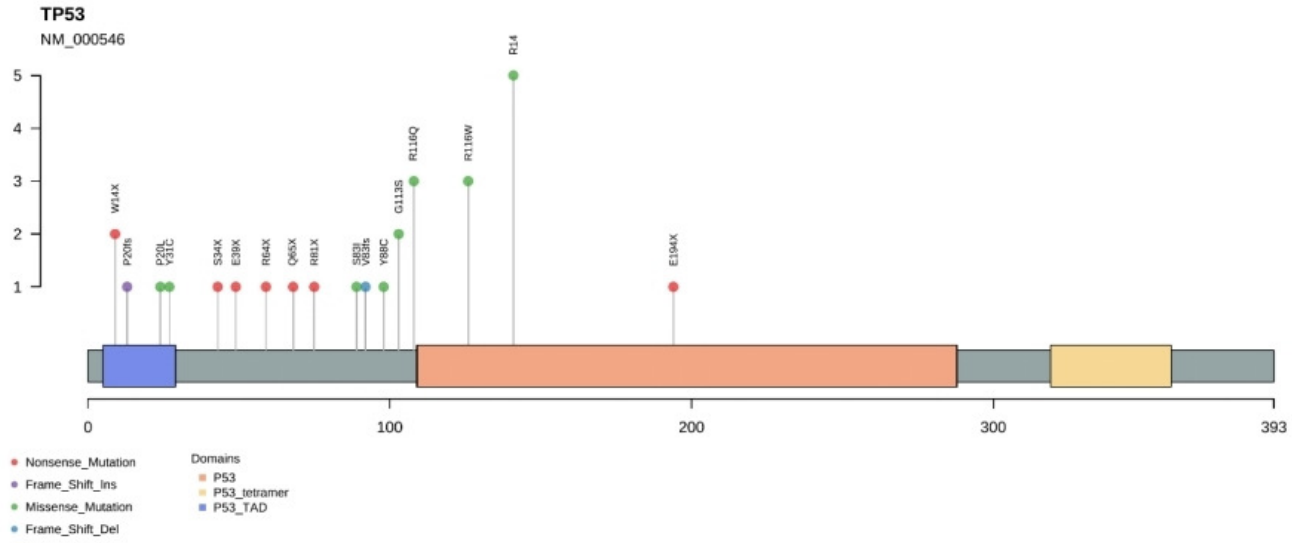
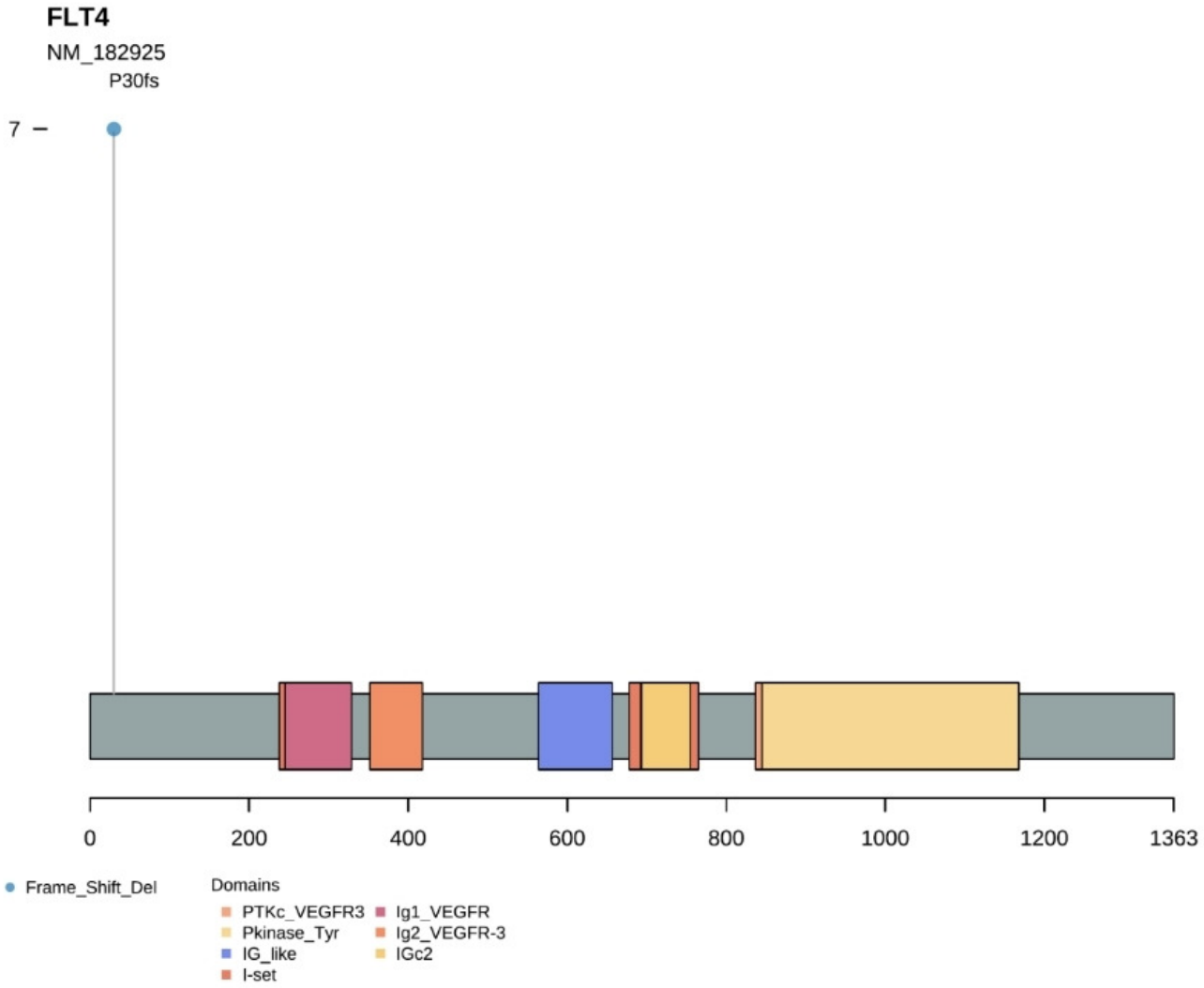


Figure 4

Needle plot of mutations. Mutations in selected significantly mutated genes across OSCC clinical specimens


TP53
NM_000546
5
4
3
2
1
W141X
P20fs
P20L
Y31C
S34X
E39X
R64X
Q65X
R81X
S83I
V83fs
Y88C
G113S
R116Q
R116W
R14
E194X
0
100
200
300
393
Nonsense_Mutation
Frame_Shift_Ins
Missense_Mutation
Frame_Shift_Del
Domains
P53
P53_tetramer
P53_TAD


FLT4
NM_182925
P30fs
7
0
200
400
600
800
1000
1200
1363
Frame_Shift_Del
Domains
PTKc_VEGFR3
Pkinase_Tyr
IG_like
I-set
Ig1_VEGFR
Ig2_VEGFR-3
IGc2