Int J Clin Exp Pathol 2020;13(7):1909-1914
www.ijcep.com /ISSN:1936-2625/IJCEP0109451

Original Article

Gene expression analysis of primary gingival cancer by whole exome sequencing in thirteen Chinese patients

Guo-Min Liu[2,4], Tian-Cheng Lu[3,4], Yun Liu[2,4], Yun-Gang Luo[1,4]

Departments of [1]Stomatology, [2]Orthopedics, The Second Hospital of Jilin University, Changchun 130041, Jilin, China; [3]Life Sciences College, Jilin Agricultural University, Changchun 130118, Jilin, China; [4]Jilin Provincial Changbai Mountain Anti-tumor Medicine Engineering Center, Changchun, Jilin, China

Received February 18, 2020; Accepted April 13, 2020; Epub July 1, 2020; Published July 15, 2020

Abstract: Objectives: Early diagnosis of and markers for gingival oral squamous cell carcinoma (OSCC) is important for effective treatment. Methods: The current study performed a whole exome sequencing of gingival OSCC tissues in thirteen Chinese patients to explore exonic mutants. Results: Eighty-five genes emerged as mutants in patients with primary gingival OSCC. CCL4L1 presented a G>A transversion at chr17 17q12, position 36212480, exon 3. KDM5B presented a T>TA insertion at chr1 1q32.1, position 202766506, exon 6. ANKRD36C presented a C>G transition at chr2 2q11.1, position 95945175, exon 18. Conclusion: These three mutants might be new markers of gingival OSCC. The finding may provide new targets to diagnose and treat gingival OSCC.

Keywords: Whole exome sequencing, gingival cancer, CCL4L1, KDM5B, ANKRD36C

Introduction

Oral cancer is a head and neck cancer involving the oral cavity [1]. Early gingival tumors invade the alveolar process and jaw bone, and teeth become loose, displaced and even fall out, complicated by local pain. Gingival oral squamous cell carcinoma (OSCC) is the most prevalent malignant tumor of the oral and maxillofacial region in China, with more than 50,000 cases worldwide every year [2, 3]. Surgery is the main clinical treatment, accompanied by postoperative radiotherapy. Radiotherapy greatly reduces the post-operative quality of life. Therefore, how to reduce the operative field of patients and provide targeted therapy for gingival OSCC without side effects is an urgent problem. If oral cancer is diagnosed in its earliest stages, treatment is generally very effective.

Oncogenes are activated as a result of gene mutation. These markers would allow early diagnosis of gingival OSCC. Studies have suggested genes marking oral cancer include CCL18, ANKRD17, KDM4A, FHIT, CALR, H3K9ac, LAMP3, Nrp2, SAMD4A, and WNT5A [4-10]. Gingival OSCC may have peculiar gene mutations unlike other oral cancers. However, information is still insufficient.

Whole exome sequencing (WES) is a genomic technique for sequencing all of the exome. It consists of two steps: the first step is to select only the subset of DNA that encodes proteins. The second step is to sequence the exonic DNA using any high-throughput DNA sequencing technology [11]. WES has been applied in academic research and the clinical diagnosis. The current study performed WES in thirteen Chinese patients with gingival OSCC to explore exonic mutants contributing to gingival OSCC.

Patients and methods

Patients

The current study included 13 patients diagnosed with primary gingival OSCC in the Jilin University Second Hospital, Department of Stomatology (Changchun, Jilin province, China) during 14 October 2016 to 24 November 2017. Their pericancerous tissues were used as controls (CK). All patients gave informed consent.

The study was approved by the Jilin University No. 2 Hospital Ethics Committee, Changchun, Jilin, China.

Exome sequencing

The gingival tumor biopsies and pericancerous tissues were removed from patients in outpatient clinic. Immediately, they were stored in liquid nitrogen. The genomes were extracted using the DNA extract kit (Invitrogen) according to the manufacturer's protocols. DNA quality was assessed by Agilent BioAnalyzer 2100. The genomic regions of whole exons were captured by microarray with synthesized oligonucleotide probes hybridized to a fragmented genomic DNA sample (Aglient SureSelect Human All Exon V6). High throughput sequencing was performed on the Illumina HiSeq 2000 system (Illumina). The datasets generated or analyzed during the current study are available from the corresponding author on request.

Data quality control and genome alignment

The quality evaluation report was generated by FastQC, and the specific quality control standards for sequencing data were as follows: (1) BWA algorithm was used to remove low quality areas at both ends of the sequence, with a quality threshold of 30. (2) Remove sequences containing joints; (3) Remove sequences containing fuzzy base N; (4) Remove sequences with length less than 60 bp. The processing results of the raw sequencing data (Raw Data) were called Clean Data for use in subsequent analysis.

In order to detect a variation, Clean Data were aligned to reference Genome. In the mutation site, the Clean Reads did not completely match with the reference Genome forming mismatches or gaps. Subsequent variants were detected on the basis of this information [12, 13]. Reference genome alignment and calibration were operated as follows: 1. The BWA software was used to align Clean Data to reference genome ftp://ftp.ensembl.org/pub/release-85/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna.primary_assembly.fa.gz., generating a *.Sam format file. 2. The Picard software was used to sort the *.Sam files according to chromosomes and loci, generating a *.bam format file. 3. The Picard software was used to mark the duplicates, which were derived from overamplification in the process of library preparation and were not used as evidence for subsequent mutation detection. 4. Reads aligning to near an insertion/deletion (In/Del) position were locally aligned again to reduce the error rate of alignment. 5. Base Quality Score Recalibration (BQSR) was performed to avoid Single Nucleotide Polymorphisms (SNP) that were caused by low Quality bases in sequencing. 6. Finally, the optimal alignment results were obtained.

Mutation detection, filtration and annotation

SNP refers to a single nucleotide variation with a mutation frequency of more than 1%. The total number of SNPs in the human genome is about 3×10^6. SNP has become a third-generation genetic marker, and many phenotypic differences, and susceptibility to drugs or diseases may be related to SNP. In/Del refers to the insertion/deletion of small fragments in a sample relative to the reference genome, and In/Del may contain one or more bases.

The main steps for the detection, filtering, and annotation of SNP and In/Del variations were as follows: 1. Based on the *.bam files that were generated in the previous step, GATK software was used to identify the SNP and In/Del variations, forming a variation detection file in the *.VCF format. 2. The SNP and In/Del sites were re-calibrated using the GATK VQSR software, and the sites through the filtering were marked as PASS. 3. The ANNOVAR software was used to annotate the results of *.VCF file, including the gene structure function, chromosome region (cytoBand, genomicSuperDups) and other mutation databases: dbSNP (http://www.ncbi.nlm.nih.gov/snp), 1000Genomes (http://www.1000genomes.org/), ESP6500 (http://evs.gs.washington.edu/EVS), SIFT (http://sift.jcvi.org/), Polyphen (http://genetics.bwh.harvard.edu/pph2/index.shtml), MutationTaster (http://www.mutationtaster.org), GERP++ (http://mendel.stanford.edu/SidowLab/downloads/gerp).

Alignment to East Asian population and examination of cancer-related mutation

The above data was aligned to the East Asian population as normal CK in the 1000 g database (1000g2015aug_eas). If a mutation was not annotated in the database, it was consid-

Table 1. Mutation types of variants passing through VQSR filtration

var_type	PASSED	FILTERED	total
In/del	259744	2634	262378
del	143951	1221	145172
ins	111761	1319	113080
unknown	4032	94	4126
SNP	977296	44852	1022148
ts	657696	26459	684155
tv	319001	18240	337241
unknown	599	153	752
total	1237040	47486	1284526

ered that the mutation did not occur in the normal East Asian population, thus reserving these filtered data. Filtered data were annotated in the OMIM (Online Mendelian Inheritance in Man) database to filter out the sites of amino acid sequence changes in the exon region and the sites of splicing mutation, obtaining the final filtered data. Considering that mutations different from the normal East Asians' database did not certainly lead to oncogenesis, the known oral cancer-related genes were examined in the WES data of 13 gingival OSCC patients. The known oral cancer marker genes included CCL18, ANKRD17, KDM4A, FHIT, CALR, H3K9ac, LAMP3, Nrp2, SAMD4A, WNT5A, which were described previously [4-10].

Results and discussion

Quality control statistics of raw sequencing data and reference genome alignment RawData were filtered to get CleanData. Quality assessment before and after data processing was generated by FastQC. The average length of retained sequences was about 142 bp. The high-scored clean data were analysed in the next step. Clean Reads did not match with the reference genome at the mutation site, forming a mismatch or gap. The results showed that the covered target was 99.8%-99.97%, indicating a good degree of matching.

Mutation detection, filtering, and annotation

SNP and In/Del mutations were filtered using VQSR software, and the results are shown in **Table 1**. SNP contained subtypes ts (transition: substitution between purine and purine, or substitution between pyrimidine and pyrimidine) and tv (transversion: substitution between purine and pyrimidine). For a mutation site of unknown subtype, the allele number was more than 2, so the software could not recognize it correctly. The number of mutations for the SNP subtype was higher than that of the In/del subtype. Insertion mutations and deletion mutations were equivalent in the number while the transition mutations were higher than the transversion mutations in the SNP subtype. Functional annotation of mutation sites and statistics of variation subtype of coding gene exon region are shown in **Tables 2** and **3**.

Alignment to normal East Asians and examination of cancer-related genes

After aligning to the normal East Asians, the obtained data were annotated in OMIM database. It showed 667 mutation sites and 85 differential genes. These genes are shown in **Table 4**. The mutation subtypes of all these 85 putative mutation genes were splicing mutations (data available if needed). A splice site mutation is a genetic mutation that inserts, deletes, or changes a number of nucleotides in the specific site at which splicing takes place during the processing of precursor messenger RNA into mature messenger RNA. In these 85 mutated genes, no genes were previously reported relating to cancer. Therefore, these genetic mutations were possible markers for gingival OSCC.

Considering that mutations different from the normal East Asians' database did not surely lead to oncogenesis, the current study examined the mutation subtype of known oral cancer-related genes in the WES data of 13 gingival OSCC patients. Consequently, no known cancer marker genes completely matched with any of the 85 putative mutated genes. Fortunately, CCL4L1, KDM5B and ANKRD36C that are listed in **Table 4** were close to three known oral cancer markers, CCL18, KDM4A, ANKRD17 that were reported [4-6].

In the results, CCL4L1 presented a G>A transversion with details of chr17 17q12, position 36212480, AVSNP rs4796195, NM_207007: exon 3:c.192-2G>A. CCL4L1, Chemokine (CC motif) ligand 4 like 1 was a lymphocyte activation gene, which is expressed higher in severe psoriasis [14]. The CCL4L1 G>A transversion might be a marker for the gingival OSCC. The

Table 2. Functional annotation of mutation sites

Func.refGene	PASSED	FILTERED	total
exonic	35748	3609	39357
exonic; splicing	15	3	18
splicing	173	43	216
UTR3	17516	603	18119
UTR5	6729	211	6940
UTR5; UTR3	32	10	42
intronic	469561	8201	477762
upstream	13488	541	14029
downstream	10235	325	10560
upstream; downstream	765	48	813
intergenic	616438	29417	645855
ncRNA_exonic	7399	1134	8533
ncRNA_exonic; splicing	3	1	4
ncRNA_splicing	32	6	38
ncRNA_UTR5	1	0	1
ncRNA_intronic	58905	3334	62239
total	1237040	47486	1284526

Table 3. Statistics of variation subtype of coding gene exon region

ExonicFunc.refGene	PASSED	FILTERED	total
synonymous_SNV	17541	1502	19043
nonsynonymous_SNV	16573	1801	18374
frameshift_deletion	238	32	270
frameshift_insertion	186	32	218
nonframeshift_deletion	411	91	502
nonframeshift_insertion	262	7	269
stopgain	175	29	204
stoploss	20	1	21
unknown	357	117	474
total	35763	3612	39375

current study linked CCL4L1 to a cancer for the first time. In the literature, similar chemokine (CC motif) ligand 18 (CCL18) is found to be involved in the remodeling of tumor microenvironment and have a critical role in oral squamous cell carcinoma, and CCL18 upregulates Slug expression to promote stem-cell like features by activating the mTOR pathway [4, 5]. Whether the function of CCL4L1 might be associated with CCL18 should be studied next.

KDM5B presented a T>TA insertion with details of chr1 1q32.1, position 202766506, AVSNP rs758523577, NM_001314042: exon 6:c.712-2->A. KDM5B, lysine demethylase 5B is a marker in various squamous cancers [15-18]. KDM5B is regulated by anti-oncogene miR424-5p, which inhibits cervical cancer cell by the Notch signaling pathway [15]. miR-194 targeting KDM5B inhibits promotes apoptosis of esophageal squamous cell carcinoma cells [16]. KDM5B (also name JARID1B) functions at SHIP1/AKT signaling in human hypopharyngeal squamous cell carcinoma [17]. Altered splicing of JARID1B (KDM5B) and its RBP2-H1 variant may be important factor increasing melanoma aggressiveness [18]. On the function of KDM5B in gingival OSCC, the current study provided a new putative mutation of ->A insertion with details of chr1 1q32.1, position 202766506, AVSNP rs758523577, NM_001314042: exon 6:c.712-2->A. This finding might provide gingival OSCC with a new marker or extend KDM5B to be a marker for gingival OSCC.

ANKRD36C presented a C>G transition with details of chr2 2q11.1, position 95945175, AVSNP rs75080810, NM_001310154: exon 18:c.1363-1C>G. ANKRD36C ankyrin repeat domain 36C and mutant ANKRD36C are related to location of gastric cancer and cases with mutant and wild-type ANKRD36C show different locations in gastric body cancer [19]. The current study found a mutant ANKRD36C with a C>G transition, which may be a marker of gingival OSCC.

Considering that "peri-cancerous" tissue might not be entirely normal (for example, might contain dysplasia) and might provide misleading results, the sequence output of gingival OSCC biopsies was not normalized to the "peri-cancerous" adjacent tissue, but simply to a normal East Asian population sequence in the current study. One customary way to normalize sequence data and begin to identify mutations would be normalizing to regular tissue such as peripheral blood. Unfortunately, we were unable to persuade patients to have their normal tissues or blood used since the gingival cancer tissues were often removed by outpatient surgery directly. Thus, normalizing to a putatively normal East Asian population would be inadequate and might overcall normal sequence variations. To proceed with a more standard exome

Table 4. List of 85 putative mutation genes aligning to normal East Asians

	ACTR8	ATP5B	ATXN3	AK7	ACLY	ATP6V1H	AP3S1
ARMC4	ACADSB	ANKRD36C	B3GAT3	BAGE	BAGE4	BAGE5	CORIN
CNOT1	CNTNAP3	C16orf52	C4orf46	CCDC113	CCDC124	CBWD3	CEACAM21
CASC1	CCL4L1	DOC2A	FOXP2	FANCD2	FOXK2	FRG1	FCGR2A
FTH1	GXYLT1	HLA-DRB5	HLA-DRB1	HLA-DQB1	HEATR1	HDGFRP2	INO80C
ITGA10	IRF5	KMT2C	KDM5B	KCTD20	LILRA1	LRMP	MUC19
MYCBP2	MYL6	MROH5	NBPF10	NPIPB9	NUP153	ORAI1	POLR3H
PYGO1	PDCD2L	PKD2L2	PRSS3	PKD2L2	POLR2A	PSPH	RAD51B
RBBP6	RAET1E	RAD17	RSU1	SKA3	SP140L	SLC43A2	STAM2
SEC24D	TDG	TPTE	TRPM3	TMED10	TMEM2	VIM	WDSUB1
XPO1	ZNF280D	ZNF98	ZFC3H1	ZSCAN30			

sequence normalization protocol and additionally increase the number of samples would improve the validity of the study in the future.

Conclusion

Eighty-five mutant genes were found in primary gingival OSCC patients. CCL4L1, KDM5B, and ANKRD36C mutants might be new markers of gingival OSCC. The finding might help to provide new targets to diagnose and treat gingival OSCC.

Acknowledgements

The paper is funded by Jilin Provincial Development and Reform Commission, Jilin province of China (grant no. 3J115AK93429; 2014G072). The project was approved by the Committee of Jilin University Ethics, Changchun, Jilin, China.

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Yun-Gang Luo, Department of Stomatology, The Second Hospital of Jilin University, 218 Ziqiang Street, Changchun 130041, Jilin, China. E-mail: luoyungang@jlu.edu.cn

References

[1] Gomez-Dantes H, Fullman N, Lamadrid-Figueroa H, Cahuana-Hurtado L, Darney B, Avila-Burgos L, Correa-Rotter R, Rivera JA, Barquera S, Gonzalez-Pier E, Aburto-Soto T, de Castro EFA, Barrientos-Gutierrez T, Basto-Abreu AC, Batis C, Borges G, Campos-Nonato I, Campuzano-Rincon JC, Cantoral-Preciado AdJ, Contreras-Manzano AG, Cuevas-Nasu L, de la Cruz-Gongora VV, Diaz-Ortega JL, Garcia-Garcia MdL, Garcia-Guerra A, Gonzalez de Cossio T, Gonzalez-Castell LD, Heredia-Pi I, Hijar-Medina MC, Jauregui A, Jimenez-Corona A, Lopez-Olmedo N, Magis-Rodriguez C, Medina-Garcia C, Medina-Mora ME, Mejia-Rodriguez F, Montanez JC, Montero P, Montoya A, Moreno-Banda GL, Pedroza-Tobias A, Perez-Padilla R, Quezada AD, Richardson-Lopez-Collada VL, Riojas-Rodriguez H, Blancas MJR, Razo-Garcia C, Romero Mendoza MP, Sanchez-Pimienta TG, Sanchez-Romero LM, Schilmann A, Servan-Mori E, Shamah-Levy T, Tellez-Rojo MM, Texcalac-Sangrador JL, Wang H, Vos T, Forouzanfar MH, Naghavi M, Lopez AD, Murray CJ and Lozano R. Dissonant health transition in the states of Mexico, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2016; 388: 2386-2402.

[2] Bernier J, Bentzen SM and Vermorken JB. Molecular therapy in head and neck oncology. Nat Rev Clin Oncol 2009; 6: 266-277.

[3] Chamorro-Petronacci C, Perez-Sayans M, Elena Padin-Iruegas M, Marichalar-Mendia X, Gallas-Torreira M and Garcia Garcia A. Differential expression of snoRNAs in oral squamous cell carcinomas: new potential diagnostic markers. J Enzyme Inhib Med Chem 2018; 33: 424-427.

[4] Wang H, Liang X, Li M, Tao X, Tai S, Fan Z, Wang Z, Cheng B and Xia J. Chemokine (CC motif) ligand 18 upregulates Slug expression to promote stem-cell like features by activating the mammalian target of rapamycin pathway in oral squamous cell carcinoma. Cancer Sci 2017; 108: 1584-1593.

[5] Jin X, Xu H, Wu X, Li T, Li J, Zhou Y, Dan H, Jiang L, Zeng X, Ji P and Chen Q. KDM4A as a prognostic marker of oral squamous cell carcinoma: Evidence from tissue microarray studies in a multicenter cohort. Oncotarget 2017; 8: 80348-80357.

[6] D'Souza W, Pradhan S and Saranath D. Multiple single nucleotide polymorphism analysis and association of specific genotypes in FHIT, SAMD4A, and ANKRD17 in Indian patients with oral cancer. Head Neck 2017; 39: 1586-1595.
[7] Harada K, Takenawa T, Ferdous T, Kuramitsu Y and Ueyama Y. Calreticulin is a novel independent prognostic factor for oral squamous cell carcinoma. Oncol Lett 2017; 13: 4857-4862.
[8] Webber LP, Wagner VP, Curra M, Vargas PA, Meurer L, Carrard VC, Squarize CH, Castilho RM and Martins MD. Hypoacetylation of acetylhistone H3 (H3K9ac) as marker of poor prognosis in oral cancer. Histopathology 2017; 71: 278-286.
[9] Lu J, Ma H, Lian S, Huang D, Lian M, Zhang Y, Huang J and Feng X. Clinical significance and prognostic value of the expression of LAMP3 in oral squamous cell carcinoma. Dis Markers 2017; 2017: 1218254.
[10] Prgomet Z, Andersson T and Lindberg P. Higher expression of WNT5A protein in oral squamous cell carcinoma compared with dysplasia and oral mucosa with a normal appearance. Eur J Oral Sci 2017; 125: 237-246.
[11] Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA and Shendure J. Targeted capture and massively parallel sequencing of 12 human exomes. Nature 2009; 461: 272-U153.
[12] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G and Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009; 25: 2078-2079.
[13] Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 2011; 27: 2987-2993.
[14] Pedrosa E, Carretero-Iglesia L, Boada A, Colobran R, Faner R, Pujol-Autonell I, Palou E, Esteve A, Pujol-Borrell R, Ferrandiz C, Juan M and Carrascosa JM. CCL4L polymorphisms and CCL4/CCL4L serum levels are associated with psoriasis severity. J Invest Dermatol 2011; 131: 1830-7.
[15] Zhou Y, An Q, Guo RX, Qiao YH, Li LX, Zhang XY and Zhao XL. miR424-5p functions as an antioncogene in cervical cancer cell growth by targeting KDM5B via the Notch signaling pathway. Life Sci 2017; 171: 9-15.
[16] Cui G, Liu D, Li W, Li Y, Liang Y, Shi W and Zhao S. miR-194 inhibits proliferation and invasion and promotes apoptosis by targeting KDM5B in esophageal squamous cell carcinoma cells. Exp Biol Med (Maywood) 2017; 242: 45-52.
[17] Zhang J, An X, Han Y, Ma R, Yang K, Zhang L, Chi J, Li W, Llobet-Navas D, Xu Y and Jiang Y. Overexpression of JARID1B promotes differentiation via SHIP1/AKT signaling in human hypopharyngeal squamous cell carcinoma. Cell Death Dis 2016; 7: e2358.
[18] Kuzbicki L, Lange D, Stanek-Widera A and Chwirot BW. Prognostic significance of RBP2-H1 variant of JARID1B in melanoma. BMC Cancer 2017; 17: 854.
[19] Cong T, Liu GX, Cui JX, Zhang KC, Chen ZD, Chen L, Wei B and Huang XH. Exome sequencing of gastric cancers screened the differences of clinicopathological phenotypes between the mutant and the wide-type of frequently mutated genes. Zhonghua Yi Xue Za Zhi 2018; 98: 2242-2245.