Whole-Exome Sequencing of Esophageal Submucosal Gland Duct Adenocarcinoma

Zhu Zhu

The first affiliated hospital of zhengzhou university

Bo Qin

The first affiliated hospital of zhengzhou university

Beibei Wang

The first affiliated hospital of zhengzhou university

Kehan Liu

The first affiliated hospital of zhengzhou university

Xiao Hu

The first affiliated hospital of zhengzhou university

Bing Chen

The first affiliated hospital of zhengzhou university

Xiaohan Yao

The first affiliated hospital of zhengzhou university

Yihui Ma

The first affiliated hospital of zhengzhou university

Zhengyang Wang (✉ iamwangzhao@163.com)

The first affiliated hospital of zhengzhou university <https://orcid.org/0000-0002-7498-2436>

Research Article

Keywords: Esophageal submucosal gland duct adenocarcinoma, whole-exome sequencing, predisposition genes, TP53

Posted Date: August 11th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-791813/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Purpose Esophageal submucosal gland duct adenocarcinoma (ESGDAC) is an extremely rare histological variant of esophageal epithelial malignant tumors. To date, only few cases of ESGDAC have been reported. No comprehensive molecular analysis of ESGDAC has been conducted. This study focusses on the whole-exome sequencing (WES) to identify somatically acquired mutations and signatures of somatic mutations in ESGDAC.

Methods Immunohistochemical analysis and whole-exome sequencing (WES) was performed in 3 ESGDAC genomes.

Results In this study, we summarized the clinical, pathological, and immunohistochemical characteristics of patients with ESGDAC, we first performed whole-exome sequencing (WES) to summarize WES coverage statistics for each patient. We also examined cancer predisposition genes, ESGDAC-associated driver genes, and significantly mutated genes (SMGs). WES revealed TP53 as a SMG in ESGDAC.

Conclusion We studied the tumor mutation profiling for the development of future targeted therapies for ESGDAC.

Introduction

Esophageal squamous cell carcinoma (ESCC) and esophageal adenocarcinoma (EAC) remain the top two prevalent esophageal epithelial malignant tumors. The majority of adenocarcinomas of the esophagus originate from the Barrett's esophagus and heterotopic gastric mucosa of the esophagus (Lam 2018). Esophageal submucosal glands (SMGs) reside within the submucosal layer of the esophagus (Nie et al. 2016). Esophageal submucosal gland duct adenoma (ESGDA), an extremely rare type of esophageal adenocarcinoma, originates from esophageal submucosal gland duct (ESGD) and exhibits similar, while different histological and immunohistochemical features compared to Barrett's-associated adenocarcinoma and salivary gland-type carcinoma (Tamura et al. 2017).

To date, a limited number of esophageal submucosal gland duct adenocarcinoma (ESGDAC) cases have been reported. The most significant histological feature of ESGDAC is the double layers of cells in structure (including CK7-positive luminal cells and p63-positive basal cells), indicating esophageal gland duct differentiation (Tamura et al. 2017). Due to the similarity of histopathological and immunohistochemical characteristics, ESGDAC is mainly misdiagnosed as other adenocarcinomas, containing a bi-layered adenocarcinomatous element, such as basaloid squamous cell carcinoma and salivary gland-type adenocarcinoma. Hence, awareness of the discrepancy is vital to make correct diagnostic and therapeutic decisions.

In order to reduce the uncertainty and misunderstanding caused by the similar pathological features, additional studies should be conducted to further characterize the molecular profile of this entity. Through sequencing of the almost all exons in genomes, the whole-exome sequencing (WES) provides new opportunities to identify mutations and to establish a novel genetic basis for rare diseases in sporadic cases (Wang L. et al. 2017). The present study aimed to summarize the clinical, histopathological, and

immunohistochemical features of patients with ESGDAC, and WES was performed to identify somatically acquired mutations and signatures of somatic mutations in 3 ESGDAC genomes.

Methods

Patients and collection of specimens

Tumor and adjacent non-tumor tissues were isolated from 5 ESGDAC patients who underwent surgical resection at the Department of Thoracic Surgery, The First Affiliated Hospital of Zhengzhou University (Zhengzhou, China) from January 2018 to October 2020. Informed consent was obtained from all patients, and the study was approved by the Research and Clinical Trial Ethics Committee of The First Affiliated Hospital of Zhengzhou University (Approval No. 2021-KS-HNSR097).

The unanimous diagnostic criteria for ESGDAC have not been presented yet. Thus, based on pathological features of reported cases, the following diagnostic criteria were considered to select eligible cases: 1. The majority of neoplastic tubules had a bi-layered or a multi-layered structure (with inner luminal and outer epithelial layers), and could be differentiated from esophageal gland ducts, which were clearly confirmed by immunohistochemical and morphological examinations; 2. The tubules exhibited infiltrative growth, and moderately-to-poorly differentiated tubules could be observed locally; 3. The absence of Barrett's esophagus or heterotopic gastric mucosa of the esophagus (Kono et al. 2019, Tamura et al. 2017). Tumors that were diagnosed as the ESGDAC should meet the above-mentioned criteria. The medical records of all cases were reviewed and analyzed.

Immunohistochemical analysis

All specimens included in this study were fixed with 10% neutral buffered formalin and embedded into paraffin. Immunohistochemical analysis was performed following EnVision method. The antibodies used in this study are listed in Table 1.

Quantification and qualification of DNA

The genomic DNA was extracted using DNeasy Blood & Tissue kit (QIAGEN, Hilden, Germany), according to the manufacturer's instructions. Degradation and contamination of DNA were monitored on a 1% agarose gel. DNA concentration was measured by Qubit® DNA Assay kit in Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA).

WES and data analysis

The total amount of each sample was 0.6 µg of genomic DNA, and the sample was fragmented into an average size of 180~280 bp. We also utilized Illumina double-end kit (Illumina, Inc., San Diego, CA, USA) for construction of DNA libraries and Agilent SureSelect Human All ExonV6 kit (Agilent Technologies Inc., Santa Clara, CA, USA) for exome capture. We used the Illumina Novaseq platform (Illumina Inc.) to complete genomic DNA sequencing, and 150 bp paired-end reads were finally generated by Genechem Bioinformatics Co., Ltd. (Beijing, China).

Through base calling, the original fluorescence image files were converted into short reads (Raw data) and were recorded in FASTQ format. We conducted quality control (QC) through in-house analysis. The following reads were discarded: a) Each read contained >10 nucleotides aligned to the adapter contamination ($\leq 10\%$ mismatches); b) In each read, more than 10% of bases were uncertain; c) In each read, the low-quality (Phred quality <5) bases exceeded 50%.

Somatic mutation calling and signatures of somatic mutations

Somatic single-nucleotide variants (SNVs) were detected by muTect (Cibulskis et al. 2013). Mutect run in a default setting by taking tumor and normal sample data as input. Somatic insertions and deletions (InDels) were analyzed by Strelka (Saunders et al. 2012). Control-FREEC was used to detect somatic copy number variations (CNVs) (Boeva et al. 2012), and Genomic Identification of Significant Targets in Cancer (GISTIC) method was used to infer recurrently amplified or deleted genomic regions. The variant data were annotated using ANNOVAR software (Wang K. et al. 2010). Somatic SNVs were annotated by the 96 possible trinucleotide context substitutions (mutated base plus its sequence context). The signatures were identified by the SomaticSignatures package according to the non-negative matrix factorization (NMF) algorithm. The identified signatures were compared to 30 consensus signatures from COSMIC by cosine similarity in R programming language.

Cancer predisposition genes (CPGs)

Germline mutations were detected by SAMtools (Li et al. 2009) in normal genome, and filtered according to the following strategies: 1. Exclude SNPs with a depth less than 10x; 2. Filter out SNP loci in dbSNP databases, while the low-frequency SNPs were retained in 1000 cases, ExAc and esp6500 databases, and mutations in the COSMIC database; 3. Filter out SNPs in intergenic, non-coding, and intron regions; 4. Filter out synonymous mutations and mutations in repetitive genomic regions; 5. According to the scores of SNPs (ljb23_sift, ljb23_pp2hvar, ljb23_pp2hdiv, and ljb23_mt databases), ineligible SNPs were filtered, when they assessed as harmful mutations in one database or as moderately harmful mutations in two or more databases.

Significantly mutated genes (SMGs) and identification of driver genes

MuSiC was used to identify SMGs by taking somatic mutations as input and comparing them against background mutation rate (Dees et al. 2012). The convolution test (CT), the Fisher's combined P-value test (FCPT), and the likelihood-ratio test (LRT) were performed on mutant genes, and the gene with Q-value < 0.2 was retained. Identification of driver genes was carried out via SMGs, Mutant Cluster Preference (CLUST bias), and functional impact (FI) score. Oncodrive CLUST calculated the preferred sites of mutation clusters, and constructed the mutational background based on the unbiased distribution of synonymous mutations to identify possible driver genes (Tamborero et al. 2013).

Results

Patients' clinical characteristics

All patients were Chinese, dominantly male (male: female =3: 2), with a median age of 65.6 (range, 61-70) years old. The tumor and adjacent non-tumor tissues were isolated from ESGDAC patients who underwent surgical resection at the Department of Thoracic Surgery, The First Affiliated Hospital of Zhengzhou University from January 2018 to October 2020. All the lesions (range, 1.5-3.5 cm) were located at the middle esophagus. The endoscopic appearance showed a submucosal tumor-like lesion. Gastric mucosal erosion and hyperemia were detected in all five cases, and three cases showed a significantly narrowed esophageal. Thoracoscopic radical resection of esophageal carcinoma was adopted in all cases. All patients, except for one case who died because of severe pulmonary infection at 8 days after surgery, were followed up and remained well without recurrence or rapid aggravation of esophageal diseases (Table 2).

Pathological features

ESGDAC was located mainly in muscularis mucosa, and infiltrated into submucosa and **muscularis propria** without venous or lymphatic invasion, while perineural invasion could be detected in all cases. A number of neoplastic tubules exhibited a bi-layered structure, with inner luminal and outer epithelial layers; others showed small solid, adenoid and irregular cell nests with poorly differentiated adenocarcinoma. The outer epithelial layers had a squamous appearance, whereas the inner luminal layers showed a columnar epithelial cell differentiation. Histologically, the tumor was composed of moderately-to-poorly differentiated adenocarcinoma with marked cellular atypia, whereas the adjacent stratified squamous showed no intraepithelial neoplasia in all cases. Scattered nests of keratinization (1 case) and goblet cell-like mucus cells in the inner layer of the tubules (4 cases) could be observed (Figure 1).

The results of immunohistochemical analysis were available for all cases with ESGDAC. It was shown that the luminal epithelial cells lining the inside of the ducts were positive for CK7 and CAM5.2, and basal cells were positive for p63 and p40. The two layers of cells were negative for S-100, SOX-10, c-Myb, and CD117, which were always positive for salivary gland-type adenocarcinoma. In all the cases, we detected overexpressed p53 protein in both the inner layer of luminal and outer layer of basal epithelial cells, indicating the occurrence of mutation of p53 gene in all the cases. Furthermore, overexpression of HER-2 (c-erbB2) was identified by immunohistochemical assay and no amplification of HER-2 was noted. The proportion of Ki-67-positive cells was high (range, 10-40%). The results of immunohistochemical assay are displayed in Figure 2.

Molecular features

WES was performed to profile the somatic mutational landscape of ESGDAC, comprising primary tumors and adjacent non-tumor tissues from 3 ESGDAC patients, because of degradation of DNA of the other two patients. To the best of our knowledge, this is the first study to perform WES on ESGDAC cases. Table 3 presents the summary of WES coverage statistics for each patient.

Somatic mutation is the biological mechanism underlying the regulation of gene expression, and it plays a significant role in the study of cancer genomics. Somatic mutation can be categorized into three groups by size: somatic SNVs, somatic InDels, and somatic CNVs (Xu 2018). Somatic SNVs were detected by muTect (Cibulskis et al. 2013). MuTect run in default setting by taking tumor and normal sample data as input. A total of 796 somatic SNVs were identified, and 705 were nonsynonymous SNVs. Patient 3 had

the highest number of somatic SNVs, which was almost twice higher than that of Patient 1 (Figure 3A). Somatic InDels were analyzed by Strelka (Saunders et al. 2012), and 37 somatic InDels were detected. Patient 2 had the highest number of somatic InDels, which was six times higher than that of Patient 1 (Figure 3B). It has been demonstrated that somatic CNVs aroused as a result of preferential selection, favoring cancer development, and the study of chromosomal deletion and amplification has markedly attracted scholars' attention (Liang et al. 2016). Control-FREEC was used to detect somatic CNVs (Boeva et al. 2012), and GISTIC was employed to infer recurrently amplified or deleted genomic regions. Patient 2 had the highest number of somatic CNVs. In contrast to the somatic SNVs and somatic InDels, the number of somatic CNVs did not significantly change among the three samples (Figure 3C).

In addition to somatic mutations, CPGs, in which deleterious germline variants are associated with the increased risk of cancer, play an important role in the pathogenesis of cancer. Detection of CPGs may enable genetic counseling and cancer surveillance for early detection, and it may lead to revision of cancer treatment protocols (Park et al. 2018). We identified 30 ESGDAC-associated CPGs via analyzing germline variants in ESGDAC cases. The genes, such as ERBB2, BRD3, AKAP9, MKL1, MYH11, COL1A1, ZFHX3, and KMT2D, were present in 2 of 3 tumors, and the remaining of the CPGs were present in 1 of 3 tumors (Figure 4).

Furthermore, we determined the type of mutational signatures for ESGDAC patients. The results revealed mutational signature 1 as the major signature, influencing ESGDAC patients. The mutational signature 1 was the result of an endogenous mutational process and it was associated with small numbers of small insertions and deletions in the majority of tissue. This signature has been found in all cancer types and in the majority of cancer samples. In addition, mutational signature 5, which exhibited transcriptional strand bias for T>C substitutions at ApTpN context and has also been found in all types of cancer and the majority of cancer samples, might be another important mutational signature, affecting ESGDAC patients (Figure 5). The above-mentioned results suggested that different biological mechanisms might be involved in mutagenesis of ESGDAC.

Finally, we compared somatic variations with the known driver genes and detected candidate driver genes from the tumor samples. A total of 12 ESGDAC-associated driver genes (TP53, ACSL6, PABPC4, AFF1, SPEN, etc.) were identified. TP53 mutation was present in all the three tumors, representing the most frequent genomic alteration in this study (Figure 6A). In addition, MuSiC was used to identify SMGs by taking somatic mutations as input and comparing them with background mutation rate (Dees et al. 2012). The CT, FCPT, and LRT were performed on mutant genes, and the gene with Q-value < 0.2 was retained. The results revealed that TP53 mutation was identified in all the three tumors, and it could be a SMG in ESGDAC (Figure 6B).

Discussion

The main epithelial malignant tumors of the esophagus are ESCC and EAC. EAC mainly arise from the Barrett's esophagus, and it is scarcely associated with heterotopic gastric mucosa of the esophagus and salivary gland-type adenocarcinoma (Mukaisho et al. 2019). EAC originated from ESGDAC is extremely rare and only few cases have been reported worldwide. At present, the diagnosis of this rare type of EAC is mainly based on patients' histopathological and immunohistochemical characteristics, while pathologists are mainly confused because of morphological similarity with other esophageal carcinomas, such as salivary

gland-type adenocarcinoma and esophageal basaloid squamous cell carcinoma (Nie et al. 2020, Tamura et al. 2017). To date, molecular features of ESGDAC have not been comprehensively analyzed.

WES routinely reveals gene mutations in cancer genomics and it also facilitates identification of new potential genetic biomarkers and targeted therapies for diagnosing diverse types of cancer (Chandler et al. 2016). We, for the first time, used WES to investigate the complete genetic landscape of ESGDAC, facilitating the development of targeted therapies. Owing to the significant role in cancer genomics, somatic mutations, including SNVs, InDels, and CNVs, were identified in patients with ESGDAC. The mutational signatures obtained for the three cases from WES could be valuable for selecting the targeted drugs and standardizing the chemotherapy regimen for ESGDAC. Several plausible ESGDAC-associated driver genes had been identified and TP53 mutation was found as SMG in ESGDAC, which was consistent with the results of immunohistochemical assay.

A number of limitations in this study should be pointed out. First, only three cases were analyzed by WES, because of the extremely rarity of the disease and DNA degradation of the other two patients. Further research with a larger sample size is required to validate our findings and to explore the complete molecular mechanism, underlying the progression to ESGDAC. This may require a statistical analysis of the relationship between the molecular features identified in the current study and therapeutic outcomes.

Conclusions

In the present research, we summarized the clinical, pathological, and immunohistochemical characteristics of patients with ESGDAC. WES was performed on 3 ESGDAC genomes to identify and analyze somatically acquired mutations, signatures of somatic mutations, CPGs, ESGDAC-associated driver genes, and SMGs. We found TP53 as the SMG in ESGDAC. We described the relevant tumor mutation profiling for development of future targeted therapies for ESGDAC.

Declarations

Declarations

Acknowledgements

We thank Dr. Guozhong Jiang for generously providing technical assistance.

Funding

This study was supported by the National Natural Science Foundation of China (No.81900791), Henan Medical Science and Technology Research Plan Joint Co-construction Project (No. 2018020080), Henan Science and Technology Project in 2019 (No.192102310322).

Conflicts of Interest

The authors declare no conflict of interest.

Availability of data and material

The data generated or analyzed in this study are included in the article.

Authors' contributions

Authors' roles: Study design: ZYW and YHM. Study conduct: ZZ and BQ. Data collection: BBW, KHL and XH. Data analysis: ZYW BC and XHY. Drafting manuscript: ZYW and ZZ. All authors Approved the final version of the manuscript.

Ethics approval

The study was approved by the Research and Clinical Trial Ethics Committee of The First Affiliated Hospital of Zhengzhou University (Approval No. 2021-KS-HNSR097).

References

1. Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, Janoueix-Lerosey I, Delattre O, Barillot E. 2012. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* 28: 423-425.
2. Chandler MR, Bilgili EP, Merner ND. 2016. A Review of Whole-Exome Sequencing Efforts Toward Hereditary Breast Cancer Susceptibility Gene Discovery. *Hum Mutat* 37: 835-846.
3. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G. 2013. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 31: 213-219.
4. Dees ND, et al. 2012. MuSiC: identifying mutational significance in cancer genomes. *Genome Res* 22: 1589-1598.
5. Kono M, Nagami Y, Ohsawa M, Fujiwara Y. 2019. Esophageal Adenocarcinoma Originating from the Esophageal Gland Duct. *Intern Med* 58: 883-884.
6. Lam AK. 2018. Histopathological Assessment for Esophageal Adenocarcinoma. *Methods Mol Biol* 1756: 67-76.
7. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078-2079.
8. Liang L, Fang JY, Xu J. 2016. Gastric cancer and gene copy number variation: emerging cancer drivers for targeted therapy. *Oncogene* 35: 1475-1482.
9. Mukaisho KI, Kanai S, Kushima R, Nakayama T, Hattori T, Sugihara H. 2019. Barretts's carcinogenesis. *Pathol Int* 69: 319-330.
10. Nie L, Li W, Xue L, Wang L, Shen Y, Fan X. 2020. Submucosal gland neoplasms of the esophagus: an update and review. *Esophagus* 17: 376-384.
11. Nie L, Wu HY, Shen YH, Fan XS, Sun Q, Huang Q, Chen J. 2016. Esophageal submucosal gland duct adenoma: a clinicopathological and immunohistochemical study with a review of the literature. *Dis*

Esophagus 29: 1048-1053.

12. Park S, Supek F, Lehner B. 2018. Systematic discovery of germline cancer predisposition genes through the identification of somatic second hits. *Nat Commun* 9: 2601.
13. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. 2012. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* 28: 1811-1817.
14. Tamborero D, et al. 2013. Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Sci Rep* 3: 2650.
15. Tamura H, Saiki H, Amano T, Yamamoto M, Hayashi S, Ando H, Doi R, Nishida T, Yamamoto K, Adachi S. 2017. Esophageal carcinoma originating in the surface epithelium with immunohistochemically proven esophageal gland duct differentiation: A case report. *World J Gastroenterol* 23: 3928-3933.
16. Wang K, Li M, Hakonarson H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38: e164.
17. Wang L, Shao Z, Chen S, Shi L, Li Z. 2017. A SLC24A2 Gene Variant Uncovered in Pancreatic Ductal Adenocarcinoma by Whole Exome Sequencing. *Tohoku J Exp Med* 241: 287-295.
18. Xu C. 2018. A review of somatic single nucleotide variant calling algorithms for next-generation sequencing data. *Comput Struct Biotechnol J* 16: 15-24.

Tables

Table1 Antibodies used for immunohistochemistry

Marker	Supplier	Dilution
CK7	Abcam, ab68459	1 : 200
CAM5.2	Gene Tech (Shanghai) Co., Ltd., 7GT209107	1 : 200
P40	Abcam, ab76158	1 : 200
P63	Abcam, ab53039	1 : 200
P53	Abcam, ab131442	1 : 200
S-100	Abcam, ab54642	1 : 200
SOX10	Santa cruz, sc-365692	1 : 200
c-Myb	Sigma-Aldrich, SAB4501936	1 : 200
CD117	Gene Tech (Shanghai) Co., Ltd., POLY	1 : 200
HER-2	Roche, 790-4493	1 : 200
Ki-67	Gene Tech (Shanghai) Co., Ltd., GM001	1 : 200

Table 2 Clinicopathological characteristics of the 5 included cases

Characteristics	Number(total)	Ratio(%)
Gender		
Male	3(5)	60
Female	2(5)	40
Age(year)		
60-65	3(5)	60
66-70	2(5)	40
Distance from incisors(cm)		
26-30	4(5)	80
31-35	1(5)	20
Maximal diameter(cm)		
≤3	4(5)	80
≥3	1(5)	20
Squamous epithelium lesions		
Yes	4(5)	80
No	1(5)	20
Symptoms		
Dysphagia	4(5)	80
Retrosternal pain	1(5)	20
Follow-up		
Several	4(5)	80
Death	1(1)	20

Table 3 Whole exome sequencing data summary for 3 patients

Sample	Patient 1(NT)	Patient 1(T)	Patient 2(NT)	Patient 2(T)	Patient 3(NT)	Patient 3(T)
Total	146,374,940 (100%)	200,187,688 (100%)	208,046,347 (100%)	178,431,514 (100%)	118,784,023 (100%)	163,468,323 (100%)
Mapped	145,743,777 (99.57%)	200,052,022 (99.93%)	207,275,542 (99.63%)	178,076,469 (99.80%)	118,535,225 (99.79%)	163,347,889 (99.93%)
Properly mapped	134,157,663 (91.65%)	197,013,762 (98.41%)	191,090,975 (91.85%)	173,593,529 (97.29%)	111,713,642 (94.05%)	161,192,169 (98.61%)
Initial bases on target	60,456,963	60,456,963	60,456,963	60,456,963	60,456,963	60,456,963
Total effective yield(Mb)	22,274.72	30,014.27	32,027.67	26,940.05	18,236.33	24,544.15
Effective sequences on target(Mb)	13,271.68	21,012.36	19,978.11	18,742.53	11,995.49	17,348.42
Fraction of effective bases on target	60%	70%	62%	70%	66%	71%
Average sequencing depth on target	220	348	330	310	198	287
Mismatch rate in target region	1%	0%	1%	1%	1%	0%
Base covered on target	60,186,958	60,310,649	60,129,829	60,179,376	60,197,554	60,289,161
Coverage of target region	100%	100%	99%	100%	100%	100%
Fraction of target covered with at least 10x	98%	99%	99%	99%	97%	99%
Fraction of target covered with at least 50x	90%	95%	93%	94%	86%	93%
Fraction of	75%	88%	85%	85%	70%	84%

target
covered
with at
least 100x

Figures

Figure 1

Patients' pathological features A. ESGDAC was exhibited with a bi-layered or a multi-layered structure, and the adjacent stratified squamous showed no intraepithelial neoplasia. B. Variably sized glands grew invasively and were diffusely dispersed in the muscular layer. C. Some solid scattered nests of keratinization without a luminal structure could be partly observed in the lesion. D. Goblet cell-like mucus cells could be seen in the inner layer of the tubules.

Figure 2

The results of ICH. The epithelial cells in the inner layer of the neoplastic tubules were positive for CK7, while were negative for p63. The cells in the outer layer were immunopositive for p63, whereas were negative for CK7. The cells in both inner and outer layers in the invasive component were negative for S-100, SOX10, c-Myb, and overexpressed p53 (magnification, 100 ×).

A

Sample	Patient 1	Patient 2	Patient 3
CDS	66	89	145
synonymous_SNP	21	22	48
missense_SNP	43	60	93
stopgain	2	5	2
stoploss	0	0	0
unknown	0	2	2
intronic	68	132	133
UTR3	4	10	6
UTR5	1	8	3
splicing	1	5	3
ncRNA_exonic	4	5	11
ncRNA_intronic	11	10	12
ncRNA_UTR3	0	0	0
ncRNA_UTR5	0	0	0
ncRNA_splicing	0	0	0
upstream	2	6	5
downstream	0	1	0
intergenic	18	16	19
Others	0	0	2
Total	175	282	339

B

Sample	Patient 1	Patient 2	Patient 3
CDS	1	5	3
frameshift_deletion	0	3	1
frameshift_insertion	0	0	0
nonframeshift_deletion	0	2	2
nonframeshift_insertion	1	0	0
stopgain	0	0	0
stoploss	0	0	0
unknown	0	0	0
intronic	2	10	10
UTR3	0	0	1
UTR5	0	1	0
splicing	0	1	1
ncRNA_exonic	0	0	0
ncRNA_intronic	0	0	0
ncRNA_UTR3	0	0	0
ncRNA_UTR5	0	0	0
ncRNA_splicing	0	0	0
upstream	0	1	0
downstream	0	0	0
intergenic	0	0	1
Others	0	0	0
Total	3	18	16

C

Sample	Gain count	Gain size	Loss count	Loss size	Total count	Total size
Patient 1	127	79424630	23	14705283	150	94129913
Patient 2	166	198595751	4	640004	170	199235755
Patient 3	162	244344438	0	0	162	244344438

Figure 3

Statistical analysis of somatic mutations. A. Somatic single-nucleotide variants (SNVs). B. Somatic insertions and deletions (InDels). C. Somatic copy number variations (CNVs).

Figure 4

Cancer predisposition genes (CPGs). The heat map illustrates 30 ESGDAC-associated CPGs in 3 ESGDAC samples. Mutation type included missense mutation (dark blue block), splice site (green block), nonsense mutation (purple block), and in frame del (light blue block).

A

B

#Gene	Indels	SNVs	Tot Muts	Sample Affect	Sample Percent(%)	P-value CT	FDR CT
TP53	0	4	4	3	100	1.21E-11	2.31E-07

Figure 6

Driver genes and significantly mutated genes. A. Driver genes represent the distribution between samples in ESGDAC. The column on top shows the mutational rate of every sample. Heat map displays the driver genes and mutation type, including missense mutation (blue block), splice site (green block), and frame shift del (yellow block). B. TP53 mutation was identified as the only significantly mutated gene (SMG).