

# Analysis of esophageal cancer-related mutations from cfDNA sequenced by Single-strand Adaptor Library Preparation sequencing

Yu Guan<sup>a</sup>, Qiong Li<sup>b</sup>, Shicai Liu<sup>c</sup>, Hongde Liu<sup>d</sup>, Jinke Wang<sup>\*</sup>

State Key Laboratory of Digital Medical Engineering, School of Biological Science and Medical Engineering, Southeast University, Nanjing 210096, China

**Abstract: Objectives:** More convenient and effective non-invasive diagnostic methods are essential for the detection and prognosis of cancer. This study aimed to mine the information in plasma cfDNA to find novel biomarkers for the diagnosis of esophageal cancer (ESCA). **Methods:** Blood samples were collected from esophageal cancer patients and healthy individuals. SALP-seq method was used to construct libraries and sequence cfDNA samples from 40 esophageal cancer patients and 10 normal cfDNA samples, and mutation analysis was performed. **Results:** Esophageal cancer related mutational signatures and 52 mutated genes were identified. Many of these genes are known cancer-related genes. Mutations in these genes were also found in 11 additional ESCA cfDNA samples. **Conclusion:** SALP-seq based cfDNA mutation analysis can obtain reliable and verifiable biomarkers for ESCA. These biomarkers provide a novel reference for the diagnosis of esophageal cancer, as well as offer novel insights into understanding the cellular and molecular mechanisms of esophageal carcinogenesis. Finally, our method provides a new avenue to explore novel cancer biomarkers.

## 1. Introduction

Cancer is caused by many factors such as genetics, environment, and lifestyle. Due to its high incidence and mortality rates, cancer has become a serious global public health issue, posing a significant threat to people's lives and health, and severely impacting their quality of life. In both China and developed countries, cancer has become a primary cause of mortality. Data provided by GLOBOCAN shows that in 2020, there were 19,292,789 new cancer cases and 9,958,133 cancer-related deaths globally [1]. According to the National Cancer Center of China, there were approximately 4,824,700 cancer cases and 2,574,200 cancer deaths in China in 2022 [2]. As per the latest cancer statistics from the American Cancer Society, it is projected that by 2024, there will be approximately 2,001,140 new cancer cases and 611,720 new cancer -caused deaths in the United States [3].

The improvement of survival rate of cancer patients is not only due to the continuous development of cancer treatment, but also the advancement of cancer diagnosis technology. Compared with those diagnosed later, people diagnosed with cancer at an early stage not only have a higher survival rate, but also a lower incidence of treatment and a higher quality of life. Therefore, the

development of more sophisticated and accurate cancer diagnostic technologies remains an important task.

In recent years, liquid biopsy has gradually become a hot spot in cancer research field. Liquid biopsy is a technique for analyzing tumor material in blood that is non-invasive and has considerable potential in the detection of pre-cancerous or early-stage cancers. As a kind of tumor substances which can be detected by liquid biopsy, cfDNA consists of extracellular nucleic acid fragments released by cells during apoptosis or necrosis. Research indicates that most cfDNA originates from the hematopoietic system of healthy people, but it is also released by relevant tissues or tumors in conditions such as pregnancy or cancer [4]. This suggests that it is possible to identify cancers by analyzing cfDNA. By analyzing plasma cfDNA, researchers can observe genetic and epigenetic changes in cells, including mutations, chromosomal rearrangements, and methylation changes. Numerous studies have abundantly demonstrated that methods based on cfDNA analysis possess high clinical potential and application value. Non-invasive prenatal testing (NIPT) can detect a wide range of dominant single-gene disorders through cfDNA analysis, providing more possibilities for screening for aneuploidy or carriers of recessive diseases. In cancer research, researchers have

<sup>a</sup> gyspace@163.com      <sup>b</sup> linda\_summerq@163.com

<sup>c</sup> liushicainj@163.com    <sup>d</sup> liuhongde@seu.edu.cn

<sup>\*</sup> Corresponding author: wangjinke@seu.edu.cn

discovered new cancer biomarkers by analyzing information such as cfDNA methylation, mutations, chromatin accessibility, providing strong support for the development of clinical cancer diagnostic methods. For example, the whole-genome sequencing (WGS) based on cfDNA establishes a comprehensive genomic mutation integration method, which can overcome the limitations of cfDNA abundance, achieve ultra-sensitive detection in low disease burden cancers, and optimize treatment [5]. Therefore, cfDNA holds important research value and application potential in exploring cancer biomarkers and cancer diagnosis.

As a common and highly aggressive tumor, esophageal cancer (ESCA) had become one of the main culprits of global cancer deaths, with over 400,000 deaths annually globally [6]. Ranking sixth in overall mortality rates among cancers, esophageal cancer is characterized by a lack of specific symptoms, leading to a poor five-year survival rate for about 40% of patients diagnosed at a terminal stage [7]. Additionally, the incidence of esophageal cancer is higher in East Asia compared to other regions [7]. Statistics show that over 50% of global esophageal cancer cases originate from China, with particularly severe conditions among vulnerable populations [8]. Endoscopic examination is used for early detection of ESCA in high-risk populations in high prevalence regions. However, due to the inherent invasiveness, inconvenience, and time-consuming nature of this method, it is challenging to implement in large-scale screenings. Furthermore, given the high recurrence rate, early metastasis tendency of esophageal cancer, limited understanding of its related biomarkers and therapeutic targets, identifying new biomarkers for ESCA diagnosis, especially those suitable for early-stage esophageal cancer diagnosis, had become an urgent requirement. Hence, it is necessary to explore a convenient, rapid non-intrusive or minimally invasive technique for esophageal cancer diagnosis.

Previously, we had developed a library construction method based on single-stranded DNA called Single-strand Adaptor Library Preparation (SALP) [9]. This method utilizes specific barcodes and adapters to construct highly efficient, high-throughput, and low-bias NGS libraries, particularly suitable for highly degraded DNA like cfDNA. In this study, we utilized cfDNA extracted from 40 esophageal cancer patients and 10 normal individuals' plasma, and constructed corresponding NGS libraries based on the SALP-seq. Through bioinformatic analysis of the sequencing data, 52 mutated genes were identified unique to esophageal cancer samples. These results may have significant implications for developing more effective methods for ESCA diagnosis.

## 2. Experimental procedure

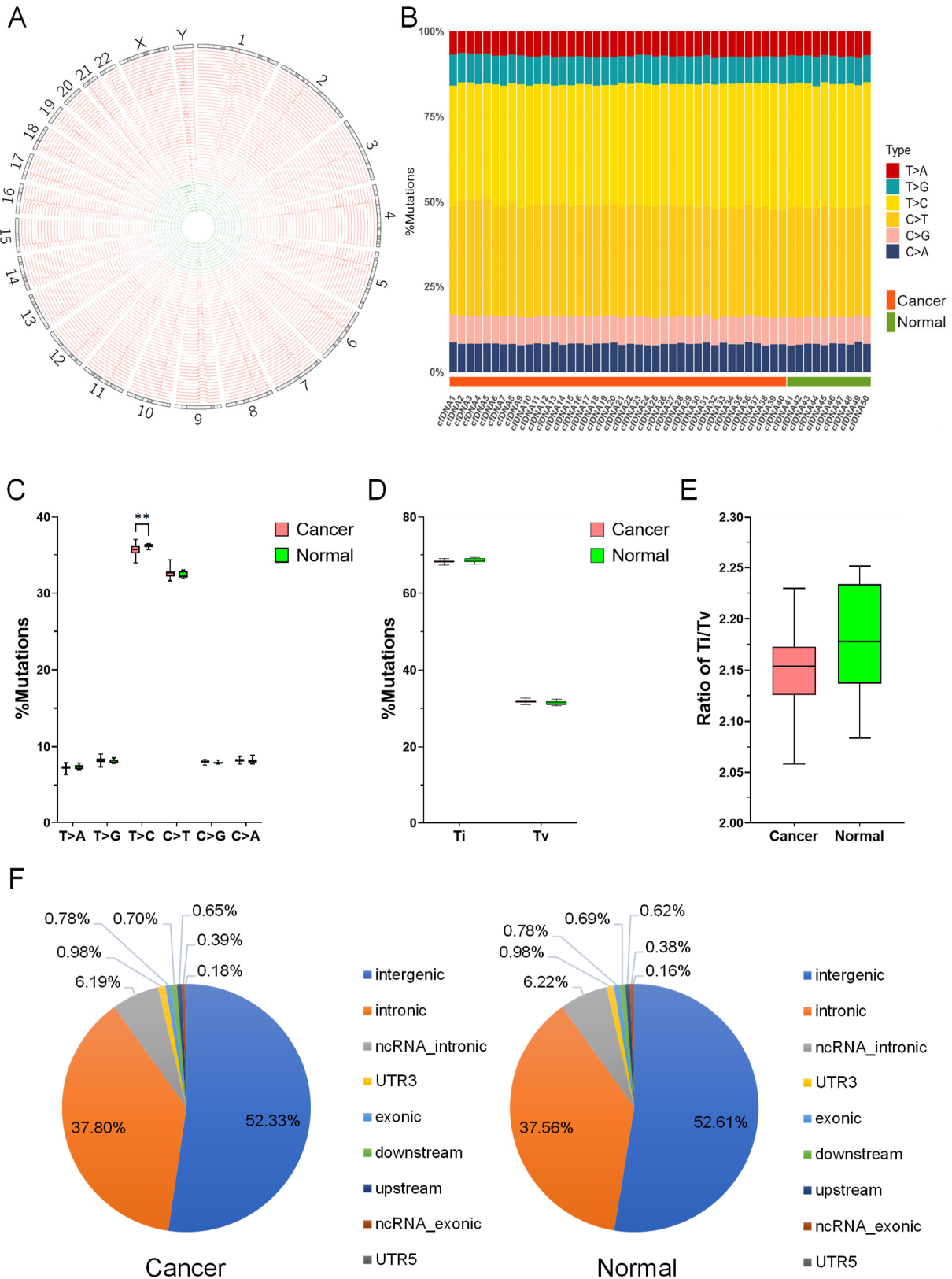
The blood samples were processed following the library construction protocol of SALP-seq [9]. The constructed libraries were subjected to paired-end sequencing on the Illumina HiSeq X10 platform (Nanjing Geneseeq) using two lanes. Subsequently, custom scripts were used to demultiplex the raw sequencing data based on the indexes,

and remove the 19 bp constant sequence and 6 bp barcodes in the 5' end of read2. Bowtie2 was employed to align all the paired-end reads to the human reference genome hg19 [9]. SNV analysis was performed using bcftools [10]. ANNOVAR was utilized for SNV annotation [11]. Functional enrichment analysis was conducted using the DAVID tool [12]. Through the Kaplan-Meier Plotter tool, the data from the TCGA database were used to conduct the Overall Survival (OS) analysis and Relapse Free Survival (RFS) analysis [13].

## 3. Results

In NIPT or liquid biopsy, mutation analysis based on targeted or whole-genome scale sequencing is widely applied. In this study, mutation analysis was conducted on the sequencing data of 50 cfDNA samples. As shown in Figure 1A, various mutations were observed at the whole-genome level in different samples. Next, we performed statistical analysis on the six types of base substitutions (T>A, T>C, T>G, C>A, C>G, and C>T) in all cfDNA samples. The results indicated differences in the distribution of these six base substitutions among individuals (Figure 1B). By comparing the distribution of the six base substitutions between cancer samples and normal samples, we found a lower frequency of T>C transition in cancer cfDNA samples compared to normal ones (Figure 1C). However, there were no significant differences in the frequencies of transitions (Ti) and transversions (Tv), as well as the Ti/Tv ratio between normal and esophageal cancer samples, although the Ti/Tv ratio in cancer samples appeared slightly lower than in normal samples (Figure 1D, E). These mutation features could potentially serve as references for ESCA liquid biopsy diagnostics. Subsequently, we used the ANNOVAR tool to annotate single nucleotide variations (SNVs) to accurately locate genomic positions. The results showed that most SNV sites distributed in intergenic regions and intronic regions (Figure 1F).

To identify all mutations involving coding sequences of genes, further analysis was conducted on each sample in this study. The results suggested that numerous genes mutated in cfDNA from both cancer and normal samples. Subsequently, we compared these mutated genes with the MSK-IMPACT panel genes to assess whether the cfDNA mutation analysis based on SALP-seq could detect clinically relevant mutations. MSK-IMPACT is the first academic or commercial tumor analysis test authorized by the U.S. Food and Drug Administration, used to identify clinically relevant somatic mutations, novel non-coding changes, and shared mutation features among common and rare tumor types [14]. As of now, the number of genes on the MSK-IMPACT panel has reached 505. Ultimately, this study identified 52 mutated genes unique to cancer samples (Table 1), indicating that these genes may play corresponding roles in ESCA. Among these 52 genes, many have been reported in studies and confirmed to be associated with cancer, such as INSR, FAT1, TP53BP1, EGFR and FLT1.



**Figure 1.** Mutation analysis of cfDNA. (A) Mutation density distribution in per 1Mb window at the whole-genome level for each sample, with sample numbers from 1 to 50 displayed from outer to inner circles. (B) Spectra of the six types of base mutations in each sample. (C) Display of SNV types in different categories of cfDNA samples (\*\* means p value < 0.01). (D & E) The distribution of transition (Ti) and transversion (Tv) for SNVs showed in box plots. (F) Distribution of SNV genomic locations in cancer and normal samples, with most SNVs distributed in intergenic regions and intronic regions.

**Table 1.** Gene symbol of 52 mutated genes

ABL1	ETAA1	LATS2	RASA1
ACVR1	FAT1	MAP3K1	REST
AGO1	FGFR2	MAP3K13	RET
APLNR	FLT1	MYCL	RRAS
ASXL1	FOXA1	NOTCH3	RUNX1
ATM	GAB2	PAX5	SETDB1
AXIN2	GRIN2A	PIK3C2G	SF3B1
BARD1	HLA-B	PIK3CA	SH2B3
BCL6	HLA-C	PIK3CG	SOX17
BRIP1	HNF1A	PLCG2	TAP1
CDH1	INSR	PREX2	TMPRSS2
CREBBP	IRS1	PTPRD	TP53BP1
EGFR	KEAP1	RAD54L	USP8

To verify the reliability of these mutated genes, we performed the same mutation analysis on the SALP-seq sequencing results of 11 additional cfDNA samples from ESCA patients. In each sample, over 80% reads mapped to hg19. Then, we successfully detected all 52 mutated genes in these 11 cfDNA samples (Figure 2B). The result validated our previous findings.

We conducted functional enrichment analysis on these 52 genes using the DAVID database. The GO analysis results showed significant associations with positive regulation of transcription from RNA polymerase II promoter, protein phosphorylation, regulation of transcription, transmembrane receptor protein tyrosine kinase signaling pathway, and regulation of apoptotic process (Figure 2C). KEGG pathway analysis revealed significant enrichment of these genes in Ras signaling pathway, Pathways in cancer, MicroRNAs in cancer, MAPK signaling pathway, among others (Figure 2D). GAD DISEASE CLASS analysis indicated significant enrichment of these genes in cancer, metabolic, developmental, and reproduction categories (Figure 2E). Furthermore, GAD DISEASE analysis showed associations of the genes with various cancers, including esophageal cancer (Figure 2E). UP KEYWORDS analysis showed significant enrichment of these genes in phosphoprotein, nucleotide-binding, and proto-oncogene categories (Figure 2F). The UP SEQ FEATURE analysis from DAVID indicated significant enrichment of these genes in disordered, polar residues, and activation loop categories (Figure 2F).

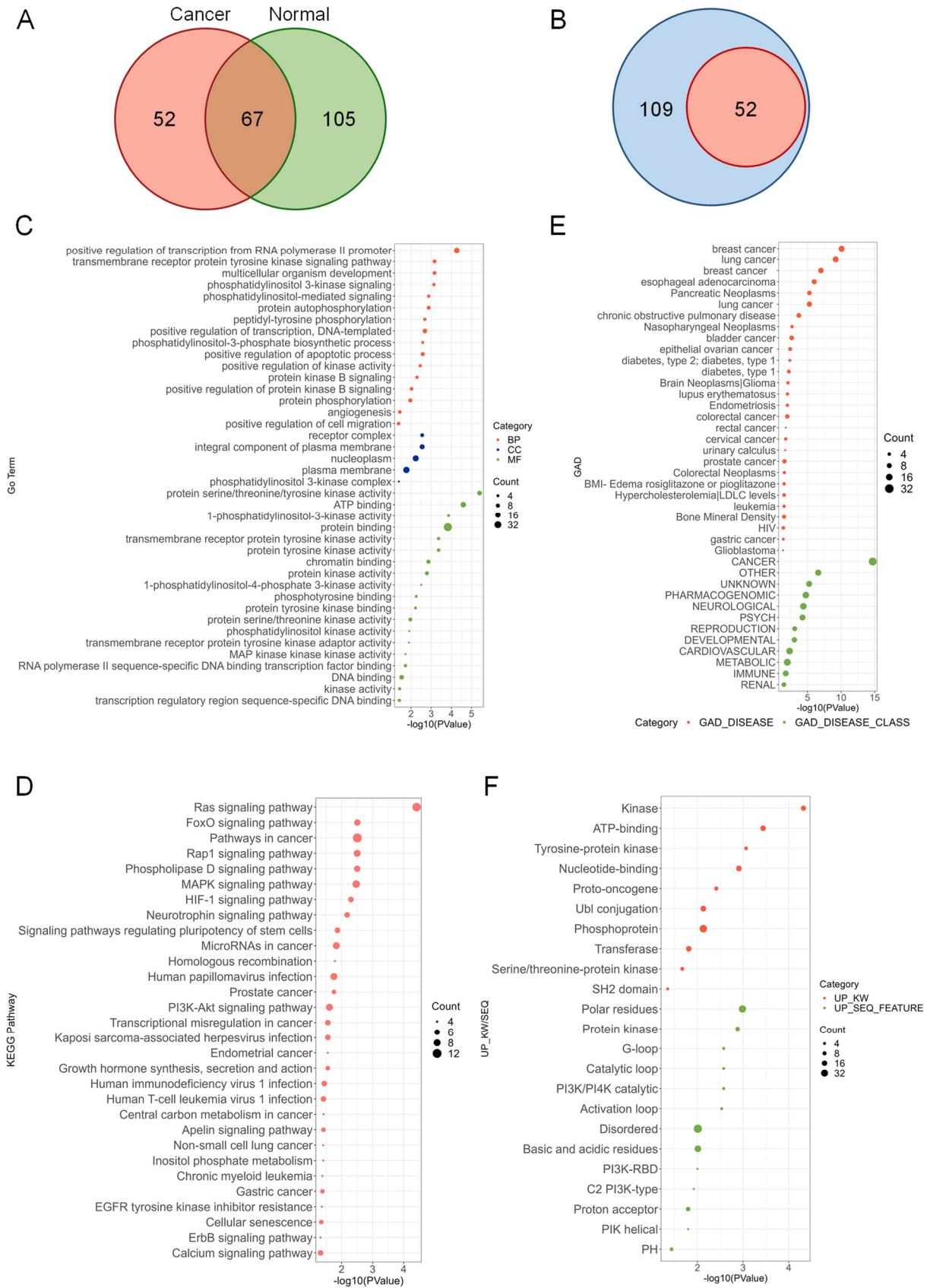
To explore the clinical association of these genes with ESCA, we used the Kaplan-Meier Plotter tool to conduct survival analysis based on data from TCGA database. The result suggested a strong correlation between the expression level s of the genes and survival status and prognosis of ESCA patients. For instance, in ESCA cases, FAT1 (OS, RFS) correlated with favorable prognosis (Figure 3A). EGFR (OS) had a correlation with good prognosis in ESCA patients (Figure 3B). In the OS analysis, the expression of INSR was associated with the prognosis of ESCA analysis (Figure 3C). HLA-B (OS) with

low expression level was correlated to the good prognosis of ESCA individuals (Figure 3D). These above results indicated that these genes had great potential and significance for clinical diagnosis and prognosis of ESCA.

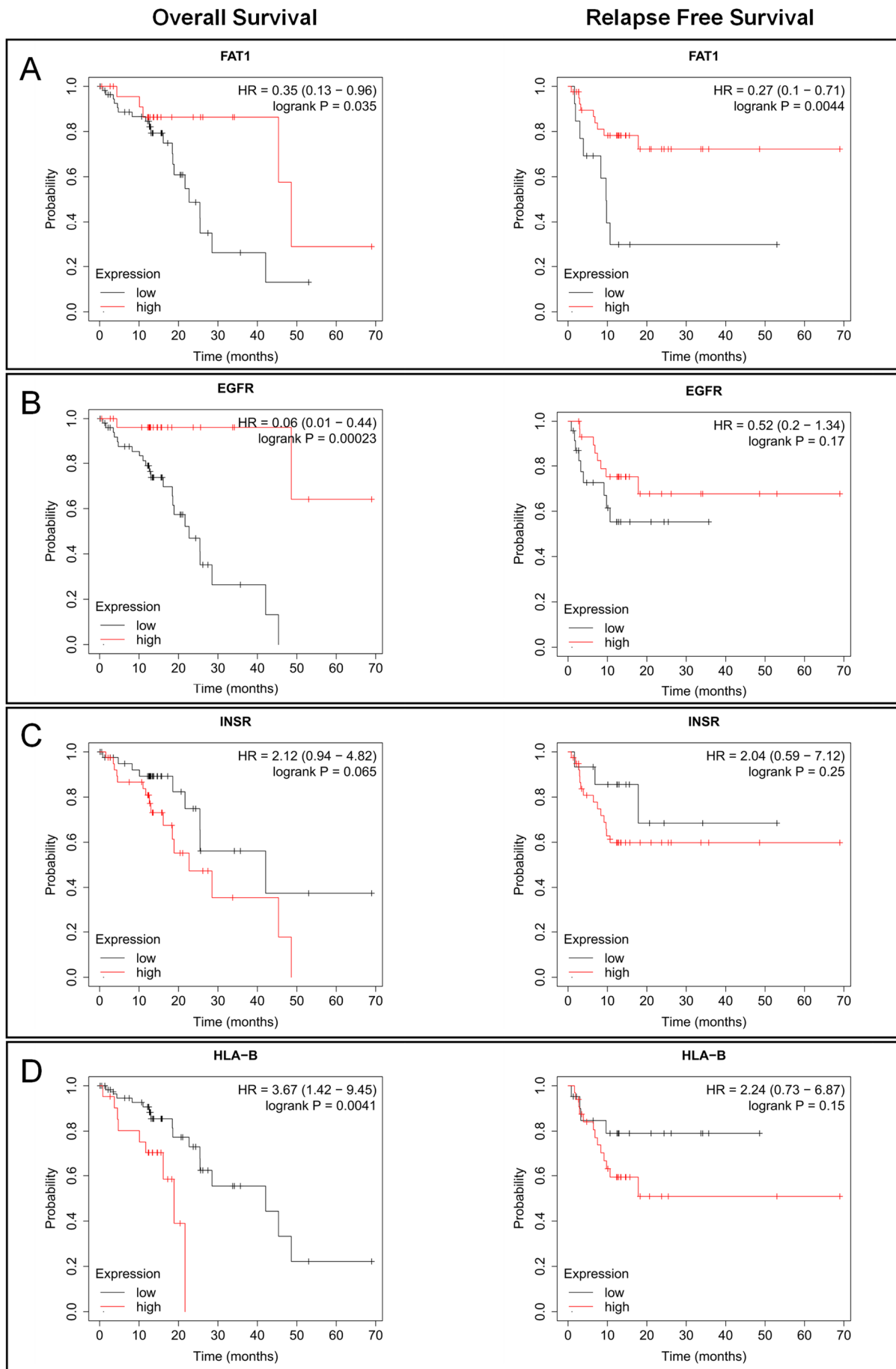
## 4. Discussion

This study utilized a laboratory-developed SALP-seq method based on NGS sequencing to explore the characteristics of mutations between esophageal cancer and normal cfDNA samples. SALP-seq is particularly well-suited for handling highly degraded DNA samples like cfDNA. This study sequenced 50 cfDNA samples, with over 80% of reads successfully mapped to the genome in each sample. Since mutations are common factors in cancer development, mutation analysis of sequencing data is a commonly used method. Through mutation analysis of the 50 cfDNA samples, we observed a significant proportion of T>C and C>T conversions in SNVs (Figure 2B). The study suggested that the percentage of T>C mutations in cfDNA samples from ESCA was significantly lower than in normal cfDNA samples (Figure 2C), which could potentially serve as a biomarker for liquid biopsy diagnosis in ESCA. Further analysis identified 52 unique mutated genes in esophageal cancer samples, many of which are well-known cancer-related mutated genes. After that, we successfully validated mutations of these genes in additional 11 ESCA cfDNA samples and analyzed the impact of the genes on prognosis of cancer patients, revealing the potential value on cancer clinical application of these biomarkers.

In human cancers, FAT1 is one of the most common mutated gene, and its loss of function promotes tumorigenesis, progression, metastasis, and stemness through inducing epithelial-to-mesenchymal transition (EMT) [15]. EGFR regulates epithelial tissue development and homeostasis. Targeted therapies against EGFR and VEGF have been applied clinically in esophageal cancer treatment [16]. INSR and related signaling pathways have been proved to be crucial in regulating cancer cell function. In patients with insulin resistance and compensatory hyperinsulinemia to overcome metabolic pathway resistance, overexpressed INSR excessively stimulates mitotic pathways and enhances cell proliferation, thereby increasing the risk of cancer [17]. The polymorphism of human leukocyte antigen (HLA) is associated with the risk of various cancers. Research on esophageal cancer patients from the Chaoshan region in southern China has shown an association between HLA-A and HLA-B related mutations and genetic susceptibility to esophageal cancer [18]. Functional enrichment analysis of these 52 genes also indicated a close connection with cancer development. Therefore, mutation analysis of cfDNA based on SALP-seq can detect mutated genes related to esophageal cancer, which can serve as potential biomarkers for esophageal cancer liquid biopsy applications in faster and simpler esophageal cancer screening methods.



**Figure 2.** Functional enrichment analysis of 52 unique mutated genes identified in cfDNA from esophageal cancer patients. (A) Exonic mutated MSK-IMPACT panel genes in different types of 50 cfDNA samples. (B) Exonic mutated MSK-IMPACT panel genes in additional 11 cfDNA samples. (C) GO analysis. (D) KEGG pathway analysis. (E) GAD DISEASE CLASS and GAD DISEASE analysis. (F) UP KEYWORDS (UP\_KW) and UP SEQ FEATURE analysis.



**Figure 3.** Overall Survival (OS) and Relapse Free Survival (RFS) analysis of genes and ESCA.

This study was conducted on a small sample basis, including 40 esophageal cancer cfDNA samples and 10 normal cfDNA samples. Therefore, future research will require a larger sample size of cfDNA samples to validate the current findings and further explore more universal and valuable biomarkers. Additionally, this study only identified mutated genes related to esophageal cancer as biomarkers, and whether these genes have esophageal cancer specificity will require further analysis of cfDNA samples from various types of cancers.

## 5. Conclusion

In this study, cfDNA from esophageal cancer patients and normal individuals were effectively analyzed through SALP-seq, and finally we identified mutation characteristics associated with esophageal cancer. These biomarkers hold potential for liquid biopsy diagnostic applications and provide significant new insights into the cellular and molecular mechanisms underlying esophageal cancer development. Mutation analysis of cfDNA based on SALP-seq can detect gene mutations present in cancer patients and can be corroborated with other literature studies, demonstrating the reliability and potential of this analytical approach, and offering significant novel insights into the clinical value of cfDNA.

## Acknowledgement

This investigation was financially funded by the project of the National Natural Science Foundation of China (62371126).

## References

- [1] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F. (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*, 71(3): 209–249.
- [2] Han, B., Zheng, R., Zeng, H., Wang, S., Sun, K., Chen, R., He, J. (2024) Cancer incidence and mortality in China, 2022. *J Nat Cancer Cent*, 10: 27.
- [3] Siegel, R. L., Giaquinto, A. N., Jemal, A. (2024) Cancer statistics, 2024. *CA: a cancer journal for clinicians*, 74(1): 12–49.
- [4] Wang, Y., Springer, S., Mulvey, C. L., Silliman, N., Schaefer, J., Sausen, M., Agrawal, N. (2015) Detection of somatic mutations and HPV in the saliva and plasma of patients with head and neck squamous cell carcinomas. *Sci Transl Med*, 7(293): 293ra104–293ra104.
- [5] Zviran, A., Schulman, R. C., Shah, M., Hill, S. T. K., Deochand, S., Khamnei, C. C., Maloney, D., Patel, K., Liao, W., Widman, A. J., Wong, P., Callahan, M. K., Ha, G., Reed, S., Rotem, D., Frederick, D., Sharova, T., Miao, B., Kim, T., Gydush, G., Landau, D. A. (2020) Genome-wide cell-free DNA mutational integration enables ultra-sensitive cancer monitoring. *Nat Med*, 26(7): 1114–1124.
- [6] Yuan, Z., Wang, X., Geng, X., Li, Y., Mu, J., Tan, F., Xue, Q., Gao, S., He, J. (2021) Liquid biopsy for esophageal cancer: Is detection of circulating cell-free DNA as a biomarker feasible? *Cancer Commun (Lond)*, 41(1): 3–15.
- [7] Siegel, R. L., Miller, K. D., Jemal, A. (2020) Cancer statistics, 2020. *CA Cancer J Clin*, 70(1): 7–30.
- [8] Zhu, H., Ma, X., Ye, T., Wang, H., Wang, Z., Liu, Q., Zhao, K. (2022) Esophageal cancer in China: Practice and research in the new era. *Int J Cancer*. 152(9): 1741–1751.
- [9] Wu, J., Dai, W., Wu, L., Wang, J. (2018) SALP, a new single-stranded DNA library preparation method especially useful for the high-throughput characterization of chromatin openness states. *BMC Genomics*, 19(1): 143.
- [10] Li H. (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics (Oxford, England)*, 27(21): 2987–2993.
- [11] Wang, K., Li, M., Hakonarson, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*, 38(16): e164.
- [12] Huang, daW., Sherman, B. T., Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 4(1): 44–57.
- [13] Nagy, Á., Munkácsy, G., Györfy, B. (2021) Pancancer survival analysis of cancer hallmark genes. *Scientific reports*, 11(1): 6047.
- [14] Zehir A, Benayed R, Shah RH, et al. (2017) Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med*, 23(6):703–713.
- [15] Pastushenko, I., Mauri, F., Song, Y., et al. (2021) Fat1 deletion promotes hybrid EMT state, tumour stemness and metastasis. *Nature*, 589(7842): 448–455.
- [16] Yang, Y. M., Hong, P., Xu, W. W., He, Q. Y., Li, B. (2020) Advances in targeted therapy for esophageal cancer. *Signal Transduct Target Ther*, 5(1): 229.
- [17] Vella, V., Milluzzo, A., Scalisi, N. M., Vigneri, P., Sciacca, L. (2018) Insulin Receptor Isoforms in Cancer. *Int J Mol Sci*, 19(11): 3615.
- [18] Hu, S. P., Zhou, G. B., Luan, J. A., Chen, Y. P., Xiao, D. W., Deng, Y. J., Huang, L. Q., Cai, K. L. (2010) Polymorphisms of HLA-A and HLA-B genes in genetic susceptibility to esophageal carcinoma in Chaoshan Han Chinese. *Dis Esophagus*, 23(1): 46–52.