Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Identification of a Novel Epithelial-mesenchymal Transition Gene Signature Regulated by *KEAP1-NRF2* Pathway in Esophageal Carcinoma

Mohamed Elshaer
Zhejiang University,School of medicine
Ahmed Hammad
Zhejiang University, School of medicine
Xiu Jun Wang
Zhejiang University,School of medicine
xiuwen Tang (xiuwentang@zju.edu.cn)
Department of Biochemistry and Department of Thoracic Surgery of the First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310003, PR China, https://orcid.org/0000-0002-6601-1234

Primary research

Keywords: TCGA, KEAP1, NRF2, epithelial-mesenchymal transition, esophageal cancer, biomarker

Posted Date: September 16th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-76086/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Background

KEAP1-NRF2 pathway alterations were identified in many cancers including, esophageal cancer (ESCA). Identifying biomarkers that are associated with mutations in this pathway will aid in defining this cancer subset; and hence in supporting precision and personalized medicine.

Methods

In this study, 182 tumor samples from the Cancer Genome Atlas (TCGA)-ESCA RNA-Seq V2 level 3 data were segregated into two groups *KEAP1-NRF2*-mutated (22) and wild-type (160).The two groups were subjected to differential gene expression analysis and we performed Gene Set Enrichment Analysis (GSEA). Then, the enriched gene set was integrated with the differentially expressed genes (DEGs) to identify a gene signature regulated by the *KEAP1-NRF2* pathway in ESCA. Furthermore, we validated the gene signature using mRNA expression data of ESCA cell lines provided by the Cancer Cell Line Encyclopedia (CCLE). The identified signature was tested in 3 independent ESCA datasets to assess its prognostic value.

Results

We identified 11 epithelial-mesenchymal transition (EMT) genes regulated by the *KEAP1-NRF2* pathway in ESCA patients. Five of the 11 genes showed significant over-expression in *KEAP1-NRF2*-mutated ESCA cell lines. In addition, over-expression of these five genes was significantly associated with poor survival in 3 independent ESCA datasets, including the TCGA-ESCA dataset.

Conclusion

Altogether, we identified a novel EMT 5-gene signature regulated by the *KEAP1-NRF2* axis and this signature is strongly associated with metastasis and drug resistance in ESCA. These 5-genes are potential biomarkers and therapeutic targets for ESCA patients in whom the *KEAP1-NRF2* pathway is altered.

Background

Esophageal cancer is the sixth most common cause of cancer death and the eighth in incidence worldwide. In fact, it accounts for 4% of cancer diagnoses and for 6% of cancer deaths. The prognosis for esophageal carcinoma is poor, with a 5-year survival rate of 19% and only 0.9% for advanced esophageal carcinoma (1).

To maintain oxidative homeostasis, cancer cells increase the transcription of antioxidant genes by acquiring either stabilizing mutations in *NFE2L2* (encoding *NRF2*, the master transcriptional regulator of the cellular antioxidant program) or by selecting for inactivating mutations in its negative regulator,

KEAP1 (2). *KEAP1* is a substrate receptor of the Cul3-RING ubiquitin ligase (CRL3) that, under physiological conditions, constitutively binds and targets *NRF2* for degradation. In response to oxidative stress, the *KEAP1-NRF2* binding is inhibited and, consequently, *NRF2* is stabilized (3).

The TCGA network has revolutionized the cancer research by enriching the cancer research community with a huge amount of cancer-related data. This revolution has enabled researchers to identify cancer driver genes, cancer dependency, prognostic biomarkers and therapeutic targets. In addition, it has enabled researchers to segregate one cancer type into subgroups in order to assist personalized and precision medicine field. Identification of genes and biological processes that are regulated by the *KEAP1-NRF2* pathway in different cancers may provide an effective approach for therapy of subset of cancers that harbor *KEAP1-NRF2* pathway alterations. Moreover, these gene signatures can be used to predict survival of patients.

In previous studies, we identified gene signatures that are regulated by the *KEAP1-NRF2* pathway in lung adenocarcinoma (LUAD) and head and neck squamous cancer (HNSC) (4–6). In the current study, we used the genomics, and transcriptomics data of the ESCA cohort from TCGA to identify a gene signature regulated by the *KEAP1-NRF2* pathway in ESCA.

Methods

Overall database selection

TCGA RNA-Seq gene expression version 2 level 3 data (Illumina HiSeq platform) for 182 ESCA tissues were downloaded from the Broad GDAC (Global Data Assembly Centers) Firehose website (http://gdac.broadinstitute.org/). All the mutation data used in the present study was obtained from CVCDAP (https://omics.bjcancer.org/cvcdap/home.do) and UCSC Xena Browser (https://xenabrowser.net/) (7, 8). Twelve percent (22 out of 185) of the TCGA-ESCA patients were found to harbor mutations in either *KEAP1*, *NRF2* or both. Then, we segregated these patients into two groups, one had 22 patients with mutations in either *KEAP1, NRF2* or both (mutated group) while the other had 160 patients with neither *KEAP1* nor *NRF2* mutations (wild-type group).

RNA-Seq data analysis

TCGA RNA-Seq gene expression version 2 level 3 data (Illumina Hiseq platform) for 182 ESCA tissues were subjected to differential gene expression analysis. Briefly, level 3 transcriptomic data was normalized by the FPKM (Fragments per Kilobase of transcript per Million mapped reads) method. All gene expression values were log-transformed to approximate the data to a normal distribution. In addition, genes with zero values in more than 25% of the patients were excluded. The differentially expressed genes (DEGs) were identified by applying the two-tailed t-test assuming unequal variance. Then, *P* values were adjusted using the FDR method. DEGs with FDR 0.05 were considered significant.

Gene Set Enrichment Analysis (GSEA)

GSEA (https://omics.bjcancer.org/) was performed to determine all significantly affected biological pathways. GSEA is a computational method that determines whether a defined set of genes shows statistically significant, concordant differences between two biological states. The primary result of the GSEA is the enrichment score (ES), which reflects the degree to which a gene set is overrepresented at the top or bottom of a ranked list of genes. The gene list metric was built using ratio between classes (biological states). A positive ES indicates gene set enrichment at the top of the ranked list (up-regulated); a negative ES indicates gene set enrichment at the bottom of the ranked list (down-regulated).

Cancer Cell Line Encyclopedia (CCLE) data analysis

RNA-seq gene expression data for 1019 cell lines were downloaded from CCLE https://portals.broadinstitute.org/ccle/data. We identified ESCA cell lines that harbor mutations in either *KEAP1, NRF2* or both using the COSMIC data base https://cancer.sanger.ac.uk/cell_lines. Then, we divided the ESCA cell lines into two groups mutated and wild-type. Differential gene expression analysis was performed as described above.

Identification of *NRF2* binding sites by *in silico* analysis

To identify the *NRF2* binding sites within the promoter regions of the putative *NRF2* regulated genes, we used the transcription factor-binding site finding tool ConTra V3 (9) with stringency core = 0.95 and similarity matrix = 0.85. The search was limited to the -1 kb upstream the transcription start site (TSS).

Survival analysis

For the identification of prognostic biomarkers, Kaplan-Meier curves were generated by using the web-based patients survival analysis tool SurvExpress (10). Log rank test $P<0.05$ was used as the cutoff for significance. The method of analysis has been discussed in a previous study (5).

Results

Overview of the mutational landscape of the *KEAP1-NRF2* pathway in cancer

First, we investigated the mutational landscape of the *KEAP1-NRF2* pathway in cancer by analyzing 11,079 TCGA samples from 33 different cancer types using the CVCDAP database. As shown in (Fig.2A), we found that the *KEAP1-NRF2* pathway was altered in many cancers, however it was altered with a percentage that was higher than 10% in only five cancer types. Lung Squamous Cell Cancer (LUSC) was the cancer type that harbor the highest percentage of *KEAP1-NRF2* pathway alterations with mutations in 24.2% (*KEAP1*, 9.92%; *NRF2*, 13.69%; both, 0.59%) of samples. LUAD came in the second place with *KEAP1-NRF2* pathway mutations in 20.82% (*KEAP1*, 17.68%; *NRF2,* 2.97%; both, 0.17%) of LUAD patients. In addition, the *KEAP1-NRF2* pathway was altered in 12.07% ESCA patients (*KEAP1*, 2.19%; *NRF2*, 9.34%; both, 0.54%), in 11.7% of uterine corpus endometrial carcinoma (UCEC) patients (*KEAP1*, 3.19%; *NRF2*, 7.04%; both, 1.48%) and in 10.1% of HNSC (*KEAP1*, 3.95%; *NRF2*, 5.49%; both, 0.659%).

Overview of the ESCA mutational landscape

The lack of prognostic biormarkers for the ESCA subgroup with *KEAP1-NRF2* mutations motivated us to focus on identifying prognostic biormakers that is regulated by *KEAP1-NRF2* in ESCA. First, we investigated the different driver mutations in the TCGA-ESCA patients and we found that *TP53* was the most frequently mutated gene in ESCA as it was found mutated in 86% of patients (Fig.2B). 42% of ESCA patients harbored *TTN* mutations which made *TTN* in the second place. *MUC16, SYNE1, CSMD3, FLAG, DNAH5, HMCN1, LRP1B, PCLO* and *RYR2* were mutated in 23%, 20%, 20%, 18%, 16%, 16%, 16%, 15% and 12% of ESCA patients, respectively. *KEAP1-NRF2* was found mutated in 12% of ESCA patients.

The TCGA-ESCA cohort included 182 patient samples. We segregated the cohort into two groups: *KEAP1-NRF2*-mutated (22 patients) and wild-type (160 patients). In order to ensure that the differences between the two groups were due to *KEAP1-NRF2* mutations, we first investigated the driver mutations in the *KEAP1-NRF2*-mutated group. We found that *TP53* was mutated in 95% of patients in this group, *NRF2* was found mutated in 85% patients, *TTN, KMTD2, MUC13, KEAP1, PATCH*, and *SACS* were found mutated in 41%, 32%, 27%, 23%, 23% and 23%, respectively (Fig.3A). Then, we performed differential gene mutation analysis between the two groups to investigate the percentages of mutations of these driver genes in the two groups. Only *NRF2* and *KEAP1* weren't mutated in wild-type group while the other driver genes were found mutated in both the *KEAP1-NRF2*-mutated and wild-type groups, with similar percentages (Fig.3B). Therefore, none of these driver genes can be considered as variables that contribute to differences between the two groups.

In order to better understand the mutational landscape of *KEAP1-NRF2* in ESCA, we used the USCS Xena browser to examine the types of mutations and their positions in the domain structure of *KEAP1* and *NRF2* proteins. As noted earlier, we found that 2.19% of TCGA-ESCA patient samples had *KEAP1* mutations while *NRF2* was mutated in 9.34% and both were mutated in 0.54%. All the detected *KEAP1* mutations were missense mutations while 77.8% (14/18) of *NRF2* mutation were missense mutations, 11.1% were intron (2/18), 5.6% (1/18) were in-frame-deletions and 5.6% (1/18) were in-frame-insertions. *KEAP1* consists of 605 amino-acids, and 3 main domains with two mutations were detected in the BTB (broad-complex, tramtrack, and bric-a-brac) domain, three in the IVR (intervening region), and one in the Kelch domain, which is essential for the binding of NRF2 (Fig.3C). In the case of *NRF2* structure, the majority of mutations (17) occurred in the crucial *KEAP1*-binding domain Neh2, and only one was found in the Neh1 domain.

Identification of genes regulated by the *KEAP1-NRF2* pathway in ESCA

In order to identify genes that are regulated by the *KEAP1-NRF2* pathway in ESCA, we subjected the TCGA-ESCA data set (22 *KEAP1-NRF2*-mutated versus 160 wild-type tumor samples) to differential gene expression analysis. We identified 896 DEGs with log FC >|1| ($p < 0.05$ with FDR adjustment) (Fig.4A). Of these DEGs, 403 were up-regulated and 493 were down-regulated (Additional file1: Table S1). Since the ultimate effect of changes in the *KEAP1-NRF2* pathway is increased activity of *NRF2*; and hence the over-expression of its target genes, it was not surprising that several *bonafideNRF2* target genes were among

the up-regulated genes including, *AKR1C1, AKR1C2, AKR1B10, GSTM2, UGT1A6, AKR1C3, G6PD, GCLC, GCLM, GSTM3, GPX2, ABCC1, OSGIN1, SRXN1*, and *TXNRD1*. The gene expression profiles of 22 *KEAP1-NRF2*-mutated and 160 wild-type ESCA patient samples were visualized on a heatmap produced by unsupervised hierarchical clustering, and major differences between the gene expression patterns enabled cluster analysis to discriminate between sample types. Significant differences or trends between *KEAP1-NRF2*-mutated and wild-type ESCA patient samples were detectable for DEGs with log FC> |1| (Fig.4B). *CES1, AKR1C1, ADH7, ALDH3A,* and *CYP4F11* were the top five up-regulated genes in *KEAP1-NRF2*-mutated ESCA patient samples (Fig.4C), while *PIGR, MUC13, TASPAN8, LGALS4*, and *OLFM4* were the top five down-regulated genes (Fig.4D).

Epithelial-mesenchymal transition is regulated by the *KEAP1-NRF2* pathway in ESCA

The expression signatures of the hallmark gene sets, each containing 50 specific gene sets, were derived by concentrating multiple gene sets from the Molecular Signatures Database to represent well-defined biological statuses or courses. GSEA was performed to determine whether the identified gene sets showed statistically notable differences between the *KEAP1-NRF2*-mutated and their wild-type counterparts groups (Additional file 2: Table S2). Interestingly, four gene sets were up-regulated in the *KEAP1-NRF2-mutated* ESCA, namely, estrogen response late, hypoxia, reactive oxygen species pathway and EMT, the 4 gene sets were greatly enriched, with FDR < 0.05 (Fig.5). The gene set with the lowest FDR, namely, EMT (FDR = 0.001), which contained 194 genes was selected for further analysis. In order to specifically identify EMT genes that is associated with the *KEAP1-NRF2* pathway in ESCA, we integrated the DEGs between *KEAP1-NRF2*-mutated and wild-type ESCA patient samples (log FC> |1.5|, FDR < 0.05) with the set of 194 EMT-enriched genes (Fig.6A) using Venny 2.1 web-based tool (http://bioinfogp.cnb.csic.es/tools/venny/index.html). Intriguingly, we found 11 common genes: *SPP1, PTHLH, WNT5A, COL11A1, COL7A1, GPC1, SNAI2, ADAM12, FBN2, PFN2*, and *IGFBP3* (Fig.6B).

Epithelial-mesenchymal transition genes were validated using ESCA cell lines

In order to validate these 11 EMT-related genes as potential *NRF2* targets, we used CCLE to download RNA-seq mRNA expression data of 1019 cell lines. Then, using the COSMIC data base we identified human ESCA cell lines that harbors *NRF2* and/or *KEAP1* mutations. We selected three cell lines (TE6, TE11 and KYSE180) as the *KEAP1-NRF2*-mutated group. In addition, we selected another three human ESCA cell lines that have neither *NRF2* nor *KEAP1* mutations (TE5, TE9 and KYSE150) as the wild-type group. Then, we carried out differential gene expression analysis between the two groups. As shown in *GSTM3, AKR1C1* and *TXNRD1* (well-known NRF2 targets) showed significant up-regulation in the mutated group compared to the wild-type counterpart, which ensures *KEAP1-NRF2* pathway alteration in the group (Fig.6C). Furthermore, we investigated the expression of these 11 EMT-related genes between the two groups. Interestingly, five of the 11 genes (*SPP1, WNT5A, PTHLH, PNF2*, and *GPC1)* were significantly up-regulated in the mutated ESCA cell lines (Fig.6D). This finding suggests that these five genes are potential *NRF2* targets. For further evidence, we investigated the presence of the putative and known antioxidant responsive elements (AREs), the NRF2 binding site, (Fig.6E) in the promoter region of

these five genes by using ConTra V3 web tool. We performed *insilico* analysis within the –1 kb upstream the transcription start site (TSS) of the 5 genes. Interestingly, we identified highly conserved *NRF2* binding sites (AREs) in the promoter regions of human *PTHLH* (positions: -71 and -916), *WNT5A* (position: -434), *SPP1* (positions: -545,-605 and -870), *PFN2* (positions:-737 and -864) and *GPC1* (position: -458) (Fig.6F).

Evaluation of prognostic power of EMT gene-signature regulated by the *KEAP1-NRF2* pathway in ESCA

In order to evaluate the prognostic power of these 5 genes (*SPP1, WNT5A, PTHLH, PFN2, and GPC1*) as an EMT-derived signature for *KEAP1-NRF2* pathway alterations in ESCA, we first analyzed overall survival in the TCGA-ESCA cohort using the SurvExpress database. A total of 184 patient samples were divided into high-risk (n = 127) and low-risk groups (n = 57) based on their expression patterns (Fig.7A). The separation of risk groups was optimized using the 'maximize risk group' option provided in the SurvExpress database. The survival probability estimates in the two risk groups were visualized as Kaplan-Meier plots. Strikingly, overall survival analysis revealed that the patients in the high-risk group had poorer survival (HR = 1.67 (CI: 1.01-2.78); Log-Rank p = 0.04443) than the low-risk group (Figure.7B). Moreover, the Rao Giddings (GSE11595) cohort (34 ESCA patient samples) showed that the expression of *SPP1, WNT5A, PTHLH, PFN2*, and *GPC1* in the high-risk group (n=17) was associated with poorer survival (HR = 6.84 (CI: 2.36 - 19.8); Log-Rank p = 3.072×e-5) than the low-risk group (n=17) (Fig.7C). In addition, we analyzed the overall survival in the Peters C.Fitzgerald (GSE19417) cohort available in the SurvExpress database. After optimized risk group separation, a total of 70 ESCA patient samples were divided into high-risk (n = 27) and low-risk groups (n = 43) based on their expression patterns (Fig.7D). The survival probability estimates in the two risk groups were represented as Kaplan-Meier plots. Similarly, overall survival analysis showed that the patients in the high-risk group had poorer survival (HR = 2.25 (CI: 1.34 - 3.79); Log-Rank p = 0.001659) than the low-risk group. As shown in Fig.5, The 5 genes were significantly over-expressed in the high risk patients compared to the low risk group. Moreover, the expression of *SPP1, WNT5A, PTHLH, PFN2*, and *GPC1* successfully discriminated the survival of the ESCA high risk group from that of the low risk group in three ESCA cohorts (288 patients). These findings indicated that over-expression of the 5 genes is associated with a poor prognosis of ESCA and presents *SPP1, WNT5A, PTHLH, PFN2*, and *GPC1* as an EMT signature based on changes in the *KEAP1-NRF2* pathway in ESCA.

Discussion

The key role of KEAP1-NRF2 pathway alterations in developing drug- and radio-resistance in ESCA is well-established. The lack of specific biomarkers for *KEAP1-NRF2* pathway alterations in ESCA motivated us to analyze TCGA-ESCA data in order to identify different biological processes that are regulated by *KEAP1-NRF2* in ESCA; hence identifying new therapeutic targets and biomarkers to predict prognosis of ESCA patients. We found that the *KEAP1-NRF2* pathway was altered in 12% of TCGA-ESCA patients. *Swada* et al., performed whole-exome sequence analysis of tumor and nontumor esophageal tissues collected from 144 patients with ESCA (11). They found that *NRF2* was mutated in 16.7% of the patients

and it was one of the most frequently mutated gene in their cohort while *KEAP1* was mutated in almost 5% of patients. We performed GSEA to detect gene sets that showed statistically-notable differences between *KEAP1- NRF2*-mutated samples and their wild-type counterparts groups. Interestingly, 4 gene sets, namely, estrogen response late, hypoxia, reactive oxygen species pathway and epithelial mesenchymal transition, were greatly enriched, with FDR < 0.05. Since the ultimate effect of *KEAP1-NRF2* pathway alterations is the stabilization of NRF2, the master transcriptional regulator of the cells antioxidant program(12), it was not surprising to find the reactive oxygen species pathway among the top pathways that show statistically notable differences between the *KEAP1- NRF2*-mutated and wild-type groups. Surprisingly, pathways such as EMT and hypoxia were greatly enriched. As EMT was more enriched, we selected EMT pathway to identify a *NRF2-KEAP1* pathway signature in ESCA. EMT, an evolutionarily conserved developmental program, has been implicated in carcinogenesis and confers metastatic properties upon cancer cells by enhancing mobility, invasion, and resistance to apoptotic stimuli. Furthermore, EMT-derived tumor cells acquire stem cell properties and exhibit marked therapeutic resistance (13). Given these attributes, the complex biological process of the EMT has been heralded as a key hallmark of carcinogenesis, and targeting EMT pathways constitutes an attractive strategy for cancer treatment (14). Recently, it has been suggested that *NRF2* contributes to malignant transformation of pancreatic duct epithelium through distinct EMT-related mechanisms accounting for an invasive phenotype (15). Furthermore, the expression of *NRF2* is correlated with the lymph node metastasis of esophageal squamous cell carcinoma and blockage of *NRF2* enhances the expression of E-cadherin , the well-known marker of epithelial cell polarity (16). In order to identify an EMT-derived signature for the *KEAP1-NRF2* pathway in ESCA, we integrated the EMT gene list obtained from GSEA with the DEGs between the mutated and wild-type groups (log FC> |1.5|, FDR < 0.05). Interestingly, 11 genes were identified (*SPP1, PTHLH, WNT5A, COL11A1, COL7A1, GPC1, SNAI2, ADAM12, FBN2, PFN2*, and *IGFBP3*). Then, we validated these 11 genes by subjecting the *KEAP1-NRF2*-muatated and wild-type ESAC cell lines to differential gene expression analysis. Intriguingly, only 5 of the 11 genes showed significant up-regulation between the two groups (*SPP1, WNT5A, PTHLH, PFN2,* and *GPC1*). Further, we evaluated the prognostic power of these 5 genes using three ESCA cohorts (288 patients), and we found that the over-expression of these five genes were associated with a poor prognosis in ESCA.

In agreement with our analysis, *SPP1* and EMT have been shown by bioinformatics analysis to have a close association in colorectal cancer (17). Moreover, Osteopontin, encoded by *SPP1* promotes the EMT in hepatocellular carcinoma through regulating vimentin (18), and high *SPP1* expression in hepatocellular carcinoma is associated with poor survival outcome (19). Additionally, up-regulation of *WNT5A* has been suggested to promote EMT and metastasis in pancreatic cancer models, which involves activation of β-catenin-dependent canonical Wnt signaling(20). Furthermore, it has been illustrated that *WNT5A* promotes EMT and metastasis in non-small-cell lung cancer (NSCLC), and high *WNT5A* expression is associated with poor prognosis in NSCLC patients (21). In addition, parathyroid hormone related-protein, encoded by *PTHLH* has been found to promote EMT in prostate and pancreatic cancers (22, 23). It has been indicated that *PTHLH* is a poor prognosis marker and promotes cell growth of HNSC. Besides, it has been illustrated that inhibition of *PFN2* hinders cell invasion and migration, as well as induces an EMT

phenotype, including increased expression of epithelial marker E-cadherin, decreased mesenchymal marker Vimentin, Snail, Slug and ZEB1, and morphological changes in ESCA cells in vitro (24). High *PFN2* expression independently predicts poor overall survival in primary HNSC and ESCA (24,25). Moreover, Over-expression of *GPC1* activates EMT which then increases invasion and migration in colorectal cancer and ESCA (26-28). Additionally, it has been suggested that *GPC1* plays an important role in regulating TGF-β-mediated EMT and stemness, and could be a potential future therapeutic target to prevent progression of gastric cancer (29). It has been pointed out that *GPC1* is over-expressed and implies a poor prognosis in several cancers including, ESCA, uterine cervical cancer, pancreatic cancer, and methothelioma (30-34). Altogether, the above evidence suggests an oncogenic role of the 5-gene signature in many cancers.

Conclusions

Our study identified an EMT-derived gene signature regulated by the *KEAP1-NRF2* pathway that is strongly associated with tumorigenesis, metastasis, and drug resistance in ESCA. This 5-gene signature provides potential biomarkers and therapeutic targets for ESCA patients in whom *KEAP1-NRF2* pathway is activated.

List Of Abbreviations

KEAP1		Kelch-like ECH-associated protein 1
NRF2		Nuclear factor erythroid 2-related 2
LUAD		Lung adenocarcinoma
NSCLC		Non-small-cell lung cancer
TCGA		The cancer genome atlas
ARE		Antioxidant response element
ESCA	ESE	Esophageal cancer
EMT		Epithelial-mesenchymal transition
CCLE		Cancer cell line encyclopedia
DEGs		Differentially expressed genes
PTHLH		Parathyroid hormone like hormone
FDR		False discovery rate
GSEA		Gene set enrichment analysis

Declarations

Funding

This work was supported by the National Natural Science Foundation of China (31571476, 31971188, and 31370772).

Availability of data and materials

The datasets used in this study are publicly available as noted in the text.

Author information

Affiliations

[1]Department of Biochemistry and Department of Thoracic Surgery of the First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310003, PR China;

[2]Department of Pharmacology and Cancer Institute, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310009, PR China;

[3]Labeled Compounds Department, Hot Labs Center, Egyptian Atomic Energy Authority, Cairo 13759, Egypt;

[4]Radiation Biology Department, National Center for Radiation Research and Technology, Egyptian Atomic Energy Authority, Cairo 13759, Egypt;

Corresponding author

Correspondenceto Xiuwen Tang

Contributions

ME conducted the study. XT supervised the study. AH and XJW assisted in data interpretation. ME and XT wrote the manuscript with assistance from all authors. All authors read and approved the manuscript.

Ethics declarations

Ethics approval and consent to participate

Not required

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

References

1. Testa U, Castelli G, Pelosi E. Esophageal Cancer: Genomic and Molecular Characterization, Stem Cell Compartment and Clonal Evolution. Medicines (Basel). 2017;4(3):67.

2. Lignitto L, LeBoeuf SE, Homer H, Jiang S, Askenazi M, Karakousi TR, et al. Nrf2 Activation Promotes Lung Cancer Metastasis by Inhibiting the Degradation of Bach1. Cell. 2019;178(2):316-29.e18.
3. Hammad A, Namani A, Elshaer M, Wang XJ, Tang X. "NRF2 addiction" in lung cancer cells and its impact on cancer therapy. Cancer Letters. 2019;467:40-9.
4. Namani A, Matiur Rahaman M, Chen M, Tang X. Gene-expression signature regulated by the KEAP1-NRF2-CUL3 axis is associated with a poor prognosis in head and neck squamous cell cancer. BMC Cancer. 2018;18(1):46.
5. Elshaer M, ElManawy AI, Hammad A, Namani A, Wang XJ, Tang X. Integrated data analysis reveals significant associations of KEAP1 mutations with DNA methylation alterations in lung adenocarcinomas. Aging (Albany NY). 2020;12(8):7183-206.
6. Namani A, Zheng Z, Wang XJ, Tang X. Systematic Identification of Multi Omics-based Biomarkers in KEAP1 Mutated TCGA Lung Adenocarcinoma. Journal of Cancer. 2019;10(27):6813-21.
7. Goldman M, Craft B, Hastie M, Repečka K, Kamath A, McDade F, et al. The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. bioRxiv. 2019:326470.
8. Guan X, Cai M, Du Y, Yang E, Ji J, Wu J. CVCDAP: an integrated platform for molecular and clinical analysis of cancer virtual cohorts. Nucleic Acids Research. 2020;48(W1):W463-W71.
9. Kreft Ł, Soete A, Hulpiau P, Botzki A, Saeys Y, De Bleser P. ConTra v3: a tool to identify transcription factor binding sites across species, update 2017. Nucleic Acids Research. 2017;45(W1):W490-W4.
10. Aguirre-Gamboa R, Gomez-Rueda H, Martínez-Ledesma E, Martínez-Torteya A, Chacolla-Huaringa R, Rodriguez-Barrientos A, et al. SurvExpress: an online biomarker validation tool and database for cancer gene expression data using survival analysis. PLoS One. 2013;8(9):e74250-e.
11. Sawada G, Niida A, Uchi R, Hirata H, Shimamura T, Suzuki Y, et al. Genomic Landscape of Esophageal Squamous Cell Carcinoma in a Japanese Population. Gastroenterology. 2016;150(5):1171-82.
12. Jaramillo MC, Zhang DD. The emerging role of the Nrf2-Keap1 signaling pathway in cancer. Genes & development. 2013;27(20):2179-91.
13. Mittal V. Epithelial Mesenchymal Transition in Tumor Metastasis. Annual Review of Pathology: Mechanisms of Disease. 2018;13(1):395-412.
14. Roche J. The Epithelial-to-Mesenchymal Transition in Cancer. Cancers (Basel). 2018;10(2):52.
15. Arfmann-Knübel S, Struck B, Genrich G, Helm O, Sipos B, Sebens S, et al. The Crosstalk between Nrf2 and TGF-β1 in the Epithelial-Mesenchymal Transition of Pancreatic Duct Epithelial Cells. PLoS One. 2015;10(7):e0132978-e.
16. Shen H, Yang Y, Xia S, Rao B, Zhang J, Wang J. Blockage of Nrf2 suppresses the migration and invasion of esophageal squamous cell carcinoma cells in hypoxic microenvironment. Diseases of the esophagus : official journal of the International Society for Diseases of the Esophagus. 2014;27(7):685-92.

17. Xu C, Sun L, Jiang C, Zhou H, Gu L, Liu Y, et al. SPP1, analyzed by bioinformatics methods, promotes the metastasis in colorectal cancer by activating EMT pathway. Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie. 2017;91:1167-77.
18. Dong Q, Zhu X, Dai C, Zhang X, Gao X, Wei J, et al. Osteopontin promotes epithelial-mesenchymal transition of hepatocellular carcinoma through regulating vimentin. Oncotarget. 2016;7(11):12997-3012.
19. Menyhárt O, Nagy Á, Győrffy B. Determining consistent prognostic biomarkers of overall survival and vascular invasion in hepatocellular carcinoma. Royal Society Open Science. 2018;5(12):181006.
20. Bo H, Zhang S, Gao L, Chen Y, Zhang J, Chang X, et al. Upregulation of WNT5A promotes epithelial-to-mesenchymal transition and metastasis of pancreatic cancer cells. BMC Cancer. 2013;13(1):496.
21. Wang B, Tang Z, Gong H, Zhu L, Liu X. WNT5A promotes epithelial-to-mesenchymal transition and metastasis in non-small-cell lung cancer. Biosci Rep. 2017;37(6):BSR20171092.
22. Ongkeko WM, Burton D, Kiang A, Abhold E, Kuo SZ, Rahimy E, et al. Parathyroid hormone related-protein promotes epithelial-to-mesenchymal transition in prostate cancer. PLoS One. 2014;9(1):e85803-e.
23. Pitarresi JR, Norgard RJ, Stanger BZ, Rustgi AK. Abstract B43: p120 catenin loss drives pancreatic cancer EMT and metastasis through activation of PTHrP-mediated calcium signaling. Cancer Research. 2019;79(24 Supplement):B43-B.
24. Cui X-B, Zhang S-M, Xu Y-X, Dang H-W, Liu C-X, Wang L-H, et al. PFN2, a novel marker of unfavorable prognosis, is a potential therapeutic target involved in esophageal squamous cell carcinoma. J Transl Med. 2016;14(1):137-.
25. Liu J, Wu Y, Wang Q, Liu X, Liao X, Pan J. Bioinformatic analysis of PFN2 dysregulation and its prognostic value in head and neck squamous carcinoma. Future Oncology. 2018;14(5):449-59.
26. Li J, Li B, Ren C, Chen Y, Guo X, Zhou L, et al. The clinical significance of circulating GPC1 positive exosomes and its regulative miRNAs in colon cancer patients. Oncotarget. 2017;8(60):101189-202.
27. Li J, Chen Y, Zhan C, Zhu J, Weng S, Dong L, et al. Glypican-1 Promotes Tumorigenesis by Regulating the PTEN/Akt/β-Catenin Signaling Pathway in Esophageal Squamous Cell Carcinoma. Digestive Diseases and Sciences. 2019;64(6):1493-502.
28. Li Y, Li M, Shats I, Krahn JM, Flake GP, Umbach DM, et al. Glypican 6 is a putative biomarker for metastatic progression of cutaneous melanoma. PLoS One. 2019;14(6):e0218067.
29. Wang S, Wu Z, Zhou M, Liao W. Effect of GPC1 on epithelial-to-mesenchymal transition and stemness and interaction with ITGB1 in gastric cancer. Journal of Clinical Oncology. 2017;35(15_suppl):e15580-e.
30. Hara H, Takahashi T, Serada S, Fujimoto M, Ohkawara T, Nakatsuka R, et al. Over-expression of glypican-1 implicates poor prognosis and their chemoresistance in oesophageal squamous cell carcinoma. British journal of cancer. 2016;115(1):66-75.
31. Matsuzaki S, Serada S, Hiramatsu K, Nojima S, Matsuzaki S, Ueda Y, et al. Anti-glypican-1 antibody-drug conjugate exhibits potent preclinical antitumor activity against glypican-1 positive uterine

cervical cancer. 2018;142(5):1056-66.
32. Duan L, Hu XQ, Feng DY, Lei SY, Hu GH. GPC-1 may serve as a predictor of perineural invasion and a prognosticator of survival in pancreatic cancer. Asian journal of surgery. 2013;36(1):7-12.
33. Amatya VJ, Kushitani K, Kai Y, Suzuki R, Miyata Y, Okada M, et al. Glypican-1 immunohistochemistry is a novel marker to differentiate epithelioid mesothelioma from lung adenocarcinoma. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc. 2018;31(5):809-15.
34. Kato D, Yaguchi T, Iwata T, Katoh Y, Morii K, Tsubota K, et al. GPC1 specific CAR-T cells eradicate established solid tumor without adverse effects and synergize with anti-PD-1 Ab. eLife. 2020;9:e49392.

Figures

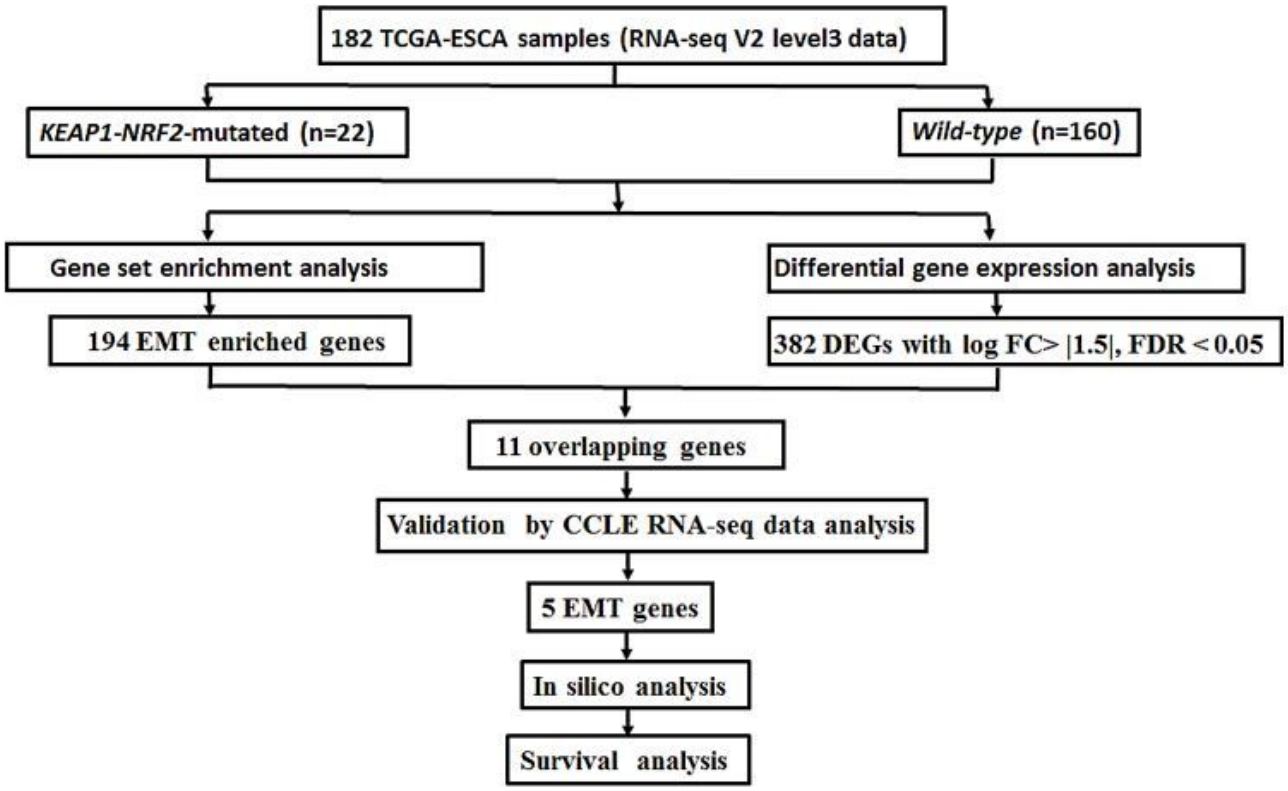

Figure 1

Schematic diagram showing the analysis overflow that was followed in this study.

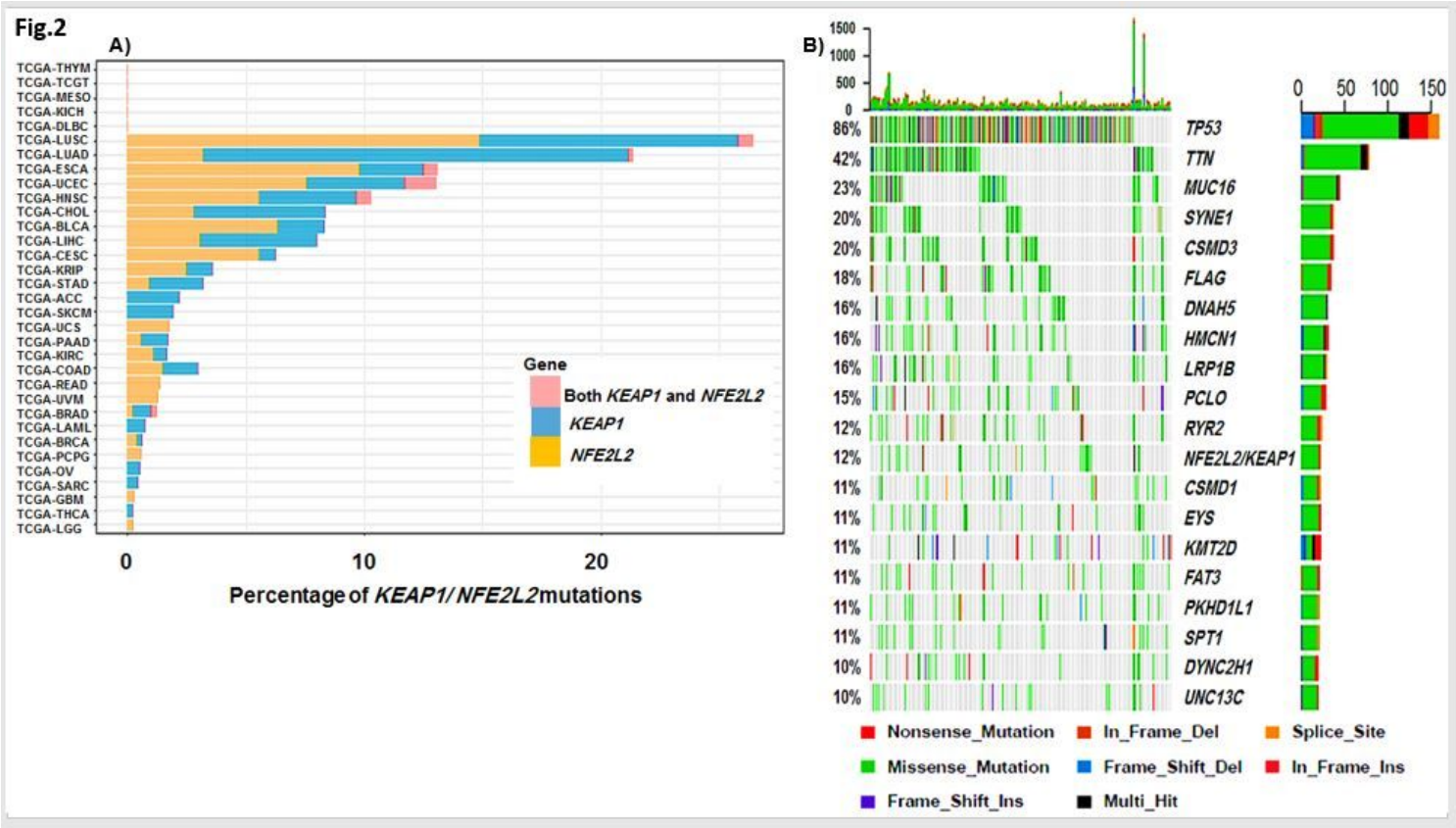

Figure 2

General mutational landscape A) Bar chart representing TCGA-pan cancer analysis of KEAP1-NRF2 pathway alterations in diffrent cancers. B) Landscape of genetic alterations across the 185 ESCA samples. The samples are sorted by mutation rates (top bars), while genes are sorted by the proportion of altered samples (left bars).

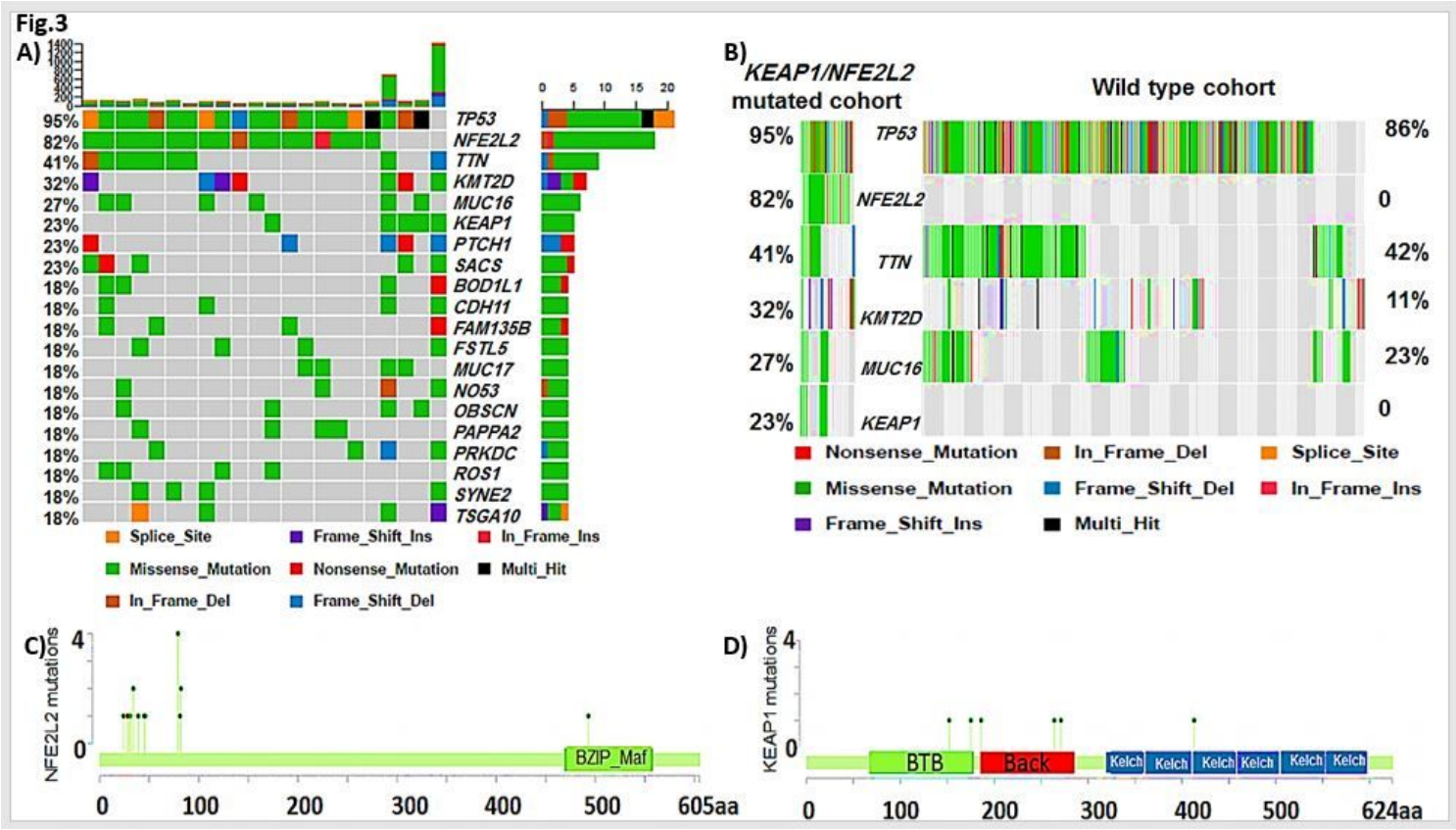

Figure 3

Mutational landscape of ESCA samples with KEAP1 and/or NRF2 mutations. A) Landscape of genetic alterations across the 22 ESCA samples with KEAP1 and/or NRF2 mutations. The samples are sorted by mutation rates (top bars), while genes are sorted by the proportion of altered samples (left bars). B) Differential mutational analysis between KEAP1-NRF2-mutated and wild-type ESCA samples. C) Lollipop plot showing the locations of mutations in the functional domains of NRF2 protein. D) Lollipop plot showing the locations of mutations in the functional domains of KEAP1 protein. The lollipops show the locations of the mutations as identified by whole-exon sequencing.

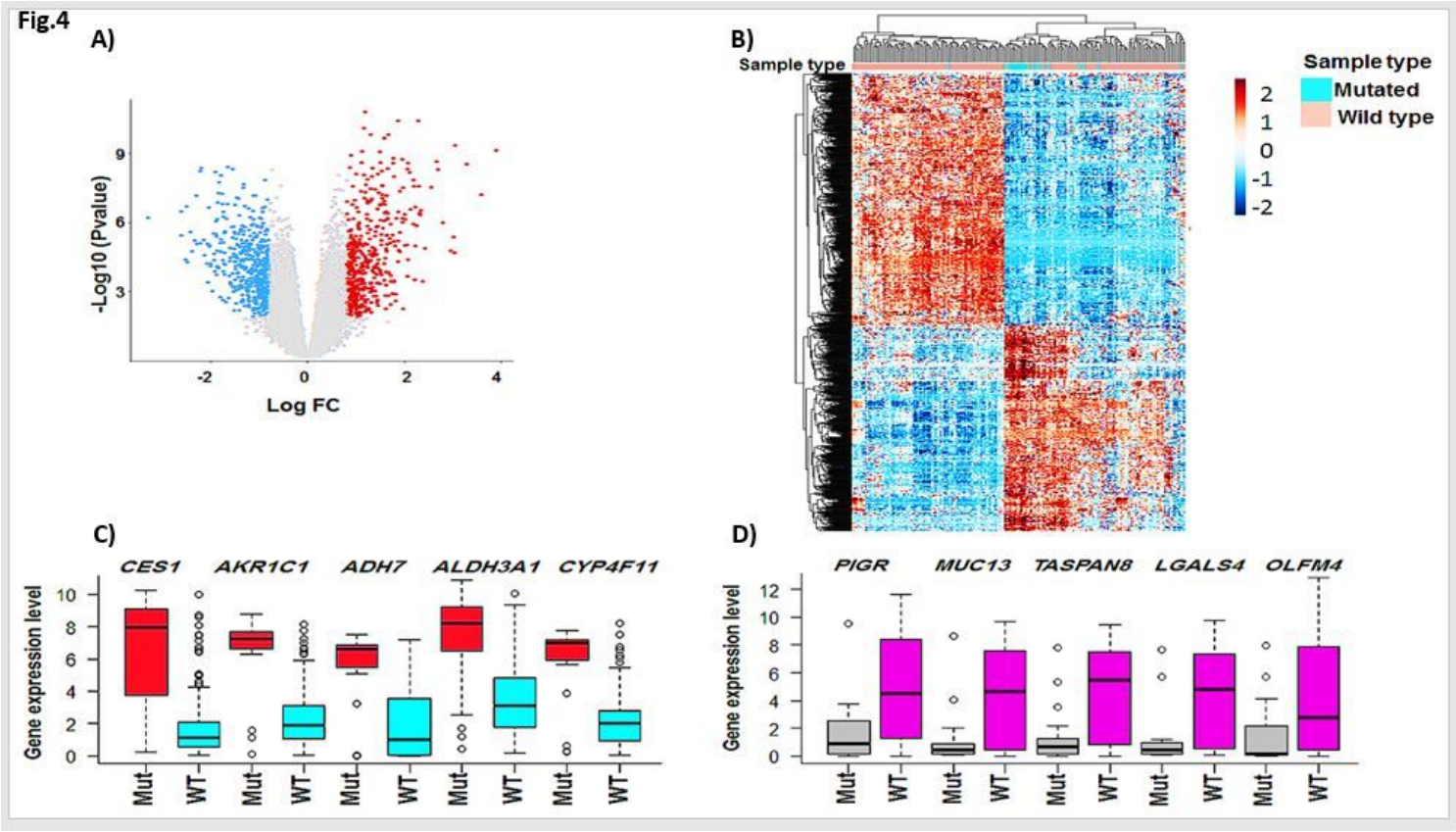

Figure 4

Differential gene expression analysis. A) Volcano plot showing the distribution of DEGs between KEAP1-NRF2-mutated and wild-type ESCA patient samples based on significance and fold change. B) Heatmap showing the top DEGs between KEAP1- NRF2 mutated and wild-type ESCA patient samples with Log FC>|1| and FDR <0.05. C) Box plots showing the top 5 overexpressed genes between KEAP1- NRF2-mutated and wild-type ESCA patient samples. D) Box plots showing the top 5 down-regulated genes between KEAP1-NRF2-mutated and wild-type ESCA patient samples. Center lines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles; outliers are represented by dots.

Figure 5

Gene set enrichment analysis. Enrichment plots of four gene sets that are importantly differentiated between KEAP1-NRF2-mutated and wild-type ESCA samples.

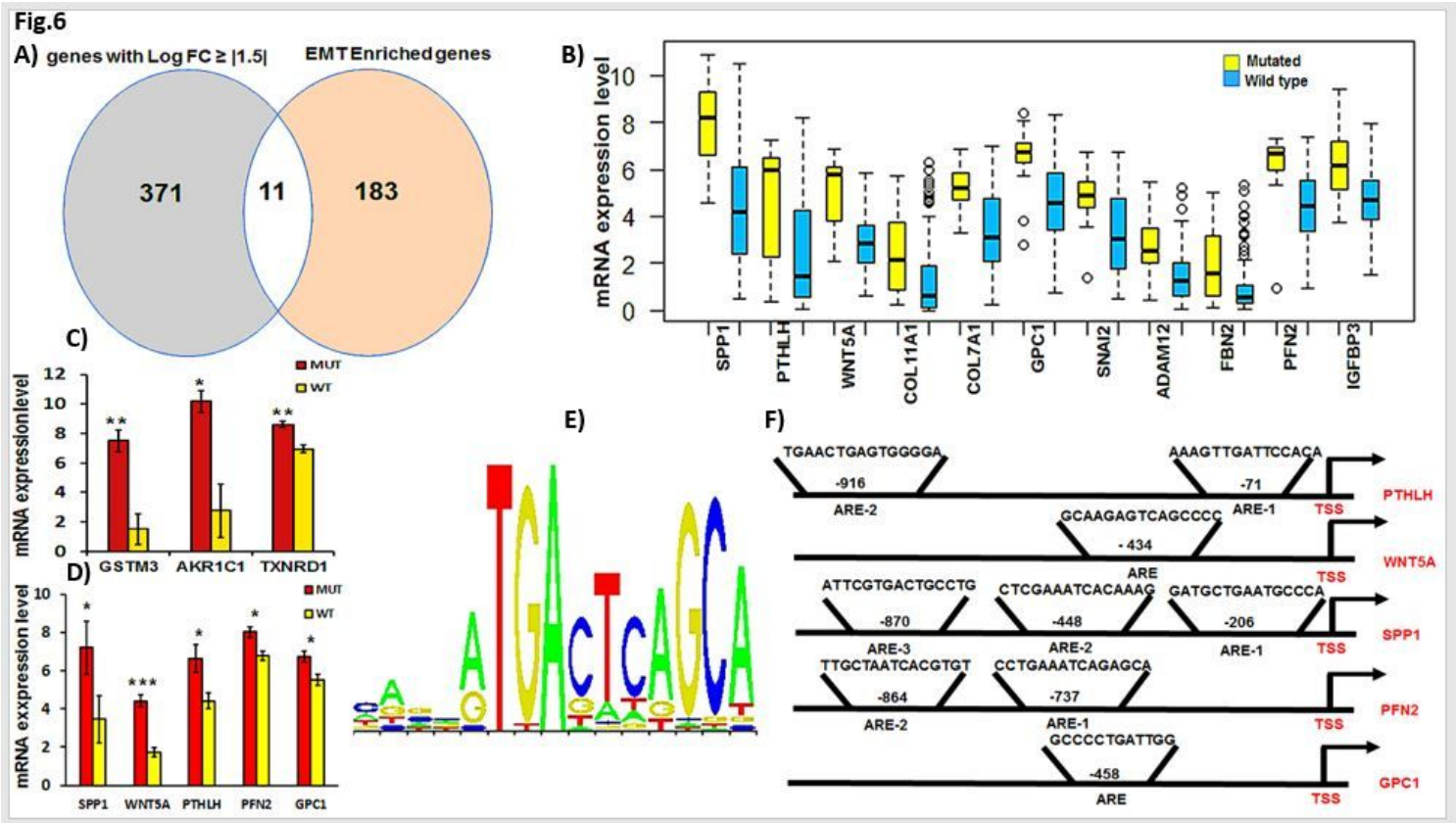

Figure 6

Identification of the EMT signature of ESCA patients with altered KEAP1-NRF2 pathway A) Venny diagram showing the overlapping between the EMT-enriched gene set and DEGs between KEAP1-NRF2-mutated and wild-type ESCA patient samples. B) Box plots showing the differential expression of 11 overlapped EMT genes between KEAP1- NRF2-mutated and wild-type ESCA patient samples. C) Bar chart showing differential mRNA expression of some well-known NRF2 targets between ESCA cell lines with KEAP1 and/or NRF2 mutations and their wild-type counterparts. D) Bar chart showing 5 EMT genes that were significantly differentially expressed between ESCA cell lines with KEAP1 and/or NRF2 mutations and their wild-type counterparts. (*p < 0.05, ** p < 0.01, *** p < 0.001). E) The NRF2 binding motif as provided by JASPER. (F) Schematic representation of the locations of insilico-predicted NRF2 binding sites (AREs) in the promoter regions of the human SPP1, WNT5A, PTHLH, PFN2, and GPC1genes.

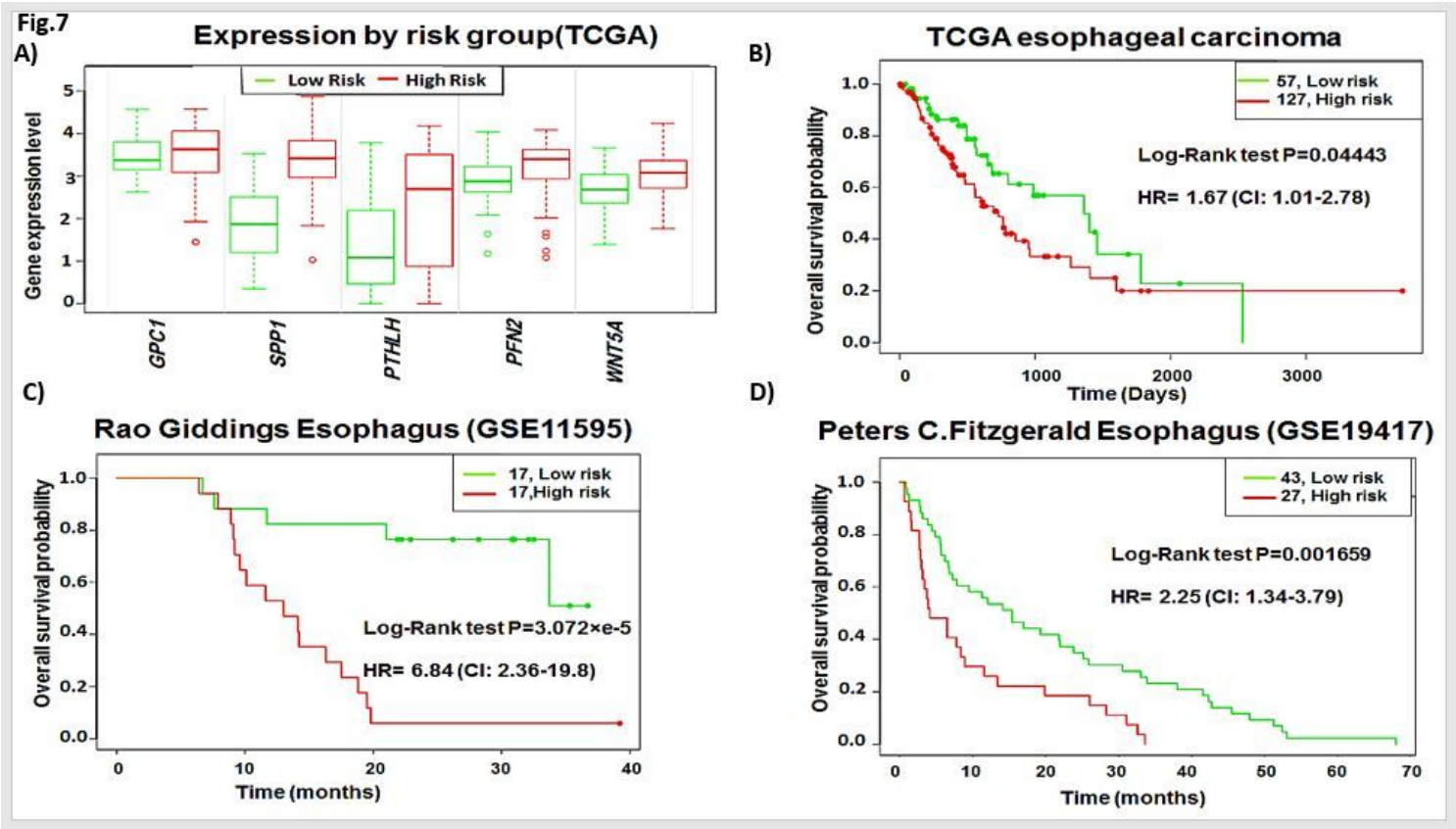

Figure 7

Five-gene signature predicts poor survival in three independent cohorts. (A) Box plots showing the expression differences of the 5-gene signature in low (green) and high (red) risk groups of TCGA–ESCA patients (y-axis, gene expression value of each gene).(B) Kaplan-Meier survival plots showing that high expression of the 5-gene signature is associated with poor survival in TCGA–ESCA patients. (C) The Rao Giddings (GSE11595) cohort. D) The Peters C.Fitzgerald (GSE19417) cohort. Red, high-risk group; green, low-risk group; top right corner inset, numbers of high- and low-risk samples (x-axis, time; y-axis, overall survival probability; HR, hazard ratio; CI, confidence interval).

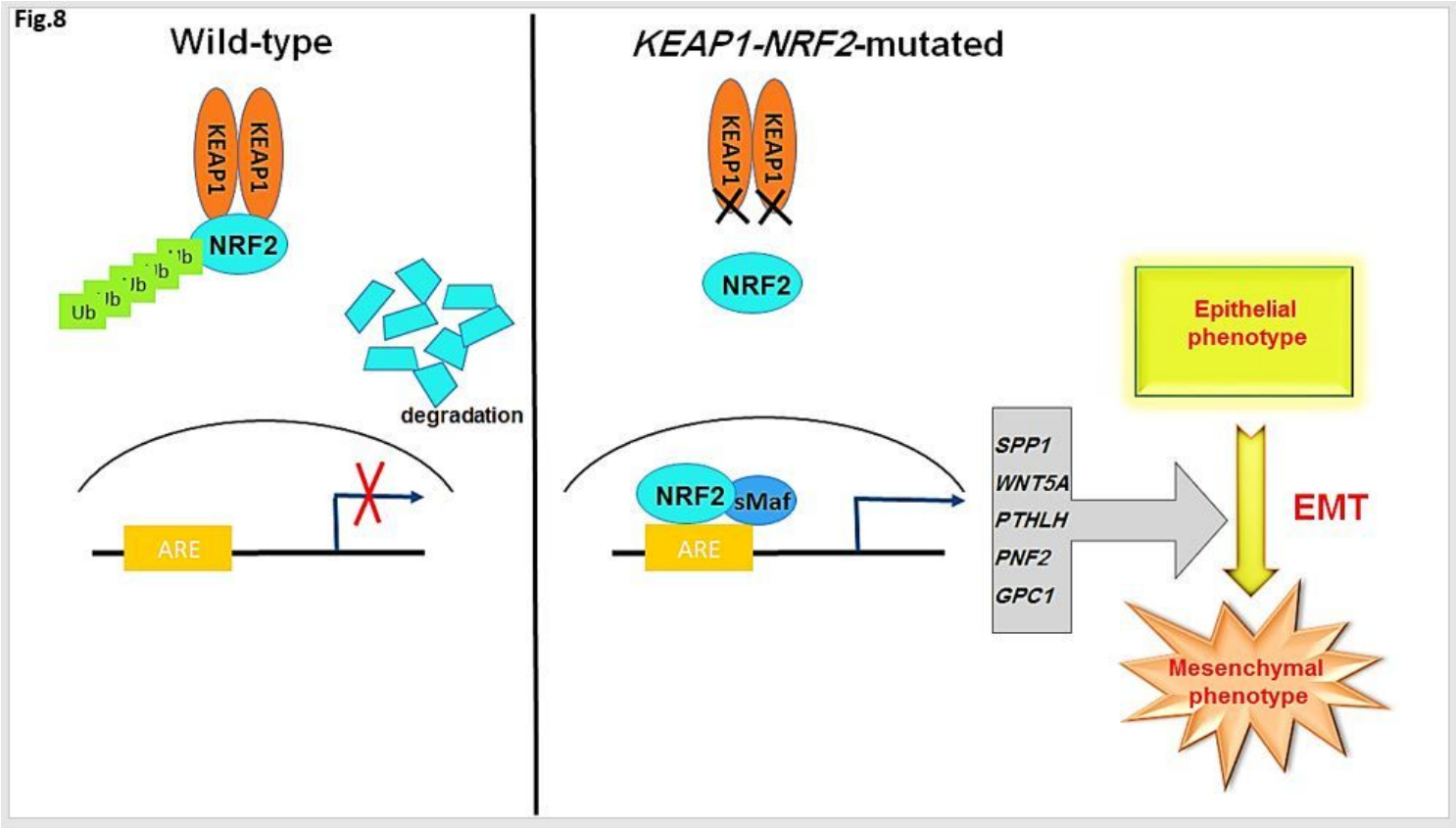

Figure 8

Schematic diagram summarizing the findings of this study.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile2.xlsx
- Additionalfile2.xlsx
- Additionalfile1.xlsx
- Additionalfile1.xlsx