OMICS RESEARCH SYMPOSIUM

Transcriptomic Analysis of histological subtypes in Esophageal Carcinoma

Dhruv Mehra, Dr. Harpreet Kaur, Elia Brodsky ,Dr. Mohit Mazumder

Research Fellow
Pine Biotech
New Orleans,
Louisiana, USA

Introduction

The Esophageal Carcinoma is the 6th most common cancer in the world, accounting for 604100 new cases and 544076 deaths in the year 2020. The disease has a diverse geographical distribution. While Asia reported the largest number of Incidence and mortality, It is also seen that the incidence of Esophageal Cancer in Men is considerably more than the Females (Sung H et al). Moreover, the 5-year survival of Esophageal Carcinoma is between 12% and 20%, owing to the fact that the disease is usually diagnosed at advanced stages and is usually resistant to the majority of therapies due to molecular heterogeneity (Napier KJ et al).

The Esophageal Carcinoma has two major histological categories called the Esophageal Adenocarcinoma (EAC) and the Esophageal Squamous Cell Carcinoma (ESCC). The ESCC subtype is the neoplasm of squamous cells, usually associated with alcohol consumption and tobacco use. It is localised to the middle and lower esophagus. The esophageal Adenocarcinoma is predominant in the lower third of the esophagus. Unlike ESCC, Esophageal Adenocarcinoma is closely associated with conditions like gastroesophageal reflux disease, untreated can lead to Barrett's Esophagus (Napier KJ et al), a pre-neoplastic condition where the squamous epithelium is replaced by columnar epithelium.

The ESCC is the most prevalent histological subtype of Esophageal Cancer in the world, developing nations such as India reports the majority of the ESCC cases (Figure-1). On the contrary, EAC is majorly diagnosed in the population of developed nations like the USA, UK (Joel H. Rubenstein & Nicolas J. Shaheen).

Treatment of Esophageal Cancer includes surgical resection, chemotherapy and radiotherapy. Although the two histological subtype exhibits similar characteristics but still in clinical studies, they are not differentiated for the purpose of therapy.

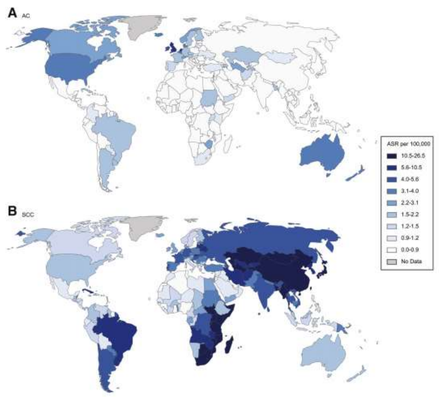

Figure 1: *Global differences in the incidence of histological subtypes of esophageal cancer , cited: Joel H. Rubenstein & Nicolas J. Shaheen*

Methods

DATASET

For this study, The Cancer Genome Atlas Esophageal Carcinoma (ESCA) RNA-seq dataset (retrieved from firebrowse) was used. Among the 198 samples analysed, there were 97 ESCC tumour samples, 88 EAC tumour samples and 13 Normal samples (Fig-2).

CLUSTERING AND CLASSIFICATION

Principal Component Analysis was performed on the dataset using the T-Bioinfo server. The expression RPKM values table (Quantile normalized and log2 transformed) was uploaded on the server and the PCA Pipeline was performed on the data to see clustering based on the subtype.

DIFFERENTIAL GENE EXPRESSION ANALYSIS AND PATHWAY ENRICHMENT ANALYSIS

The data was analysed using the T-bioinfo Server. The Differential Gene expression analysis was performed using the DEseq2 algorithm on the T-bioinfo server, for the pathway enrichment, GSEA enrichment and Pathway Enrichment algorithms on the T-bioinfo were used (streamlined with DEseq2).
The three different analyses were performed for understanding diverse affairs involved in the subtypes:
- Differential Gene Expression Analysis was performed between the 88 EAC samples and 97 ESCC samples. For the analysis, raw read counts were used.
- Differential Gene Expression Analysis and Pathway Enrichment analysis was done between two groups:
  - 13 Normal samples and 88 EAC samples
  - 13 Normal samples and 97 ESCC samples

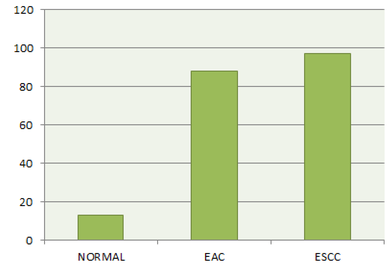

Figure 2: *Bar graph signifying number of samples in each category*

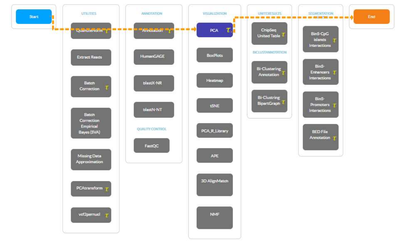

Figure 5 : *Principal Component Analysis Pipeline Graph*

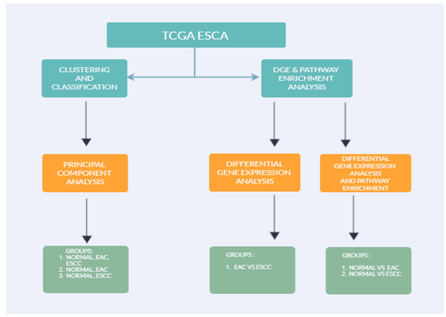

Figure 4 : *Analysis Flow Chart*

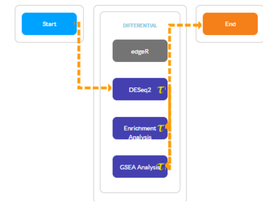

Figure 5 : *DEG and Pathway Enrichment Analysis Pipeline Graph*

Results

Result of Clustering

The PCA performed on 198 samples based on the histological subtype. The two histological subtypes name EAC and ESCC separated along the X-axis (PC1), while the normal samples were mixed with EAC (Figure 6). Another PCA plotted between Normal and individual subtype showed two separate clusters (Figure 7 & 8).

Result of Differential Gene Expression Analysis (EAC VS ESCC)

A total of 3633 Genes were found to be Differentially Expressed between the subtypes (with a P-adjusted value of >0.05 and log Fc Value greater than 1.5 or less than -1.5). It was also seen that 43 lncRNAs genes and 66 miRNAs genes were also differentially expressed between the two subtype (Figure 9).

Result of DGE and Pathway Enrichment

1. Normal VS EAC

DGE analysis revealed 1807 Up-regulated genes (with a P-adjusted value of >0.05 and log Fc Value greater than 1.5) and 1807 down-regulated genes (with a P-adjusted value of >0.05 and log Fc Value less than -1.5). While the pathway enrichment revealed that the majority of the Genes were enriched in fat metabolism pathways, chemokine signalling, Cell cycle-related pathways (Figure 10).

2. Normal VS ESCC

DGE analysis revealed 2277 Up-regulated Genes (with a P-adjusted value of >0.05 and log Fc Value greater than 1.5) and 2250 down-regulated genes (with a P-adjusted value of >0.05 and log Fc Value less than -1.5). Pathway enrichment analysis of majority of T cel ways, DNA ways (

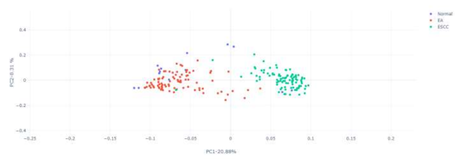

Figure 6 : *PCA (Normal, EAC and ESCC)*

Figure 7 : *PCA (Normal & EAC)*

Figure 8: *PCA (Normal & ESCC)*

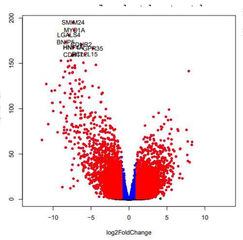

Figure 9 : *Volcano Plot*

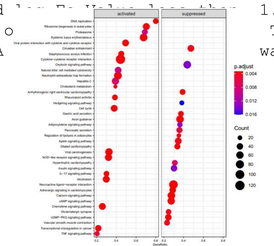

Figure 10 : *Pathway Enrichment Analysis, EAC subtype*

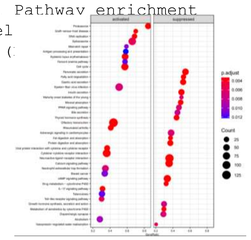

Figure 11 : *Pathway Enrichment Analysis, ESCC subtype*

Conclusions

A number of differentially expressed genes were discovered between the two subtypes. It was seen that a number of differentially expressed long non-coding RNA genes and miRNAs were discovered. Both lncRNA and miRNA plays an important role in Gene regulation thus, making them a new class of targets for drug discovery (Matsui M and Corey DR).

HOTAIR or HOX Transcript Antisense Intergenic RNA, an important lncRNA found in cancer usually involved in cancer progression and drug resistance, was found to be up-regulated in Esophageal adenocarcinoma (Rajagopal T et al). HOTTIP OR HOXA transcript at the distal tip, an important prognostic marker in various cancers was up-regulated in EAC (Fig-13).

A number of long intergenic non coding RNAs were found to be up-regulated in ESCC including LINC01549 which is an important lncRNA in Hepatocellular carcinoma (Ye J et al) (Fig-12).

As we discovered a number of lncRNAs and miRNA differentially expressed between the two histological subtypes, we can work towards the path of Precision Medicine based on the histology.

The Pathway enrichment revealed some of the similarities between EAC and ESCC. While both shared up-regulated pathways related to DNA confirmation change, Cell cycle regulation, Signalling pathway, the Lipid metabolism related pathways were activated in both.

SYMBOL	GENENAME	SYMBOL	GENENAME
CCDC140	CCDC140 long non-coding RNA	MIR1231	microRNA 1231
CCDC26	CCDC26 long non-coding RNA	MIR1260A	microRNA 1260a
LINC00051	long intergenic non-protein coding RNA 51	MIR1908	microRNA 1908
LINC00052	long intergenic non-protein coding RNA 52	MIR1910	microRNA 1910
LINC00161	long intergenic non-protein coding RNA 161	MIR203A	microRNA 203a
LINC00163	long intergenic non-protein coding RNA 163	MIR203B	microRNA 203b
LINC00173	long intergenic non-protein coding RNA 173	MIR205HG	MIR205 host gene
LINC00184	long intergenic non-protein coding RNA 184	MIR2117	microRNA 2117
LINC00240	long intergenic non-protein coding RNA 240	MIR2355	microRNA 2355
LINC00269	long intergenic non-protein coding RNA 269	MIR31	microRNA 31
LINC00303	long intergenic non-protein coding RNA 303	MIR31HG	MIR31 host gene
LINC00320	long intergenic non-protein coding RNA 320	MIR3660	microRNA 3660
LINC00347	long intergenic non-protein coding RNA 347	MIR3666	microRNA 3666
LINC00461	long intergenic non-protein coding RNA 461	MIR4252	microRNA 4252
LINC00486	long intergenic non-protein coding RNA 486	MIR4497	microRNA 4497
LINC00592	long intergenic non-protein coding RNA 592	MIR4671	microRNA 4671
LINC00606	long intergenic non-protein coding RNA 606	MIR4675	microRNA 4675
LINC00615	long intergenic non-protein coding RNA 615	MIR497	microRNA 497
LINC00896	long intergenic non-protein coding RNA 896	MIR548K	microRNA 548k
LINC01363	long intergenic non-protein coding RNA 1363	MIR575	microRNA 575
LINC01549	long intergenic non-protein coding RNA 1549	MIR708	microRNA 708
LINC01565	long intergenic non-protein coding RNA 1565	MIR936	microRNA 936
LINC02872	long intergenic non-protein coding RNA 2872	MIR944	microRNA 944

Figure 12: *miRNAs and lncRNAs upregulated in ESCC subtype*

SYMBOL	GENENAME	SYMBOL	GENENAME
HOTAIR	HOX transcript antisense RNA	MIR1236	microRNA 1236
HOTTIP	HOXA distal transcript antisense RNA	MIR1257	microRNA 1257
LINC00114	long intergenic non-protein coding RNA 114	MIR126	microRNA 126
LINC00261	long intergenic non-protein coding RNA 261	MIR1293	microRNA 1293
LINC00323	long intergenic non-protein coding RNA 323	MIR135A1	microRNA 135a-1
LINC00336	long intergenic non-protein coding RNA 336	MIR192	microRNA 192
LINC00473	long intergenic non-protein coding RNA 473	MIR196A1	microRNA 196a-1
LINC00482	long intergenic non-protein coding RNA 482	MIR2276	microRNA 2276
LINC00494	long intergenic non-protein coding RNA 494	MIR3131	microRNA 3131
LINC00511	long intergenic non-protein coding RNA 511	MIR3185	microRNA 3185
LINC00574	long intergenic non-protein coding RNA 574	MIR3189	microRNA 3189
LINC00619	long intergenic non-protein coding RNA 619	MIR3197	microRNA 3197
LINC01555	long intergenic non-protein coding RNA 1555	MIR326	microRNA 326
LINC01558	long intergenic non-protein coding RNA 1558	MIR3621	microRNA 3621
LINC01559	long intergenic non-protein coding RNA 1559	MIR3646	microRNA 3646

Figure 13: *miRNAs and lncRNAs upregulated in EAC subtype*

REFERENCE:
Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021 Feb 4.

Napier KJ, Scheerer M, Misra S. Esophageal cancer: A Review of epidemiology, pathogenesis, staging workup and treatment modalities. World J Gastrointest Oncol 2014;6:112-20

Joel H. Rubenstein & Nicolas J. Shaheen, Epidemiology, Diagnosis and Management of Wsophageal Adenocarcinoma, Gasteroenterology, vol 149, issue 2

Matsui M, Corey DR. Non-coding RNAs as drug targets. *Nat Rev Drug Discov*. 2017;16(3):167-179. doi:10.1038/nrd.2016.117

Rajagopal T, Talluri S, Akshaya RL, Dunna NR. HOTAIR LncRNA: A novel oncogenic propellant in human cancer. Clin Chim Acta. 2020 Apr;503:1-18. doi: 10.1016/j.cca.2019.12.028. Epub 2019 Dec 31. PMID: 31901481.

Ye J, Li H, Wei J, Luo Y, Liu H, Zhang J, Luo X. Risk Scoring System based on lncRNA Expression for Predicting Survival in Hepatocellular Carcinoma with Cirrhosis. Asian Pac J Cancer Prev. 2020 Jun 1;21(6):1787-1795. doi: 10.31557/APJCP.2020.21.6.1787. PMID: 32592379; PMCID: PMC7568908.

Project Mentors- **Dr. Harpreet Kaur, Elia Brodsky ,Dr. Mohit Mazumder**