



## Data in brief

# RNA-Seq analysis of human cell lines established from normal and neoplastic esophageal squamous epithelium



Ichiro Akagi<sup>a</sup>, Osamu Ishibashi<sup>b,\*</sup>

<sup>a</sup> Tokyo Mid-town Medical Center, Japan

<sup>b</sup> Laboratory of Biological Macromolecules, Graduate School of Life and Environmental Sciences, Osaka Prefecture University, Japan

## ARTICLE INFO

## Article history:

Received 27 January 2017

Accepted 2 February 2017

Available online 5 February 2017

## Keywords:

RNA-Seq

Esophageal cancer

## ABSTRACT

Esophageal cancer (EC) is the eighth most common cancer globally in 2012 and predominantly occurs in the man (Enzinger and Mayer, 2003; Conteduca et al., 2012). EC is classified mainly into two types, esophageal squamous cell carcinoma (ESCC) and esophageal adenocarcinoma, accounting for 60–70% and 20–30% of all EC cases, respectively. In a previous statistical study it was reported that the numbers of new EC cases and EC-related deaths worldwide in 2008 were estimated to be 482,300 and 406,800, respectively (Jemal et al., 2011). This high mortality rate is largely due to the characteristics of EC such as frequent distant/local metastasis and poor subjective symptoms leading to difficulty with early diagnosis. Patients affected with EC diagnosed at late stages mostly have unsatisfactory prognosis, even though various therapeutic options are available. Therefore, there is an urgent need to develop effective methods that enable the early detection of EC (Orringer, 1993), prompting us to search for novel biomarkers for EC. Here, we provide datasets from RNA-Seq analysis of Het-1A, a normal human esophageal squamous cell line (Stoner et al., 1991), and TE-1, TE-5, and TE-8, which are well-, poorly-, and moderately-differentiated ESCC-derived cell lines, respectively (Nishihira et al., 1993). The raw data of these experiments have been deposited at DNA Data Bank of Japan (DDBJ) under the accession IDs DRR084199, DRR084200, DRR084201, and DRR084202.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Specifications	
Organism/cell line/tissue	Human cell lines derived from normal (Het-1A) and neoplastic (TE-1, TE-5, and TE-8) esophageal squamous epithelium
Sex	n/a
Sequencer or array type	Illumina HiSeq2500
Data format	Normalized reads mapped to <i>Homo sapiens</i> GRCh38
Experimental factors	Normal and neoplastic human esophageal squamous cell lines
Experimental features	Well (TE-1), poorly (TE-5), and moderately (TE-8) differentiated neoplastic cell lines established from ESCC at stages II, IV, and III, respectively. Het-1A is an SV40 T-antigen-immortalized human esophageal epithelial cells.
Consent	n/a
Sample source location	DDBJ accession IDs DRR084199, DRR084200, DRR084201, and DRR084202. The summary of this project is available on <a href="http://trace.ddbj.nig.ac.jp/BPSearch/bioprotect?acc=PRJDB5493">http://trace.ddbj.nig.ac.jp/BPSearch/bioprotect?acc=PRJDB5493</a> .

## 1. Direct link to deposited data

Deposited data are available here:

<http://trace.ddbj.nig.ac.jp/DRAsearch/run?acc=DRR084199>

<http://trace.ddbj.nig.ac.jp/DRAsearch/run?acc=DRR084200>

<http://trace.ddbj.nig.ac.jp/DRAsearch/run?acc=DRR084201>

<http://trace.ddbj.nig.ac.jp/DRAsearch/run?acc=DRR084202>

## 2. Experimental design, materials and methods

### 2.1. Experimental design, cell culture and RNA isolation

Neoplastic cell lines used in this study were TE-1, TE-5, and TE-8, which were established from well-, poorly-, and moderately-differentiated ESCC, respectively [1]. These cell lines were purchased from RIKEN Cell Bank (Tsukuba, Japan) and cultured in RPMI-1640 medium supplemented with 10% fetal bovine serum. Het-1A, an SV40 large T antigen-harboring normal human esophageal squamous cell line [2], was purchased from American Type Culture Collection (Manassas, VA, USA) and cultured using the BEGM kit (Lonza, Basel, Switzerland) but without addition of the GA-1000 (gentamycin-amphotericin B mix)

\* Corresponding author.

E-mail address: [ishibashi@biochem.osakafu-u.ac.jp](mailto:ishibashi@biochem.osakafu-u.ac.jp) (O. Ishibashi).

**Table 1**  
Overview of sequencing results and statistics.

Sample	Raw reads	Called bases (Mbp)	%Q30 <sup>a</sup>	Average quality
Het-1A	53,656,462	6707	93.55	35.25
TE-1	53,331,576	6666	93.27	35.16
TE-5	58,934,174	7367	93.24	35.16
TE-8	65,921,940	8240	93.28	35.16

<sup>a</sup> Quality score 30, an index to predict the probability of an error in base calling.

**Table 2**  
Overview of mapping.

Sample	Raw reads (M)	Cleaned reads (Mbp)	Reads mapped (M)	Mapping rates (%)
Het-1A	53.7	44.3	40.6	91.5
TE-1	53.3	42.9	38.8	90.5
TE-5	58.9	47.0	42.8	91.0
TE-8	65.9	54.6	50.3	92.1

provided with the kit. Total RNA was isolated from these cells using ISOGEN reagent (Wako, Osaka, Japan) according to the manufacturer's protocol. RNA purity was evaluated by the RNA integrity number (RIN), a representative index to assess RNA quality determined using the Agilent 2100 bioanalyzer (Agilent, Santa Clara, CA, USA). The RIN value of all RNA samples used in this study was shown to be 10. The RNA samples were then subjected to ribosomal and mitochondrial RNA depletion using Ribo-Zero rRNA Removal Kit (Illumina).

2.2. Sequencing, mapping and normalization

Multiplex RNA-seq libraries were generated for samples using TruSeq Stranded Total Library Preparation Kit (Illumina). Libraries were generated according to TruSeq Stranded Total RNA Low Sample Protocol. Sequencing was carried out on HiSeq2500 (Illumina) according to the protocol of 2 × 125 bp run. As an output of this sequencing, 53 million to 66 million raw reads were obtained from the individual

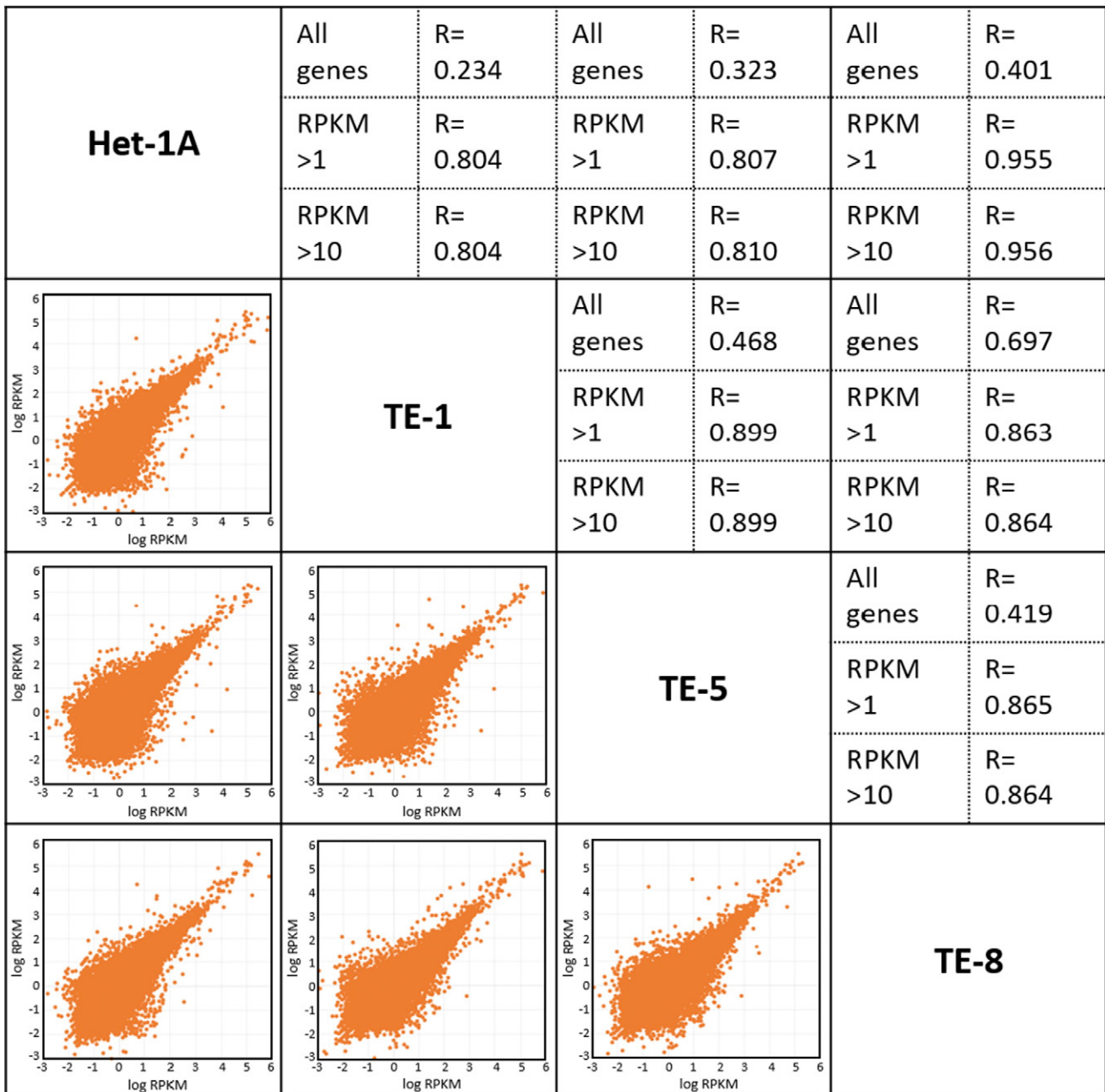


Fig. 1. Pairwise correlation of log-transformed FPKM values.

cell lines (Table 1). Trimming and filtering of the reads to remove adaptor sequences and low-quality nucleotides were performed using Trimmomatic (ver. 0.32) (<http://www.usadellab.org/cms/?page=trimmomatic>).

The filtered reads were aligned to the Ensembl human genome GRCh38.p7 as a reference ([ftp://ftp.ensembl.org/pub/release-85/fasta/homo\\_sapiens/dna/](ftp://ftp.ensembl.org/pub/release-85/fasta/homo_sapiens/dna/)) using Bowtie (ver. 2–2.2.2) and Tophat (ver. 2.0.11). For the individual cell lines, 90.3–92.3% of the filtered reads were mapped to the genome. The statistics of sequencing and mapping results are summarized in Table 1 and Table 2, respectively.

### 2.3. Expression analysis

Assembly of transcriptomes from RNA-Seq data and quantification of their expression levels were performed using Cufflinks (ver. 2.2.1). Fragments per kilobase of exon per million fragments mapped (FPKM) values are used as an index of expression level. Differential gene expression among the cell lines were analyzed using Cuffdiff (ver. 2.2.1).

### 2.4. Correlations of FPKM values per sample

Correlations of normalized FPKM values between the samples were assessed (Fig. 1). Using FPKM values of all annotated genes, the values of correlation coefficients between Het-1A and TE-1/5/8 are calculated to be 0.234 (Het-1A vs. TE-1), 0.323 (Het-1A vs. TE-5), and 0.401 (Het-

1A vs. TE-8). Correlation coefficients among the three ESCC-derived cell lines range from 0.419 to 0.697. These values possibly indicate that the individual ESCC-derived cell lines show a more similar gene expression pattern to each other than to Het-1A. When analysis objects are filtered with the criteria of FPKM > 1 or > 10 for both of paired samples, the values of correlation coefficients increase to 0.804–0.956.

### Conflict of interest

The authors declare that there are no competing interests.

### Acknowledgements

This study was supported by Grants-in-Aid (24791449) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT). The RNA-Seq experiment and subsequent bioinformatical analysis in this study were supported by Eurofins Genomics (Tokyo, Japan).

### References

- [1] T. Nishihira, Y. Hashimoto, M. Katayama, S. Mori, T. Kuroki, Molecular and cellular features of esophageal cancer cells. *J. Cancer Res. Clin. Oncol.* 119 (1993) 441–449.
- [2] G.D. Stoner, M.E. Kaighn, R.R. Reddel, J.H. Resau, D. Bowman, Z. Naito, N. Matsukura, M. You, A.J. Galati, C.C. Harris, Establishment and characterization of SV40 T-antigen immortalized human esophageal epithelial cells. *Cancer Res.* 51 (1) (1991) 365–371.