Identification of Alternative Splicing Characteristic Associated with Clear-Cell Ovarian Cancer from Paired Normal and Tumor Tissues

Publisher: IEEE

7 Author(s) Yu-Ting Huang ; Meng-Shin Shiao ; Chen-An Tasi ; Kuer-Yuan Lan ; Chieh-Hsi Li… All Authors

Abstract: Alternative splicing of messenger RNAs (mRNAs) is a common and conserved biological process in eukaryotes. The aberrancy or disruption of different alternative splicing f... **View more**

Metadata

Abstract:
Alternative splicing of messenger RNAs (mRNAs) is a common and conserved biological process in eukaryotes. The aberrancy or disruption of different alternative splicing forms may cause alterations of cell functions and result in diseases. It is proposed that alternative splicing may play a critical role in the mechanisms of carcinogenesis. By studying a large dataset in The Cancer Genome Atlas database, a recent study showed that alternative splicing, particularly exon-exclusion, is a powerful prognosis factor in serious subtype of ovarian cancer. However, the characteristics of alternative splicing has not been studied in other subtypes. In this study, we focus on the alternative splicing events, i.e. single exon-inclusion or -exclusion, in clear-cell ovarian cancer subtype. The subtype appears to have particularly high incidence in Asians comparing to Europeans and Americans and tend to be drug resistant. Transcriptomes were obtained from tumors and their paired-normal tissues from five patients. PSI-values, which represent the proportions of alternative splicing events of an exon, were calculated in both tumor and paired-normal tissues. Differences of PSI-values between tumors and paired-normal were examined by a significant test based on Conditional Beta Regression model. In total, we identified ~200 exons covering 52 genes with significant differences between cancer and paired-normal tissue ($p <; 0.001$) including gene ERP29 and PAM which were previously identified in serous subtypes.

Published in: 2019 International Conference on Machine Learning and Cybernetics (ICMLC)

Date of Conference: 7-10 July 2019

Date Added to IEEE *Xplore*: 06 January 2020

ISBN Information:

ISSN Information:

INSPEC Accession Number: 19278309

DOI: 10.1109/ICMLC48188.2019.8949316

Publisher: IEEE

Conference Location: Kobe, Japan, Japan

SECTION 1.
Introduction

Although advancing knowledge about ovarian cancer (OvCa) has been hindered by substantial disease heterogeneity and uncertainties about tumor tissues of origin, understanding of the disease has evolved rapidly in recent years. Most of the OvCa tumors have epithelial origins (90%) which can be further subdivided into type I (clear-cell, endometrioid, and mucinous tumors) and type II (high-grade serous) based on different molecular and histopathological signatures [1], [2]. In Asian countries, percentage of the most drug-resistant subtype, clear-cell, is significantly higher than the western countries which is estimated to be relatively rare (4-5%) in Europe and in the United States but common (20%) in Japan [3]. It is believed to be associated with the higher incidences of endometriosis [4]. In Thailand, clear-cell accounted for around 15% of all the cases (Ramathibodi Hospital Cancer Registry 2016 report). Interestingly, type I epithelial OvCa were found to be highly associated with the incidences of endometriosis in most of the cases while type II epithelial OvCa was recently proposed to be originated from the epithelium of the fimbriae of fallopian tubes [5].

Alternative splicing of mRNA increases the diversity of transcriptomes and proteomes of eukaryotes and it was proposed to be associated with various kinds of cancers from literature [6], [7]. Recently, alternative splicing has been studied in serous subtype of OvCA with a large set of data from The Cancer Genome Atlas (TCGA) database in order to understand whether it is associated with prognosis of the disease [8]. However, it has not been studied in the clear-cell subtype, which is the most abundant subtype in Asian populations.

In order to understand the characteristics and mechanisms of alternative splicing, particularly *single exon-inclusionor -exclusion*, in clear-cell subtype, we identified events from RNA-seq data of clear-cell OvCa tumors and their paired-normal tissues from five patients. We aim to identify exons/genes with significant differential exon utilization between tumors and paired-normal tissues.

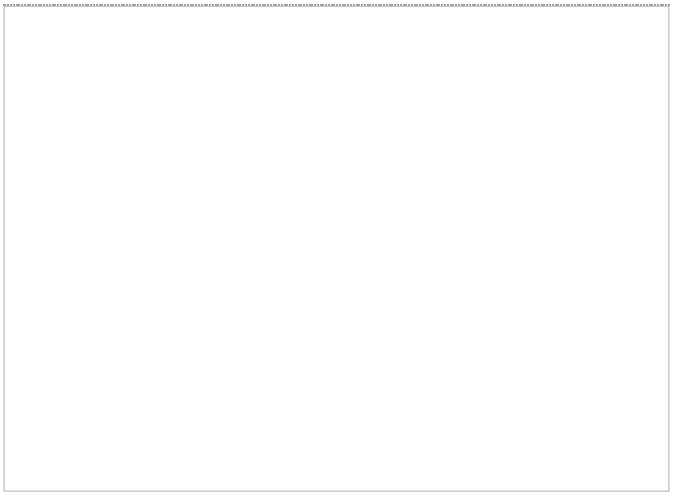
SECTION 2.
Materials and Methods

2.1. mRNA Samples

All clinical samples were collected, processed and sequenced by Faculty of Medicine Ramathibodi Hospital, Mahidol University, Thailand followed **IRB** protocols MURA2018/529. A total of 10 samples were collected from five patients with clear-cell OvCa. Tumors and their paired-normal tissues were both collected from each patient. For each sample, *5ug* of total RNA was used for library preparations. RNA-Seq libraries were constructed by Truseq RNA Sample preparation kit v2 (illumina) (cat# RS-122-2002). Sequencing was performed on Nextseq (patient 1 and 2) or HiSeq 4000 (patient 3–5). The file size of each

sample is around 1.5G to 2.1G which contains around 20~40 million 150~200 bp paired-end reads based on different sequencing platforms (Table 1, # of reads).

Table 1. Mapping results of each sample. Rows and columns represent samples and mapping metrics, respectively. id naming: patientid.case, N = normal sample, C = cancer sample. Uni.map refers to the proportion of reads mapped to one locus. Mulmapmeans reads mapped to multiple loci.



2.2. Mrna Mapping

Reference genome hg38 and gene annotation table was retrieved from UCSC for mapping. Spliced transcript alignment is conducted by STAR 2.5.1a [9] in two steps, where $sam indicates sample name:

1. Build genome index

   STAR −−runModegenomeGenerate −−

   genomeDirSTARgenomelndex −−

   genomeFastaFilesgenome.fa −−sjdbGTFfilegenes.gtf−−

   sjdbOverhang 100

2. Map reads based on the index

   STAR −−genomeDirSTARgenomelndex−−

   readFilesCommandzcat −−outSAMtype BAM

   SortedByCoordinate −−

   outFilterlntronMotifsRemoveNoncanonical −−twopassMode

   Basic −−runThreadN 8 −−readFilesln $saml.fastq.gz

   $sam2.fastq.gz−−outFileNamePrefix$sam

Finally, the expression of the transcriptome is quantified in the output file, *SJ.out.tab* [10]. The mapping step was done parallelly by submitting each sample to the TAIWANIA 1 cluster of the National High-Speed Network and Computing Center. Each job finished in half an hour with 48 GB memory.

2.3. Quantification of Alternative Splicing

Isoform preference is quantified in exon level by *percent spliced-in* (PSI/Ψ values which is the number of inclusion reads divided by the sum of inclusion and exclusion reads (two times) (Figure 2). Take gene *ERP29* as an example, two isoforms related to the exon 2 were identified (Figure 1) in which NM_006817 (isoform 1) includes exon 2 but NM_001034025 (isoform 2) excludes exon 2. The PSI of the exon 2 is calculated as the ratio between the number of inclusion reads (170+300)

and total reads (170+300+2*55), which equals to 0.81 (Figure 2). In another word, the higher PSI-values indicate the higher percentage of the exon-inclusion in the gene.

For each individual sample, we used *outrigger [11]* to create *de novo* alternative splicing annotation and quantified PSI of the events by the following three steps where $sam indicates sample name.

1. Build an index of splicing events using STAR's output junction reads and the annotation outrigger index ––output $outrigger_output$sam ––sj-out-tab sam_dirsam/SJ.out.tab ––gtf $genome/genes.gtf

2. Ensure that the splicing events found all have the correct splicing sites outrigger validate ––output $outrigger_output$sam ––genome hg38 ––fasta $genome/genome.fa

3. Calculate PSI values using the splicing event index built from Step1. outrigger PSI ––output $outrigger_output$sam

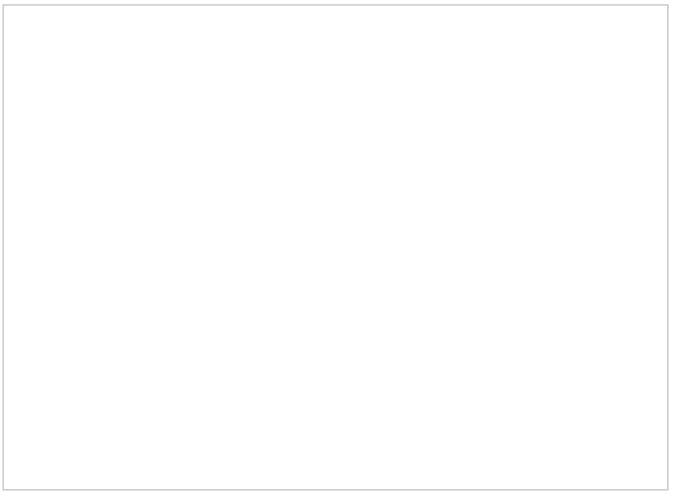


Figure 1.
Two transcriptions of ERP29 related with the second exon, in which NM_006817 (isoform1) includes exon2 but NM_001034025 (isoform2) excludes exon2.the formula of psi. The PSI of the exon 2 is 0.81 (= (170+300)/(170+300+2*55)).

2.4. Identification of Significant Differential PSI Between Clear-Cell Tumors and Paired-Normal Tissues

Candidate exons were excluded if their PSIs were missing in any of the sample. The differential PSI values of each pair of samples were then calculated. In order to consider overall differential PSI distribution, we used the R package *glmmTMB* to perform *Conditional Beta Regression model* for significant test.

SECTION 3.
Results

52 Exons with Significant Differential PSI

After mapping with the reference genome, each sample has about 90% unique mappable rate (Table 1, UniMap&MuIMap, detailed metrics in Supp. Table1). A total of 222 exons were identified with alternative splicing events by outrigger (the complete list in Supp. Table2). In the end, there are 52 exons showing significantly differential PSIs with p-value < 0.001 (the complete list in Supp. Table3). Among 52 exons, *EFEMP 1* shows highest differential PSI. Here, we demonstrate the mapping results of *EFEMP1* with IGV [12] (Figure 2).

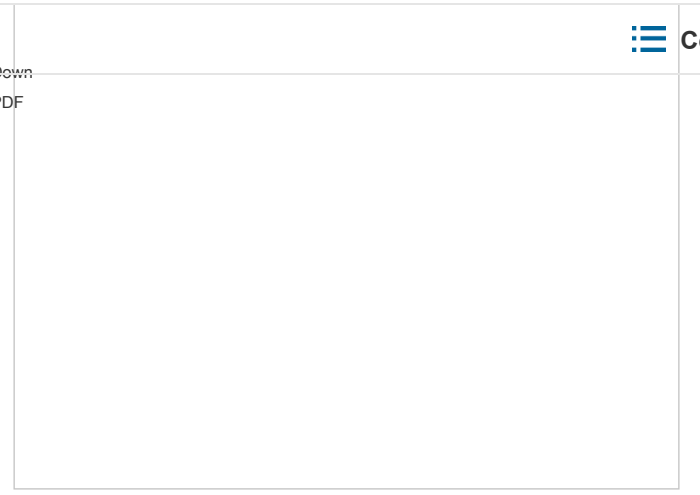


Figure 2.
The visualization of efemp1 read mapping by igv. The upper shows patient4.c in range 0~41 compared with the below, patient4.n, in range 0~219. It confirms that the exon of efemp1 tends to exclude in the cancer cell.

3.2. Representative Genes with Significant PSI Values

Zhu et al. (2018) analyzed 408 samples of high-grade serous subtype of OvCa from TCGA database. In this study, seven types of alternative splicing events were curated. Among the seven types, the authors proposed that exon-exclusion may be the most significant favorable prognostic factor (Hazard ratio (HR) < 1) in serous OvCa [8]. Although we don't have a large number of samples in this study, we have special sets of cancer vs. normal-paired samples which have not obtained in any other study. Therefore, the results may well represent specific characteristics of exon-inclusion and -exclusion in clear-cell subtype OvCa in Thai patients.

In Zhu et al. (2018), the authors listed top 20 genes showing exon-exclusion with the most significant association of prognosis in high-grade serous OvCa. Two of the genes identified in Zhu et al. (2018), i.e., *ERP29* and *PAM*, were also observed in our study. This indicates that the alternative splicing events of the two genes might be common in both subtypes and in different populations.

In addition, exon-exclusion of *ERP29* and *PAM* shows opposite trends in Zhu et al. (2018) [8]. Exon-exclusion in *ERP29* shows favorable prognosis (HR < 1, higher survival rate) while that in *PAM* shows less favorable prognosis (HR > 1, lower survival rate) in high-grade serous OvCa. In our study, PSI-values are lower in all the normal-paired samples in *ERP29 (*$p < 0.0001$) which indicates a lower incidences of exon-exclusion in cancer samples (Table 2). In contrast, PSI-values are higher in 4 out of 5 normal-paired samples in *PAM (*$p < 0.05$), which indicates a higher incidences of exon-exclusion in cancer tissues. These results show that the exon-inclusion and -exclusion patterns might be related with carcinogenesis differently in different genes.

Table 2. The commonly found gene set with our PSI analysis.

SECTION 4.
Conclusions

We have identified significant exon usage preferences of 52 genes regarding the subtype, clear-cell OvCa. Two of them have been reported in the literature, i.e. *ERP29* and *PAM*, in high-grade serous OvCa. In the future, we may further study the function of the alternative spliced transcripts showing significant differential PSI-values between tumors and paired-normal tissues. We hope the study will gain more insights into the mechanisms of alternative splicing in clear-cell OvCa subtype.

Supplemental Table: additional results in supplementary tables S 1-S3, available in https://figshare.com/s/6353cd2dlege5elea5d0.

ACKNOWLEDGMENTS

This paper is supported by the Taiwan Ministry of Science and Technology [106-2221-E-004-011-MY2 to J.-M.C.; 107-2813-C-004-048-E to C.-H.L.]. We acknowledge support “The Human Project from Mind, Brain and Learning” of NCCU from the Higher Education Sprout Project by the Ministry of Education in Taiwan. Weare grateful to the National Center for High-performance Computing for computer time and facilities.

Authors
Figures
References
Keywords
Metrics

Contents

Down

PDF