Oral Oncology 49 (2013) 604–610

Contents lists available at SciVerse ScienceDirect

Oral Oncology

journal homepage: www.elsevier.com/locate/oraloncology

The landscape of alternative splicing in buccal mucosa squamous cell carcinoma

Tejas M. Shah [a], Amrutlal K. Patel [a], Vaibhav D. Bhatt [a], Ajai K. Tripathi [a], Siddharth Shah [b], Vangipuram Shankar [b], Chaitanya G. Joshi [a,*]

[a] Department of Animal Biotechnology, College of Veterinary Science & Animal Husbandry, Anand Agricultural University, Anand-388 001, Gujarat, India
[b] MS Patel Cancer Center, Pramukhswami Medical College, Karamsad, Anand-388 001, Gujarat, India

ARTICLE INFO

Article history:
Received 15 September 2012
Received in revised form 6 March 2013
Accepted 9 March 2013
Available online 6 April 2013

Keywords:
BMSCC cancer
Transcriptome
Alternative splicing (AS)
RT-PCR
Pseudogene
lncRNA

SUMMARY

Objectives: Alternative splicing (AS) is a key regulatory mechanism in the process of protein synthesis generating transcriptome and proteome diversity. In this study, we attempted to identify alternative splicing in a pair of BMSCC cancer and adjacent normal tissue using RNAseq datasets and also assessed the potential of these datasets to provide quantitative measurements for alternative splicing levels.
Materials and methods: We performed high-throughput sequencing of buccal mucosal cancer and healthy tissue cDNA library which resulted in a transcriptome map of BMSCC cancer. RNAseq analysis was performed to assess alternative splicing complexity in cancer tissue and to search splice junction sequences that represent candidate 'new' splicing events. The splice junctions were predicted by SpliceMap software and putative assembled transcripts validated using the RT-PCR. We also analyzed the coding potential of alternative spliced candidate by HMMER.
Results: We detected a total of 11 novel splice junctions derived mostly from alternate 5' splice site; including two of them which contained new translation initiation sites (TISs). We have identified novel IgG pseudogene and a fusion transcript of MEMO1 and RPS9, which were further confirmed by PCR from genomic DNA. We also found novel putative long non-coding RNA (lncRNA), which is antisense to SPINK5 gene. The coding potential of these AS variants revealed that alternative splicing caused premature termination, insertion/deletion of amino acid (s) or formation of novel N-terminus.
Conclusions: Differential splicing of these novel AS variants between cancer and adjacent normal tissue suggests their involvement in BMSCC cancer development and progression.

Introduction

Head and neck squamous cell carcinoma (HNSCC), including oral squamous cell carcinoma (OSCC), is the sixth most common cancers in men, accounting for approximately 6% of all incident cancers.[1] OSCC is the most common neoplasia among HNSCC and is found frequently at the sites such as cheek, gum, and tongue in oral cavity.[2] Although cigarette and alcohol are considered as two major risk factors of oral carcinogenesis[3], occurrence of oral cancer was proven to be tightly associated with tobacco and betel quid chewing in India and South Asia.[4] Buccal mucosa squamous cell carcinoma (BMSCC) is a rare but aggressive form of OSCC, associated with high rate of loco-regional recurrence and poor survival.[5] Molecular profiles of oral cancers vary throughout the world and are influenced by both aetiological factors and ethnicity.[6] Most genome-wide studies on OSCC have been carried out on various intra-oral sites that are associated with different aetiological agents.

Each regulatory point in the control of gene expression is subject to profound alterations during the development of most, if not all, cancers[7–10], but none of them provides the potential for more diverse outcomes than alternative splicing (AS), which is considered to be a key factor underlying increased cellular and functional complexity in higher eukaryotes.[11–13] AS, the alternative selection of splice sites present within a pre-mRNA, leads to the production of different mRNA isoforms from a single gene and thus alter the composition and function of the encoded protein. The plasticity offered by AS to remodel the proteome provides opportunities for cancer cells to subvert the processes to produce proteins that leads to the growth and metastasis of cancer. All kinds of cellular processes in cancer cell appear to be affected by changes in AS, including metabolism, apoptosis, cell cycle control, invasion, and metastasis, as well as angiogenesis.[14,15] Genome-wide approaches have revealed that tumorigenesis often involves large-scale alterations in AS.[16] Such approaches have been valuable in providing insight into the regulation of splicing in cancer, and have even proven useful in the classification of tumors.[10,17,18]

However, because of the limited depth of coverage and sensitivity afforded by conventional sequencing and microarray profiling

* Corresponding author. Tel./fax: +91 2692 261201.
E-mail address: cgjoshi@rediffmail.com (C.G. Joshi).

1368-8375/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.oraloncology.2013.03.431

methods, the extent of alternative splicing is not known.[19] High-throughput or 'next generation' sequencing technologies offer the potential to address this question[20], and several recent studies have applied analyses of short cDNA read (mRNAseq) data from these technologies to survey alternative splicing in various cancers.[21–24]

Recently, we have studied differential transcriptome profiling in buccal cancer and normal tissues by RNAseq.[25] In this study, we used Roche 454 pyrosequencing data to survey alternative splicing in a pair of BMSCC cancer and adjacent normal tissue using RNAseq datasets and also assessed the potential of these datasets to provide quantitative measurements for alternative splicing levels.

Materials and methods

To assess alternative splicing complexity in cancer tissue, we used mRNAseq datasets of our previous study[25] on BMSCC, to search splice junction sequences that represent candidate 'new' splicing events. Informed written consent was obtained from participant of the study.

Raw reads filtering

The criteria used to filter raw reads were: Remove reads with sequence adaptors; Remove reads with more than 8% 'N' bases; Remove low quality reads, which have quality score <20. All subsequent analysis was carried out on these good quality reads in FASTQ format.

Read alignment to the reference human genome

A total of 104867 and 128369 single-end reads were obtained from BMSCC cancer and adjacent normal tissues by 454 GS-FLX sequencing, respectively. Read mapping to the human reference genome build GRCh37/hg19 and splice junction detection was done using SpliceMap[26] with default parameters.

Splice junction detection

We used SpliceMap software[26] to predict splice junctions. We filtered the splice junctions originally detected (using alignment with two mismatches) by two criteria: quality of the alignment and coverage of splice junction. However, the coverage of RNAseq data was not enough to include only those junctions supported by at least two reads, because majority of the novel splice junctions detected were supported by only one read. To diminish the number of potential false positives, we performed prediction of splice junctions without mismatches in addition to the original alignment of two mismatches. The above approach allowed us to discard splice junctions that were predicted by erroneous alignment, especially in regions where several di-nucleotides resembling splice sites were found in tandem. We filtered out those splice junctions that were not predicted in the alignment of no mismatches and that were in proximity with another better supported junction by <10 nt. The remaining splice junctions were used for further analysis if they were supported by at least two alignments in the original list (alignment with two mismatches).[27]

We have defined as genic splice junctions as those junctions which were inside the coordinates of annotated genes in human genome build GRCh37/ hg19 and have the same strand as the gene in question. Those splice junctions which were inside gene coordinates but were in the opposite strand were named as antisense. Finally, those junctions which were not inside any known gene coordinates were called intergenic.[28]

Validation of putative assembled transcripts using the RT-PCR

Total RNA was isolated using an RNeasy Kit (Invitrogen) with DNase I digestion according to the manufacturer's instructions. RNA integrity was verified on a Bioanalyzer 2100 (Agilent Technologies). cDNA was synthesized from total RNA using Superscript III (Invitrogen) and random primers (Invitrogen). The oligonucleotide primers used for novel alternative splice event are listed in Table 3. PCR was performed using 1X PCR Master Mix (Fermentas) on GeneAmp® PCR System 9700 (Applied Biosystem), with cycling parameters: 95 °C for 10 min followed by 35 cycles each of 95 °C for 45 s, annealing at 55 or 60 °C for 45 s, and 72 °C for 45 s, and a final extension at 72 °C for 10 min and PCR products were resolved on a 1.5% agarose gel. Quantitative Real-time PCR (qPCR) was performed using SYBR Green Mastermix (Applied Biosystems) on a 7500 FAST Real-Time PCR System (Applied Biosystems) to validate the expression of novel alternative splicing event.

Analysis of coding potential of alternative spliced candidate

We have analyzed the coding potential of alternative spliced candidate. The putative transcripts with novel splice junctions were derived by incorporating novel splice junctions after alignment with genomic and reference mRNA transcript sequences. Translation was performed using the frame of the neighboring exon annotated in human reference genome build GRCh37/hg19. The resulting translated sequences were used to search Pfam protein domains employing Hidden Markov models (HMMs) with HMMER (Pfam HMM database release 26.0).

Results

Single-end reads were aligned to the human reference genome buildGRCh37/hg19 (UCSC) using Bowtie[29], and included in the SpliceMap software v3.3.5.2.[26] We mapped nearly 75% and 74% of the quality trimmed reads to the reference genome from cancer and normal sample, respectively.

Detection level of novel splice- junctions in cancer

For splice junction detection, we used SpliceMap.[26] Our results showed that the majority of the splice junctions reside in annotated genes (1517 and 917 in cancer and adjacent normal sample, respectively) and, in protein-coding genes, highlighting the potential of splice junctions to affect the resulting protein sequence. From the total of predicted splice junctions, 13 (0.99%) were novel, not present in annotated hg19 Database and adjacent normal sample (Table 1). However, two junctions were excluded from the further analysis as their supported reads were not matched with junctions predicted by SpliceMap as confirmed by BLAT analysis. One novel junction with alternative 3′ splice site in IMPA2 was found only in cancer and supported by two reads (Fig. 1), whereas other novel splice junctions were supported by only one read (Table 1). Novel splice junctions present only in cancer were analyzed on UCSC genome browser of human genome build GRCh37/hg19 via custom track, which consisted five 5′ AS, three exon skipping, two3′ AS, one 5′ as well as 3′ AS, one 5′ as well as 3′ AS with exon skipping and one 3′ AS with exon skipping (Fig. S1 and Table 1). One AS variant was found antisense to SPINK5 where none of the gene/transcript was annotated for the respective chromosome coordinate on antisense to SPINK5 in hg19 genome. A total of 11 (two excluded) AS variants identified in this study, of which nine variants matched with spliced EST database, whereas four variants were novel as they did not match with spliced EST database. AS with intron retention was not detected in any event. We found that use of alternate 5′ splice sites is more than the use of alternate 3′