Analysis of alternative splicing events by RNA-seq data in glioblastoma: model cases

Vladimir Babenko, Anatoly Bragin, Irina Medvedeva, Irina Chadaeva, Yuri Orlov

Diffuse gliomas are the most common type of intracranial malignant neoplasm, and account for >60% of all primary brain tumors [1]. Based on the classification of nervous system tumors by the World Health Organization, diffuse gliomas are classified into seven principal categories: Diffuse astrocytoma (grade II), oligodendroglioma (grade II), oligoastrocytoma (grade II), anaplastic astrocytoma, anaplastic oligodendroglioma and oligoastrocytoma (grade III) and glioblastoma (GBM, grade IV) [1]. Of these, GBM is the most common and aggressive type of primary brain tumor, accounting for 80% of malignant astrocytomas. GBM may develop rapidly without the diagnosis of a less malignant precursor lesion, and this is termed primary or de novo GBM. It may also develop slowly through progression from a pre-existing low-grade glioma, in which case it is termed a secondary GBM. Despite advances in neurosurgery, chemotherapy and radiotherapy, glioma commonly has a poor prognosis. Therefore, it is critical that the genetic pathways underlying the development of this type of cancer are defined.

In the current study, the gene expression profiles of GBM were obtained on cell culture samples. The cell culture samples of primary glioblastoma were isolated following [2] and processed for RNA extraction. The significant differentially expressed exon-level probes and their corresponding genes were identified using a combination of the splicing index (SI) method. A previous study indicated that tumor-specific alternative splicing is important in the regulation of gene expression and corresponding protein functions during cancer development [3]. Multiple alternative splicing transcripts have been identified as progression markers, including generalized splicing abnormalities and tumor- and stage-specific events.

We used set of computer tools applied recently to analysis of gene expression in laboratory animals [4]. RNA-seq sequencing of cell samples was done using Illumina HiSeq 1500. RNA quality was tested on Bioanalyzer 2100 (Agilent) following the Illumina protocol. Sequencing depth consisted of at least 10 mol reads for each sample. The files obtained in “fastq” format were used for the reads mapping onto the human reference genome using Tophat2 aligner. Input reads were filtered and trimmed by Trimmomatic software [5]. Trimmomatic omits technical sequences and sequences with low quality from input data. We use Tophat2 program [6] for mapping reads on the rodent reference genomes. Tophat2 is built on ultrafast, memory-efficient program Bowtie [7] and can identify splice junctions. Files with genome annotations were downloaded from UCSC genome browser. The bam-files - received from Tophat output - were used for detection of differentially expressed genes and alternative splicing analysis in the samples. We have analysed gene expression by Cufflinks v2.0.2 programs [6]. Cufflinks tool provides information about differently expressed genes between the

samples. Gene expression levels were estimated in FPKM (fragments per kilobase per million reads) values. The detection of splice variants and differential splicing was performed using rMATS software [8].

To analyze alternative splicing events in the transcriptomics data MATS (multivariate analysis of transcript splicing) and rMATS (replicate MATS) shown to be effective tools [9]. We implemented a Python-based application to tackle rMATs annotation. The rMATs output lacks the information on the short isoforms spectra of particular exon. We implemented the add-on that, based on 3 subsequent exons locations used in alternative exons identification, retrieves from RefGene database short isoforms IDs lacking the exon. While it’s not possible to relate exon skipping event to the particular full length isoform due to the short reads length, we can still speculate on tissue specificity of the exon skipping events. rMATS program [8 - Shen et al., 2014] was used to assess the difference in alternative splicing profiles between Normal brain (NB) tissue (cortex) and secondary glioblastoma (GBM) tissue.

We identified 69 gene loci with highly significant differential isoforms expression. The total number of significant alternative splicing (Exon Skipping) events comprised 107 instances with FDR<0.01 (160 with FDR<0.05). The major GO entries were cytoskeleton and intracellular (cytoplasmic) related genes. We found also genes of nuclear pore complex as differentially expressed in NGB cell culture sample.