RNA-Seq profiling reveals aberrant RNA splicing in patient with adult acute myeloid leukemia during treatment

X.-Y. LI, X. YAO¹, S.-N. LI², A.-L. SUO², Z.-P. RUAN², X. LIANG², Y. KONG², W.-G. ZHANG³, Y. YAO²

Department of Orthopaedic Surgery, The First Affiliated Hospital of Xi’an Jiaotong University, Xi’an 710061, Shaanxi, China; ¹Department of General Surgery, The Center Hospital of Xi’an, Xi’an 710061, Shaanxi, China; ²Department of Oncology, The First Affiliated Hospital of Xi’an Jiaotong University, Xi’an 710061, Shaanxi, China; ³Department of Hematology, The Second Affiliated Hospital of Xi’an Jiaotong University, Xi’an 710061, Shaanxi, China

Abstract. – BACKGROUND: Multiple genetic alterations that affect the process of acute myeloid leukemia (AML) have been discovered, and more evidence also indicates that aberrant splicing plays an important role in cancer.

MATERIALS AND METHODS: We present a RNA-Seq profiling of an AML patient with complete remission after treatment, to analyze the aberrant splicing of genes during treatment. We sequenced 3.97 and 3.32 Gbp clean data of the AML and remission sample, respectively. Firstly, by analyzing biomarkers associated with AML, to assist normal clinical tests, we confirmed that the patient was anormal karyotype, with NPM1 and IDH2 mutations and deregulation patterns of related genes, such as BAALC, ERG, MN1 and HOX family. Then, we performed alternative splicing detection of the AML and remission sample.

RESULTS: We detected 91 differentially splicing events in 68 differentially splicing genes (DSGs) by mixture of isoforms (MISO). Considerig Psi values (Ψ) and confidence intervals, 25 differentially expressed isoforms were identified as more confident isoforms, which were associated with RNA processing, cellular macromolecule catabolic process and DNA binding according to GO enrichment analysis. An exon2-skipping event in oncogene FOS were detected and validated in this study. FOS has a critical function in regulating cell proliferation, differentiation and transformation. The exon2-skipping isoform of FOS was increased significantly after treatment.

CONCLUSIONS: All the data and information of RNA-Seq provides highly accurate and comprehensive supplements to conventional clinical tests of AML. Moreover, the splicing aberrations would be another source for biomarker and even therapeutic target discovery. More information of splicing may also assist the better understanding of leukemogenesis.

Key Words: RNA-Seq, AML, Splicing, Biomarkers.

Introduction

Acute myeloid leukemia (AML) is one kind of aggressive clonal hematopoietic disorders, with the differentiation of hematopoietic progenitor cells blocked and the normal regulation of proliferation disturbed.

Inherited and acquired gene mutations and expression dysregulations have been identified that affect the process of AML, especially in the large group (40-50%) of cytogenetically normal (CN) AML. Through the karyotyping of AML cells is routinely used in clinical for predicting the outcome in AML patients in clinical practice, much more genetic markers have also entered clinical practice for stratification of patients and risk assessment. For example, mutations identified in the nucleophosmin 1 (NPM1), the FMS-related tyrosine kinase 3 (FLT3) and the CCAAT/enhancer binding protein α gene (CEPBA) are used as prognostic implications and for clinical guidance of targeted therapy. And the dysregulations of the brain and acute leukemia, cytoplasmic (BAALC), the meningioma 1(MNI) and the v-ets erythroblastosis virus E26 oncogene homolog, avian (ERG) also affect the complete remission (CR) rate and overall survival (OS).

Besides of genetic aberrations described above, alternative RNA splicing (AS) also plays a critical role in pre-mRNA processing during cell proliferation and differentiation. Most studies have focused on AS of individual genes in AML patients, such as AML1, GFI1B. And some of them can be used as prognostic or diagnostic biomarkers in other cancers.
In this study, we performed RNA-Seq to evaluate genome-wide splicing events of blood samples from a patient who had AML with maturation (AML-M2) before and after treatment with complete remission. We sequenced 3.97 and 3.32 Gb clean data of the AML and remission sample, respectively. We firstly confirmed the karyotype and genotype of the patient, as well as gene expression patterns, by colligating the results of transcriptome analysis and clinical testing. Then, splicing events were detected by MISO (mixture of isoforms)\textsuperscript{13}. Differentially expressed isoforms were further identified according to Bayes factors and other filtering criteria, when comparing the AML and remission samples. Interested isoforms were validated by RT-Q-PCR. Overall, our study identified genome-wide splicing events in AML patients during treatment. More information of splicing may also assist the better understanding of leukemogenesis. And the splicing aberrations would be another source to explore the potential biomarkers which can be used in clinical and personal medicine.

**Materials and Methods**

**Case Information**

The AML sample from peripheral blood was collected from a 47-year-old patient, diagnosed with AML M2 in March 2012. The patient was informed consent and ethical approval for scientific use of the sample including genetic studies was obtained. After induction therapy using the fludarabine and high-dose cytarabine and G-CSF (FLAG) protocol, complete remission was achieved. After leukocyte recovery in April 2012, a remission sample from peripheral blood was taken.

**Sample Preparation**

Total RNA was extracted from blood samples by Trizol (Invitrogen, Carlsband, CA, USA, Cat. No.: 15596-026). The sequencing libraries were prepared using TruSeq RNA Sample Prep Kit (Illumine, San Diego, CA, USA, Cat. No.: FC-122-1001). Libraries were sequenced on an Illumina Genome Analyzer IIx with 115-bp pair-end read length.

**Expression Analysis**

Reads were aligned to the Hg19 genome assembly and transcripts from the Ensembl database using TopHat v.1.4.0\textsuperscript{14} and Bowtie v.0.12.7\textsuperscript{19}. The result of read mapping was visualized by IGV v.2.0.26\textsuperscript{16, 17}. Transcript abundance and differential gene expression were calculated by Cufflinks v.1.3.0\textsuperscript{14}. Gene expression values were normalized using fragments per kilobase of exon per million mapped reads (FPKM). With comparison of AML and remission samples, differential expressed genes (DEGs) were identified with the threshold of their q-value < 0.05. Up- or down regulation of DEGs were determined by fold change. GO and KEGG pathway annotation and enrichment analysis were performed by DAVID\textsuperscript{18}. The enriched GO categories and KEGG pathways were identified with a false discovery rate (FDR) of 5%. The results of enrichment analysis were visualized by AmiGO.

**SNV Calling**

Read alignment were performed using the Burrows-Wheeler Aligner (BWA)\textsuperscript{19}, with the reference of hg19 genome and extension exon dataset. Raw Single-Nucleotide Polymorphisms (SNPs) were called by Sequence Alignment/Map (SAM) tools\textsuperscript{20} and Single-Nucleotide Variants (SNVs) were identified by VarScan\textsuperscript{21} by comparison of the AML and remission samples. Functional analysis of SNVs was performed with the Variant Effect Predictor (VEP) tool from Ensembl.

**Alternative Splicing Event Analysis**

Alternative splicing events were detected by MISO (The Mixture of Isoforms)\textsuperscript{13}. Differentially expressed isoforms were further identified according to Bayes factors calculated by MISO and filtering criteria, Psi values (Ψ) and confidence intervals, when comparing the AML and remission samples.

**Alternative Splicing Validation**

The primers were designed using Primer 5 software (PREMIER Biosoft International, Palo Alto, CA, USA). Primers were used at a final concentration of 330 nM using Eurogentec qPCR mastermix (Eurogentec, Seraing, Belgium) in a total volume of 15 ul, using the following thermal profile: 95°C for 10 min followed by 40 cycles of 95°C for 10 s, 62°C for 1 min and 72°C for 30 s. All RT-Q-PCR experiments were performed using a StepOnePlus thermocycler (Applied Biosystems, Darmstadt, Germany).
**Results**

**RNA-Seq Analysis Released the Differences Between AML and Remission Sample**

We sequenced 3.97 and 3.32 Gbp clean data of the AML and remission sample, respectively, on an Illumina GA IIx sequencer (Illumina). The Hg19 genome assembly and transcripts from the Ensembl database were used as reference sequence. Read mapping to the reference genome was performed with the TopHat software. Approximately 25.06M (72.57%) and 20.58M (71.06%) reads were mapped to the reference in the AML and remission sample, respectively. And the average coverage was greater than 20 in both samples (Table I).

Gene expression analysis was performed with Cufflinks, and the FPKM (fragments per kilobase of exon per million) value of each gene was calculated. More than 19,000 genes were detected in both samples and FPKM values of the AML and remission samples ranged from 0 to 533,518 and 138,131, respectively (Table I). The evenness and integrity of read distribution were evaluated. And the correlation of the gene expression levels between the samples as shown by a correlation coefficient of 0.7575.

The pairwise comparison of the AML and remission samples resulted in the identification of a total of 4,148 DEGs; 68% of them were down regulated and 32% were up-regulated. GO and KEGG pathway enrichment analysis were performed by DAVID\(^ {18} \), and DEGs were categorized in 140 GO terms and enriched in 26 KEGG pathways, and most of them were related to immune response and signal transduction.

Single-nucleotide variants (SNVs) were detected by the VarScan\(^ {21} \), 29,881 SNPs were detected in the first place. Quality filtering resulted in a set of 752 somatic variations, while 97 out of them were identified as missense mutations.

**Further Analysis Assisted Conventional Clinical Tests**

Conventional cytogenetic analysis revealed that 3 out 11 cells presented a susceptible abnormal karyotype (46, XX, 5p(-)?[3]). But according to the analysis of RNA-Seq data, there were no significant difference of gene expression level in chr5p, which suggested the normal karyotype of the patient (Figure 1).

For mutation testing, mutations in FLT3/ITD and CEBPA, point mutation of C-KIT D816V, and gene fusion of BCR and ABL were excluded in diagnostic tests, a mutation in NPM1 gene was detected. In the RNA-Seq results, a deletion in exon 12 of the gene NPM1 were found in both AML and remission samples, while no other mutations were detected in genes previously tested in diagnosis. Besides, IDH2R140 mutation was found in tumor sample. And no other mutations of AML related genes were found in this study, including IDH1, WT1, RUNX1, MLL, NRAS, TP53, TET2 and ASXL1\(^ {11} \). Besides genetic mutations, expression patterns of specific genes were found mostly deregulated (Table SI), such as HOX cluster genes (a homeodomain-containing family of transcription factors).

**Alternative Splicing Analysis and Validation**

To identify the differences of transcript isoforms by alternative splicing in AML and remission samples, we employed the software MISO. Approximately 50,193 and 38,344 alternative splicing events were detected in the AML and remission sample, respectively (Table II).

Furthermore, we identified 91 differentially splicing events in 68 differentially splicing genes (DSGs) by performing the comparison of the two samples. According to the sashimi plots, we estimated the more confident differentially expressed isoforms with the difference of Psi values (Ψ) and narrower confidence intervals. Finally 25 isoforms (Table SII) were selected to apply the further analysis. With GO annotation, most of

---

**Table I. Statistics of the reads obtained and their mapping on HG19 genome.**

<table>
<thead>
<tr>
<th>Samples</th>
<th>Total reads (M)</th>
<th>Total nucleotides</th>
<th>Mapped reads (M)</th>
<th>Mapped rate (%)</th>
<th>Coverage</th>
<th>GC (%)</th>
<th>Q20 (%)</th>
<th>N (%)</th>
<th>Expressed genes</th>
<th>Min FPKM value</th>
<th>Max FPKM value</th>
</tr>
</thead>
</table>
RN A-Seq Reveals RNA Splicing in AML  

Figure 1. The Identification of karyotype based on expression analysis of RNA-Seq. The gene expression level of chrY shows no expression in AML and remission samples, while the expression of chr5p doesn’t show significant difference when comparing with the T1 and R1 control. The region of chromosome zoomed is in the red frame. And the gene expression level below is shown with the blue bar, respectively.

them were associated with RNA processing, cellular macromolecule catabolic process and DNA binding.

An exon2-skipping event in oncogene FOS (FBJ murine osteosarcoma viral oncogene homolog) was detected (Figure 2). FOS has a critical function in regulating cell proliferation, differentiation and transformation. We further performed RT-Q-PCR to validate the expression of FOS isoforms in both AML and remission samples. The primers we designed were listed in Table IIIA, and the expression levels of isoforms in patients were shown in Table IIIB.
Table II. The statistic of alternative splicing events detected in the AML and complete remission (CR) samples.

<table>
<thead>
<tr>
<th>Sample</th>
<th>SE</th>
<th>AS5S</th>
<th>AFE</th>
<th>MXE</th>
<th>RI</th>
<th>TUTR</th>
<th>ALE</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS events</td>
<td>AML</td>
<td>18,584</td>
<td>5,113</td>
<td>8,647</td>
<td>1,222</td>
<td>2,619</td>
<td>1,661</td>
<td>6,404</td>
</tr>
<tr>
<td></td>
<td>CR</td>
<td>13,372</td>
<td>3,462</td>
<td>4,612</td>
<td>6,794</td>
<td>908</td>
<td>2,032</td>
<td>1,409</td>
</tr>
<tr>
<td>Comparable events</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AML vs. CR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Differentially splicing</td>
<td>AML vs. CR</td>
<td>8</td>
<td>0</td>
<td>2</td>
<td>30</td>
<td>1</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Differentially splicing</td>
<td>AML vs. CR</td>
<td>8</td>
<td>0</td>
<td>2</td>
<td>15</td>
<td>1</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Filtered by Psi and</td>
<td>AML vs. CR</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>confidence intervals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Discussion

Our study provides a comprehensive insight into transcriptome changes of the AML-M2 patient during the treatment. The NGS approaches have been proved powerful for detecting genetic alterations and chromosomal lesions in academic and clinical studies. RNA-Seq gathered more detailed and precise information than the conventional diagnostic tests. We firstly confirmed that...
RNA-Seq Reveals RNA Splicing in AML

Table III. RT-Q-PCR validation for the isoforms of FOS.

<table>
<thead>
<tr>
<th>AML</th>
<th>CR</th>
<th>Expression ratio</th>
<th>Fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>WL_Ct_mean</td>
<td>E2S_Ct_mean</td>
<td>WL_Ct_mean</td>
</tr>
<tr>
<td>Sample 1</td>
<td>23.257</td>
<td>23.450</td>
<td>19.443</td>
</tr>
<tr>
<td>Sample 2</td>
<td>23.553</td>
<td>23.723</td>
<td>18.637</td>
</tr>
</tbody>
</table>

*WL, whole length transcript; E2S, exon2-skipping transcript; Expression ratio, expression of WL/expression of E2S; Fold change, expression ratio in CR/expression ratio in AML.

Conclusions

Our findings suggest that splicing aberrations would be another source for biomarker and even therapeutic target discovery. And further studies are still undergoing to evaluate the interested alternative splicing in AML.

Acknowledgements

This study was supported by the Science and Technology Research Foundation of Shaanxi Province, P.R. China (No.2013K12-03-12).

Conflict of Interest

The Authors declare that there are no conflicts of interest.
References


