Genomics Data 2 (2014) 20–23

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

Genome-wide analysis of SRSF10-regulated alternative splicing by deep sequencing of chicken transcriptome

Xuexia Zhou [1], Wenwu Wu [1], Ning Wei, Yuanming Cheng, Zhiqin Xie, Ying Feng *

Key Laboratory of Food Safety Research, Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

ARTICLE INFO

Article history:
Received 12 February 2014
Received in revised form 20 February 2014
Accepted 20 February 2014
Available online 12 March 2014

Keywords:
Alternative splicing
SRSF10
RNA-seq
Bioinformatics

ABSTRACT

Splicing factor SRSF10 is known to function as a sequence-specific splicing activator that is capable of regulating alternative splicing both in vitro and in vivo. We recently used an RNA-seq approach coupled with bioinformatics analysis to identify the extensive splicing network regulated by SRSF10 in chicken cells. We found that SRSF10 promoted both exon inclusion and exclusion. Functionally, many of the SRSF10-verified alternative exons are linked to pathways of response to external stimulus. Here we describe in detail the experimental design, bioinformatics analysis and GO/pathway enrichment analysis of SRSF10-regulated genes to correspond with our data in the Gene Expression Omnibus with accession number GSE53354. Our data thus provide a resource for studying regulation of alternative splicing in vivo that underlines biological functions of splicing regulatory proteins in cells.

Specifications	
Organism/cell line/tissue	*Gallus gallus*/DT40 cell line
Sequencer or array type	Illumina HiSeq™ 2000 system
Data format	Raw data: fastq files, analyzed data: txt file
Experimental factors	Wild type (WT) vs. SRSF10-deficient (KO) chicken DT40 cells
Experimental features	RNA-seq and bioinformatics analysis
Sample source location	Shanghai 200031, China

Direct link to deposited data

Deposited data can be found here: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE53354.

Experimental design, materials and methods

Sample preparation and RNA-seq analysis

It is known that phosphorylated SRSF10 functions as a sequence-specific splicing activator that is capable of regulating alternative splicing both in vitro and in vivo [1,2]. To understand the global profile of the alternative splicing network regulated by SRSF10 in vivo, we carried out an RNA-seq approach to analyze the effects of SRSF10 depletion on splicing regulation in chicken DT40 cells. The general strategy for sample preparation for RNA-seq was provided in Fig. 1A and the detailed experimental procedures were as follows. Wild type (WT) and SRSF10-knockout (KO) DT40 cells were cultured in a standard medium to 90% confluence. Total RNA was isolated from these cells, and cDNA libraries were prepared individually by performing a series of experiments, including poly(A) enrichment, RNA fragmentation, random hexamer-primed cDNA synthesis, linker ligation, size selection and PCR amplification. After selecting approximately 200-bp fragments by agarose gel electrophoresis, Illumina paired-end sequencing adapters were ligated to the DNA fragments for PCR amplification and sequencing with an Illumina HiSeq™ 2000 system.

The bioinformatics analysis pipeline for RNA-seq data was provided in Fig. 1B. Briefly, images generated by sequencers were converted by base calling into nucleotide sequences, which were called raw reads. Then raw reads were trimmed for adaptor sequence, and masked for low-complexity or low-quality sequence to get clean reads with high quality. The clean reads were then aligned to the reference chicken genome (release: galGal4) and the corresponding recently annotated genes (galGal4.0) using the TopHat program (v2.0.6) [3]. In the alignment, preliminary experiments were performed to optimize the alignment parameters to provide the most information on the alignment events. Cufflinks [4] was used to construct KO and WT DT40 transcriptomes and Cuffmerge merges the transcriptomes with the reference gene annotations (galGal4.0) to generate a new comprehensive transcriptome. Finally, a Java program called ASD (alternative splicing detector, available at http://www.novelbio.com/asd/ASD.html) was developed to detect and assess alternative splicing events between the two samples [5].

* Corresponding author. Tel./fax: +86 21 54920965.
E-mail address: fengying@sibs.ac.cn (Y. Feng).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.gdata.2014.02.001

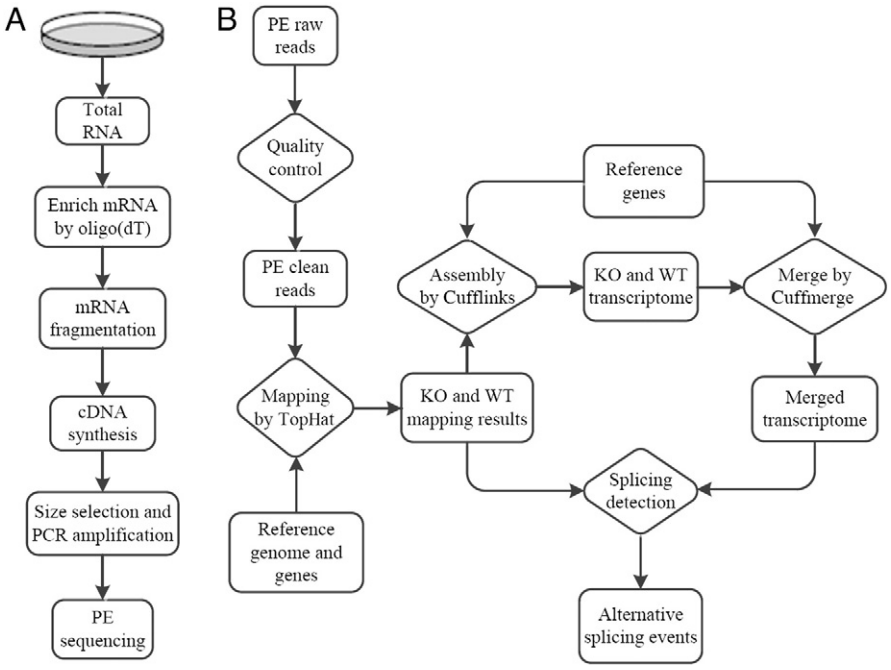

Fig. 1. Experimental and bioinformatics pipelines. (A) General strategy for RNA sequencing. Total RNA was isolated from wild type (WT) and SRSF10-deficient (KO) DT40 cells and then cDNA libraries were constructed for paired-end sequencing. (B) Bioinformatics analysis pipelines. TopHat was used to map clean reads to the Gal4 reference genome and transcriptome. Cufflinks and Cuffmerge were then used to construct and merge DT40 transcriptomes. The merged transcriptome and the read counts in exonic reads and junction reads were then further analyzed using our Java program ASD to identify differential alternative splicing events between WT and KO samples.

SRSF10 both activates and represses exon inclusion

There are five major modes of alternative splicing (AS) events described in metazoan organisms, including cassette exon, alternative 3′ splice site, alternative 5′ splice site, mutually exclusive exon and retained intron (Fig. 2A). With a p value cutoff of <0.05, we were able to identify 167 events belonging to the five major modes of AS from a total number of 10,652 splicing events that changed significantly between the WT and the KO samples by ASD. Specifically, the majority of the affected splicing events (130) belonged to the cassette exon

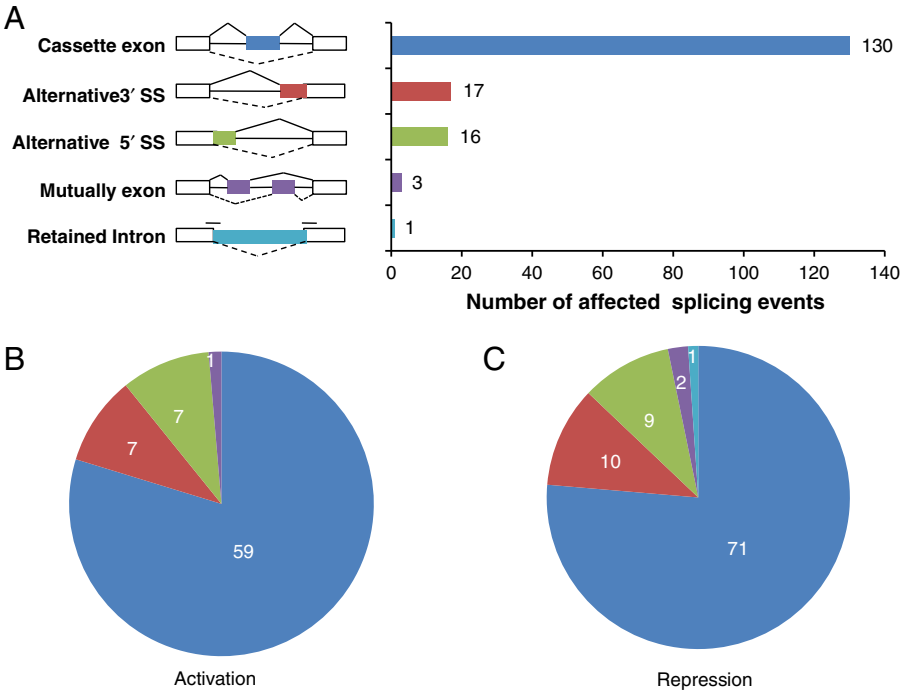

Fig. 2. SRSF10 affected AS events in DT40 cells. (A) Diagrams of the exclusion and inclusion isoforms for the five modes of AS events that were examined. White boxes, flanking constitutive exons; colored boxes, alternative spliced exons/regions; solid lines, splice junctions supporting the inclusion isoform; dotted lines, splice junctions supporting the exclusion isoform. Bar graph shows the number of affected (p value < 0.05) splicing event. In the affected 167 events, pie charts show the alternative splicing events regulated by SRSF10 activation (B) and repression (C) with the number of events in brackets.

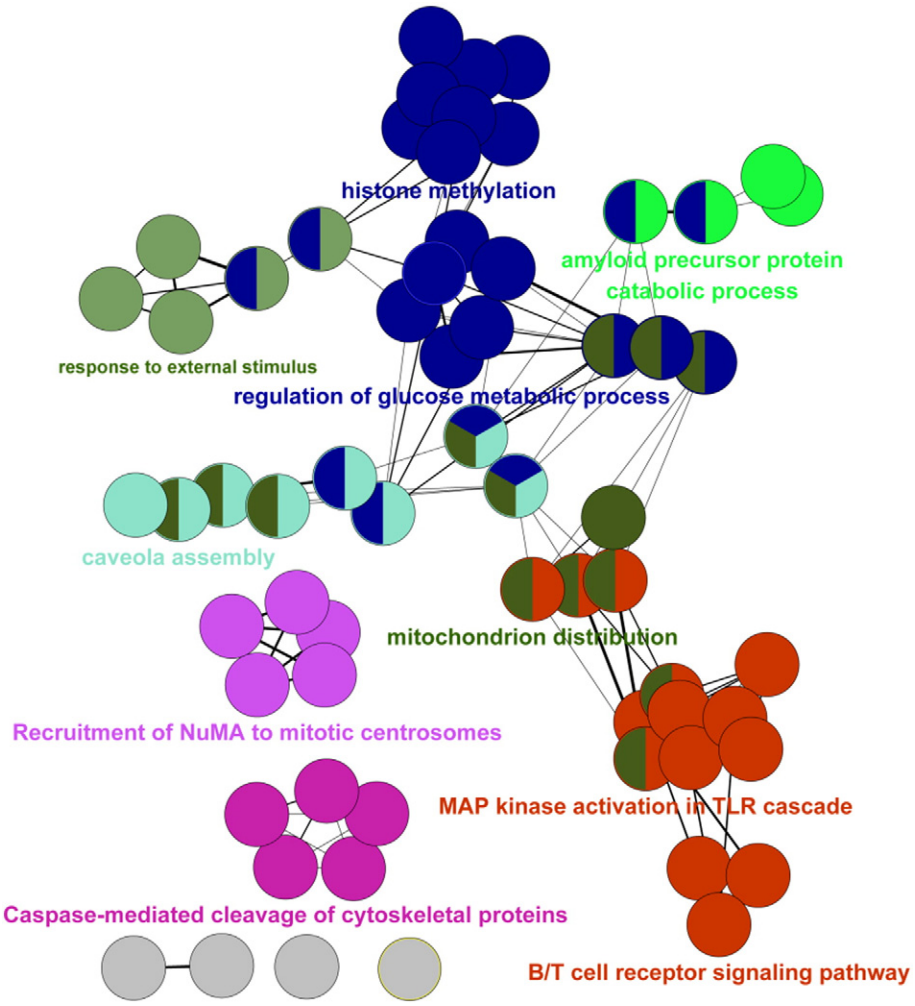

Fig. 3. ClueGO enrichment analysis of SRSF10-regulated genes. To reduce redundancy of GO terms, fusion option was selected. GO/KEGG pathway functionally grouped networks with terms are indicated as nodes (Benjamini–Hochberg p value < 0.05) linked by their kappa score level (≥0.35), where only the label of the most significant term per group is shown. Functionally related groups partially overlap and ungrouped terms are shown in gray.

category. Strikingly, analytical results by ASD indicated that SRSF10 could both activate and repress exon inclusion (Fig. 2B). RT-PCR validation of SRSF10-affected splicing events further demonstrated this prediction. In addition, the selection of alternative 3′ and 5′ splice sites was similarly affected by SRSF10. Taken together, these results provide strong evidence that SRSF10 can promote both exon inclusion and exclusion in vivo.

Function enrichment analysis of SRSF10-affected genes

To understand how loss of SRSF10 affects cellular biological processes, we next performed gene ontology (GO) and pathway analysis of SRSF10-affected genes by using ClueGO [6]. Function enrichment analysis indicated that these affected spliced genes are involved in multiple processes, including histone methylation, response to external stimulus, glucose metabolic processes and mitochondrion distribution (Fig. 3). These results illustrated that SRSF10 might play an important role in a wide range of biological processes by controlling its downstream splicing targets.

Discussion

We present here a unique dataset of RNA-seq results in which we globally analyzed SRSF10-regulated AS events based on deep sequencing of the chicken transcriptome. We found that SRSF10 is involved in all of the common modes of AS, with cassette exons being the most frequent targets for SRSF10 regulation. We also found that SRSF10 could both activate and repress exon inclusion in vivo. Finally, SRSF10-affected splicing genes were involved in multiple processes, indicating a role of SRSF10 in these processes. This dataset is of high quality and will facilitate future study of SRSF10 function in vivo and will also improve understanding of AS regulation in vivo.

Acknowledgements

We thank Dr. Manley (Columbia University) for his generous gifts of SRSF10 KO cells and anti-SRSF10 antibodies. This work was supported by grants from the Ministry of Science and Technology of China [2012CB524900]; the National Natural Science Foundation [31370786 and 31170753] and the One Hundred Talents Program of the Chinese Academy of Sciences.

References

[1] Y. Feng, M.T. Valley, J. Lazar, A.L. Yang, R.T. Bronson, S. Firestein, W.A. Coetzee, J.L. Manley, SRp38 regulates alternative splicing and is required for Ca(2+) handling in the embryonic heart. Dev. Cell 16 (2009) 528–538.
[2] Y. Feng, M. Chen, J.L. Manley, Phosphorylation switches the general splicing repressor SRp38 to a sequence-specific activator. Nat. Struct. Mol. Biol. 15 (2008) 1040–1048.
[3] C. Trapnell, L. Pachter, S.L. Salzberg, TopHat: discovering splice junctions with RNA-seq. Bioinformatics 25 (2009) 1105–1111.

[4] C. Trapnell, B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. van Baren, S.L. Salzberg, B.J. Wold, L. Pachter, Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28 (2010) 511–515.
[5] X. Zhou, W. Wu, H. Li, Y. Cheng, N. Wei, J. Zong, X. Feng, Z. Xie, D. Chen, J.L. Manley, et al., Transcriptome analysis of alternative splicing events regulated by SRSF10 reveals position-dependent splicing modulation. Nucleic Acids Res. (2014), http://dx.doi.org/10.1093/nar/gkt1387.
[6] G. Bindea, B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W.H. Fridman, F. Pages, Z. Trajanoski, J. Galon, ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 25 (2009) 1091–1093.