Identification of Tat-SF1 cellular targets by exon array analysis reveals dual roles in transcription and splicing

HEATHER B. MILLER,[1,2,7] TIMOTHY J. ROBINSON,[1,2,3,4] RALUCA GORDÂN,[5,8] ALEXANDER J. HARTEMINK,[5] and MARIANO A. GARCIA-BLANCO[1,2,6]

[1]Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, North Carolina 27710, USA
[2]Center for RNA Biology, Duke University Medical Center, Durham, North Carolina 27710, USA
[3]Medical Scientist Training Program, Duke University Medical Center, Durham, North Carolina 27710, USA
[4]Program in Molecular Cancer Biology, Duke University Medical Center, Durham, North Carolina 27710, USA
[5]Department of Computer Science, Duke University, Durham, North Carolina 27708, USA
[6]Department of Medicine, Duke University Medical Center, Durham, North Carolina 27710, USA

ABSTRACT

Tat specific factor 1 (Tat-SF1) interacts with components of both the transcription and splicing machineries and has been classified as a transcription-splicing factor. Although its function as an HIV-1 dependency factor has been investigated, relatively little is known about the cellular functions of Tat-SF1. To identify target genes of Tat-SF1, we utilized a combination of RNAi and exon-specific microarrays. These arrays, which survey genome-wide changes in transcript and individual exon levels, revealed 450 genes with transcript level changes upon Tat-SF1 depletion. Strikingly, 98% of these target genes were down-regulated upon depletion, indicating that Tat-SF1 generally activates gene expression. We also identified 89 genes that showed differential exon level changes after Tat-SF1 depletion. The 89 genes showed evidence of many different types of alternative exon use consistent with the regulation of transcription initiation sites and RNA processing. Minimal overlap between genes with transcript-level and exon-level changes suggests that Tat-SF1 does not functionally couple transcription and splicing. Biological processes significantly enriched with transcript- and exon-level targets include the cell cycle and nucleic acid metabolism; the insulin signaling pathway was enriched with Tat-SF1 transcript-level targets but not exon-level targets. Additionally, a hexamer, ATGCCG, was over-represented in the promoter region of genes showing changes in transcription initiation upon Tat-SF1 depletion. This may represent a novel motif that Tat-SF1 recognizes during transcription. Together, these findings suggest that Tat-SF1 functions independently in transcription and splicing of cellular genes.

Keywords: Tat-SF1; RNA splicing; alternative RNA processing; transcription; exon array

INTRODUCTION

Tat specific factor 1 (Tat-SF1) was first discovered over a decade ago as a cellular protein required for Tat-specific, TAR-dependent activation of HIV-1 transcription in vitro (Zhou and Sharp 1995, 1996; Li and Green 1998; Wu-Baer et al. 1998; Parada and Roeder 1999). Immunodepletion of this protein from nuclear extracts resulted in a reduction in Tat transactivation (Zhou and Sharp 1996; Li and Green

1998; Parada and Roeder 1999); overexpression of Tat-SF1 resulted in a small increase in Tat transactivation, which was mostly due to a decrease in the basal level of LTR-driven transcripts (Zhou and Sharp 1996). We re-examined the role of Tat-SF1 using methods not available when this protein was first identified and concluded that Tat-SF1 is not required for Tat transactivation of HIV-1 transcription in vivo (Miller et al. 2009). We confirmed recent work (Brass et al. 2008) that Tat-SF1 is required for efficient HIV-1 propagation and showed that this host factor is required to maintain the relative levels of unspliced and spliced viral RNAs (Miller et al. 2009).

Although Tat-SF1 has been studied primarily as a cofactor for HIV-1 gene regulation, some reports investigate its cellular function. Tat-SF1 has been proposed to be a general transcriptional elongation factor based on the loss of elongation competence in nuclear extracts depleted of Tat-SF1 and the enhancement of elongation competence by Tat-SF1

complementation (Li and Green 1998; Parada and Roeder 1999). Indeed, Tat-SF1 associates with RNA polymerase II (RNAPII) (Parada and Roeder 1999), Tat, P-TEFb (Zhou et al. 1998; Fong and Zhou 2000), hSPT5, the RAP30 protein of TFIIF (Kim et al. 1999), and TCERG1 (Smith et al. 2004; Sanchez-Alvarez et al. 2006). Very recently, Tat-SF1 has been identified as a factor that facilitates RNAPII elongation, along with DSIF/hSPT4-hSPT5 and Paf1 (Chen et al. 2009). On the other hand, Kim and colleagues showed that neither transcription from three different promoters was affected by Tat-SF1 immunodepletion, nor was transcription activation by Gal4-VP16 affected by Tat-SF1 overexpression (Kim et al. 1999).

In addition to associating with transcription factors, Tat-SF1 interacts with several components of the spliceosome, such as snRNP proteins U1 70K and U2B″, and Sm proteins B and B′. Tat-SF1 also associates with all five spliceosomal U snRNAs, and this interaction depends on its RNA recognition motifs (Fong and Zhou 2001). Moreover, the yeast homolog of Tat-SF1, CUS2, helps refold U2 snRNAs to aid in prespliceosome assembly (Yan et al. 1998). It remains to be determined if Tat-SF1 is a functional homolog of CUS2.

We now know that many steps in pre-mRNA processing occur cotranscriptionally (for review, see Hirose and Manley 2000; Goldstrohm et al. 2001; Bentley 2002; Maniatis and Reed 2002; Neugebauer 2002; Orphanides and Reinberg 2002; Proudfoot et al. 2002; Kornblihtt et al. 2004; Bentley 2005; Pandit et al. 2008). Several proteins, including Tat-SF1, have been proposed to couple transcription and splicing via their interactions with factors of both machineries (Fong and Zhou 2001). Indeed, large RNAPII-containing complexes that associate with 5′-splice sites contain Tat-SF1 (Kameoka et al. 2004). Nevertheless, functional coupling by Tat-SF1 has not been demonstrated.

The aforementioned studies indicate that the functions of Tat-SF1 remain largely unknown. To investigate Tat-SF1 function, we identified the cellular targets of Tat-SF1 using a combination of RNA interference (RNAi) and genome-scale splicing analysis. The advent of exon array technology now allows an unbiased survey of the human transcriptome for changes in both transcript and exon levels. Many of the reports utilizing exon arrays have focused on experimental comparisons of gene expression data across different microarray platforms (Abdueva et al. 2007; Robinson and Speed 2007; Bemmo et al. 2008). Others have used exon arrays to investigate alternative splicing differences between normal and diseased tissues (Gardina et al. 2006; French et al. 2007), in the presence of a stimulus (McKee et al. 2007), or instances of tissue-specific alternative splicing (Clark et al. 2007). To date, relatively few

transcription and/or splicing factors' functions have been investigated by exon array analysis (Hung et al. 2008; Oberdoerffer et al. 2008; Xing et al. 2008; Sun and Li 2009; Warzecha et al. 2009). In the work presented here, Tat-SF1 depletion led to many changes in overall transcript levels, with the overwhelming majority of these genes showing decreased expression. Additionally, targets were identified with differential exon levels, which could be attributed to changes in transcript initiation, splicing of internal exons, polyadenylation, and/or stability of different mRNA isoforms.

RESULTS

Tat-SF1 depletion results in widespread changes in both transcription and splicing

Since Tat-SF1 has been implicated in both transcription and splicing, Affymetrix GeneChip Human Exon 1.0ST Arrays are an attractive technology to use for studying its function because they survey both transcript-level changes (due to altered transcription and/or stability) and exon-level changes (due to alternative transcript initiation, alternative pre-mRNA processing, and/or differential isoform stability) (Gardina et al. 2006). Using these arrays, we explored global changes in the human transcriptome occurring after Tat-SF1 depletion using T-Rex-293 cell lines that stably expressed green fluorescent protein (GFP) and tetracycline-inducible Tat-SF1-specific shRNAs (Miller et al. 2009). To confirm knockdown, cell lysates were collected after 72 h of tetracycline induction and analyzed by Western blotting. Biological triplicates show that Tat-SF1 was efficiently depleted in both Tat-SF1(A) and Tat-SF1(B) cells compared to empty vector and GFP-shRNA control cells (Fig. 1A). The polypyrimidine tract binding protein (PTB) was used as a loading control (Fig. 1A). Quantitative RT-PCR confirmed that knockdown with Tat-SF1(A) and Tat-SF1(B) shRNAs also led to a decrease in *Tat-SF1* RNA levels (see below and the upcoming Fig. 5A). Total RNA from these cells was processed for hybridization on the aforementioned exon arrays. After Robust Multichip Averaging normalization and background correction, the probe sets on
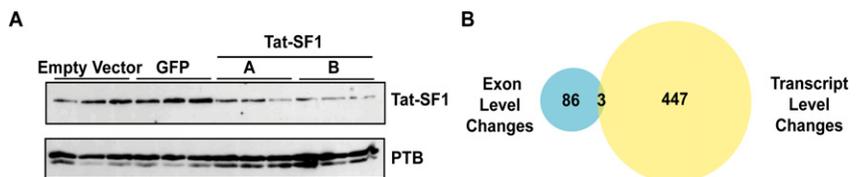

FIGURE 1. Tat-SF1 depletion results in significant changes in both transcript and differential exon levels. (*A*) Western blot analysis of Tat-SF1 depletion in stable T-Rex-293 cells. Tat-SF1 knockdown was assessed in triplicate samples by probing with anti-Tat-SF1 antibody. Anti-PTB served as a loading control. (*B*) Summary of exon array results. Exon array analysis comparing control cells (empty vector and GFP shRNA) and Tat-SF1 knockdown cells [both Tat-SF1(A) and Tat-SF1(B) shRNAs] resulted in 89 differential exon-level changes and 450 transcript-level changes ($P \le 1 \times 10^{-2}$). Three genes showed evidence of both transcript-level and differential exon-level changes.
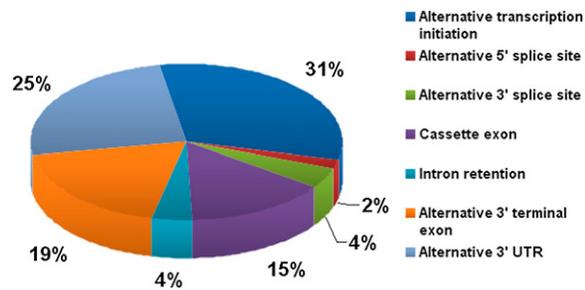
FIGURE 2. Tat-SF1 regulates a number of alternative processing events. Eighty-nine exon-level targets were manually inspected for evidence of a known alternative processing event. Forty-eight of these showed one clear hypothesis and were categorized.

these chips were further filtered to reduce the detection of false positives (see Materials and Methods; Supplemental Table 1). Transcript-level changes were analyzed using XRAY (Biotique Systems, Inc.) and exon-level changes were analyzed using a modified version of SplicerAV (Robinson et al. 2010) (see Materials and Methods).

In response to Tat-SF1 depletion, overall transcript levels changed for 450 genes ($P \leq 1 \times 10^{-2}$) out of 7317 genes tested (Fig. 1B; Supplemental Table 2). The analysis described here uses data comparing pooled controls (two empty vector and three GFP shRNA samples) vs. pooled experimentals (three shRNA A and three shRNA B samples). Very similar data were obtained if each of the two Tat-SF1 shRNAs were considered separately (data not shown); this indicates that the effects observed were likely due to Tat-SF1 knockdown and not to idiosyncratic effects of the shRNAs. The overwhelming majority of these genes (442 of 450) showed decreased transcript levels compared to control cells. As expected, the arrays showed that Tat-SF1 depleted cells showed a 1.5-fold decrease in *Tat-SF1* transcript levels. Differential exon levels were detected in 89 genes ($P \leq 1 \times 10^{-2}$) out of 7317 genes tested (Fig. 1B; Supplemental Table 3). Only three genes showed evidence of changes in both transcript and exon levels, which was not higher than would be expected by chance alone ($P = 0.19$). Transcript- and exon-level changes were independently assigned to each category and therefore the low overlap is not due to artificial constraints of the analysis. These data support a role for Tat-SF1 in both transcription and splicing, but do not support a role in coupling of the two processes.

We investigated the genes that showed changes in transcript and exon levels (termed here transcript-level targets and exon-level targets) in parallel, first focusing on the exon-level events regulated by Tat-SF1.

Tat-SF1 influences differential exon levels at both the 5′ and 3′ ends of transcripts

With the large number of differential exon levels detected by exon array analysis, we wanted to determine if any one type of

alternative processing occurred more frequently than another in response to Tat-SF1 depletion. Genes showing Tat-SF1 depletion-mediated exon-level changes (89 genes) were interrogated to determine whether or not the predicted exon-level changes corresponded to a UCSC known alternative transcript. Among those that did (55 genes) we analyzed only those where the change could be explained by one clear hypothesis (48 genes). Seven different categories of alternative processing were found among Tat-SF1 exon-level targets (Fig. 2). The largest percentage of targets showed results consistent with the utilization of alternative transcription initiation sites (15 genes; 31%), followed by alternative 3′ UTRs (12 genes; 25%), alternative 3′ terminal exons (nine genes; 19%), cassette exons (seven genes; 15%), intron retention (two genes; 4%), alternative 3′ splice sites (two genes; 4%), and alternative 5′ splice sites (one gene; 2%).

Exon-level targets identified by the exon arrays were experimentally validated using RT-qPCR. Of the six targets tested, three were validated. Homer homolog 2 (*HOMER2*), the highest-ranked target by splice score, has four different transcript isoforms; two of these utilize an alternative 3′ splice site. The exon arrays showed an increase in probe-set intensity that corresponded to an exon with an alternative 3′ splice site. Indeed, amplification with primers specific to that exon showed an increase in the relative quantity of this transcript isoform in Tat-SF1 knockdown cells (Fig. 3A). An alternative transcription initiation event was depicted by probe sets specific to Fanconi anemia, complementation group G (*FANCG*). Three UCSC transcripts represent *FANCG*, one of which has an alternative first exon. Primers that specifically amplified the canonical, upstream first exon confirmed that these levels were increased upon Tat-SF1 knockdown (Fig. 3B). Finally, the TAP binding protein (*TAPBP*) is represented by six UCSC transcripts, one with a retained intron. Exon arrays showed an increase in probe-set intensity for that intron. Primers designed to anneal to

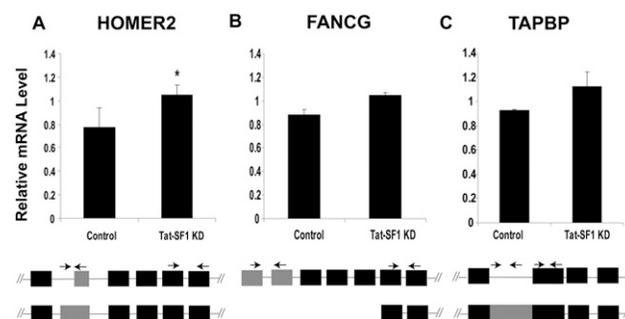**FIGURE 3.** Tat-SF1 is involved in both transcription initiation and alternative processing decisions. (*A*) RT-qPCR validation of *HOMER2*. Primer locations are indicated with arrows. The transcript variant of interest (gray) was quantified relative to an mRNA common to all transcript variants. Error bars show standard error of the mean. (*) $P \leq 0.05$. (*B*) RT-qPCR validation of *FANCG*. (*C*) RT-qPCR validation of *TAPBP*.

that intron validated that increased levels of this variant were present upon Tat-SF1 knockdown (Fig. 3C).

The hexamer ATGCCG is over-represented among Tat-SF1 alternative initiation targets

A bias in Tat-SF1 depleted cells for changes in transcription initiation led us to focus on these targets to acquire more mechanistic detail on how Tat-SF1 is involved in initiation site choice. Evidence for a *cis*-acting element required for the Tat-SF1 function has not previously been published, so we sought to determine if a sequence was over-represented among target genes with alternative transcription initiation events. The sequence ATGCCG was enriched in the promoter regions of the 15 Tat-SF1 target genes showing alternative initiation, as compared to 15 negative control sequences (see Materials and Methods). This hexamer was found in 10 of the 15 sequences analyzed, but was only found in two of the negative control sequences (hypergeometric $P = 4 \times 10^{-3}$). Although other sequences were also over-represented in these regions with very significant P-values (Supplemental Table 4), we were particularly interested in ATGCCG for several reasons. First, this hexamer was frequently located within 2 kb upstream of the most upstream alternative first exon (seven out of 10 target gene regions) (Fig. 4). The hexamer was found in the sense orientation at least once in five of the target gene regions and in the antisense orientation in six of the target gene regions. ATGCCG was never found within an exon. Finally, this motif was not over-represented in the promoter regions of all Tat-SF1 exon-level target genes (hypergeometric $P = 1 \times 10^{-1}$). Analysis of the 15 alternative initiation targets along with 33 other targets represent-
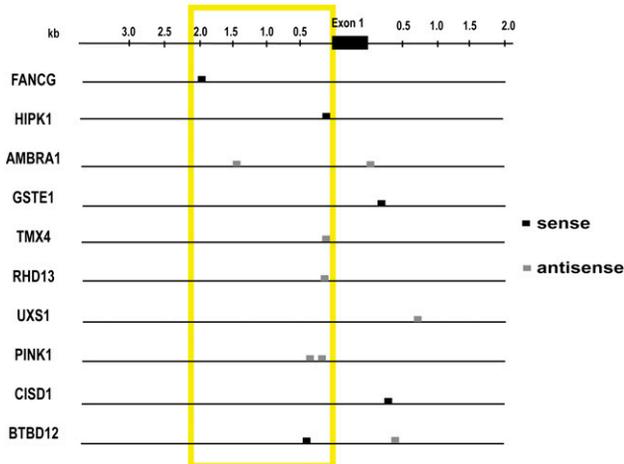

FIGURE 5. Tat-SF1 positively regulates overall transcript levels of many genes. (*A*) RT-qPCR validation of *Tat-SF1*. The target gene level was quantified relative to *GAPDH*. Error bars show standard error of the mean. (*) $P \leq 0.05$. (**) $P \leq 0.01$. (*B–F*) RT-qPCR validations of *TMEM54, HDGF2, BOP1, DCI*, and *CCND1*, respectively.

ing various categories of alternative processing showed that ATGCCG was found in only 21 of the 48 targets examined (Supplemental Table 4). These data suggest that the ATGCCG motif does not play a role in all Tat-SF1 exon-level regulated events, but plays a role in Tat-SF1 regulation of transcription initiation.

Tat-SF1 positively regulates a large number of overall transcript levels

In addition to the exon-level targets described above, 450 targets were identified that showed a change in overall transcript levels with Tat-SF1 knockdown. As described previously, 98% of these were down-regulated. Although fold changes were small, we were able to validate six out of nine tested. RT-qPCR confirmed that *Tat-SF1* levels were decreased more than twofold in knockdown compared to control cells (Fig. 5A). Dodecenoyl-CoA isomerase (*DCI*) showed the largest down-regulation by exon array (2.3-fold), followed by a 1.5-fold change in transmembrane protein 54 (*TMEM54*), hepatoma-derived growth factor-related protein 2 (*HDGF2*), block in proliferation 1 (*BOP1*), and cyclin D1 (*CCND1*) (Fig. 5B–F). These validated changes show that Tat-SF1 knockdown results in reproducible down-regulation of many cellular genes.

Tat-SF1 regulates transcription and alternative processing of genes involved in the cell cycle and nucleic acid metabolism

Analysis of gene ontology biological processes showed that genes annotated as affecting cell signaling, host defense, translation, and the cell cycle were significantly enriched among Tat-SF1 transcript-level targets (Table 1). Several

FIGURE 4. The hexamer ATGCCG is over-represented in the promoter region of Tat-SF1 alternative initiation targets. Sequences surrounding exon 1 of Tat-SF1 exon-level targets that were categorized as alternative transcription initiation events. Exon 1 represents the most upstream exon found in the transcript isoforms. Locations of ATGCCG are indicated with black boxes and CGGCAT (reverse complement) are indicated with gray boxes.
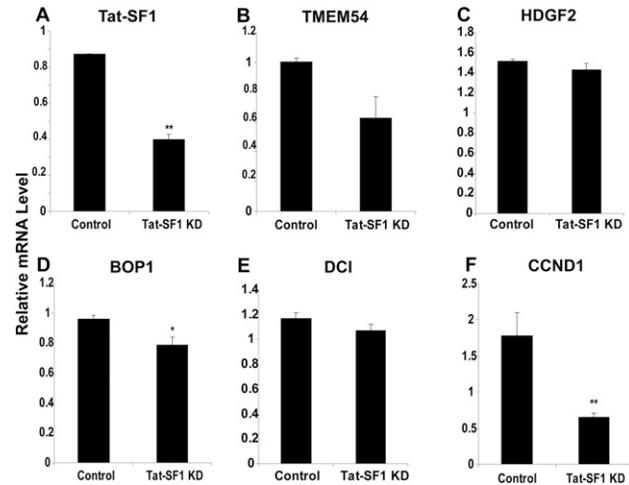
TABLE 1. Gene ontology (GO); cellular processes over-represented among Tat-SF1 transcript-level targets

Annotation	Tat-SF1 targets	Total possible	P-value
GO:0007166 [5]: cell surface receptor linked signal transduction	17	1421	2×10^{-4}
GO:0044249 [5]: cellular biosynthesis	45	944	4×10^{-4}
GO:0043037 [7]: translation	14	163	9×10^{-4}
GO:0007049 [5]: cell cycle	36	713	1×10^{-3}
GO:0006952 [5]: defense response	8	838	2×10^{-3}
GO:0043207 [5]: response to external biotic stimulus	3	502	5×10^{-3}
GO:0000278 [6]: mitotic cell cycle	14	196	7×10^{-3}
GO:0044260 [5]: cellular macromolecule metabolism	103	2986	8×10^{-3}
GO:0006415 [8]: translational termination	3	8	1×10^{-2}
GO:0007243 [6]: protein kinase cascade	15	252	2×10^{-2}
GO:0006139 [5]: nucleobase, nucleoside, nucleotide and nucleic acid metabolism	99	2967	3×10^{-2}

Criteria: tree branch level ≥ 5 (if ≤50% overlap with tree branch of higher level, both GO categories were included), Tat-SF1 target genes in GO category ≥ 2, P-value ≤ 5×10^{-2}.

lambda (POLL). This polymerase is involved in DNA repair processes (Aoufouchi et al. 2000) and interacts with two Tat-SF1 exon-level targets identified here: leucine zipper, down-regulated in cancer 1 and proliferating cell nuclear antigen.

DISCUSSION

Previous attempts to determine the cellular functions of Tat-SF1 were limited to immunodepletions and overexpression analyses (Zhou and Sharp 1996; Li and Green 1998; Kim et al. 1999). Newer technologies allowed us to reinvestigate the role of Tat-SF1 in cellular gene expression and expand on the information available for this relatively understudied protein. In this work, RNAi was used in conjunction with splicing-sensitive microarrays to gather new information about the cellular functions of a candidate transcription-splicing factor.

biological processes were significantly enriched for Tat-SF1 exon-level targets (Table 2), with response to a DNA damage stimulus being the most significant ($P = 3 \times 10^{-4}$). Interestingly, both Tat-SF1 transcript- and exon-level targets were enriched for cell cycle and nucleic acid metabolism related categories.

The top biological pathway enriched for Tat-SF1 transcript-level target genes was insulin signaling ($P = 1 \times 10^{-2}$) (Table 3). The target genes identified as components of the insulin signaling pathway primarily encode kinases and phosphatases (Table 4). Another signaling-related biological pathway, cytokine–cytokine receptor interaction, was also enriched for Tat-SF1 transcript-level target genes ($P = 1 \times 10^{-2}$). Pyrimidine metabolism was found to be the most significantly enriched biological pathway among Tat-SF1 exon-level targets ($P = 5 \times 10^{-3}$) (Table 5). Finally, known interactions between human proteins and those identified as Tat-SF1 targets were analyzed computationally to determine if any proteins were enriched among these targets. Tat-SF1 transcript-level targets showed enrichment for CDC2 and SKP2 interacting proteins ($P = 1 \times 10^{-2}$) (Table 6), both of which are involved in cell cycle regulation. This analysis also revealed protein interactors of Tat-SF1 exon-level targets (Table 7). Interestingly, both Fanconi anemia, complementation group G and F (FANCG and FANCF, respectively), were identified as Tat-SF1 targets. These protein products make up part of the FA protein network, which is involved in response to DNA damage (Gurtan and D'Andrea 2006; Levitus et al. 2006; Patel and Joenje 2007; Wang 2007). The most significantly over-represented protein interactor was polymerase (DNA-directed),

Several findings emerge from the summary of changes detected by the exon arrays. First, we observed that Tat-SF1 depletion changed the overall transcript levels of 450 out of 7317 genes tested. Ninety-eight percent of these were down-regulated, consistent with Tat-SF1 positively regulating gene expression, most likely transcription. The modest effects on transcript levels we observed are consistent with recent data of the Handa lab that demonstrated complex regulation of elongation by Tat-SF1, hSPT5, and the Paf1 complex (Chen et al. 2009). Although previous studies have concluded that Tat-SF1 is a general elongation factor (Li and Green 1998; Parada and Roeder 1999), our studies suggest that most genes are not sensitive to significant depletion of Tat-SF1.

Furthermore, our data suggest that Tat-SF1 alters the use of alternative first exons, either by directing transcription initiation events or by regulating the stability of two RNA isoforms that initiate at two different sites. The former would imply a role for Tat-SF1 at transcription initiation. Tat-SF1 has been shown to directly bind at least one transcription

TABLE 2. Gene ontology (GO); cellular processes over-represented among Tat-SF1 exon-level targets

Annotation	Tat-SF1 targets	Total possible	P-value
GO:0006974 [5]: response to DNA damage stimulus	6	201	3×10^{-4}
GO:0006259 [5]: DNA metabolism	7	538	2×10^{-3}
GO:0000075 [7]: cell cycle checkpoint	2	34	2×10^{-2}
GO:0006139 [5]: nucleobase, nucleoside, nucleotide and nucleic acid metabolism	20	2967	3×10^{-2}

Criteria: tree branch level ≥ 5 (if ≤50% overlap with tree branch of higher level, both GO categories were included), Tat-SF1 target genes in GO category ≥ 2, P-value ≤ 5×10^{-2}.

TABLE 3. Biological pathways over-represented among Tat-SF1 transcript-level targets

Annotation	Tat-SF1 targets	Total possible	P-value
path:hsa04910: insulin signaling pathway	9	144	1×10^{-2}
path:hsa04060: cytokine-cytokine receptor interaction	2	257	1×10^{-2}

Criteria: Tat-SF1 target genes in pathway ≥ 2, P-value $\leq 5 \times 10^{-2}$.

initiation factor, the RAP30 subunit of TFIIF (Kim et al. 1999). TFIIF is associated with RNAPII during preinitiation complex assembly, initiation, and elongation (Tan et al. 1994, 1995). We speculate that Tat-SF1, in a RAP30:RNAPII complex, could help signal transcription initiation. It is possible then that upon Tat-SF1 depletion, alternative promoters are used, resulting in transcripts with different first exons. This finding also demonstrates how exon arrays can delineate changes in transcription and alternative processing, providing additional detail about gene expression changes that would have most likely been missed by conventional gene expression arrays.

A second explanation for a 5′ end bias is that Tat-SF1 may differentially regulate the stability of RNA isoforms with alternative first exons. While we cannot rule out this second explanation, we favor the earlier one related to initiation choice.

Exclusion of exon-level changes categorized as alternative transcription initiation events still leaves 74 target genes with Tat-SF1-mediated effects on RNA processing, only two of which also exhibit changes at the overall transcript level. Given the insignificant overlap between alternatively processed targets with transcript-level targets, our study suggests that Tat-SF1 acts independently on transcription and RNA processing.

Whereas the conclusions above represent our best interpretation of the data obtained, it is important to point out several caveats of the analysis. First, the methodology used here cannot distinguish between direct and indirect effects of Tat-SF1. This distinction will require extensive testing of each of the hypotheses generated from the data. Second, validation of exon array data by independent methods (e.g., RT-PCR) requires careful attention to probe-set behavior and the isoform changes it may predict. Exon arrays are susceptible to "edge bias effect" that may lead to many false positives in the 5′ and 3′ ends of genes (Bemmo et al. 2008). Probe sets located at the 5′ and 3′ ends of genes can respond in a nonuniform fashion (hypothesized to be due to the nature of promoter sequences and the sample labeling protocol), thus falsely indicating an alternative

processing/initiation event. This behavior was seen in instances with a very large fold change in transcript levels between two groups, which was not the case in our analysis, however. Nevertheless, this observation does emphasize the importance of visually inspecting exon array probe-set behavior and comparing it to annotated isoforms in a genome browser to make an informed hypothesis regarding alternative exon use. Exon array false positives can also arise from nonexpressed probes or probes that do not vary in their expression. To minimize these artifacts, several filtering strategies were employed (see Materials and Methods). Although the filters help to reduce the number of false positives, false positives are inevitable in experiments such as these. As analysis of exon array data continues to move forward, alternative statistical approaches may prove to be superior at minimizing false positives (Laajala et al. 2009). In our analysis, we employed a variant of a program described previously, SplicerAV (Robinson et al. 2010), which preferentially detected changes in alternative mRNA processing detected by probe sets changing in opposite directions. This more conservative approach yielded a 62% overlap between predicted exon-level changes and previously known transcript variants (55 of 89) vs. 20% by XRAY (Biotique Systems, Inc.) analysis. In this study, we based validation experiments on the interpretation of probe-set intensity changes, which in practice, were limited by the annotation available on genome browsers. Whereas we previously found a 33% validation of exon-level changes predicted using XRAY, the validation rate using SplicerAV was 50%, which is typical of exon array analysis of alternative processing (Gardina et al. 2006; Clark et al. 2007; French et al. 2007; Hung et al. 2008; Moore and Silver 2008). These studies suggest that the SplicerAV program is a better choice for the analysis of exon array data and we are currently improving the program for open distribution.

TABLE 4. Tat-SF1 transcript-level targets belonging to the insulin signaling pathway

Gene symbol	Annotation	P-value transcript-level change
EIF4E2	Eukaryotic translation initiation factor 4E family member 2	3×10^{-3}
FLOT1	Flotillin 1	7×10^{-4}
FLOT2	Flotillin 2	5×10^{-3}
IKBKB	Inhibitor of kappa light polypeptide gene enhancer in B-cells kinase beta	6×10^{-3}
INPPL1	Inositol polyphosphate phosphatase-like 1	2×10^{-3}
MAP2K2	Mitogen-activated protein kinase kinase 2	2×10^{-3}
MAPK3	Mitogen-activated protein kinase 3	2×10^{-3}
PKM2	Pyruvate kinase muscle	1×10^{-2}
PRKAR2A	Protein kinase cAMP-dependent regulatory type II alpha	2×10^{-3}

TABLE 5. Biological pathways over-represented among Tat-SF1 exon-level targets

Annotation	Tat-SF1 targets	Total possible	P-value
path:hsa00240: pyrimidine metabolism	3	89	5×10^{-3}
path:hsa00051: fructose and mannose metabolism	2	52	1×10^{-2}
path:hsa04510: focal adhesion	4	234	2×10^{-2}

Criteria: Tat-SF1 target genes in pathway ≥ 2, P-value $\leq 5 \times 10^{-2}$.

MATERIALS AND METHODS

RNAi-mediated depletion of Tat-SF1

T-Rex-293 cells stably expressing GFP and expressing one of two shRNAs specific for Tat-SF1 (A or B), GFP, or an empty vector control, were cultured and induced with tetracycline as described previously (Miller et al. 2009). After inducing cells in three independent experiments, cell lysates were prepared for Western blot analysis by freeze–thaw lysis. Depletion of Tat-SF1 was assessed by probing membranes with anti-Tat-SF1 (Research Genetics Inc.) and anti-PTB as a loading control, followed by anti-rabbit secondary antibody (Amersham). Proteins were detected with SuperSignal West Chemiluminescence Substrate (Pierce).

Sample preparation and exon array processing

Total RNA was isolated from T-Rex-293 cells with RNeasy Mini kits (Qiagen) and treated with DNase I (Ambion). One microgram of RNA from each of the 12 samples (biological triplicates of empty vector, anti-GFP, anti-Tat-SF1(A) and anti-Tat-SF1(B)) was subjected to the Whole Transcript Sense Target Labeling Assay, as described by the manufacturer (Affymetrix). At several points in the labeling protocol, sample quality and yield was assessed using an Agilent Lab-on-a-Chip 2100 Bioanalyzer. Samples were hybridized to Affymetrix GeneChip Human Exon 1.0ST Arrays.

Normalization and background correction

Quality control on each of the 12 CEL files was performed using Affymetrix Expression Console and XRAY (Biotique Systems, Inc.). In our analysis, one CEL file, belonging to the third replicate of the empty vector control condition, was eliminated from subsequent analyses based on quality control metrics. XRAY was used to normalize and background-correct each CEL file. The files were grouped into either the control group (two empty vector controls plus three GFP knockdown controls) or the Tat-SF1 knockdown group [three Tat-SF1(A) plus three Tat-SF1 (B)]. The 6,553,590 probes were manipulated into the analysis values as follows. Probes with a GC count less than six and greater than 17 were excluded from the analysis. Probe scores were then transformed by taking the natural logarithm of 0.1 plus the probe score. Each probe score was corrected for background by subtracting the median expression score of background probes with similar GC content.

Probe-set expression scores and filtering

The HumanExon 1.0ST array contains 1,404,693 probe sets (typically, but not always, sets of four probes). The expression score for a probe set was defined to be the median of its probe expression scores, and probe sets with fewer than three probes (that pass all of the tests defined above) were excluded from further analysis. Probe-set reliability is ranked from more to less reliable as Core, Extended, or Full. For this analysis, only Core probe sets were analyzed. Nonexpressed probe sets and low-variance probes were also excluded from the analysis to minimize the number of false-positives.

The 73,293 remaining probe sets corresponding to 7394 transcript clusters (genes) were passed on to a mixed model, nested analysis of variance (ANOVA) (Montgomery 2006) to detect overall transcript level differences between the control and Tat-SF1 knockdown groups using XRAY.

Changes in mRNA processing were conducted using a variant of a previously described program, SplicerAV (Robinson et al. 2010). First, all probe sets were mapped to known UCSC genome browser genes, leaving 7317 genes available for analysis for changes in alternative mRNA processing. Within these genes, highly correlated probe sets were combined (Pearson correlation > 0.7 across the 11 samples) to simplify SplicerAV's analysis of the increased number of probe sets found on the exon arrays. Following this prefiltering step, SplicerAV was used to analyze these probe sets for differential exon level changes within each gene and then rank each gene in order of descending splice score (Robinson et al. 2010).

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (Edgar et al. 2002) and are accessible through GEO Series accession number GSE19927 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE19927).

Computational analyses

The GATHER gene annotation tool (http://gather.genome.duke.edu/) (Chang and Nevins 2006) was used to examine the Tat-SF1 transcript- and exon-level target genes for over-representation in gene ontology categories, biological pathways, and protein interactions. The criteria for filtering these categories are listed below each table. Categorization of alternative processing events was performed by manually inspecting hits from SplicerAV on the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway).

TABLE 6. Protein interactors over-represented among Tat-SF1 transcript-level targets

Annotation	Tat-SF1 targets	Total possible	P-value
CDC2: cell division cycle 2, G1 to S and G2 to M	7	37	1×10^{-2}
SKP2: S-phase kinase-associated protein 2 (p45)	5	18	1×10^{-2}
HNRPA0: heterogeneous nuclear ribonucleoprotein A0	2	2	5×10^{-2}
ZNF297: zinc finger protein 297	2	2	5×10^{-2}
MYOCD: myocardin	2	2	5×10^{-2}

Criteria: Tat-SF1 target genes in group ≥ 2, P-value $\leq 5 \times 10^{-2}$.

TABLE 7. Protein interactors over-represented among Tat-SF1 exon-level targets

Annotation	Tat-SF1 targets	Total possible	*P*-value
POLL: polymerase (DNA directed), lambda	2	2	$<1 \times 10^{-4}$
DPPA4: developmental pluripotency associated 4	2	7	3×10^{-3}
FANCC: Fanconi anemia, complementation group C	2	9	5×10^{-3}
FANCA: Fanconi anemia, complementation group A	2	11	8×10^{-3}
CD9: CD9 antigen (p24)	2	15	1×10^{-2}
ARNT: aryl hydrocarbon receptor nuclear translocator	2	16	2×10^{-2}

Criteria: Tat-SF1 target genes in group ≥ 2, *P*-value $\leq 5 \times 10^{-2}$.

De novo motif searching

All 15 exon-level target genes categorized as alternative initiation events were analyzed against 15 genes known to have alternative promoters, but in our experiments did not show evidence of Tat-SF1 regulation. All 33 exon-level target genes categorized as alternative processing events were then included in the analysis, along with 33 genes that showed no evidence of an alternative promoter and in our experiments did not show evidence of Tat-SF1 regulation. The alternative initiation target sequences were composed of the alternative promoter sequence displayed on the AltEvents track (150 nucleotides [nt]) of the UCSC Genome Browser and the 4000 flanking nucleotides (2000 nt both 5′ and 3′ for a total of 4150 nt). The alternative processing target sequences were also composed of 4150 nt surrounding the transcription start site. Sequences were analyzed by a hypergeometric enrichment analysis. The hypergeometric *P*-values were computed using the phyper function in R (Johnson et al. 1992).

Experimental validation by quantitative RT-PCR

Each of the exon-level targets was visually inspected using the UCSC Genome Browser to see if probe-set intensity changes between control and Tat-SF1 knockdown groups corresponded to a known or predicted RNA isoform (RefSeq or UCSC gene). An isoform ratio of 1.5 or more was considered acceptable for experimental validation. Transcript-level targets were ranked in order of ascending *P*-value, and required a fold change of 1.5 or more.

T-Rex-293 cells (Empty vector, GFP, Tat-SF1(A) and Tat-SF1(B)) were induced with tetracycline in experiments independent from the exon array experiments and total RNA was isolated using RNeasy Mini kits with on-column DNase digestion (Qiagen). RNA purity and yield was assessed with a NanoDrop (Thermo Scientific). cDNA synthesis was performed with the iScript Select cDNA Synthesis Kit (Bio-Rad). Briefly, 5 µg of total RNA and random hexamer primers were used in a 100-µL reaction according to the manufacturer's instructions. Negative control reactions (without reverse transcriptase) were also performed. cDNA was diluted 1:50 for use in qPCR reactions. All primers were designed to be 75–200 base pairs long using Primer3Plus (Untergasser et al. 2007), and specificity was confirmed using the in silico PCR tool on the UCSC Genome Bioinformatics site (http://genome.ucsc.edu/cgi-bin/hgPcr?command=start). Both transcript- and exon-level target validations utilized intron-flanking primers wherever possible to eliminate genomic DNA amplification. Primers used in Figures 3 and 5 are listed in Supplemental Table 5. GAPDH was selected as a reference gene for transcript level validations because its mRNA

level did not change upon Tat-SF1 knockdown by exon array or RT-qPCR analysis (data not shown). Quantitative PCR reactions were performed in a MyiQ Single-Color Real-Time Detection System (Bio-Rad) using iQ SYBR Green Supermix according to manufacturer's instructions (Bio-Rad). Duplicate reactions were performed with each primer pair for each cDNA. Empty vector and GFP samples constituted the control group and Tat-SF1(A) and (B) samples constituted the knockdown group, consistent with the exon array analysis. The following thermal cycler program was used: 3 min of denaturation at 95°C, 40 cycles of amplification (15 sec of denaturation at 95°C, 30 sec of annealing at 60°C), followed by a melt curve analysis. Melt curves and agarose gel analysis confirmed the presence of one specific product of the correct size and the absence of primer dimers. For transcript-level targets, the target mRNA was quantified relative to GAPDH by the formula ($E_{\text{GAPDH}}^{\text{Ct(GAPDH)}}/E_{\text{target}}^{\text{Ct(target)}}$) (Pfaffl 2001), where E is PCR efficiency, and Ct is the threshold value for amplification. For exon-level targets, the transcript variant of interest was quantified relative to an mRNA common to all transcript variants using the same formula. All mRNA levels were then normalized to the level from an independent well. Error bars represent standard error of the mean. REST 2009 software (Qiagen) was used to test for statistically significant changes in gene expression between the control and knockdown groups through a randomization test (Pfaffl et al. 2002).

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

REFERENCES

Abdueva D, Wing MR, Schaub B, Triche TJ. 2007. Experimental comparison and evaluation of the Affymetrix exon and U133Plus2 GeneChip arrays. *PLoS ONE* **2:** e913. doi: 10.1371/journal.pone.0000913.

Aoufouchi S, Flatter E, Dahan A, Faili A, Bertocci B, Storck S, Delbos F, Cocea L, Gupta N, Weill JC, et al. 2000. Two novel human and mouse DNA polymerases of the polX family. *Nucleic Acids Res* **28:** 3684–3693.

Bemmo A, Benovoy D, Kwan T, Gaffney DJ, Jensen RV, Majewski J. 2008. Gene expression and isoform variation analysis using Affymetrix exon arrays. *BMC Genomics* **9:** 529 doi: 10.1186-1471-2164-9-529.

Bentley D. 2002. The mRNA assembly line: transcription and processing machines in the same factory. *Curr Opin Cell Biol* **14:** 336–342.

Bentley DL. 2005. Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr Opin Cell Biol* **17:** 251–256.

Brass AL, Dykxhoorn DM, Benita Y, Yan N, Engelman A, Xavier RJ, Lieberman J, Elledge SJ. 2008. Identification of host proteins required for HIV infection through a functional genomic screen. *Science* **319:** 921–926.

Chang JT, Nevins JR. 2006. GATHER: a systems approach to interpreting genomic signatures. *Bioinformatics* **22:** 2926–2933.

Chen Y, Yamaguchi Y, Tsugeno Y, Yamamoto J, Yamada T, Nakamura M, Hisatake K, Handa H. 2009. DSIF, the Paf1 complex, and Tat-SF1 have nonredundant, cooperative roles in RNA polymerase II elongation. *Genes Dev* **23:** 2765–2777.

Clark TA, Schweitzer AC, Chen TX, Staples MK, Lu G, Wang H, Williams A, Blume JE. 2007. Discovery of tissue-specific exons using comprehensive human exon microarrays. *Genome Biol* **8:** R64. doi: 10.1186-gb-2007-8-4-r64.

Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30:** 207–210.

Fong YW, Zhou Q. 2000. Relief of two built-in autoinhibitory mechanisms in P-TEFb is required for assembly of a multicomponent transcription elongation complex at the human immunodeficiency virus type 1 promoter. *Mol Cell Biol* **20:** 5897–5907.

Fong YW, Zhou Q. 2001. Stimulatory effect of splicing factors on transcriptional elongation. *Nature* **414:** 929–933.

French PJ, Peeters J, Horsman S, Duijm E, Siccama I, van den Bent MJ, Luider TM, Kros JM, van der Spek P, Sillevis Smitt PA. 2007. Identification of differentially regulated splice variants and novel exons in glial brain tumors using exon expression arrays. *Cancer Res* **67:** 5635–5642.

Gardina PJ, Clark TA, Shimada B, Staples MK, Yang Q, Veitch J, Schweitzer A, Awad T, Sugnet C, Dee S, et al. 2006. Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. *BMC Genomics* **7:** 325. doi: 10.1186/1471-2164-7-325.

Goldstrohm AC, Greenleaf AL, Garcia-Blanco MA. 2001. Co-transcriptional splicing of pre-messenger RNAs: considerations for the mechanism of alternative splicing. *Gene* **277:** 31–47.

Gurtan AM, D'Andrea AD. 2006. Dedicated to the core: understanding the Fanconi anemia complex. *DNA Repair (Amst)* **5:** 1119–1125.

Hirose Y, Manley JL. 2000. RNA polymerase II and the integration of nuclear events. *Genes Dev* **14:** 1415–1429.

Hung LH, Heiner M, Hui J, Schreiner S, Benes V, Bindereif A. 2008. Diverse roles of hnRNP L in mammalian mRNA processing: a combined microarray and RNAi analysis. *RNA* **14:** 284–296.

Johnson NL, Kotz S, Kemp AW. 1992. *Univariate discrete distributions*, 2nd ed. Wiley, New York.

Kameoka S, Duque P, Konarska MM. 2004. p54(nrb) associates with the 5′ splice site within large transcription/splicing complexes. *EMBO J* **23:** 1782–1791.

Kim JB, Yamaguchi Y, Wada T, Handa H, Sharp PA. 1999. Tat-SF1 protein associates with RAP30 and human SPT5 proteins. *Mol Cell Biol* **19:** 5960–5968.

Kornblihtt AR, de la Mata M, Fededa JP, Munoz MJ, Nogues G. 2004. Multiple links between transcription and splicing. *RNA* **10:** 1489–1498.

Laajala E, Aittokallio T, Lahesmaa R, Elo LL. 2009. Probe-level estimation improves the detection of differential splicing in Affymetrix exon array studies. *Genome Biol* **10:** R77. doi: 10-1186/gb-2009-10-7-r77.

Levitus M, Joenje H, de Winter JP. 2006. The Fanconi anemia pathway of genomic maintenance. *Cell Oncol* **28:** 3–29.

Li XY, Green MR. 1998. The HIV-1 Tat cellular coactivator Tat-SF1 is a general transcription elongation factor. *Genes Dev* **12:** 2992–2996.

Maniatis T, Reed R. 2002. An extensive network of coupling among gene expression machines. *Nature* **416:** 499–506.

McKee AE, Neretti N, Carvalho LE, Meyer CA, Fox EA, Brodsky AS, Silver PA. 2007. Exon expression profiling reveals stimulus-mediated exon use in neural cells. *Genome Biol* **8:** R159. doi: 10-1186/gb-2007-8-8-r159.

Miller HB, Saunders KO, Tomaras GD, Garcia-Blanco MA. 2009. Tat-SF1 is not required for Tat transactivation but does regulate the relative levels of unspliced and spliced HIV-1 RNAs. *PLoS ONE* **4:** e5710. doi: 10.1371/journal.pone.0005710.

Montgomery DC. 2006. *Design and analysis of experiments*, 5th ed. Wiley and Sons, Hoboken, NJ.

Moore MJ, Silver PA. 2008. Global analysis of mRNA splicing. *RNA* **14:** 197–203.

Neugebauer KM. 2002. On the importance of being co-transcriptional. *J Cell Sci* **115:** 3865–3871.

Oberdoerffer S, Moita LF, Neems D, Freitas RP, Hacohen N, Rao A. 2008. Regulation of CD45 alternative splicing by heterogeneous ribonucleoprotein, hnRNPLL. *Science* **321:** 686–691.

Orphanides G, Reinberg D. 2002. A unified theory of gene expression. *Cell* **108:** 439–451.

Pandit S, Wang D, Fu XD. 2008. Functional integration of transcriptional and RNA processing machineries. *Curr Opin Cell Biol* **20:** 260–265.

Parada CA, Roeder RG. 1999. A novel RNA polymerase II-containing complex potentiates Tat-enhanced HIV-1 transcription. *EMBO J* **18:** 3688–3701.

Patel KJ, Joenje H. 2007. Fanconi anemia and DNA replication repair. *DNA Repair (Amst)* **6:** 885–890.

Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29:** e45. doi: 10.1093/nar/29.9.e45.

Pfaffl MW, Horgan GW, Dempfle L. 2002. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res* **30:** e36. doi: 10.1093/nar/30.9.e36.

Proudfoot NJ, Furger A, Dye MJ. 2002. Integrating mRNA processing with transcription. *Cell* **108:** 501–512.

Robinson MD, Speed TP. 2007. A comparison of Affymetrix gene expression arrays. *BMC Bioinformatics* **8:** 449. doi: 10.1186/1471-2105-8-449.

Robinson TJ, Dinan MA, Dewhirst M, Garcia-Blanco MA, Pearson JL. 2010. SplicerAV: a tool for mining microarray expression data for changes in RNA processing. *BMC Bioinformatics* **11:** 108. doi: 10.1186/1471-2105-11-108

Sanchez-Alvarez M, Goldstrohm AC, Garcia-Blanco MA, Sune C. 2006. Human transcription elongation factor CA150 localizes to splicing factor-rich nuclear speckles and assembles transcription and splicing components into complexes through its amino and carboxyl regions. *Mol Cell Biol* **26:** 4998–5014.

Smith MJ, Kulkarni S, Pawson T. 2004. FF domains of CA150 bind transcription and splicing factors through multiple weak interactions. *Mol Cell Biol* **24:** 9274–9285.

Sun J, Li R. 2009. Human negative elongation factor activates transcription and regulates alternative transcription initiation. *J Biol Chem* **285:** 6443–6452.

Tan S, Aso T, Conaway RC, Conaway JW. 1994. Roles for both the RAP30 and RAP74 subunits of transcription factor IIF in transcription initiation and elongation by RNA polymerase II. *J Biol Chem* **269:** 25684–25691.

Tan S, Conaway RC, Conaway JW. 1995. Dissection of transcription factor TFIIF functional domains required for initiation and elongation. *Proc Natl Acad Sci* **92:** 6042–6046.

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* **35:** W71-W74.

Wang W. 2007. Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins. *Nat Rev Genet* **8:** 735–748.

Warzecha CC, Shen S, Xing Y, Carstens RP. 2009. The epithelial splicing factors ESRP1 and ESRP2 positively and negatively

regulate diverse types of alternative splicing events. *RNA Biol* **6:** 546–562.

Wu-Baer F, Lane WS, Gaynor RB. 1998. Role of the human homolog of the yeast transcription factor SPT5 in HIV-1 Tat-activation. *J Mol Biol* **277:** 179–197.

Xing Y, Stoilov P, Kapur K, Han A, Jiang H, Shen S, Black DL, Wong WH. 2008. MADS: a new and improved method for analysis of differential alternative splicing by exon-tiling microarrays. *RNA* **14:** 1470–1479.

Yan D, Perriman R, Igel H, Howe KJ, Neville M, Ares M Jr. 1998. CUS2, a yeast homolog of human Tat-SF1, rescues function of

misfolded U2 through an unusual RNA recognition motif. *Mol Cell Biol* **18:** 5000–5009.

Zhou Q, Sharp PA. 1995. Novel mechanism and factor for regulation by HIV-1 Tat. *EMBO J* **14:** 321–328.

Zhou Q, Sharp PA. 1996. Tat-SF1: cofactor for stimulation of transcriptional elongation by HIV-1 Tat. *Science* **274:** 605–610.

Zhou Q, Chen D, Pierstorff E, Luo K. 1998. Transcription elongation factor P-TEFb mediates Tat activation of HIV-1 transcription at multiple stages. *EMBO J* **17:** 3681–3691.