Published online 21 December 2018

Nucleic Acids Research, 2019, Vol. 47, No. 4 ***e24***
doi: 10.1093/nar/gky1288

A cell-based splicing reporter system to identify regulators of cis-splicing between adjacent genes

Katarzyna Chwalenia[1,†], Fujun Qin[1,†], Sandeep Singh[1] and Hui Li[1,2,3,*]

[1]Department of Pathology, School of Medicine, University of Virginia, Charlottesville, VA 22908, USA, [2]Department of Biochemistry and Molecular Genetics, School of Medicine, University of Virginia, Charlottesville, VA 22908, USA and [3]School of Basic Medical Sciences, Zhengzhou University, Zhengzhou, Henan 450001, China

Received May 04, 2018; Revised November 14, 2018; Editorial Decision December 13, 2018; Accepted December 14, 2018

ABSTRACT

Chimeric RNAs generated by cis-splicing between adjacent genes (cis-SAGe) are increasingly recognized as a widespread phenomenon. These chimeric messenger RNAs are present in normal human cells, and are also detected in various cancers. The mechanisms for how this group of chimeras is formed are not yet clear, in part due to the lack of a tractable system for their experimental investigation. Here we developed a fast, easy and versatile cell-based reporter system to identify regulators of cis-SAGe. The reporter, consisting of four main cassettes, simultaneously measures the effects of a candidate regulator on cis-SAGe and canonical splicing. Using this cell-based assay, we screened 102 candidate factors involved in RNA pol II cleavage and termination, elongation, splicing, alternative splicing and R-loop formation. We discovered that two factors, SRRM1 and SF3B1, affect not only cis-SAGe chimeras, but also other types of chimeric RNAs in a genome-wide fashion. This system can be used for studying trans-acting factors and cis-acting sequence elements and factors, as well as for screening small molecule inhibitors.

INTRODUCTION

Chimeric RNAs resulting from cis-splicing between adjacent genes (cis-SAGe) are composed of exons from two distinct neighbor genes transcribed in the same direction. Historically, these chimeric RNAs have been called transcription-induced chimeras (1,2), tandem RNA chimeras (3), conjoined genes (4,5) and read-through fusions (6). To distinguish the intergenic cis-splicing from trans-splicing events, and to avoid confusion of the process of skipping the stop codon during translation (classic read-through), we prefer the term 'cis-splicing of adjacent genes' (7–11). They were once considered rare in mammalian cells, with only a handful of examples experimentally identified; however, they have more recently been found to be widely present in numerous cells and tissues (6,8,10–13). They may also be misregulated in cancer, and thus represent an underappreciated repertoire for cancer biomarkers (14–18). Even though the number of discovered cis-SAGe chimeric RNAs keeps increasing, the mechanism for their formation is poorly understood. We have hypothesized that at least three conditions must be met for cis-SAGe to occur: (i) the primary transcript of the upstream gene has to be active; (ii) the primary transcript has to pass through the gene boundary and read into the downstream gene; and (iii) alternative splicing must be allowed, as most cis-SAGe fusions tend to skip the last exon(s) of the 5′ gene and the beginning exon(s) of the 3′ gene (7). Specific factors, such as CCCTC-binding factor (CTCF), which binds to the insulators between neighboring genes and have been shown to affect at least some cis-SAGe chimeric RNAs (7). However, no systematic approach to identify modulators of cis-SAGe events has yet been developed.

In this method paper, we describe a novel, efficient and easy reporter system to identify potential factors that may regulate the cis-SAGe process. We modeled the reporter system after a widely expressed cis-SAGe RNA, *CTNNBIP1-CLSTN1*, and inserted fragments of the parental genes in between the two pieces of renilla luciferase open reading frame (ORF). Only when cis-SAGe occurs will a mature renilla luciferase be expressed. As a control for regular splicing, we built in a firefly luciferase ORF, interrupted only by a β-globin/immunoglobulin intron. Using a custom small interfering RNA (siRNA) library from Dharmacon SMARTpool, we screened possible factors previously reported in the literature that are involved in RNA pol II cleavage, termination, elongation, splicing, alternative splicing and other RNA processings. Two RNA processing factors, SF3B1 and SRRM1, were identified to have preferential effects on renilla over firefly luciferase, thus implicating potential function in cis-SAGe events. Global impacts on cis-SAGe chimeric RNAs were also investigated by RNA-sequencing cells in which these two factors were perturbed.

*To whom correspondence should be addressed. Tel: +1 434 982 6680; Fax: +1 434 243 7244; Email: hl9r@virginia.edu
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

© The Author(s) 2018. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

MATERIALS AND METHODS

Cell culture

HEK293T cells were maintained in Dulbecco's Modified Eagle's Medium with 4500 mg/l glucose (Gibco), supplemented with 10% fetal bovine serum and 1% Pen/Strep solution (Hyclone). Cells were cultivated at 37°C in 5% CO_2 humidity.

cis-SAGe reporter system construction and stable cell line derivation

An out-of-frame renilla luciferase sequence driven by an EF1α promoter was split and 100 bp of the beginning and end on intron5, and exon6 of the *CTNNBIP1* gene, followed by 100 bp of the beginning and end of exon1, and intron1 of the *CLSTN1* gene, driven by the EF1α promoter introduced. The described construct was cloned into a pGL4.16-CMV-LUC2CP/intron/ARE backbone plasmid, which contains a split firefly luciferase ORF separated by a hybrid β-globin/Immunoglobulin intron (19). HEK 293T cells were transfected with the reporter construct using PEI (Fisher Scientific), and selected using hygromycin for 2 weeks.

siRNA transfection

Custom siRNA SMARTpool libraries were ordered from Dharmacon™, and the transfection was performed using Lipofectamine® RNAiMAX Transfection Reagent (Thermo Fisher Scientific) according to the manufacturer's instruction.

RNA extraction and qRT-PCR

RNA was extracted from cell lines using TRIzol Reagent (Life Technologies) according to the manufacturer's instruction. RNA samples were analyzed on a NanoDrop (Thermo Scientific), and 3 ug RNA was used for complementary DNA (cDNA) synthesis. All RNA samples in this study were treated with DNAse I (NEB), followed by standard reverse transcription using the SensiFAST cDNA Synthesis Kit (Bioline) according to the manufacturer's instructions. Specific primers (Supplementary Table S2) were used to detect RNA expression in the corresponding cDNA samples. Quantitative reverse transcription polymerase chain reaction (qRT-PCR) was performed as described previously (7,8,16).

Luciferase assay

Firefly and renilla luciferase activities were measured using the Dual-Luciferase® Reporter Assay System (Promega). Cells were lysed using 500 μl 1× Passive Lysis Buffer diluted in phosphate-buffered saline according to the vendor's instructions. Cell lysates were collected and briefly centrifuged to get rid of debris. The assay was conducted according to the manufacturer's instruction.

RNA-sequencing, bioinformatics and statistics

HEK293T cells transfected with siCT, si*SF3B1* and si*SRRM1* were harvested 2 days after transfection. RNAs were extracted and purified. Raw RNA-Seq reads were filtered to obtain high-quality reads using NGSQC toolkit (20). The EricScript software (version 0.5.5) (21) with default parameters was used to predict gene fusion events using high-quality filtered reads as input. Chimeric RNAs with high confidence of prediction based on EricScore (≥0.6) were retained. We filtered out M/M fusions, which may represent artifacts due to template switching (22). The output files from EricScript software were processed using in-house perl scripts. Circos plots were made using the Circos software (version 0.69.6) (23). Evaluation of significance was performed using *t*-test.

RESULTS

A model reporter system

CTNNBIP1-CLSTN1 is a chimeric RNA frequently detected in multiple tissues and cells (22). The two neighboring parental genes are located on chr.1p36, and are separated by ∼24 kb. They both transcribe in the same direction, making them candidates for cis-SAGe. The chimeric RNA retains the first 5 exons of *CTNNBIP1* and last 17 exons of *CLSTN1,* and is induced via *CTCF* silencing (7). In addition, we could detect fragments of transcripts in-between the two genes, which is consistent with its cis-SAGe nature. To further confirm its read-through mechanism, we designed an assay to detect the precursor messenger RNA (mRNA) (Figure 1A). In this experiment, a reverse primer annealing to exon2 of *CLSTN1* was used to perform reverse transcription. We then used three primer pairs designed to amplify fragments of cDNA covering exon4 and intron4, exon5 and intron5 and exon6 and intron6 of *CTNNBIP1*. To eliminate potential DNA contamination, RNA was treated with DNaseI before the assay. To confirm that the signal was not due to remaining DNA contaminants, we included controls with no AMV-RT enzyme. Amplicons were only detected with the AMV-RT enzyme (Figure 1B), confirming that the *CTNNBIP1-CLSTN1* chimeric RNA is a product of cis-SAGe.

We then used the *CTNNBIP1-CLSTN1* chimeric RNA as a model to construct a dual-luciferase reporter (Figure 1C). The reporter contains four cassettes. The first cassette contains the 5′ portion of a renilla luciferase gene (Reni-) driven by an EFS promoter. The renilla coding sequence was split at a position that mimics the end of exon5 of *CTNNBIP1* (CCAGG). We added in 100 bp from the beginning and ending sequence of intron5, followed by 100 bp from the beginning and ending sequence of exon6 of *CTNNBIP1*. The ending sequence of exon6 contains the canonical polyadenylation signal sequence (AATAAA). The second cassette, which is driven by another EFS promoter, contains 100 bp from the beginning and ending sequence of *CLSTN1* exon1, and 100 bp from the beginning and ending sequences of intron1, followed by the 3′ section of the renilla luciferase gene. After the renilla coding sequence, we inserted three sets of 'ATTTA' RNA degradation sequence to prevent measurement of accumu-

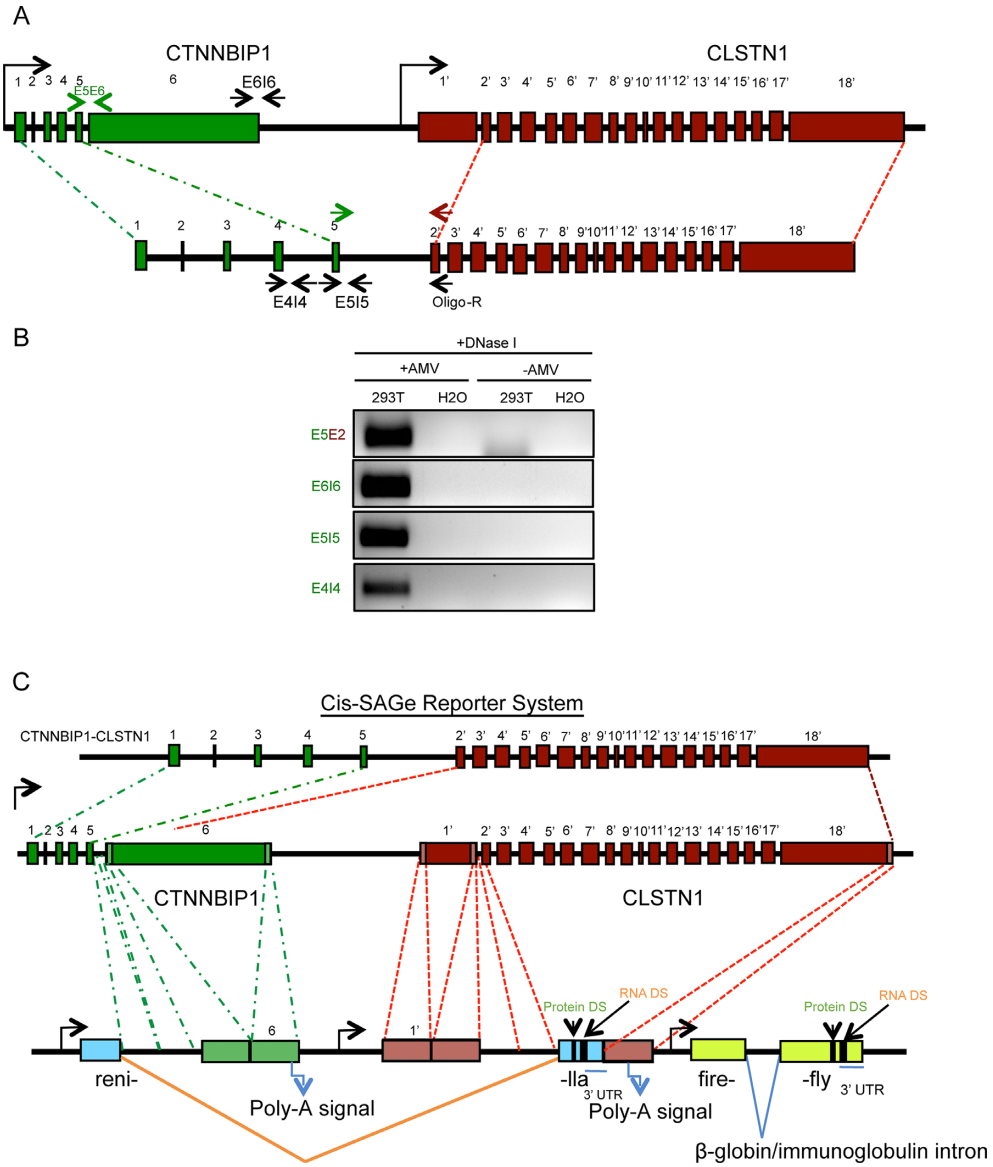


Figure 1. The reporter for cis-SAGe. (**A**) The *CTNNBIP1-CLSTN1* chimera is composed of the first 5 exons of *CTNNBIP1* and last 17 exons of *CLSTN1*. Reverse primer annealing to exon2 (E2) of *CLSTN1* was used for reverse transcription. For RT-PCR, pairs of specific primers spanning exon4 and intron4 (E4I4), exon5 and intron5 (E5I5), and exon6 and intron6 (E6I6) of *CTNNBIP1* were used to validate the presence of a precursor transcript covering both genes. (**B**) The *CTNNBIP1-CLSTN1* chimeric RNA is a cis-SAGe chimeric RNA. Signals were only seen in the samples with AMV-RT. DNA contamination was eliminated by DNaseI treatment before the assay (+DNaseI). Further controls without the AMV-RT enzyme confirmed that the signals were not due to remaining DNA contaminants. (**C**) The cis-SAGe reporter system modeled after the *CTNNBIP1-CLSTN1* chimeric RNA. The renilla luciferase ORF was interrupted by the terminal part of *CTNNBIP1* and the beginning of *CLSTN1*, and broken into reni- and –lla parts. When the cis-SAGe formation of *CTNNBIP1-CLSTN1* occurs, the intermittent region is spliced out, giving rise to mature, in-frame renilla transcripts (orange lines). A β-globin/immunoglobulin intron containing the firefly gene (fire- and -fly) was used as an internal control. Protein and RNA destabilization signals (RNA DS, and protein DS including CL1 and PEST) were introduced to enable quick response.

lation, followed by 100 bp of the end of *CLSTN1*, also containing the canonical polyadenylation signal sequence (AATAAA). The third cassette was borrowed from CMV-LUC2CP/intron/ARE plasmid (19), which includes a split firefly luciferase gene (Fire- and -fly), separated by a β -globin/Immunoglobulin intron that has been optimized to splice with high efficiency (24). This cassette is regulated under a CMV promoter, and also contains three copies of the 'ATTTA' RNA degradation signal, as well as an SV40 polyA sequence after the firefly luciferase ORF. The fourth cassette is a hygromycin resistant gene driven by an SV40 promoter and synthetic polyA sequence (not shown in the figure). At the C terminus of both renilla (cassette2) and firefly (cassette3) protein-coding sequences, we inserted protein degradation sequences (DS) (protein DS CL1 and PEST), preventing measurement of accumulation and rendering the system for a quick response. Restriction enzyme sites were also built in front of and after cassettes one and two to allow changes, or for inserting additional elements. As we have shown that *CTCF* silencing induced *CTNNBIP1-CLSTN1*

and that the *CTCF* binding insulator sequence is too far from the parental genes; therefore, we did not include the insulator sequence in this construct.

The 5′ part of the renilla luciferase fragment is out of its reading frame. In the absence of transcriptional read-through and splicing-out of the intermittent fragments, the renilla luciferase will not be expressed. Firefly luciferase serves as an internal control for regular splicing. The system can be applied to all cells. As a proof of principle, we selected human embryonic kidney cells (HEK293T) cells because of their high-transfection rate. HEK293T cells were transfected with the reporter plasmid. Cells stably expressing the plasmid were selected with hygromycin. Since we aim to identify modulators, which induce or suppress cis-SAGe events, the system must have a basal level of expression for reference. Indeed, we observed a low-level expression of renilla luciferase (Figure 2A), and a higher level of firefly luciferase (Figure 2B). Consistently, qRT-PCR detected the expression of both luciferase genes at the RNA level (Supplementary Figure S1).

Screening for putative cis-SAGe trans-acting regulators

Even though several hypotheses have been proposed (5,25–28), the detailed mechanism for cis-SAGe formation is currently unknown. Specific factors regulating the process are yet to be discovered. To identify trans-acting regulators that play a role in cis-SAGe regulation, we conducted an extensive literature search, and identified potential contributing factors. These include SR proteins important for constitutive and alternative splicing (19,29,30), intronic and exonic splicing enhancers (ESEs) (30), RNA polymerase II cleavage and termination (31–34) and elongation factors (35,36), as well as other alternative splicing regulators (30,37–41). In addition, we included proteins involved in R-loop formation and dissolution, as the R-loop structure has been reported to play a role in transcriptional termination and splicing (42). In total, over 100 candidates were selected (Supplementary Table S1). We then built a custom siRNA library using Dharmacon SMARTpool siRNAs, which contains four different siRNA duplexes for each target. The siRNA library screening was performed using the Dual-Luciferase Reporter Assay.

Based on the ratio of renilla and firefly luciferase readout in comparison to the control (siGENOME non-targeting control), the factors can be grouped into three categories by using a 1.5-fold cut-off: unchanged, induced and repressed (Figure 2C). Interestingly, only one factor, SRRM1, was grouped into the repressed category. Since we aimed to identify specific factors that regulate cis-SAGe, but not canonical splicing, we focused on factors that induced or repressed renilla luciferase, with no significant changes in firefly luciferase.

Cleavage and polyadenylation specificity factors (CPSF), as well as cleavage stimulation factor subunits (CSTF) are critical for the cleavage of freshly synthesized pre-mRNA (43). Therefore, their silencing could result in RNA polymerase II reading-through gene boundaries. In our experimental setup, the knockdown of several CPSF family members (Figure 2D) exerted a variable effect on cis-SAGe events. *CPSF1* knockdown resulted in the highest increase in the renilla/firefly ratio, although this change can be mostly attributed to reduced firefly expression. In the context of the CSTF family, only the *CSTF3* knock-down contributed to a significant change in the renilla/firefly ratio, also mainly due to changes in firefly activity (Figure 2D). *CSTF1* knock-down resulted in a induction of both renilla and firefly expression (Figure 2D), suggesting that the silencing of *CSTF1* affects both canonical splicing, as well as cis-SAGe. Silencing of Senataxin (*SETX*), which is involved in R-loop resolution and transcription termination (42), showed no significant change at any level of measurement (ratio, renilla or firefly signals) (Figure 2E).

Two factors caught our attention and became the focuses of our downstream study: *SF3B1*, the factor that upon silencing had the highest fold change in renilla, with no obvious change in firefly (Figure 2C and G); and *SRRM1*, which is the only factor that upon silencing repressed renilla, with no apparent change in firefly (Figure 2C and F).

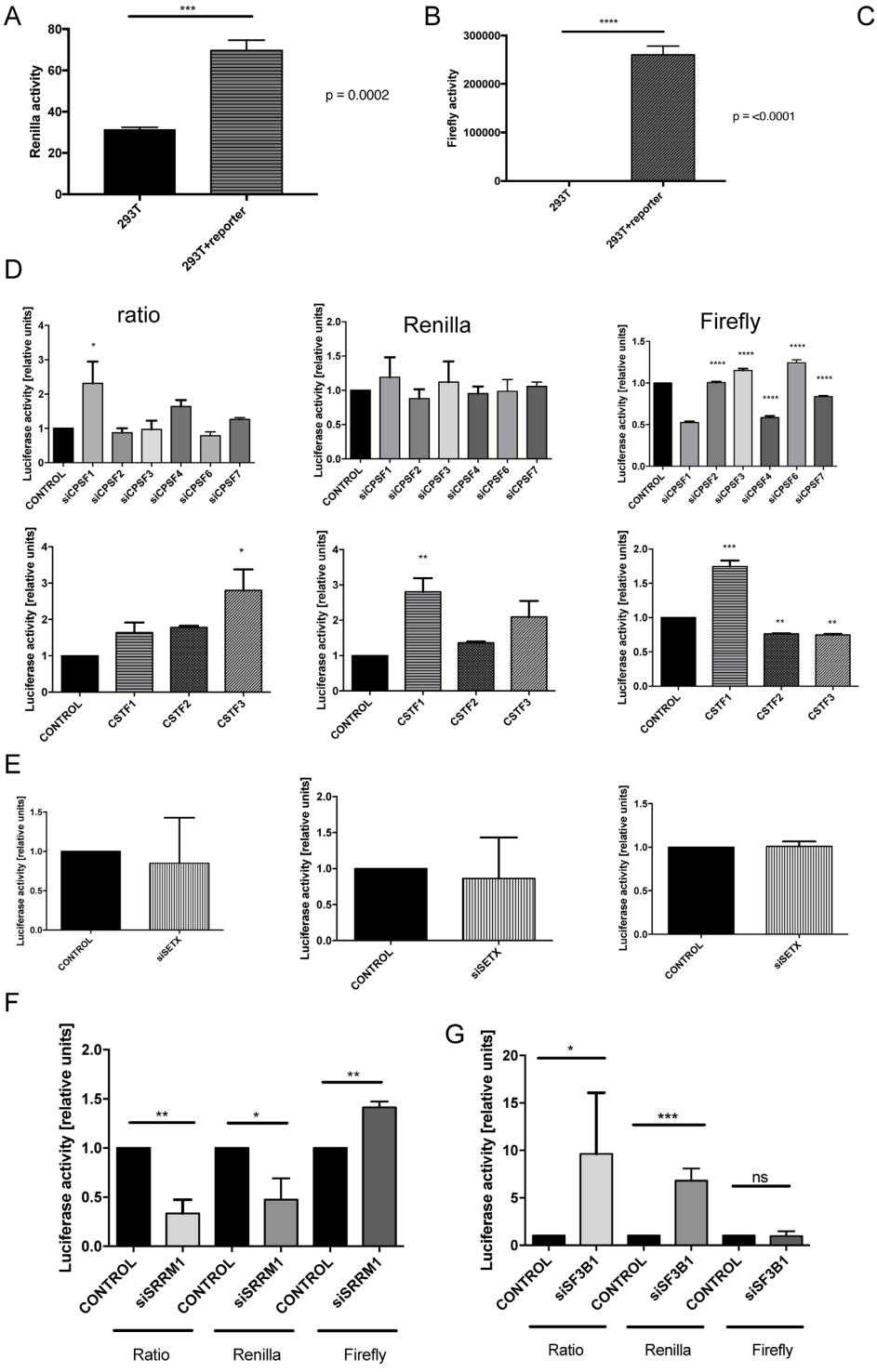
Effect of silencing SF3B1 and SRRM1 on cis-SAGe

To confirm that the changes do happen at the RNA level, we extracted RNAs from cells transfected with SMARTpool siRNAs targeting *SF3B1* and *SRRM1*. Both genes were effectively silenced (Supplementary Figure S2). Consistent with the results of the dual-luciferase assay, knock-down of *SRRM1* dramatically reduced the mRNA levels of the renilla transcript, and to a much lesser extent the firefly transcript (Figure 3A). In *SF3B1* knocked down cells, the mRNA level of renilla luciferase was significantly upregulated, with only a slight increase in firefly gene expression (Figure 3B).

We then examined the effect of silencing these two factors on cis-SAGe chimeric RNAs. We selected a few widely expressed cis-SAGe fusions, including: *CTNNBIP1-CLSTN1*, *DUS4L-BCAP29*, *CLN6-CALML*, *UBA2-WTIP* and *SLC29A1-HSP90AB1*, based on literatures (9–11,13,16,22,44,45), and measured their fusion RNA levels relative to internal control *GAPDH* (Figure 3C and D). Silencing *SRRM1* resulted in a reduction of the majority of five cis-SAGe fusion RNAs, consistent with the luciferase assay of si*SRRM1*. However, si*SF3B1* disagreed with its results from the luciferase assay. This may be due to the missing cis-acting sequence elements in the reporter system, or perhaps, more candidates need to be tested. To examine the effect on the alternative choice between canonical splicing and cis-SAGe for the 5′ parental gene, we compared the effect of the siRNAs on the chimeras relative to the wild-type of the 5′ parental genes, we found that three chimeras were favourably downregulated by siSRRM1, and three were upregulated by siSF3B1 (Supplementary Figure S3).

To test whether the effect on cis-SAGe chimeric RNAs is also true on other types of chimeric RNAs, we examined one interchromosomal chimeric RNA, *C15orf57-CBX3* and one intrachromosomal chimeric RNA with one gene in between, *CTBS-GNG5* (22). si*SRRM1* also repressed these two chimeras (Figure 3E), whereas si*SF3B1* had mixed results (Figure 3F).

The same results were observed from biological repetitive knocking down of *SRRM1* and *SF3B1* (Supplemen-



	ratio	renilla	firefly
CONTROL	1.0	1.0	1.0
CPSF6	0.8*	1.0	1.2
CPSF7	1.3	1.2	0.8*
CSTF3	2.8	2.1	0.7**
CTR9	3.4	2.1	0.6**
DICER	6.6	3.3	0.5***
FUS	5.7	1.9	0.7
RBFOX2	1.6*	1.9**	1.2
SF3B1	4.7*	6.7*	1.4**
SNRNP200	2.5*	2.9*	1.1*
SRRM2	1.9	1.6	0.8**
SRSF2	1.7	2.2	1.3
SRSF3	2.1	2.9	1.4*
SRSF4	2.7	2.6	1.0
SRSF5	2.0	1.7	0.9*
SRSF9	1.8	1.6	0.9**
U2AF1	2.0***	1.8**	0.9
CSTF1	1.6	2.8	1.7*
DHX8	1.9	3.0	1.7*
DROSHA	7.4	1.9	0.3**
DXO	20.3*	1.7	0.1***
FIP1L1	8.9	3.6	0.5*
PAPOLA	7.9	3.7	0.5**
PCF11	27**	1.6	0.1***
WDR33	7.6	3.1	0.4**
AICDA	3.3*	1.4	0.4**
CHD1	8.4*	1.0	0.1***
CTCF	5.6**	1.2*	0.2***
HNRNPA1	6.4*	1.1	0.2**
MBNL1	3.3*	1.4*	0.4***
NCBP1	6.7*	1.4	0.2***
POLR2A	109.7**	1.1	0.0***
POLR2C	2.7	0.5	0.2***
RPAP1	1.7*	0.7	0.4**
SUPT5H	5.1*	0.7	0.1***
ASCC1	1.8	1.1	0.6*
CAMK2A	1.8	1.1	0.6*
CELF1	2	1.5	0.8**
CELF2	1.7	1.4	0.8*
CPSF1	2.3	1.2	0.5**
CPSF4	1.6	1.0	0.6**
CSTF2	1.8	1.4*	0.8**
DHX9	1.8	1.3	0.7*
HNRNPF	2.2*	1.5	0.7**
NUDT21	2.7	1.5	0.5**
PAF1	2.8*	1.5	0.5**
PRBP1	1.6*	1.3*	0.8*
RNASEH1	2.3	1.2	0.5**
SMN1	2.0	1.1	0.6**
SRSF11	2.3	1.3	0.6***
SUB1	2*	1.4	0.7**
SUPT4H1	1.6*	1.0	0.6**
XRN2	1.8	0.8	0.5***
SRRM1	0.3*	0.5	1.4**
NELFB	1.2	1.6*	1.4*
PRPF40B	1.2*	1.7*	1.3*
SRSF6	1.4	1.9	1.5
HNRNPL	1.3	2.8**	2.2**
POLR2JZ	0.6*	2.0	3.2**
DDX5	0.9	1.5	1.7**
SF1	0.5**	0.9	1.7**
SF3A1	0.6**	0.9	1.6***
SRSF1	0.9	1.4	1.6**
PRPF2	1.4	0.6	0.4
PRPF39	1.4*	0.6***	0.4***
RBFOX1	1.4*	0.6*	0.4***
RBM39	1.2	1.1	0.4
ASCC2	0.8	0.8	1.0
CPSF2	0.9	0.9	1.0
CPSF3	1.0	1.1	1.2
DDHX38	1.0	1.0	1.0
DDX39B	0.8	0.7	0.9
DDX46	0.6*	0.8	1.4**
ELAVL1	1.1	0.7*	0.7**
HNRNPH1	1.5	1.3	0.9
ITPR1	1.5*	1.5	1.0
MBNL2	1.0	0.7	0.7*
PRKCA	1.1	1.4	1.4
PRMT5	1.5	1.4	0.9
PRPF40A	0.8	0.7*	0.9
PRPF6	1.0	0.7**	0.7
RBFOX3	0.6***	0.5**	0.7*
RBM5	0.6**	0.5**	0.9*
RNPS1	1.3	1.2	0.9
RPAP2	1.1	1.2	1.1*
RTF1	1.2	1.3	1.1*
SCAF4	0.9	1.0	1.1**
SCAF8	0.9	0.7	0.8
SETD2	0.8*	0.6*	0.8
SETX	0.8	0.9	1.0
SFPQ	1.5***	1.5	1.0
SNRNP70	1.3	1.5	1.2**
SNRPA	1.1	0.7	0.6***
SNRPC	1.4	1.0	0.7**
SREK1	0.7	0.8	1.2*
SRSF10	1.2	1.0	0.8**
SRSF12	0.6	0.8	1.3**
SRSF7	0.7	0.7	1.0
SRSF8	1.5	1.2	0.8**
SSU72	1.5*	1.1	0.8*
TRA2A	1.2	1.3	1.0
U2AF2	1.5	1.1	0.7**

up
down
unchanged

Figure 2. Screening for trans-acting regulators of cis-SAGe. (**A** and **B**) The basal level expression of renilla and firefly luciferase in the cis-SAGe reporter system. HEK293T cells were transfected with the reporter construct and selected with hygromycin. The basal level of renilla (A) and firefly (**B**) luciferase was detected via the Dual-Luciferase Assay. (**C**) Cells stably expressing the reporter were transfected with siRNAs (SMARTpool Dharmacon library) targeting various, potential cis-SAGe candidates. Renilla and firefly luciferase, as well as their ratio, were measured and normalized against a control siRNA (siGENOME non-targeting control). Results are highlighted for a signal increase (red), decrease (green) and no change (gray) (1.5-fold as a cut-off). Detailed luciferase assay results are shown for the CPSF family (**D**) and Sentaxin (STEX) (**E**) as well as for *SRRM1* (**F**) and *SF3B1* (**G**). $*P < 0.05$, $**P < 0.01$, $***P < 0.001$

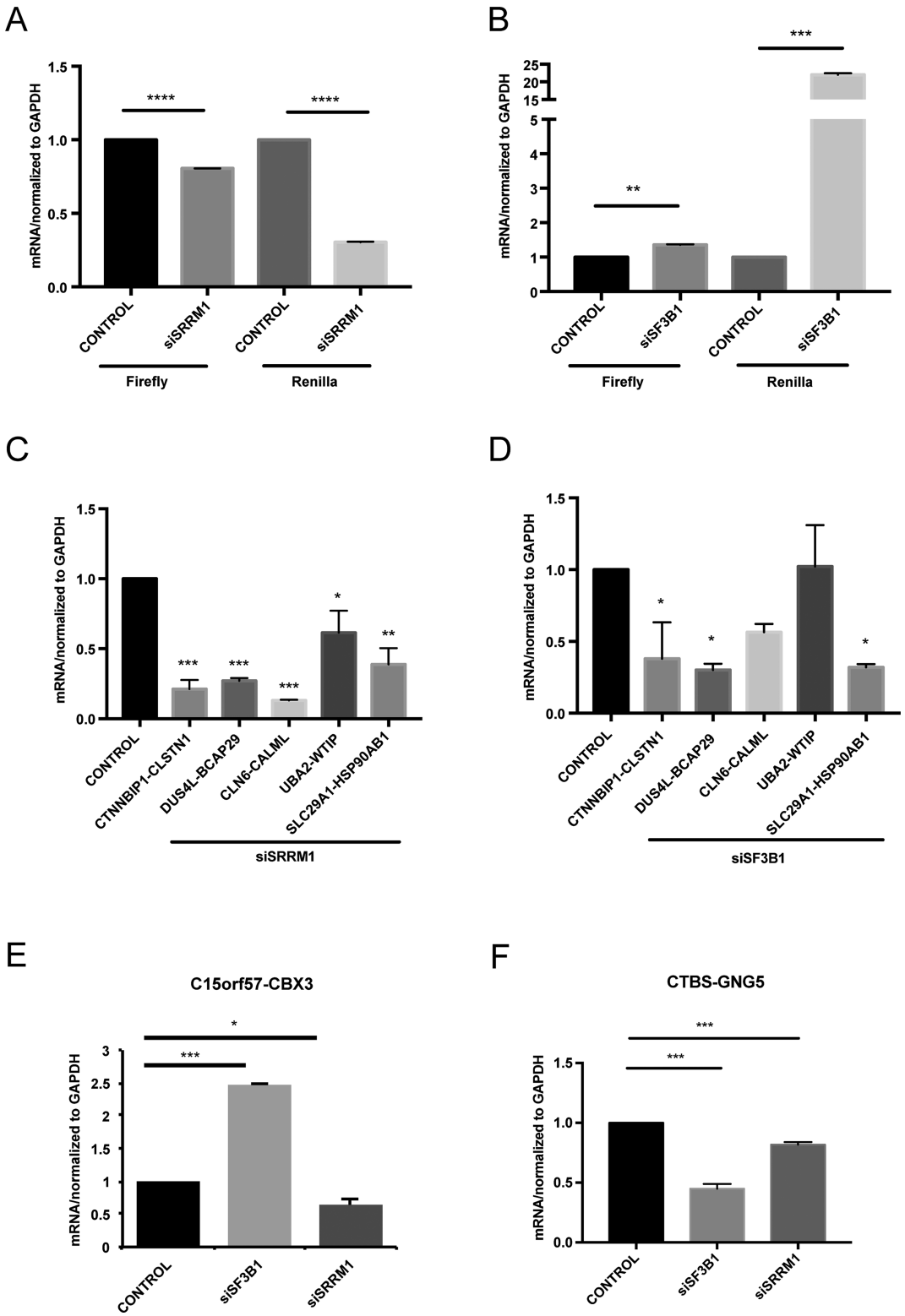


Figure 3. Validation of the effect of *SF3B1* and *SRRM1* on chimeric RNAs. (**A** and **B**) Changes in renilla and firefly luciferase at the mRNA level after *SRRM1* and *SF3B1* silencing. To confirm that the changes happen at the RNA level, we extracted RNAs from the cells transfected with SMARTpool siRNAs targeting *SRRM* and *SF3B1*. Knocking down of *SRRM1* reduced the mRNA levels of the renilla transcript, and to a much lesser extent the firefly transcript (A). *SF3B1* knocking down upregulated renilla luciferase expression with a slight difference in firefly transcript (B). (**C** and **D**) Expressions of several cis-SAGE chimeric RNAs were measured by qRT-PCR in *SRRM1* silencing (C) and *SF3B1* silencing (D) cells. (**E** and **F**) Interchromosomal chimeric RNA, *C15orf57-CBX3,* (E) and intra-others chimeric RNA, *CTBS-GNG5* (F) were downregulated by *SRRM1* silencing, while *SF3B1* knocking down gave mixed results. Expressions of the target transcripts were normalized against that of the internal control, *GAPDH*, and further normalized to the level in the control siRNA transfected cells. $*P < 0.05$, $** P < 0.01$, $***P < 0.001$

tary Figure S4A and B). Additionally, various effects on the *CTNNBIP1-CLSTN1* fusion RNA were investigated by siRNAs against different factors including those involved in splicing, chromatin remodeling and other factors (Supplementary Figure S5).

Genome-wide effect on chimeric RNAs upon silencing of *SRRM1* and *SF3B1*

We performed paired-end RNA-Seq to evaluate the silencing of *SSRM1* and *SF3B* in HEK293T cells on chimeric RNAs throughout the genome. Chimeric RNAs identified by EricScript software (21) were divided into three groups based on the chromosomal location of the parental genes: inter-chr, two parental genes located on separate chromosomes; intra-ss-0gap, two parental genes located on the same chromosome, transcribing in the same direction, and with no genes in between; and intra-others, whose parental genes are on the same chromosome, but with genes in between, or are transcribing in the opposite direction (Figure 4A). The chimeras of intra-ss-0gap are candidates of cis-SAGe. Two biological repeats were conducted. Consistent with the luciferase assay and qRT-PCR validation, majority of chimeric RNAs were downregulated across all three categories in si*SRRM1*, including all 21 intra-ss-0gap chimeric RNAs, 28 out of 37 inter-chr and eight out of 11 intra-others (Figure 4B). In contrast, majority of chimeric RNAs were upregulated in siSF3B1, most obvious for intra-ss-0gap (34 out of 49), less obvious in intra-others (14 out of 23), and no change in inter-chr (19 out of 39) (Figure 4C). We selected two intra-ss-0gap chimeras in each group, and confirmed their trend by qRT-PCR (Supplementary Figure S6).

DISCUSSION

Chimeric RNAs, composed of transcripts from neighboring genes are becoming increasingly recognized as a widespread phenomenon (1–3,6–11,16,17,46,47). Their formation contributes not only to transcriptome diversity (1,2), but also plays a significant role in cellular homeostasis (1,27,30,31). To date, several possible mechanisms of cis-SAGe have been postulated. One of them proposes mutation in polyadenylation signals, allowing transcription machinery to read through gene boundaries (5,26). Others focus on transcriptional slippage, caused by short homologous sequences (48), transcription anti-termination (27), or the torsional stress of DNA strands (49). However, the exact mechanism of cis-SAGe chimeras formation remains elusive. To facilitate high-throughput screening and identification of cis-SAGe modulators, we created a dual-luciferase reporter system; a system in which the renilla luciferase ORF was interrupted by portions of the genes, *CTNNBIP1* and *CLSTN1*, and used an intron-containing firefly luciferase ORF as a control. Extensive screening of a variety of RNA processing regulators including SR proteins (19,29,30), intronic and ESEs (30), RNA polymerase II cleavage and termination (31–34), elongation factors (35,36) and different alternative splicing regulators (30,37–41) allowed us to identify trans-acting factors, particularly those regulating cis-SAGe events. This system can also be used to evaluate cis-acting factors, including specific sequences or motifs. In addition, the system can be used to assess environmental factors, viral infections, and even to screen small molecule inhibitors. For example, a recent study revealed the induction of transcriptions downstream of genes (DoGs) under osmotic stress (18,50,51), which may contribute to the regulation of cis-SAGe chimeric RNAs. We built in additional RNA and protein destabilization sequences, ideal for studying the effect of such osmotic stress, which often requires a rapid readout within a few hours. These features are also critical when used for small molecule screening. Here, we used HEK293T cells as a proof of concept, but the system can be readily applied to other cells. We envisioned a typical flow of experiments using our system to investigate the perturbation of candidate factors on cis-SAGe chimeric RNA(s) (Supplementary Figure S7).

We anticipate that some of our findings may be restricted to the *CTNNBIP1-CLSTN1* fusion, we used as the model. However, the results from RNA-sequencing on genome-wide chimeric RNAs are consistent with the luciferase assays, i.e. *SRRM1* silencing reduced, and *SF3B1* silencing enhanced chimeric RNAs.

SRRM1 (also known as Srm160) (30,39,52) plays a role in splicing co-activation (53). *SRRM1* is important for constitutive, as well as ESE dependent, splicing to take place (54,55). We found that silencing SRRM1 resulted in a significant decrease of renilla luciferase in the reporter construct. The same trend was confirmed with five cis-SAGe RNAs. Interestingly, two chimeric RNAs from other categories were also suppressed upon its silencing. Consistently, RNA-Seq data analysis also revealed that silencing *SRRM1* greatly reduced all three groups of chimeric RNAs, especially in the category of cis-SAGe candidates, suggesting its ubiquitous role in regulating chimeric RNAs.

SF3B1 is a splicing factor, commonly mutated in cancer cells (56–58). Knocking it down greatly enhanced renilla luciferase expression in our reporter construct. However, the trend was not observed in the few chimeric RNAs we selected for further validation. This inconsistency may be due to the missing sequence elements in our reporter system. Nevertheless, RNA-Seq data obtained from cells treated with si*SF3B1* showed more upregulation in cis-SAGe chimeras, while had smaller effect on the chimeras of the other two categories. *SF3B1,* as a part of the U2 component of the spliceosome, plays a role in splice site recognition (59). It was shown that *SF3B1* knockdown has a significant impact on splicing, and induces alternative splice site choice, exon skipping or intron retention in myeloid cell lines (57). Moreover, dissociation of *SF3B1* from the U2 complex also affects alternative splicing (60). In the context of our study, lack of *SF3B1* may direct cells toward increased transcription DoGs, followed by aberrant splice site choice resulting in increased cis-SAGe chimeras occurrence. Most interestingly, we observed an obvious increase of *SRRM1* expression in all three siRNAs specifically targeting *SF3B1*, but no obvious changes on the expression of *SF3B1* under the depletion of *SRRM1* (Supplementary Figure S8A and B). This finding indicates that in some situations, *SRRM1* might be a downstream target of *SF3B1*, consistent with the opposite effect on at least some chimeric fusion RNAs. Inhibition of *SF3B1,* resulting in

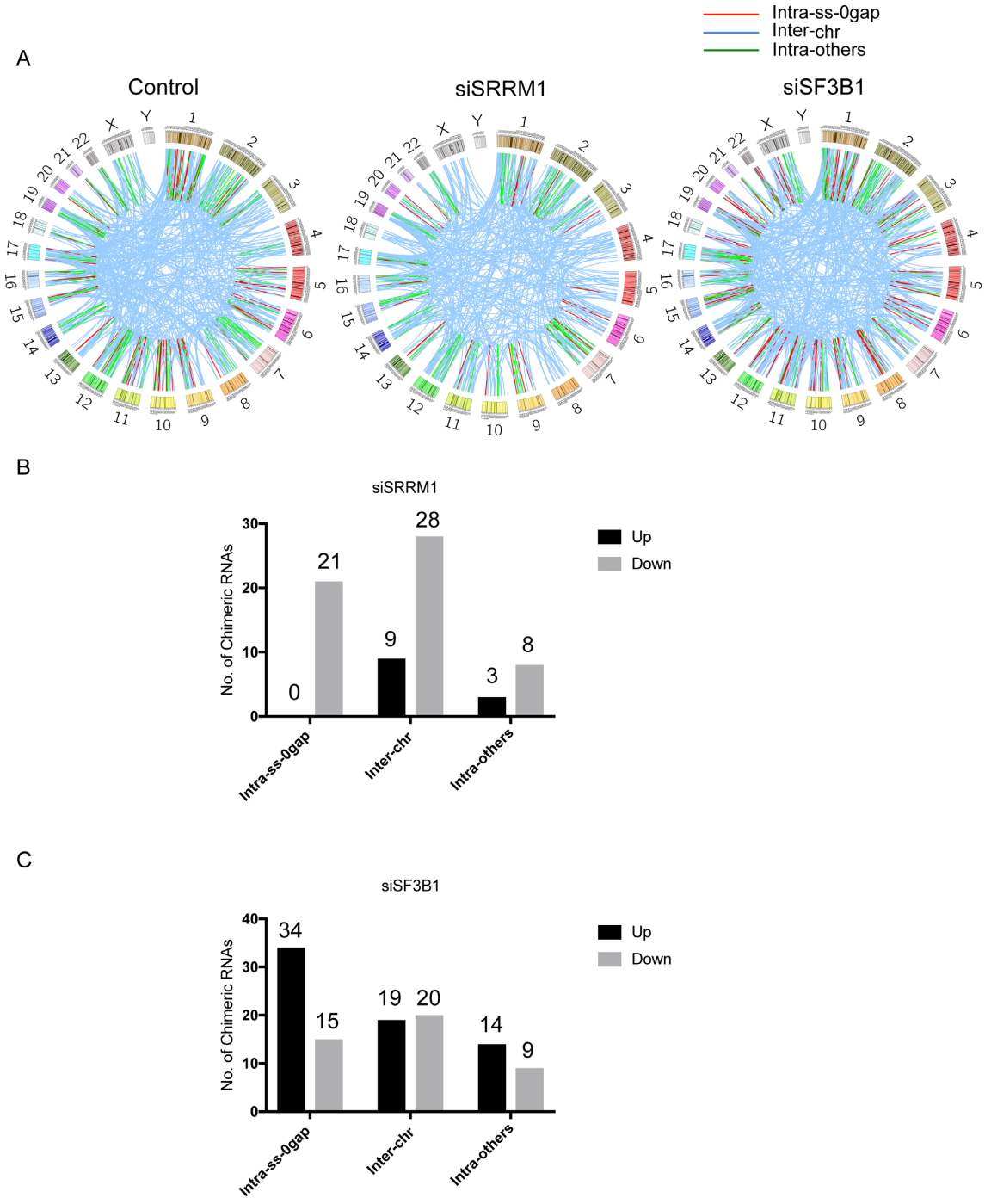


Figure 4. The genome-wide effect caused by *SRRM1*, and *SF3B1* silencing. Paired-end RNA-Seq data evaluation of *SRRM1*, and *SF3B1* silencing. Chimeric RNAs analyzed by the EricScript software were divided into three groups: Inter-chr, intra-ss-0gap and intra-others. (**A**) Circos plots depict the landscape of chimeric RNAs. Parental genes involved in forming a chimeric RNA are joined by a line. (**B**) (**C**) Histograms showing the number of chimeric RNAs up and downregulated in each of the three categories in siSRRM1 transfected cells (B), or siSF3B1 transfected cells (C) compared with control siRNA transfected cells. Two-fold cut-off was used.

massive aberrant exon shipping (61) and intron retention, can increase the 'read-through' between exons, and could potentially be the cause of enhanced level of cis-SAGe fusion RNAs (Supplementary Figure S8C). Given its frequent misregulation in cancer, it is possible that cancers with the *SF3B1* mutation may have abnormal cis-SAGe fusion RNA profiles. Our system can easily be used to study the effects of the particular *SF3B1* mutant on cis-SAGe RNAs.

We focused on the group of factors that induced, or repressed the renilla luciferase specifically, with no significant changes in the firefly luciferase level. However, the assumption that such factors only regulate cis-SAGe, but not regular splicing, may miss some genuine regulators of cis-SAGe. For instance, reduced cleavage may contribute to transcriptional read-through. One example is the cleavage stimulation factor, *CSTF1*, which was not investigated further because silencing it induced both renilla and firefly luciferases. On the other hand, if a factor affects renilla and firefly luciferase expression cassettes differently besides splicing, it would also result in false positive discoveries. This is the reason that further validation is always needed.

CONCLUSIONS

Despite the recognition of cis-SAGe chimeric RNAs as widespread natural phenomena and their implications for cancer, the mechanisms of their generation are far from being clear. We have developed a rapid-response, easy and versatile cell-based system that can be used to study the cis-SAGe process, as well as identified potential regulators for the process. Using this system, we discovered that two factors, *SRRM1* and *SF3B1*, affect cis-SAGe in opposite directions. Consistently, the genome-wide study also revealed the same trend.

DATA AVAILABILITY

The raw and processed RNA-sequencing data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE113101.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Loryn Facemire, Justin Elfman and Emily Lin for the English editing. The high-performance computing systems and services used in this work were provided by the Data Science Institute and the other Computation and Data Resource Exchange (CADRE) partner organizations at the University of Virginia.

Authors' contributions: K.C. and F.Q. carried out experiments, conducted the analyses, interpreted results and wrote and/or revised manuscript. S.S. conducted bioinformatic analyses. H.L. conceived and supervised the project and manuscript.

FUNDING

Funding for open access charge: Startup

Conflict of interest statement. None declared.

REFERENCES

1. Akiva,P., Toporik,A., Edelheit,S., Peretz,Y., Diber,A., Shemesh,R., Novik,A. and Sorek,R. (2006) Transcription-mediated gene fusion in the human genome. *Genome Res.*, **16**, 30–36.
2. Parra,G., Reymond,A., Dabbouseh,N., Dermitzakis,E.T., Castelo,R., Thomson,T.M., Antonarakis,S.E. and Guigo,R. (2006) Tandem chimerism as a means to increase protein complexity in the human genome. *Genome Res.*, **16**, 37–44.
3. Greger,L., Su,J., Rung,J., Ferreira,P.G., Geuvadis,C., Lappalainen,T., Dermitzakis,E.T. and Brazma,A. (2014) Tandem RNA chimeras contribute to transcriptome diversity in human population and are associated with intronic genetic variants. *PLoS One*, **9**, e104567.
4. Prakash,T., Sharma,V.K., Adati,N., Ozawa,R., Kumar,N., Nishida,Y., Fujikake,T., Takeda,T. and Taylor,T.D. (2010) Expression of conjoined genes: another mechanism for gene regulation in eukaryotes. *PLoS One*, **5**, e13284.
5. Kim,R.N., Kim,A., Choi,S.H., Kim,D.S., Nam,S.H., Kim,D.W., Kim,D.W., Kang,A., Kim,M.Y., Park,K.H. *et al.* (2012) Novel mechanism of conjoined gene formation in the human genome. *Funct. Integr. Genomics*, **12**, 45–61.
6. Varley,K.E., Gertz,J., Roberts,B.S., Davis,N.S., Bowling,K.M., Kirby,M.K., Nesmith,A.S., Oliver,P.G., Grizzle,W.E., Forero,A. *et al.* (2014) Recurrent read-through fusion transcripts in breast cancer. *Breast Cancer Res. Treat*, **146**, 287–297.
7. Qin,F., Song,Z., Babiceanu,M., Song,Y., Facemire,L., Singh,R., Adli,M. and Li,H. (2015) Discovery of CTCF-sensitive Cis-spliced fusion RNAs between adjacent genes in human prostate cells. *PLoS Genet.*, **11**, e1005001.
8. Qin,F., Song,Z., Chang,M., Song,Y., Frierson,H. and Li,H. (2016) Recurrent cis-SAGe chimeric RNA, D2HGDH-GAL3ST2, in prostate cancer. *Cancer Lett.*, **380**, 39–46.
9. Qin,F., Song,Y., Zhang,Y., Facemire,L., Frierson,H. and Li,H. (2016) Role of CTCF in regulating SLC45A3-ELK4 Chimeric RNA. *PLoS One*, **11**, e0150382.
10. Tang,Y., Qin,F., Liu,A. and Li,H. (2017) Recurrent fusion RNA DUS4L-BCAP29 in non-cancer human tissues and cells. *Oncotarget*, **8**, 31415–31423.
11. Jia,Y., Xie,Z. and Li,H. (2016) Intergenically spliced chimeric RNAs in cancer. *Trends Cancer*, **2**, 475–484.
12. Pflueger,D., Mittmann,C., Dehler,S., Rubin,M.A., Moch,H. and Schraml,P. (2015) Functional characterization of BC039389-GATM and KLK4-KRSP1 chimeric read-through transcripts which are up-regulated in renal cell cancer. *BMC Genomics*, **16**, 247.
13. Kim,H.P., Cho,G.A., Han,S.W., Shin,J.Y., Jeong,E.G., Song,S.H., Lee,W.C., Lee,K.H., Bang,D., Seo,J.S. *et al.* (2014) Novel fusion transcripts in human gastric cancer revealed by transcriptome analysis. *Oncogene*, **33**, 5434–5441.
14. Huang,H.Y., Ladanyi,M. and Soslow,R.A. (2004) Molecular detection of JAZF1-JJAZ1 gene fusion in endometrial stromal neoplasms with classic and variant histology: evidence for genetic heterogeneity. *Am. J. Surg. Pathol.*, **28**, 224–232.
15. Nucci,M.R., Harburger,D., Koontz,J., Cin,Dal and Sklar,J. (2007) Molecular analysis of the JAZF1-JJAZ1 gene fusion by RT-PCR and fluorescence in situ hybridization in endometrial stromal neoplasms. *Am. J. Surg. Pathol.*, **31**, 65–70.
16. Zhang,Y., Gong,M., Yuan,H., Park,H.G., Frierson,H.F. and Li,H. (2012) Chimeric transcript generated by cis-splicing of adjacent genes regulates prostate cancer cell proliferation. *Cancer Discov.*, **2**, 598–607.
17. Li,Z., Qin,F. and Li,H. (2018) Chimeric RNAs and their implications in cancer. *Curr. Opin. Genet. Dev.*, **48**, 36–43.
18. Chwalenia,K., Qin,F., Singh,S., Tangtrongstittikul,P. and Li,H. (2017) Connections between transcription downstream of genes and cis-SAGe chimeric RNA. *Genes (Basel)*, **8**, E338.
19. Younis,I., Berg,M., Kaida,D., Dittmar,K., Wang,C. and Dreyfuss,G. (2010) Rapid-response splicing reporter screens identify differential regulators of constitutive and alternative splicing. *Mol. Cell. Biol.*, **30**, 1718–1728.
20. Patel,R.K. and Jain,M. (2012) NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One*, **7**, e30619.

21. Benelli,M., Pescucci,C., Marseglia,G., Severgnini,M., Torricelli,F. and Magi,A. (2012) Discovering chimeric transcripts in paired-end RNA-seq data by using EricScript. *Bioinformatics*, **28**, 3232–3239.
22. Babiceanu,M., Qin,F., Xie,Z., Jia,Y., Lopez,K., Janus,N., Facemire,L., Kumar,S., Pang,Y., Qi,Y. *et al.* (2016) Recurrent chimeric fusion RNAs in non-cancer tissues and cells. *Nucleic Acids Res.*, **44**, 2859–2872.
23. Krzywinski,M., Schein,J., Birol,I., Connors,J., Gascoyne,R., Horsman,D., Jones,S.J. and Marra,M.A. (2009) Circos: an information aesthetic for comparative genomics. *Genome Res.*, **19**, 1639–1645.
24. Senapathy,P., Shapiro,M.B. and Harris,N.L. (1990) Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. *Methods Enzymol.*, **183**, 252–278.
25. Chwalenia,K., Facemire,L. and Li,H. (2017) Chimeric RNAs in cancer and normal physiology. *Wiley Interdiscip. Rev. RNA*, **8**, e1427.
26. Kowalski,P.E., Freeman,J.D. and Mager,D.L. (1999) Intergenic splicing between a HERV-H endogenous retrovirus and two adjacent human genes. *Genomics*, **57**, 371–379.
27. Batista,L., Bourachot,B., Mateescu,B., Reyal,F. and Mechta-Grigoriou,F. (2016) Regulation of miR-200c/141 expression by intergenic DNA-looping and transcriptional read-through. *Nat. Commun.*, **7**, 8959.
28. Eul,J., Graessmann,M. and Graessmann,A. (1996) Trans-splicing and alternative-tandem-cis-splicing: two ways by which mammalian cells generate a truncated SV40 T-antigen. *Nucleic Acids Res.*, **24**, 1653–1661.
29. Long,J.C. and Caceres,J.F. (2009) The SR protein family of splicing factors: master regulators of gene expression. *Biochem. J.*, **417**, 15–27.
30. Chen,M. and Manley,J.L. (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.*, **10**, 741–754.
31. Kuehner,J.N., Pearson,E.L. and Moore,C. (2011) Unravelling the means to an end: RNA polymerase II transcription termination. *Nat. Rev. Mol. Cell Biol.*, **12**, 283–294.
32. Zhao,D.Y., Gish,G., Braunschweig,U., Li,Y., Ni,Z., Schmitges,F.W., Zhong,G., Liu,K., Li,W., Moffat,J. *et al.* (2016) SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. *Nature*, **529**, 48–53.
33. Loya,T.J. and Reines,D. (2016) Recent advances in understanding transcription termination by RNA polymerase II [version 1; referees: 2 approved]. *F1000Res*, **5**, 1478.
34. Proudfoot,N.J. (2016) Transcriptional termination in mammals: stopping the RNA polymerase II juggernaut. *Science*, **352**, aad9926.
35. Zhang,Z., Klatt,A., Gilmour,D.S. and Henderson,A.J. (2007) Negative elongation factor NELF represses human immunodeficiency virus transcription by pausing the RNA polymerase II complex. *J. Biol. Chem.*, **282**, 16981–16988.
36. Wada,T., Takagi,T., Yamaguchi,Y., Ferdous,A., Imai,T., Hirose,S., Sugimoto,S., Yano,K., Hartzog,G.A., Winston,F. *et al.* (1998) DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev.*, **12**, 343–356.
37. Ladd,A.N., Charlet,N. and Cooper,T.A. (2001) The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. *Mol. Cell. Biol.*, **21**, 1285–1296.
38. Black,D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.*, **72**, 291–336.
39. David,C.J. and Manley,J.L. (2008) The search for alternative splicing regulators: new approaches offer a path to a splicing code. *Genes Dev.*, **22**, 279–285.
40. Park,J.W., Parisky,K., Celotto,A.M., Reenan,R.A. and Graveley,B.R. (2004) Identification of alternative splicing regulators by RNA interference in Drosophila. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 15974–15979.
41. Smith,C.W. and Valcarcel,J. (2000) Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.*, **25**, 381–388.
42. Skourti-Stathaki,K. and Proudfoot,N.J. (2014) A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev.*, **28**, 1384–1396.
43. Keller,W., Bienroth,S., Lang,K.M. and Christofori,G. (1991) Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3′ processing signal AAUAAA. *EMBO J.*, **10**, 4241–4249.
44. Yue,F., Cheng,Y., Breschi,A., Vierstra,J., Wu,W., Ryba,T., Sandstrom,R., Ma,Z., Davis,C., Pope,B.D. *et al.* (2014) A comparative encyclopedia of DNA elements in the mouse genome. *Nature*, **515**, 355–364.
45. Kumar-Sinha,C., Kalyana-Sundaram,S. and Chinnaiyan,A.M. (2012) SLC45A3-ELK4 chimera in prostate cancer: spotlight on cis-splicing. *Cancer Discov.*, **2**, 582–585.
46. Kekeeva,T., Tanas,A., Kanygina,A., Alexeev,D., Shikeeva,A., Zavalishina,L., Andreeva,Y., Frank,G.A. and Zaletaev,D. (2016) Novel fusion transcripts in bladder cancer identified by RNA-seq. *Cancer Lett.*, **374**, 224–228.
47. Oshlack,A., Robinson,M.D. and Young,M.D. (2010) From RNA-seq reads to differential expression results. *Genome Biol.*, **11**, 220.
48. Li,X., Zhao,L., Jiang,H. and Wang,W. (2009) Short homologous sequences are strongly associated with the generation of chimeric RNAs in eukaryotes. *J. Mol. Evol.*, **68**, 56–65.
49. Meyer,S. and Beslon,G. (2014) Torsion-mediated interaction between adjacent genes. *PLoS Comput. Biol.*, **10**, e1003785.
50. Vilborg,A., Passarelli,M.C., Yario,T.A., Tycowski,K.T. and Steitz,J.A. (2015) Widespread inducible transcription downstream of human genes. *Mol. Cell*, **59**, 449–461.
51. Vilborg,A. and Steitz,J.A. (2017) Readthrough transcription: how are DoGs made and what do they do? *RNA Biol.*, **14**, 632–636.
52. McCracken,S., Lambermon,M. and Blencowe,B.J. (2002) SRm160 splicing coactivator promotes transcript 3′-end cleavage. *Mol. Cell. Biol.*, **22**, 148–160.
53. Blencowe,B.J., Issner,R., Nickerson,J.A. and Sharp,P.A. (1998) A coactivator of pre-mRNA splicing. *Genes Dev.*, **12**, 996–1009.
54. Blencowe,B.J., Bauren,G., Eldridge,A.G., Issner,R., Nickerson,J.A., Rosonina,E. and Sharp,P.A. (2000) The SRm160/300 splicing coactivator subunits. *RNA*, **6**, 111–120.
55. Longman,D., McGarvey,T., McCracken,S., Johnstone,I.L., Blencowe,B.J. and Caceres,J.F. (2001) Multiple interactions between SRm160 and SR family proteins in enhancer-dependent splicing and development of C. elegans. *Curr. Biol.*, **11**, 1923–1933.
56. Alsafadi,S., Houy,A., Battistella,A., Popova,T., Wassef,M., Henry,E., Tirode,F., Constantinou,A., Piperno-Neumann,S., Roman-Roman,S. *et al.* (2016) Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nat. Commun.*, **7**, 10615.
57. Dolatshad,H., Pellagatti,A., Fernandez-Mercado,M., Yip,B.H., Malcovati,L., Attwood,M., Przychodzen,B., Sahgal,N., Kanapin,A.A., Lockstone,H. *et al.* (2015) Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells. *Leukemia*, **29**, 1092–1093.
58. Inoue,D. and Abdel-Wahab,O. (2016) Modeling SF3B1 mutations in cancer: advances, challenges, and opportunities. *Cancer Cell*, **30**, 371–373.
59. Kfir,N., Lev-Maor,G., Glaich,O., Alajem,A., Datta,A., Sze,S.K., Meshorer,E. and Ast,G. (2015) SF3B1 association with chromatin determines splicing outcomes. *Cell Rep.*, **11**, 618–629.
60. Convertini,P., Shen,M., Potter,P.M., Palacios,G., Lagisetti,C., de la Grange,P., Horbinski,C., Fondufe-Mittendorf,Y.N., Webb,T.R. and Stamm,S. (2014) Sudemycin E influences alternative splicing and changes chromatin modifications. *Nucleic Acids Res.*, **42**, 4947–4961.
61. Wu,G., Fan,L., Edmonson,M.N., Shaw,T., Boggs,K., Easton,J., Rusch,M.C., Webb,T.R., Zhang,J. and Potter,P.M. (2018) Inhibition of SF3B1 by molecules targeting the spliceosome results in massive aberrant exon skipping. *RNA*, **24**, 1056–1066.