

# Genome-wide evidence for an essential role of the human Staf/ZNF143 transcription factor in bidirectional transcription

Yannick-Noël Anno<sup>1,2</sup>, Evelyne Myslinski<sup>2</sup>, Richard Patryk Ngondo-Mbongo<sup>2</sup>, Alain Krol<sup>2</sup>, Olivier Poch<sup>1</sup>, Odile Lecompte<sup>1</sup> and Philippe Carbon<sup>2,\*</sup>

<sup>1</sup>Department of Structural Biology and Genomics, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Institut National de la Santé et de la Recherche Médicale (INSERM), The Centre National de la Recherche Scientifique (CNRS), UMR7104, F-67400 Illkirch, Université de Strasbourg, F-67000 Strasbourg and <sup>2</sup>Architecture et Réactivité de l'ARN, Université de Strasbourg, CNRS, IBMC, 15 rue René Descartes, 67084 Strasbourg, France

Received October 29, 2010; Revised November 30, 2010; Accepted December 4, 2010

## ABSTRACT

In the human genome, ~10% of the genes are arranged head to head so that their transcription start sites reside within <1 kbp on opposite strands. In this configuration, a bidirectional promoter generally drives expression of the two genes. How bidirectional expression is performed from these particular promoters constitutes a puzzling question. Here, by a combination of *in silico* and biochemical approaches, we demonstrate that hStaf/ZNF143 is involved in controlling expression from a subset of divergent gene pairs. The binding sites for hStaf/ZNF143 (SBS) are over-represented in bidirectional versus unidirectional promoters. Chromatin immunoprecipitation assays with a significant set of bidirectional promoters containing putative SBS revealed that 93% of them are associated with hStaf/ZNF143. Expression of dual reporter genes directed by bidirectional promoters are dependent on the SBS integrity and requires hStaf/ZNF143. Furthermore, in some cases, functional SBS are located in bidirectional promoters of gene pairs encoding a noncoding RNA and a protein gene. Remarkably, hStaf/ZNF143 *per se* exhibits an inherently bidirectional transcription activity, and together our data provide the demonstration that hStaf/ZNF143 is indeed a transcription factor controlling the expression of divergent protein–protein and protein–non-coding RNA gene pairs.

## INTRODUCTION

Despite the vast genomic space, a substantial fraction of human genes (11%) are arranged as bidirectional gene pairs in a head to head, divergent fashion and controlled by bidirectional promoters (1,2). A bidirectional promoter was defined as an intergenic region containing <1 kb of DNA which is flanked by the transcription start sites (TSS) of two genes in opposite directions (1,2). Thus the closely located bidirectional pairs are overrepresented in the human genome and this abundance has been observed across several mammalian genomes (3). For a vast majority of human bidirectional promoters where the associated genes have mammalian orthologs the bidirectional arrangement is conserved suggesting that it is functionally important (1,2). Indeed, the expression pattern of divergent gene pairs are more correlated than those of randomly paired genes (2). Most of the bidirectional promoters lack TATA element, are GC rich and exhibit enriched colocalization with CpG islands (2,4). Bidirectional promoters display also a mirror sequence composition with an overrepresentation of G and T on one side of the promoter, and C and A on the other, on the genomic plus strand (5). Divergent promoters are enriched in binding site consensus sequences of a small group of transcription factors including GABPA/NRF2, NRF1, NFY and YY1 while the majority of known vertebrate binding motifs are underrepresented in bidirectional promoters (6). Further study has shown that the GABPA/NRF2 factor is involved in the regulation of transcription from divergent gene pairs (7). The work described in this article establishes that the human transcription factor Staf (hStaf, also called ZNF143) acts also

\*To whom correspondence should be addressed. Tel: +33 3 88417064; Fax: +33 388602218; Email: P.Carbon@ibmc-cnrs.unistra.fr

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

as an essential factor in transcription from bidirectional promoters. Staf, originally identified in *Xenopus laevis* as the transcription activator of the tRNA<sup>Sec</sup> gene, also controls expression of snRNA and snRNA-type genes (8–11). Seven contiguous zinc fingers of the C2–H2 type contain the Staf DNA binding domain and only zinc fingers 1–6 establish base-specific contacts with 18 bp in Staf–DNA complexes (12–14). Furthermore, hStaf/ZNF143 has been reported to regulate transcription of the SCARNA2 gene and of 10 protein coding genes (15–17 and references therein). Lastly, a genome-wide analysis led us to identify 1175 hStaf/ZNF143 binding sites (SBS) distributed in 938 promoters of mammalian protein genes, strongly suggesting that hStaf/ZNF143 is involved in the transcriptional regulation of a very large number of protein coding genes (18). In the present study, we conducted bioinformatic and biochemical analysis at the whole genome scale which demonstrated that functional SBS are overrepresented in bidirectional compared to unidirectional promoters. Chromatin immunoprecipitation assays (ChIP) on 100 representative bidirectional promoters containing putative SBS revealed that 93% of them are associated with hStaf/ZNF143. Using a combination of luciferase, RTase assays, hStaf/ZNF143 overexpression and silencing, we demonstrated that hStaf/ZNF143 is clearly involved in the bidirectional expression directed by the promoters of protein–protein and protein–non-coding RNA gene pairs. Furthermore we also showed that, in dual reporter genes, the sole presence of an SBS isolated from unidirectional promoters was sufficient to direct bidirectional transcription, demonstrating that hStaf/ZNF143 *per se* exhibits an inherent bidirectional activity.

## MATERIALS AND METHODS

### Identification of bidirectional gene pairs

Bidirectional gene pairs were identified by skimming through the gene annotation set constituted by Genomic Context database (GeCo) (Y.-N. Anno *et al.*, manuscript in preparation) and extracting all pairs of genes in inverted orientations separated by <1 kbp, regardless of their gene type. The Genomic Context database is related to human genome NCBI build 36 (hg18) and was built by computing ‘refGene’ (proteins), ‘rnaGene’ (snRNA, snoRNA, tRNA, rRNA, scaRNA), kgXref tables from the University of Santa Cruz California (UCSC, www.genome.ucsc.edu), ‘mirna’, ‘mirna\_literature\_references’, ‘mirna\_mature’ and ‘literature\_references’ tables from Sanger Institute and piRNA’s file from the piRNAdatabank. It was implemented locally in a high-speed DB2 architecture called Biological Integration and Retrieval of Data (BIRD), allowing to quickly address the whole gene set (19). For each bidirectional promoter, 10 randomly chosen unidirectional promoters of the same size were extracted from GeCo, assessing that a gene with no other gene in the vicinity of its TSS (1 kbp) is a unidirectional promoter.

### Sequence analysis and identification of potential SBS in promoters

Prediction of Staf binding sites in bidirectional and unidirectional promoters was performed with a position-specific scoring matrix (PSSM) built using 347 experimentally validated binding sites (18). Matrix generation and scan of sequences were done using the Genomatix suite, with MatDefine and MatInspector programs (20), respectively. Significant overrepresentation of SBS in bidirectional promoters was evaluated using the  $\chi^2$  test for two proportions, with  $\alpha = 0.05$  and 1 degree of freedom (df). The test was repeated 10 times on corresponding random sets (Supplementary Table S1) and assessed in 100% of the cases a significantly higher rate of predicted SBS in bidirectional promoters.

### Reporter constructs

Thirteen human bidirectional promoters (extended intergenes EI1–EI13) ranging from 229 to 751 bp were PCR amplified from human genomic DNA (list of primers and chromosomal localization available as Supplementary Table S2) and cloned into pGEMTeasy (Promega). BamHI and SacI sites were included in the forward and reverse primers, respectively, used for PCR amplification of EI11 and EI12. KpnI sites were included in the primers used in EI13 amplification. The dual luciferase reporters constructs containing EI1–EI10 were obtained as follows. EcoRI fragments isolated from pGEMTeasy constructs containing EI1–EI10 were subcloned into EcoRI-cut pFRLN50 dual luciferase reporter vector generated by insertion of the NcoI linker CATGGGAGCTCGAATTCGAGCTC in the NcoI site of the pFRL vector (21). The luciferase reporter constructs containing the EI11 and EI12 fragments were generated by subcloning the BamHI-SacI and NotI-SacI fragments isolated from pGEMTeasy-EI11 and EI12 into BamHI-SacI and NotI-SacI cut pFlashI luciferase reporter vector (SynapSys), respectively. The KpnI fragment isolated from pGEMTeasy-EI13 was subcloned in KpnI-cleaved pU6/Hae/RA.2/EcoRV construct (11,22). The resulting BamHI-SacI fragment containing EI13, followed by a 137 bp spacer derived from the  $\beta$ -globin gene, was subcloned into BamHI-SacI cut pFlashI luciferase reporter vector. PCR primers were designed so that the PCR products containing the EI1–EI11 and EI13 promoters do not contain the AUG initiation codon of the endogenous gene. Instead, translation initiation is ensured by the AUG of the luciferase reporter genes. For the luciferase construct containing the EI12 bidirectional promoter, the endogenous AUG of the C19orf6 gene was in frame with the AUG of the firefly luciferase. MaxiU6 genes were created by inserting the GACCTCG AGGCGGTTTC sequence at position 87 of the wt U6. In all mutant versions of the SBS, the NNCCCR sequence (N standing for any nucleotide, R for A or G) at positions 1–6 of the SBS was replaced by GGT TTC. The dual luciferase vector with part of the BUB1B promoter containing two SBS and the associated ACTACAA motif was prepared as in Myslinski *et al.* (16). The 271 bp DNA fragment (Ch15: 38 240 250–38 240 520, in hg18) was amplified

by PCR and inserted as an EcoRI fragment into the EcoRI-cut pFRLN50. All mutants were generated with the QuickChange II XL site-directed mutagenesis kit (Stratagene) and verified by DNA sequencing.

#### ChIP assay and semi-quantitative PCR analysis

The ChIP procedure and the PCR analysis was performed as described in Myslinski *et al.* (18) using a rabbit polyclonal antibody against a C-terminal epitope of Staf. For the negative control, we used the PP1–PP4 primer pairs hybridizing to unique regions located at 2.4, 2.1, 6.5 and 2.5 kbp upstream of the tRNA<sup>Sec</sup>, U4 ATAC, GAPDH and BUB1B genes and generating PCR products of 235, 211, 174, 226 bp, respectively. The primer sequences used in this study are available on request.

#### Transfection and bidirectional promoter activity assay

HeLa cells ( $5 \times 10^3$  cells/well in 96-well plates) were cotransfected using Lipofectamine 2000 with 200 ng of each experimental luciferase construct (containing wt or mutated E11–E113 promoters) and 100 ng of the pCH110 internal control plasmid. After 24 h, cells lysates were prepared with passive lysis buffer (Promega) and assayed for  $\beta$ -galactosidase and luciferase activities. The firefly and *Renilla* luciferase assays (promoters E11–E110) were performed using the dual-luciferase TM Reporter assay system (Promega) on a GloMax<sup>TM</sup> 96 Plate Luminometer (Promega). The firefly luciferase assay (promoters E11–E113) was performed using the Luciferase assay system (Promega). The luciferase activities were normalized to the  $\beta$ -galactosidase activity. Each transfection experiment and luciferase assay were done at least in triplicate. For U6 snRNA and RPPH1 (H1RNA) gene expression analysis, HeLa cells ( $9 \times 10^5$  cells) were cotransfected using Lipofectamine 2000 with 8.4  $\mu$ g of wt or mutant luciferase constructs containing the E111 to E113 promoters with 600 ng of maxi5S RNA as internal standard. Cells were collected after 48 h and total RNA was extracted using TRIAGENT (Euromedex). Total RNA was analyzed by primer extension of two labeled oligonucleotides, one complementary to positions +88/+104 of the maxiU6 (human wt U6 gene numbering) and the other to positions +112/+129 of the maxi5S. The extended products were separated on a 6% denaturing gel and quantitated with a Fuji Bioimage Analyzer. The yield of extended maxiU6 was normalized to that of extended maxi5S. Each transfection experiment and reverse transcriptase assay were done at least in triplicate. H1 RNA gene expression from the E113 promoter was monitored as described in Myslinski *et al.* (11).

#### Transfection of siRNA, overexpression, reverse transcription and RT–qPCR

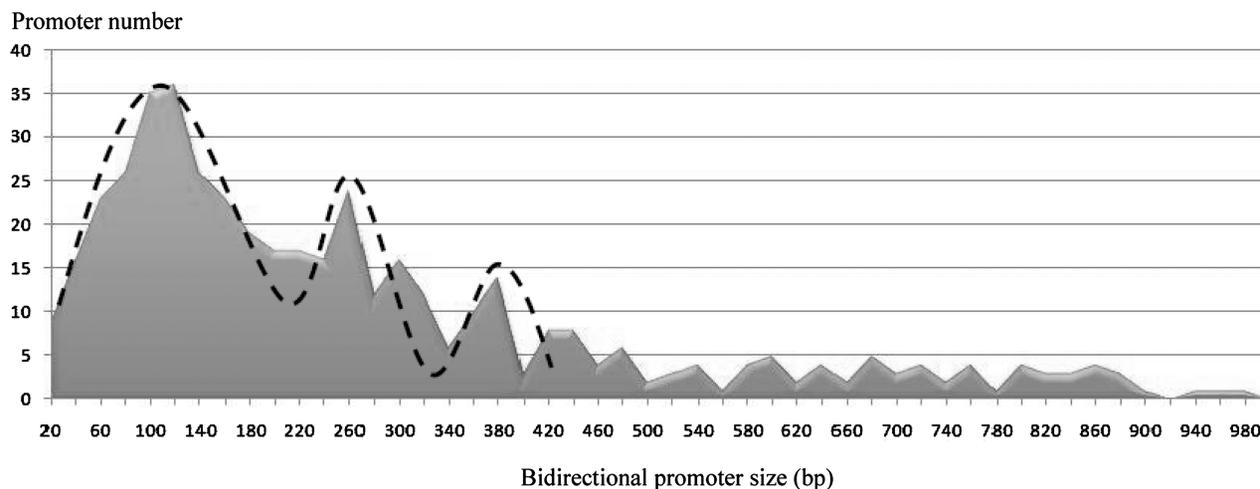
The hStaf/ZNF143-specific siRNAs GCUGGAAGAUG GUACCACAGCUUUAU (siRNA1), GGGCAUUUGCC AGUGCAACAAAUUA (siRNA2), GGAACGCACUC UGUUGCUAUGGUUA (siRNA3) or control-siRNA (Invitrogen) were transfected into HeLa cells using Lipofectamine 2000 according to the manufacturer's instructions (Invitrogen). To obtain hStaf/ZNF143

overexpression, HeLa cells ( $5 \times 10^5$  cells) were transfected with Lipofectamine 2000. Cells received 3  $\mu$ g of pcDNA3.1-hStaf/ZNF143 (10) with 500 ng of pCH110 as internal control. The total amount of DNA was kept constant (4  $\mu$ g) with carrier DNA. Total RNA was isolated using TRIAGENT (Euromedex) from cells harvested 48 h (overexpression experiment) or 72 h (siRNA transfection) post transfection, treated by RNase-free DNase and reverse transcribed using oligo (N)9 primer. cDNA was amplified by qPCR with gene specific primers on a Stratagene Mx3005P PCR system (Agilent Technologies) using EvaGreen qPCR Mix Plus (Euromedex). We used Primer3 software (<http://frodo.wi.mit.edu/primer3/>) to design primers so that the final amplicon was 100–150 bp. The primer sequences used in this study are available on request. All reactions were carried out in triplicate. Relative gene expression was calculated using the  $\Delta C_t$  method following the manufacturer's instructions.

## RESULTS

#### Promoter size and genes associated in bidirectional promoters

We extracted from the Genomic Context database (GeCo) pairs of human genes that are the closest neighbors on opposite strands and with their TSS separated by <1 kbp ('Materials and Methods' section). In this search, we included only those genes whose transcripts were not predicted to overlap at the 5'-ends. As defined by Trinklein *et al.* (2), we considered the region between the two TSS as a putative bidirectional promoter. Therefore, the constituted set contained 839 bidirectional promoters and 1678 genes (Supplementary Table S3, part 1). Different interesting features regarding the size of the bidirectional promoters, distribution of the sizes and the length of the genes in the divergent gene pairs arose from our study. First, considering the size of the bidirectional promoters, Figure 1 shows the number of bidirectional promoters with a definite length plotted versus their length. We found a bias toward a small size: 83% of the promoters harbor a size <500 bp. Second, the size distribution of the bidirectional promoters is not uniform. Indeed different pools are present with preferential sizes of 80–140, 240–280 and 360–420 bp corresponding to 25, 10 and 7% of the total bidirectional promoter set, respectively. Third, a bias toward small size genes associated to divergent promoters was observed compared to unidirectional promoters: 40.1 versus 112.8 kbp (Supplementary Table S3, part 2). The bidirectional promoter set is constituted at 95% (801 out of the 839) of protein coding gene pairs (c–c gene pairs). The remainder is distributed into two small groups: 26 combine a non-coding (nc) RNA gene with a protein coding partner (nc–c gene pairs), and 12 contain ncRNA genes only (nc–nc gene pairs). All members of the nc–nc subgroup are entirely constituted of two tRNA genes but tRNA genes were also identified in 17 of the nc–c pairs. The ncRNA genes present in the other nine nc–c gene pairs are: MRP-RNA, H1-RNA, SRP-RNA, 7SK RNA,



**Figure 1.** Features of bidirectional promoter sizes in the human genome. The length of the bidirectional promoter in base pair is plotted versus the frequency of its occurrence (promoter number for a 10 bp interval). Dotted line indicates the position of the promoters pools.

scaRNA17 (U91), U13 snoRNA, U6.2 snRNA, U6.9 snRNA and U12 snRNA (Supplementary Table S3, part 1). Thus, the closely located divergent gene pairs and the associated bidirectional promoters harbor a biased small size, suggesting a putative functional importance of this particular feature.

#### Sequence analysis and staf binding site identification in bidirectional promoters

The hStaf/ZNF143 Staf binding site (SBS) consists of a 18 bp sequence with a highly conserved consensus sequence CCCR at positions 3–6, a highly conserved C at position 12 and a more degenerate sequence at positions 1–2, 7–11 and 13–18 (9,12,18). The strategy to identify SBS in bidirectional promoters was the following. We collected a pool of 347 binding sites experimentally validated by ChIP (18) and converted them into a position-specific scoring matrix (PSSM) using MatDefine programs of the Genomatix suite (20) (Supplementary Figure S1). We then used MatInspector to search matrix matches in the sequences of intergenes and in the first 400 bp of the transcribed sequences of two genes partners. The search was extended to the first 400 bp of the genes because it is well established that the first exon and 5' part of the first intron of genes contain transcription factors binding sites (TFBS) (23). With an optimal cutoff score of 0.82, the search identified 601 SBS above the cutoff and which are distributed in 392 potential bidirectional promoters (392 out of 839: 46.7% of the bidirectional promoters) (Figure 2A, Supplementary Tables S3 and S4). Among these 392 pairs, 11 combine a ncRNA gene with a protein coding partner (nc–c gene pairs) and two combine two ncRNA genes (nc–nc gene pairs) (Figure 2A and Supplementary Table S5). The two nc–nc gene pairs contain each two tRNA genes, tRNA<sup>Gly-TCC</sup>–tRNA<sup>Trp-CCA</sup> and tRNA<sup>Arg-TCG</sup>–tRNA<sup>Arg-CCT</sup>. Among the 11 nc–c gene pairs, six of them contain a tRNA gene and in the other five nc–c gene pairs, the ncRNA gene partner is the RNase MRP RNA, RNase P RNA

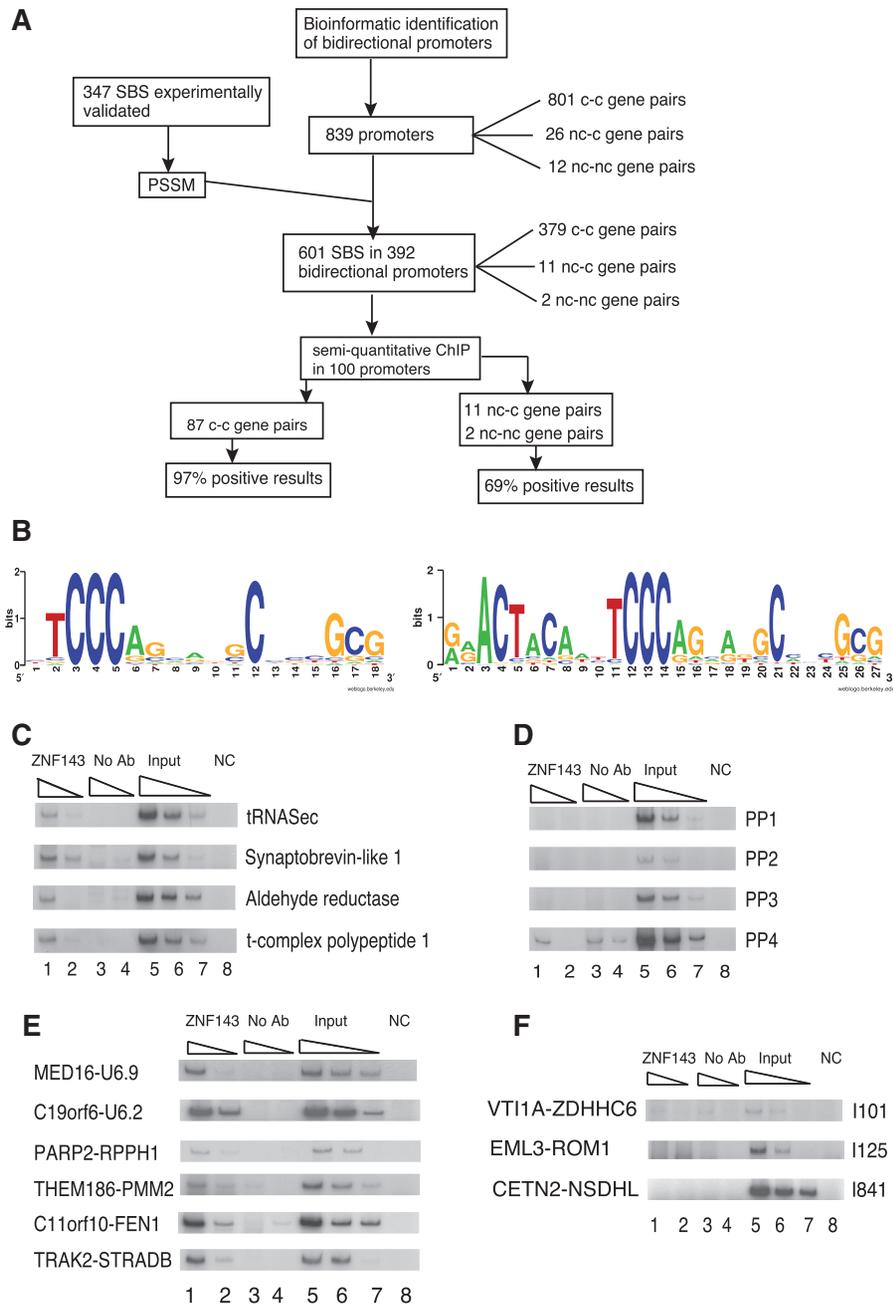
(H1RNA), SRP-RNA (RNA 7SL1), snRNAU6.2 and snRNAU6.9 (Supplementary Table S5).

Among 126 (21%) of the 601 identified SBS, inspection of sequences adjacent to the identified SBS revealed the presence, immediately upstream of the SBS, of the 9-bp RRCTAYRN motif or a one bp variant (Figure 2B). Among the 392 promoters identified as containing an SBS, 203 (51.8%) harbored at least one SBS in the intergene, the percentage being 24.2% (203/839) when considering the whole set of bidirectional promoters at the onset of this study (Supplementary Table S3, part 1). A control set was generated by random extraction from unidirectional promoters of a bp number equivalent to the total bp number in the intergenic region of the 839 bidirectional promoters. The search of SBS in this and nine other different control sets showed that 10% of unidirectional promoters contain at least one SBS.

Altogether, these results indicate that hStaf/ZNF143 binding sites are present in at least 46.7% of the bidirectional promoters with a 2.4-fold enrichment compared to unidirectional promoters.

#### *In vivo* occupancy of bidirectional promoters by hStaf/ZNF143

To determine whether hStaf/ZNF143 is indeed associated *in vivo* with the bidirectional promoters identified as containing putative SBS, a chromatin immunoprecipitation assay was performed in HeLa cells. We examined 186 putative SBS with scores ranging from 0.82 to 0.99 (listed in Supplementary Table S6, scores in Supplementary Table S4), contained in 100 (25.5%) of the 392 promoters identified as containing SBS. The 100 selected promoters correspond to 87 pairs of protein coding genes only and to the whole set of c–nc (11 pairs) and nc–nc gene pairs (2 pairs). The 87 pairs were extracted with a random choice from the bidirectional promoter set (Supplementary Table S4). After immunoprecipitation with an antibody specifically recognizing hStaf/ZNF143, enrichment of the promoters was monitored by



**Figure 2.** Characterization of bidirectional promoters that are target for hStaf/ZNF143 and detection of promoter occupancy *in vivo* by chromatin immunoprecipitation (ChIP). (A) Outline showing identification of the target promoters. The results of the ChIP and PCR assays performed on 100 promoters (87 protein coding pairs, 11 protein coding–nc RNA pairs and two noncoding–noncoding RNA pairs) are indicated. (B) SBS motif analysis with sequence logos depicting nucleotide distribution in the SBS (left panel) and in the SBS with the 5'-associated submotif (right panel) (C–F). Binding of the endogenous hStaf/ZNF143 to genomic sites in HeLa cells was analyzed by ChIP. Genomic DNA fragments, recovered from input material or immunoprecipitated with hStaf/ZNF143 antibody or no-antibody, were subjected to semi-quantitative PCR amplification in the presence of ( $\alpha^{32}$ P)-dCTP with specific primer pairs. Lanes 1, 2 and 3, 4: serial dilutions of DNA immunoprecipitated with anti-hStaf/ZNF143 or no-antibody (no Ab), respectively. Lanes 5–7: serial dilutions of input material. Lane 8 (NC, negative control): PCR lacking the chromatin DNA template. Positive (C) and negative (D) controls for ChIP assays. Promoters used are indicated on the right of panel C. PCR products generated with the PP1–PP4 primer pairs originate from unique regions 2.4, 2.1, 6.5 and 2.5 kbp upstream of the tRNA<sup>Sec</sup>, U4 ATAC, GAPDH and BUB1B genes, respectively. Typical positive and three negative results obtained from bidirectional promoter regions are depicted in E and F, respectively. Bidirectional promoters are indicated by gene names.

semi-quantitative PCR amplification with primers amplifying the regions surrounding the putative SBS. The specificity of the ChIP reaction was monitored by PCR amplification of four promoters recognized by hStaf/

ZNF143 (Figure 2C, tRNA<sup>Sec</sup>, synaptobrevin-like1, aldehyde reductase and t-complex polypeptide 1 promoters as positive controls) and four DNA fragments which do not contain any hStaf/ZNF143 binding site

(Figure 2D, negative controls). Each DNA sequence was tested with each of the three templates obtained from anti-hStaf/ZNF143, control ChIP and input chromatin. We tested two dilutions of DNA isolated with (Figure 2C–F, lanes 1 and 2; Supplementary Figure S2A) or without antibody (Figure 2C–F, lanes 3 and 4; Supplementary Figure S2A). In addition, a serial dilution of the input material was analyzed to demonstrate that the PCR was quantitative within a linear range of amplification (Figure 2C–F, lanes 5–7; Supplementary Figure S2A). As expected, the positive controls yielded a signal of intensity higher with anti-hStaf/ZNF143 than in the no-antibody control (Figure 2C, compare lanes 1,2 and 3,4). In contrast, no specific signal could be obtained with the primer pairs PP1 to PP4 amplifying DNA sequences lying several kbp upstream of the tRNA<sup>Sec</sup>, U4 ATAC, GAPDH and BUB1B genes because these remote regions were not expected to interact with hStaf/ZNF143 (compare lanes 1,2 and 3,4 in Figure 2D). Among the 87 PCR amplifications involving promoters of c–c gene pairs (Figure 2E and F, and Supplementary Figure S2A), three of them (I101, I125 and I841; Figure 2F) were close to background level but the remaining 84 provided clear positive signals (Figure 2E and Supplementary Figure S2A; Table S6). These experiments showed that 97% (84 out of 87) of the bidirectional promoters of the c–c gene pairs tested did contain genuine hStaf/ZNF143 binding sites. In an additional control, we amplified from ChIP samples a DNA region corresponding to bidirectional promoters of c–c gene pairs identified *in silico* but lacking SBS. Of the 10 promoters that were tested, none provided positive amplification (Supplementary Figure S2B), suggesting that the hStaf/ZNF143 sequence is necessary for protein binding *in vivo*. More contrasted results were obtained with the PCR amplifications of nc–nc and nc–c gene pair promoters. Among the 13 that

were tested, only 9 (69%) provided positive amplification: LTA4H-tRNA<sup>Asp-GTC</sup>, PARP2-RPPH1, PRMT5-tRNA<sup>Arg-ACG</sup>, RPS29-SRP RNA, POLG-tRNA<sup>Arg-TCG</sup>, PSMB3-tRNA<sup>AsnGTT</sup>, MED16-snrRNAU6.9, C19orf6-snrRNAU6.2, SLC27A4-tRNA<sup>Arg-TCT</sup> (Figure 2E; Supplementary Figure S2A and Table S6). This will be further discussed.

Taken together, these results demonstrate the robustness of the computational screens and reveal the very high prevalence of *bona fide* direct targets of the hStaf/ZNF143 transcription factor in bidirectional promoters of c–c gene pairs.

#### Functional activity of the hStaf/ZNF143 binding site in bidirectional expression of protein gene pairs

We assessed the ability of 10 of the bidirectional promoters (associated with hStaf/ZNF143) to initiate transcription from both directions. These were EI1 to EI10 and are constituted of pairs of protein coding genes only (Table 1 and Supplementary Table S6). DNA fragments containing the full intergenic region lying between the TSS and parts of the first exon of the protein genes were inserted into a dual reporter vector (21) in which the transcription activities in the two opposite directions could be tested simultaneously via the readout of the firefly and *Renilla* luciferases (Table 1, Figure 3 and Supplementary Figure S3). HeLa cells were transiently transfected with the different constructs and the luciferase activities of the resulting cells extracts was measured. Analysis of the firefly and *Renilla* luciferase activities, normalized to that of the  $\beta$ -galactosidase control (Figure 3 and Supplementary Figure S3), showed that all the tested constructs bore bidirectional activity. The transcription activity depending on the promoter tested was from 9.2- to 100-fold in the *Renilla*, and from 6- to 160-fold in the firefly luciferase direction that of the promoter-less empty

**Table 1.** Characteristics of the 13 bidirectional promoters analyzed and effects of the SBS substitutions on the direction of transcription

EIn	Intergene (I)	Genes in pair	SBS number in EIn	SBS mutants number in EIn	Transcriptional effect of SBS mutation (in order of SBS in EIn)
EI1	I124	C11orf10–FEN1	2	2	b–b
EI2	I202	KNTC1–RSRC2	1	1	b
EI3	I252	C14orf133–AHSA1	3	3	b–b–u <sup>d</sup> (AHSA1)
EI4	I298	TMEM186–PMM2	3	2	b–b
EI5	I512	TRAK2–STRADB	2	2	u <sup>i</sup> (STRADB)–b
EI6	I562	ZMAT5–UCRC	4	3	u <sup>i</sup> (ZMAT5)–b–b
EI7	I659	COL4A3BP–POLK	2	2	b*–b
EI8	I758	DOCK4–ZNF277	1	1	b
EI9	I775	NUDCD1–ENY2	4	4	u <sup>d</sup> (ENY2)–u <sup>d</sup> (ENY2)–b–b
EI10	I797	ZNF189–MRPL50	2	2	b–b
EI11	I418	MED16–U6.9	3	2	b**–b**
EI13	I419	C19orf6–U6.2	2	2	b**–b**
EI13	I221	PARP2–RPPH1	1	1	b

EIn: name of the extended intergene. The intergene name refers to Supplementary Table S3. The numbers of SBS identified and SBS mutated in the intergene are mentioned. The effect of SBS substitution on the direction of transcription is indicated.

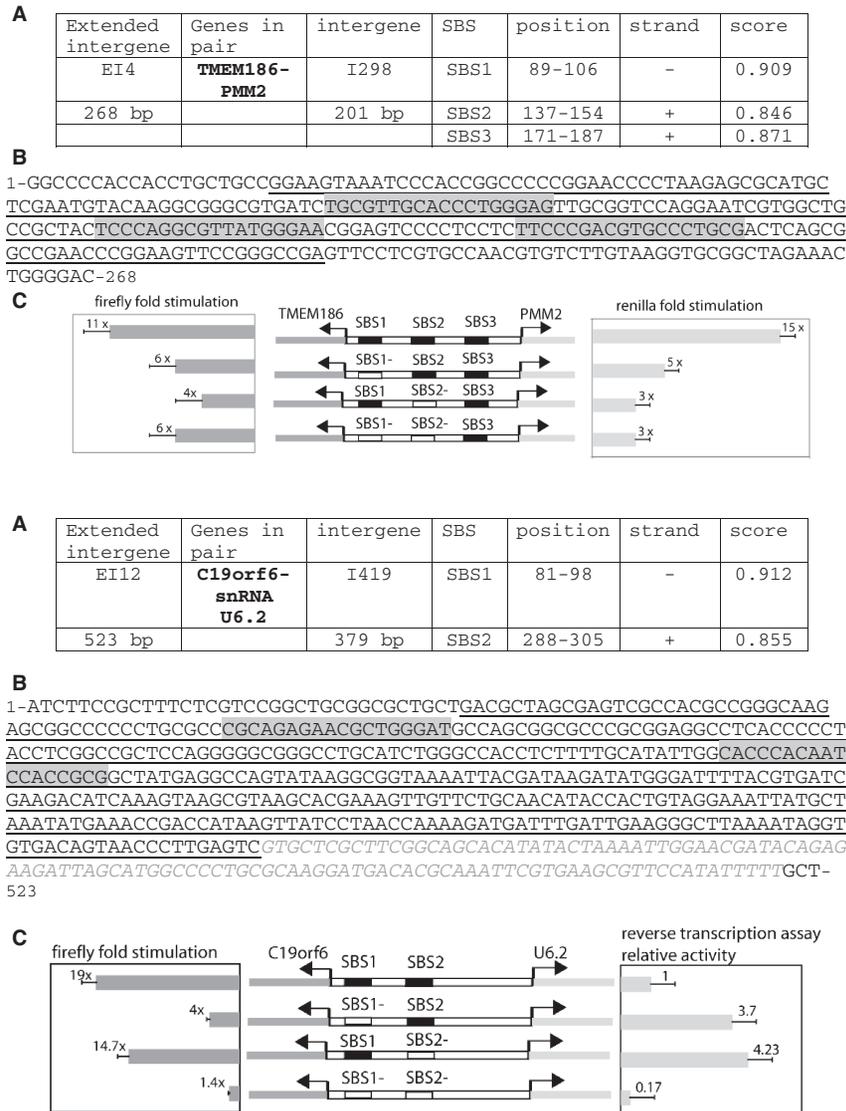
b, bidirectional effect with decrease in both directions.

u<sup>d</sup> (XXX), unidirectional effect with decrease in the direction of gene in parenthesis.

u<sup>i</sup> (XXX), unidirectional effect with increase in the direction of gene in parenthesis.

b\*, decreased and increased efficiencies in COL4A3BP and POLK directions, respectively.

b\*\*, decreased and increased efficiencies in protein gene and U6 directions, respectively.



**Figure 3.** The human TMEM186–PMM2 and C19orf6–snRNAU6.2 intergenic regions act as bidirectional promoters. Mutational analysis of the promoter identified the functional importance of the SBS in bidirectional transcription. (A) Characteristics of the extended intergenes EI4 and EI12 (length and name of genes in pair), intergenes (name and length) and SBS (position, orientation and score) are indicated. (B) Sequences of the EI4 (TMEM186–PMM2) and EI12 (C19orf6–snRNAU6.2) regions cloned into the dual and firefly luciferase reporters, respectively. The sequence of the intergene is underlined, the SBS are highlighted in gray and the U6 gene is in italics. (C) Schematic diagrams of the reporter genes used in the transfection assays. Mutations (–) are indicated; the relative transcription activities of the different constructs are indicated, representing either the luciferase activity/empty vector ratio or the relative activity in the reverse transcription assay for the U6 snRNA expression. The yield of extended maxiU6 was normalized to that of the extended maxiS5 added as the internal standard. Values are mean  $\pm$  SE (error bars),  $n = 3$  independent experiments with two replicates/group.

vector (Figure 3 and Supplementary Figure S3). The *in silico* data found that the fragments inserted in the dual luciferase reporter vector contain 1–3 SBS, with 17 SBS in total (Supplementary Table S4). However, visual inspection of the sequence revealed the presence of seven additional putative SBS, in fragments EI1, EI3, EI6, EI9 and EI10, that were unlisted by the bioinformatic screen (SBS with a score below 0.82; SBS1 in EI1, SBS1 and SBS2 in EI3, SBS4 in EI6, SBS3 and SBS4 in EI9 and SBS2 in EI10; Supplementary Figure S3). To determine whether the 24 SBS (17+7) do play a role in bidirectional transcription, the effect of substitutions at positions 1–6 in the SBS core sequence NNCCCR was tested on expression of

the reporter gene. Such substitutions were known to completely abolish formation of the DNA–protein complex (9,16). Twenty-two of the 24 SBS identified in the 10 promoters were mutated singly, and simultaneous mutations of SBS1 and SBS2 were engineered in EI1, EI4 and EI7 where they reside in the same promoter (Supplementary Figure S3). All single mutations had an effect on transcription efficiency. Five substituted SBS altered transcription unidirectionally whereas it was affected bidirectionally for 17 of them (77%) (Supplementary Figure S3 and Table 1). In the 10 promoters tested, at least one SBS was involved in the bidirectional control of transcription. For 16 of the 17 mutants with

bidirectionally-altered transcription, the promoter activity decreased simultaneously in both directions, demonstrating that these SBS are involved in up-regulating both genes in the pair. Interestingly, in the case of SBS1 in EI7, the SBS mutation decreased promoter activity in the COL4A3BP direction and increased it in the POLK direction (Table 1 and Supplementary Figure S3). In the five substitutions leading to unidirectional alteration of transcription, three of them exhibited a decrease: SBS3 in EI3, AHSA1 direction; SBS1 and SBS2 in EI9, ENY2 direction; and two a unidirectional increase: SBS1 in EI5, STRADB direction; SBS1 in EI6, ZMAT5 direction (Table 1 and Supplementary Figure S3). The transcriptional status of the mutants containing two simultaneously mutated SBS (SBS1 and SBS2 in EI1, EI3 and EI7) essentially recapitulates the effects observed with the single mutants. Taken together, these results demonstrate the functional importance of the hStaf/ZNF143 binding sites in bidirectional transcription of divergent genes.

#### **The simultaneous presence of two SBS in the promoters of the C19orf6–U6.2 and MED16–U6.9 gene pairs down-regulated expression of the U6.2 and U6.9 snRNA genes**

The activity of three hStaf/ZNF143-containing bidirectional promoters associating a protein and a ncRNA gene was tested by transient transfection assays. The gene pairs were MED16–U6.9, C19orf6–U6.2 and PARP2–RPPH1. The U6.9 and U6.2 ncRNA genes encode U6 snRNAs (24), RPPH1 coding for H1 RNA, the RNA component of the RNase P (from EI11 to EI13 in Table 1), all three genes being RNA polymerase III-dependent. To monitor expression from these promoters, constructs were engineered as follows. For MED16–U6.9 (EI11) and C19orf6–U6.2 (EI12), we transfected constructs containing part of the first exon of MED16 or C19orf6, the full intergenic region between the transcriptional start sites, and the U6.2 or U6.9 genes. In these constructs, parts of the protein coding genes are placed in front of the firefly luciferase reporter (EI12 and EI11 in Figure 3 and Supplementary Figure S3), and a 16 bp fragment was inserted into both U6 genes to distinguish transient expression of the U6 snRNAs from that of the endogenous genes. For PARP2–RPPH1, the construct contained part of the first exon of PARP2, and the full intergenic region placed between the luciferase reporter and a chimeric gene consisting of a 137 bp spacer derived from the  $\beta$ -globin gene followed by an efficient RNA polymerase III termination site (11,22). After transfection, expression in the PARP2 direction was analyzed by the luciferase assay, that toward the RPPH1 gene by RNase protection assay of an antisense RNA probe normalized to the expression of an  $\alpha$ -globin mRNA included as the internal standard. The results of these experiments clearly establish that the three promoters possess a bidirectional transcription activity (Figure 3 and Supplementary Figure S3). To further address whether the SBS identified *in silico* within these promoters (1–3 SBS in PARP2–RPPH1, C19orf6–U6.2 and MED16–U6.9) are indeed involved in the

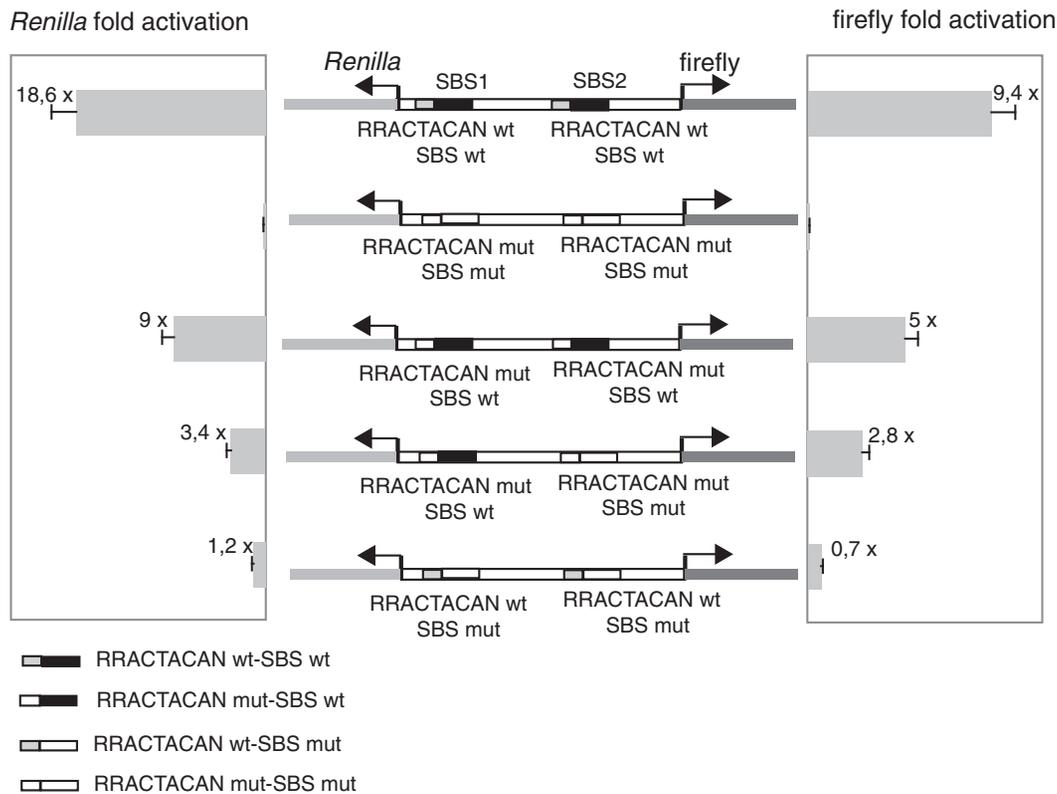
bidirectional activity, we examined again the effects of SBS substitutions on transcription activity. Mutation of the single SBS in PARP2–RPPH1 decreased expression in both directions (Supplementary Figure S3). Substitution of either SBS1 or SBS2 in MED16–U6.9, C19orf6–U6.2 decreased expression in the direction of the protein coding genes but, surprisingly, increased expression in the U6 direction. As expected, however, the simultaneous mutations of both SBS dramatically decreased the promoter activity in the U6 snRNA direction (Figure 3 and Supplementary Figure S3). Taken together, our data indicate that the three PARP2–RPPH1, C19orf6–U6.2 and MED16–U6.9 intergenic regions act as bidirectional promoters and that the simultaneous presence of two SBS in C19orf6–U6.2 and MED16–U6.9 clearly down-regulated the expression of the U6.2 and U6.9 genes.

#### **A DNA fragment containing an SBS, with or without the associated RRACACAN motif, is sufficient for bidirectional promoter activity**

Having shown that hStaf/ZNF143 binding and bidirectional transcription activity are correlated, we asked whether an SBS-containing DNA fragment isolated from a promoter driving unidirectional transcription is sufficient to lead to bidirectional transcription. To do this, we selected a 271 bp fragment from the BUB1B promoter that contains two functional SBS associated to the RRACACAN motif that was identified in our previous studies (16). HeLa cells were transiently transfected with constructs containing wt or mutant versions of the BUB1B promoter fragment inserted into a dual luciferase reporter vector. Figure 4 shows that the wt fragment actually possessed bidirectional promoter activity. However, the simultaneous mutations of the two RRACACAN and SBS motifs disabled completely the promoter, demonstrating that the bidirectional activity is directly due to the presence of the SBS and/or associated RRACACAN motifs (Figure 4). The bidirectional activity dropped ~14-fold following the simultaneous mutation of the two SBS whereas that of both RRACACAN motifs led to a 2-fold reduction (Figure 4). Lastly, we tested the transcription capacity of a construct containing only one SBS without the associated RRACACAN motif. Figure 4 shows that the presence of the sole SBS motif is sufficient for bidirectional activity. These results establish without ambiguity that the SBS is an element with sufficient intrinsic ability to direct efficient transcription activity in the direct and reverse orientations.

#### **Knock-down and overexpression of hStaf/ZNF143 altered expression from bidirectional gene pairs**

Finally, we wished to examine the functional importance of hStaf/ZNF143 in controlling the expression of gene pairs *in vivo*. To this end, we first measured the abundance of the corresponding mRNAs in HeLa cells in which expression of hStaf/ZNF143 was knocked-down by RNAi. Cells were treated with a mixture of siRNAs and the actual reduction of the hStaf/ZNF143 protein level was monitored by western blot. A decrease of >80% of hStaf/ZNF143 was obtained 72 h post-transfection; this



**Figure 4.** The staf binding site (SBS) bears the capacity to direct bidirectional transcription. Schematic diagrams of the reporter genes used in transfection assays and dual luciferase activities. The BUB1B promoter fragment  $-464/-194$  (relative to the translation initiation codon) containing two SBS (SBS1 and SBS2) and two RRACTACAN motifs was inserted into the dual luciferase vector. Boxed are RRACTACAN (light gray) and SBS (solid box) elements. mut RRACTACAN (open box) and mut SBS (open box) constructs contain ACTAC to TATGG and CCCA to TTTC substitutions, respectively. The relative transcription activities of the different constructs are indicated, representing the firefly and *Renilla* luciferase activity/empty vector ratios. Values are mean  $\pm$  SE (error bars),  $n = 3$  independent experiments with two replicates/group.

effect was specific because no change was observed with siRNA-control treated cells (Figure 5A). The hStaf/ZNF143 mRNA level also dropped specifically to  $>85\%$  in siRNA-treated cells (Figure 5B). The steady-state level of mRNAs produced by expression of the 10 gene pairs C11orf10–FEN1, KNTC1–RSRC2, C14orf133–AHSA1, TMEM186–PMM2, TRAK2–STRADB, ZMAT5–UCRC, COL4A3BP–POLK, DOCK4–ZNF277, NUDCD1–ENY2 and ZNF189–MRPL50 was measured by RT–qPCR in 72 h treated cells (Figure 5C). The bidirectional promoters in these gene pairs are contacted by hStaf/ZNF143 *in vivo* and gene expression was dependent on SBS integrity in transient transfection assays. Four independent RNAi experiments indicated that the mRNA levels obtained from the 10 gene pairs decreased between 63 and 97% in hStaf/ZNF143-depleted cells versus non-depleted cells (Figure 5C).

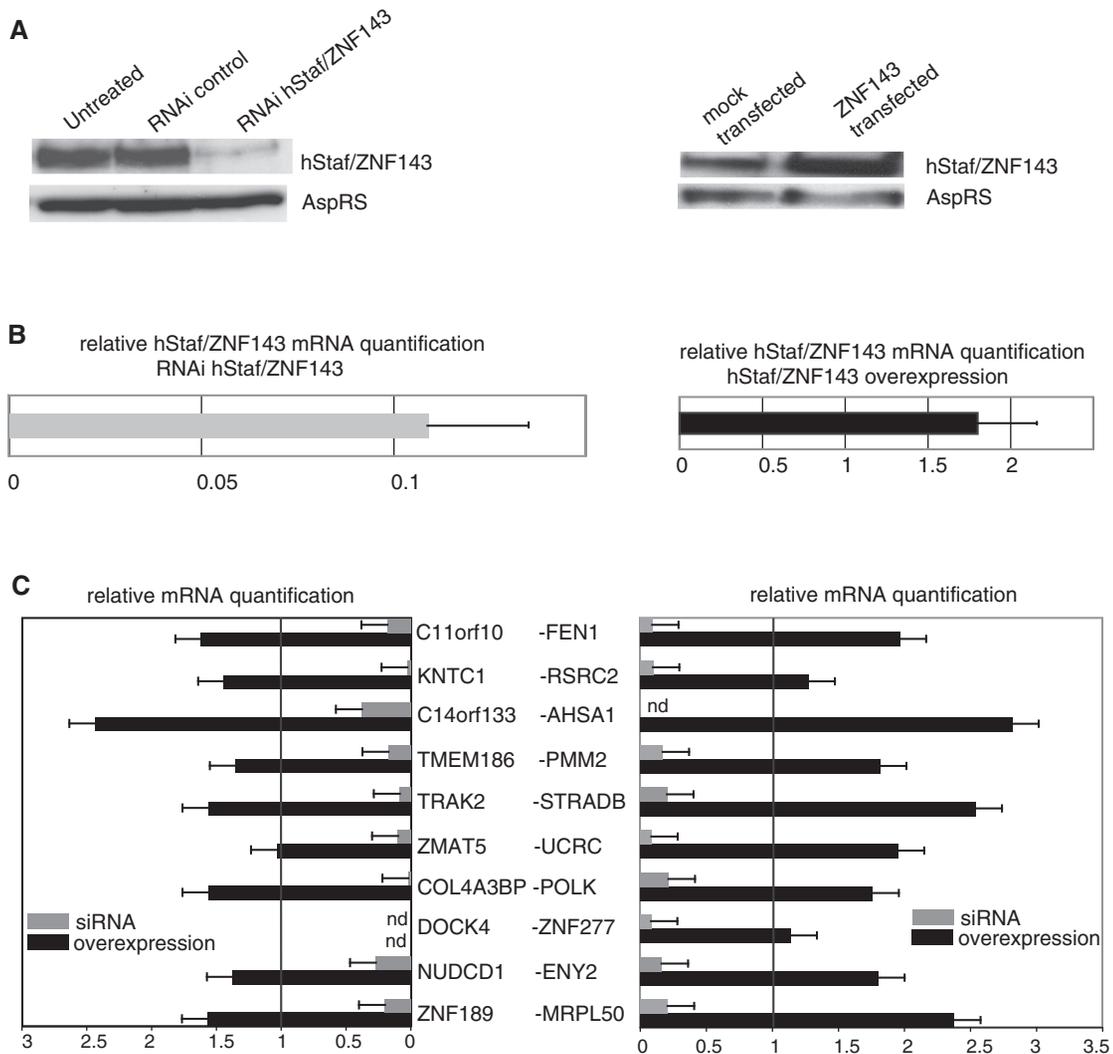
Next, we tested the effect of hStaf/ZNF143 overexpression on the mRNA level of the same gene pair set. Forty-eight hours after transfection of the recombinant vector expressing hStaf/ZNF143, we detected an  $\sim 2.5$ - and 1.8-fold increase of the hStaf/ZNF143 and hStaf/ZNF143 mRNA levels, respectively (Figure 5A and B). By measuring the mRNA content arising from 19 genes from 10 gene pairs in hStaf/ZNF143 overexpressing cells, we found a significant increase

(1.2–2.75-fold) in the level of a vast majority of the genes (18/19: 94%) (Figure 5C).

We conclude from these experiments that (i) variation in the mRNA levels from all the analyzed genes, following hStaf/ZNF143 depletion or overexpression, is consistent with the data obtained from *in vivo* functional assays (ChIP and luciferase assays); (ii) hStaf/ZNF143 positively controls expression of divergent gene pairs.

## DISCUSSION

In this article, with a higher quality data set of bidirectional promoters that we extracted from the human genome, we highlighted new characteristics of this particular promoter class and their associated genes. A bias toward the small size of genes associated to bidirectional promoters versus unidirectional promoters, and another one concerning the small size of bidirectional promoters were evidenced. It should be mentioned here that the peaks observed in the length distribution of the bidirectional promoters are correlated to the span of DNA wrapped around 1, 2 and 3 nucleosomes (one nucleosome spans 147 bp of DNA). Furthermore, by a genome scale analysis we have (i) demonstrated that the hStaf/ZNF143 binding sites are overrepresented in bidirectional promoters; (ii) defined the functional significance of its



**Figure 5.** Effects of the hStaf/ZNF143 knock-down and overexpression on the mRNA level produced by divergent gene pairs. (A) hStaf/ZNF143 protein was knocked-down in HeLa cells by RNAi. Cells, either treated with RNAi control or treated with hStaf/ZNF143 siRNAs were harvested 72 h after transfection. hStaf/ZNF143 protein overexpression was obtained by transfecting HeLa cells with a DNA construct containing the hStaf/ZNF143 cDNA under the control of the CMV promoter. Total cellular protein were subjected to western blotting with polyclonal antibodies against a C-terminal epitope of hStaf/ZNF143 or anti-AspRS as the internal control. (B) Total RNA was extracted from 72 h treated and control cells, and relative quantification of hStaf/ZNF143 mRNA was carried out by quantitative RT-PCR. The bar indicated the relative content of mRNA normalized to U3 snoRNA (endogenous control) in treated with respect to control cells, fixed as 1. Values are expressed as a ratio and results are mean  $\pm$  SD ( $n = 8$ ). Statistical analysis was performed using paired two-tailed Student's *t*-test. (C) Relative mRNA quantification by quantitative RT-PCR in 10 divergent gene pairs. The measures were performed for one of the partners only in the two gene pairs C14orf133-AHSA1 and DOCK4-ZNF277. The undetermined partner is marked as nd, not determined. Values are expressed as in (B).

recruitment in the expression control of divergent gene pairs. Our findings that hStaf/ZNF143 binds to at least 47% of the bidirectional promoters in the human genome, and that hStaf/ZNF143 Binding Sites (SBS) are 2.4-fold overrepresented in these promoters compared to unidirectional ones, constitute strong evidence that hStaf/ZNF143 is an essential regulator of bidirectional transcription. Direct experimental validation by ChIP on 87 promoters of protein coding gene pairs (c-c) yielded 97% success, indicating that very few of the identified sites were false positives. Importantly, the high rate of success demonstrates the robustness of the bioinformatic screen using the PSSM established from a pool of 347 SBS that were previously experimentally validated (18). Surprisingly, an

equivalent percentage of hStaf/ZNF143 binding was not observed with the promoters of c-nc and nc-nc gene pairs. Among the 13 tested promoters, only nine c-nc gene pairs (69%) provided positive amplification by ChIP experiments. A possible explanation might be that the identified SBS were located downstream of several of the ncRNA genes. Indeed, the window used for the bioinformatic screen covering the intergene and the first 400 bp downstream of the TSS, it could have encompassed in several cases a ncRNA gene which is of small size compared to a protein coding gene. This suggests that the presence of the SBS *per se* is not sufficient to recruit hStaf/ZNF143 but that the binding requires a motif located in the context of the gene promoter. Five of the

nine c–nc bidirectional promoters containing hStaf/ZNF143 were shown to associate a protein coding gene with a tRNA gene (LTA4H-tRNA<sup>Asp-GTC</sup>, PRMT5-tRNA<sup>Arg-ACG</sup>, POLG-tRNA<sup>Arg-TCG</sup>, PSMB3-tRNA<sup>AsnGTT</sup>, SLC27A4-tRNA<sup>Arg-TCT</sup>). This suggests a possible involvement of hStaf/ZNF143 in the transcription of these tRNA genes by RNA polymerase III. In the other four identified c-nc gene pairs, the ncRNA partner is either RPPH1 (the RNase P RNA), the SRP-RNA (RNA 7SL1) or a U6 snRNA gene. The presence of a bidirectional promoter connecting the RPPH1 and PARP2 mouse genes was previously described but without establishing the involvement of hStaf/ZNF143 in the control of gene expression (24). Concerning the SRP-RNA, this is the first report which, to our knowledge, described that the SRP RNA (RNA 7SL1) and RPS29 genes share the same promoter bound to the hSTAF/ZNF143 factor. In the human genome, five functional U6 snRNA genes have been identified (25) and we showed here that the U6.2 and U6.9 genes are associated to a protein coding gene (C19orf6–snRNAU6.2 and MED16–snRNAU6.9) in a bidirectional promoter containing two functional SBS. We had previously demonstrated that RNA polymerase III transcription of the human U6.1 gene requires Staf, and that the transcriptional activation is performed via a single SBS (9,26). In the case of the bidirectional promoters, the mutation of either of the two SBS caused a reduction of the transcription activity in the direction of the protein gene. This result was expected as is the simultaneous mutation of each SBS which reduced transcription activity in both directions. However, and much to our surprise, the mutation of one single SBS, whatever it was, led to enhanced U6 expression. This surprising result must be examined in the context of the snRNA gene promoter architecture. In fact, all the snRNA gene promoters characterized so far contain only one single SBS (26) and our report is the first describing the presence of two SBS in the promoter of an snRNA gene. In the light of our data, it looks as if the presence of two SBS is incompatible with the optimal formation of transcription complexes and is detrimental to U6 expression.

Our bioinformatic screen showed that 21% of the identified SBS were associated with the 9bp functional RRACTACAN motif. The transcription level of constructs containing only the RRACTACAN motif is very low. Paradoxically, mutation of the RRACTACAN motif in the presence of a wt SBS yielded a 50% drop of transcription efficiency, suggesting an important role of this motif, but only in the context of the SBS. It may well be that this motif serves as a binding site for an unknown transcription factor. It might also induce a particular DNA local structure enhancing the affinity of hStaf/ZNF143 to its cognate sequence. The RRACTACAN sequence is particularly interesting in the light of the work describing the discovery of motifs overrepresented in human bidirectional promoters (6). This analysis yielded five motifs, including the NRF-1, GABPA, YY1 and NFY transcription factor binding sites, and the fifth one of sequence ACTACANNTCCC with no known

identified transcription factor. It was predicted that the ACTACANNTCCC sequence and the cognate transcription factor would play an important role in regulating bidirectional transcription. Our work established clearly that ACTACAN indeed corresponds to the 3'-part of the RRACTACAN motif that we identified and that NTCCC represents SBS residues 1–5. Lin *et al.* (6) did not report residues 6–18 of the SBS as overrepresented in bidirectional promoters. This is very likely because of the lower sequence conservation in this part of the SBS. A similar motif with the ACTAYRNNNCCCR consensus sequence was previously reported by Xie *et al.* (27) who ranked it fourth among 174 motifs in terms of conservation across several mammalian transcription factors. In the present work, we showed that a significant part (20%) of the identified SBS is linked to the RRACTACAN motif. If this motif is the target for an unknown transcription factor, we propose that this factor is also an essential actor working in conjunction with hStaf/ZNF143 to regulate bidirectional transcription.

Finally, we have also demonstrated that insertion of the unidirectional BUB1B promoter [containing functional SBS (16)] into the dual luciferase reporter was sufficient to induce bidirectional transcription. The observation that a unidirectional promoter can drive transcription in two opposite directions should be examined in the light of recent data which collectively established that promoters can support divergent initiation of transcription but productive elongation occurs only in the direction of the genes (28–34). In an earlier work, we had demonstrated that hStaf/ZNF143 binds to and controls expression from a large number of unidirectional promoters (18). The presence of a single SBS is sufficient to direct bidirectional activity and this intrinsic ability of hStaf/ZNF143 suggests that hStaf/ZNF143 can be involved in divergent transcription from unidirectional promoters (28–34). The transcription factor GABPA/NRF2 (7) also bears the capacity to direct bidirectional transcription and, to our knowledge, hStaf/ZNF143 and GABPA/NRF2 are the two sole factors described so far with this particular ability.

In conclusion, we have shown that hStaf/ZNF143 binds to and governs expression from a majority of bidirectional promoters. The picture emerging from our current and previous studies strengthens the role of hStaf/ZNF143 as an essential factor for controlling gene expression in humans.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We are grateful to H.T. Jacobs (University of Tampere, Finland) for the gift of the pFRLN50 reporter vector and M. Frugier for anti-AspRS. We also thank C. Graber and T. Strub, two undergraduate students, for their involvement in the preparation of several constructs; S. Baudrey and A. Schweigert for valuable technical assistance. We

are also grateful to E. Westhof for his interest in our study.

## FUNDING

Ministère de l'Enseignement Supérieur et de la Recherche, Allocation de Recherche (to Y.-N.A and R.P.N.-M). Funding for open access charge: CNRS.

*Conflict of interest statement.* None declared.

## REFERENCES

- Adachi,N. and Lieber,M.R. (2002) Bidirectional gene organization: a common architectural feature of the human genome. *Cell*, **109**, 807–809.
- Trinklein,N.D., Aldred,S.F., Hartman,S.J., Schroeder,D.I., Otililar,R.P. and Myers,R.M. (2004) An abundance of bidirectional promoters in the human genome. *Genome Res.*, **14**, 62–66.
- Koyanagi,K.O., Hagiwara,M., Itoh,T., Gojobori,T. and Imanishi,T. (2005) Comparative genomics of bidirectional gene pairs and its implications for the evolution of a transcriptional regulation system. *Gene*, **353**, 169–176.
- Yang,M.Q. and Elnitski,L.L. (2008) Diversity of core promoter elements comprising human bidirectional promoters. *BMC Genomics*, **9**(Suppl. 2), S3.
- Engstrom,P.G., Suzuki,H., Ninomiya,N., Akalin,A., Sessa,L., Lavorgna,G., Brozzi,A., Luzi,L., Tan,S.L., Yang,L. *et al.* (2006) Complex Loci in human and mouse genomes. *PLoS Genet.*, **2**, e47.
- Lin,J.M., Collins,P.J., Trinklein,N.D., Fu,Y., Xi,H., Myers,R.M. and Weng,Z. (2007) Transcription factor binding and modified histones in human bidirectional promoters. *Genome Res.*, **17**, 818–827.
- Collins,P.J., Kobayashi,Y., Nguyen,L., Trinklein,N.D. and Myers,R.M. (2007) The ets-related transcription factor GABP directs bidirectional transcription. *PLoS Genet.*, **3**, e208.
- Schuster,C., Myslinski,E., Krol,A. and Carbon,P. (1995) Staf, a novel zinc finger protein that activates the RNA polymerase III promoter of the selenocysteine tRNA gene. *EMBO J.*, **14**, 3777–3787.
- Schaub,M., Myslinski,E., Schuster,C., Krol,A. and Carbon,P. (1997) Staf, a promiscuous activator for enhanced transcription by RNA polymerases II and III. *EMBO J.*, **16**, 173–181.
- Myslinski,E., Krol,A. and Carbon,P. (1998) ZNF76 and ZNF143 are two human homologs of the transcriptional activator Staf. *J. Biol. Chem.*, **273**, 21998–22006.
- Myslinski,E., Ame,J.C., Krol,A. and Carbon,P. (2001) An unusually compact external promoter for RNA polymerase III transcription of the human H1RNA gene. *Nucleic Acids Res.*, **29**, 2502–2509.
- Schaub,M., Krol,A. and Carbon,P. (1999) Flexible zinc finger requirement for binding of the transcriptional activator staf to U6 small nuclear RNA and tRNA(Sec) promoters. *J. Biol. Chem.*, **274**, 24241–24249.
- Schaub,M., Myslinski,E., Krol,A. and Carbon,P. (1999) Maximization of selenocysteine tRNA and U6 small nuclear RNA transcriptional activation achieved by flexible utilization of a Staf zinc finger. *J. Biol. Chem.*, **274**, 25042–25050.
- Schaub,M., Krol,A. and Carbon,P. (2000) Structural organization of Staf-DNA complexes. *Nucleic Acids Res.*, **28**, 2114–2121.
- Gerard,M.A., Krol,A. and Carbon,P. (2007) Transcription factor hStaf/ZNF143 is required for expression of the human TFAM gene. *Gene*, **401**, 145–153.
- Myslinski,E., Gerard,M.A., Krol,A. and Carbon,P. (2007) Transcription of the human cell cycle regulated BUB1B gene requires hStaf/ZNF143. *Nucleic Acids Res.*, **35**, 3453–3464.
- Gerard,M.A., Myslinski,E., Chylak,N., Baudrey,S., Krol,A. and Carbon,P. (2010) The scaRNA2 is produced by an independent transcription unit and its processing is directed by the encoding region. *Nucleic Acids Res.*, **38**, 370–381.
- Myslinski,E., Gerard,M.A., Krol,A. and Carbon,P. (2006) A genome scale location analysis of human Staf/ZNF143-binding sites suggests a widespread role for human Staf/ZNF143 in mammalian promoters. *J. Biol. Chem.*, **281**, 39953–39962.
- Nguyen,H., Berthommier,G., Friedrich,A., Poidevin,L., Ripp,R., Moulinier,L. and Poch,O. (2008) *Introduction du nouveau centre de données biomédicales Décrypton*, CORIA 2008, Hermes Edition.
- Cartharius,K., Frech,K., Grote,K., Klocke,B., Haltmeier,M., Klingenhoff,A., Frisch,M., Bayerlein,M. and Werner,T. (2005) MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics*, **21**, 2933–2942.
- Zanotto,E., Shah,Z.H. and Jacobs,H.T. (2007) The bidirectional promoter of two genes for the mitochondrial translational apparatus in mouse is regulated by an array of CCAAT boxes interacting with the transcription factor NF-Y. *Nucleic Acids Res.*, **35**, 664–677.
- Lobo,S.M. and Hernandez,N. (1989) A 7 bp mutation converts a human RNA polymerase II snRNA promoter into an RNA polymerase III promoter. *Cell*, **58**, 55–67.
- Sandelin,A., Carninci,P., Lenhard,B., Ponjavic,J., Hayashizaki,Y. and Hume,D.A. (2007) Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat. Rev. Genet.*, **8**, 424–436.
- Ame,J.C., Schreiber,V., Fraulob,V., Dolle,P., de Murcia,G. and Niedergang,C.P. (2001) A bidirectional promoter connects the poly(ADP-ribose) polymerase 2 (PARP-2) gene to the gene for RNase P RNA. structure and expression of the mouse PARP-2 gene. *J. Biol. Chem.*, **276**, 11092–11099.
- Domitrovich,A.M. and Kunkel,G.R. (2003) Multiple, dispersed human U6 small nuclear RNA genes with varied transcriptional efficiencies. *Nucleic Acids Res.*, **31**, 2344–2352.
- Hernandez,N. (2001) Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *J. Biol. Chem.*, **276**, 26733–26736.
- Xie,X., Lu,J., Kulbokas,E.J., Golub,T.R., Mootha,V., Lindblad-Toh,K., Lander,E.S. and Kellis,M. (2005) Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature*, **434**, 338–345.
- Core,L.J. and Lis,J.T. (2008) Transcription regulation through promoter-proximal pausing of RNA polymerase II. *Science*, **319**, 1791–1792.
- Preker,P., Nielsen,J., Kammler,S., Lykke-Andersen,S., Christensen,M.S., Mapendano,C.K., Schierup,M.H. and Jensen,T.H. (2008) RNA exosome depletion reveals transcription upstream of active human promoters. *Science*, **322**, 1851–1854.
- Seila,A.C., Calabrese,J.M., Levine,S.S., Yeo,G.W., Rahl,P.B., Flynn,R.A., Young,R.A. and Sharp,P.A. (2008) Divergent transcription from active promoters. *Science*, **322**, 1849–1851.
- Jacquier,A. (2009) The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nat. Rev. Genet.*, **10**, 833–844.
- Neil,H., Malabat,C., d'Aubenton-Carafa,Y., Xu,Z., Steinmetz,L.M. and Jacquier,A. (2009) Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature*, **457**, 1038–1042.
- Preker,P., Nielsen,J., Schierup,M.H. and Jensen,T.H. (2009) RNA polymerase plays both sides: vivid and bidirectional transcription around and upstream of active promoters. *Cell Cycle*, **8**, 1106–1107.
- Xu,Z., Wei,W., Gagneur,J., Perocchi,F., Clauder-Munster,S., Camblong,J., Guffanti,E., Stutz,F., Huber,W. and Steinmetz,L.M. (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature*, **457**, 1033–1037.