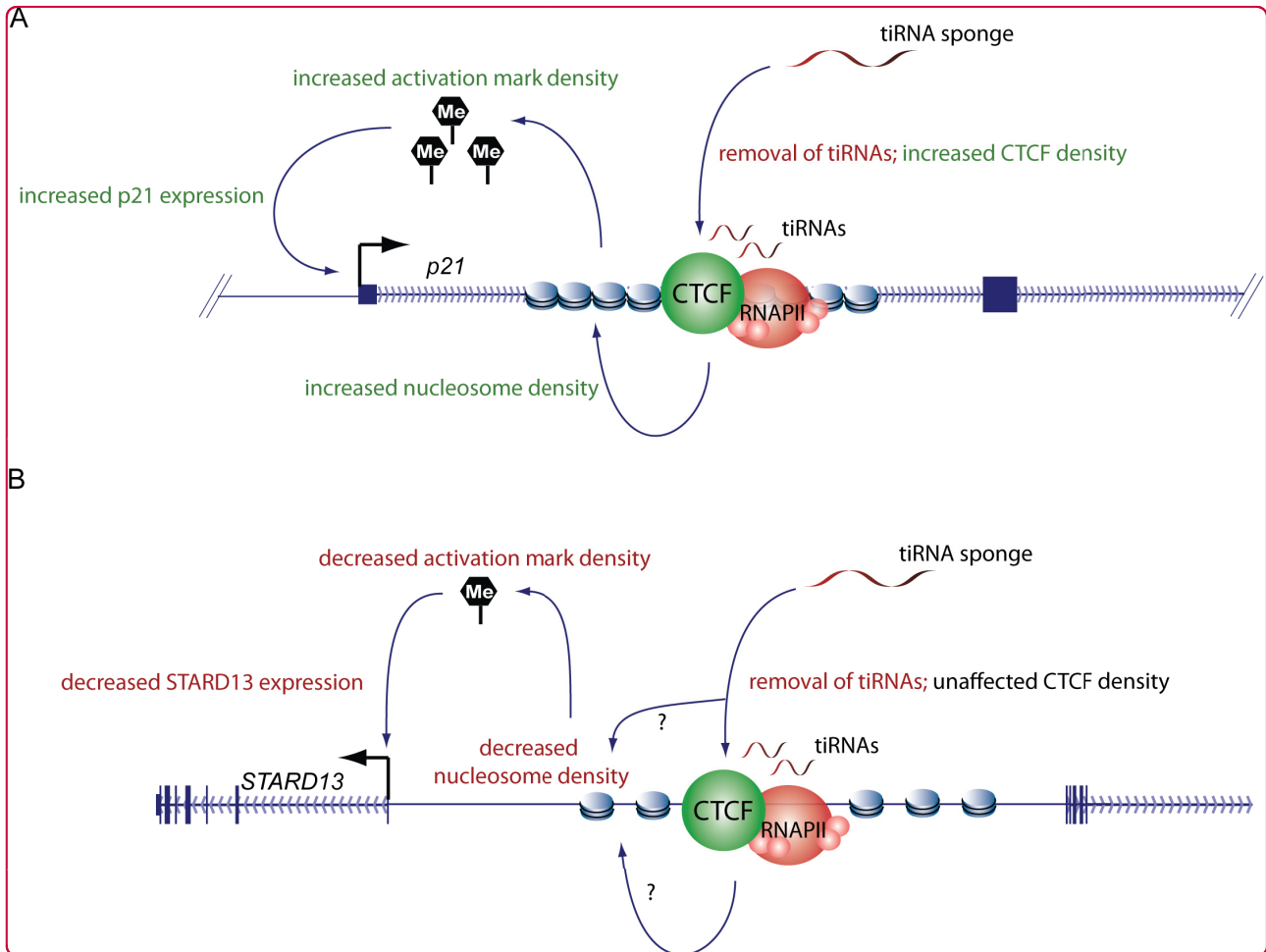


# Epigenetics & Chromatin



## The relationship between transcription initiation RNAs and CCCTC-binding factor (CTCF) localization

Taft *et al.*

RESEARCH

Open Access

# The relationship between transcription initiation RNAs and CCCTC-binding factor (CTCF) localization

Ryan J Taft<sup>1</sup>, Peter G Hawkins<sup>2,3</sup>, John S Mattick<sup>1</sup> and Kevin V Morris<sup>2\*</sup>

## Abstract

**Background:** Transcription initiation RNAs (tiRNAs) are nuclear localized 18 nucleotide RNAs derived from sequences immediately downstream of RNA polymerase II (RNAPII) transcription start sites. Previous reports have shown that tiRNAs are intimately correlated with gene expression, RNA polymerase II binding and behaviors, and epigenetic marks associated with transcription initiation, but not elongation.

**Results:** In the present work, we show that tiRNAs are commonly found at genomic CCCTC-binding factor (CTCF) binding sites in human and mouse, and that CTCF sites that colocalize with RNAPII are highly enriched for tiRNAs. To directly investigate the relationship between tiRNAs and CTCF we examined tiRNAs originating near the intronic CTCF binding site in the human tumor suppressor gene, *p21* (cyclin-dependent kinase inhibitor 1A gene, also known as *CDKN1A*). Inhibition of CTCF-proximal tiRNAs resulted in increased CTCF localization and increased *p21* expression, while overexpression of CTCF-proximal tiRNA mimics decreased CTCF localization and *p21* expression. We also found that tiRNA-regulated CTCF binding influences the levels of trimethylated H3K27 at the alternate upstream *p21* promoter, and affects the levels of alternate *p21* (*p21<sup>alt</sup>*) transcripts. Extending these studies to another randomly selected locus with conserved CTCF binding we found that depletion of tiRNA alters nucleosome density proximal to sites of tiRNA biogenesis.

**Conclusions:** Taken together, these data suggest that tiRNAs modulate local epigenetic structure, which in turn regulates CTCF localization.

## Background

In addition to mRNAs, it is now clear that most eukaryotic genic loci generate a complex network of overlapping short (<200) and long non-protein coding RNA (ncRNA) species [1-3]. This growing catalog of ncRNAs includes a host of small RNA (sRNA) transcripts proximal to transcription start sites (TSSs) [4], some of which are capped [5] or associate with polycomb repressive complex (PCR2) components [6,7]. We have recently described a set of nuclear localized dominantly 18 nucleotide tiny RNAs that are generated from regions immediately downstream of TSSs and conserved across metazoa but absent in fungi and plants [8,9]. We

have previously suggested that these transcription initiation RNAs (tiRNAs) may be connected to epigenetic regulation in light of the fact that (i) they originate from the same position relative to the +1 nucleosome in evolutionarily distant animals, suggesting that expression of tiRNAs is influenced by the position of the nucleosome, or vice versa, (ii) they are intimately connected with peaks of RNA polymerase II (RNAPII) binding and levels of gene expression, and (iii) they are enriched genome-wide at chromatin marks associated with the activation, but not elongation, of transcription [4,8,9]. We have also observed that tiRNAs are enriched at sites of CCCTC-binding factor (CTCF) [9], an enigmatic epigenetic regulator that has been recently dubbed 'the master weaver' of the genome [10].

CTCF is a highly conserved zinc finger protein associated with a diverse set of phenomena including epigenetic insulation, imprinting and transcriptional

\* Correspondence: kmorris@scripps.edu

<sup>2</sup>Department of Molecular and Experimental Medicine, The Kellogg School of Science and Technology, The Scripps Research Institute, La Jolla, CA 92037, USA

Full list of author information is available at the end of the article

regulation [10-13]. Intriguingly, CTCF has been shown to both positively and negatively regulate gene expression in a gene-specific and context-specific manner [10,14]. A resolution to this apparent incongruity has recently been proposed: CTCF does not directly influence the surrounding genes or transcriptional machinery, but rather acts as a three-dimensional orchestrator of chromatin architecture [10]. CTCF's involvement in a wide range of epigenetic phenomena appears to be the secondary, but undoubtedly regulated, effects of its ability to form specific intrachromosomal and interchromosomal connections [10].

CTCF has been shown to regulate the expression of several tumor suppressor genes, including *p21* (informal gene name for the cyclin-dependent kinase inhibitor 1A gene, *CDKN1A*) [11,15] and *p16* (INK4a), the latter by insulating the promoter from silent-state histone modifications such as H3K27 trimethylation (H3K27me3) [16]. CTCF has also recently been shown to be involved in the epigenetic regulation of *frataxin* (*FXN*), a gene mutated and silenced in Friedreich ataxia, which causes progressive damage to the nervous system [17]. Loss of CTCF binding in the 5' untranslated region (UTR) of *FXN* leads to a deficiency of the *FXN* transcript, an increase in *FXN antisense transcript 1*, and heterochromatin formation involving the +1 nucleosome [17]. Given that tiRNAs and RNAPII are intimately connected, and that there is increasing evidence that CTCF and RNAPII are coupled together (see below) [18-22], we speculated that tiRNAs at CTCF-binding sites might be involved in the alteration of local chromatin states, and therefore transcript expression, via indirect regulation CTCF.

## Results and discussion

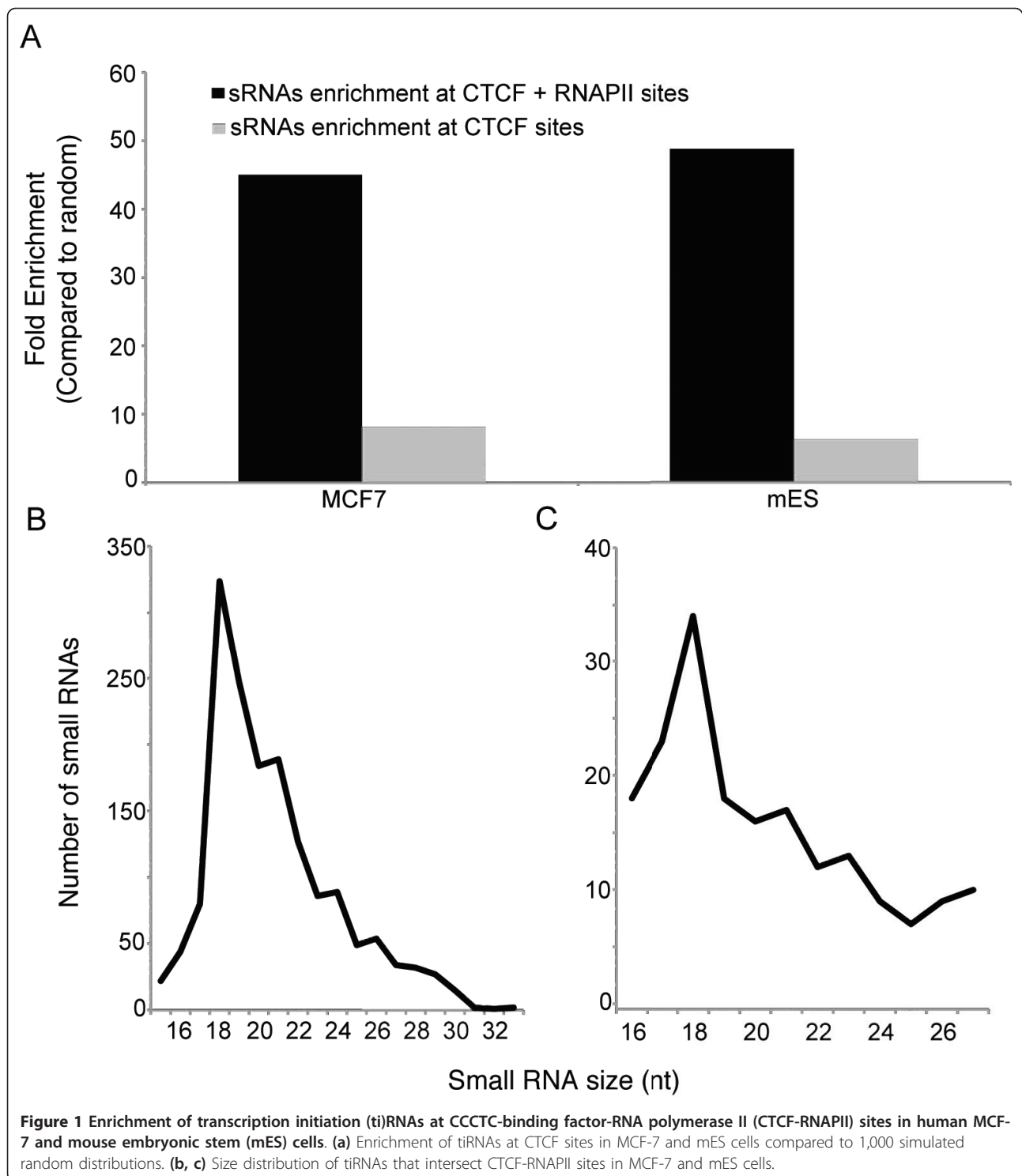
We have previously shown that tiRNAs isolated from THP-1 cells (a human monocytic leukemia cell line) are systematically enriched at white blood cell CTCF binding sites [9]. To examine if this relationship is preserved across cell and tissue types, and multiple species, we interrogated small RNA enrichments at CTCF binding sites in MCF-7 breast cancer cells and mouse embryonic stem cells (mESCs) (for a full list of data sources please see (Additional file 1, Table S1).

Consistent with prior work we found that tiRNAs derived from both MCF-7 and mESCs are enriched approximately sixfold at CTCF binding sites that sit outside TSSs or other annotated genomic features (see Methods and Figure 1a), and show the characteristic 18 nucleotide tiRNA peak (Figure 1b, c). When CTCF binding sites were further refined to include only sites coincident with RNA polymerase II binding (CTCF-RNAPII sites), tiRNA enrichments increased considerably, to approximately 45-fold. Indeed, more than 50%

and 20% of MCF-7 and mESC CTCF-RNAPII sites intersect with tiRNAs, respectively (Tables 1 and 2). This relationship appears to bridge the reports indicating that tiRNA biogenesis is a direct result of RNAPII backtracking and nascent transcript cleavage [4,8,9], and recent studies showing that CTCF is directly involved in RNAPII function. Indeed, it has now become clear that (i) a subpopulation of CTCF directly interacts with the large subunit of RNAPII through its phosphorylated C-terminal tail [21,22], (ii) that in some cases a single CTCF site is both necessary and sufficient to drive RNAPII transcription in the absence of canonical promoters by recruitment of RNAPII [21,22], and (iii) that CTCF specificity for, and regulation of, transcriptionally competent complexes also extends to RNA polymerase I [18-20].

To examine if the association between tiRNAs, CTCF and RNAPII extends beyond MCF-7 and mESCs we identified CTCF sites conserved across an additional eight human cell lines (GM12878, HepG2, HMEC, HSMM, HUVEC, K562, NHEK, and NHLF cells) [23,24] and RNAPII sites conserved across three (K562, GM12878, and HUVEC), and intersected them with nuclear and cytoplasmic small RNAs (sRNAs) from THP-1 and 5-8f cells (a nasopharyngeal carcinoma cell line [25]) and MCF-7 total sRNAs. Despite the fact that these datasets are derived from disparate origins, nuclear sRNAs from THP-1 and 5-8f are 33-fold and 16-fold enriched, respectively, and total sRNAs from MCF cells are 31-fold enriched at conserved CTCF-RNAPII sites (Additional file 2, Figure S1a). Additionally, like the MCF-7 and mESC datasets discussed above, the small RNAs that overlap CTCF-RNAPII sites are dominantly 18 nucleotides, indicating they are tiRNAs (Additional file 2, Figure S1b-d). Overall, greater than 10% of the conserved CTCF sites, and 60% of conserved CTCF-RNAPII sites, overlap with sequences that generate tiRNAs (Additional file 3, Table S2). To further ensure that the tiRNA enrichment at CTCF-RNAPII sites was robust we parsed the conserved human CTCF sites into two groups by origin, 'cancer' and 'normal', and removed all CTCF sites that overlapped with TSSs, repeat masker annotations and small RNAs. Using the most robust RNAPII datasets in each group (MCF-7 and HUVEC for 'cancer' and 'normal', respectively), we found that this dramatically reduced set still shows robust enrichment for tiRNAs at CTCF-RNAPII sites (Additional file 4, Figure S2).

To experimentally interrogate the tiRNA-CTCF-RNAPII relationship we queried for sites in clinically relevant genes and identified a CTCF-RNAPII site with tiRNAs in the first intron of *p21*, which is conserved across both multiple human cell types (Figure 2) and mammalian species (Additional file 5, Figure S3). *CDKN1A/p21*



is a significant tumor suppressor that acts at the G1 checkpoint to inhibit cell cycle progression [26-29], and its downregulation (but not mutation) is a common feature of many cancers [30-34]. In addition to *p21* mRNA, the *p21* locus encodes a number of other transcripts,

including alternative *p21* transcripts (*p21<sup>alt</sup>*) that originate from a unique promoter located approximately 2 kb upstream of the canonical *p21* transcription start site and include the majority of the *p21* coding regions in their final spliced products [35], and a long non-coding

**Table 1 MCF-7 CCCTC-binding factor (CTCF), RNA polymerase II (RNAPII) and CTCF-RNAPII site intersections with small RNAs (sRNAs)**

	Total number of MCF-7 sites	Number that overlap with sRNAs (%)	Total number of sRNAs overlapped (percentage total sRNAs)
RNAPII	10,821	1,165 (11)	2,333 (2)
CTCF	23,857	2,047 (9)	4,731 (4)
CTCF + RNAPII	936	468 (50)	1,609 (1.6)

antisense RNA (*bx332409*) that regulates local epigenetic states [36] (Figure 2).

The *p21* locus encodes two tiRNA clusters, one at the TSS (tss-tiRNAs) and the other at the CTCF-RNAPII site (CTCF proximal (cp)-tiRNAs) that are antisense to one another. The tss-tiRNAs are sense to the gene (as observed generally), while the cp-tiRNAs are antisense. Both overlap distinct peaks of RNAPII binding, suggesting that their biogenesis is tied to RNAPII molecules heading in opposite directions, possibly linked to nucleosome position [4], and this reinforces our previous finding that tiRNAs are found at sites of active RNAPII transcription initiation outside of canonical transcription start sites (Figure 2).

To investigate the function of *p21* tiRNAs, we utilized short antisense 'sponge' RNAs [37] that were designed to bind and inhibit tss-tiRNAs and cp-tiRNAs (Figure 2). MCF-7 cells transfected with the cp-tiRNA sponge demonstrated a significant increase of *p21* mRNA and *p21<sup>alt</sup>* expression, as measured by quantitative PCR (qPCR) (Figure 3a, b). In contrast, the tss-tiRNA sponge did not exhibit a detectable effect on *p21* expression (Figure 3a, b), and thus cp-tiRNAs became the focus of the remainder of this study.

As reverse transcription in the qPCR samples was not specifically primed (Figure 3a, b), these transcripts might represent sense and/or antisense transcripts associated with these regions [36] or any of the plethora of splice variants. To determine the extent of the effect that the cp-tiRNA sponge has on relative sense and antisense *p21* transcript levels, strand-specific reverse transcription PCR (RT-PCR) was performed. Upon treatment with the cp-tiRNA sponge, *p21* mRNA, sense *p21<sup>alt</sup>*, and antisense *p21<sup>alt</sup>* transcript levels increased, whereas transcripts antisense to *p21* mRNA were

unaffected (Figure 3c, d). These data indicate that CTCF-proximal tiRNAs may be involved in the negative regulation of *p21*.

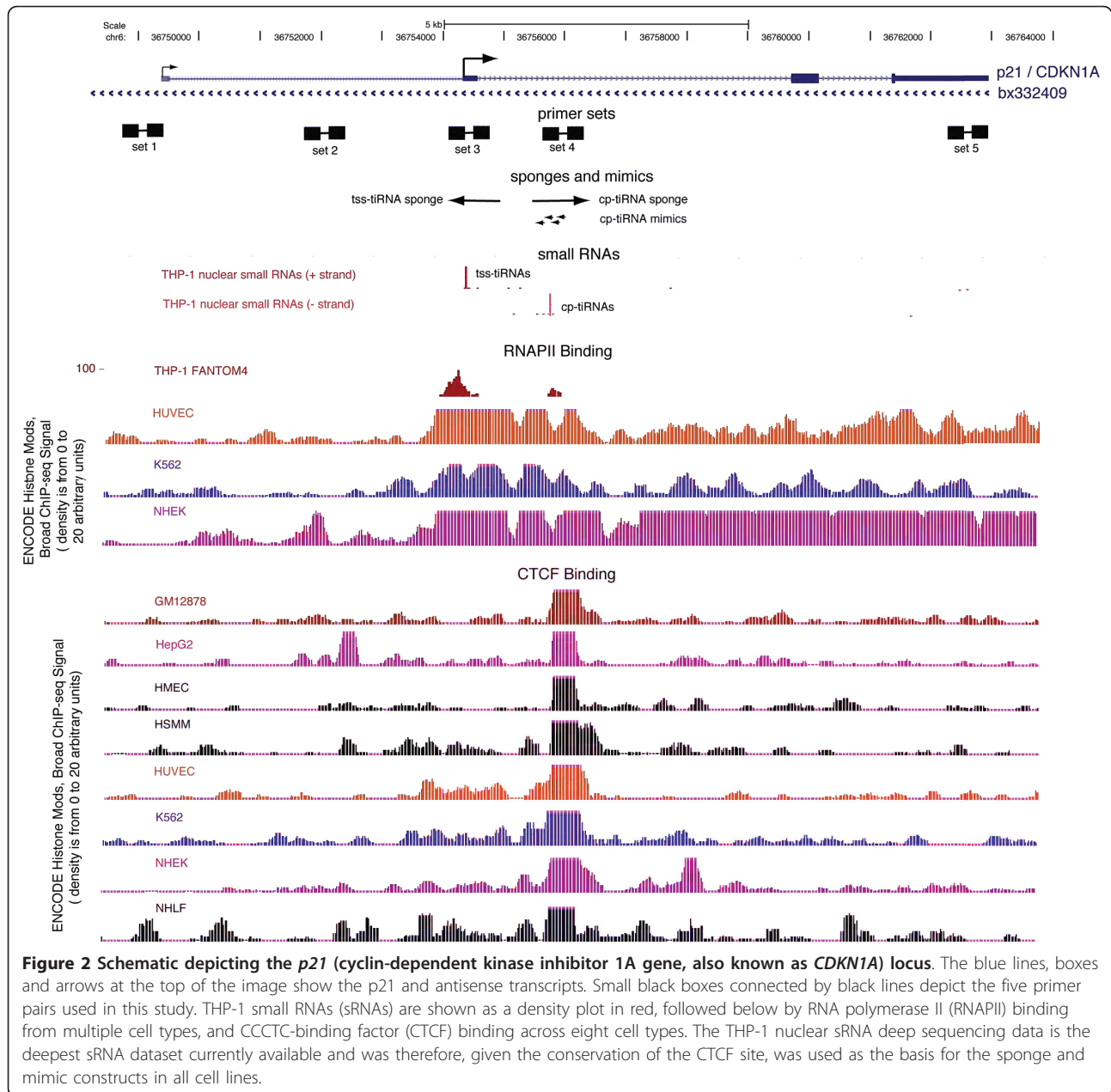
We next performed the reciprocal experiment testing the effect that overexpression of CTCF-proximal tiRNA mimics has on *p21* expression. Consistent with our speculation that tiRNAs are connected to transcriptional regulation, we found that overexpression of a set of four cp-tiRNA mimics resulted in a marked reduction of the *p21* mRNA (Figure 3e). To confirm that the effect of the cp-tiRNA sponges and mimics was not restricted to MCF-7 cells we repeated these experiments in THP-1 cells and found that the principal results were recapitulated (Additional file 6, Figure S4), indicating that cp-tiRNAs have a regulatory effect on *p21* transcription in multiple human cell systems.

To further investigate the effects of cp-tiRNA sponge and mimics on *p21* transcription, elongating forms of RNAPII were assessed by chromatin immunoprecipitation-PCR (ChIP-PCR). The only signal increase appeared in regions overlapping *p21<sup>alt</sup>*, although that increase was modest (Figure 4), suggesting that cp-tiRNAs do not function by affecting local RNAPII densities, but rather by directly or indirectly modulating local chromatin architecture.

To explore this possibility we examined the effects of cp-tiRNAs sponge and mimic constructs on CTCF localization, and on epigenetic marks at various locations within the *p21* locus by ChIP. The density of the silent state chromatin mark, H3K27me3, did not change upon introduction of cp-tiRNA sponge or mimic constructs at their perfectly complementary target sites (that is, at sites of tiRNA biogenesis; Figure 5a, b), as would be expected if the cp-tiRNA mimic or sponges were themselves altering local chromatin

**Table 2 Mouse embryonic stem (mES) CCCTC-binding factor (CTCF), RNA polymerase II (RNAPII) and CTCF-RNAPII site intersections with small RNAs (sRNAs)**

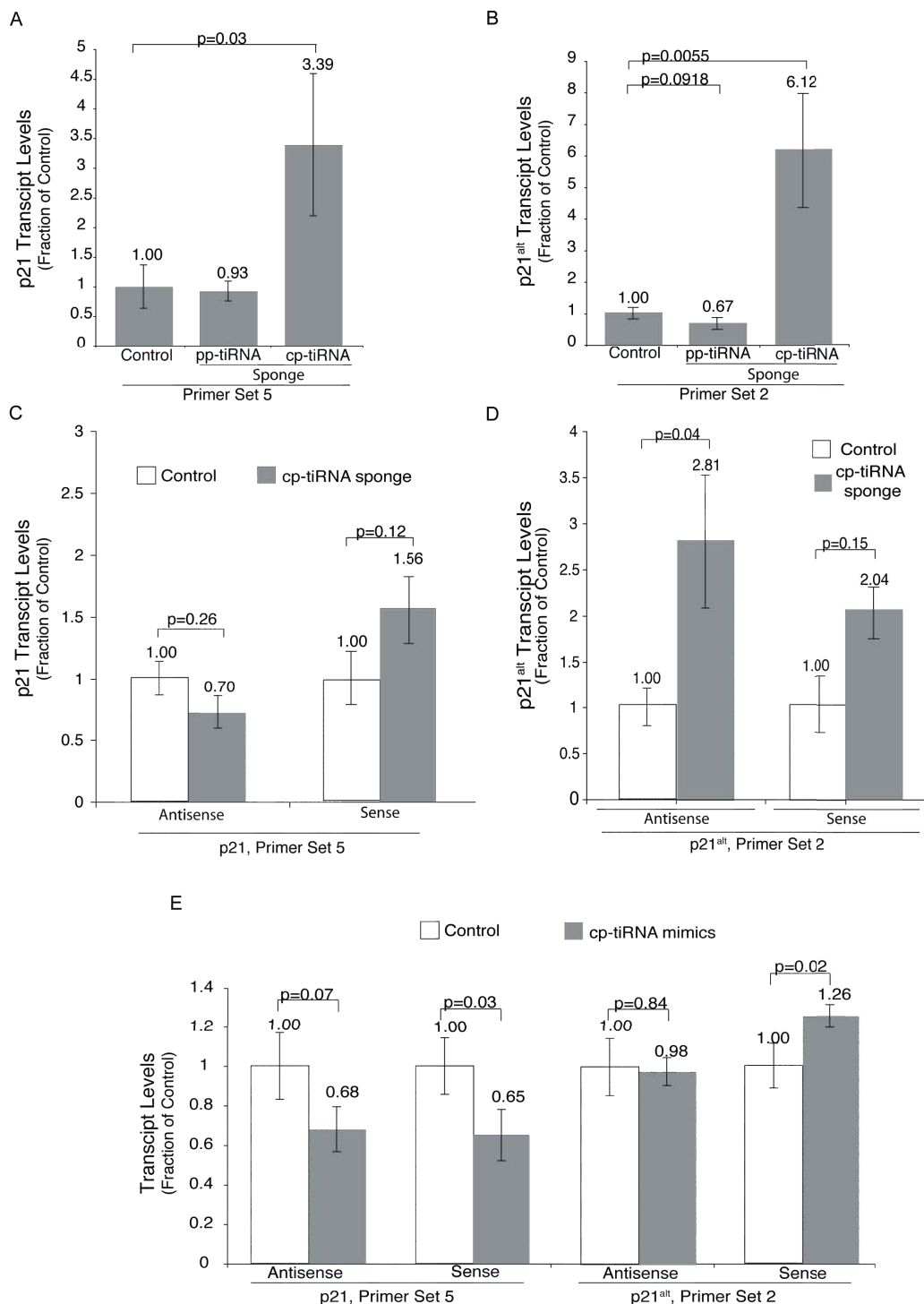
	Total number of mES sites	Number that overlap with sRNAs (%)	Total number of sRNAs overlapped (percentage total sRNAs)
RNAPII	8,421	575 (7)	1,775 (3)
CTCF	13,469	589 (4)	981 (1.5)
CTCF + RNAPII	434	86 (20)	186 (0.3)



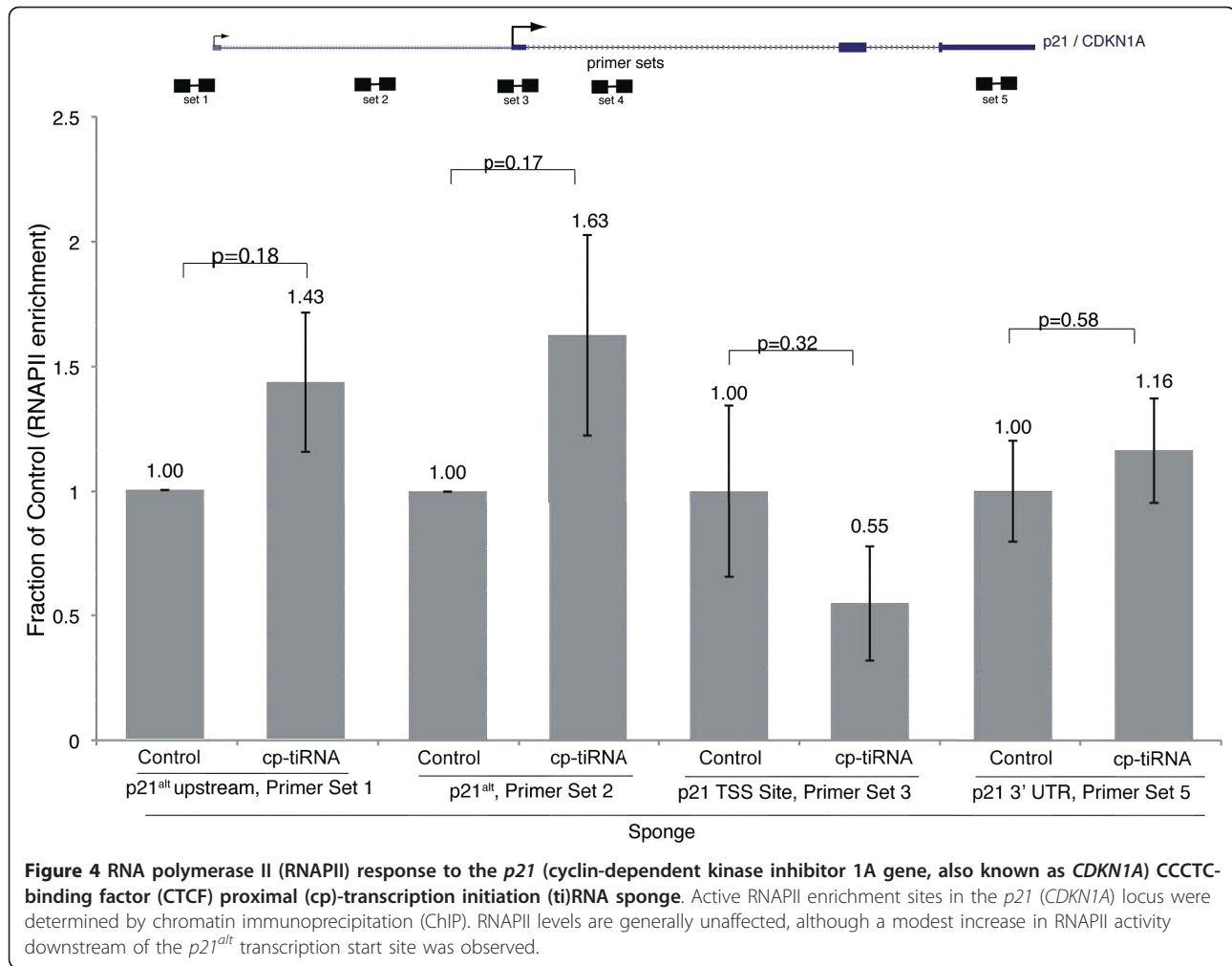
status, as has been observed previously with small non-coding RNAs associated with transcriptional gene silencing [38]. However, H3K27me3 levels upstream of the *p21<sup>alt</sup>* transcription start site were decreased upon cp-tiRNA sponge treatment (Figure 5c). Given that the distance between the site of tiRNA biogenesis and the *p21<sup>alt</sup>* promoter is greater than 6 kilobases, we speculated that these effects are facilitated by tiRNA-mediated regulation of other epigenetic regulators capable of acting at long distances.

Consistent with this, treatment with the cp-tiRNA sponge resulted in a significant increase in CTCF

binding (Figure 5a), and overexpression of the cp-tiRNA mimics exhibited a significant decrease of CTCF binding (Figure 5b). This indicates that the effect of cp-tiRNAs on *p21* transcription is directly related to its ability to modulate CTCF binding, which may be involved in three-dimensional (re)ordering of the *p21* locus. Indeed, western blot analysis showed that *p21* protein levels were increased in samples treated with the cp-tiRNA sponge, and decreased in samples treated with the cp-tiRNA mimic constructs (Figure 5d, e). Taken together, these data suggest that one function of *p21* cp-tiRNAs may be to inhibit CTCF binding to the *p21* gene,



**Figure 3** The effects of transcription initiation (ti)RNA sponges and mimics on *p21* (cyclin-dependent kinase inhibitor 1A gene, also known as *CDKN1A*) expression. (a) Transfection by the CCCTC-binding factor (CTCF) proximal (cp)-tiRNA sponge resulted in an increase in *p21* mRNA and (b) *p21<sup>alt</sup>* transcript levels. Reverse transcription was non-specifically primed and cDNA was analyzed by quantitative (qPCR) using primer set 2 for *p21* mRNA and primer set 1 for *p21<sup>alt</sup>* RNA. (c) Further validation of the results in (a) and (b) were confirmed by transfection with the cp-tiRNA sponge followed by strand-specific quantitative reverse transcription (qRT)-PCR which showed an increase in both antisense and sense *p21<sup>alt</sup>* transcript levels, and (d) an increase in sense *p21* mRNA. (e) Transfection of cp-tiRNA mimics resulted in a decrease in both sense classical and alternate *p21* transcripts. (a-e) Samples were analyzed at 72 h post transfection. The averages of samples transfected in triplicate are shown, error bars represent the standard errors of the means, and *P* values from paired t tests are shown.

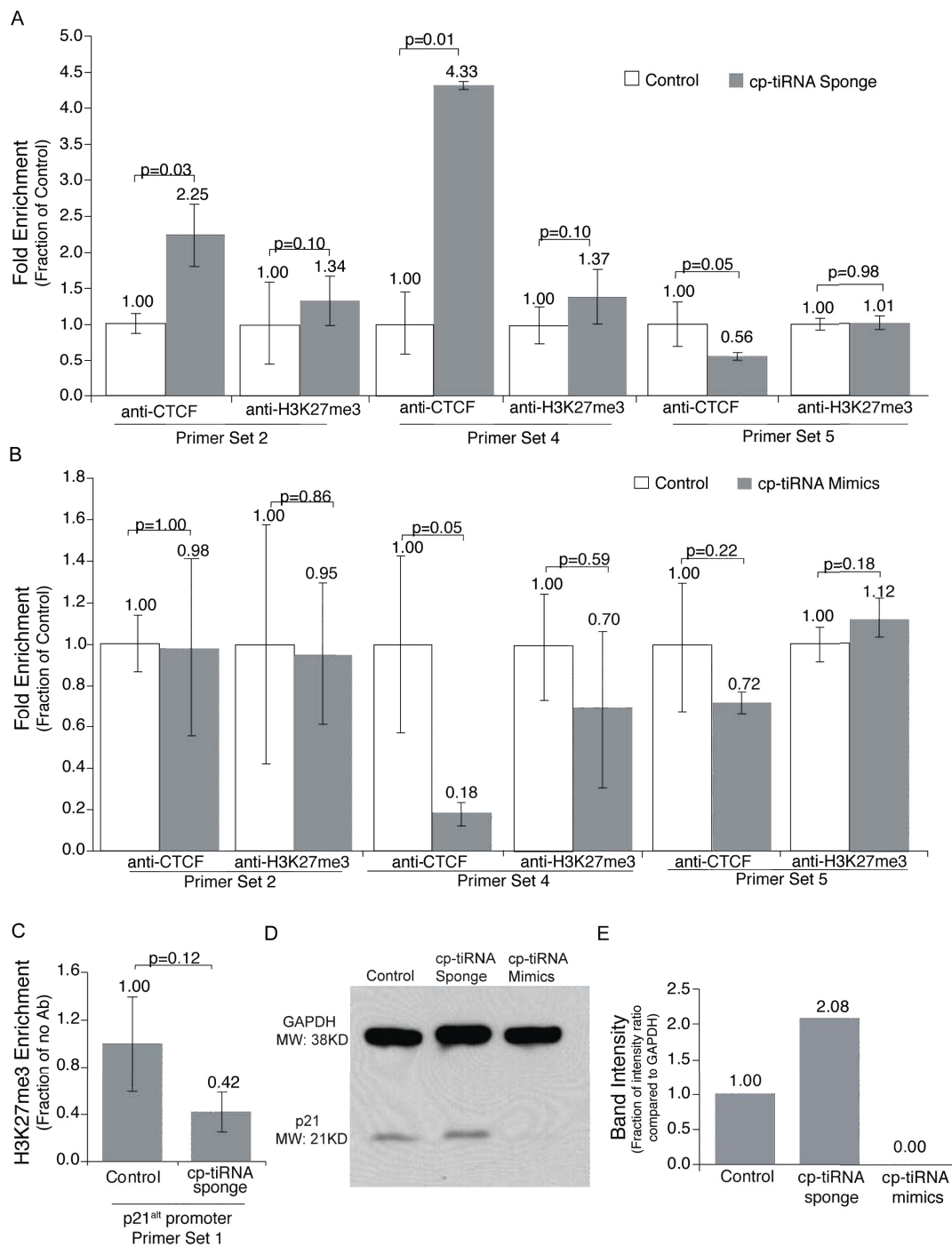


possibly as a means to repress transcription and downstream translation.

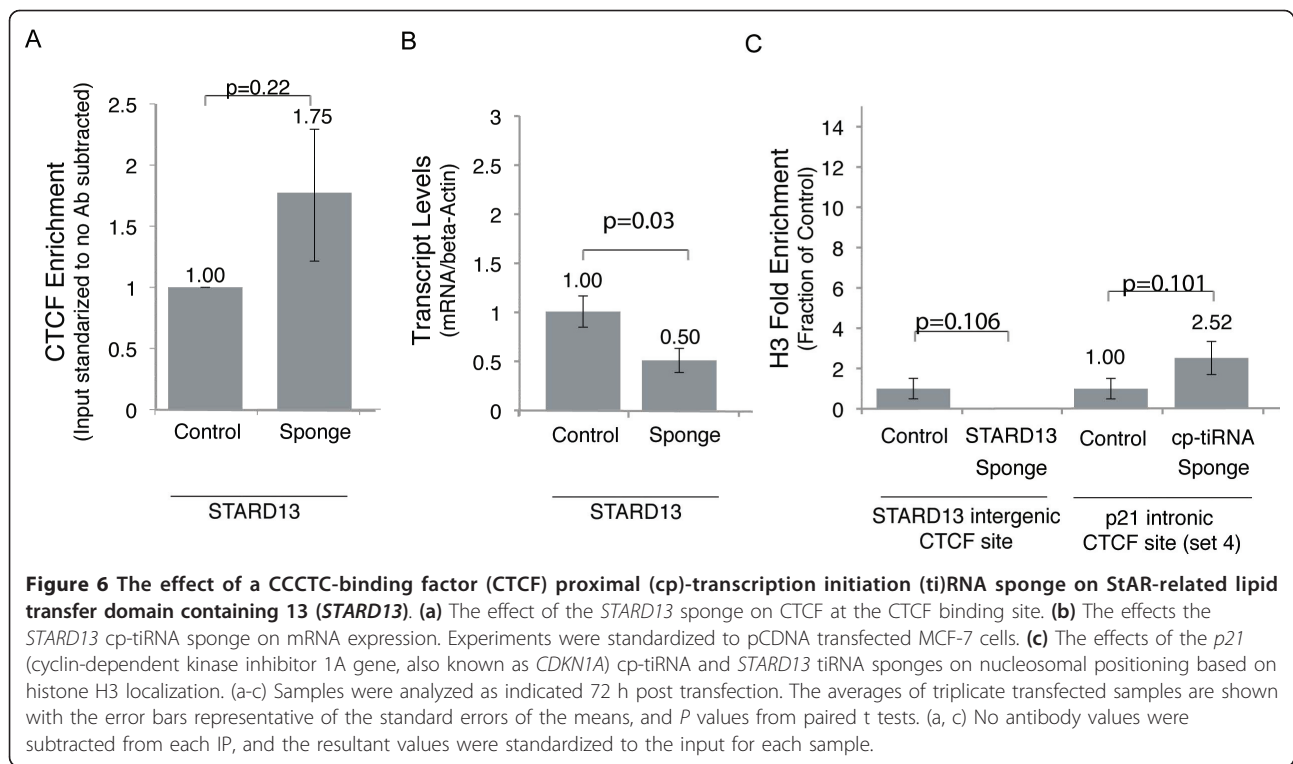
To test whether cp-tiRNAs can modulate CTCF binding at other loci we generated sponges for cp-tiRNAs derived from an intergenic region downstream of the *C2orf42* (*Homo sapiens* chromosome 2 open reading frame 42), and an intergenic site upstream of StAR-related lipid transfer domain containing 13 (*STARD13*) (Additional file 7, Figure S5). To ensure that selection bias did not affect our study, these sites were chosen at random from approximately 900 sites with strong CTCF binding and tiRNA conservation across cell lines (see Methods). Examination of the *C2orf42* site revealed no significant effect of tiRNA sponges (Additional file 8, Figure S6). However, we observed that *STARD13* cp-tiRNA sponges resulted in a reduction in *STARD13* mRNA expression, in spite of the fact that CTCF binding was largely unaffected (Figure 6a, b). This cp-tiRNA-mediated sponge effect is contrary to that observed for *p21*, which strongly increased *p21* expression. To

further investigate this we examined local nucleosome density at both loci and found that the *p21* cp-tiRNA sponges induced increased nucleosomal localization, while the *STARD13* cp-tiRNA induced a decrease in nucleosomal localization (Figure 6c). This is consistent with our hypothesis that cp-tiRNAs mimics and sponges facilitate condition dependent small-scale rearrangements to nucleosome order, and that this in turn leads to large-scale chromatin reorganization orchestrated by CTCF or other DNA binding and chromatin modifying complexes. Indeed, recent work has shown that an array of up to 20 well positioned nucleosomes enriched for the transcription initiation mark H3K4me3 flank CTCF sites, a phenomenon previously only observed downstream of TSSs [39]. This finding not only potentially explains why tiRNAs are frequently found at CTCF sites, but also suggests that the contradictory *p21* and *STARD13* tiRNA sponge effects may result from changes to the local density of chromatin activating marks (Figure 7).





**Figure 5** CCCTC-binding factor (CTCF) binding and H3K27 trimethylation (H3K27me3) at *p21* (cyclin-dependent kinase inhibitor 1A gene, also known as *CDKN1A*) in response to the CTCF proximal (cp)-transcription initiation (ti)RNA sponge and mimics. **(a)** Transfection with the cp-tiRNA sponge resulted in an enrichment of CTCF at the conserved CTCF binding site. H3K27me3 was unaffected at the canonical *p21* promoter or in the *p21*-coding region. **(b)** Transfection with cp-tiRNA mimics resulted in a decrease of CTCF at the CTCF binding site. H3K27me3 was generally unaffected. **(c)** Transfection with the cp-tiRNA sponge resulted in a decrease of H3K27me3 at the alternate *p21* promoter, consistent with an increase in transcriptional activity. **(d)** As measured by western blot, samples transfected with the cp-tiRNA sponge showed an increase of p21 protein levels, while samples transfected with the cp-tiRNA mimics showed a decrease. **(e)** Differences in p21 protein levels from the western blot in (d) quantified using ImageJ. (a-d) Samples were analyzed as indicated 72 h post transfection. No antibody values were subtracted from each immunoprecipitation, and the resultant values were standardized to the input each sample. The averages of samples transfected in triplicate are shown, error bars represent the standard errors of the means, and *P* values from paired t tests are shown.



The mechanism by which tiRNAs inhibit CTCF localization is unclear, although there are several obvious possibilities: (i) cp-tiRNAs spanning the CTCF binding site may coat local chromatin by binding nascent transcripts [36] or chromatin associated RNAs [40-42], which could sterically hinder CTCF from accessing its binding site; (ii) cp-tiRNAs may directly interact with CTCF and inhibit CTCF binding, although attempts to immunoprecipitate CTCF with biotin-linked cp-tiRNAs were unsuccessful (data not shown); (iii) cp-tiRNAs may bind to regulatory elements including cis-acting ncRNAs (for example, *bx332409* at *p21*) or polycomb group components and direct their action to specific sites; or (iv) cp-tiRNAs may serve as sequence-specific markers for chromatin modification complexes.

## Conclusions

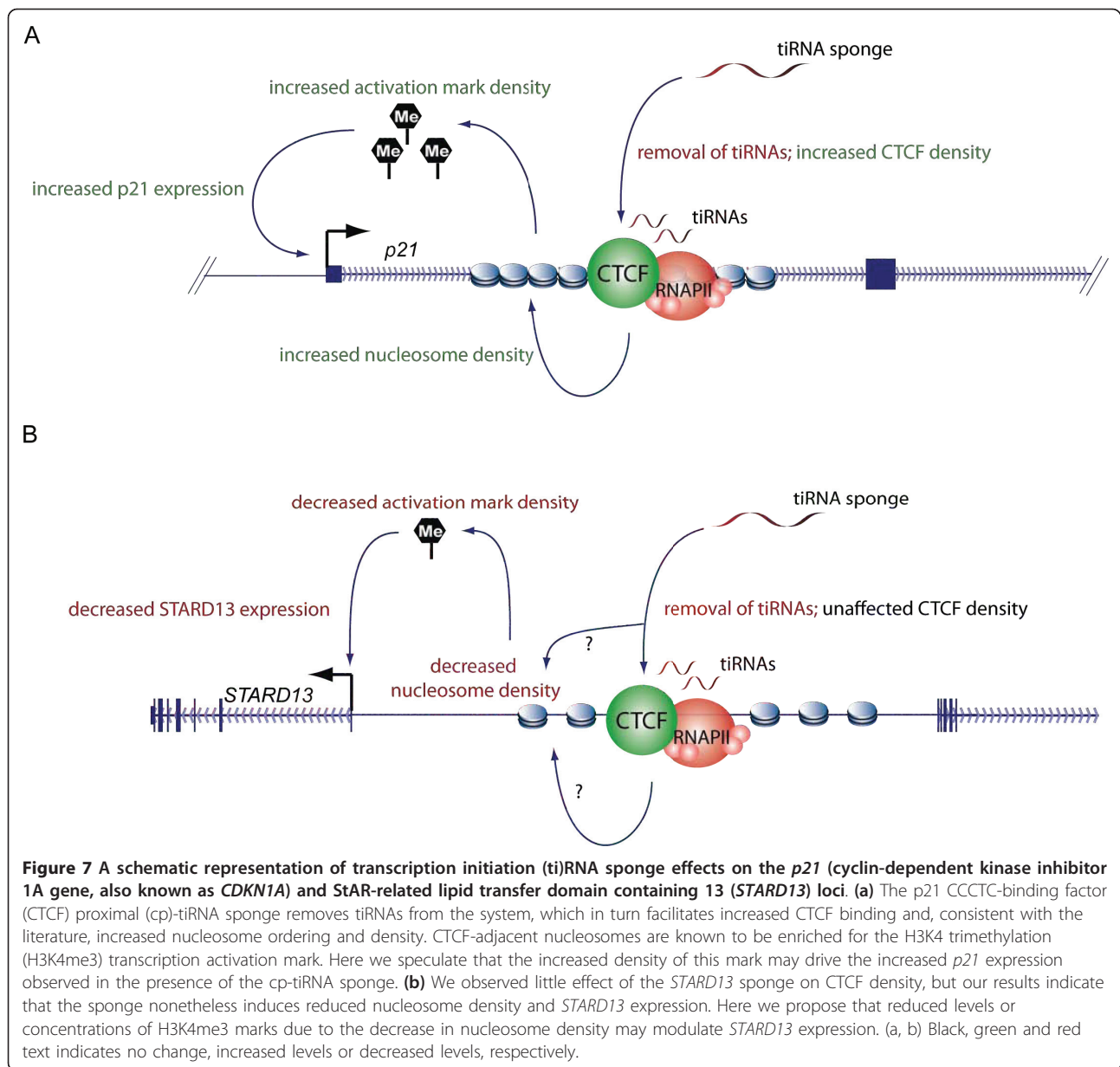
The data presented here indicate that cp-tiRNAs can have a powerful effect on CTCF binding and local transcription. Indeed, tiRNA-mediated modulation of CTCF binding at *p21* not only reduces *p21* mRNA and protein levels, but also appears to affect chromatin state and expression of *p21<sup>alt</sup>* transcripts. This suggests that at some loci the role of tiRNAs, whose biogenesis is connected to RNAPII activity and progression, may be to modulate (presumably indirectly) local chromatin states, which in turn regulates the binding of other factors including CTCF. Indeed, the relationship between

tiRNAs and epigenetic structures may indicate a self-reinforcing feedback loop wherein the RNAPII-nucleosome interaction generates tiRNAs, which in turn serve to mark (directly or indirectly) nucleosome positions and/or epigenetic state. Although the mechanism of tiRNA action is still elusive, this work is the first to report a role for tiRNAs in gene regulation, and shows that at least a subset of tiRNAs are functional modulators of CTCF, which may lead to the development of novel RNA-based therapeutics that target epigenetic regulation of gene structure and transcription.

## Methods

### Bioinformatic analyses

Bioinformatic analyses were performed on a local high-performance computer at the UQ Institute for Molecular Bioscience that houses a mirror of the UCSC Genome Browser [43]. We used a suite of in-house AWK, C, Perl, and Python scripts and UCSC backend tools. All small RNA, CTCF binding and RNAPII binding data were obtained from publicly available sources and are listed in detail in Additional file 1, Table S1. For all ChIP-seq datasets the available peak calls were used, except in the case of the mES RNAPII data where peaks were defined as regions with signal greater than 3 SDs from the mean. All intersections were performed using a modified version of the UCSC backend tools *bedIntersect* or *overlapSelect*. A minimum of 1 bp of overlap



was required, but generally >50% of any given feature intersected with another. The relative enrichment of small RNAs at CTCF sites with or without RNAPII coverage was computed using an in-house (Perl) bootstrapping program over 1,000 iterations as previously described [9]. Bootstrapping was constrained such that the randomized placements of small RNAs excluded TSSs, known small RNA annotations, repeat masker annotations and genome assembly gaps. For the analysis of the MCF-7, mES, and human CTCF data grouped into 'cancer' and 'normal' the data was parsed so that data points that intersected with known small RNA annotations, the 500 bp adjacent to TSSs, repeat masker annotations or Ensembl annotations less than 300

nucleotides were removed. The conserved CTCF site data set were generated by taking all the peaks identified by Broad/ENCODE as significant across all eight sources (see Additional file 1, Table S1) and intersecting them against one another using the UCSC backend tool overlapSelect. Overlapping CTCF features (called peaks) were collapsed into one concordant set of coordinates using the UCSC tool featureBits. Small RNA size distributions were computed as previously described [9].

#### Cell culture and transfection

Experiments were conducted in MCF-7 and THP-1 cells cultured at 37°C and 5% CO<sub>2</sub> in Dulbecco's modified Eagle's medium or RPMI 1640 (THP-1 cells only)

supplemented with 10% fetal bovine serum and penicillin-streptomycin (Life Technologies, Carlsbad, California, USA). Plasmid transfections of MCF-7 cells were performed using Lipofectamine 2000 (Life Technologies, Carlsbad, California, USA) per the manufacturer's guidelines at a concentration of  $1 \mu\text{g}/1 \times 10^6$  cells. Plasmid transfections of THP-1 cells were carried out using the Neon electroporation system (Life Technologies, Carlsbad, CA, USA), per manufacturer's guidelines.

#### Sponge and mimic plasmid construction

Sponge plasmids: reverse complement sequences to the promoter-proximal and CTCF-proximal tiRNA sites in p21 were amplified from MCF-7 genomic DNA by polymerase chain reaction (PCR) using primers with *Bgl*III restriction sites on either end (Additional file 3, Table S2). Amplified DNA was digested by *Bgl*III and ligated into the similarly cut pCDNA3 U6M2 plasmid. The parent vector, pCDNA3 U6M2, was used as a control in sponge experiments. The resulting sponge sequences were: p21 promoter-proximal tiRNA sponge (pp-tiRNA) 5'CCGCCGGCCCCGGGGTCCCCTGTTGTCTGCCGCGCTCTCTCACCTCCTCTGAGTGCCTCGGT GCC TCGGCGAATCCGCGCCCAGCTCCGGCTCCACAA GGAAGTACTTCGGCAGCTGCTCACA CCTCAG CTGG3' and p21 CTCF-proximal tiRNA sponge (cp-tiRNA) 5'GGGCTCAGGGCTTCCCGACCTAGCCCA-GATTCCCTCTCCGAAAGCTACAGGGCTGAGCGGA GCAGGGGGGCGAGTCGCCCCCTGGGGCGCCGCC GCCTGGCGCGGACCACAGCGCGTCCCTCT CCGTCCAAACCCCTGGGGGACACTTGCGCCCTTTCGTGAGGAAAAGCATCTTGGAGCTGG GTTAGGAACT TGGGGCGCCCAGGCAGCTTCCCCTCTCCTTGC CTCCTCCACGTCGCGTTTCT GGGAGGACTTGC-GAGCGGTTTTGTTTTCGTTGCTCCCGTCTATTTT-TATTTTCCAGGGATCTGA CT3'.

Sponge constructs for C2Orf41 and STARD13 were constructed by annealing the respective oligonucleotides together (Additional file 3, Table S2), top strand oligo (200  $\mu\text{M}$ ): 5  $\mu\text{l}$  bottom strand oligo (200  $\mu\text{M}$ ): 5  $\mu\text{l}$ , 10  $\times$  Oligo Annealing Buffer (BLOCK-iT kit\*, Life Technologies, Carlsbad, California, USA): 2  $\mu\text{l}$ , DNase/RNase-free water: 8  $\mu\text{l}$ . The reaction was heated to 95°C for 4 min, cooled to room temperature and treated to polynucleotide kinase (PNK) (1  $\mu\text{l}$  PNK (10,000 U/ml), 1  $\mu\text{l}$  T4 ligase buffer, 4  $\mu\text{l}$  dsOligo, 4  $\mu\text{l}$  water, at 37°C for 30 min). The oligonucleotides were purified using minelute PCR purification kit and ligated to pCDNA3 U6M2 as described above.

#### Mimic plasmids

A negative control and four sequences of CTCF-proximal tiRNAs were synthesized (Integrated DNA Technologies) and cloned into the BLOCK-iT U6 RNAi Entry

Vector (Life Technologies, Carlsbad, CA, USA) per manufacturer's guidelines. The mimics, four predominant cp-tiRNA sequences (Additional file 9, Table S3) were independently cloned into the U6 driven BLOCK-iT system (Life Technologies, Carlsbad, CA, USA). The resultant plasmids were transfected into MCF-7 cells as described previously [36].

#### Quantitative strand-specific PCR (qPCR)

RNA was extracted (RNeasy Qiacube, Qiagen), DNase treated (TURBO DNase, Ambion), reverse transcribed (Reverse Transcriptase Core Kit, Eurogentec) using the non-specific or indicated primers (for strand-specific RT-PCR), and analyzed by qPCR using indicated primers (Kapa Sybr Fast Universal qPCR Kit, Kapa Biosystems, Woburn, MA, USA) (Additional file 9, Table S3). In strand-specific RT-PCR, reverse transcription is primed with a gene specific forward or reverse primer alone, thereby generating cDNA of specifically the antisense or sense strand of the targeted region respectively. Controls for this assay are reverse transcription or template RNA in the absence of any primer. Quantitative PCR (qPCR) is then performed using forward and reverse primers, yielding amplicons that represent sense or antisense transcripts overlapping that region with the control no primer RT sample values subtracted as background from the directional RT primed samples.

#### ChIP

ChIP assays were performed as previously described [44]. DNA was immunoprecipitated using anti-RNAPII phosphor-S2 (AbCam no. ab5095, AbCam Cambridge, MA USA), anti-H3K27me3 (Cell Signaling no. 9756S, Danvers, MA, USA), or anti-CTCF (Santa Cruz Biotechnology no. sc-15914, Santa Cruz, CA, USA) antibody bound complexes were then pulled down using magnetic Dynabeads Protein A (Life Technologies, Carlsbad, CA, USA). DNA was then recovered by phenol/chloroform extraction and analyzed by qPCR using indicated primers (Additional file 9, Table S3) (Kapa Sybr Fast Universal qPCR Kit, Kapa Biosystems, Woburn, MA, USA).

#### Western blot

Cells were lysed in modified RIPA buffer (25 mM Tris HCl, pH 7.5, 15 mM NaCl, 1% Nonidet P-40, 1% NaD, and 0.1% SDS) and separated on a NuPAGE 4% to 12% BisTris gel (Life Technologies, Carlsbad, CA, USA). Proteins were transferred to a nitrocellulose membrane which was blocked with 5% milk for 1 h and then incubated overnight at 4°C with anti-p21 (Cell Signaling no. 2946) and anti-glyceraldehyde 3-phosphate dehydrogenase (GAPDH) (Millipore no. MAB374, Billerica, MA, USA) antibodies. The membrane was then washed (10

mM Tris HCl, pH 7.5, 50 mM NaCl, 0.075% Tween 20) and incubated with an anti-mouse horseradish peroxidase-conjugated secondary antibody for 1 h at room temperature (Upstate no. 12-349, Billerica, MA, USA). The membrane was then washed, treated with chemiluminescent detection reagent (HyGLO, Denville Metuchen, NJ, USA), and exposed to film. Blot density of a binary image of Figure 5d was calculated using ImageJ. Results were standardized to GAPDH and expressed as fractions of control values.

## Additional material

**Additional file 1: Table S1.** Small RNA (sRNA), CCCTC-binding factor (CTCF) chromatin immunoprecipitation (ChIP)-seq and RNA polymerase II (RNAPII) ChIP-seq datasets. <sup>†</sup>Data obtained from ENCODE/UCSC and can be retrieved from the publicly available ENCODE UCSC genome browser (<http://genome.ucsc.edu/ENCODE/>) [5,9]. <sup>\*</sup>All data sets with identifiers beginning with GSE can be found at that the NCBI Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>). MCF-7 CTCF ChIP-seq data can be found in the EBI Array Express Archive (<http://www.ebi.ac.uk/arrayexpress/>).

**Additional file 2: Figure S1.** Enrichment of transcription initiation (ti) RNAs at sites of conserved CCCTC-binding factor (CTCF) binding genome-wide. CTCF sites were obtained from the ENCODE Broad Institute UCSC Histone Mods tracks. Only those with significant peaks conserved across GM12878, HepG2, HMEC, HSMM, HUVEC, K562, NHEK and NHLF cell lines were considered in this analysis. **(a)** Enrichment of small RNAs at all conserved CTCF sites (gray), and conserved CTCF sites with evidence of conserved RNA polymerase II (RNAPII) binding in HUVEC, K562 and NHEK cells (black). **(b-d)** The size distribution of small RNAs found at conserved CTCF sites with evidence of RNAPII binding in (b) THP-1, (c) 5-8f and (d) MCF cells. In (b) and (c) nuclear small RNAs are shown in black, and cytoplasmic small RNAs are depicted in gray.

**Additional file 3: Table S2.** Conserved CCCTC-binding factor (CTCF) and CTCF-RNA polymerase II (RNAPII) site intersection with small RNAs (sRNAs).

**Additional file 4: Figure S2.** Enrichment of transcription initiation (ti) RNAs at sites of subsets of conserved CCCTC-binding factor (CTCF) binding sites genome-wide. CTCF sites were parsed to (i) exclude any that mapped within 500 bp of a TSSs or overlapped repeat masker, small RNA or Ensembl gene annotations less than 300 nucleotides, (ii) into groups by cell type where 'cancer' was derived from MCF-7, K562 and HepG2 data and normal was derived from GM12878, HUVEC, HMEC, HSMM, NHEK and NHLF data, and (iii) intersected with the most robust RNA polymerase II (RNAPII) data for each group (MCF-7 and HUVEC, for cancer and normal, respectively). Note that tiRNAs are highly enriched at CTCF-RNAPII sites even in these highly reduced sets.

**Additional file 5: Figure S3.** The intronic *p21* (cyclin-dependent kinase inhibitor 1A gene, also known as *CDKN1A*) CCCTC-binding factor-RNA polymerase II (CTCF-RNAPII) site is highly conserved in mammals. In the top panel the *p21/CDKN1A* locus in mouse is shown. The bottom panel is a focused view of the intronic CTCF-RNAPII site. Note the conserved CTCF site, RNAPII binding, and high conservation of the site itself and the sequence immediately to the left (5' with respect to *p21*), which is the site of transcription initiation (ti)RNA biogenesis in humans.

**Additional file 6: Figure S4.** *p21* (cyclin-dependent kinase inhibitor 1A gene, also known as *CDKN1A*) CCCTC-binding factor (CTCF) proximal (cp)-transcription initiation (ti)RNA RNA sponge and mimics effects are conserved in THP-1 cells. Samples were prepared and analyzed identical to those shown in Figure 3. Note the increase in *p21* mRNA levels in response to the cp-tiRNA sponge, and the decrease in expression in response to cp-tiRNA mimics.

**Additional file 7: Figure S5.** A schematic of the CCCTC-binding factor (CTCF) sites proximal to *C2orf42* and StAR-related lipid transfer domain

containing 13 (*STARD13*). **(a, b)** Gene models are shown at the top of each panel, followed by the collapsed density of small RNAs in three datasets, CTCF binding density in nine human cell types, and RNA polymerase II (RNAPII) binding in three human cell lines. In (a) the conserved *C2orf42* CTCF site is approximately 5 kb downstream of the 3' untranslated region (UTR). The CTCF site in (b) sits approximately halfway between *STARD13* and *RFC3*.

**Additional file 8: Figure S6.** The effect of the CCCTC-binding factor (CTCF) proximal (cp)-transcription initiation (ti)RNA RNA sponge on CTCF sites proximal to *C2orf42*. **(a)** The effects of the *C2orf42* sponge on CTCF localization. **(b)** The effects of *C2orf41* tiRNA sponges on mRNA expression. Experiments were standardized to pCDNA transfected MCF-7 cells. **(c)** The effects of the *C2orf41* sponge on histone H3 localization. Samples were analyzed as indicated 72 h post transfection. The averages of triplicate transfected samples are shown with the error bars representative of the standard errors of the means, and *P* values from paired *t* tests. (a, c) No antibody values were subtracted from each IP, and the resultant values were standardized to the input for each sample.

**Additional file 9: Table S3.** Oligonucleotides used for cloning, quantitative reverse transcription (qRT)-PCR, and chromatin immunoprecipitation (ChIP) analysis.

## Acknowledgements

This project was supported by NIH R01 grants HL083473-02 and AI084406-01A1 (KVM) a University of Queensland Postdoctoral Research Fellowship (RJT) and ARC Discovery Project Grant DP DP0988851 and ARC Federation Fellowship FF0561986 (JSM).

## Author details

<sup>1</sup>Institute for Molecular Bioscience, University of Queensland, St Lucia, Queensland 4072, Australia. <sup>2</sup>Department of Molecular and Experimental Medicine, The Kellogg School of Science and Technology, The Scripps Research Institute, La Jolla, CA 92037, USA. <sup>3</sup>The Kellogg School of Science and Technology, The Scripps Research Institute, La Jolla, CA 92037, USA.

## Authors' contributions

RJT performed the bioinformatic analyses, helped design the laboratory experiments and wrote the manuscript. PGH performed the generated and tested tiRNA mimics and sponges and helped write the manuscript. JSM assisted in study design and direction. KVM designed the laboratory experiments, performed tiRNA sponge and mimic experiments and wrote the manuscript. All authors read and approved the final manuscript.

## Competing interests

RJT and JSM have an ownership stake in a patent concerning tiRNAs and their diagnostic and therapeutic uses (International Patent No. AU2009/000423).

Received: 1 April 2011 Accepted: 3 August 2011

Published: 3 August 2011

## References

1. Kapranov P, Willingham AT, Gingeras TR: **Genome-wide transcription and the implications for genomic organization.** *Nat Rev Genet* 2007, **8**:413-423.
2. Mattick JS, Taft RJ, Faulkner GJ: **A global view of genomic information-moving beyond the gene and the master regulator.** *Trends Genet* 2010, **26**:21-28.
3. Mercer TR, Dinger ME, Mattick JS: **Long non-coding RNAs: insights into functions.** *Nat Rev Genet* 2009, **10**:155-159.
4. Taft RJ, Kaplan CD, Simons C, Mattick JS: **Evolution, biogenesis and function of promoter-associated RNAs.** *Cell Cycle* 2009, **8**:2332-2338.
5. Affymetrix/Cold Spring Harbor Laboratory ENCODE Transcriptome Project: **Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs.** *Nature* 2009, **457**:1028-1032.
6. Kanhere A, Viiri K, Araújo CC, Rasaiyaah J, Bowman RD, Whyte WA, Pereira CF, Brookes E, Walker K, Bell GW, Pombo A, Fisher AG, Young RA, Jenner RG: **Short RNAs are transcribed from repressed polycomb target**

- genes and interact with polycomb repressive complex-2. *Mol Cell* 2010, **38**:675-688.
7. Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, Song JJ, Kingston RE, Borowsky M, Lee JT: **Genome-wide identification of polycomb-associated RNAs by RIP-seq.** *Mol Cell* 2010, **40**:939-953.
  8. Taft RJ, Glazov EA, Cloonan N, Simons C, Stephen S, Faulkner GJ, Lassmann T, Forrest AR, Grimmond SM, Schroder K, Irvine K, Arakawa T, Nakamura M, Kubosaki A, Hayashida K, Kawazu C, Murata M, Nishiyori H, Fukuda S, Kawai J, Daub CO, Hume DA, Suzuki H, Orlando V, Carninci P, Hayashizaki Y, Mattick JS: **Tiny RNAs associated with transcription start sites in animals.** *Nat Genet* 2009, **41**:572-578.
  9. Taft RJ, Simons C, Nahkuri S, Oey H, Korbie DJ, Mercer TR, Holst J, Ritchie W, Wong JJ, Rasko JE, Rokhsar DS, Degnan BM, Mattick JS: **Nuclear-localized tiny RNAs are associated with transcription initiation and splice sites in metazoans.** *Nat Struct Mol Biol* 2010, **17**:1030-1034.
  10. Phillips JE, Corces VG: **CTCF: master weaver of the genome.** *Cell* 2009, **137**:1194-1211.
  11. Filippova GN: **Genetics and epigenetics of the multifunctional protein CTCF.** *Curr Top Dev Biol* 2008, **80**:337-360.
  12. Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenkov VV, Ren B: **Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome.** *Cell* 2007, **128**:1231-1245.
  13. Wallace JA, Felsenfeld G: **We gather together: insulators and genome organization.** *Curr Opin Genet Dev* 2007, **17**:400-407.
  14. Ohlsson R, Renkawitz R, Lobanenkov V: **CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease.** *Trends Genet* 2001, **17**:520-527.
  15. Qi CF, Martensson A, Mattioli M, Dalla-Favera R, Lobanenkov VV, Morse HC: **CTCF functions as a critical regulator of cell-cycle arrest and death after ligation of the B cell receptor on immature B cells.** *Proc Natl Acad Sci USA* 2003, **100**:633-638.
  16. Witcher M, Emerson BM: **Epigenetic silencing of the p16(INK4a) tumor suppressor is associated with loss of CTCF binding and a chromatin boundary.** *Mol Cell* 2009, **34**:271-284.
  17. De Biase I, Chutake YK, Rindler PM, Bidichandani SI: **Epigenetic silencing in Friedreich ataxia is associated with depletion of CTCF (CCCTC-binding factor) and antisense transcription.** *PLoS One* 2009, **4**:e7914.
  18. Zentner GE, Saiakhova A, Manaenkov P, Adams MD, Scacheri PC: **Integrative genomic analysis of human ribosomal DNA.** *Nucleic Acids Res* 2011, **39**:4949-4960.
  19. van de Nobelen S, Rosa-Garrido M, Leers J, Heath H, Souchit W, Joosen L, Jonkers I, Demmers J, van der Reijden M, Torrano V, Grosveld F, Delgado MD, Renkawitz R, Galjart N, Sleutels F: **CTCF regulates the local epigenetic state of ribosomal DNA repeats.** *Epigenetics Chromatin* 2010, **3**:19.
  20. Guerrero PA, Maggert KA: **The CCCTC-binding factor (CTCF) of *Drosophila* contributes to the regulation of the ribosomal DNA and nucleolar stability.** *PLoS One* 2011, **6**:e16401.
  21. Klenova EM, Chernukhin IV, El-Kady A, Lee RE, Pugacheva EM, Loukinov DI, Goodwin GH, Delgado D, Filippova GN, León J, Morse HC, Neiman PE, Lobanenkov VV: **Functional phosphorylation sites in the C-terminal region of the multivalent multifunctional transcriptional factor CTCF.** *Mol Cell Biol* 2001, **21**:2221-2234.
  22. Chernukhin I, Shamsuddin S, Kang SY, Bergström R, Kwon YW, Yu W, Whitehead J, Mukhopadhyay R, Docquier F, Farrar D, Morrison I, Vigneron M, Wu SY, Chiang CM, Loukinov D, Lobanenkov V, Ohlsson R, Klenova E: **CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide.** *Mol Cell Biol* 2007, **27**:1631-1648.
  23. ENCODE Project Consortium, Myers RM, Stamatoyannopoulos J, Snyder M, Dunham I, Hardison RC, Bernstein BE, Gingeras TR, Kent WJ, Birney E, Wold B, Crawford GE: **A user's guide to the encyclopedia of DNA elements (ENCODE).** *PLoS Biol* 2011, **9**:e1001046.
  24. Rosenbloom KR, Dreszer TR, Pheasant M, Barber GP, Meyer LR, Pohl A, Raney BJ, Wang T, Hinrichs AS, Zweig AS, Fujita PA, Learned K, Rhead B, Smith KE, Kuhn RM, Karolchik D, Haussler D, Kent WJ: **ENCODE whole-genome data in the UCSC Genome Browser.** *Nucleic Acids Res* 2010, **38**:D620-625.
  25. Liao JY, Ma LM, Guo YH, Zhang YC, Zhou H, Shao P, Chen YQ, Qu LH: **Deep sequencing of human nuclear and cytoplasmic small RNAs reveals an unexpectedly complex subcellular distribution of miRNAs and tRNA 3' trailers.** *PLoS One* 2010, **5**:e10563.
  26. el-Deiry WS, Tokino T, Velculescu VE, Levy DB, Parsons R, Trent JM, Lin D, Mercer WE, Kinzler KW, Vogelstein B: **WAF1, a potential mediator of p53 tumor suppression.** *Cell* 1993, **75**:817-825.
  27. Gu Y, Turck CW, Morgan DO: **Inhibition of CDK2 activity *in vivo* by an associated 20K regulatory subunit.** *Nature* 1993, **366**:707-710.
  28. Harper JW, Adami GR, Wei N, Keyomarsi K, Elledge SJ: **The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases.** *Cell* 1993, **75**:805-816.
  29. Xiong Y, Hannon GJ, Zhang H, Casso D, Kobayashi R, Beach D: **p21 is a universal inhibitor of cyclin kinases.** *Nature* 1993, **366**:701-704.
  30. Pellikainen MJ, Pekola TT, Ropponen KM, Kataja VV, Kellokoski JK, Eskelinen MJ, Kosma VM: **p21WAF1 expression in invasive breast cancer and its association with p53, AP-2, cell proliferation, and prognosis.** *J Clin Pathol* 2003, **56**:214-220.
  31. Pinto AE, Andre S, Laranjeira C, Soares J: **Correlations of cell cycle regulators (p53, p21, pRb and mdm2) and c-erbB-2 with biological markers of proliferation and overall survival in breast cancer.** *Pathology* 2005, **37**:45-50.
  32. Shiohara M, el-Deiry WS, Wada M, Nakamaki T, Takeuchi S, Yang R, Chen DL, Vogelstein B, Koefler HP: **Absence of WAF1 mutations in a variety of human malignancies.** *Blood* 1994, **84**:3781-3784.
  33. Tiezzi DG, Andrade JM, Ribeiro-Silva A, Zola FE, Marana HR, Tiezzi MG: **HER-2, p53, p21 and hormonal receptors proteins expression as predictive factors of response and prognosis in locally advanced breast cancer treated with neoadjuvant docetaxel plus epirubicin combination.** *BMC Cancer* 2007, **7**:36.
  34. Xiong Y, Zhang H, Beach D: **Subunit rearrangement of the cyclin-dependent kinases is associated with cellular transformation.** *Genes Dev* 1993, **7**:1572-1583.
  35. Radhakrishnan SK, Gierut J, Gartel AL: **Multiple alternate p21 transcripts are regulated by p53 in human cells.** *Oncogene* 2006, **25**:1812-1815.
  36. Morris KV, Santoso S, Turner AM, Pastorci C, Hawkins PG: **Bidirectional transcription directs both transcriptional gene activation and suppression in human cells.** *PLoS Genet* 2008, **4**:e1000258.
  37. Ebert MS, Neilson JR, Sharp PA: **MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells.** *Nat Methods* 2007, **4**:721-726.
  38. Malecova B, Morris KV: **Transcriptional gene silencing through epigenetic changes mediated by non-coding RNAs.** *Curr Opin Mol Ther* 2010, **12**:214-222.
  39. Fu Y, Sinha M, Peterson CL, Weng Z: **The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome.** *PLoS Genet* 2008, **4**:e1000138.
  40. Mondal T, Rasmussen M, Pandey GK, Isaksson A, Kanduri C: **Characterization of the RNA content of chromatin.** *Genome Res* 2010, **20**:899-907.
  41. Rodriguez-Campos A, Azorin F: **RNA is an integral component of chromatin that contributes to its structural organization.** *PLoS One* 2007, **2**:e1182.
  42. Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY: **Long noncoding RNA as modular scaffold of histone modification complexes.** *Science* 2010, **329**:689-693.
  43. Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, Goldman M, Barber GP, Clawson H, Coelho A, Diekhans M, Dreszer TR, Gardine BM, Harte RA, Hillman-Jackson J, Hsu F, Kirkup V, Kuhn RM, Learned K, Li CH, Meyer LR, Pohl A, Raney BJ, Rosenbloom KR, Smith KE, Haussler D, Kent WJ: **The UCSC Genome Browser database: update 2011.** *Nucleic Acids Res* 2011, **39**:D876-882.
  44. Han J, Kim D, Morris KV: **Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells.** *Proc Natl Acad Sci USA* 2007, **104**:12422-12427.

doi:10.1186/1756-8935-4-13

**Cite this article as:** Taft et al.: The relationship between transcription initiation RNAs and CCCTC-binding factor (CTCF) localization. *Epigenetics & Chromatin* 2011 **4**:13.