CTCF blocks anti-sense transcription initiation at divergent gene promoters

Jing Luan[1], Camille M. Syrett[2¶], Marit W. Vermunt[2], Allison Coté[3,4], Jacob M. Tome[5§], Haoyue Zhang[2†], Anran Huang[2], Jennifer M. Luppino[4], Cheryl A. Keller[6], Belinda M. Giardine[6], Shiping Zhang[7], Margaret C. Dunagin[3], Zhe Zhang[7], Eric F. Joyce[4], John T. Lis[5], Arjun Raj[3,4], Ross C. Hardison[6], Gerd A. Blobel[2,8*]

[1]Medical Scientist Training Program, Perelman School of Medicine, University of Pennsylvania, Philadelphia, USA

[2]Division of Hematology, The Children's Hospital of Philadelphia, Philadelphia, USA

[3]Department of Bioengineering, University of Pennsylvania, Philadelphia, USA

[4]Department of Genetics, University of Pennsylvania, Philadelphia, USA

[5]Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA

[6]Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA

[7]Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, USA

[8]Lead contact

¶ present affiliation: Clarion Healthcare, LLC, Boston, MA, USA

§ present affiliation: Shape Therapeutics Inc, Seattle, WA, USA

† present affiliation: Institute of Molecular Physiology, Shenzhen Bay Laboratory, Shenzhen, Guangdong, China

*Correspondence: BLOBEL@chop.edu

Abstract

Transcription at most promoters is divergent, initiating at closely spaced oppositely oriented core promoters to produce sense transcripts along with often unstable upstream antisense (uasTrx). How antisense transcription is regulated and to what extent it is coordinated with sense transcription is largely unknown. Here by combining acute degradation of the multi-functional transcription factor CTCF and nascent transcription measurements, we find that CTCF specifically suppresses antisense but not sense transcription at hundreds of divergent promoters, the great majority of which bear proximal CTCF binding sites. Genome editing, chromatin conformation studies, and high-resolution transcript mapping revealed that precisely positioned CTCF directly suppresses the initiation of uasTrx, in a manner independent of its chromatin architectural function. Primary transcript RNA FISH revealed co-bursting of sense and anti-sense transcripts is disfavored, suggesting CTCF-regulated competition for transcription initiation. In sum, CTCF shapes the transcriptional landscape in part by suppressing upstream antisense transcription.

Main Text

Divergent transcription at active promoters is prevalent among eukaryotes, producing upstream antisense transcripts (uasTrx) that are rapidly processed and tend to be short lived[1-3]. Divergent promoters are nucleosome-depleted region densely occupied by transcription factors. They typically harbor two distinct core promoters positioned in inverted orientations, instructing the assembly of separate transcription pre-initiation complexes (PICs) that transcribe along opposite DNA strands[4-7]. Transcriptional outputs by divergent promoters in both orientations are generally concordant, suggesting co-regulation[1,2,8-10]. In certain cases, however, sense and antisense transcription appears to be anti-correlated[11]. It thus remains unclear whether and how divergent transcription is coordinated spatially and temporally. On one hand, divergent transcription may be cooperative, as simultaneous presence of two PICs may help maintain nucleosome-depleted regions and allow for efficient transcription factor recruitment[5,12]. On the other, divergent PICs may compete for common transcription activators or physical space, thus rendering co-occurrence unfavorable[13].

CTCF (CCCTC-binding factor) was first identified as a transcription factor and was later recognized to also shape genome topology together with the cohesin protein complex[14]. CTCF depletion is known to cause genome-wide architectural perturbation but limited changes in the transcription of coding genes[15-23]. However, the mammalian genome is ubiquitously expressed, producing abundant noncoding transcripts that have now gained increasing recognition as functional[24]. Whether and how CTCF affects the noncoding transcriptome remains to be explored experimentally.

We performed precision nuclear run-on sequencing (PRO-seq) in the mouse murine erythroid cell line G1E-ER4 in which both *CTCF* alleles have been modified to bear an auxin-inducible degron (AID) that allows for rapid CTCF degradation[23]. PRO-seq interrogates nascent transcription in a strand-specific manner at single base-pair resolution[25]. Overall, we observed limited perturbation of annotated transcripts after acute CTCF depletion[23]. Notably, however, at 376 active promoters we observed a significant increase in uasTrx production (Fig. 1a-c and Supplementary Data 1). These changes were corroborated by ChIP-seq (chromatin immunoprecipitation sequencing) of RNA polymerase II subunit A (POLR2A) and RT-qPCR at 3 select loci (Extended Data Fig. 1a,b). UasTrx were heterogenous in size, with the median being

1956 nucleotides (Extended Data Fig. 1c). The most 5' ends of these transcripts initiated upstream of sense transcription start sites (TSSs). The average distance of the most frequently used uasTrx start sites from sense TSSs was approximately 100 bp (Fig. 1d), which is similar to that of divergent promoters found in other mammalian cells where this distance was ~110 bp[5,13]. CTCF depletion led to increases only in the antisense direction leaving sense transcription ostensibly unperturbed, suggesting that CTCF promotes the directionality of divergent promoters by exerting strand-specific transcription repression (Fig. 1d and Extended Data Fig. 1b,d,e,f).

Different terms have been used to describe antisense transcription from divergent promoters in eukaryotes, including cryptic unannotated transcripts (CUTs), stable unannotated transcripts (SUTs), and Xrn1-sensitive unstable transcripts (XUTs) in yeast, as well as PROMoter uPstream Transcripts (PROMPTs) and "upstream divergent transcripts" in higher eukaryotes[1,3,9,26,27]. The uasTrx that we found to be repressed by CTCF may represent a subset of PROMPs/upstream divergent transcripts.

Promoters with up-regulated uasTrx are enriched with proximal (mostly <100bp from annotated TSSs) CTCF binding, reminiscent of an earlier finding observed across multiple human cell lines (Fig. 1e and Extended Data Fig. 2a)[28]. Notably, only a fraction of CTCF-bound promoters (319 out of 1,846) increased uasTrx production in response to CTCF loss, but those tended to have a stronger CTCF binding intensity (Extended Data Fig. 2b). However, CTCF binding reduction and uasTrx gains were only weakly correlated (Extended Data Fig. 2c). Because strong CBSs tend to be conserved across cell types[15,23], we assessed CTCF occupancy across mouse tissues[29]. Indeed, CBSs at uasTrx regulatory sites were more tissue-invariant, indicating that uasTrx repression may be a conserved feature (Extended Data Fig. 2d). To assess whether CTCF functions in a similar way in other species and tissues, we measured uasTrx changes upon CTCF depletion in the human colorectal carcinoma cell line HCT-116 and again found 199 uasTrx to be up-regulated without significantly affecting sense transcription (Extended Data Fig. 3a). We also examined published data sets in mouse embryonic stem cells (mESCs) and observed a similar number of up-regulated uasTrx (Extended Data Fig. 4a,b)[30]. Antisense changes in both cell types were similarly associated with strong promoter-proximal CTCF binding (Extended Data Fig. 3b and Extended Data Fig. 4c) and a lack of sense perturbation (Extended Data Fig. 3c-f and Extended Data Fig. 4d-f). Lastly, up-regulated uasTrx in mESCs were silenced upon CTCF recovery following auxin removal (Extended Data Fig.

4b,g). Hence, CTCF represses uasTrx at numerous genes across multiple species and cell lineages.

Because promoter-proximal CTCF only suppresses a subset of the uasTrx, we examined features that determine uasTrx regulation by CTCF. In addition to being enriched with strong CBSs, promoters with up-regulated uasTrx harbored high levels of cohesin, a protein complex central to genome folding[31,32], compared to those that were unchanged upon CTCF depletion (Extended Data Fig. 5a). In addition, they are enriched at chromatin loop anchors and chromatin domain boundaries (Fig. 1f,g and Extended Data Fig. 5b,c). The associated sense transcripts also tend to be housekeeping genes, which are frequently found at domain boundaries[33] (Extended Data Fig. 5d). Finally, in yeast, chromatin looping ("gene loops") was implicated in the control of transcription directionality[34]. Therefore, we interrogated the possibility of CTCF controls transcription directionality by regulation architectural functions.

We first determined whether the repressive effects of CTCF on uasTrx were direct by disrupting CTCF binding at TSS-proximal CTCF binding site (CBS) at three loci (Fig. 2a), *Ahcyl1*, *Azi2* and *Rps3a1*, through CRISPR/Cas9-mediated genome editing (Fig. 2b, Extended Data Fig. 7a and Extended Data Fig. 7-8a)[35]. Upon disruption of TSS-proximal CTCF binding, uasTrx production became elevated while sense transcription remained unperturbed, demonstrating that CTCF binding directly constrains uasTrx production (Fig. 2c-e, Extended Data Fig. 6a,b, Extended Data Fig. 7a-d,f and Extended Data Fig. 8b-d).

Most chromatin boundaries are occupied by CTCF; however, a large fraction of CTCF sites is not associated with domain boundaries or measurable chromatin loops[33]. We thus employed 4C-seq (Circularized Chromosome Conformation Capture sequencing) to determine whether CTCF-bound promoters engage in long-range looped interactions[36]. We focused on the 2 loci, *Ahcyl* and *Azi2*, where uasTrx was strongly and directly suppressed by CTCF, and found significant looping interactions with distant CBSs (Fig. 2b and Extended Data Fig. 7a). Upon CTCF depletion, these loops were strongly diminished, indicating that CBSs are indeed involved in architectural functions at these 2 genes. In light of prior studies in yeast invoking gene looping as a mechanism to constrain uasTrx, we assessed whether CTCF's architectural function constrains uasTrx production[34]. Inspection of the 4C-seq tracks identified the most prominent loop anchors, which we disrupted via CRISPR-Cas9 mediated genome editing in a manner that

preserved promoter-proximal CTCF binding. At the *Ahcyl1* gene, deletion of the distal CTCF site (Dist A) that is associated with the most prominent loop promoter loop (Dist A) led to loss of 4C-seq contacts (Fig. 2c and Extended Data Fig. 6c-e) but no change in uasTrx production (Fig. 2d,e). However, since some contacts remained, we removed two additional CBSs at 4C-seq contact sites (Dist B and Dist C), which further reduced interactions with the promoter proximal CBS (Fig. 2c). None of these perturbations increased uasTrx production (Fig. 2 d,e). At the *Azi2* locus, deletion of distal loop anchors (Dist A and Dist B) but not promoter-proximal CBS led to significant architectural perturbations (Extended Data Fig. 7b-e). In contrast to promoter-primal CBS removal, Dist A/Dist B deletions failed to increase uasTrx production (Extended Data Fig. 7f). Of note, neither CTCF depletion nor CBS removal at the promoters of the *Ahcyl1* and *Azi2* genes detectably increased contacts between uasTrx promoters and surrounding putative enhancers (not shown). This argues against promoter-proximal CBSs functioning as enhancer blocking insulators. Together, these results separate the uasTrx repressive effects of CBSs from their architectural involvement at these loci.

Promoter-proximal CTCF sites involved in inhibition of uasTrx generation are enriched for cohesin (Extended Data Fig. 5a). As an independent means to assess a possible role of CTCF/cohesin-associated loops in regulating uasTrx production, we globally disrupted looped contacts by depleting Nipbl in HCT-116 cells, a cohesin-loading factor[37], and interrogated transcriptional changes. PRO-seq experiments in Nipbl deficient cells revealed minimal uasTrx upregulation (Extended Data Fig. 9a). Finally, we analyzed published data sets in HCT-116 cells in which transient depletion of the cohesin component Rad21 was previously shown to cause genome-wide chromatin organization disruption[38]. Again, we did not observe strand-specific uasTrx changes. Instead, hundreds of genes underwent concomitant changes in both sense and antisense directions (Extended Data Fig. 9b-e), and were not enriched with strong CTCF or Rad21 binding at their promoters (Extended Data Fig. 9f). Together, three orthogonal approaches demonstrate that CTCF inhibits uasTrx directly and proximally, and independently of its architectural functions.

The process of transcription involves multiple steps, including initiation, pausing of RNA polymerase II (Pol II) after transcribing the first 20-60 nucleotides, and release of Pol II into the gene body (GB). CTCF was previously reported as capable of repressing pause-release in the sense direction[39] and was also implicated in impeding Pol II elongation in the GB[40,41]. To

determine the CTCF-controlled step(s) in uasTrx transcription, we took advantage of the single base-pair resolution afforded by PRO-seq and examined the distribution of CTCF motifs relative to the 5' and 3' PRO-seq signals which allows assessment of transcription initiation and stalling, respectively. Only active promoters with proximal (±100bp relative to TSS) CTCF binding sites harboring high-confidence CTCF motifs (motif score>75) were included in the analysis to ensure precise prediction of CTCF positioning (Extended Data Fig. 10a and Supplementary Data 2). Changes in transcription initiation and stalling would be expected to give rise to distinct PRO-seq patterns. Specifically, blockade of Pol II processivity would show significant accumulation of 3' PRO-seq signals (i.e. paused Pol II) upstream of CTCF motifs, which would then get released upon CTCF depletion (Fig. 3a, "stalling"). Release from CTCF-mediated blockade on transcription initiation, on the other hand, would reveal enrichment of 5' PRO-seq signal extending from the motif to the end of uasTrx after CTCF removal (Fig. 3a, "Initiation blockade").

The measured 5' PRO-seq changes triggered by CTCF loss indicate that CTCF impacts antisense transcription initiation (Fig. 3b). Strikingly, CTCF is consistently positioned ~20bp downstream of uasTrx initiation sites at affected promoters, reminiscent of a previous observation that CTCF tends to reside at the borders of transcription initiation clusters[42] (Fig. 3b,c). This distinct spatial arrangement is in stark contrast to the much more variable distribution around unperturbed promoters (Fig. 3d,e). A fraction (120 of 1201) of the unperturbed promoters did have CBSs downstream proximally (Extended Data Fig. 10b, "Downstream proximal"). However, a closer look revealed an upward trend of uasTrx production even though they were not included in the perturbed group because of thresholding (Extended Data Fig. 10c,d). Thus, uasTrx appears to be linked to a particular positioning pattern of CTCF. Finally, 3' PRO-seq reads accumulated downstream, rather than upstream, of CTCF motifs, indicating that Pol II can successfully pass through CTCF without stalling (Fig. 3f and Extended Data Fig. 10e-g). Altogether, the evidence points to CTCF repressing uasTrx transcription through initiation inhibition rather than Pol II stalling, which is consistent with our recent observation that CTCF binding does not strongly interfere with Pol II processivity in the gene body[23].

Transcription is known to occur in bursts, with burst frequency and amplitude subject to modulation[43-45]. To investigate the effects of CTCF on bursting, and whether sense and antisense

transcription are coordinated, we employed single-molecule fluorescence in situ hybridization (smFISH) to quantify at the *Ahcyl1* and *Rps3a1* loci 1) transcription burst size (i.e. amplitude), 2) burst fraction (related to burst frequency), and 3) co-burst frequency (Fig. 4a). CTCF depletion led to no significant changes in burst fraction or size on the sense strand, consistent with bulk PRO-seq readouts. Antisense transcription, on the other hand, underwent significant increases in burst fraction but minimal changes in burst size, suggesting that CTCF mainly affects antisense burst frequency without altering sense transcription dynamics (Fig. 4b,c).

To interrogate sense/antisense burst coordination, we quantified the frequency at which both strands burst alone or together before and after CTCF depletion. At baseline, sense/antisense co-bursting occurred at a minimal number of alleles that is significantly less than expected (i.e. the product of sense and antisense burst fractions), suggesting that co-bursting is highly disfavored (Fig. 4d,e and Extended Data Fig. 11a). Upon CTCF removal, co-burst frequency increased significantly (Extended Data Fig. 11b) but still less frequently than would be expected if these events were independent of each other (Extended Data Fig. 11h). It is important to note that the results are confounded by the unexpectedly long half-lives (>4hr) of uasTrx at both loci (Extended Data Fig. 11c-g), which causes uasTrx transcripts to persist after completion of a burst, thus reducing temporal resolution of smFISH and inflating signal overlap. Regardless, sense and antisense bursts appear to be anti-coordinated temporally when transcribing from the same divergent promoter, suggesting competition between sense and antisense transcription initiation.

A variety of factors have been shown to affect uasTrx transcription, including R-loop formation, oncoprotein MYC, transcription elongation factor SPT6, transcription factor Rap1, looped contacts, histone modifications, and chromatin remodeling proteins (ex. Mot1, Ino80, NC2)[34,46-51]. In many instances, perturbations were accompanied by changes in the sense counterparts, which is in contrast to the present findings and suggests that CTCF functions through mechanism(s) distinct from those previously reported. On the other hand, the CAF-1 complex and histone H3K56 acetylation have been shown to suppress antisense transcription without significantly perturbing sense transcription in yeast[8], but it remains to be tested whether a similar process is operational in mammalian cells and whether CTCF is involved.

Our smFISH results show that CTCF removal increases antisense burst fraction. Since CTCF can block enhancer-promoter contacts[52,53] and since enhancers can increase burst fraction[54], it is conceivable that CTCF loss leads to illegitimate enhancer contacts. However, we did not observe increased long-range contacts upon CTCF loss. Combined with our 5' transcript mapping, this indicates that CTCF inhibits uasTrx production locally at the step of transcription initiation, possibly by preventing PIC formation. The dynamic relationship between sense and antisense transcriptional bursts has not been investigated previously. Single-molecule RNA-FISH at the two genes revealed that co-bursting of divergent transcripts is disfavored, suggesting that at higher temporal resolution the oppositely oriented core promoters may compete at the level of transcription initiation. The mechanisms underlining this competition are unclear, but may include steric hindrance and/or local DNA structure alterations, where supercoiling from transcription in one direction impacts transcription dynamics of the other[55,56].

The competitive relationship of transcriptional bursting was unexpected since at the PRO-seq level no significant reduction in sense transcription was observed upon uasTrx upregulation. We speculate that compensatory mechanisms may buffer against reduction in sense transcription in cases where maintenance of normal gene expression is essential. Finally, although divergent transcription is largely concordant in population-based assays[1,2,8-10], that concordance might be a reflection of overall promoter strength rather than a direct coordination of sense/antisense core promoters.

CTCF at gene promoters has been invoked to facilitate communication with enhancers[16,57]. Nevertheless, CTCF (previously also known as NeP1) was originally shown to function as a direct transcriptional repressor in reporter gene assays[39,58], either alone or perhaps by aiding the adjacent binding of a distinct repressor molecule[58]. The CTCF function uncovered here is novel and distinct in that it blocks initiation selectively of uasTrx production at hundreds of genes without significantly impacting sense transcription. Whether CTCF inhibits chromatin binding of PIC components directly by steric hindrance, by recruiting co-repressors, or by facilitating the binding of neighboring repressor molecules remains to be determined. Regardless, our study demonstrates that CTCF can play separate and independent roles in both genome architecture and transcriptional regulation, even at sites with architectural connectivity.

In sum, we uncovered a novel role for CTCF as direct and selective repressor of uasTrx production, independently of its architectural functions, which expands CTCF's role in the control of the non-coding genome.

References

1. Seila, A. C. *et al.* Divergent transcription from active promoters. *Science (New York, N.Y.)* **322,** 1849–1851 (2008).
2. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science* **322,** 1845–1848 (2008).
3. Preker, P. *et al.* RNA exosome depletion reveals transcription upstream of active human promoters. *Science (New York, N.Y.)* **322,** 1851–1854 (2008).
4. Rhee, H. S. & Pugh, B. F. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483,** 295–301 (2012).
5. Scruggs, B. S. *et al.* Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. *Mol. Cell* **58,** 1101–1112 (2015).
6. Andersson, R., Sandelin, A. & Danko, C. G. A unified architecture of transcriptional regulatory elements. *Trends in Genetics* **31,** 426–433 (2015).
7. Duttke, S. H. C. *et al.* Human promoters are intrinsically directional. *Mol. Cell* **57,** 674–684 (2015).
8. Marquardt, S. *et al.* A Chromatin-Based Mechanism for Limiting Divergent Noncoding Transcription. *Cell* **158,** 462 (2014).
9. Xu, Z. *et al.* Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457,** 1033–1037 (2009).
10. Kapranov, P. *et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science (New York, N.Y.)* **316,** 1484–1488 (2007).
11. Trinklein, N. D. *et al.* An abundance of bidirectional promoters in the human genome. *Genome Res.* **14,** 62–66 (2004).
12. Seila, A. C., Core, L. J., Lis, J. T. & Sharp, P. A. Divergent transcription: a new feature of active promoters. *Cell Cycle* **8,** 2557–2564 (2009).
13. Core, L. J. *et al.* Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet* **46,** 1311–1320 (2014).
14. Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137,** 1194–1211 (2009).
15. Khoury, A. *et al.* Constitutively bound CTCF sites maintain 3D chromatin architecture and long-range epigenetically regulated domains. *Nat Comms* **11,** 54 (2020).
16. Kubo, N. *et al.* Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. *Nature Structural & Molecular Biology* **459,** 108 (2021).
17. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485,** 381–385 (2012).
18. Thiecke, M. J. *et al.* Cohesin-Dependent and -Independent Mechanisms Mediate Chromosomal Contacts between Promoters and Enhancers. *Cell Rep* **32,** 107929 (2020).
19. Wutz, G. *et al.* Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* **36,** 3573–3599 (2017).
20. Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proceedings of the National Academy of Sciences of the United States of America* **111,** 996–1001 (2014).
21. Busslinger, G. A. *et al.* Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* **544,** 503–507 (2017).

22. Hyle, J. *et al.* Acute depletion of CTCF directly affects MYC regulation through loss of enhancer-promoter looping. *Nucleic Acids Research* **47,** 6699–6713 (2019).
23. Luan, J. *et al.* Distinct properties and functions of CTCF revealed by a rapidly inducible degron system. *Cell Rep* **34,** 108783 (2021).
24. Mattick, J. S. & Makunin, I. V. Non-coding RNA. *Hum Mol Genet* **15 Spec No 1,** R17–29 (2006).
25. Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science (New York, N.Y.)* **339,** 950–953 (2013).
26. Wyers, F. *et al.* Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* **121,** 725–737 (2005).
27. van Dijk, E. L. *et al.* XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* **475,** 114–117 (2011).
28. Bornelöv, S., Komorowski, J. & Wadelius, C. Different distribution of histone modifications in genes with unidirectional and bidirectional transcription and a role of CTCF and cohesin in directing transcription. *BMC Genomics 2013 14:1* **16,** 300 (2015).
29. Shen, Y. *et al.* A map of the cis-regulatory sequences in the mouse genome. *Nature* **488,** 116–120 (2012).
30. Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169,** 930–944.e22 (2017).
31. Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep* **15,** 2038–2049 (2016).
32. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences of the United States of America* **112,** E6456–65 (2015).
33. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).
34. Tan-Wong, S. M. *et al.* Gene loops enhance transcriptional directionality. *Science (New York, N.Y.)* **338,** 671–675 (2012).
35. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8,** 2281–2308 (2013).
36. van de Werken, H. J. G. *et al.* Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat Meth* **9,** 969–972 (2012).
37. Schwarzer, W. *et al.* Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551,** 51–56 (2017).
38. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171,** 305–320.e24 (2017).
39. Filippova, G. N. *et al.* An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Molecular and Cellular Biology* **16,** 2802–2813 (1996).
40. Shukla, S. *et al.* CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479,** 74–79 (2011).
41. Mayer, A. *et al.* Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* **161,** 541–554 (2015).

42. Tome, J. M., Tippens, N. D. & Lis, J. T. Single-molecule nascent RNA sequencing identifies regulatory domain architecture at promoters and enhancers. *Nat Genet* **50,** 1533–1541 (2018).
43. Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-time kinetics of gene activity in individual bacteria. *Cell* **123,** 1025–1036 (2005).
44. Chubb, J. R., Trcek, T., Shenoy, S. M. & Singer, R. H. Transcriptional pulsing of a developmental gene. *Current Biology* **16,** 1018–1025 (2006).
45. Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* **4,** e309 (2006).
46. Wu, A. C. K. *et al.* Repression of Divergent Noncoding Transcription by a Sequence-Specific Transcription Factor. *Mol. Cell* **72,** 942–954.e7 (2018).
47. Sansó, M. *et al.* Cdk9 and H2Bub1 signal to Clr6-CII/Rpd3S to suppress aberrant antisense transcription. *Nucleic Acids Research* **48,** 7154–7168 (2020).
48. Baluapuri, A. *et al.* MYC Recruits SPT5 to RNA Polymerase II to Promote Processive Transcription Elongation. *Mol. Cell* **74,** 674–687.e11 (2019).
49. Xue, Y. *et al.* Mot1, Ino80C, and NC2 Function Coordinately to Regulate Pervasive Transcription in Yeast and Mammals. *Mol. Cell* **67,** 594–607.e4 (2017).
50. Tan-Wong, S. M., Dhir, S. & Proudfoot, N. J. R-Loops Promote Antisense Transcription across the Mammalian Genome. *Mol. Cell* **76,** 600–616.e6 (2019).
51. Nojima, T. *et al.* Deregulated Expression of Mammalian lncRNA through Loss of SPT6 Induces R-Loop Formation, Replication Stress, and Cellular Senescence. *Mol. Cell* **72,** 970–984.e7 (2018).
52. Hou, C., Zhao, H., Tanimoto, K. & Dean, A. CTCF-dependent enhancer-blocking by alternative chromatin loop formation. *Proceedings of the National Academy of Sciences of the United States of America* **105,** 20398–20403 (2008).
53. Hsu, S. C. *et al.* The BET Protein BRD2 Cooperates with CTCF to Enforce Transcriptional and Architectural Boundaries. *Mol. Cell* **66,** 102–116.e7 (2017).
54. Bartman, C. R., Hsu, S. C., Hsiung, C. C.-S., Raj, A. & Blobel, G. A. Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Mol. Cell* **62,** 237–247 (2016).
55. Lim, H. M., Lewis, D. E. A., Lee, H. J., Liu, M. & Adhya, S. Effect of varying the supercoiling of DNA on transcription and its regulation. *Biochemistry* **42,** 10718–10725 (2003).
56. Peter, B. J. *et al.* Genomic transcriptional response to loss of chromosomal supercoiling in Escherichia coli. **5,** R87 (2004).
57. Lee, J., Krivega, I., Dale, R. K. & Dean, A. The LDB1 Complex Co-opts CTCF for Erythroid Lineage-Specific Long-Range Enhancer Interactions. *Cell Rep* **19,** 2490–2502 (2017).
58. Baniahmad, A., Steiner, C., Köhne, A. C. & Renkawitz, R. Modular structure of a chicken lysozyme silencer: involvement of an unusual thyroid hormone receptor binding site. *Cell* **61,** 505–514 (1990).
59. Weiss, M. J., Yu, C. & Orkin, S. H. Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line. *Molecular and Cellular Biology* **17,** 1642–1651 (1997).
60. Luppino, J. M. *et al.* Cohesin promotes stochastic domain intermingling to ensure proper regulation of boundary-proximal genes. *Nat Genet* **52,** 840–848 (2020).

61. Stonestrom, A. J. *et al.* Functions of BET proteins in erythroid gene expression. *Blood* **125,** 2825–2834 (2015).
62. Cong, L. & Zhang, F. Genome engineering using CRISPR-Cas9 system. *Methods Mol. Biol.* **1239,** 197–217 (2015).
63. Reimer, K. A., Mimoso, C. A., Adelman, K. & Neugebauer, K. M. Co-transcriptional splicing regulates 3' end cleavage during mammalian erythropoiesis. *Mol. Cell* **81,** 998–1012.e7 (2021).
64. Letting, D. L., Chen, Y.-Y., Rakowski, C., Reedy, S. & Blobel, G. A. Context-dependent regulation of GATA-1 by friend of GATA-1. *Proc Natl Acad Sci USA* **101,** 476–481 (2004).
65. Splinter, E., de Wit, E., van de Werken, H. J. G., Klous, P. & de Laat, W. Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. *Methods* **58,** 221–230 (2012).
66. van de Werken, H. J. G. *et al.* 4C technology: protocols and data analysis. *Methods Enzymol* **513,** 89–112 (2012).
67. Femino, A. M., Fay, F. S., Fogarty, K. & Singer, R. H. Visualization of single RNA transcripts in situ. *Science* **280,** 585–590 (1998).
68. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **15,** 550 (2014).
69. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Meth* **9,** 357–359 (2012).
70. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). **9,** R137 (2008).
71. Raj, A., Rifkin, S. A., Andersen, E. & van Oudenaarden, A. Variability in gene expression underlies incomplete penetrance. *Nature* **463,** 913–918 (2010).

Acknowledgments

We are grateful to Hardison, Raj, Lis, and Blobel laboratories for insightful discussions.

Funding

This work was supported by NIH grants R01 DK054937 to G.A.B., R24 DK106766 to G.A.B. and R.C.H., T32 HL07439 to C.M.S., R01GM121613 to R.C.H, and U01 DK127405 to G.A.B. and A.R.

Author contributions

J.L. and G.A.B. conceived the study and designed experiments. J.L. and C.A.K. performed ChIP-seq experiments; J.L., M.W.V., and B.M.G. performed ChIP-seq analysis. J.L., C.M.S, and A.H. performed CRISPR editing experiments and 4C experiments. 4C results were analyzed by J.L. and S.Z. J.M.L. performed PRO-seq experiments; J.L. and Z.Z. analyzed PRO-seq results with advice from J.M.T. and J.T.L. C.M.S., M.G., and A.H. performed single-molecule FISH experiments, with data analyzed by C.M.S., J.L., and A.C. under the supervision of A.R. J.L. and G.A.B. wrote the manuscript with input from all authors.

Competing interests: The authors declare no competing interests.

Data and materials availability

All sequencing and processed data have been deposited at GEO under accession GSE173442, GSE173443, GSE173444.

Figure 1

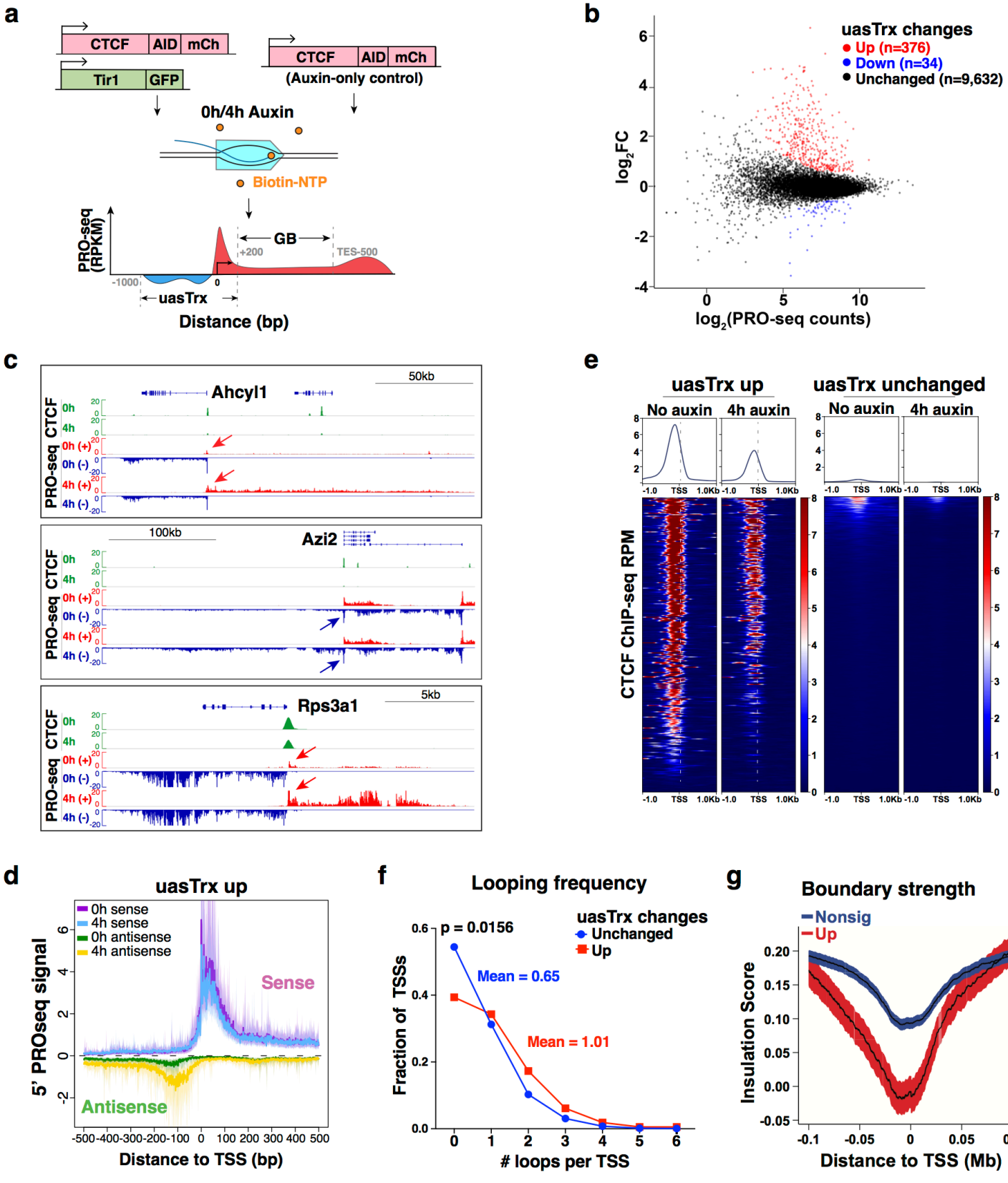


Fig. 1 | Transient CTCF depletion leads to widespread antisense transcription up-regulation at divergent promoters. a, Schematics of PRO-seq experiment and quantification strategy. **b,** PRO-seq MA plot of control versus CTCF-depleted cells on the antisense strand (-1000bp to +200 relative to annotated TSS) in G1E-ER4s. Differentially expressed transcripts highlighted in color. **c,** Genome browser views of CTCF ChIP-seq and PRO-seq signals at *Ahcyl1*, *Azi2* and *Rps3a1* loci. Arrows point to uasTrx, with colors indicating strandedness. **d,** Metaplot of sense and antisense 5' PRO-seq signals at activated uasTrx, centered at annotated TSSs and plotted with respect to sense orientation. Solid lines and shades show bootstrapped estimates of average signals and the 12.5/87.5 percentiles, respectively. **e,** Row-linked heatmaps showing CTCF occupancy at active promoters (up n=376; unchanged n=9,632), grouped by uasTrx changes upon CTCF depletion, sorted by occupancy level and shown with respect to sense orientation. **f,** Frequency of looping interactions engaged by all gained and unchanged uasTrx. *P* value calculated by Wilcoxon signed-rank test. **g,** Averaged insulation score centered at annotated TSSs over 0.2Mb window, grouped by uasTrx changes, and plotted with respect to sense orientation.

Figure 2

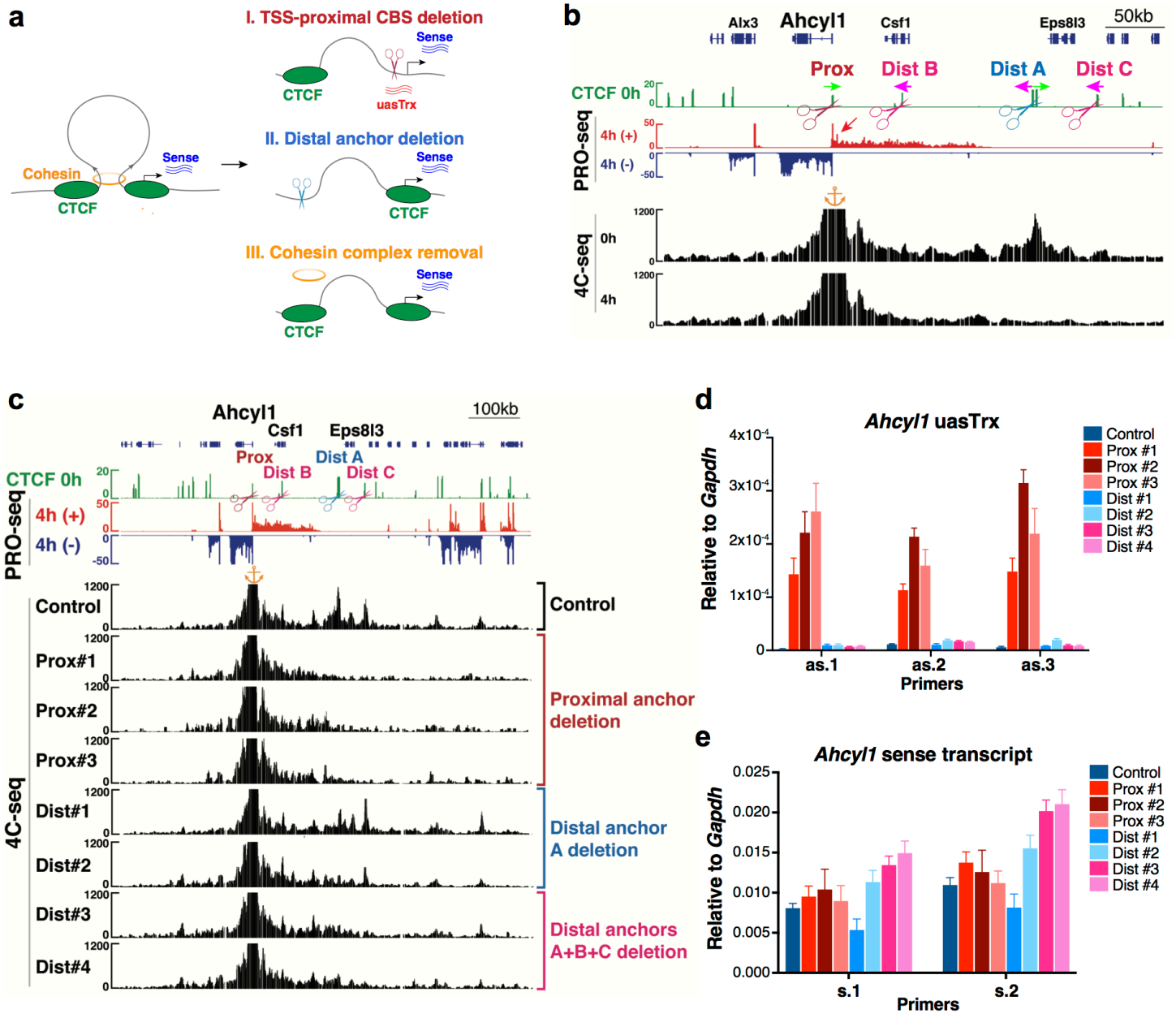


Fig. 2 | CTCF inhibits uasTrx directly and proximally, independent of its architectural functions. a, Illustration of experimental strategy and summarized findings in **Fig. 2** and **Extended Data Figs. 6-9. b,** Genome browser views of CTCF ChIP-seq, PRO-seq and 4C-seq signals at *Ahcyl1*. 4C-seq anchored at *Ahcyl1* promoter. Colored arrows above ChIP-seq track indicate CTCF motif directionality. Red arrow points to elevated uasTrx after CTCF depletion. Scissors point to regions disrupted by CRISPR/Cas9-mediated genome editing, one at CBS proximal to *Ahcyl1* promoter and the others at a distal CBSs engaging in loop contacts with the promoter. Orange anchor indicates 4C-seq viewpoint. **c,** Genome browser tracks of bulk CTCF ChIP-seq and PRO-seq and representative 4C-seq profiles of control and edited clones with indicated regions disrupted. Similar observations were made in 2 additional independent 4C-seq experiments and not shown. Orange anchor indicates 4C-seq viewpoint. **d,** RT-qPCR of *Ahcyl1* uasTrx in control and edited clones. Transcripts were normalized to *Gapdh* (error bar: SEM; n=3-4). **e,** Same as (**d**) but of nascent *Ahcyl1* sense transcripts. Prox, TSS-proximal CBS. Dist, distal anchor.

Figure 3

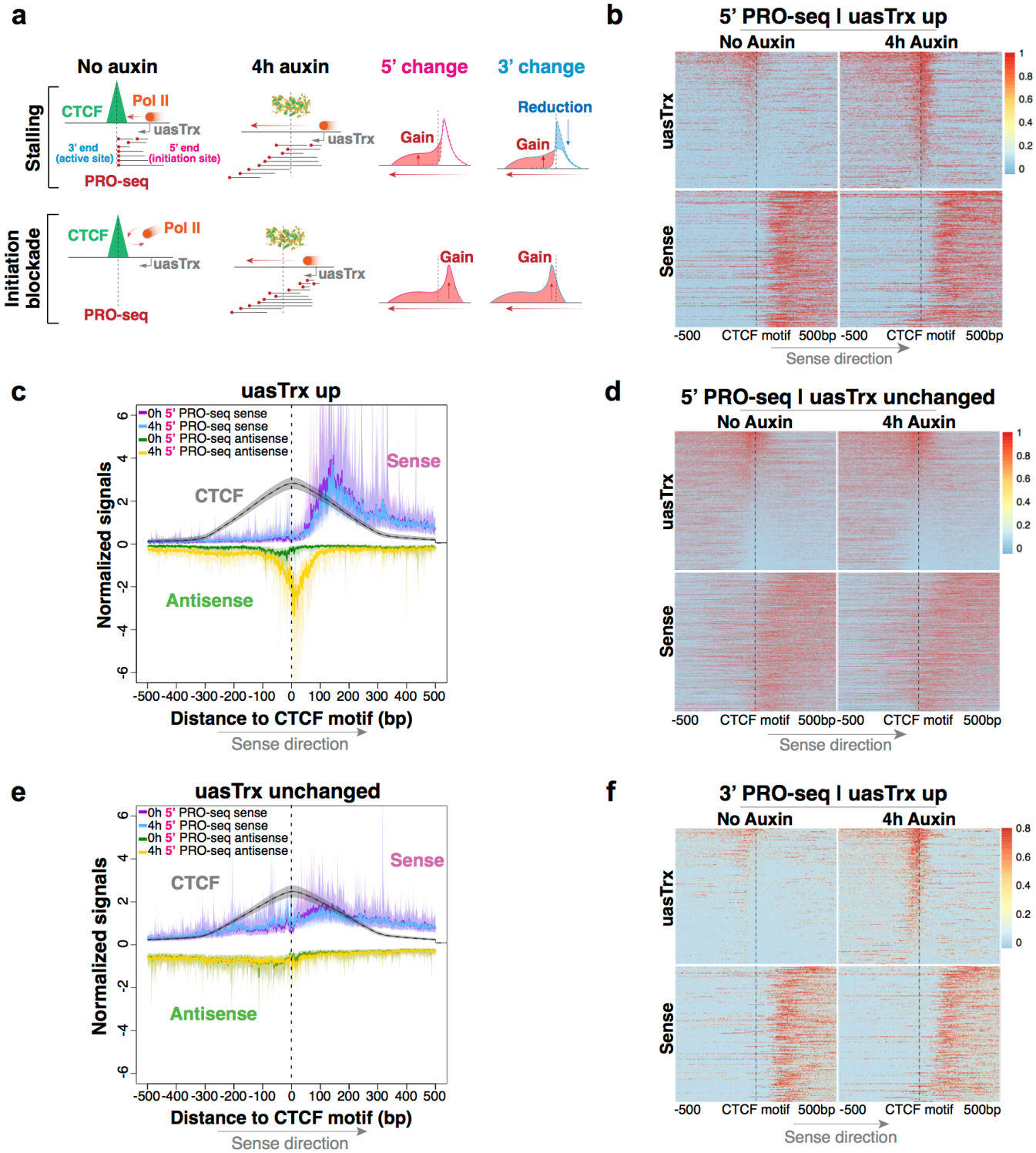


Fig. 3 | CTCF inhibits antisense transcription initiation through proximal binding. a, Model illustrating expected 5' and 3' PRO-seq distribution when CTCF blocks transcription processivity or initiation. **b,** 5' PRO-seq heatmap at affected active promoters (n=298) that exhibit proximal CTCF binding and high-confidence CTCF motif(s) (motif prediction score>75), centered at CTCF motifs, sorted by mean antisense signal densities over the center 200bp and shown with respect to sense orientation. Black line highlights CTCF motif locations. **c,** Metaplot centered by CTCF motifs summarizing 5' PRO-seq and CTCF signals shown in **b**. Solid lines and shades show bootstrapped estimates of average signals and the 12.5/87.5 percentiles, respectively. **d,** Same as (**b**) but at unaffected promoters (n=1,201) that satisfy the same CTCF criteria. **e,** Same as (**c**) but summarizing sites in **d**. **f,** Same as (**b**), but plotting 3' PRO-seq signals at activated uasTrx.

Figure 4

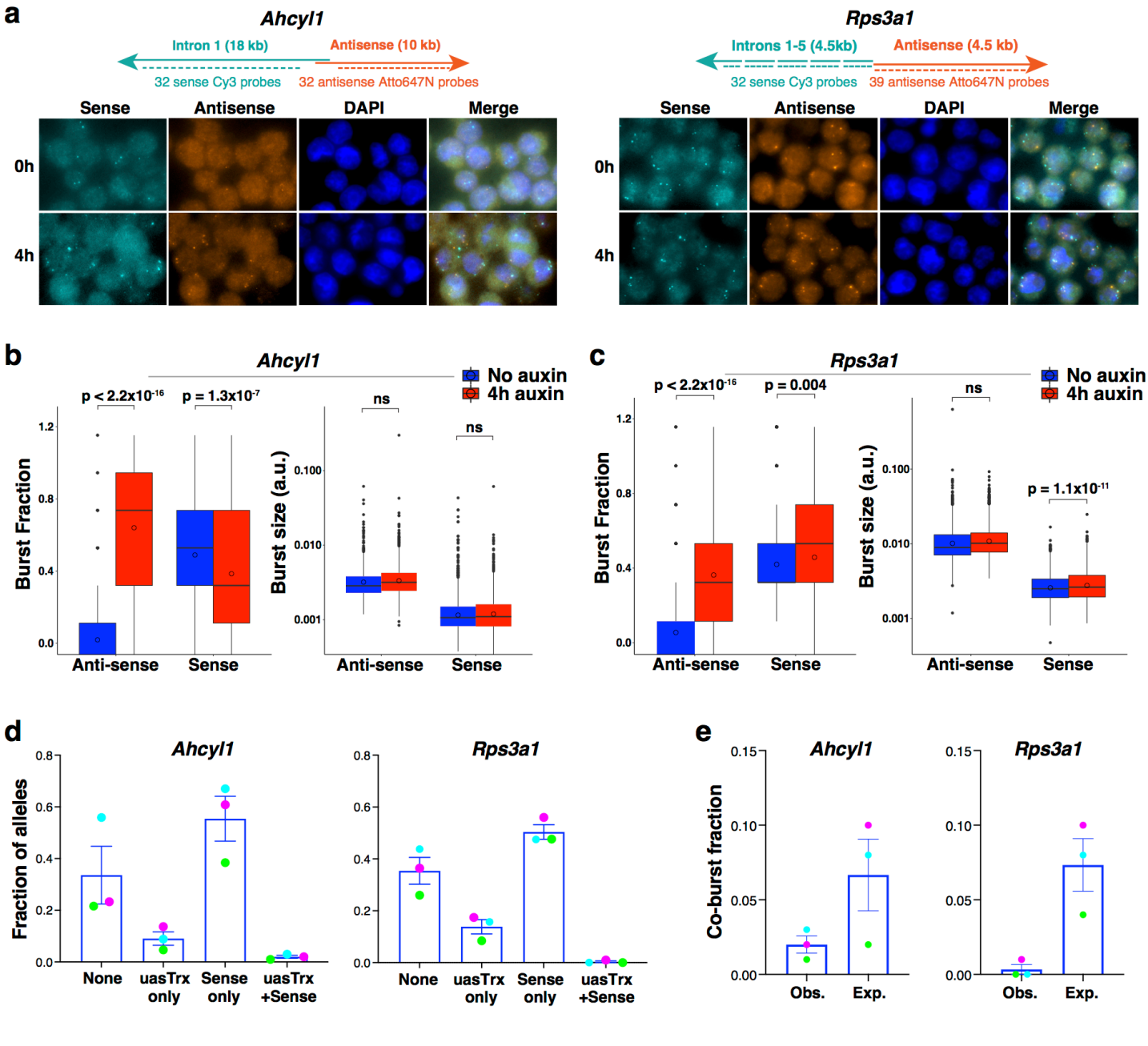


Fig. 4 | CTCF mainly regulates antisense burst fraction; sense and antisense bursts appear to compete temporally at divergent promoters. a, Top, maps of FISH probes targeting sense and antisense nascent transcripts at *Rps3a1* and *Ahcyl1* loci. Bottom, representative FISH images before and after CTCF depletion. **b,** Left, box plots showing antisense and sense burst fractions before and after CTCF depletion at *Ahcyl1*. Right, box plot showing antisense and sense burst sizes before and after CTCF depletion. n=3 biological replicates. *P* values were calculated by two-sample *t*-test. **c,** Same as (**b**) but at *Rps3a1*. **d,** Left, fraction of *Ahcyl1* alleles with 4 different sense/antisense bursting status at baseline (error bar: SEM; n=3). Right, same as left but at *Rps3a1*. Biological replicates matched by dot colors. **e,** Left, predicted and observed co-burst fraction at *Ahcyl1* at baseline (error bar: SEM; n=3). Right, same as left but at *Rps3a1*. Biological replicates matched by dot colors.

Methods

Experiments. *Cell culture and maintenance.* G1E-ER4 is an established murine erythroblast cell line[59]. G1E-ER4 cells were grown in IMDM+15% FBS, penicillin/ streptomycin, Kit ligand, monothioglycerol and erythropoietin in a standard tissue culture incubator at 37ºC with 5% CO_2. Cells were maintained at density below 1 million/ml at all times. Transient CTCF depletion in G1E-ER4 cells was induced by 1mM auxin in culture. Nascent RNA half-life was assessed by quantifying transcript levels via smFISH and RT-qPCR after transcription blockade for 0h, 4h and 6h with 75uM DRB. HCT-116 cells were cultured in McCoy's 5A medium supplemented with 10% fetal bovine serum, 2 mM L-glutamine, 100 U ml^{-1} penicillin, and 100 μg ml^{-1} streptomycin at 37°C with 5% CO2.

siRNA-mediated CTCF/Nipbl depletion. RNAi was performed in HCT-116 cells as previously described using the same published guide sequences[60] with a final siRNA concentration of 50 nM (non-targeting control, NIPBL) or 150 nM (CTCF). Cells were harvested after 72 hr treatment.

CRISPR-Cas9-mediated genome editing. We performed all CRISPR editing in a previously established Cas9-TagBFP expressing G1E-ER4 cell line to enhance editing efficiency[23]. All sgRNA encoding oligonucleotides were inserted into a retroviral U6-sgRNA-PGK-GFP expression vector[61] using a BsmBI restriction site and transfected into cells by Amaxa II electroporator (Lonza; program G-016) and Amax II Cell Line Nucleofector Kit (R) (Lonza, VCA-1001). GFP+ cells were sorted by FACS 24h post-transfection, followed by single-cell clone screening and genotyping by Sanger sequencing. All guide RNA sequences were obtained using CRISPR design tool (https://zlab.bio/guide-design-resources)[62]. Guide sequences are listed in Extended Data Table 2.

PRO-seq library preparation. PRO-seq libraries in G1E-ER4 was performed as described previously[23]. For each library, 50 million cells were used with 2 million Drosophila Schneider 2 (S2) cells added as spike-in to control for potential global bias associated with library scaling. Fragments longer than 140bp from the PCR-amplified library were selected and sequenced (2x75bp) on the Illumina NextSeq 500 platform according to manufacturer's instructions to a depth of ~100 million/library.

PRO-seq libraries in HCT-116 were performed by the Nascent Transcriptomics Core at Harvard Medical School, Boston, MA. Specifically, aliquots of frozen (-80°C) permeabilized cells were thawed on ice and pipetted gently to fully resuspend. For each sample, 1 million permeabilized cells were used, with 50,000 permeabilized Drosophila S2 added for normalization. Nuclear run on assays and library preparation were performed as described[63] with following modifications: 2X nuclear run-on buffer consisted of 10 mM Tris (pH 8), 10 mM MgCl2, 1 mM DTT, 300mM KCl, 40uM/ea biotin-11-NTPs

(Perkin Elmer), 0.8U/uL SuperaseIN (Thermo), 1% sarkosyl. Run-on reactions were performed at 37°C. Adenylated 3' adapter was prepared using the 5' DNA adenylation kit (NEB) and ligated using T4 RNA ligase 2, truncated KQ (NEB, per manufacturer's instructions with 15% PEG-8000 final) and incubated at 16°C overnight. 180uL of betaine blocking buffer (1.42g of betaine brought to 10mL with binding buffer supplemented to 0.6 uM blocking oligo (TCCGACGATCCCACGTTCCCGTGG/3InvdT/)) was mixed with ligations and incubated 5 min at 65°C and 2 min on ice prior to addition of streptavidin beads. After T4 polynucleotide kinase (NEB) treatment, beads were washed once each with high salt, low salt, and blocking oligo wash (0.25X T4 RNA ligase buffer (NEB), 0.3uM blocking oligo) solutions and resuspended in 5' adapter mix (10 pmol 5' adapter, 30 pmol blocking oligo, water). 5' adapter ligation was per Reimer but with 15% PEG-8000 final. Eluted cDNA was amplified 5-cycles (NEBNext Ultra II Q5 master mix (NEB) with Illumina TruSeq PCR primers RP-1 and RPI-X) following the manufacturer's suggested cycling protocol for library construction. A portion of preCR was serially diluted and for test amplification to determine optimal amplification of final libraries. Pooled libraries were sequenced using the Illumina NovaSeq platform.

RNA extraction, cDNA synthesis and RT-qPCR. Cells were harvested in buffer RLT Plus (Qiagen, Cat # 1053393) with lysate homogenized using QIAshredders (Qiagen, Cat # 79656), followed by RNA purification with RNeasy Mini Kit that included an on-column DNase treatment step (Qiagen, Cat. #74106). Complementary DNA (cDNA) was synthesized with iScript Supermix (Bio-Rad, Cat. #1708841). Quantitative polymerase chain reaction (qPCR) was performed using Power SYBR Green kit (Invitrogen; 4368577) with signals detected by ViiA7 System (Life Technologies). Primers used for RT-qPCR are listed in Extended Data Table 3.

ChIP-seq library preparation. Chromatin immunoprecipitation (ChIP) was performed as previously described[64]. Antibodies include: CTCF (Millipore; 07-729), POLR2A (Cell Signaling; Cat#14958), IgG from rabbit serum (Sigma; 15006). Quantitative polymerase chain reaction (qPCR) was performed using Power SYBR Green kit (Invitrogen; 4368577) with signals detected by ViiA7 System (Life Technologies). ChIP-seq libraries were prepared using Illumina's TruSeq ChIP sample preparation kit (Illumina, Cat#IP-202-1012) according to manufacturer's specifications, with the addition of size selection (left side at 0.9x, right side at 0.6x) using SPRIselect beads (Beckman Coulter, Cat#B23318). Library size was determined (average 351 bp, range 333-372 bp) using the Agilent Bioanalyzer 2100, followed by quantitation using real-time PCR using the KAPA Library Quant Kit for Illumina (KAPA Biosystems; Cat#KK4835). Libraries were then pooled and sequenced (1x75bp) on the Illumina NextSeq

500 platform according to manufacturer's instructions. Bclfastq2 v 2.15.04 (default parameters) was used to convert reads to fastq. Primers used for RT-qPCR are listed in Extended Data Table 4.

4C-seq sample preparation. The 4C experiments were performed as previously described using DpnII and Csp6I as restriction enzymes[65,66]. Sequencing was done on Illumina Hiseq 2000 genome sequencer with reads mapped onto mm9. Reads mapping to multiple fragment ends were removed, and 4C coverage was computed by averaging mapped reads in running windows of 41 fragment ends. Amplification primers for each view point are listed in Extended Data Table 5. Quality of all libraries meet the previously described standards[66] based on the *cis*/overall ratio and the percentage of covered fragends within 0.2Mb window around the viewpoints.

smFISH imaging. Single-molecule RNA FISH was performed as previously described[45,67]. All sense probes used were complementary to introns of gene of interest and are listed in Extended Data Table 6. Briefly, cells were fixed in 1.85% formaldehyde for 10 min at room temperature, and stored in 70% ethanol at 4ºC. Pools of fluorophore-conjugated FISH probes were hybridized to samples overnight, followed by DAPI staining and washes performed in suspension. Cells were cytospun onto slides for imaging on a Nikon Ti-E inverted fluorescence microscope using a 100x Plan-Apo objective (numerical aperture of 1.43), a cooled CCD camera (Pixis 1024B from Princeton Instruments), and filter sets SP102v1 (Chroma), SP104v2 (Chroma), and 31000v2 (Chroma) for Cy3, Atto647N, and DAPI, respectively. Slides were imaged in 36 optical z sections at intervals of 0.35 microns with 1 s exposure time for Cy3/Atto647N and 35 ms for DAPI.

Analysis. *PRO-seq quantification.* Read alignment and identification of active transcripts have been described in detail previously[23]. An arbitrary window of +200bp relative to Refseq-annotated TSS to -500 bp relative to TES (transcription end site) was used to quantify sense transcript levels to avoid any confounding effects associated with promoter-proximal pausing. A window of -1000bp to +200bp relative to TSS was selected to quantify uasTrx changes unless noted otherwise. Differential expression analysis was performed using paired DESeq2 method[68] with FDR<0.05 & fold-change>1.5 as thresholds. Each up-regulated uasTrx in G1E-ER4s was confirmed visually to rule out false positives such as run-throughs from nearby up-regulated genes. For analysis of PRO-seq datasets published in Rao et al. 2017, only active genes identified by the authors were included for characterization.

The start and end sites of uasTrx were annotated as follows: 1) Reads less than 100bp long were extended to 100bp from the 3' end to "smooth over" PRO-seq signals. 2) Regions overlapping any known transcripts were masked. 3) Global averaged sequencing depth was obtained by dividing all mapped reads

over the entire genome. 4) Unbroken regions starting within 500bp of the annotated TSSs on the antisense strand and with sequencing depth exceeding global average were counted as part of uasTrx and taken into consideration for length estimates.

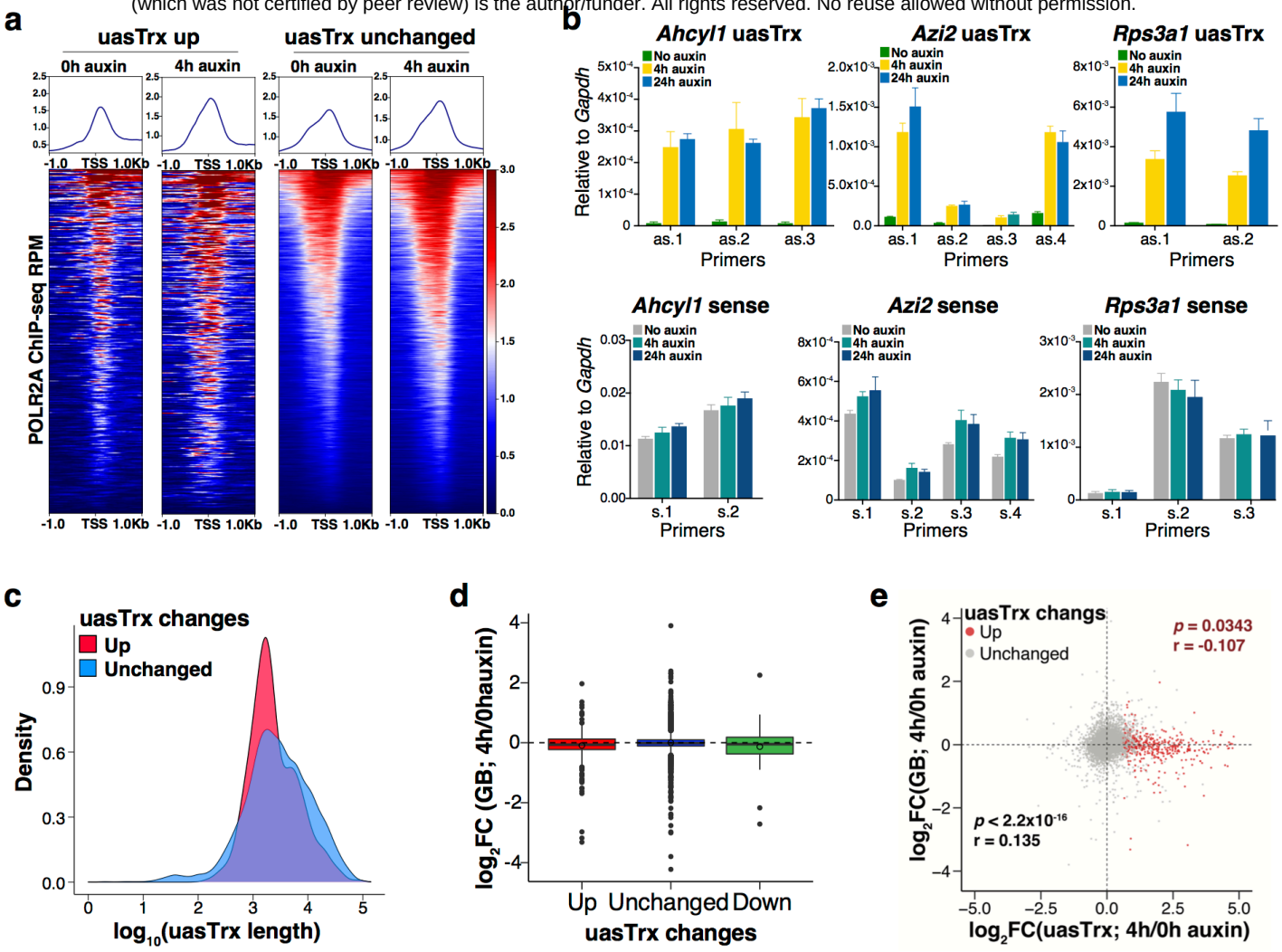
RNA-seq quantification. A window of -2000bp to -50bp relative to annotated TSSs was used to quantify uasTrx in unstranded RNA-seq datasets published in Nora et al. 2017 to minimize inclusion of sense signals. DESeq2 was applied to read count matrix to evaluate differential expression between groups.

ChIP-seq analysis. Bowtie 1.1.0 was used to align sequences to the mm9 reference genome[69]. Reads with more than one mismatch or multiple alignments were excluded. Significantly enriched regions were called using MACS2 version 2.1.0[70] with the following parameters: p = 10 5, extsize = 300 and local lambda = 100,000 using whole-cell extract input controls. Reads for the bigwigs were RPM normalized.

smFISH image analysis. Nuclear boundaries were segmented manually from DAPI images, with RNA spots localized and quantified using custom software written in MATLAB[71]. Transcription sites were identified by bright nuclear intron spots; fluorescence intensities of transcription sites were determined by 2D Gaussian fitting on processed image data. Subsequent analysis was performed in R. To identify sense and antisense co-transcription status, a wide range of sense-antisense distance thresholds were tested, ranging from 1 pixel (our resolution limit) to 10 pixels (1.3μm). Almost all distance thresholds yielded similar results. Results shown in Fig. 3 and Extended Data Fig. 11 are based on distance threshold of 3 pixels (0.39μm).

Gene ontology analysis. Gene ontology (GO) analysis was performed using PANTHER overrepresentation test (release 20210224) against all Mus musculus genes in the database as background. The Fisher's exact test was performed with FDR correction. GO Ontology database DOI: 10.5281/zenodo.4495804 (released 2021-02-01).

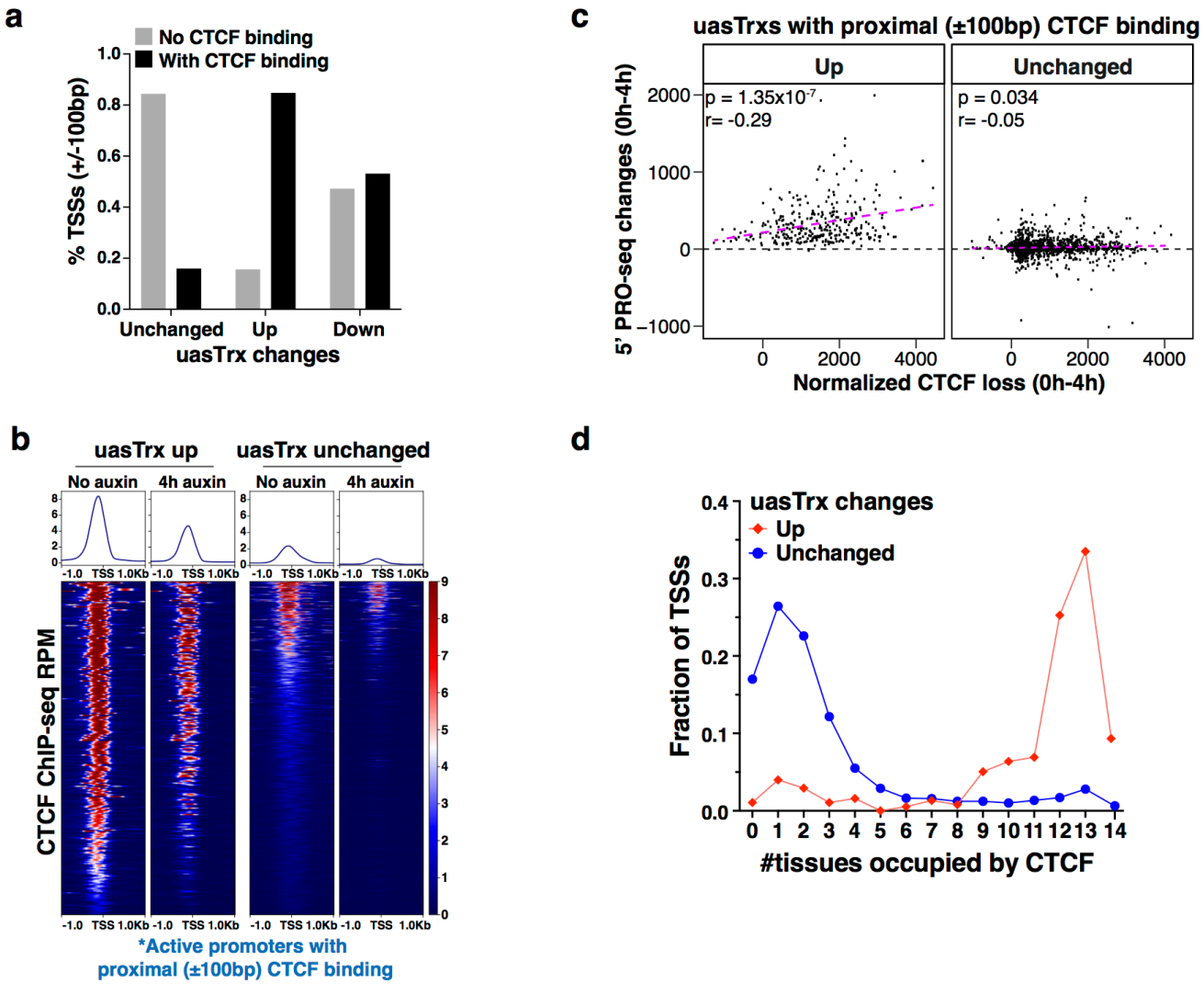
Metaplots. All metaplots were generated as previously described[42] and show estimated average signals and the 87.5 and 12.5 percentiles obtained from bootstrapping.



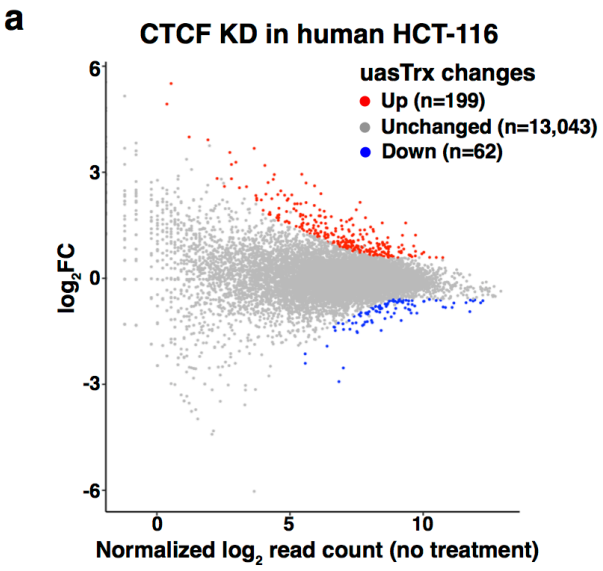
f

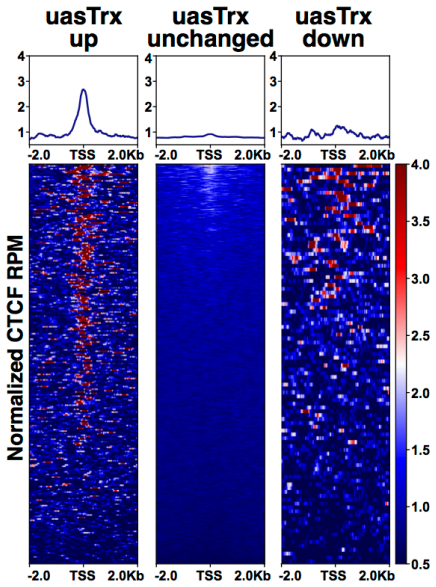
uasTrx changes	GB changes: Up	Unchanged	Down
Up	1.3%	94.9%	3.7%
Unchanged	0.5%	99%	0.5%

Extended Data Fig.1 | Transient CTCF depletion leads to widespread uasTrx up-regulation at divergent promoters. a, Row-linked heatmaps showing POLR2A occupancy at active promoters, grouped by antisense changes (up n= 376; unchanged n=9,632) upon CTCF depletion, sorted by occupancy level, and shown with respect to sense orientation. **b,** Top, RT-qPCR of uasTrx at indicated loci at indicated time points after CTCF depletion. Transcripts were normalized to *Gapdh* (error bar: SEM; n=3-4). Bottom, same as top but quantifying nascent sense transcripts. **c,** Distribution of uasTrx lengths, grouped by changes in response to CTCF depletion. **d,** Log-transformed PRO-seq fold changes in GB after CTCF depletion, grouped by uasTrx changes. **e,** Scatterplot comparing transcriptional changes in GB versus uasTrx. Data points grouped and colored based on uasTrx changes. **f,** Transcriptional changes in uasTrx and GB after CTCF depletion.



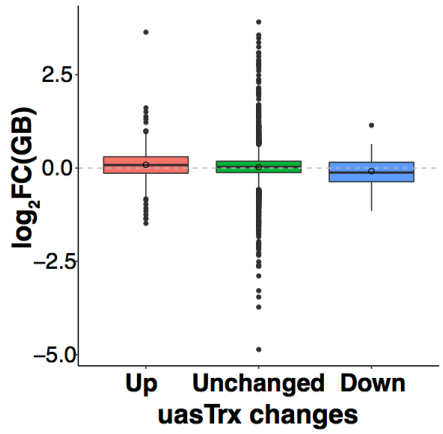
Extended Data Fig.2 | CBSs proximal to activated uasTrx exhibit distinct features. a, Percentage of promoters with and without proximal (±100bp) CBSs as a function of uasTrx changes. **b,** Heatmaps showing CTCF occupancy at active promoters with proximal (±100bp) CTCF binding (up n=319; unchanged n=1,527), sorted by occupancy level, and shown with respect to sense orientation. **c,** Correlation between 5' PRO-seq changes and CTCF loss at uasTrx with proximal (±100bp) CTCF binding. Linear regression line shown in magenta. *P* value was calculated by Spearman rank correlation test; r is the correlation coefficient. **d,** Fraction of TSSs detected in the indicated numbers of mouse tissues where CTCF binds in proximity (within ±100bp), grouped by uasTrx changes.

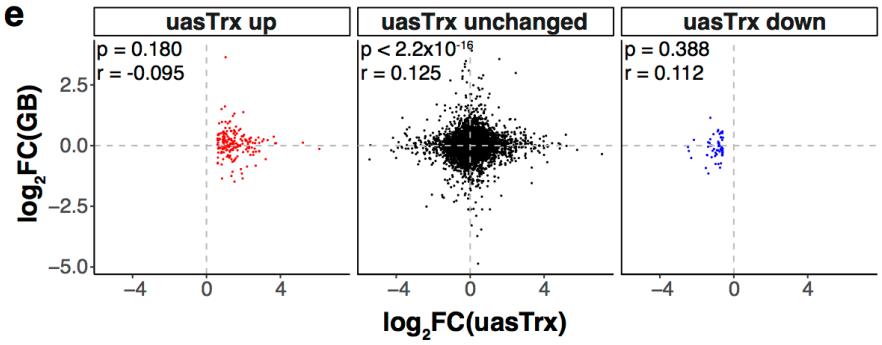


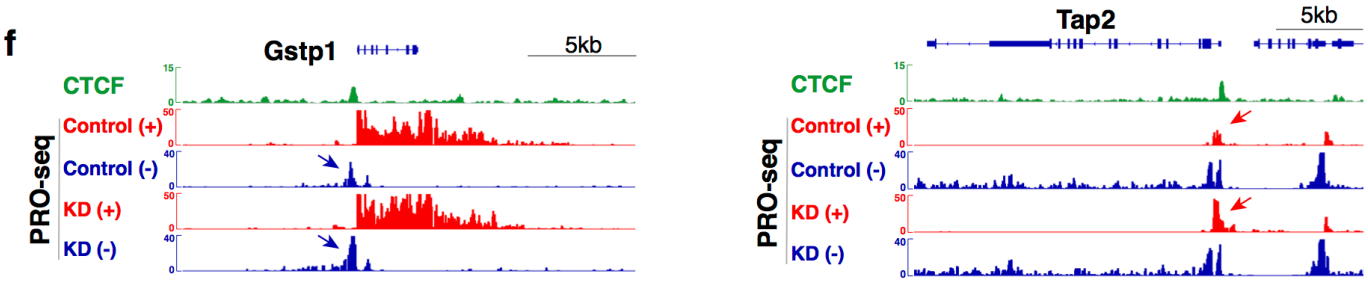


c

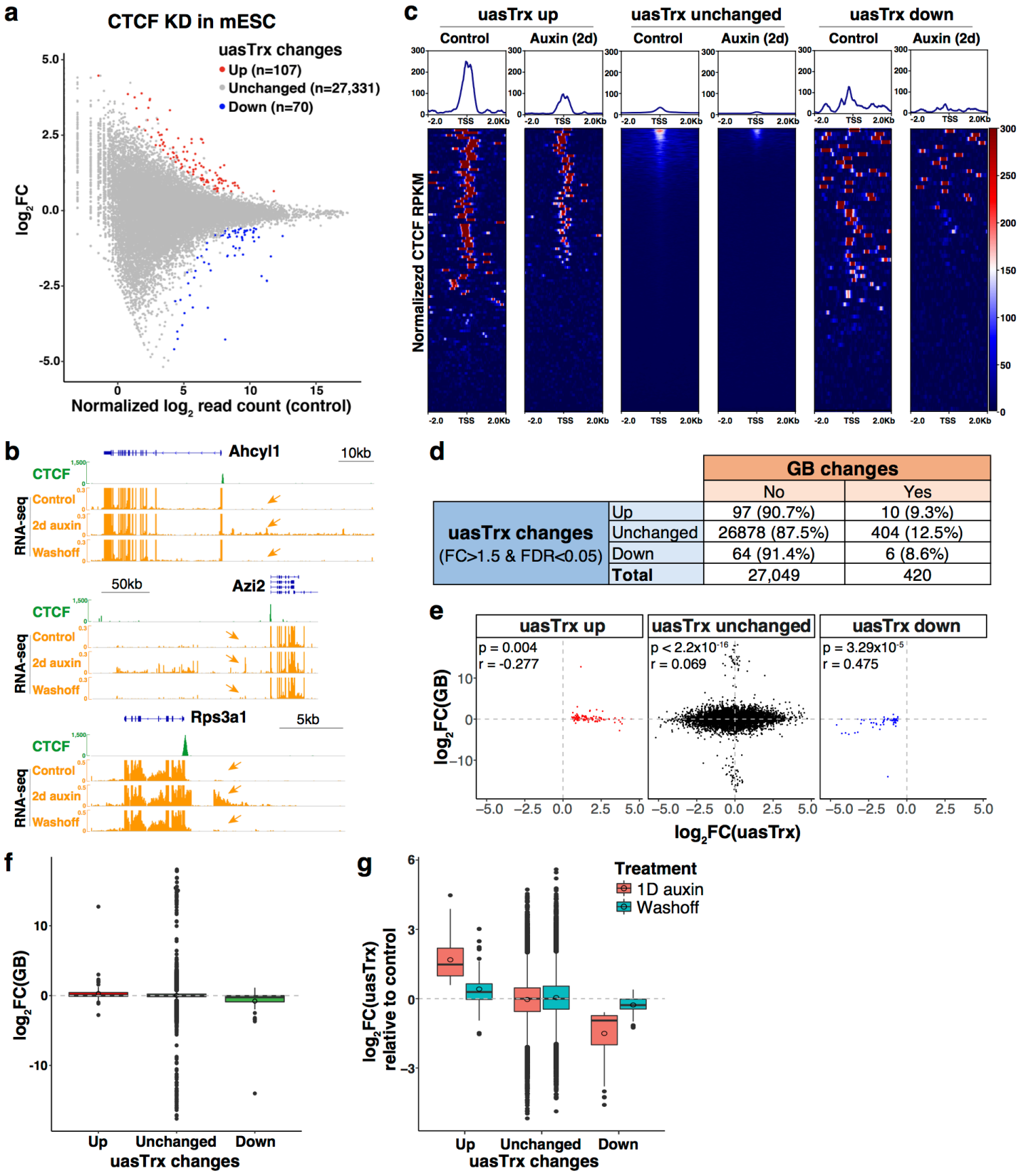
		GB changes (FC>1.5 & FDR<0.05)		
		Up	Unchanged	Down
uasTrx changes (FC>1.5 & FDR<0.05)	Up	15 (4.0%)	176 (88.4%)	8 (7.5%)
	Unchanged	236 (1.8%)	12,549 (96.2%)	258 (2.0%)
	Down	2 (3.2%)	55 (88.7%)	5 (8.1%)
	Total	253	12,780	271



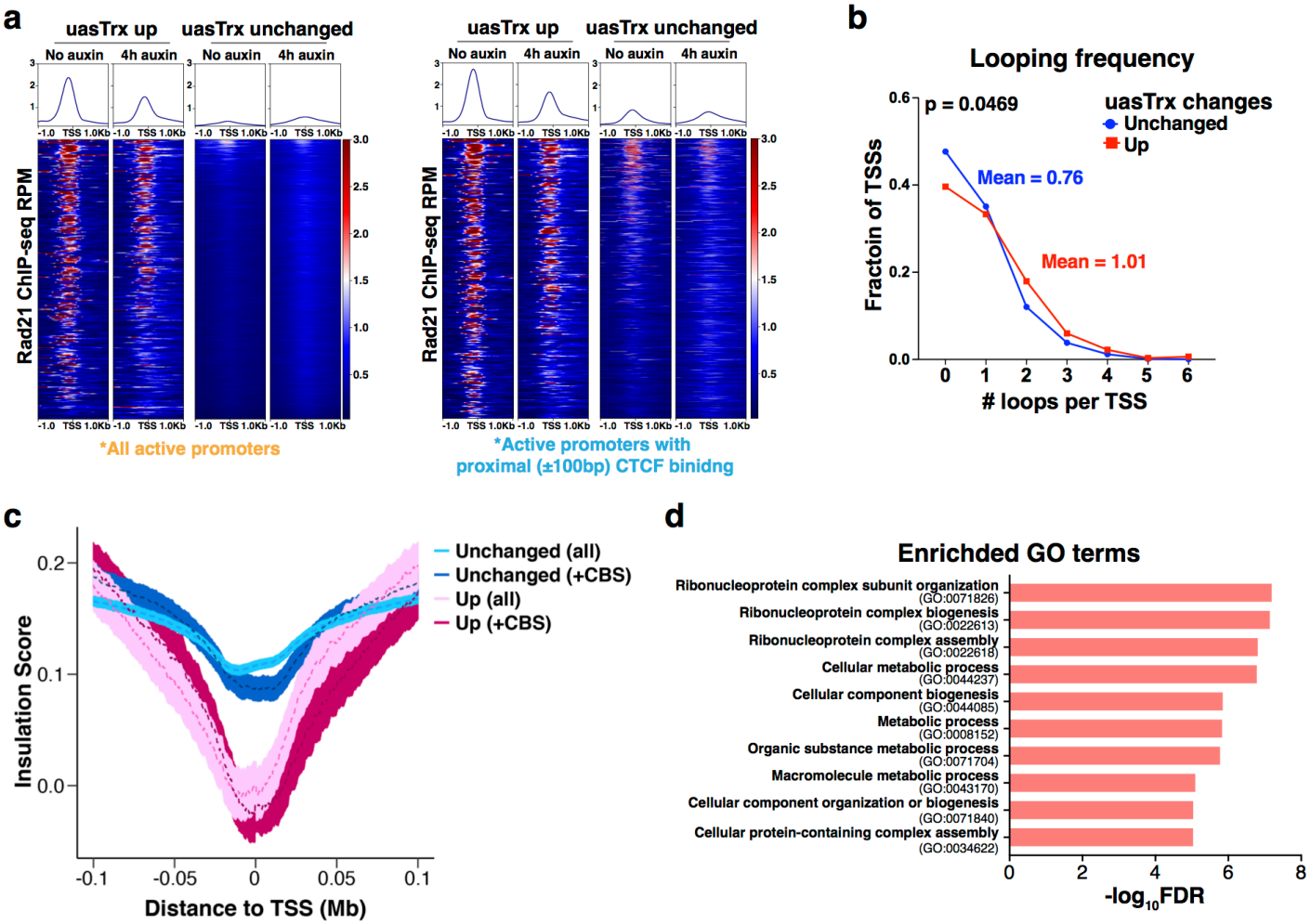




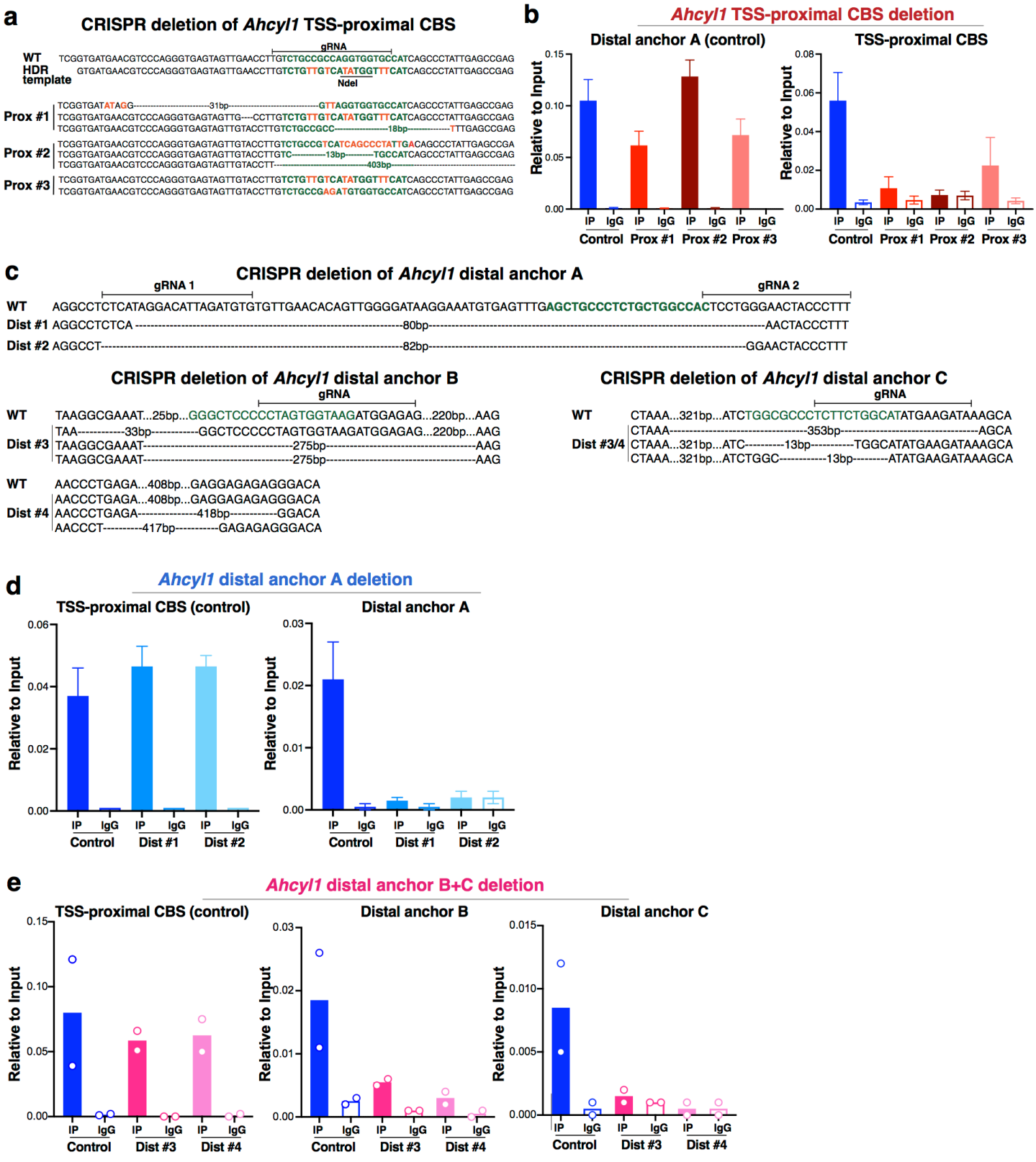
Extended Data Fig.3 | Transient CTCF depletion in human HCT-116 leads to similar antisense transcriptional changes. a, PRO-seq MA plot of control versus CTCF-depleted cells on the antisense strand (-1000bp to +200 relative to annotated TSS) in human HCT-116 cells. Differentially expressed transcripts highlighted in color. **b,** Row-linked heatmaps showing CTCF occupancy at active promoters, grouped by uasTrx changes, sorted by binding enrichment levels, and shown with respect to sense orientation. **c,** Transcriptional changes in uasTrx and GB after CTCF depletion. **d,** Boxplot showing log-transformed PRO-seq fold changes in GB. **e,** Scatterplot showing log-transformed PRO-seq fold changes in GB and uasTrx. **f,** Brower views of CTCF ChIP-seq (mm9 liftover from Rao et al., 2014) and PRO-seq signals at *Gstp1* and *Tap2* loci. Arrows highlight location of CTCF-repressed uasTrx. Arrow color indicates uasTrx strandedness. KD, knockdown.



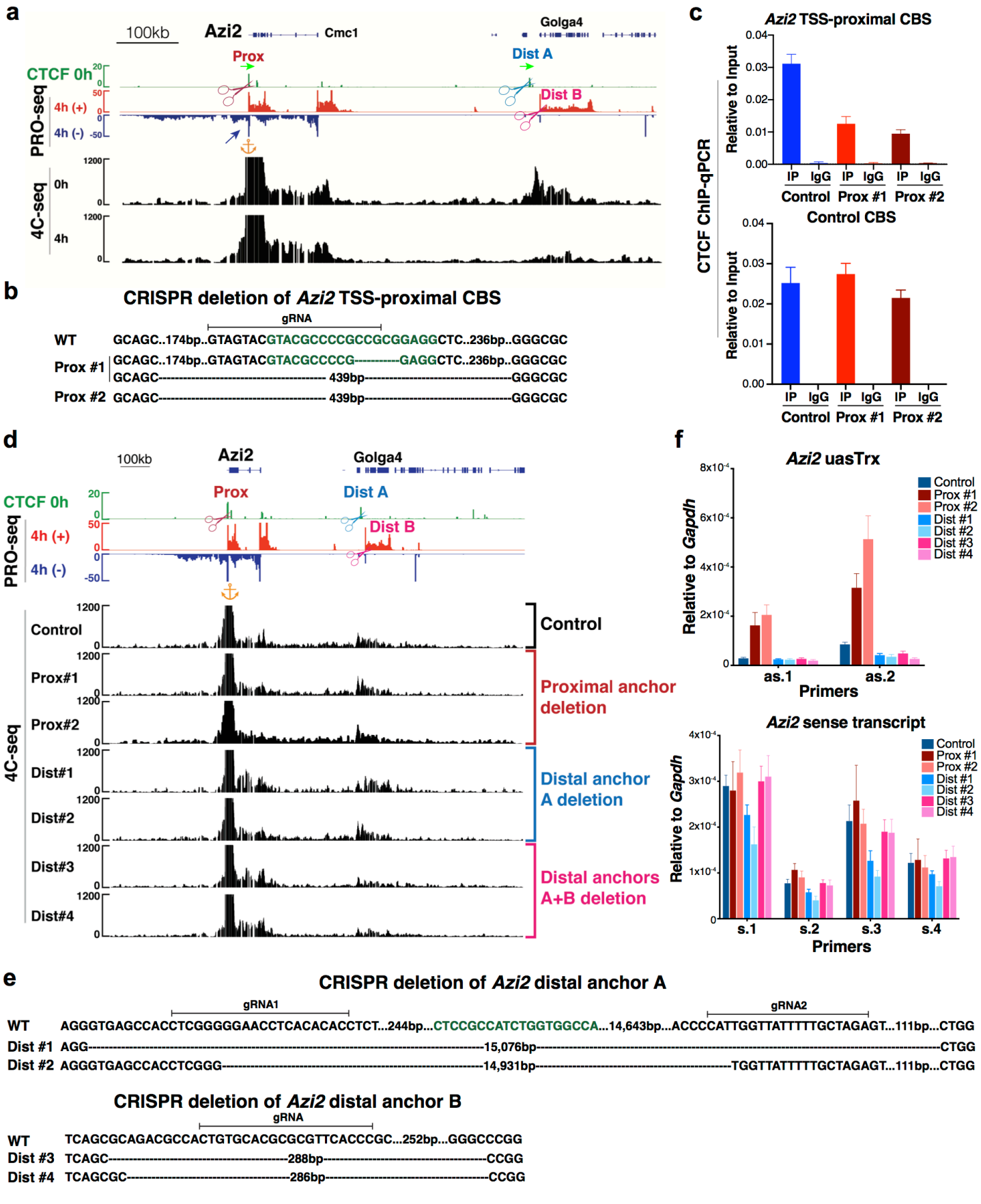
Extended Data Fig.4 | Transient CTCF depletion in mESC leads to similar antisense transcriptional changes. a, RNA-seq MA plot of control versus CTCF-depleted cells on the antisense strand (-1000bp to +200 relative to annotated TSS) in mESC. Differentially expressed transcripts highlighted in color. **b,** Brower views of CTCF ChIP-seq and RNA-seq signals at *Ahcyl1*, *Azi2* and *Rps3a1* loci. Arrows highlight signals upstream of TSS indicative of antisense transcription. **c,** Row-linked heatmap showing CTCF occupancy at active promoters, grouped by uasTrx changes and shown with respect to sense orientation. **d,** Transcriptional changes in uasTrx and GB after CTCF depletion. **e,** Correlation between uasTrx and GBs changes in RNA-seq upon CTCF depletion. *P* value was calculated by Spearman rank correlation test; r is the correlation coefficient. **f,** Log-transformed RNA-seq fold changes in GB after CTCF depletion over control. **g,** Log-transformed RNA-seq fold change in uasTrx in indicated conditions over control. Note the repression of elevated uasTrx after auxin washoff.



Extended Data Fig.5 | Affected promoters are associated with architectural features. a, Left, row-linked heatmaps showing Rad21 occupancy at all activated (n=376) and unaffected (n=9,632) active promoters, grouped by CTCF depletion-elicited uasTrx changes, sorted in the same order as **Fig. 1e,** and shown with respect to sense orientation. Right, same as left except only plotting those with proximal (±100bp) CTCF binding (up n=319; unchanged n=1,527). **b,** Distribution of looping frequencies of up-regulated versus unchanged uasTrx with proximal (±100bp) CTCF binding. *P* value calculated by Wilcoxon signed-rank test. **c,** Averaged insulation score centered at annotated TSS over 0.2Mb window, plotted with respect to sense orientation, and grouped by uasTrx changes and whether CTCF binds proximally (±100bp; "+CBS"). **d,** Gene ontology terms enriched at genes with activated uasTrx.

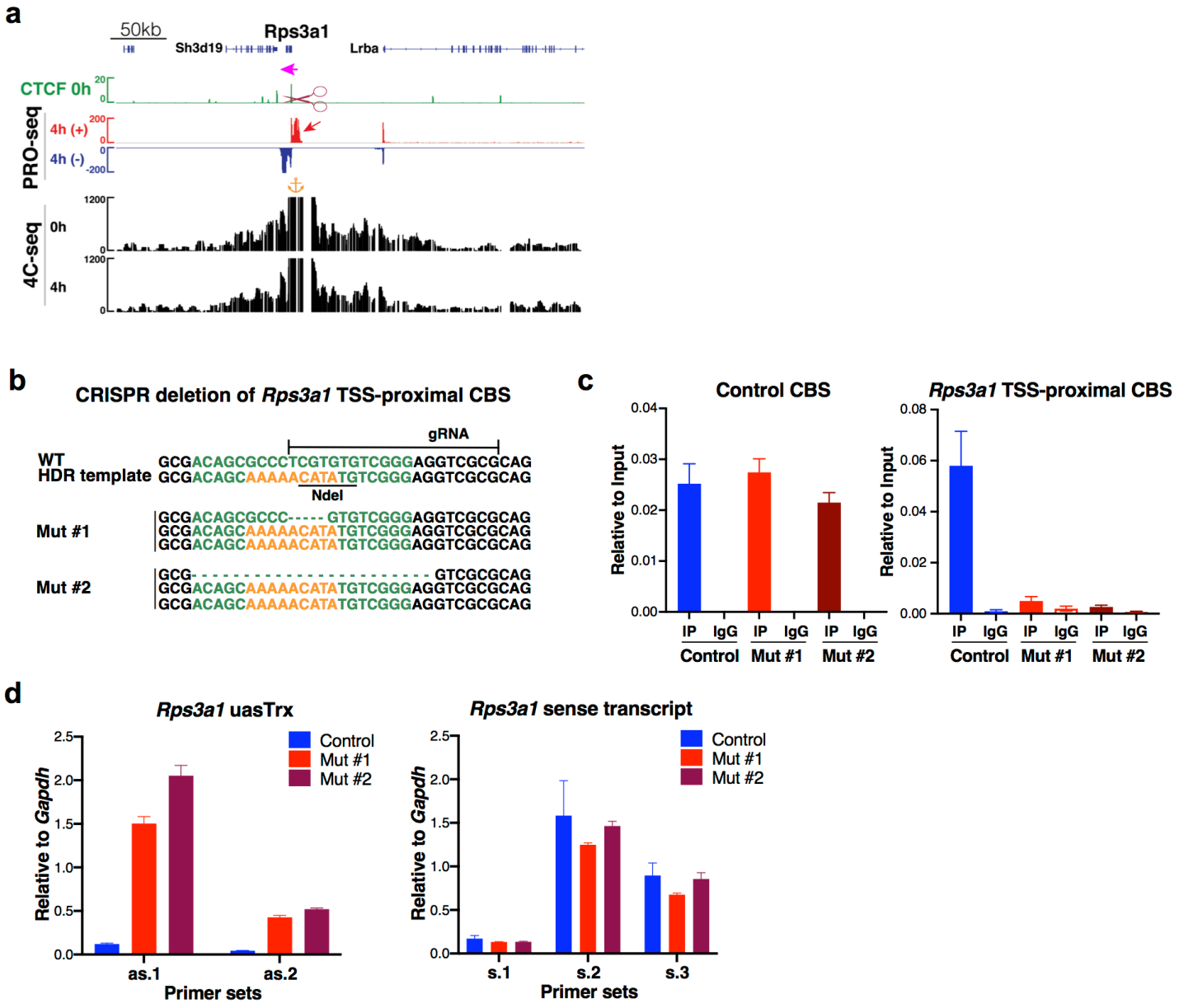


Extended Data Fig.6 | CRISPR/Cas9-mediated genome editing disrupts CTCF binding at *Ahcyl1*. a, Genotype of edited clones shown in **Fig. 2c**. Predicted CTCF motif highlighted in green. **b,** Left, CTCF ChIP-qPCR showing abrogation of CTCF binding at TSS-proximal CBS in mutants shown in **Fig. 2c**. Right, distal CBS served as a control for ChIP efficiency (error bar: SEM; n=3). **c,** Genotype of distally edited clones shown in **Fig. 2c**. Predicted CTCF motif highlighted in green. **d,** Left, TSS-proximal CBS served as a control for ChIP efficiency. Right, CTCF ChIP-qPCR showing abrogation of CTCF binding at distal anchor A in distal clones #1 and 2 shown in **Fig. 2c** (error bar: SEM; n=3). **e,** Left, TSS-proximal CBS served as a control for ChIP efficiency. Middle, CTCF ChIP-qPCR showing abrogation of CTCF binding at distal anchor B in distal clones #3 and 4 shown in **Fig. 2c** (n=2). Right, same as middle but measuring binding at distal anchor C.

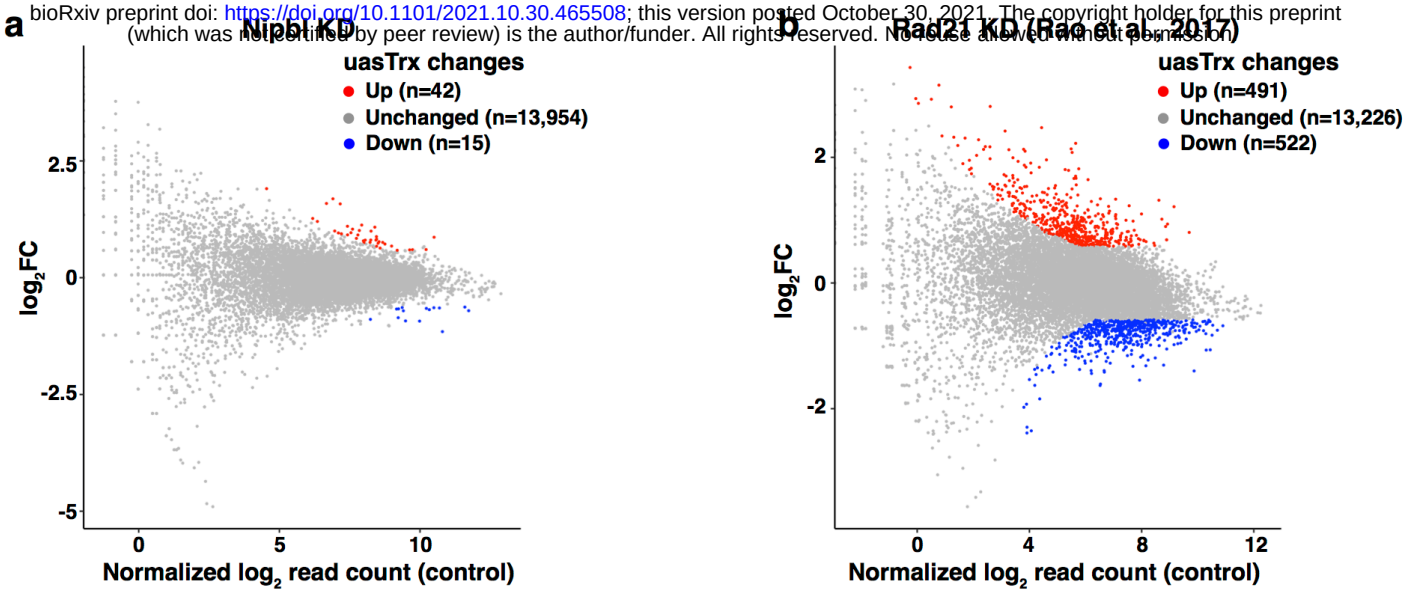


Extended Data Fig.7 | CRISPR/Cas9-mediated deletion of TSS-proximal CBS, but not distal loop anchors, at *Azi2* leads to uasTrx activation. a, Genome browser tracks of CTCF ChIP-seq, PRO-seq and 4C-seq at *Azi2* locus, with uasTrx highlighted by dark blue arrow and CRISPR targeted regions indicated by scissors. Green arrows indicate CTCF motif directionality. Orange anchor shows 4C-seq viewpoint. **b,** Genotype of TSS-proximally edited clones. Predicted CTCF motif highlighted in green. **c,** Top, ChIP-qPCR confirming disruption of CTCF binding in mutants (error bar: SEM; n=3). Bottom, ChIP-qPCR at an independent locus

controlling for ChIP efficiency (error bar: SEM; n=3). **d,** Left, representative 4C-seq profiles of control/mutant clones with edited regions indicated. Genome browser tracks of bulk CTCF ChIP-seq and PRO-seq shown on top. Similar observations were made in 2-3 independent 4C-seq experiments. Orange anchor indicates 4C-seq viewpoint. Scissor indicates edited region. Middle and right, RT-qPCR of nascent antisense and sense transcripts in WT/mutant clones (error bar: SEM; n=4). **e,** Genotype of mutants with distal anchor(s) disrupted. Predicted CTCF motif highlighted in green. **f,** RT-qPCR of *Azi2* uasTrx and sense primary transcripts in control and edited clones. Transcripts were normalized to *Gapdh* (error bar: SEM; n=3-4). Prox, TSS-proximal CBS. Dist, distal anchor.

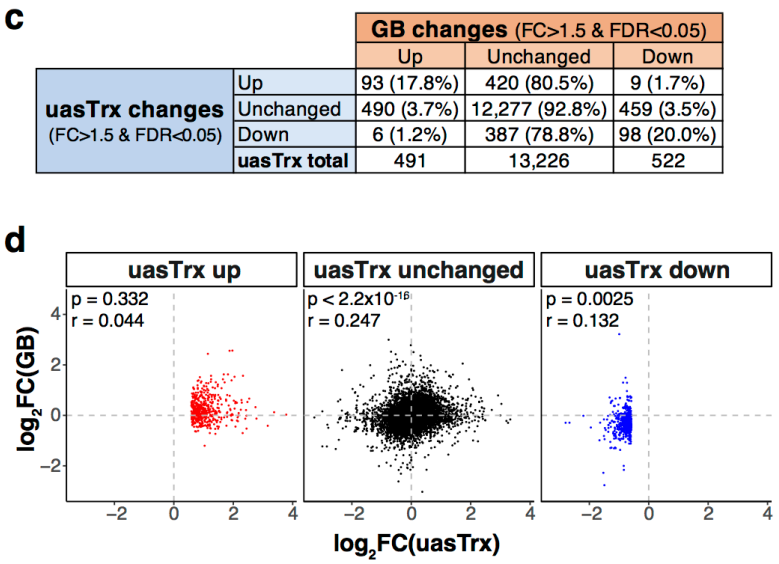


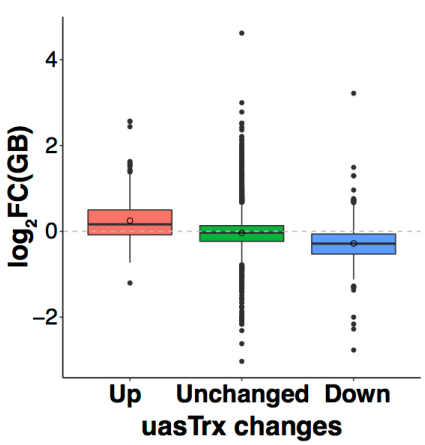
Extended Data Fig.8 | CRISPR/Cas9-mediated deletion of TSS-proximal CBS at *Rps3a1* leads to uasTrx up-regulation. a, Genome browser tracks of CTCF ChIP-seq, PRO-seq and 4C-seq at *Rps3a1* locus, with elevated uasTrx highlighted by red arrow and edited region indicated by scissor. Orange anchor indicates 4C-seq viewpoint. Magenta arrow above ChIP-seq track indicates CTCF motif directionality. **b,** Genotype of mutants after CRISPR/Cas9-mediated deletion of TSS-proximal CBS. Predicted CTCF motif highlighted in green. **c,** Left, ChIP-qPCR confirming disruption of CTCF binding in mutant clones (error bar: SEM; n=3). Right, ChIP-qPCR at an independent locus controlling for ChIP efficiency (error bar: SEM; n=3). **d,** RT-qPCR of nascent uasTrx and sense transcripts in control/mutant clones (error bar: SEM; n=3).

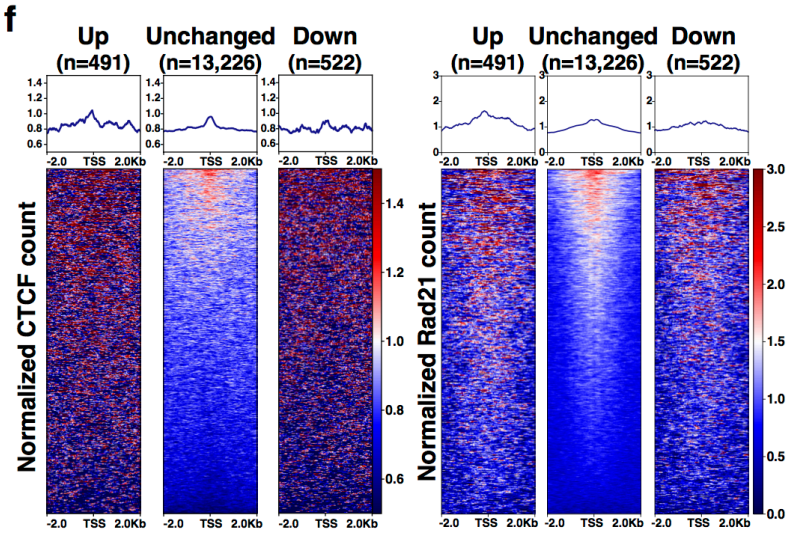


c

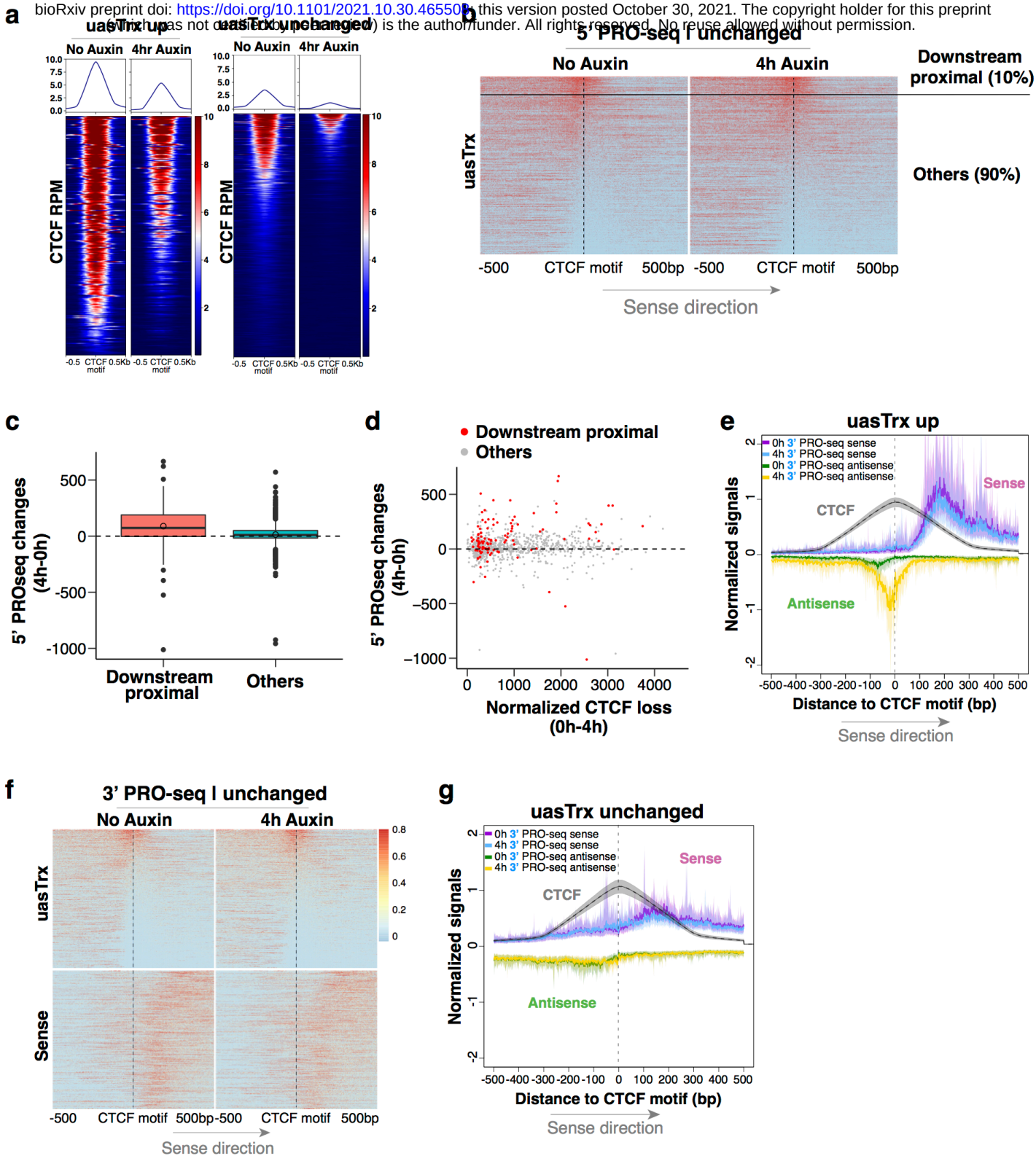
		GB changes (FC>1.5 & FDR<0.05)		
		Up	Unchanged	Down
uasTrx changes (FC>1.5 & FDR<0.05)	Up	93 (17.8%)	420 (80.5%)	9 (1.7%)
	Unchanged	490 (3.7%)	12,277 (92.8%)	459 (3.5%)
	Down	6 (1.2%)	387 (78.8%)	98 (20.0%)
	uasTrx total	491	13,226	522







Extended Data Fig.9 | Removal of chromatin-bound cohesin does not recapitulate CTCF-induced uasTrx changes. a, PRO-seq MA plot of control versus Nipbl-depleted cells on uasTrx expression (-1000bp to +200 relative to annotated TSS). Differentially expressed transcripts highlighted in color. **b,** Same as **a** but of Rad21-depleted cells. **c,** Table showing the number and percentage of uasTrx and GB changes after Rad21 depletion. **d,** Scatterplot comparing log-transformed 5' PRO-seq fold changes in uasTrx and GB. *P* value was calculated by Spearman rank correlation test; r is the correlation coefficient. **e,** Boxplot showing log-transformed PRO-seq fold changes in GBs after Rad21 depletion. **f,** Left, row-linked heatmap showing CTCF occupancy at active promoters, grouped by uasTrx changes after Rad21 depletion, sorted by occupancy levels, and shown with respect to sense orientation. Right, same as left, but plotting Rad21 occupancy. Note that neither CTCF nor Rad21 is enriched at up-regulated uasTrx.



Extended Data Fig.10 | CTCF inhibits antisense transcription initiation through precise positioning. a, Left, row-linked CTCF heatmap at affected active promoters that harbor proximal (±100bp) CTCF binding and high-confidence CTCF motif scores (>75), centered at CTCF motifs, grouped by mean signal densities over center 200bp, and shown with respect to sense orientation. Right, same as left but at unaffected active promoters meeting the same CTCF criteria. **b,** 5' PRO-seq heatmap at unchanged promoters shown in **Fig. 3d** with a portion of sites (10%; "downstream proximal") manually picked from the rest ("others"), which demonstrate similar CTCF distribution relative to 5' PRO-seq signals as **Fig. 3b**. **c,** Related to **b**, plotting PRO-seq changes in uasTrx at unaffected promoters, grouped based on CTCF positioning relative to 5' PRO-seq signals. **d,** Related to **b**, comparing uasTrx changes and CTCF binding loss at unaffected promoters, grouped based on CTCF positioning relative to 5' PRO-seq signals. **e,** Metaplot summarizing 3' PRO-seq and CTCF signals shown in **Fig. 3f**. Solid lines and shades show bootstrapped estimates of average signals and the 12.5/87.5 percentiles, respectively. **f,** Same as **Fig. 3f**, but at unaffected promoters. **g,** Same as **e**, but summarizing **f**.

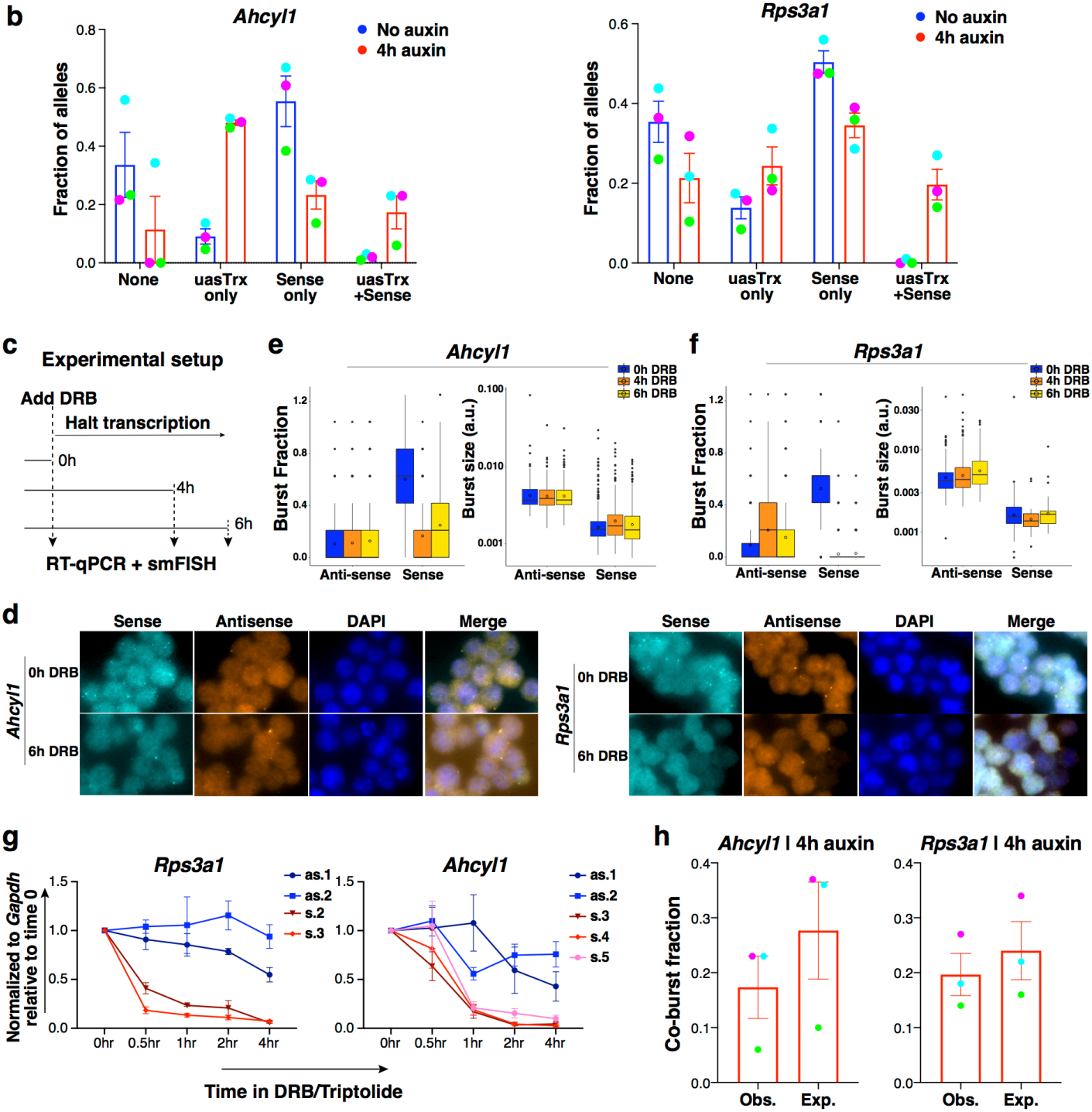
a

Ahcyl1 allele counts

None	Antisense only	Sense only	Both	Allele count	Condition	Replicate
283	116	878	33	1310	No auxin	1
391	229	1019	36	1675	No auxin	2
636	53	437	12	1138	No auxin	3
0	557	329	267	1153	4h auxin	1
0	665	372	304	1341	4h auxin	2
234	316	93	39	682	4h auxin	3

Rps3a1 allele counts

None	Antisense only	Sense only	Both	Allele number	Condition	Replicate
675	291	883	4	1853	No auxin	1
585	391	1258	12	2246	No auxin	2
625	120	677	4	1426	No auxin	3
365	209	412	161	1147	4h auxin	1
182	592	503	480	1757	4h auxin	2
357	347	640	298	1642	4h auxin	3



Extended Data Fig.11 | CTCF inhibits antisense burst fraction; sense/antisense co-bursting is disfavored. a, Table showing raw smFISH allele counts. **b,** Observed fractions of alleles where sense/antisense burst independently or simultaneously at *Ahcyl1* and *Rps3a1* (error bar: SEM; n=3). Biological replicates matched by dot colors. (**c,** Experimental outline for RNA half-life estimation. **d,** Representative smFISH images before and after DRB treatment at *Ahcyl1* and *Rps3a1*. **e,** Left, box plot showing antisense and sense burst fractions at *Ahcyl1* before and after DRB treatment. Right, same as left but quantifying burst sizes. n=3 biological replicates. *P* values were calculated by two-sample *t*-test. **f,** Same as (**e** but at *Rps3a1*. **g,** RT-qPCR measuring nascent sense and antisense transcript levels at *Ahcyl1* and *Rps3a1* before and after DRB treatment. Transcripts were normalized to *Gapdh* and plotted relative to time 0 (error bar: SEM; n=4). **h,** Predicted and observed co-burst fractions at *Ahcyl1* and *Rps3a1* after 4h auxin treatment (error bar: SEM; n=3). Biological replicates matched by dot colors.

Extended Data Table 1. Top gene ontology terms.

GO biological process	# Reference	# Dataset	Expected	Fold Enrichment	+/-	raw *P* value	FDR
Developmental process	5554	54	26.02	2.08	+	4.36E-09	6.86E-05
Negative regulation of biological process	5264	46	24.66	1.87	+	4.34E-06	1.37E-02
Regulation of multicellular organismal process	2787	31	13.06	2.37	+	3.55E-06	1.40E-02
Multicellular organism development	4788	44	22.43	1.96	+	1.96E-06	1.54E-02
Anatomical structure development	5166	46	24.2	1.9	+	3.37E-06	1.77E-02
System development	4176	39	19.56	1.99	+	7.42E-06	1.95E-02
Regulation of developmental process	2521	28	11.81	2.37	+	1.35E-05	2.36E-02
Cell differentiation	3560	35	16.68	2.1	+	1.06E-05	2.38E-02
Cellular developmental process	3611	35	16.92	2.07	+	1.31E-05	2.58E-02
Positive regulation of gene expression, epigenetic	29	4	0.14	29.44	+	1.64E-05	2.59E-02

Extended Data Table 2. Oligos for sgRNAs.

Name	**Sequence - Forward**	**Sequence - Reverse**
Ahcyl1 proximal CBS guide	CACCGAACCTTGTCTGCCGCCAGG	AAACCCTGGCGGCAGACAAGGTTC
Ahcyl1 distal anchor A guide 1	CACCGCACATCTAATGTCCTATGAG	AAACCTCATAGGACATTAGATGTGC
Ahcyl1 distal anchor A guide 2	CACCGAAAGGGTAGTTCCCAGGAG	AAACCTCCTGGGAACTACCCTTTC
Ahcyl1 distal anchor B guide	CACCGAAGTAAGGCGAAATAAGTGA	AAACTCACTTATTTCGCCTTACTTC
Ahcyl1 distal anchor C guide	CACCGTATCTTCATATGCCAGAAGA	AAACTCTTCTGGCATATGAAGATAC
Azi2 proximal CBS guide	CACCGTAGTACGTACGCCCCGCCG	AAACCGGCGGGGCGTACGTACTAC
Azi2 distal anchor A guide 1	CACCGAGGTGTGTGAGGTTCCCCCG	AAACCGGGGGAACCTCACACACCTC
Azi2 distal anchor A guide 2	CACCGTCTAGCAAAAATAACCAATG	AAACCATTGGTTATTTTTGCTAGAC
Rps3a1 proximal CBS guide	CACCGCGCGACCTCCCGACACACGA	AAACTCGTGTGTCGGGAGGTCGCGC
Rps3a1 proximal CBS ssODN	GCAGGCTCCCATGCACAGCCTAGCCAGCCCGCGGCCCGCTGCGCGACCTCCCGAC ATATGTTTTTGCTGTCGCAGCACGGCGGGAAACGAGCCGCCAGCTAGTGTCGCGA GATTCTTAAA	

Extended Data Table 3. RT-qPCR primers.

Target	Sequence - Forward	Sequence - Revere	Coordinates (mm9)
Ahcyl1.as1	CAGGGTGAGTAGTTGAACCTTG	AGACCCTTAGTCTTGTGGCATC	chr3:107499644-107499741
Ahcyl1.as2	TGGATTGATGAGGAGCTGAG	GTTAAATCCTGTGGCAGAGTCC	chr3:107526394-107526469
Ahcyl1.as3	CCACTGGACATGATGATAGGC	CCTTTGGCTTGAGTCTTTGC	chr3:107526049-107526144
Ahcyl1.s1	GCTCTGATTTCACTCAGGAAACG	GTGCGTACAGCCCACTATTTTA	chr3:107476944-107477018
Ahcyl1.s2	AATAGTGGGCTGTACGCACAT	GCAGCTACTTCATTCTGAGTTGA	chr3:107476371-107476962
Ahcyl1.s3	GCTTTCTGTCGAACCTTTGC	AGCCTTGGGGATTAACTGCT	chr3:107498014-107498211
Ahcyl1.s4	GCTTGTTGTGCTGGACTTGA	CCCCTCCAGGATTTGTTTTT	chr3:107493421-107493619
Ahcyl1.s5	CAGGACCTCTGGGAGATCAG	TTCCTAAAATTCGGCGTCAC	chr3:107480993-107481144
Azi2.as1	TCATCGGCTTCCTGGAATAG	ATGGCTCATGGTTCTGAAGG	chr9:117813405-117813511
Azi2.as2	TCACGGAATCCCAGTTGC	CGGAGTGGAGCTGAGACAG	chr9:117949338-117949447
Azi2.s1	AGGCACATGAGAAACACAGC	TATTCCTCACATGCCCACAC	chr9:117952938-117953015
Azi2.s2	CAGCGTGCTGTCTTCATTTG	CCAGAGGGATGGTTTTCAAAG	chr9:117955664-117955734
Azi2.s3	TGCTGTGTTGCCTCTGAAAG	TGTGAGCAGGGGAAGAAAAG	chr9:117956408-117956493
Azi2.s4	AAGCTAGCTGGCTGGTTTTG	TGAATTCCACGTAGCCTTGG	chr9:117954728-117954806
Gapdh	AGGTTGTCTCCTGCGACTTCA	CCAGGAAATGAGCTTGACAAAG	chr6:125112229-125112329
Rps3a1.as1	GCTGGCTAGGCTGTGCAT	CGGAAACCACAAGAAACCTG	chr3:85946693-85946767
Rps3a1.as2	CTCGTGTGTCGGGAGGTC	GAAGTGGGTTGAGCATCTCTG	chr3:85946659-85946748
Rps3a1.s1	CCGCGCGATCCGCCA	CAAGAACAAGCGCCTGACGA	chr3:85946410-85946480
Rps3a1.s2	AATGGATCGACCCTGGATGG	AAAGTGGTCGGGAGTGTTGTT	chr3:85945712-85945786
Rps3a1.s3	GAGCAAATACCCATCGGTCG	CCCCAAAACCATTATGAGCTG	chr3:85943120-85943212

Extended Data Table 4. ChIP-qPCR primers.

Target	**Sequence - Forward**	**Sequence - Revere**	**Coordinates (mm9)**
Ahcyl1 TSS-proximal CBS	TAAGGGTAGAGGGCGGAGAC	GGAGTCACACTCGGCTCAAT	chr3:107499485-107499713
Ahcyl1 distal anchor A	TTTGCCAACTGGTCCTTTTC	TCCAATTTATCCAGGCCAGA	chr3:107666932-107667188
Ahcyl1 distal anchor B	ACTTGGCTAAGCATGCTCCT	GCACATTCCCCAATTAATCC	chr3:107558030-107558123
Ahcyl1 distal anchor C	ATGGTTCACAGCCACTGCTT	TGAAGGAGCTTCCTCGGGTA	chr3:107721267-107721346
Control CBS	CCACACAGGCAGTCTTGAAA	GCAAGCCCTAACGCATAGAA	chr3:107670064-107670280
Azi2 TSS-proximal CBS	TCATGGGACCTGTAGTACGC	GAAGCCGGACCTGAAGACTA	chr9:117949568-117949644
Rps3a1 TSS-proximal CBS	TTTCTTGTGGTTTCCGTTGC	AAGCCCATGGTCTAGGGAAG	chr3:85946752-85946835

Extended Data Table 5. 4C primers.

Target	**Sequence - Reading**	**Sequence - Non-reading**
Ahcyl1 promoter	TTGATATGCCATCTTCCCGA	CACAGTTTCTGGATTCTACTGTGTA
Azi2 promoter	CATTTAAGACGATGGAGTGATC	ACAAAGTGAGACATCTTCAAGA
Rps3a1 promoter	GGTAGGGAGGCAGAAGATC	TGTCAGATACGGGTTTTCTC

Extended Data Table 6. smFISH primers.

Ahcyl1						
	Sequence	**Name**	**chr**	**strand**	**start_mm9**	**end_mm9**
Sense	cttcagtaagatccaggagg	mAhcyl1FirstIntron_1	chr3	+	107498845	107498864
	ttccaagagccatagaagga	mAhcyl1FirstIntron_2	chr3	+	107498275	107498294
	accctacaaatactgtaggc	mAhcyl1FirstIntron_3	chr3	+	107498072	107498091
	gatatttactgctgttggca	mAhcyl1FirstIntron_4	chr3	+	107497338	107497357
	accaacaagccctgtaatat	mAhcyl1FirstIntron_5	chr3	+	107496531	107496550
	cacttaggggataccaatt	mAhcyl1FirstIntron_6	chr3	+	107496467	107496486
	ggtacaatgtacacttctcc	mAhcyl1FirstIntron_7	chr3	+	107495612	107495631
	ggaggttaacaataccacca	mAhcyl1FirstIntron_8	chr3	+	107494811	107494830
	cacagttgctcaaagtctgg	mAhcyl1FirstIntron_9	chr3	+	107494736	107494755
	actgtgtgtcagaccttata	mAhcyl1FirstIntron_10	chr3	+	107494574	107494593
	gtaaggatagctctgagctg	mAhcyl1FirstIntron_11	chr3	+	107494281	107494300
	cctcaaacttgcaacaacct	mAhcyl1FirstIntron_12	chr3	+	107494144	107494163
	cagagccctaacatacattc	mAhcyl1FirstIntron_13	chr3	+	107494095	107494114
	tcccagctggcaaataagaa	mAhcyl1FirstIntron_14	chr3	+	107493335	107493354
	ctcatattccccaaatagga	mAhcyl1FirstIntron_15	chr3	+	107493137	107493156
	aaggaacatttgggctgtgg	mAhcyl1FirstIntron_16	chr3	+	107491725	107491744
	atgggcatggaaagttctca	mAhcyl1FirstIntron_17	chr3	+	107491665	107491684
	ggtacatgatcatatcctct	mAhcyl1FirstIntron_18	chr3	+	107491433	107491452
	cttttggggagtactttctg	mAhcyl1FirstIntron_19	chr3	+	107490921	107490940
	aactaggtggggaaagcagt	mAhcyl1FirstIntron_20	chr3	+	107490693	107490712
	caagaaagccgggaaggact	mAhcyl1FirstIntron_21	chr3	+	107489850	107489869
	ccaagctagcttactgtatt	mAhcyl1FirstIntron_22	chr3	+	107488981	107489000
	aagtagtgtttcctggaagc	mAhcyl1FirstIntron_23	chr3	+	107488525	107488544
	ctcagcactagtaactgtcg	mAhcyl1FirstIntron_24	chr3	+	107487251	107487270
	tccactctcagattaacagc	mAhcyl1FirstIntron_25	chr3	+	107484728	107484747
	aagtaccccaagtacaactc	mAhcyl1FirstIntron_26	chr3	+	107484705	107484724
	ttactactatgtgcagtgct	mAhcyl1FirstIntron_27	chr3	+	107484125	107484144
	atatttcttcagcaaccgga	mAhcyl1FirstIntron_28	chr3	+	107484003	107484022
	agtcaagagttccttagtgg	mAhcyl1FirstIntron_29	chr3	+	107483722	107483741
	agaaagggagagcctgtttt	mAhcyl1FirstIntron_30	chr3	+	107483258	107483277
	agacagaaacctgcgtgttt	mAhcyl1FirstIntron_31	chr3	+	107483179	107483198
	tctggaatcaatcggcagtt	mAhcyl1FirstIntron_32	chr3	+	107483017	107483036
Antisense	cggcagacaaggttcaacta	mAhcyl1Antisense_1	chr3	-	107499653	107499672
	gtgtagaaaccagactgctc	mAhcyl1Antisense_2	chr3	-	107500001	107500020
	ttgtaaacagacaaggccca	mAhcyl1Antisense_3	chr3	-	107500053	107500072
	ctacactcactgagagtgga	mAhcyl1Antisense_4	chr3	-	107500226	107500245
	cagaaactgtgcgagtccaa	mAhcyl1Antisense_5	chr3	-	107500261	107500280
	cttattctcagcatagtggg	mAhcyl1Antisense_6	chr3	-	107500931	107500950
	catctcctgcaagagcaaac	mAhcyl1Antisense_7	chr3	-	107501890	107501909
	atagctttgtcacgggattg	mAhcyl1Antisense_8	chr3	-	107503157	107503176
	gagacctacagaacctgtac	mAhcyl1Antisense_9	chr3	-	107503532	107503551
	atgtcaatggtgcctctaag	mAhcyl1Antisense_10	chr3	-	107504416	107504435
	ggaagaaccagcataagggg	mAhcyl1Antisense_11	chr3	-	107504539	107504558
	aaggcagagaggtccttaat	mAhcyl1Antisense_12	chr3	-	107505133	107505152
	ttaagtcatcaaagccctcg	mAhcyl1Antisense_13	chr3	-	107505414	107505433
	ctttaaaacctccagttgct	mAhcyl1Antisense_14	chr3	-	107505525	107505544
	gctgtggaaaattgcagctt	mAhcyl1Antisense_15	chr3	-	107506601	107506620

	agagggtaaggaaggctgac	mAhcyl1Antisense_16	chr3	-	107506741	107506760
	ttggaaatccagactggcat	mAhcyl1Antisense_17	chr3	-	107506898	107506917
	agtaggcctgcagaaaatgg	mAhcyl1Antisense_18	chr3	-	107507266	107507285
	gaggacactgtgtcttcata	mAhcyl1Antisense_19	chr3	-	107507362	107507381
	attccagactatttcaagcc	mAhcyl1Antisense_20	chr3	-	107507508	107507527
	cctctatttctctcatggaa	mAhcyl1Antisense_21	chr3	-	107507722	107507741
	gcattttcaagtccaactca	mAhcyl1Antisense_22	chr3	-	107507783	107507802
	agacagacacatgtccacag	mAhcyl1Antisense_23	chr3	-	107507935	107507954
	gtcaaccatatggcttaagc	mAhcyl1Antisense_24	chr3	-	107508152	107508171
	actgacgtgatggtcatcta	mAhcyl1Antisense_25	chr3	-	107508238	107508257
	taatttctttcctgccaact	mAhcyl1Antisense_26	chr3	-	107508300	107508319
	gacactgagactcagatggg	mAhcyl1Antisense_27	chr3	-	107509199	107509218
	ggaagttggtttacgtaggg	mAhcyl1Antisense_28	chr3	-	107509237	107509256
	gtatgagctgtaatctcacc	mAhcyl1Antisense_29	chr3	-	107509648	107509667
	agcagcactgtatattcagt	mAhcyl1Antisense_30	chr3	-	107509786	107509805
	acattttcctacagcttcat	mAhcyl1Antisense_31	chr3	-	107510310	107510329
	aatagtcttcccagatgact	mAhcyl1Antisense_32	chr3	-	107510361	107510380

Rps3a1

	Sequence	**Name**	**chr**	**strand**	**start**	**end**
Sense	gcggtgttaaaaagtgccac	mRps3a1Introns_1	chr3	+	85946143	85946162
	ccagcacacaagttgacatg	mRps3a1Introns_2	chr3	+	85946086	85946105
	catcacacaggcagtactat	mRps3a1Introns_3	chr3	+	85945958	85945977
	tgcacttaacacgacggacg	mRps3a1Introns_4	chr3	+	85945850	85945869
	ttcttatccttcagacttct	mRps3a1Introns_5	chr3	+	85945794	85945813
	taatatgcacatgccacagc	mRps3a1Introns_6	chr3	+	85945565	85945584
	cactgggctgacatttcaat	mRps3a1Introns_7	chr3	+	85945511	85945530
	ctgccacatctgtactaaga	mRps3a1Introns_8	chr3	+	85945461	85945480
	tactgccaacacctaaacgt	mRps3a1Introns_9	chr3	+	85945410	85945429
	ctccattccatgttttgata	mRps3a1Introns_10	chr3	+	85945063	85945082
	gtggccgtaacaagcaatac	mRps3a1Introns_11	chr3	+	85944926	85944945
	tagagcacctttttctctatc	mRps3a1Introns_12	chr3	+	85944843	85944862
	tccactgtcagacaactgag	mRps3a1Introns_13	chr3	+	85944814	85944833
	aaggacagaaggggaccgaa	mRps3a1Introns_14	chr3	+	85944777	85944796
	atttccctacatgtcatttc	mRps3a1Introns_15	chr3	+	85944694	85944713
	tgaacagctttcccttgaat	mRps3a1Introns_16	chr3	+	85944642	85944661
	tcatcactcattcccatata	mRps3a1Introns_17	chr3	+	85944599	85944618
	ccggaaacaaaatcctggca	mRps3a1Introns_18	chr3	+	85944440	85944459
	gaccatttttgtggcaaagc	mRps3a1Introns_19	chr3	+	85944302	85944321
	atcctttcttgaaagagggc	mRps3a1Introns_20	chr3	+	85944138	85944157
	gctgtacatgacagatgtct	mRps3a1Introns_21	chr3	+	85943866	85943885
	tagtttcagtactccaagca	mRps3a1Introns_22	chr3	+	85943784	85943803
	atttatgtcaacatccctca	mRps3a1Introns_23	chr3	+	85943458	85943477
	cgtttatggcacttccttaa	mRps3a1Introns_24	chr3	+	85943363	85943382
	tcaagcccacgtgacaaata	mRps3a1Introns_25	chr3	+	85943340	85943359
	agctcataatggttttgggg	mRps3a1Introns_26	chr3	+	85943193	85943212
	aactccacctcttctttaac	mRps3a1Introns_27	chr3	+	85942881	85942900
	gtaatgttatcagtggccat	mRps3a1Introns_28	chr3	+	85942727	85942746
	taaccacctgtatactgcat	mRps3a1Introns_29	chr3	+	85942690	85942709
	ttccttatgctttttgtct	mRps3a1Introns_30	chr3	+	85942662	85942681
	aaaaccacgtacttggttcc	mRps3a1Introns_31	chr3	+	85942347	85942366
	aacctaaactacagcccgaa	mRps3a1Introns_32	chr3	+	85942076	85942095

Antisense	catctctgacaatctgcagg	mRps3a1asHT1_1	chr3	-	85946716	85946735
	ccacaagaaacctgaagtgg	mRps3a1asHT1_2	chr3	-	85946742	85946761
	gagtagtcagcacgtgaaca	mRps3a1asHT2_1	chr3	-	85946995	85947014
	agttcacaacacagccaagc	mRps3a1asHT2_2	chr3	-	85947019	85947038
	gctaagcctctttttcatg	mRps3a1asHT2_3	chr3	-	85947041	85947060
	caaagaaaagccccgagggc	mRps3a1asHT3_1	chr3	-	85949274	85949293
	actaccaagaagagcctgag	mRps3a1asHT3_2	chr3	-	85949296	85949315
	gagggtttggtgggacttaa	mRps3a1asHT3_3	chr3	-	85949318	85949337
	atggaagctatgctgtagct	mRps3a1asHT3_4	chr3	-	85949340	85949359
	ttaagagttaaaggcttgcc	mRps3a1asHT3_5	chr3	-	85949362	85949381
	ggtaggaggactaggaggga	mRps3a1asHT3_6	chr3	-	85949384	85949403
	agggagattggctcagaggg	mRps3a1asHT3_7	chr3	-	85949408	85949427
	ttttgggtgggactaaaggg	mRps3a1asHT3_8	chr3	-	85949430	85949449
	atttgtggagttactgagga	mRps3a1asHT3_9	chr3	-	85949452	85949471
	ggagacatgacagacaactg	mRps3a1asHT3_10	chr3	-	85949474	85949493
	atagagtcacacagtaagga	mRps3a1asHT3_11	chr3	-	85949496	85949515
	gaggatacacttgggtagtt	mRps3a1asHT3_12	chr3	-	85949520	85949539
	caggcctcatcttaacgatc	mRps3a1asHT3_13	chr3	-	85949543	85949562
	atctctatcacttacttaca	mRps3a1asHT3_14	chr3	-	85949566	85949585
	gatctattttcagattacct	mRps3a1asHT3_15	chr3	-	85949663	85949682
	gtttctgtcttctgaattgt	mRps3a1asHT3_16	chr3	-	85949697	85949716
	tacaagggattctacgccac	mRps3a1asHT4_1	chr3	-	85953429	85953448
	agattggacctacggtcttg	mRps3a1asHT4_2	chr3	-	85953458	85953477
	cagtgcaatgttgctgtgcg	mRps3a1asHT4_3	chr3	-	85953480	85953499
	tctgagcactctcagctaag	mRps3a1asHT4_4	chr3	-	85953508	85953527
	atggtggagaacaacagtga	mRps3a1asHT4_5	chr3	-	85953535	85953554
	attattttgtaggagtccag	mRps3a1asHT4_6	chr3	-	85953557	85953576
	cttcaggggaatgtaagtaa	mRps3a1asHT4_7	chr3	-	85953598	85953617
	tgtaagcaagggaatcgtat	mRps3a1asHT4_8	chr3	-	85953620	85953639
	cagctaaggagggtgaggga	mRps3a1asHT4_9	chr3	-	85953642	85953661
	ggaccggcttcatattgaaa	mRps3a1asHT4_10	chr3	-	85953699	85953718
	gctaacaaggtcttcgggaa	mRps3a1asHT4_11	chr3	-	85953749	85953768
	agatgggaaaaaggccacgg	mRps3a1asHT4_12	chr3	-	85953825	85953844
	ctatgttctgagagatgggg	mRps3a1asHT4_13	chr3	-	85953848	85953867
	ctgatgtcttctaaccttct	mRps3a1asHT4_14	chr3	-	85953899	85953918
	ctgggctgattcacagataa	mRps3a1asHT4_15	chr3	-	85953926	85953945
	gctcctaagattattctgat	mRps3a1asHT4_16	chr3	-	85953948	85953967
	cttaatacttgttgtcttgt	mRps3a1asHT4_17	chr3	-	85953970	85953989
	aactacacttttatcagggg	mRps3a1asHT4_18	chr3	-	85954093	85954112