TET-catalyzed oxidation of intragenic 5-methylcytosine regulates CTCF-dependent alternative splicing

Ryan J Marina[†], David Sturgill[†], Marc A Bailly^{†,‡}, Morgan Thenoz[†], Garima Varma[#], Maria F Prigge, Kyster K Nanan, Sanjeev Shukla[§], Nazmul Haque[¶] & Shalini Oberdoerffer^{*}

Abstract

Intragenic 5-methylcytosine and CTCF mediate opposing effects on pre-mRNA splicing: CTCF promotes inclusion of weak upstream exons through RNA polymerase II pausing, whereas 5-methylcytosine evicts CTCF, leading to exon exclusion. However, the mechanisms governing dynamic DNA methylation at CTCF-binding sites were unclear. Here, we reveal the methylcytosine dioxygenases TET1 and TET2 as active regulators of CTCF-mediated alternative splicing through conversion of 5-methylcytosine to its oxidation derivatives. 5-hydroxymethylcytosine and 5-carboxymethylcytosine are enriched at an intragenic CTCF-binding sites in the *CD45* model gene and are associated with alternative exon inclusion. Reduced TET levels culminate in increased 5-methylcytosine, resulting in CTCF eviction and exon exclusion. *In vitro* analyses establish the oxidation derivatives are not sufficient to stimulate splicing, but efficiently promote CTCF association. We further show genome-wide that reciprocal exchange of 5-hydroxymethylcytosine and 5-methylcytosine at downstream CTCF-binding sites is a general feature of alternative splicing in naïve and activated CD4⁺ T cells. These findings significantly expand our current concept of the pre-mRNA “splicing code” to include dynamic intragenic DNA methylation catalyzed by the TET proteins.

Keywords alternative splicing; CTCF; DNA methylation; TET1; TET2

Subject Categories Chromatin, Epigenetics, Genomics & Functional Genomics; RNA Biology; Transcription

DOI 10.15252/embj.201593235 | Received 9 October 2015 | Revised 12 November 2015 | Accepted 25 November 2015 | Published online 28 December 2015

The EMBO Journal (2016) 35: 335–355

Introduction

Genes in higher eukaryotes are characterized by numerous short coding exons interrupted by relatively long non-coding introns. As genes are transcribed by RNA polymerase II (pol II), introns are excised by the spliceosomal complex, which recognizes short consensus sequences at exon–intron boundaries (reviewed in Black (2003)). Exon–intron architecture serves as a platform for transcriptome diversification through alternative pre-mRNA splicing and greater than 95% of human genes produce alternative transcripts (Pan *et al*, 2008; Wang *et al*, 2008). To minimize potentially deleterious aberrant splicing events, spliceosome assembly is coordinated at many levels. In addition to a vast network of RNA-binding proteins that recognize *cis*-elements encoded within nascent transcripts (reviewed in Matlin *et al* (2005)), co-transcriptional assembly of splicing factors at newly synthesized splice sites promotes appropriate ligation of contiguous exons (Pandya-Jones & Black, 2009; Tilgner *et al*, 2012; Bentley, 2014). Co-transcriptional splicing further allows for kinetic regulation of alternative splicing, wherein variations in the pol II elongation rate can shift the co-temporaneous availability of competing splice sites for regulatory factor binding (de la Mata *et al*, 2003; Munoz *et al*, 2009; Close *et al*, 2012; Dujardin *et al*, 2014).

An emerging concept in co-transcriptional splicing is the potential for extensive cross talk between the splicing machinery and the chromatin structure of the transcribed DNA template (reviewed in Haque & Oberdoerffer (2014)). Exonic DNA presents a distinct chromatin landscape characterized by increased nucleosome occupancy, elevated DNA methylation, and specific histone modifications relative to intronic DNA, raising the possibility that chromatin may poise exons for spliceosome recognition (Andersson *et al*, 2009; Hodges *et al*, 2009; Kolasinska-Zwierz *et al*, 2009; Lister *et al*, 2009; Schwartz *et al*, 2009; Spies *et al*, 2009; Tilgner *et al*, 2009). Chromatin structure has been shown to impact splicing decisions through modulation of pol II elongation and through

Center for Cancer Research, Laboratory of Receptor Biology and Gene Expression, National Cancer Institute, Bethesda, MD, USA

^{*}Corresponding author. Tel: +1 301 594 8151; Fax: +1 301 496 4951; E-mail: shalini.oberdoerffer@nih.gov

[†]These authors contributed equally to this work

[‡]Present address: Merck Research Laboratory, Palo Alto, CA, USA

[#]Present address: Department of Pathology, Anatomy and Cell Biology, Thomas Jefferson University, Philadelphia, PA, USA

[§]Present address: Department of Biological Sciences, Indian Institute of Science Education and Research (IISER), Bhopal, India

[¶]Present address: National Heart, Lung and Blood Institute, Bethesda, MD, USA

recruitment of RNA-binding proteins to their sites of action through interaction with chromatin-binding proteins (Sims *et al*, 2007; Andersson *et al*, 2009; Chodavarapu *et al*, 2010; Luco *et al*, 2010; Churchman & Weissman, 2011; Pradeepa *et al*, 2012; Kwak *et al*, 2013; Weber *et al*, 2014). Accordingly, efforts to modulate chromatin structure have revealed global alterations in splicing patterns, thus broadening the pre-mRNA splicing “code” to include intragenic chromatin structure (Schor *et al*, 2009, 2013; Luco *et al*, 2010; Hnilicova *et al*, 2011; Saint-Andre *et al*, 2011; Ameyar-Zazoua *et al*, 2012; Patrick *et al*, 2015).

Of the described intragenic chromatin features, DNA methylation shows particularly robust partitioning to exons (Lister *et al*, 2009; Feng *et al*, 2010; Zemach *et al*, 2010), though the significance to gene expression remains unclear. In mammalian genomes, methylation primarily occurs in a symmetric context on CpG dinucleotides (Lister *et al*, 2009). Once established by the *de novo* DNA methyltransferases DNMT3a/b in early embryogenesis, methylation patterns are preserved by the DNMT1 enzyme, which recognizes hemimethylated DNA and ensures site-specific propagation in the newly synthesized strand (Li & Zhang, 2014). Intriguingly, while promoter methylation is associated with transcriptional silencing, DNA methylation is globally enriched within gene bodies, where it is positively associated with active transcription (Klose & Bird, 2006; Lister *et al*, 2009; Wu *et al*, 2011). Notably, exons that are included in spliced mRNA show a higher level of methylation than their excluded counterparts (Choi, 2010; Maunakea *et al*, 2013). Conversely, DNA methylation is depleted at introns, intronless genes, and pseudoeons (Lyko *et al*, 2010; Gelfman *et al*, 2013). Together, these findings suggest that DNA methylation plays a fundamental role in “marking” exonic DNA for recognition by the spliceosome. However, experimental modulation of DNA methylation results in a limited number of alternative splicing events, with no clear directional bias to increased or decreased inclusion, thereby challenging a direct role for methylation in spliceosome recruitment (Li-Byarlay *et al*, 2013; Maunakea *et al*, 2013; Yearim *et al*, 2015). We recently established a context-dependent association between intragenic DNA methylation and alternative pre-mRNA splicing that is achieved through variable binding of the methyl-sensitive zinc-finger protein CTCF-binding factor (CTCF). Binding of CTCF to intragenic DNA promotes local pol II pausing and favors inclusion of weak upstream exons in spliced mRNA through kinetic regulation. In contrast, overlapping 5-methylcytosine (5mC) evicts CTCF, leading to loss of pol II accumulation and consequent exon exclusion (Shukla *et al*, 2011). These studies begin to uncover a dual role for DNA methylation in both the initiation and downstream processing of pol II gene products through variable methylation at promoters and gene bodies, respectively.

While widespread changes in methylation are typically limited to undifferentiated cell types, methylome analysis across diverse tissues identified increased differential methylation at CpG-dense intragenic regions in peripheral immune cells (Deaton *et al*, 2011). To query the underlying principles that define methyl-sensitive alternative exons, here we examine the cause and consequence of dynamic gene-body methylation in peripheral lymphocytes. Through detailed examination of a CTCF-dependent methyl-sensitive exon in the immune-specific *CD45* model gene, we reveal that the alpha-ketoglutarate-dependent TET (ten-eleven

translocation) proteins, TET1 and TET2, positively regulate CTCF-dependent exon inclusion in naïve lymphocytes through successive oxidation of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine and its downstream derivatives (5hmC or collectively 5oxiC) (Tahiliani *et al*, 2009; Ito *et al*, 2011). We further show that alternative splicing events in naïve versus activated CD4⁺ T cells are globally enriched for differential 5mC and 5oxiC at proximal CTCF-binding sites. Markedly, 5mC and 5hmC display overlapping distribution patterns within the intragenic landscape, wherein both modifications of cytosine are enriched at exons relative to introns (Ficz *et al*, 2011; Pastor *et al*, 2011; Khare *et al*, 2012). We thus conceive that similar to the role for TET proteins in modulating gene expression from promoters, alterations in the intragenic distribution of 5oxiC play a fundamental role in regulating pre-mRNA processing through variable association of methyl-sensitive DNA-binding proteins, as we show here for CTCF.

Results

Detection of 5-hydroxymethylcytosine at *CD45* exon 5 in CTCF-binding cell types

To gain insight into methyl-sensitive alternative splicing, we examined the mechanisms that support differential methylation at an established sensitive exon. Variable exclusion of exons 4–6 from transcripts encoding the protein tyrosine phosphatase *CD45* (*PTPRC*) is a robust model for developmentally regulated alternative splicing in the immune system. *CD45* initiates signaling through antigen receptors by dephosphorylating the inhibitory tyrosine on Src family kinases, and responds to activation through stepwise exclusion of the tandem variable exons (Martinez & Lynch, 2013). In general, the larger exon-4-containing isoforms (CD45RA⁺, or B220⁺ in B cells) are expressed early in peripheral lymphocyte development, whereas the shortest isoform that excludes all three variable exons (CD45RO) is found on terminally differentiated lymphocytes (Hermiston *et al*, 2003). We and others formerly established that exclusion of exons 4 and 6 from *CD45* transcripts (the A and C exons, respectively) results from activation-induced upregulation of the splicing repressor, hnRNPL (Oberdoerffer *et al*, 2008; Topp *et al*, 2008; Motta-Mena *et al*, 2010). In contrast, exclusion of exon 5 (the B exon) from *CD45* mRNA is achieved through loss of CTCF-dependent pol II pausing in the terminal stages of lymphocyte development due to the emergence of 5mC at exon 5 DNA (Shukla *et al*, 2011) (Appendix Fig S1).

To address the mechanisms leading to increased 5mC at *CD45* exon 5 DNA in differentiated lymphocytes, we established the methyl composition at base pair resolution in previously characterized B-cell lines and primary T cells (Fig 1A) (Oberdoerffer *et al*, 2008; Shukla *et al*, 2011). *CD45* exon 5 contains 9 CpGs that can be symmetrically methylated. As expected based on our previous findings, bisulfite sequencing in exon 5-excluding BL41 RB low (BL-E5 (–)) B cells and activated CD4⁺ T cells, which fail to bind CTCF, showed uniform methylation at all CpGs (Fig 1B). Similarly, exon 5-including BJAB B cells, which bind CTCF at exon 5 DNA, were unmethylated. However, exon 5-including BL41 RB high (BL-E5(+)) B cells and naïve peripheral T cells showed uniform methylation including within the CTCF-binding site (Fig 1B). This was surprising

given the demonstration that overlapping 5-methylcytosine ablates CTCF binding (Bell & Felsenfeld, 2000), yet both these cell types were previously determined to bind CTCF at exon 5 DNA (Shukla et al, 2011). Cytosines outside of a CpG context were fully detected as thymines, indicating that persistence of cytosine was not due to incomplete bisulfite conversion (Fig 1B, inset). CTCF-binding site methylation in exon 5-including BL-E5(+) cells was further confirmed through Southern blot using a methyl-sensitive restriction enzyme (BsaHI). In agreement with the bisulfite sequencing results, hybridization to a 5' radiolabeled probe showed that the CTCF-binding site is unmethylated in BJAB cells (3.8 kb), whereas it is methylated in BL-E5(+) and BL-E5(−) cells (7.6 kb) (Fig 1C).

Mechanisms to assess methylation, including bisulfite sequencing and most restriction digests, often fail to discriminate between 5mC and the closely related oxidation product 5hmC. We thus asked whether overlapping detection of methylation and CTCF in exon 5-including cells reflected the presence of 5hmC. To this end, we performed methylated DNA immunoprecipitation (MedIP) with antibodies specific to 5mC or 5hmC in BL-E5(+) and BL-E5(−) isogenic B cells. MedIP confirmed enriched 5mC at *CD45* exon 5 DNA in BL-E5(−) relative to BL-E5(+) cells (Fig 1D, left) and further revealed reciprocally elevated 5hmC at exon 5 in BL-E5(+) relative to BL-E5(−) cells (Fig 1D, right). Methylation did not substantially deviate at the contiguous alternative exons, which possess fewer

Figure 1. Exonic 5-hydroxymethylcytosine correlates with exon inclusion in *CD45* mRNA.

A Schematic representation of *CD45* exon 5 splicing and CTCF binding in B-cell lines and primary T cells, as previously characterized.

B Bisulfite sequencing of *CD45* exon 5 DNA in cell types that either include or exclude exon 5 from *CD45* mRNA, as indicated. Closed versus open circles indicate the locations of methylated or unmethylated CpGs, respectively. A partial sequencing chromatogram from naïve peripheral T cells demonstrates both bisulfite conversion efficiency and the persistence of methylated cytosines within the CTCF-binding site.

C Southern blot of genomic DNA double-digested with BsaHI (methyl-sensitive) and BglII (methyl-insensitive) hybridized to a radiolabeled 5' probe. The absence of BsaHI digestion in BL-E5(+) cells confirmed CTCF-binding site methylation.

D Methylated DNA immunoprecipitation (MedIP) with antibodies to 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) in BL41 B cells sorted and maintained based on distinct exon 5 inclusion levels. qPCR with primers against the *CD45* gene body, as indicated.

E 5hmC MedIP as in (D), but with peripheral human CD4⁺ T cells sorted on the basis of cell-surface expression of exon 5-containing *CD45* isoforms.

Data information: MedIP and values represent means $\pm$ SD from three independent IP experiments. Two-tailed Student's *t*-test comparing the specified values.

Source data are available online for this figure.

CpGs (2 per exon) and were not previously defined to be methyl-sensitive, nor at an upstream CTCF-binding site in *CD45* intron 2 that does not contain overlapping CpGs (Fig 1D) (Shukla *et al*, 2011). Likewise, mature human $CD4^+$ T cells (exon 5-excluding) are characterized by increased 5mC at exon 5 DNA relative to naïve cells (exon 5-including) (Shukla *et al*, 2011). Here, we find reciprocally increased 5hmC at exon 5 in the naïve relative to mature $CD4^+$ T-cell population (Fig 1E, Appendix Fig S1). These data confirm that the seemingly fixed overall methylation level in our exon 5-including versus exon 5-excluding isogenic cellular contexts is associated with a change in the ratio of 5hmC to 5mC at the CTCF-regulated exon.

TET1 and TET2 promote *CD45* exon 5 inclusion by antagonizing CTCF eviction

5hmC is a stable oxidation derivative in a pathway to active demethylation catalyzed by the TET proteins. To determine whether detection of 5hmC at *CD45* DNA reflects an active role in inclusion of exon 5 in processed mRNA, we modulated TET levels in 5hmC-rich BL-E5(+) cells (Fig 1D). qRT-PCR and Western blot revealed robust TET1 and TET2, but relatively little TET3 expression in BL41 cells (Appendix Fig S2). To examine the effect of forced 5hmC reduction on exon 5 splicing in BL-E5(+) cells, we performed shRNA-mediated depletion of all three TET proteins (Fig 2A, Appendix Fig S2). MedIP showed reduced 5hmC at *CD45* DNA in response to TET1 and TET2, but not TET3 depletion, perhaps owing to the already low level of TET3 in BL-E5(+) cells (Fig 2B, Appendix Fig S2). Remarkably, the extent of 5hmC depletion in TET1 and TET2 knockdown cells mirrored the degree of exon 5 repression in *CD45* transcripts as assessed through qRT-PCR with junction-spanning primers, suggesting a tight correlation between these parameters (Fig 2C).

As modulation of 5hmC has been shown to influence overall transcription, we further examined the broader context of *CD45* isoform expression. Cell-surface staining with an antibody to total *CD45* (pan-*CD45*) showed a minor decrease in overall *CD45* expression in response to TET1 and TET2 depletion (Fig 2D). In contrast, staining with the *CD45RB* antibody, which detects all exon 5-containing variants, showed a substantial reduction in TET1- and TET2-depleted cells relative to control that was commensurate to the degree of 5hmC reduction (Fig 2B and D). This effect was specific for exon 5, as *CD45RA* staining, which detects all exon 4-containing isoforms,

reflected the pan-*CD45* results (Fig 2D). Importantly, TET3 depletion in BL-E5(+) cells, which did not decrease 5hmC at exon 5, also did not alter exon 5 inclusion in *CD45* transcripts nor overall *CD45* expression (Appendix Fig S2). Likewise, depletion of TET1 and TET2 in unmethylated BJAB cells, which lack a substrate for TET activity, had no effect on *CD45* splicing or expression (Fig 2E, Appendix Fig S2). These findings support the conclusion that reduced exon 5 inclusion in *CD45* mRNA in response to TET1 and TET2 depletion is a function of perturbed *CD45* gene-body methylation, rather than an indirect effect of altered *CD45* expression.

As we previously showed that exon 5 inclusion in *CD45* mRNA is a consequence of CTCF binding at exon 5 DNA, we further examined whether loss of 5hmC in TET-depleted cells was coupled to altered CTCF association (Shukla *et al*, 2011). We thus performed large-scale RNAi in BL-E5(+) cells with the less penetrant TET1 hairpin, which resulted in a moderate decrease in *CD45* exon 5 inclusion, but produced the highest rate of cell survival (Appendix Fig S3). MedIP in TET1-depleted cells demonstrated decreased 5hmC with reciprocally increased 5mC at exon 5 and further showed no change in the overall low level of methylation at two CpG-encompassing regions of the *CD45* promoter relative to control (Fig 2F). Chromatin immunoprecipitation (ChIP) with an antibody to CTCF validated that elevated 5mC at exon 5 DNA in TET1-depleted BL-E5(+) cells manifested in reduced CTCF binding as compared to control-transduced cells (Fig 2F, right). Together, these RNAi data establish that the TET proteins facilitate CTCF-dependent inclusion of exon 5 in spliced mRNA by actively antagonizing overlapping exon 5 DNA methylation and further confirm that the observed effects are not secondary to altered 5mC at the *CD45* promoter.

T-cell activation-induced exon 5 exclusion is associated with reduced 5hmC and nuclear TET levels

Given the above indication that exchange of 5hmC for 5mC influences *CD45* alternative splicing through modulation of CTCF binding, we next examined the mechanistic basis for this conversion during lymphocyte development. While the signaling cascades leading to exon 5 exclusion in peripheral B cells are unclear, T cells can be induced to exclude exon 5 through direct ligation of the T-cell receptor (TCR). To this end, $CD4^+$ T cells were isolated from human peripheral blood and stimulated with agonist antibodies to the TCR

Figure 2. Reduced 5hmC through TET1 or TET2 depletion leads to decreased exon 5 inclusion in *CD45* transcripts.

- A BL-E5(+) cells were individually transduced with shRNA directed against *TET1* (TET1-sh1 and TET1-sh3), *TET2* (TET2-sh4) or control shRNA. Knockdown efficiency was assessed through qRT-PCR relative to expression of the housekeeping gene *RPS16*.
- B 5hmC MedIP and qPCR for *CD45* exon 5 DNA relative to input in control-transduced and *TET1*- or *TET2*-depleted BL-E5(+) cells.
- C qRT-PCR in sh*TET1*, sh*TET2*, and control-transduced BL-E5(+) cells, showing reduced detection of exon 5/6 junctions relative to a constitutive *CD45* junction (exon 7/8) in TET-depleted cells.
- D Cell-surface staining for *CD45RA* (exon 4-containing isoforms), *CD45RB* (exon 5-containing isoforms), and total *CD45* (pan) in *TET1*- or *TET2*-depleted BL-E5(+) cells. RNAi against either TET protein specifically reduced detection of exon 5-containing *CD45* isoforms relative to control.
- E *TET1* and *TET2* RNAi in unmethylated BJAB cells. qRT-PCR for knockdown efficiency (top) and cell-surface staining for exon 5-containing *CD45* isoforms (*CD45RB*, bottom) relative to control-transduced cells.
- F Large-scale depletion of *TET1* in BL-E5(+) cells. MedIP (left and center) and CTCF-ChIP (right) followed by qPCR with the indicated *CD45* primers in sh*TET1* (red, throughout) and sh-control (gray, throughout) cells. MedIP showed decreased 5hmC and increased 5mC, whereas CTCF-ChIP indicated reduced binding to *CD45* exon 5 in response to *TET1* depletion, as compared to control-transduced cells. Rabbit Ig ChIP served as a control.

Data information: ChIP and MedIP values represent means $\pm$ SD from three independent IP experiments. qRT-PCR values represent means $\pm$ SD from three independent cDNAs. Two-tailed Student's *t*-test comparing *TET1*- and *TET2*-depleted values to the sh-control.

Figure 2.

(anti-CD3 and anti-CD28) in the presence of interleukin-2 (Fig 3A). To quantify exon 5 inclusion independent of variations in alternative exons 4 and 6, we performed sub-saturating RT-PCR with primers extending from the exon 2/3 junction through exon 7 in the presence of dCTP α - 32 P. Phosphorimaging analysis revealed an approximate 2.5-fold reduction in exon 5 inclusion in response to activation as determined through the sum of the exon 5-containing *CD45* variants versus the sum of all *CD45* variants (Fig 3B, Appendix Fig S4).

Having confirmed robust exclusion of exon 5 in response to *in vitro* stimulation, we next examined whether and how *CD4*⁺ T-cell activation influences *CD45* DNA. Consistent with results from sorted *CD4*⁺ T cells (Fig 1E), MedIP showed a donor-independent decrease in 5hmC and increase in 5mC at exon 5 DNA following *in vitro* activation of naïve *CD4*⁺ T cells (Fig 3C, Appendix Fig S4). In contrast, the lower level of methylation at the surrounding exons

and CpG-containing regions of the *CD45* promoter remained relatively fixed (Fig 3C). Additionally, CTCF-ChIP showed reduced binding at exon 5 in activated relative to naïve *CD4*⁺ T cells (Fig 3D), whereas the CpG-less intron 2 site showed increased interaction following stimulation, demonstrating that CTCF binding to DNA is not generally compromised by activation (Appendix Fig S4). Thus, *in vitro* activation of *CD4*⁺ T cells recapitulates the determined changes in *CD45* exon 5 DNA methylation and pre-mRNA splicing that occur during the naïve to mature transition *in vivo*.

Building on our observations in TET1/2-depleted BL-E5(+) cells, to query the mechanisms supporting dynamic methylation and alternative splicing during lymphocyte development, we explored direct regulation of the TET proteins. Although partially functionally redundant, the TET proteins display some tissue and target site specificity (Hon *et al*, 2014; Huang *et al*, 2014; Wu & Zhang, 2014). Examination of individual TET transcripts in naïve *CD4*⁺ T cells

Figure 3. Activation-induced exclusion of exon 5 from *CD45* mRNA correlates with decreased exonic 5hmC.

A Naïve *CD4*⁺ T cells (black bars, throughout) were isolated from peripheral blood and activated *in vitro* with antibodies against CD3 and CD28 (red bars, throughout).
 B Phosphorimaging analysis of *CD45* RT-PCR products generated with primers extending from the exon 2/3 junction through exon 7 in the presence of dCTP α - 32 P. Percent exon 5 inclusion was determined as the sum of the exon 5-containing *CD45* variants versus the sum of all *CD45* variants.
 C MedIP with antibodies specific to 5hmC (left) or 5mC (right) followed by qPCR with the indicated *CD45* primers in naïve and activated *CD4*⁺ T cells.
 D CTCF-ChIP in naïve and activated T cells and relative qPCR for *CD45* DNA. Rabbit IgG ChIP served as a control.
 E qRT-PCR for TET family members normalized to the average of *RPS16*, *Bact*, and *Ubch5b* transcripts in naïve and activated *CD4*⁺ T-cell cDNA.
 F TET1 and TET2 Western blot in nuclear lysates from naïve and activated *CD4*⁺ T cells and p84 immunoblot as a loading control.
 G TET1-ChIP in naïve and activated T cells and qPCR for *CD45* DNA, relative to mouse IgG control ChIP.

Data information: MedIP and ChIP values represent means $\pm$ SD from three independent IP experiments. qRT-PCR values represent means $\pm$ SD from three independent cDNAs. Two-tailed Student's *t*-test comparing naïve and activated values.

Source data are available online for this figure.

revealed abundant *TET1* and *TET2* mRNA, but relatively little *TET3* (Fig 3E). Activation caused a substantial reduction in *TET1* and *TET3* transcripts, whereas *TET2* mRNA was marginally increased (Fig 3E). Consistent with the qRT-PCR results, TET1 immunoblot showed decreased expression in activated relative to naïve nuclear lysates (Fig 3F, Appendix Fig S4). However, in contrast to *TET2* mRNA, TET2 protein was significantly ablated in activated nuclear lysates, suggestive of posttranslational regulation (Fig 3F). Indeed, TET2 immunoblot in whole-cell extracts showed the emergence of two lower molecular weight species in activated lysates that are consistent with previously described caspase-dependent TET2 cleavage products (Appendix Fig S4) (Ko *et al*, 2013). Whereas we were unable to find antibodies against TET2 that produced reliable ChIP results (Appendix Fig S4), TET1-ChIP revealed abundant binding to *CD45* exon 5 in naïve $CD4^+$ T cells that was completely ablated following *in vitro* activation (Fig 3G). We can thus conclude that the mechanism supporting decreased 5hmC at *CD45* exon 5 in activated $CD4^+$ T cells is reduced local availability of the TET proteins. These primary T-cell data reinforce a model in which developmental regulation of exon 5 skipping results from a reduced ability to convert 5mC to its oxidation products and consequent CTCF eviction.

Methylated *CD45* minigenes confirm effect of 5hmC on exon 5 splicing

The sum of our data implies that the developmental emergence of 5mC at *CD45* DNA is preceded by a TET-facilitated, CTCF-bound state. However, co-detection of CTCF and 5hmC at *CD45* DNA called for re-examination of the precise determinants promoting pol II pausing and associated inclusion of exon 5 in *CD45* transcripts. To uncouple the specific contribution of CTCF from additional unknown TET-catalyzed 5oxiC effects, we turned to a *CD45* minigene to validate the proposed associations in a heterologous and mutable setting. We previously produced a mammalian expression vector that encompasses ~10 kilobases of human *CD45* (*hCD45*) genomic DNA extending from intron 3 through intron 7 with either a wild-type or mutated CTCF-binding site (I3-I7 and I3-I7*CTCF, respectively) (Shukla *et al*, 2011) (Fig 4A). Transfection of these

minigenes into fibroblast cell lines, which do not express endogenous *CD45*, confirmed that CTCF is required for exon 5 inclusion (Shukla *et al*, 2011): the I3-I7 minigene bound CTCF and was associated with pol II pausing and the I3-I7*CTCF minigene failed on both accounts and showed a dramatic decrease in exon 5 inclusion relative to the wild-type minigene.

To elucidate the independent contributions of CTCF and 5hmC on exon 5 inclusion through obstruction of pol II elongation, we layered methylation into the wild-type and mutant minigene analysis. We reasoned that stable incorporation of CpG-methylated minigene DNA into a chromosomal context would allow for propagation through DNMT1 activity and would further result in stochastic or integration site-dependent conversion to 5oxiC through endogenous TET function. Methylation of the I3-I7 minigenes was accomplished with the M.SssI CpG methyltransferase (Renbaum *et al*, 1990) and was confirmed through methyl-sensitive BsaHI restriction digest (Fig 4A). Methylated and control unmethylated I3-I7 minigenes were transfected into Chinese hamster ovary (CHO) cells and selected to generate stable clones. To correct for copy number variability between clones, data are presented as either a ratio or in reference to another location within the *CD45* minigene. Combined 5mC and 5hmC MedIP confirmed that a substantial portion of the I3-I7 minigene DNA was targeted for oxidation upon introduction into CHO cells (Fig 4B). On the whole, the wild-type clones showed slightly elevated 5hmC relative to the mutant clones, but this effect was not statistically significant (Appendix Fig S5). Oxidation of 5mC at the minigene DNA is further supported by elevated exon 5 inclusion in I3-I7- relative to I3-I7*CTCF-derived *CD45* transcripts, suggestive of CTCF interaction with the competent binding site (Fig 4C).

Wild-type and mutated CHO *hCD45* clones were selected for a detailed molecular analysis based on a high rate of conversion to 5hmC (Fig 4B, red and blue bars, respectively, referred to as WT-hmC and Mut-hmC henceforth), comparable minigene copy number and similar exon 4 mRNA expression (Fig 4D, Appendix Fig S5). Bisulfite sequencing validated that the minigene DNA retained methylation following integration and propagation in the CHO genome (Fig 4E). Note that CpG content in the CTCF-binding site

Figure 4. Methylated *CD45* minigenes support role for 5mC oxidation in exon inclusion through CTCF binding.

- Plasmids containing human *CD45* (*hCD45*) genomic DNA from intron 3 through intron 7 with a wild-type (I3-I7) or mutated exon 5 CTCF-binding site (I3-I7*CTCF) were methylated with M.SssI and confirmed through BsaHI digest. Methylated and unmethylated minigenes were transfected into CHO cells for stable clone generation.
- 5hmC and 5mC MedIP were performed in stable clones derived from methylated I3-I7 and I3-I7*CTCF minigenes. Results from individual clones are plotted as a percent of total methylation ($n = 32$ clones) as determined through input normalized IP values for 5hmC or 5mC divided by the sum of the values for 5hmC and 5mC. Methylation in the I3-I7 and I3-I7*CTCF clones used for downstream analysis is indicated by the red and blue bars (WT-hmC and Mut-hmC clones), respectively.
- qRT-PCR for relative exon 5 inclusion in methylated I3-I7 and I3-I7*CTCF minigene-derived mRNA. Results are shown for the average of all the I3-I7 and I3-I7*CTCF clones ($n = 16$ clones of each) as well as individually for the WT-hmC and Mut-hmC clones that were selected for downstream analysis. Values represent mean $\pm$ SEM for the pooled data or the mean $\pm$ SD for the individual clones. Two-tailed Student's *t*-test comparing the I3-I7 to I3-I7*CTCF values.
- qPCR of minigene exon 5 relative to CHO *GAPDH* for comparison of *hCD45* copy number in the WT-hmC and Mut-hmC clones. Values represent means $\pm$ SD from 3 replicates. Two-tailed Student's *t*-test comparing the WT-hmC and Mut-hmC values.
- Bisulfite sequencing of *hCD45* exon 5 DNA from the WT-hmC and Mut-hmC clones to confirm the persistence of methylation at the intact and mutated CTCF-binding site following stable insertion into the CHO genome.
- CTCF-ChIP in the WT-hmC and Mut-hmC clones and qPCR for *hCD45* exons, as indicated. ChIP values represent means $\pm$ SD from three independent IP experiments. Two-tailed Student's *t*-test comparing WT-hmC and Mut-hmC values.
- Pol II-ChIP in the WT-hmC and Mut-hmC clones, and a WT-C CHO clone generated with an unmethylated *hCD45* minigene. qPCR for *hCD45* DNA as indicated. ChIP values represent means $\pm$ SD from three independent IP experiments. * $P = 0.013$, ** $P = 0.0089$, two-tailed Student's *t*-test comparing the Wt-hmC and Mut-hmC values.
- Summary of the WT-hmC and Mut-hmC data illustrating elevated detection of CTCF and pol II at the wild-type minigene exon 5, relative to the mutated CTCF-binding site.

Source data are available online for this figure.

Figure 4.

was not perturbed in the mutant minigene (Fig 4E). As in the overall population of clones, WT-hmC minigene-derived transcripts showed substantially increased exon 5 inclusion compared to the Mut-hmC clone, as determined through qRT-PCR with junction-spanning primers (Fig 4C). Consistent with our demonstration that CTCF binding to exon 5 DNA is a major mechanism of exon 5 inclusion (Shukla *et al*, 2011), CTCF-ChIP confirmed binding to exon 5 in the WT-hmC, but not in the Mut-hmC clone (Fig 4F). In addition to validating co-detection of 5hmC and CTCF at the minigene DNA, comparison of these hydroxymethylated *hCD45* clones differing only in the CTCF-binding site sequence illustrates that 5hmC does not directly mediate exon 5 inclusion, but rather acts through facilitating CTCF binding.

As we previously showed that CTCF promotes exon inclusion through kinetic regulation (Shukla *et al*, 2011), to further deconvolute the contributions of 5oxiC and CTCF to exon 5 inclusion, we examined pol II occupancy at the minigene DNA. To this end, we performed pol II-ChIP in the WT-hmC and Mut-hmC clones, as well as an unmethylated wild-type *hCD45* CHO clone generated from an unmethylated minigene (Appendix Fig S5, referred to as WT-C henceforth). Comparison of pol II profiles across the minigene DNA from these clones with distinct methylation status and CTCF binding capacity clearly demonstrates that CTCF and not 5hmC is the main determinant for pol II pausing at *CD45* exon 5. Irrespective of methylation status, both clones with an intact CTCF-binding site showed impressive overlap in their pol II profiles proximal to the CTCF-binding site with increased occupancy at exon 5 DNA relative to the mutant clone (Fig 4G, Appendix Fig S5). Through these combined CHO *hCD45* pol II data, we reason that 5hmC does not directly promote kinetic regulation of exon 5 inclusion, but instead facilitates CTCF-associated pol II pausing through the oxidation of overlapping methylcytosine residues (Fig 4H).

To solidify the role of TET-dependent 5hmC in *CD45* alternative splicing, we performed siRNA-mediated depletion of *Tet1* and *Tet2* in the WT-C and WT-hmC clones (Fig 5A and B), with the expectation that only the hydroxymethylated DNA would be susceptible to effects directly related to minigene methylation. MedIP in the *Tet1/2*-depleted and control cells demonstrated both a high basal level of 5hmC in the WT-hmC clone relative to WT-C, but also a dramatic decrease in 5hmC at the minigene DNA following reduction of *Tet1* and *Tet2* relative to non-target siRNA (Fig 5C). 5mC MedIP and bisulfite sequencing further indicated stable and persistent methylation following *Tet1/2* depletion, suggesting that loss of 5hmC in the WT-hmC clone is associated with a reciprocal increase in 5mC (Appendix Fig S5). Accordingly, decreased 5hmC in the *Tet1/2*-depleted WT-hmC clone was associated with a marked reduction in CTCF binding to the minigene DNA relative to control, as assessed by ChIP (Fig 5D). Importantly, siRNA against *Tet1/2* did not alter CTCF binding to the unmethylated WT-C minigene relative to non-target, thereby demonstrating specificity to the presence of methylated template (Fig 5D). We next examined the impact of altered Tet expression on minigene-derived mRNA. qRT-PCR indicated an overall increase in minigene transcription in the *Tet1/2*-depleted WT-C clone relative to siNT (Appendix Fig S5). To circumvent transcriptional biases, we examined the change in per exon level expression in *Tet1/2*-depleted cells relative to siNT controls and observed a substantial decrease in exon 5, but not exons 4 and 6 inclusion in WT-hmC cells depleted of *Tet1* and *Tet2*. In contrast, the WT-C

clone showed a mild increase in exon 5 inclusion upon *Tet1/2* depletion (Fig 5E). Combined with the primary and cell line data, we conclude that a major mechanism of *CD45* alternative splicing in developing lymphocytes involves a failure to target the TET proteins to exon 5 DNA, resulting in increased 5mC, loss of CTCF-dependent polymerase pausing and associated exclusion of exon 5 from *CD45* mRNA.

CTCF interacts with carboxylated DNA at *CD45* exon 5

The above establishes a functional relationship between 5hmC and *CD45* exon 5 inclusion, but whether 5hmC directly interacts with CTCF or represents an intermediate to a CTCF-binding state was unclear. While the inhibitory effect of 5mC on CTCF binding has been demonstrated in a variety of settings (Bell & Felsenfeld, 2000; Nguyen *et al*, 2008; Phillips & Corces, 2009), interaction with 5hmC, 5fC, and 5caC has not been formally tested. To clarify the relationship between CTCF and 5oxiC species, we pursued electrophoretic mobility shift assay (EMSA) with purified CTCF and human *CD45* exon 5 DNA. FLAG-tagged CTCF was immunoprecipitated from HEK293T cells, and purity was confirmed through silver stain and CTCF Western blot (Fig 6A). Preliminary EMSA with probes of increasing length centered around the CTCF-binding site identified 60 base pairs as the minimum sequence requirement for complex formation. Extending this sequence further resulted in increasingly robust interaction with CTCF (Appendix Fig S6). Specificity of binding to probes of varying length was confirmed through mutation of the CTCF-binding site, which fully ablated complex formation (Fig 6B, Appendix Fig S6). Consistent with reports detailing an inability of CTCF to discriminate between unmethylated and methylated sequences in certain contexts, particularly below a certain CpG threshold, CpG methylation of the minimal 60-mer *CD45* probe DNA was insufficient to confer methyl sensitivity (Appendix Fig S6) (Feldmann *et al*, 2013; Teif *et al*, 2014; Batlle-Lopez *et al*, 2015).

To specifically address the interaction between CTCF and variably methylated *CD45* DNA, we established conditions that conform to the *in vivo* observations, wherein overlapping 5mC acted to evict CTCF. 72-base pair exon 5 probes were generated through PCR in the presence of modified deoxynucleotide triphosphatases (dNTPs) corresponding to 5mC, 5hmC, 5fC, or 5caC, increasing the level of methylation overlapping and flanking the CTCF-binding site to a total of 14 modified residues (Appendix Fig S6). EMSA with the labeled 72-mers and purified CTCF resulted in abundant binding to the unmethylated sequence, but not to the 5mC-containing probe. However, to our surprise, CTCF showed minimum interaction with the 5hmC- and 5fC-containing probes, but instead formed a robust complex in the presence of 5caC-containing DNA (Fig 6C). Specificity of interaction with 5caC-containing DNA was confirmed through supershift with antibody to CTCF (Fig 6D). Additionally, the unmethylated and 5caC complexes were ablated in the presence of either unmethylated or 5caC-containing competitor (Fig 6D). Competition was more effective in the presence of the 5caC competitor, implying a subtle distinction between the two complexes (Fig 6E). Binding to 5caC-containing DNA was not an artifact of contaminating protein co-purification: EMSA with recombinant CTCF from a commercial source showed the same efficient interaction with 5caC-containing DNA that was *de facto* stronger than

Figure 5. *Tet1/2* depletion in wild-type *hCD45* CHO clones confirms specificity to methylated substrates.

A, B Unmethylated (WT-C) and hydroxymethylated (WT-hmC) I3-17 clones were transfected with siRNA against *Tet1* and *Tet2* (si*Tet1/2*) or non-target siRNA (siNT) (schematic in A). qRT-PCR (A) and Western blot (B) confirmed knockdown efficiency relative to CHO *GAPDH* and the p84 loading control, respectively.

C 5hmC MedIP and qPCR for human *CD45* exon 5 in si*Tet1/2* and siNT WT-C and WT-hmC clones, relative to a minigene region lacking CpGs.

D CTCF-ChIP and qPCR for human *CD45* exon 5 in si*Tet1/2* and siNT WT-C and WT-hmC CHO clones. Rabbit IgG ChIP served as a control.

E qRT-PCR with human *CD45* primers in si*Tet1/2* and siNT WT-C and WT-hmC CHO clones. Per each condition, exon level values are normalized to exon 6 and summarized data are presented relative to siNT to distinguish splicing changes from transcriptional variability.

Data information: Values represent means $\pm$ SD from three independent IP experiments or cDNA preparations. Two-tailed Student's *t*-test comparing the indicated values.

Source data are available online for this figure.

interaction with unmethylated DNA (Appendix Fig S6). Likewise, commercial synthesis of probes incorporating 5caC at the three CpGs directly overlapping the core CTCF-binding site strengthened the otherwise weak complex formed with the 41-base pair probe (Appendix Fig S6). Furthermore, EMSA examination of a distinct CTCF-binding site present in the *KCNA2B* gene showed the same binding tendencies: CTCF failed to bind 5mC, 5hmC, and 5fC, but was enhanced in the presence of 5caC-containing relative to

unmethylated DNA, suggesting generality in 5caC-binding preferences (Appendix Fig S6).

These EMSA results imply that detection of 5hmC at *CD45* DNA in CTCF-binding cellular populations reflects its role as an intermediate to the downstream oxidation products 5caC and cytosine. Consistent with this possibility, bisulfite pyrosequencing in *CD4*⁺ T cells revealed overall higher conversion efficiency at exon 5 relative to exon 4 DNA, indicative of the increased presence of

Figure 6. Mammalian CTCF interacts with 5caC-containing CD45 DNA.

A Silver stain (left) and CTCF immunoblot (right) of 3xFLAG-CTCF purified from transfected HEK293T cells.

B EMSA with purified CTCF and PCR-generated CD45 exon 5 probes with a wild-type or mutated CTCF-binding site. Cold competition was performed with unlabeled wild-type probe.

C EMSA with purified CTCF and PCR-generated 72-bp probes symmetrically flanking the exon 5 CTCF-binding site. Substitution of modified dCTP allowed for variable CTCF-binding site methylation. Cold competition was performed with unmethylated wild-type probe.

D EMSA with purified CTCF and PCR-generated 72-bp exon 5 probes generated in the presence of cytosine or 5caC. Cold competition was performed with unmethylated or 5caC-containing wild-type probe, as indicated. Supershift was accomplished with anti-CTCF antibody.

E Quantification of EMSA from (D). The volume of free probe and CTCF complexed probe was determined with ImageQuant software (Molecular Diagnostics). Percent shift represents the complex volume divided by the sum of complex plus free probe.

F MedIP with antibody specific to 5caC followed by qPCR with the indicated CD45 primers in naïve and activated CD4⁺ T cells. Values represent means $\pm$ SD from three independent IP experiments. Two-tailed Student's t-test comparing the indicated values.

Source data are available online for this figure.

5fC, 5caC or cytosine within exon 5 (Appendix Fig S7). While 5fC MedIP was not enriched above background (Appendix Fig S7), MedIP with antibody to 5caC showed substantial accumulation at *CD45* exon 5 in naïve $CD4^+$ T cells (Fig 6F). Furthermore, concomitant with the observed reductions in 5hmC and CTCF, T-cell activation resulted in decreased 5caC levels at *CD45* DNA (Figs 3 and 6F). Likewise, 5caC MedIP in the WT-hmC and Mut-hmC *hCD45* CHO clones supports overlapping 5caC at the CTCF-binding site and further establishes that CTCF binding to 5caC is not non-specific, but rather depends on an intact binding site (Appendix Fig S7). While these combined EMSA and MedIP results are not definitive of CTCF binding preferences *in vivo*, they strongly support a role for TET-catalyzed 5caC in *CD45* alternative pre-mRNA splicing regulation through direct binding of CTCF.

Exons differentially spliced in peripheral T cells are enriched for proximal CTCF binding

The above data from the *CD45* model system identifies TET-catalyzed oxidation of methylcytosine as a novel regulator of alternative pre-mRNA splicing that operates through antagonizing intragenic CTCF eviction. To examine the universality of these findings in a physiologically relevant cellular system, we generated genome-wide data in naïve and activated $CD4^+$ T cells, where we observed an overall decrease in nuclear TET levels upon stimulation (Figs 3F and 7A, Table EV1). As T-cell activation results in numerous changes in splicing that are directly attributed to RNA-binding protein activity (Cho *et al*, 2014; Kafasla *et al*, 2014), we performed iterative filtering through combined RNA-, CTCF-ChIP-, and MedIP-seq analysis to specifically query methyl-sensitive CTCF-associated splicing. Alternatively spliced exons were defined from RNA-seq data as exons that were differentially included during the naïve to activated transition independent of altered transcription (Table EV2). To facilitate downstream analysis, we focused on internal skipped exons that showed a change in percent spliced in (Δ PSI) of ≥ 0.1 ("activation-included") or ≤ -0.1 ("activation-excluded") in response to T-cell activation (Fig 7B, collectively referred to as " Δ AS"). Constitutively spliced exons and annotated alternative exons that were not differentially included in response to stimulation ("AS, not Δ ") served as unchanged controls (Fig 7B).

Alternatively spliced exons were examined for proximity to CTCF-ChIP-seq peaks in naïve and activated $CD4^+$ T cells (Appendix Fig S8, Table EV3). Examination of CTCF binding at *CD45* exon 5 showed the expected trend, with greater binding in naïve relative to activated cells (Appendix Fig S8). As we previously showed that CTCF binding downstream of alternative exons influences splicing from a distance (Shukla *et al*, 2011), we examined sites within 1.5 kb of the interrogated exon. On the whole, alternatively spliced (Δ AS) exons were enriched for overlapping or proximal CTCF as compared to unchanged alternative (AS, not Δ) and constitutive exons (Appendix Fig S8, Table EV4). Consistent with our demonstration that CTCF regulates inclusion of upstream exons through modulation of pol II elongation, enrichment persisted when considering just CTCF sites that are downstream of alternatively spliced exons (Fig 7C). However, enrichment was most pronounced in the upstream CTCF category (Fig 7C), where we previously failed to see a genome-wide association with alternative splicing as determined through averaged Δ PSI after CTCF depletion (Shukla *et al*,

2011). Additional sub-categorization of CTCF sites based on the directional change in proximal exon splicing did not reveal clear correlations between activation-induced splicing changes and CTCF location (Appendix Fig S8). Together, these data illustrate that in the absence of information regarding differential CTCF binding, there is no apparent relationship between CTCF location and directional changes in splicing. By extension, they further suggest that T-cell activation does not produce a uniform effect on CTCF binding through unilateral changes in overlapping methylation.

Reciprocal methylation in exon-proximal CTCF-binding sites correlates with splicing regulation

To specifically address the impact of differential methylation on alternative splicing through CTCF modulation, iterative filtering was applied to the genome-wide data to identify binding sites with evidence of activation-induced 5hmC/5mC exchange. Although the presented EMSA data suggest that CTCF interacts with 5caC-containing DNA at *CD45* exon 5, 5caC cellular levels are low, such that previous mapping studies have relied on depletion of thymine-DNA glycosylase (TDG) to stabilize 5caC in genomic DNA (He *et al*, 2011; Ito *et al*, 2011; Shen *et al*, 2013; Wang *et al*, 2015). Importantly, these investigations found that locations of induced 5caC upon TDG knockdown strongly overlap 5hmC signals from control cells (Shen *et al*, 2013). In addition, our reanalysis of TDG knockdown 5caC MedIP-seq detected increased signal within CTCF-ChIP-seq peaks (Shen *et al*, 2013; Yue *et al*, 2014) (Appendix Fig S8), consistent with the notion that CTCF sites are general locations of dynamic TET activity.

Given indications that 5hmC is an accurate metric for TET activity, 5hmC MedIP-seq was performed to identify CTCF-binding sites with evidence of TET-catalyzed 5mC oxidation in $CD4^+$ T cells. CTCF-binding sites located within 1.5 kb of the defined exonic categories were assessed for overlapping 5hmC. As compared to random genomic bins of equal breadth, 5hmC was generally elevated at CTCF-binding sites, with enrichment most pronounced at exon-proximal CTCF sites (Appendix Fig S8). CTCF-binding sites with no evidence of 5hmC in either naïve or activated $CD4^+$ T cells were omitted from subsequent analysis. The remaining 5hmC(+) sites were examined for activation-induced changes in 5hmC and the net association with 5mC MedIP-seq (Fig 7D). Analysis of overlapping 5mC at CTCF-binding sites with activation-increased 5hmC showed no association with 5mC loss or gain in all exon-proximal categories (Fig 7E). In contrast, CTCF-binding sites with activation-decreased 5hmC proximal to Δ AS exons were significantly more likely to gain reciprocal 5mC relative to sites proximal to unchanged exons (Fig 7E, Table EV5). Conversely, Δ AS proximal CTCF sites with reduced 5hmC were less likely to be associated with decreased or unchanged 5mC (Fig 7E, Appendix Fig S8). Enrichment of 5hmC-decreased/5mC-increased CTCF-binding sites proximal to alternatively spliced exons suggests that methyl-dependent eviction of CTCF is a general mechanism of splicing regulation in peripheral $CD4^+$ T cells.

To examine the relationship between differential CTCF-binding site methylation and alternative pre-mRNA splicing in greater detail, Δ AS proximal CTCF-binding sites with reciprocal 5hmC/5mC were further subcategorized as either upstream or downstream of the queried exon (Fig 7F). Sites with decreased 5hmC and increased 5mC were considered as potential sites of CTCF eviction (blue),

Figure 7. Genomewide analysis of differential methylation at exon-proximal CTCF-binding sites.

- A Schematic of genomewide experiments performed on naive and activated CD4⁺ T cells.
- B Volcano plot of MISO results for all annotated skipped exons. ΔAS exons (in red) have Bayes factor ≥ 10 and a ΔPSI absolute value ≥ 0.10 . Within the ΔAS exons, negative ΔPSI values indicate activation-excluded exons, whereas positive ΔPSIs reflect activation-included exons. Unchanged alternative exons (AS, not Δ, dark gray) have a Bayes factor < 10 and a ΔPSI absolute value < 0.10 .
- C Proximity of exons to CTCF. Genomic coordinates corresponding to alternatively spliced (ΔAS, red), unchanged alternative (AS, not Δ, gray), expressed constitutive (black), or all annotated constitutive exons (white) were intersected with empirically defined CTCF-binding sites in naive and activated CD4⁺ T cells. Exons were queried for CTCF sites that directly overlap or are located within 1.5 kb upstream or downstream, as indicated. Fisher's exact test.
- D CTCF-binding sites (CBS) with overlapping 5hmC in CD4⁺ T cells were examined for activation-induced changes in 5hmC as determined by MedIP-seq. Results for exon-proximal CTCF sites are shown, as are all intragenic CTCF sites and random intragenic bins. Input-adjusted changes in methylation $\geq 10\%$ are indicated.
- E Exon-proximal CTCF-binding sites with activation-induced differential 5hmC were assessed for reciprocal changes in 5mC. The percent of sites with differential 5hmC and 5mC was sorted according to the directional change in methylation. The significance of differences between sites adjacent to different exon classes is indicated (one-sided Fisher's exact test).
- F ΔAS proximal CTCF-binding sites with reciprocal 5hmC/5mC were examined for the impact on alternative splicing. Sites with relatively increased 5hmC and decreased 5mC (shown in orange) were considered as potential sites of CTCF gain, whereas sites with decreased 5hmC and increased 5mC (shown in blue) were considered as sites of CTCF loss. Exons were sub-categorized based on CTCF location (upstream or downstream) and assessed for directional changes in splicing as either activation-included (stippled boxes) or activation-excluded (solid boxes). Pie charts show the relationship between affected exon splicing categories and reciprocal methylation at upstream or downstream CTCF-binding sites as percent of exons (bottom left). Median $\pm$ standard error of ΔPSI values for exons with up- or downstream CTCF and increased or decreased relative 5hmC are plotted (right), mean values are given above.

whereas sites with increased 5hmC and decreased 5mC were considered as locations of CTCF gain (orange). While the number of events remaining after high-stringency filtering was small (34.7% of ΔAS with proximal CTCF and 3.9% of all ΔAS exons), separation into these distinct categories allowed for the emergence of clear

tendencies. Alternative exons with downstream CTCF showed strong concordance with the previously established model: 100% of activation-excluded exons were associated with decreased 5hmC and increased 5mC, consistent with loss of kinetic regulation through CTCF eviction (Fig 7F, Table EV6). Likewise, 100% of

downstream CTCF-binding sites with activation-increased 5hmC and activation-decreased 5mC (conducive to CTCF binding) were associated with inclusion of the upstream Δ AS exon (Fig 7F). In contrast, upstream CTCF-binding sites showed no association between directional changes in 5hmC and impact on downstream alternative splicing: locations with increased relative 5hmC or increased relative 5mC were both slightly skewed to downstream exon inclusion (Fig 7F). These tendencies were echoed in the overall Δ PSIs per category, wherein alternative exons with downstream CTCF showed clear partitioning between CTCF sites with relative decreased 5hmC (negative mean and median Δ PSI) versus those with relative increased 5hmC (positive mean and median Δ PSI) (Fig 7F). In contrast, differential methylation at upstream CTCF sites did not associate with downstream splicing: the overall impact on Δ PSI was level in both methylation categories with a mean Δ PSI of zero (Fig 7F). qPCR validation of the genomewide predictions showed strong agreement in independently processed MedIP, CTCF-ChIP, and RNA samples from patient donors, and further confirmed decreased 5caC concomitant with decreased 5hmC (Appendix Fig S9). In sum, through high-stringency iterative filtering in naïve and activated CD4⁺ T-cell sequencing data, we provide evidence that oxidation of methylcytosine at CTCF-binding sites globally regulates upstream exon inclusion. In addition, enrichment of CTCF binding upstream of alternative exons suggests an expansion of our model wherein CTCF regulates upstream exons in a unidirectional fashion, but downstream exons bidirectionally through a currently unknown mechanism.

Discussion

Recent reports identifying enhanced differential methylation within the intragenic landscape has led to increased focus on a potential role for gene-body DNA methylation in the regulation of gene expression. Unlike promoters, intragenic methylation positively correlates with expression and this relationship extends to the exon level: DNA methylation at expressed exons is elevated relative to introns, pseudoexons, and excluded alternative exons (Lister *et al*, 2009; Choi, 2010; Feng *et al*, 2010; Lyko *et al*, 2010; Zemach *et al*, 2010; Khare *et al*, 2012; Gelfman *et al*, 2013; Maunakea *et al*, 2013; Tsagaratou *et al*, 2014). Exonic enrichment has been demonstrated for both 5mC and 5hmC, but the functional significance of 5mC oxidation has remained unclear (Ficz *et al*, 2011; Pastor *et al*, 2011; Khare *et al*, 2012). Here, we report TET-catalyzed dynamic DNA methylation as a key regulator of CTCF-dependent alternative pre-mRNA splicing. Focusing on *CD45* alternative splicing, we show that TET1 and TET2 promote inclusion of a CTCF-dependent exon through oxidation of methylcytosine at the corresponding DNA. Conversely, a reduction in TET levels increases 5mC and exon exclusion through CTCF eviction. We further show that 5oxiC does not directly impact pol II pausing and exon inclusion, but rather enables CTCF binding by antagonizing overlapping 5mC.

Biochemical analysis of CTCF association with methylated DNA revealed a surprising preference for 5caC-containing DNA (Fig 6). While not definitive of CTCF binding preferences *in vivo*, these results reveal an unanticipated layer of complexity in CTCF/DNA interaction. The association between CTCF and 5caC-containing DNA is unlikely to be restricted to the tested sequences: an unbiased

survey for proteins that show preferential interaction with distinct cytosine species in CpG modified oligonucleotides identified CTCF as a 5caC-specific reader (Spruijt *et al*, 2013). In addition, CTCF binding in the presence of 5caC is not likely to be an artifact of linear DNA templates: CTCF is known to occupy nucleosome-free DNA *in vivo* (Ong & Corces, 2014), and 5caC was detected at CTCF-binding sites in primary T cells (Fig 6F, Appendix Fig S9). This latter demonstration is particularly relevant given that 5caC is extremely rare/transient in genomic DNA (He *et al*, 2011; Ito *et al*, 2011), implicating efficient CTCF/5caC interaction as a potential protective mechanism against demethylation through TDG-associated base excision repair. From the perspective of pre-mRNA splicing, genomewide mapping in TDG-depleted ES cells revealed increased 5caC at exonic sequences of actively transcribed genes, thus highlighting intragenic DNA as an enriched location for dynamic DNA methylation catalyzed by the TET proteins (Shen *et al*, 2013; Wang *et al*, 2015). Of particular relevance to our study, a recent report linked elevated intragenic 5caC following TDG depletion to reduced RNA polymerase II elongation (Wang *et al*, 2015). Our reanalysis of these 5caC data and available CTCF-ChIP-seq from the same ES cell populations found elevated 5caC central to these CTCF-binding sites (Appendix Fig S8). In addition to establishing TET-catalyzed oxidation as a general feature of CTCF-binding sites, given the known impact of CTCF on pol II (Wada *et al*, 2009; Shukla *et al*, 2011), this analysis raises the possibility that part of the reduced elongation may be due to enhanced CTCF binding as 5caC levels increase.

Of note, while the presented EMSA data indicate that CTCF prefers 5caC-containing DNA, they further demonstrate that CTCF is able to interact with 5mC and 5hmC CpG methylation (Fig 6C, Appendix Fig S6). Likewise, several studies have revealed CTCF binding to methylated DNA *in vivo* (5mC or 5hmC), particularly at intragenic locations with low CpG content (Stadler *et al*, 2011; Feldmann *et al*, 2013; Teif *et al*, 2014). These observations call into question whether co-detection of CTCF and 5hmC at *CD45* DNA and in the genomewide analysis represents a legitimate binding state, or a transition intermediate to 5caC-binding DNA. Alternatively, perhaps CTCF is capable of interacting with low methylated regions in intragenic environments but generates a more favorable binding context through recruitment of the TET proteins. The demonstrations that CTCF interacts with and recruits the TET proteins to genomic DNA and global studies detailing co-localization of CTCF and 5hmC bolster this premise (Feldmann *et al*, 2013; Sun *et al*, 2013; Dubois-Chevalier *et al*, 2014; Teif *et al*, 2014). Notably, *CD45* minigenes with a competent CTCF-binding site showed a mild enrichment for 5hmC relative to mutated minigenes, albeit not significant (Appendix Fig S5). While the mechanistic basis for preferential CTCF interaction with oxidated 5mC derivatives is not yet clear, possibilities include alterations in DNA structure that expose the CTCF-binding site, increased electrophysical interaction due to the enhanced negative charge of 5caC, and/or intrinsic disfavoring of nucleosome assembly at 5caC-rich DNA. Cumulatively, these observations suggest that CTCF could both respond to and shape the intragenic epigenome to effect regulated changes in pre-mRNA processing.

Our demonstrations at *CD45* DNA in peripheral CD4⁺ lymphocytes further served as the basis for genomewide analysis of TET-catalyzed 5mC oxidation on pre-mRNA alternative splicing.

Building on the observation that unmethylated and 5oxiC-containing *CD45* DNA were both able to bind CTCF and promote exon inclusion, but only the 5oxiC substrate was associated with alternative splicing through variations in TET activity, we examined whether differential 5oxiC plays a generalized role in alternative splicing through modulation of CTCF. Given the relative abundance of 5hmC in the genome and evidence that 5hmC is readily detected at established sites of further oxidation (Shen *et al*, 2013), we utilized 5hmC MedIP as an accurate metric for locations of active TET-catalyzed 5mC oxidation. As CD4⁺ T-cell stimulation activates a number of splicing regulatory pathways (Cho *et al*, 2014; Kafasla *et al*, 2014; Martinez *et al*, 2015), we invoked high-stringency iterative filtering to distinguish direct effects related to methylation at CTCF sites from indirect effects. The resulting locations with downstream CTCF showed strong adherence to the established model for CTCF function: increased relative 5hmC was associated with upstream exon inclusion, whereas increased relative 5mC was associated with exclusion, consistent with CTCF binding and eviction, respectively. Surprisingly, however, CTCF binding was most enriched upstream of alternative exons and was not associated with any clear effects on downstream alternative splicing. Notably, several recent studies have uncovered location-dependent associations between chromatin-interacting factors and alternative splicing. In addition to CTCF, the HP1 proteins and Ago1 have both been implicated in direct regulation of alternative splicing through intragenic binding (Allo *et al*, 2014; Agirre *et al*, 2015; Yearim *et al*, 2015). In the case of the HP1 proteins, upstream or overlapping binding mediated distinct effects on alternative splicing that were further dependent on overlapping methylation (Yearim *et al*, 2015). The proposed model for chromatin-directed modulation of downstream splicing involves kinetic regulation of trans-factor binding at the upstream introns (repressors in the cited examples) through transient obstruction of pol II elongation (Dujardin *et al*, 2014; Yearim *et al*, 2015). Taken together, we envision that CTCF upstream of exons can promote or inhibit downstream exon inclusion through kinetically favoring binding of positive or negative acting trans-factors, respectively.

In summary, we reveal dynamic DNA methylation catalyzed by the TET proteins as a novel modulator in the complex pre-mRNA splicing code. Our description of an epigenetic splicing switch based on the relative ratio of 5mC to its oxidation derivatives begins to reveal a function for widespread intragenic DNA methylation in the variable assembly of DNA-binding regulators of splicing. While our study focuses on CTCF, we anticipate that intragenic DNA methylation will be revealed to influence a diverse network of DNA-binding proteins that promote or inhibit splicing. Combined with recent descriptions of methyl-binding and methyl-inhibited proteins that are specific for 5mC, 5hmC, 5fC, or 5caC (Spruijt *et al*, 2013) and our observation that focused changes in intragenic DNA methylation do not affect overall transcription, DNA methylation is uniquely positioned to contribute to the combinatorial logic of regulated alternative pre-mRNA splicing through sequence-dependent shifts in methyl-sensitive transcription factor binding. We thus propose that the precise mechanisms that regulate gene expression through dynamic methylation at promoters operate within gene bodies to regulate pre-mRNA processing through variable binding of a putative repertoire of methyl-sensitive DNA-binding proteins, as we show for CTCF.

Materials and Methods

Cell culture

BJAB and BL41 cells were maintained in RPMI supplemented with 10% FBS and 1% L-glutamine. CHO cells were cultured in F12 supplemented with 10% FBS and 1% L-glutamine. Neomycin was added to a final concentration of 1 mg/ml for selection of stable clones. HEK293T cells were grown in DMEM supplemented with 10% FBS and 1% L-glutamine. All cell lines tested mycoplasma negative (12/2011). Naïve peripheral CD4⁺ T cells were purified from human blood through Ficoll Paque and magnetic bead negative depletion according to the manufacturer's directions (Miltenyi Biotech, 130-094-131). Naïve CD4⁺ T cells were cultured in RPMI supplemented with 10% FBS, 55 mM β -mercaptoethanol, and 30 units/ml human interleukin-2 (IL-2). Cells were either left to rest in IL-2-containing media (naïve) or activated with antibodies directed against CD3 and CD28 receptors for 2–3 days. Activation was confirmed by cell-surface staining with the CD45RO antibody, and naïve and activated T cells were harvested in parallel for downstream applications.

Lentiviral transduction

pLKO.1 lentiviral constructs encoding shRNA against human *TET1*, *TET2*, and *TET3* as well as shGFP and shRFP as controls were purchased from Open Biosystems. Lentivirus production and cell transduction were achieved as previously described (Shukla *et al*, 2011). Briefly, lentivirus was produced in 293T cells through co-transfection with VSV-G and gag/pol plasmids with Lipofectamine 2000 (Invitrogen). Lymphocytes were transduced with concentrated virus through spin infection in the presence of 8 μ g/ml protamine sulfate and were collected on day 6 or later.

Stable clone generation

Generation of pC1-Neo encoding *CD45* minigenes with wild-type and mutated CTCF-binding sites was previously described (Shukla *et al*, 2011). Methylation of *CD45* minigenes was performed *in vitro* with M.SssI CpG methyltransferase (New England Biolabs) according to the manufacturer's suggestions. Uniform plasmid methylation was confirmed with the methyl-sensitive restriction enzyme, BsaHI (NEB). Methylated and unmethylated plasmids were linearized with AhdI (NEB) and transfected into Chinese hamster ovary cells with Lipofectamine 2000 (Invitrogen). Individual clones were selected with neomycin-containing media (1 mg/ml) and amplified for further analysis.

siRNA

CHO-K1-selected clones containing methylated or unmethylated I3-I7 minigenes were subjected to two rounds of *Tet1/Tet2* smart-pool On-Target Plus siRNA (Dharmacon, Thermo Scientific) (10 nM) using Dharmafect 1 reagent (Dharmacon Fisher). Transfections were separated by 48 h. Cells were then harvested at 72 h after the first round of treatment for RNA, DNA, ChIP, and protein samples.

Antibodies

The following antibodies were used for flow cytometry: CD45RO [UCHL1] (eBioscience), CD45RA [MEM-56] (ExBio), CD45RB [MT4] (BD Pharmingen), and pan-CD45 [HI30] (BD Pharmingen). Staining of CD45 isoforms was performed in separate tubes to avoid competition for antibody binding. The following antibodies were used for Western blot: anti-CTCF [D31H2] (Cell signaling), anti-Tet1 [GT949] (GeneTex, for CHO blots), anti-TET1 [61443] (active Motif, for human blots), anti-Tet2 [179-050] (Diagenode), anti-p65 RelA [610869] (BD Bioscience), anti-p84 [5E10] (GeneTex), and anti-FLAG (Sigma). Cells were lysed in RIPA buffer and anti-RelA and anti-p84 immunoblotting served as loading controls for protein levels. The following antibodies were used for ChIP and MedIP: anti-CTCF [D31H2] (Cell Signaling), anti-RNA polymerase II [05-623] (Millipore), anti-Tet1 [09-872] (Millipore), anti-5-methylcytosine [A-1014 (Epigentek); 006-500 (Diagenode) for MedIP-seq], anti-5-hydroxymethylcytosine [A-1018 (Epigentek); 39769 (Active Motif) for MedIP-seq], anti-5-carboxylcytosine [RM24 1-A3] (Abcam), normal rabbit IgG [2729] (Cell Signaling), and normal mouse IgG [12-371] (Millipore).

RT-PCR

RNA was isolated with the Qiagen RNeasy Mini Kit, and reverse transcription was performed with SuperScript III (Invitrogen) according to the manufacturer's instructions. Quantitative RT-PCR was performed in the presence of SYBR green reagent (Roche), and amplification was performed on the Roche 480 or 96 Light Cycler. Cycle thresholds were normalized to *GAPDH*, *RPS16*, and *YES*, or surrounding exon levels as indicated. Radioactive PCR was performed in a 25- μ l reaction containing 0.5 μ Ci dCTP α - 32 P (PerkinElmer) for 21–24 cycles. PCR products were separated on a native 6% polyacrylamide gel at 1 $\times$ TBE. The gel was directly exposed to a phosphor screen and analyzed with a PhosphorImager Storm860 and ImageQuant software (v5.2, Molecular Dynamics).

Bisulfite sequencing

Genomic DNA isolation was performed with the Zymo research Quick gDNA Midiprep kit (D3100), according to the manufacturer's instructions. Sodium bisulfate conversion of denatured genomic DNA was achieved with the EpiTect Bisulfite Kit (Qiagen), according to the manufacturer's instructions. Eluted DNA was amplified using methylation-specific primers as indicated in Table EV7. For bisulfite pyrosequencing, purified de-cross-linked DNA isolated from BL-E5(+) cell CTCF-ChIP and input was converted with the EpiTect Kit (Qiagen) and sequenced at the NCI Advanced Technology Research Facility (ATRF).

Primers

See Table EV7 for primer and oligo sequences.

Southern blotting

Ten micrograms of genomic DNA was digested with BsaHI and BglII in the presence of 4 mM spermidine overnight. Digested DNA

samples were resolved through 0.8% agarose gel electrophoresis and sequentially treated with 0.25% HCl, alkali solution, and neutralization solution. DNA transfer to nylon membrane was performed overnight with a Whatman TurboBlotter transfer system (G&E). The membrane was UV cross-linked and then prehybridized for 2 h followed by overnight hybridization with dCTP α - 32 P-labeled probe (Prime-It II Random Primer Labeling Kit, Stratagene). The membrane was exposed to a phosphor screen for 4 h.

Chromatin immunoprecipitation (ChIP)

ChIP in primary and lymphocyte cell lines was performed as previously described (Shukla *et al*, 2011). When working with CHO cells transfected with the *CD45* minigene, CTCF-ChIP and RNA polymerase II-ChIP were performed with the following optimized protocol. Three million cells were cross-linked for 10 min in 1% formaldehyde at room temperature (RT), and quenched by adding glycine to a final concentration of 0.125 M for 5 min at RT. Cells were washed in PBS and resuspended in buffer containing 50 mM HEPES-KOH (pH 7.5), 140 mM NaCl, 1 mM EDTA, 1.0% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, and protease inhibitors (Thermo Scientific). Sonication of DNA was performed in an ultra sonicator water bath (Bioruptor) using 15 cycles of 30 s "on" and 30 s "off". Samples were centrifuged at 10,000 g for 8 min at 4°C. Three micrograms of chromatin was immunoprecipitated by adding the anti-CTCF, anti-RNA polymerase II, or normal rabbit IgG followed by overnight incubation at 4°C. Thirty microliters of Magna ChIP™ Protein A Magnetic Beads (Millipore) were added and incubated for 1 h at 4°C. Beads were washed sequentially in low salt (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), high salt (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), LiCl buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.25 M LiCl, 1% NP-40, 1% Na-deoxycholate), and TE buffer. Beads were eluted in 110 μ l elution buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS, 50 mM NaHCO₃) and treated with 1 μ l RNase A (1 mg/ml Ambion) at 37°C for 30 min. Cross-linking was reversed and proteins were degraded by the addition of 1 μ l proteinase K (20 mg/ml Ambion) and incubation at 65°C for 4 h. Eluted DNA was purified with Qiaquick PCR purification (Qiagen), according to the manufacturer instructions. Immunoprecipitated DNA and 5% input DNA were analyzed by SYBR-Green qPCR. Relative detection was calculated according to the formula $2^{(C_t^{\text{input}} - C_t^{\text{IP}})}$.

Methylated DNA immunoprecipitation (MedIP)

MedIP in primary and lymphocyte cell lines was performed as previously described (Shukla *et al*, 2011). When working with CHO cells transfected with the *CD45* minigene, MedIPs were performed with the following optimized protocol. Genomic DNA was fragmented through restriction enzyme digestion with Bsu36I and PpuMI (NEB). IP buffer was added to 1 μ g of digested DNA to a final volume of 100 μ l. DNA was denatured for 10 min at 95°C and immediately transferred to an ice bath for 10 min. To each IP reaction, 2 μ g of antibody targeted against 5-methylcytosine (Epigentek, A-1014) or 5-hydroxymethylcytosine (Epigentek, A-1018) was added and incubated overnight at 4°C with shaking. One percent of total DNA was kept for input. Following incubation, 25 μ l of Magna ChIP Protein A

Magnetic Beads (Millipore) were added and further incubated for 1.5 h at 4°C. Beads were washed three times with 500 µl IP buffer and eluted in elution buffer supplemented with proteinase K and incubated for 3 h at 55°C with rigorous shaking. Supernatants were purified with the Qiaquick PCR purification kit (Qiagen). Quantitative analysis of methylated regions was assessed through qPCR with SYBR-Green (Roche). Relative detection was calculated according to the formula $2^{(C_i^{\text{input}} - C_i^{\text{IP}})}$. For relative minigene methylation (Fig 4B), percent methylation was determined through $\left[\frac{2^{(C_i^{\text{input}} - C_i^{\text{hmC or mC IP}})}}{2^{(C_i^{\text{input}} - C_i^{\text{hmC IP}})} + 2^{(C_i^{\text{input}} - C_i^{\text{mC IP}})}} \right] \times 100$.

Preparation of nuclear extracts

To generate nuclear extracts, naïve and activated CD4⁺ T cells were pelleted, washed once in cold PBS, resuspended in 100 µl hypotonic lysis buffer (10 mM HEPES pH 7.9, 10 mM KCl, 0.1 mM EDTA) supplemented with 2 mM Na₃VO₄ and 1× Halt protease inhibitor cocktail (Thermo Scientific), and placed on ice for 10 min. NP-40 was then added to a final concentration of 0.4% and vortexed vigorously for 15 s. Samples were centrifuged for 6 min at 3,300 g at 4°C, and pellets were resuspended in 60 µl nuclear lysis buffer (20 mM HEPES pH 7.9, 0.4 M NaCl, 1 mM EDTA, 10% glycerol) supplemented with 2 mM Na₃VO₄ and 1× Halt protease inhibitor cocktail. Lysates were passed through a 23-gauge needle 4 times and left on ice for 15 min. After vortexing for 15 s, samples were centrifuged at maximum speed (20,000 g) for 5 min at 4°C and supernatants containing nuclear proteins were transferred to fresh tubes. Protein concentrations were determined with the Pierce BCA Protein Assay (Thermo Scientific). About 7.5–15 µg of nuclear lysates was resolved with a 3–8% NuPage Tris-Acetate gel (Invitrogen) and transferred to PVDF.

CTCF purification

HEK293T cells were transfected with plasmid encoding 3xFLAG-tagged CTCF. Forty-eight hours after transfection, cells were washed twice with ice-cold 1× PBS and lysed (50 mM Tris-HCl pH 7.4, 300 mM NaCl, 1 mM EDTA, 1% Triton X-100, 4 µM ZnCl₂, 2 mM Na₃VO₄, 1× Halt protease inhibitor cocktail [Thermo Scientific]) for 20 min with rotation at 4°C. Lysates were subjected to 20 cycles of sonication for 30 s at 4°C in a bioruptor (Diagenode) and centrifuged at 15,000 rpm for 15 min at 4°C. Anti-FLAG-M2-coated agarose beads (Sigma) were added to the supernatants containing soluble proteins and incubated with mixing at 4°C for 3 h. Beads were washed 3× with 10 bead volumes of ice-cold wash buffer (50 mM Tris-HCl pH 7.4, 300 mM NaCl, 1 mM EDTA, 4 µM ZnCl₂) supplemented with 1× Halt protease inhibitor cocktail. 3xFLAG-CTCF was eluted through 20-min incubation with two bead volumes of elution buffer (125 mM Tris-HCl pH 7.4, 250 mM NaCl) supplemented with 100 µg/ml of 3X-FLAG peptide (Sigma). Beads were removed by centrifugation at 15,000 rpm for 2 min at 4°C, and supernatant containing purified CTCF was flash-frozen in storage buffer (100 mM Tris-HCl pH 7.4, 200 mM NaCl, 4 mM DTT, 10% glycerol).

Electrophoretic mobility shift assay (EMSA)

Double-stranded EMSA probes corresponding to CD45 exon 5 DNA were generated through PCR using modified dCTP (TriLink) or

commercially synthesized (unmethylated or CpG methylated) (GeneLink). See Table EV7 for probe sequence information. To prepare the commercial probes for EMSA, complementary ssDNA probes were resuspended in annealing buffer (10 mM Tris-HCl pH 8, 50 mM NaCl, 1 mM EDTA), denatured for 10 min at 99°C followed by a linear cool down to 25°C at 1°C/min. All probes were radiolabeled using T4 polynucleotide kinase (NEB) for 15 min at 37°C in the presence of 100 µCi of dATP γ-³²P (3,000 Ci/mmol, PerkinElmer). Free dATP γ-³²P was removed by a G50 (GE) column. For the commercial probes, a RecJ (NEB) digestion step was added to remove remaining traces of ssDNA. Prepared probes were quantified through a scintillation counter (Beckman) and confirmed through native gel electrophoresis. For EMSA, radiolabeled probes were incubated on ice for 30 min with 0.05 to 0.5 µg CTCF (3xFLAG-CTCF from 293T, or recombinant GST-CTCF, Abnova, as indicated) in binding buffer (10 mM Tris-HCl pH 7.5, 100 mM KCl, 0.1% NP-40, 0.1 mM ZnSO₄²⁻, 2 mM DTT, 0.1 mg/ml BSA, 1 µg/ml poly-dIdC) in the presence or absence of 10-fold excess of cold competitor or 1 µg of anti-CTCF antibody (Cell signaling). Samples were then resolved through a 6% polyacrylamide gel. Following electrophoresis, the gel was exposed to a phosphor screen and analyzed with PhosphorImager Storm860 and ImageQuant software (v5.2, Molecular Dynamics).

RNA-Seq

Library construction and sequencing

Total RNA isolated with the Qiagen RNeasy kit was used for library construction with the Illumina TruSeq RNA Protocol RS-930-2001 protocol (Illumina, San Diego, CA), generating a poly-A⁺ sequencing library via poly-T oligo-attached magnetic beads. Libraries were sequenced on one lane (two indexed barcodes) of a HiSeq2000 instrument with TruSeq v3.0 chemistry (Illumina, San Diego, CA) for 101 cycles in paired-end mode.

Alignments and feature quantification

Reads were mapped to the human reference genome assembly (hg19) using TopHat v.2.0.11 (Trapnell *et al*, 2010) with Bowtie2 v.2.2.1 (Langmead & Salzberg, 2012). Parameters used were unique mapping only (-g 1) and supplying the UCSC genome annotation for hg19 from iGenomes (Illumina, San Diego, CA). A summary of experiments and read depth is provided in Table EV1.

Differential splicing was calculated using MISO (Katz *et al*, 2010). Splicing event annotation was retrieved from the MISO website (Human genome (hg19) alternative events v2.0, http://genes.mit.edu/burgelab/miso/annotations/ver2/miso_annotations_hg19_v2.zip). A conservative cutoff of Bayes factor ≥ 10 and ΔPSI absolute value ≥ 0.10 was used to classify alternative exons as differentially spliced. For convenience, short identifiers for each unique skipped exon event were generated in the form ("MISO:SE:xxx"). Alternative exons and their respective flanking exons were partitioned into separate coordinate files using an in-house perl script. Complete MISO results for skipped exons are provided in Table EV2. Any given exon may be a part of > 1 annotated splicing event. To arrive at a single splicing measurement for each unique exon, we generated a collapsed or "flattened" version of the splicing results (Table EV2). For each unique exon, we selected all splicing events where the exon is defined as alternative. From these events, we

subsetted events where the Bayes factor was ≥ 10 . From this subset, we chose the event with the highest magnitude splicing difference as the representative measurement for the exon. If no event had a Bayes factor ≥ 10 , the event with the highest Bayes factor was used as the representative measurement.

Exon centric analyses

Classification of constitutive exons is based on the *annotate_junctions* program within the Spanki program (Sturgill *et al*, 2013), which classifies gene models into relevant splicing categories. Expressed constitutive exons were defined as internal exons (excluding first and last exons) of constitutively spliced genes (those having > 1 intron and no annotated alternative splicing); further subsetted to expressed genes by selecting exons within genes with a minimum expression level of 10 FPKM. The “all constitutive exon” group was defined as all constitutive exons defined by the HexEvent database (Busch & Hertel, 2013).

ChIP-seq

Library construction and sequencing

Sequencing libraries were constructed from DNA samples with the Illumina TruSeq V3 library construction protocol. Libraries were multiplexed four samples per lane on six lanes of an Illumina HiSeq2000 instrument using TruSeq V3 chemistry, and sequenced for 50 cycles in single-end mode.

Read mapping and peak calling

Reads were mapped to the human genome with Bowtie2 v.2.1.0 (Langmead & Salzberg, 2012) with default parameters, producing output in SAM format. The default parameters allowed at most one alignment per read. The reference assembly used was hg19 from UCSC, via Igenomes (Illumina, San Diego CA). SAM files were converted to BAM format and indexed using Samtools (Li *et al*, 2009). A summary of mapped reads is provided in Table EV1. Enrichment in each ChIP sample relative to background was calculated using MACS2 (v. 2.1.0.20140616) “callpeak” with default parameters (Feng *et al*, 2012). Non-specific IgG ChIP was used as background for CTCF-ChIP, and input was used as background for MedIP. Complete results for CTCF-ChIP analysis, including coordinates and signal quantifications, are in Table EV3.

We defined common CTCF-bound regions to make comparisons of CTCF binding and methylation between cell types. Peak sizes were slightly larger in activated cells, likely due to their more accessible chromatin. The average of the median peak size in the two cell types was ~ 300 base pairs (bps), so we considered this the optimum width for common CTCF-binding site (CBS). We examined the distance of each peak in one cell type to its nearest neighbor in the other cell type. The median distance was 20 bp, reaching an asymptote at 150 bp. We therefore chose to merge peak calls into a common CBS when their peak summits were within 150 bp, with start and end positions ± 150 bp from the midpoint between their summits. In this manner, we defined 28,152 common CBSs of 301 bp, or ± 150 bp from a single-base summit.

MedIP data analysis

In addition to peak calling using MACS2, MedIP data were processed by alternative specialized methods to identify differential

methylation and hydroxymethylation. We generated sequencing depth and input normalized genome coverage tracks in log2 scale for each MedIP experiment in bigwig format with deepTools (Ramirez *et al*, 2014). We next extracted the mean enrichment value within CTCF-binding sites and classified differential methylation within features as mean enrichment values $> 10\%$ different (in log2 scale, a difference between ratios of > 0.1375). Additionally, we used MEDIPS (Lienhard *et al*, 2014) to identify sites with zero methylation. Sites with an RPKM = 0 were classified as unmethylated and excluded from directional analysis. Results for 5hmC and 5mC MedIP analysis are in Table EV5.

Processed data from 5caC MedIP-Seq experiments were downloaded from GEO in wiggle format (GSE46111) (Shen *et al*, 2013). Results in the shTDG sample are presented in Appendix Fig S7. CTCF-binding sites defined in the same tissue (Mouse ES-E14 cells) were obtained from ENCODE (Accession ENCSCR362VNF) (Yue *et al*, 2014). Binned MedIP-Seq signal within CTCF sites was calculated with deepTools (Ramirez *et al*, 2014).

Correlation analysis of ChIP results and splicing

Proximity analysis between exons and CTCF sites was performed using the Bedtools windowBed function (Quinlan & Hall, 2010). We considered CTCF sites and exons proximal when they overlapped or were within 1.5 kb. Our methylation analysis focused on unique CBS, so that sites were not counted more than once. We noted that CBS and their proximal cassette exons were more frequently one-to-one than constitutive exons. In other words, for constitutive exon-proximal sites, the number of unique CBS (804) is much lower than the number of exons proximal to them (1,574) because of one-to-many relationships within a subset of exon-dense constitutively spliced genes. All analyses used empirically determined CTCF-binding sites from our ChIP-Seq experiments.

Data deposition

All the high-throughput sequencing datasets used in this study were publicly available in the Gene Expression Omnibus (GEO) under accession number GSE74850.

Expanded View for this article is available online.

Acknowledgements

We thank K. Lynch, T. Misteli, and P. Oberdoerffer for critical reading of this manuscript. We also thank A. Rao for helpful discussions and feedback. This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD (<http://biowulf.nih.gov>). This work is supported by the Intramural Research Program of NIH, the National Cancer Institute, the Center for Cancer Research.

Author contributions

MAB, RJM, MT, SS, NH, and SO designed research. RJM, MAB, MT, KKN, GV, SS, NH, and MFP performed experiments and analyzed data. DS performed computational analysis and wrote the paper. SO supervised the project and wrote the paper.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Agirre E, Bellora N, Allo M, Pages A, Bertucci P, Kornblihtt AR, Eyraes E (2015) A chromatin code for alternative splicing involving a putative association between CTCF and HP1alpha proteins. *BMC Biol* 13: 31
- Allo M, Agirre E, Bessonov S, Bertucci P, Gomez Acuna L, Buggiano V, Bellora N, Singh B, Petrillo E, Blaustein M, Minana B, Dujardin G, Pozzi B, Pelisch F, Bechara E, Agafonov DE, Srebrow A, Luhrmann R, Valcarcel J, Eyraes E et al (2014) Argonaute-1 binds transcriptional enhancers and controls constitutive and alternative splicing in human cells. *Proc Natl Acad Sci USA* 111: 15622–15629
- Ameyar-Zazoua M, Rachez C, Souidi M, Robin P, Fritsch L, Young R, Morozova N, Fenouil R, Descostes N, Andrau JC, Mathieu J, Hamiche A, Ait-Si-Ali S, Muchardt C, Batsche E, Harel-Bellan A (2012) Argonaute proteins couple chromatin silencing to alternative splicing. *Nat Struct Mol Biol* 19: 998–1004
- Andersson R, Enroth S, Rada-Iglesias A, Wadelius C, Komorowski J (2009) Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res* 19: 1732–1741
- Battle-Lopez A, Cortiguera MG, Rosa-Garrido M, Blanco R, del Cerro E, Torrono V, Wagner SD, Delgado MD (2015) Novel CTCF binding at a site in exon1A of BCL6 is associated with active histone marks and a transcriptionally active locus. *Oncogene* 34: 246–256
- Bell AC, Felsenfeld G (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* 405: 482–485
- Bentley DL (2014) Coupling mRNA processing with transcription in time and space. *Nat Rev Genet* 15: 163–175
- Black DL (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 72: 291–336
- Busch A, Hertel KJ (2013) HEXEvent: a database of human EXon splicing events. *Nucleic Acids Res* 41: D118–D124
- Cho V, Mei Y, Sanny A, Chan S, Enders A, Bertram EM, Tan A, Goodnow CC, Andrews TD (2014) The RNA-binding protein hnRNPL induces a T cell alternative splicing program delineated by differential intron retention in polyadenylated RNA. *Genome Biol* 15: R26
- Chodavarapu RK, Feng S, Bernatavichute YV, Chen PY, Stroud H, Yu Y, Hetzel JA, Kuo F, Kim J, Cokus SJ, Casero D, Bernal M, Huijser P, Clark AT, Kramer U, Merchant SS, Zhang X, Jacobsen SE, Pellegrini M (2010) Relationship between nucleosome positioning and DNA methylation. *Nature* 466: 388–392
- Choi JK (2010) Contrasting chromatin organization of CpG islands and exons in the human genome. *Genome Biol* 11: R70
- Churchman LS, Weissman JS (2011) Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 469: 368–373
- Close P, East P, Dirac-Svejstrup AB, Hartmann H, Heron M, Maslen S, Chariot A, Soding J, Skehel M, Svejstrup JQ (2012) DBIRD complex integrates alternative mRNA splicing with RNA polymerase II transcript elongation. *Nature* 484: 386–389
- Deaton AM, Webb S, Kerr AR, Illingworth RS, Guy J, Andrews R, Bird A (2011) Cell type-specific DNA methylation at intragenic CpG islands in the immune system. *Genome Res* 21: 1074–1086
- Dubois-Chevalier J, Oger F, Dehondt H, Firmin FF, Gheeraert C, Staels B, Lefebvre P, Eeckhoutte J (2014) A dynamic CTCF chromatin binding landscape promotes DNA hydroxymethylation and transcriptional induction of adipocyte differentiation. *Nucleic Acids Res* 42: 10943–10959
- Dujardin G, Lafaille C, de la Mata M, Marasco LE, Munoz MJ, Le Jossic-Corcos C, Corcos L, Kornblihtt AR (2014) How slow RNA polymerase II elongation favors alternative exon skipping. *Mol Cell* 54: 683–690
- Feldmann A, Ivanek R, Murr R, Gaidatzis D, Burger L, Schubeler D (2013) Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLoS Genet* 9: e1003994
- Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME, Ukomadu C, Sadler KC, Pradhan S, Pellegrini M, Jacobsen SE (2010) Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci USA* 107: 8689–8694
- Feng J, Liu T, Qin B, Zhang Y, Liu XS (2012) Identifying ChIP-seq enrichment using MACS. *Nat Protoc* 7: 1728–1740
- Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, Marques CJ, Andrews S, Reik W (2011) Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* 473: 398–402
- Gelfman S, Cohen N, Yearim A, Ast G (2013) DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure. *Genome Res* 23: 789–799
- Haque N, Oberdoerffer S (2014) Chromatin and splicing. *Methods Mol Biol* 1126: 97–113
- He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, Li X, Dai Q, Song CX, Zhang K, He C, Xu GL (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333: 1303–1307
- Hermiston ML, Xu Z, Weiss A (2003) CD45: a critical regulator of signaling thresholds in immune cells. *Annu Rev Immunol* 21: 107–137
- Hnilicova J, Hozeifi S, Duskova E, Icha J, Tomankova T, Stanek D (2011) Histone deacetylase activity modulates alternative splicing. *PLoS One* 6: e16727
- Hodges E, Smith AD, Kendall J, Xuan Z, Ravi K, Rooks M, Zhang MQ, Ye K, Bhattacharjee A, Brizuela L, McCombie WR, Wigler M, Hannon GJ, Hicks JB (2009) High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. *Genome Res* 19: 1593–1605
- Hon GC, Song CX, Du T, Jin F, Selvaraj S, Lee AY, Yen CA, Ye Z, Mao SQ, Wang BA, Kuan S, Edsall LE, Zhao BS, Xu GL, He C, Ren B (2014) 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Mol Cell* 56: 286–297
- Huang Y, Chavez L, Chang X, Wang X, Pastor WA, Kang J, Zepeda-Martinez JA, Pape UJ, Jacobsen SE, Peters B, Rao A (2014) Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proc Natl Acad Sci USA* 111: 1361–1366
- Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333: 1300–1303
- Kafasla P, Skliris A, Kontoyiannis DL (2014) Post-transcriptional coordination of immunological responses by RNA-binding proteins. *Nat Immunol* 15: 492–502
- Katz Y, Wang ET, Airolidi EM, Burge CB (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods* 7: 1009–1015
- Khare T, Pai S, Koncevicius K, Pal M, Kriukiene E, Liutkeviciute Z, Irimia M, Jia P, Ptak C, Xia M, Tice R, Tochigi M, Morera S, Nazarians A, Belsham D, Wong AH, Blencowe BJ, Wang SC, Kapranov P, Kustra R et al (2012) 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. *Nat Struct Mol Biol* 19: 1037–1043
- Klose RJ, Bird AP (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* 31: 89–97
- Ko M, An J, Bandukwala HS, Chavez L, Aijo T, Pastor WA, Segal MF, Li H, Koh KP, Lahdesmaki H, Hogan PG, Aravind L, Rao A (2013) Modulation of TET2

- expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* 497: 122–126
- Kolasinska-Zwiercz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J (2009) Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat Genet* 41: 376–381
- Kwak H, Fuda NJ, Core LJ, Lis JT (2013) Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* 339: 950–953
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9: 357–359
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078–2079
- Li E, Zhang Y (2014) DNA methylation in mammals. *Cold Spring Harb Perspect Biol* 6: a019133
- Li-Byarlay H, Li Y, Stroud H, Feng S, Newman TC, Kaneda M, Hou KK, Worley KC, Elisk CG, Wickline SA, Jacobsen SE, Ma J, Robinson GE (2013) RNA interference knockdown of DNA methyl-transferase 3 affects gene alternative splicing in the honey bee. *Proc Natl Acad Sci USA* 110: 12750–12755
- Lienhard M, Grimm C, Morkel M, Herwig R, Chavez L (2014) MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. *Bioinformatics* 30: 284–286
- Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462: 315–322
- Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T (2010) Regulation of alternative splicing by histone modifications. *Science* 327: 996–1000
- Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, Maleszka R (2010) The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol* 8: e1000506
- Martinez NM, Lynch KW (2013) Control of alternative splicing in immune responses: many regulators, many predictions, much still to learn. *Immunol Rev* 253: 216–236
- Martinez NM, Agosto L, Qiu J, Mallory MJ, Gazzara MR, Barash Y, Fu XD, Lynch KW (2015) Widespread JNK-dependent alternative splicing induces a positive feedback loop through CELF2-mediated regulation of MKK7 during T-cell activation. *Genes Dev* 29: 2054–2066
- de la Mata M, Alonso CR, Kadener S, Fededa JP, Blaustein M, Pelisch F, Cramer P, Bentley D, Kornblihtt AR (2003) A slow RNA polymerase II affects alternative splicing in vivo. *Mol Cell* 12: 525–532
- Matlin AJ, Clark F, Smith CW (2005) Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol* 6: 386–398
- Maunakea AK, Chepelev I, Cui K, Zhao K (2013) Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res* 23: 1256–1269
- Motta-Mena LB, Heyd F, Lynch KW (2010) Context-dependent regulatory mechanism of the splicing factor hnRNP L. *Mol Cell* 37: 223–234
- Munoz MJ, Perez Santangelo MS, Paronetto MP, de la Mata M, Pelisch F, Boireau S, Glover-Cutter K, Ben-Dov C, Blaustein M, Lozano JJ, Bird G, Bentley D, Bertrand E, Kornblihtt AR (2009) DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. *Cell* 137: 708–720
- Nguyen P, Cui H, Bisht KS, Sun L, Patel K, Lee RS, Kugoh H, Oshimura M, Feinberg AP, Gius D (2008) CTCFL/BORIS is a methylation-independent DNA-binding protein that preferentially binds to the paternal H19 differentially methylated region. *Cancer Res* 68: 5546–5551
- Oberdoerffer S, Moita LF, Neems D, Freitas RP, Hacohen N, Rao A (2008) Regulation of CD45 alternative splicing by heterogeneous ribonucleoprotein, hnRNPL. *Science* 321: 686–691
- Ong CT, Corces VG (2014) CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet* 15: 234–246
- Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 40: 1413–1415
- Pandya-Jones A, Black DL (2009) Co-transcriptional splicing of constitutive and alternative exons. *RNA* 15: 1896–1908
- Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, McLoughlin EM, Brudno Y, Mahapatra S, Kapranov P, Tahiliani M, Daley GQ, Liu XS, Ecker JR, Milos PM, Agarwal S, Rao A (2011) Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* 473: 394–397
- Patrick KL, Ryan CJ, Xu J, Lipp JJ, Nissen KE, Roguev A, Shales M, Krogan NJ, Guthrie C (2015) Genetic interaction mapping reveals a role for the SWI/SNF nucleosome remodeler in spliceosome activation in fission yeast. *PLoS Genet* 11: e1005074
- Phillips JE, Corces VG (2009) CTCF: master weaver of the genome. *Cell* 137: 1194–1211
- Pradeepa MM, Sutherland HG, Ule J, Grimes GR, Bickmore WA (2012) Psp1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. *PLoS Genet* 8: e1002717
- Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841–842
- Ramirez F, Dunder F, Diehl S, Gruning BA, Manke T (2014) deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 42: W187–W191
- Renbaum P, Abrahamove D, Fainsod A, Wilson GG, Rottem S, Razin A (1990) Cloning, characterization, and expression in *Escherichia coli* of the gene coding for the CpG DNA methylase from *Spiroplasma* sp. strain MQ1 (M.SssI). *Nucleic Acids Res* 18: 1145–1152
- Saint-Andre V, Batsche E, Rachez C, Muchardt C (2011) Histone H3 lysine 9 trimethylation and HP1gamma favor inclusion of alternative exons. *Nat Struct Mol Biol* 18: 337–344
- Schor IE, Rascovan N, Pelisch F, Allo M, Kornblihtt AR (2009) Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. *Proc Natl Acad Sci USA* 106: 4325–4330
- Schor IE, Fiszbein A, Petrillo E, Kornblihtt AR (2013) Intragenic epigenetic changes modulate NCAM alternative splicing in neuronal differentiation. *EMBO J* 32: 2264–2274
- Schwartz S, Meshorer E, Ast G (2009) Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol* 16: 990–995
- Shen L, Wu H, Diep D, Yamaguchi S, D'Alessio AC, Fung HL, Zhang K, Zhang Y (2013) Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* 153: 692–706
- Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R, Oberdoerffer S (2011) CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 479: 74–79
- Sims RJ 3rd, Millhouse S, Chen CF, Lewis BA, Erdjument-Bromage H, Tempst P, Manley JL, Reinberg D (2007) Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. *Mol Cell* 28: 665–676
- Spies N, Nielsen CB, Padgett RA, Burge CB (2009) Biased chromatin signatures around polyadenylation sites and exons. *Mol Cell* 36: 245–254

- Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PW, Bauer C, Munzel M, Wagner M, Muller M, Khan F, Eberl HC, Mensinga A, Brinkman AB, Lephikov K, Muller U, Walter J, Boelens R, van Ingen H, Leonhardt H, Carell T *et al* (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* 152: 1146–1159
- Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D, Tiwari VK, Schubeler D (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480: 490–495
- Sturgill D, Malone JH, Sun X, Smith HE, Rabinow L, Samson ML, Oliver B (2013) Design of RNA splicing analysis null models for post hoc filtering of Drosophila head RNA-Seq data with the splicing analysis kit (Spanki). *BMC Bioinformatics* 14: 320
- Sun Z, Terragni J, Borgaro JC, Liu Y, Yu L, Guan S, Wang H, Sun D, Cheng X, Zhu Z, Pradhan S, Zheng Y (2013) High-resolution enzymatic mapping of genomic 5-hydroxymethylcytosine in mouse embryonic stem cells. *Cell Rep* 3: 567–576
- Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324: 930–935
- Teif VB, Beshnova DA, Vainshtein Y, Marth C, Mallm JP, Hofer T, Rippe K (2014) Nucleosome repositioning links DNA (de)methylation and differential CTCF binding during stem cell development. *Genome Res* 24: 1285–1295
- Tilgner H, Nikolaou C, Althammer S, Sammeth M, Beato M, Valcarcel J, Guigo R (2009) Nucleosome positioning as a determinant of exon recognition. *Nat Struct Mol Biol* 16: 996–1001
- Tilgner H, Knowles DG, Johnson R, Davis CA, Chakraborty S, Djebali S, Curado J, Snyder M, Gingeras TR, Guigo R (2012) Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res* 22: 1616–1625
- Topp JD, Jackson J, Melton AA, Lynch KW (2008) A cell-based screen for splicing regulators identifies hnRNP LL as a distinct signal-induced repressor of CD45 variable exon 4. *RNA* 14: 2038–2049
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28: 511–515
- Tsagaratou A, Aijo T, Lio CW, Yue X, Huang Y, Jacobsen SE, Lahdesmaki H, Rao A (2014) Dissecting the dynamic changes of 5-hydroxymethylcytosine in T-cell development and differentiation. *Proc Natl Acad Sci USA* 111: E3306–E3315
- Wada Y, Ohta Y, Xu M, Tsutsumi S, Minami T, Inoue K, Komura D, Kitakami J, Oshida N, Papantonis A, Izumi A, Kobayashi M, Meguro H, Kanki Y, Mimura I, Yamamoto K, Mataka C, Hamakubo T, Shirahige K, Aburatani H *et al* (2009) A wave of nascent transcription on activated human genes. *Proc Natl Acad Sci USA* 106: 18357–18361
- Wang ET, Sandberg R, Luo S, Khrebukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* 456: 470–476
- Wang L, Zhou Y, Xu L, Xiao R, Lu X, Chen L, Chong J, Li H, He C, Fu XD, Wang D (2015) Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. *Nature* 523: 621–625
- Weber CM, Ramachandran S, Henikoff S (2014) Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. *Mol Cell* 53: 819–830
- Wu H, D'Alessio AC, Ito S, Wang Z, Cui K, Zhao K, Sun YE, Zhang Y (2011) Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev* 25: 679–684
- Wu H, Zhang Y (2014) Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell* 156: 45–68
- Yearim A, Gelfman S, Shayevitch R, Melcer S, Glaich O, Mallm JP, Nissim-Rafin M, Cohen AH, Rippe K, Meshorer E, Ast G (2015) HP1 is involved in regulating the global impact of DNA methylation on alternative splicing. *Cell Rep* 10: 1122–1134
- Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, Sandstrom R, Ma Z, Davis C, Pope BD, Shen Y, Pervouchine DD, Djebali S, Thurman RE, Kaul R, Rynes E, Kirilusha A, Marinov GK, Williams BA, Trout D *et al* (2014) A comparative encyclopedia of DNA elements in the mouse genome. *Nature* 515: 355–364
- Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328: 916–919