RNA BIOLOGY
https://doi.org/10.1080/15476286.2020.1796052

Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

Ctcf haploinsufficiency mediates intron retention in a tissue-specific manner

Adel B Alharbi [a,b,c,d], Ulf Schmitz [a,b,c], Amy D Marshall[a], Darya Vanichkina [a,c,e], Rajini Nagarajah[a], Melissa Vellozzi[a,b], Justin JL Wong [c,f], Charles G Bailey [a,c], and John EJ Rasko [a,c,g]

[a]Gene & Stem Cell Therapy Program Centenary Institute, The University of Sydney, Camperdown, Australia; [b]Computational BioMedicine Laboratory Centenary Institute, The University of Sydney, Camperdown, Australia; [c]Faculty of Medicine and Health, The University of Sydney, Camperdown, Australia; [d]Department of Laboratory Medicine, Faculty of Applied Medical Sciences, Umm Al-Qura University, Makkah, Saudi Arabia; [e]Sydney Informatics Hub, University of Sydney, Darlington, Australia; [f]Epigenetics and RNA Biology Program Centenary Institute, The University of Sydney, Camperdown, Australia; [g]Cell & Molecular Therapies, Royal Prince Alfred Hospital, Camperdown, Australia

ABSTRACT

CTCF is a master regulator of gene transcription and chromatin organisation with occupancy at thousands of DNA target sites genome-wide. While CTCF is essential for cell survival, CTCF haploinsufficiency is associated with tumour development and hypermethylation. Increasing evidence demonstrates CTCF as a key player in several mechanisms regulating alternative splicing (AS), however, the genome-wide impact of *Ctcf* dosage on AS has not been investigated.

We examined the effect of *Ctcf* haploinsufficiency on gene expression and AS in five tissues from *Ctcf* hemizygous (*Ctcf*$^{+/-}$) mice. Reduced *Ctcf* levels caused distinct tissue-specific differences in gene expression and AS in all tissues. An increase in intron retention (IR) was observed in *Ctcf*$^{+/-}$ liver and kidney. In liver, this specifically impacted genes associated with cytoskeletal organisation, splicing and metabolism. Strikingly, most differentially retained introns were short, with a high GC content and enriched in Ctcf binding sites in their proximal upstream genomic region. This study provides new insights into the effects of *CTCF* haploinsufficiency on organ transcriptomes and the role of CTCF in AS regulation.

ARTICLE HISTORY
Received 5 February 2020
Revised 28 June 2020
Accepted 9 July 2020

KEYWORDS
Alternative splicing; CTCF; exon skipping; gene expression; haploinsufficiency; intron retention

Introduction

CCCTC-binding factor (CTCF) is a highly conserved, multivalent, 11-zinc finger DNA-and RNA-binding protein, which occupies thousands of conserved target sites distributed across vertebrate genomes and co-ordinates topologically associating domain (TAD) formation [1–5]. Numerous studies have shown that CTCF regulates diverse biological functions including transcriptional activation and repression, insulation, high-order chromatin organisation, X-chromosome inactivation, pre-mRNA splicing, and DNA methylation [6,7]. CTCF binding occurs in a tissue-specific manner, and the binding sites themselves are highly conserved among different species [2–5,8,9]. Over 30% and 50% of all CTCF binding sites are located in intronic and intergenic regions, respectively [2,5,9].

The ubiquitously expressed transcription factor CTCF exhibits differential expression in mammalian tissues [8,10,11], which require specific CTCF levels for growth, differentiation and development [12–14]. Homozygous deletion of *Ctcf* causes early embryonic lethality in mice [15], while mice harbouring a hemizygous deletion of *Ctcf* (*Ctcf*$^{+/-}$) exhibit delayed post-natal growth and development [15,16]. Sustained *Ctcf* haploinsufficiency in mice has various pathophysiological implications including spontaneous widespread tumour formation in diverse tissues as well as genome-wide aberrant hypermethylation [17]. Consistent with that, we have shown that CTCF acts as a haploinsufficient tumour suppressor [16,18,19]. Nevertheless, the underlying link between *Ctcf* haploinsufficiency and spontaneous tumour formation remains to be fully elucidated.

Compelling evidence has linked CTCF to alternative splicing (AS) regulation due to its direct or indirect role in modulating splicing decisions via complex mechanisms including RNA Polymerase II (Pol II) elongation, transcriptional regulation of splicing factors, DNA methylation and chromatin organisation [20–25]. Despite its potential importance, the impact of altered CTCF dosage on AS has not been investigated at a global scale. More than 90% of human multi-exonic genes undergo AS in a tissue-specific manner [26,27]. While normal AS contributes to transcriptome and proteome diversity, aberrant AS can have deleterious effects leading to the development of several pathological complications including cancer [28–30]. Therefore, identifying potential modulators of AS is critical to fully comprehend the aetiology and molecular pathophysiology of cancer.

In this study, we conducted transcriptomic analyses in *Ctcf*$^{+/-}$ mice to explore the impact of *Ctcf* haploinsufficiency in AS regulation in brain, kidney, liver, muscle, and spleen. We found that *Ctcf* haploinsufficiency induces changes to gene expression and AS that are strikingly tissue-specific. Intron retention (IR), for example, is highly upregulated in *Ctcf* haploinsufficient liver and kidney compared to wildtype (WT) mice. Differential IR could be mediated by Ctcf binding sites located up- and downstream of retained introns.

CONTACT John Ej Rasko j.rasko@centenary.org.au Gene & Stem Cell Therapy Program Centenary Institute, The University of Sydney, Camperdown, NSW 2050, Australia

Supplemental data for this article can be accessed here.

Results

Ctcf mediates tissue-specific gene expression and alternative splicing

Transcriptomic and proteomic studies have revealed that CTCF is expressed in all mammalian tissues at levels that vary by up to 20-fold [10,11]. To assess the dosage-dependent impact of CTCF on gene regulation and AS, we examined a *Ctcf* haploinsufficient mouse model harbouring a hemizygous deletion of *Ctcf*. In this model, the entire coding region of one *Ctcf* allele is replaced with an expression cassette containing the *pgk* promoter and neomycin gene ($Ctcf^{+/pgkneo}$, herein referred to as $Ctcf^{+/-}$ for simplicity) (Fig. 1A, Supplementary **Fig. 1**). We selected five tissues from different body systems for which the relative CTCF expression between tissues is consistent for human and mouse (Supplementary **Fig. 2**), including lymphatic (spleen), nervous (brain), urinary (kidney), muscular (quadriceps femoris) and digestive (liver) systems. We isolated all five tissues from 11-week-old female mice comprising $Ctcf^{+/-}$ and WT littermates, and validated the reduction of *Ctcf* mRNA (by 36–41%) and Ctcf protein (the 130 kDa species by 18–59%) expression in $Ctcf^{+/-}$ mice by RT-qPCR and Western blotting, respectively (Supplementary **Fig. 3**). Interestingly, we observed Ctcf expression compensation at the protein level in all tissues except spleen, in which the Ctcf expression was less than 40% of WT Ctcf protein levels. Apart from the spleen, the increase in Ctcf expression from DNA to RNA and protein suggests that a post-transcriptional dosage compensation of *Ctcf* may have occurred (Fig. 1B).

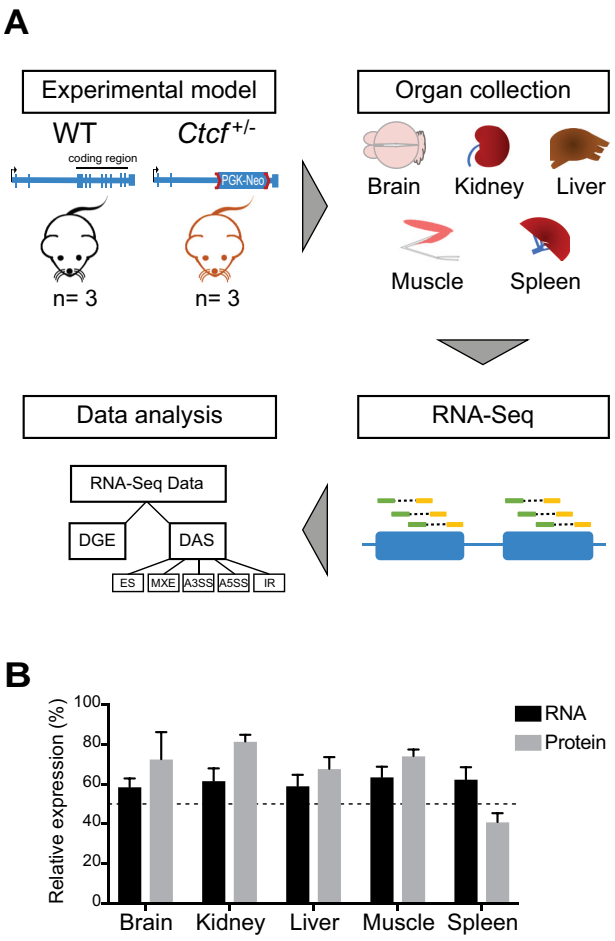

Figure 1. *Ctcf* haploinsufficiency as a model to study Ctcf-mediated transcriptome changes.

(A) Schematic illustration of the study design highlighting the mouse model, selected organs and types of data analyses. The RNA-seq data analyses performed in this study were differential gene expression (DGE) and differential alternative splicing (DAS) including exon sipping (ES), mutually exclusive exon (MXE), alternative 3′ splice site (A3SS), alternative 5′ splice site (A5SS) and intron retention (IR). (B) Summary of *Ctcf* mRNA and protein expression in all five tissues from WT and $Ctcf^{+/-}$ mice (see Supplementary **Fig. 3** for specific data and statistical analysis). The bars represent the mean ± SEM. The dashed line represents the expected level of Ctcf in the *Ctcf* haploinsufficient mice (50%).

We performed transcriptome analysis in all five tissues to examine differential gene regulation and differential AS between WT and $Ctcf^{+/-}$ mice. *Ctcf* transcript read counts assessed by RNA-seq in $Ctcf^{+/-}$ mice were significantly decreased in all tissues (33–39%) compared to WT mice (Fig. 2A). Differential gene expression analysis revealed that *Ctcf* haploinsufficiency exerts distinct tissue-specific effects. We detected at least 400 differentially expressed genes (DEGs) per tissue (a total of 3,746 DEGs in all tissues) (Supplementary **Table 1**). In all tissues other than liver, we detected an overall increase in gene expression (Fig. 2B), supporting Ctcf's role as a transcriptional repressor [6].

Apart from *Ctcf*, only six genes (*Carmil2, Gm38250, Gsto2, H2-Q7, Thap12* and *Zcwpw1*) were consistently differentially expressed in all five tissues (Fig. 2C). *Carmil2*, located adjacent to *Ctcf* at 8qD3, was upregulated in $Ctcf^{+/-}$ mice, which was also observed in a similar hemizygous $Ctcf^{+/-}$ mouse model [31]. *ZCWPW1* in humans is located within a CTCF-mediated TAD and contains a CTCF-bound intronic enhancer. A polymorphism (rs1476679) in this enhancer has been associated with increased disease risk in late onset-Alzheimer's disease by affecting CTCF binding and chromatin topology [32,33]. A relationship between *Ctcf* and the other four genes has not been established yet.

Gene Ontology (GO)-based annotation enrichment analysis of all the significant DEGs revealed a strong enrichment of tissue-specific biological processes (Supplementary **Fig. 4**). GO terms related to metabolism, immune response, and cellular component organisation were commonly affected by *Ctcf* haploinsufficiency (Fig. 2D). These data confirm that a decrease in Ctcf dosage through haploinsufficiency can impact gene expression in a tissue-specific manner.

To examine the effect of *Ctcf* haploinsufficiency on the AS landscape, we analysed differential AS events in $Ctcf^{+/-}$ using rMATS [34]. We found that exon skipping (ES) is the most abundant form of differential AS detected in all tissues (Supplementary **Table 2**). Notably, we observed a significant increase in IR events in the liver and kidney of $Ctcf^{+/-}$ mice (Fig. 2E), which is in contrast to all other forms of AS. However, a decrease in Ctcf-mediated IR events was only detected in $Ctcf^{+/-}$ muscle. Overall, our analysis shows that *Ctcf* haploinsufficiency not only affects gene expression but also perturbs the AS landscape in a tissue-specific manner.

Characterisation of intron retention in $Ctcf^{+/-}$ liver

To confirm that *Ctcf* haploinsufficiency modulates IR in a tissue-specific manner, we used IRFinder, a software

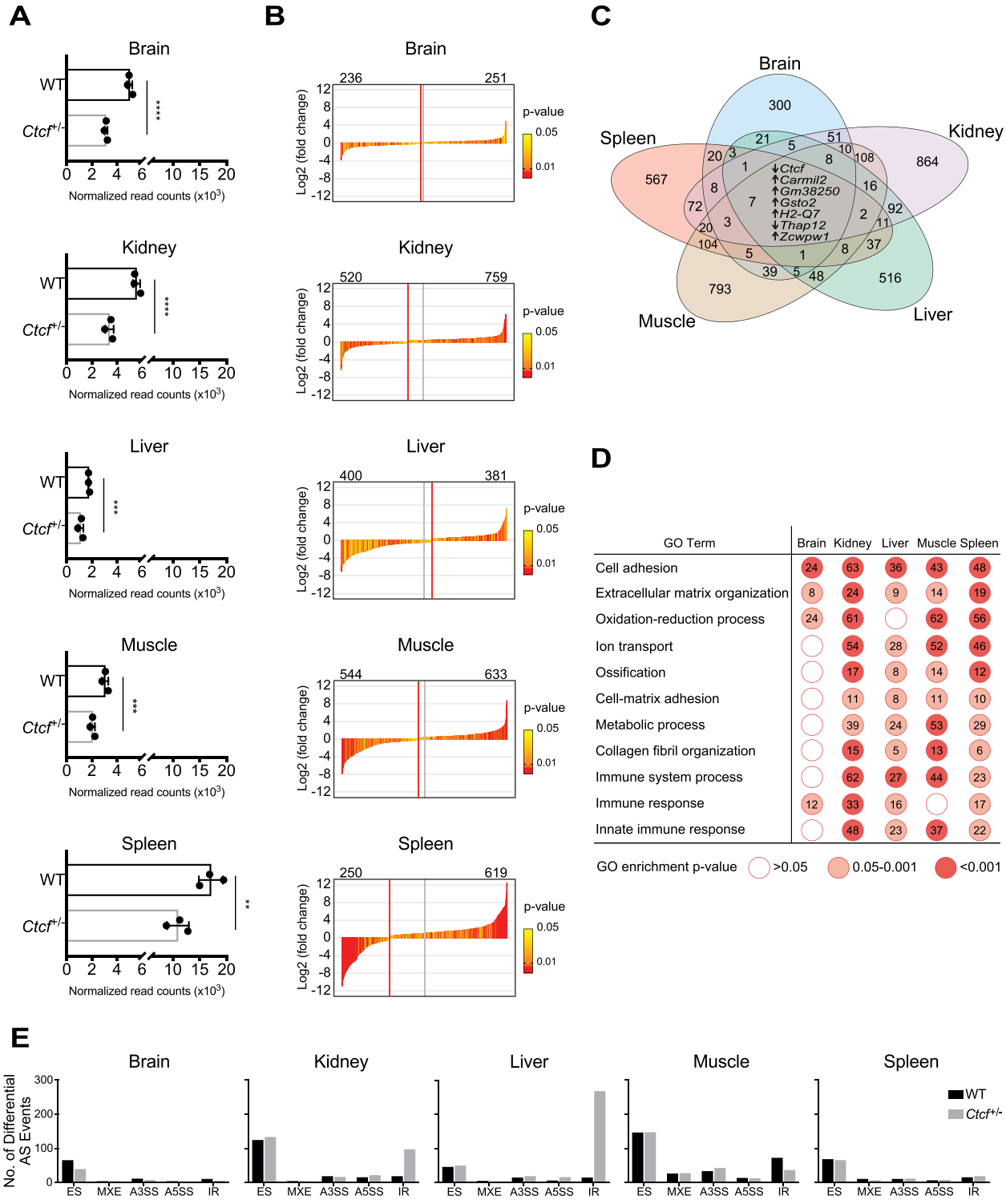

Figure 2. Differential gene expression and alternative splicing in *Ctcf*$^{+/-}$ mice.

(A) *Ctcf* mRNA expression from five tissues as normalised read counts. The bars represent the mean ± SEM. Wald test within DESeq2 was used to determine significance denoted by ** ($p < 0.01$), *** ($p < 0.001$), **** ($p < 0.0001$). (B) Waterfall plots showing the log2 (fold changes) of all significant DEGs ($p < 0.05$, Wald test using DESeq2) in each tissue. Each vertical bar represents a single DEG which is coloured based on the *p*-value. The total numbers of downregulated (left) and upregulated (right) genes are indicated on top of each plot. The central vertical line is drawn at half of the total number of DEGs while the red vertical line indicates the '0' inflection point. Liver was the only tissue not exhibiting a reduction in overall gene expression. (C) Venn diagram illustrating the overlap of DEGs among the five tissues. Arrows next to the common genes represent the change in their expression as up- or downregulated in all the five tissues. (D) Enriched functional annotations associated with DEGs common in at least four out of five tissues. Number of DEGs associated with each GO term is indicated within each circle. (E) Bar charts showing the number of significant differential AS events (FDR < 0.05, BH) detected in WT and *Ctcf*$^{+/-}$ tissues. *Ctcf*$^{+/-}$ liver exhibited a striking increase in IR. ES – exon skipping; MXE – mutually exclusive exon; A3SS – alternative 3′ splice site, A5SS – alternative 5′ splice site; IR – intron retention.

package specifically developed for the detection and quantification of IR [35–37]. We confirmed that IR is increased in *Ctcf*$^{+/-}$ liver and kidney with a total of 60 and 43 upregulated IR events, respectively. Only 1 and 5 IR events were downregulated in *Ctcf*$^{+/-}$ liver and kidney, respectively (Fig. 3A, Supplementary **Table 3**). Moreover, we detected that in some mRNA transcripts (5 in liver and 1 in kidney) multiple introns were differentially retained in the same transcript. There were no differential IR events common to all tissues.

As *Ctcf* haploinsufficient liver exhibited the most dramatic upregulation of IR, we explored these IR events for further validation and characterisation. First, we calculated the IR-ratio for all IR events (Supplementary **Fig. 5A**). The IR-ratio is the proportion of intron-retaining mRNA transcripts compared to all mRNA transcripts from the same gene. Consistent with previous studies, we considered IR events with an IR-ratio ≥0.1 as biologically significant [35,38]. Since we previously showed that IR can reduce the expression of the host gene via nonsense-mediated decay (NMD) [36], we examined the intronic expression as well as expression of host mRNA. We observed that 65% of differential intron-retaining genes, whether up- or downregulated, exhibited increased gene expression (Supplementary **Fig. 5B**). This suggests that *Ctcf* haploinsufficiency triggers both the upregulation of expression and IR-ratios of intron-retaining genes in liver.

Next, we selected five candidates exhibiting increased IR, including *Gadd45g, Plk3, Rfx5, Ppox* and *Setd4* (Fig. 3B), for validation by RT-qPCR. We confirmed a significant increase in retained introns in all five genes (Fig. 3C); however, the impact on host mRNA expression varied (Fig. 3D). We have recently summarised all possible fates of intron-retaining transcripts [39,40]. Apart from their degradation via NMD, fates include the nuclear detention of intron-retaining transcripts, the synthesis of novel protein isoforms, or their cellular translocation. To characterise the liver-specific intron-retaining genes, we performed annotation enrichment analysis and found that the affected genes are associated with biological processes and pathways related to metabolism, cytoskeletal organisation, RNA splicing and mRNA processing (Fig. 3E). These findings show that *Ctcf* haploinsufficiency mediates IR of specific introns in liver genes involved in splicing- and metabolism-related processes. In addition, we found a significant association between *Ctcf*$^{+/-}$-regulated DEGs and IR (p-value = 0.01, Fisher's exact test).

We next examined the molecular features of significantly differential IR events (n = 60) identified in *Ctcf*$^{+/-}$ liver compared to expressed non-retained introns (n = 1,160) and found that differentially retained introns are shorter (p-value = 3.3e-6, Fig. 3F) and have a higher GC content than non-retained introns (p-value = 7.7e-12, Fig. 3G), which is consistent with our previous reports [36,38,41]. To examine whether there is a correlation between these features and *Ctcf*$^{+/-}$ -mediated IR, we divided differentially retained introns and expressed non-retained introns, based on cut-off values of the median length (431 nt) and the mean GC content (45%) of all introns, into four groups: (i) short/GC-rich, (ii) short/GC-poor, (iii) long/GC-rich, (iv), and long/GC-poor (Supplementary **Fig. 6 and Table 4**). Interestingly, we found that 35 out of 60 (58.3%) differentially retained introns fall into the 'short/GC-rich' category, which constitutes 9% of all expressed short/GC-rich introns (p-value 2.5e-5, Fisher's exact test).

Enrichment of Ctcf binding sites proximal to differentially retained introns

CTCF ChIP-seq studies have found that approximately 30% and 50% of CTCF binding sites are located in intronic and intergenic regions, respectively [2,5,9]. It has been shown that the intragenic localisation of CTCF binding sites influences pre-mRNA splicing decisions leading to alternative exon inclusion [20,21,25]. To determine whether Ctcf binding sites are enriched in the proximity of differentially retained introns in liver, we analysed a publicly available ChIP-seq dataset from C57BL/6 normal mouse liver (E-MTAB-5769) [42]. Knowing that regions containing CTCF binding sites are mostly characterised by high GC content [5,43], we propose that GC-rich introns proximal to these regions are predisposed to IR particularly if they are short. Since short/GC-rich introns constitute more than half of the Ctcf-mediated differentially retained introns (35 out of 60), we merged the other classes (i.e. short/GC-poor, long/GC-rich, and long/GC-poor) of differentially retained introns into one group (named: other).

We examined Ctcf binding sites within the intron bodies as well as 200–50,000 nt upstream/downstream of differentially retained and expressed non-retained introns. The highest fold increase in Ctcf binding sites was observed upstream of differentially retained short/GC-rich introns (Fig. 4A). However, all differentially retained introns have a 2-fold increase in Ctcf binding sites in their ≤3,000 nt upstream region. These binding sites (n = 32) are primarily located in introns (50%), followed by intergenic regions (31.25%), and exons (18.75%, Fig. 4A). In the intron bodies, we observed a significantly higher percentage of the differentially retained short/GC-rich introns harbouring Ctcf binding sites compared to non-retained introns (11% vs. 4%, respectively, p-value = 0.02, Fisher's exact test). However, the total number of introns harbouring Ctcf binding sites (retained: 9 short/GC-rich and 0 others; non-retained: 23 short/GC-rich and 107 others) was low. Overall, these findings suggest that the *Ctcf* haploinsufficiency-induced IR changes in liver mostly affect short, GC-rich introns with proximal Ctcf binding sites. We then compared our results from liver to ChIP-seq data from brain, kidney and spleen (GSE36027) but found no enrichment of IR-associated Ctcf binding sites in these data, suggesting that IR in *Ctcf*$^{+/-}$ liver is indeed regulated by Ctcf (Supplementary **Fig. 7**).

Since CTCF is known for its global role in transcriptional regulation and insulation [6], we wanted to determine whether manipulating Ctcf expression *in vivo* perturbs splicing factor expression. By assessing 371 splicing factors (i.e. genes annotated with the GO term 'RNA splicing'; GO:0008380), we detected only minor changes in the expression of splicing factors (Supplementary **Fig. 8A**). Interestingly, we found that in *Ctcf*$^{+/-}$ liver a larger number of splicing factors, including *Srsf1, Prpf40b, Thoc1, Khsrp* and *Zpr1*, exhibit retained introns (Supplementary **Fig. 8B**). Given the

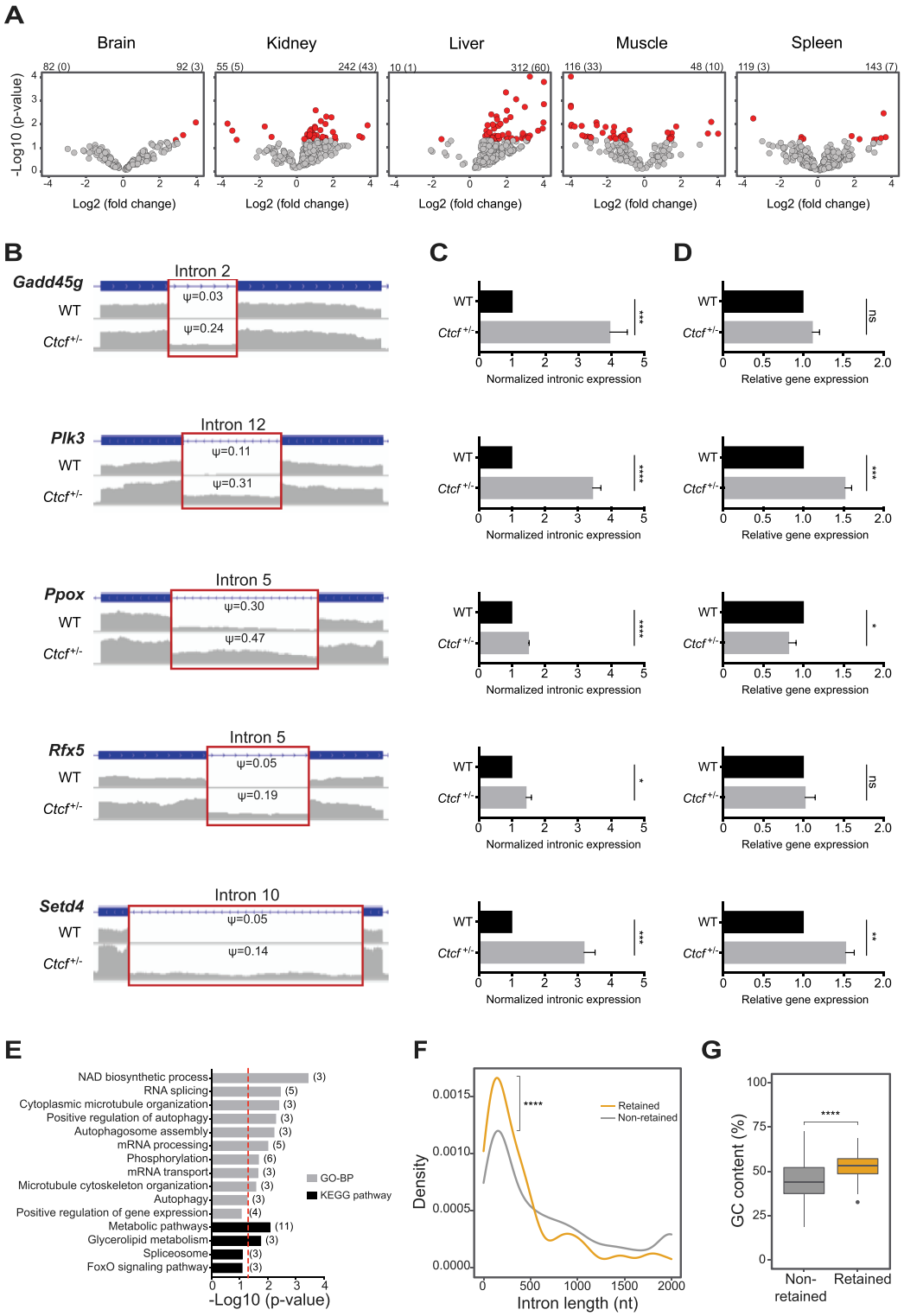

Figure 3. Validation and features of differentially retained introns in liver.

(A) Volcano plots showing differentially retained introns in mouse tissues. Significant differential IR events ($p < 0.05$, Wald test using IRFinder and DESeq2) are coloured in red. The numbers on top represent introns with downregulated (top-left) and upregulated (top-right) IR and the number of significantly differentially retained introns in parentheses. (B) Coverage plots of five differentially retained introns selected for validation. The IR-ratio (ψ) of each retained intron is indicated inside the red box. (C) Bar charts showing the expression of the same five selected differentially retained introns normalised to the expression of their flanking exons. Unpaired two-tailed Student's *t*-test was used to determine significance denoted by ns (not significant), * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$), **** ($p < 0.0001$). (D) Bar charts showing the relative expression of the same five genes harbouring the selected differentially retained introns. (E) Enriched functional annotations associated with genes of the 60 significantly differentially retained introns in *Ctcf*$^{+/-}$ liver. The dashed vertical red line represents *p*-value = 0.05. In parentheses is the number of genes associated with each GO term, where only GO terms with gene number >2 were included. (F) Density plot showing distribution of intron lengths of differentially retained introns and expressed non-retained introns in *Ctcf*$^{+/-}$ liver. Out-of-bounds intron length values (> 2000 bp) were re-scaled and plotted at the maximum value. Significance is denoted by **** (*p*-value < 0.0001, *t*-test). (G) Box plot comparing the GC content between differentially retained and expressed non-retained introns in *Ctcf*$^{+/-}$ liver.

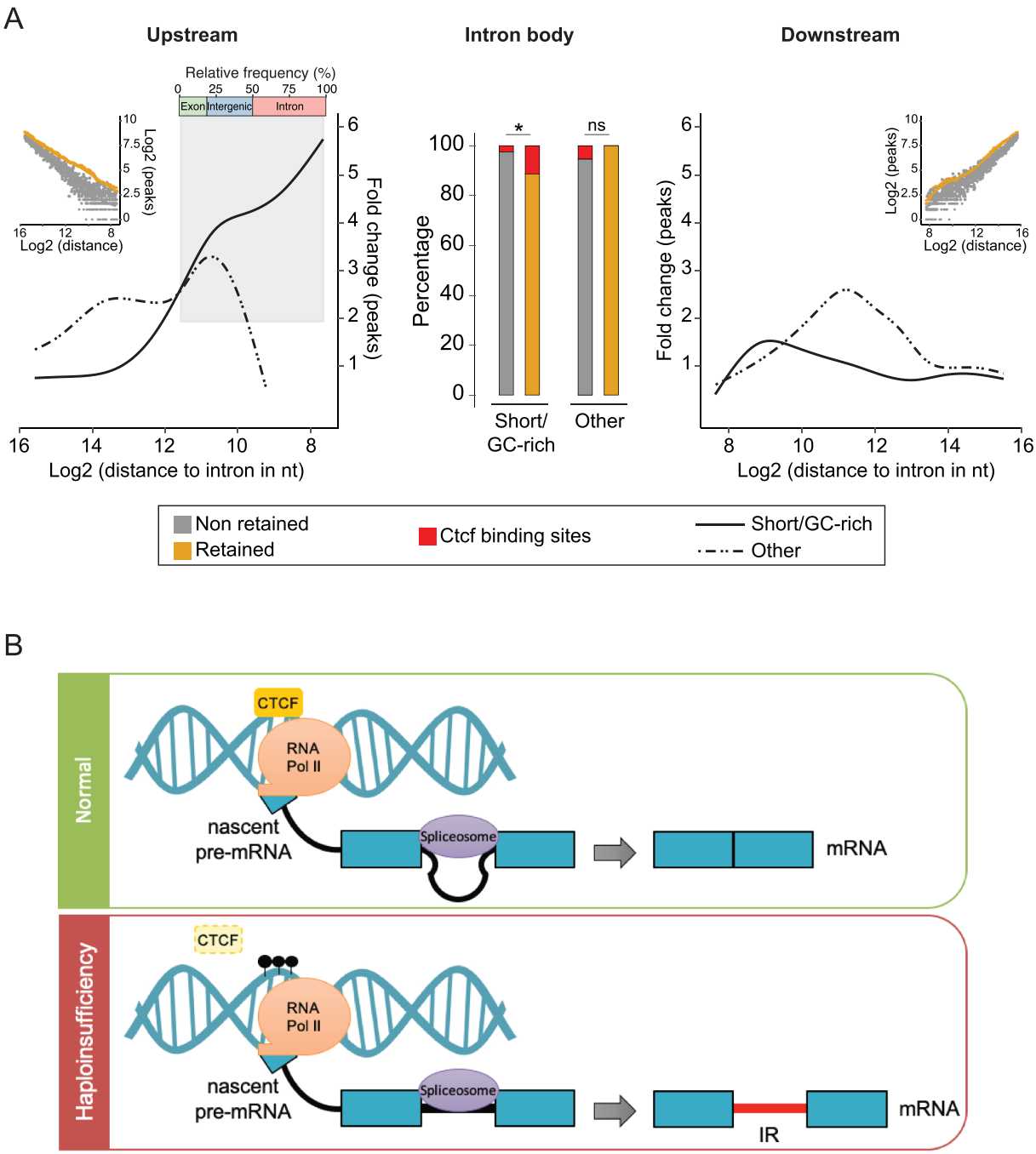

Figure 4. Enrichment of Ctcf binding sites proximal to differentially retained introns.

(A) Ctcf binding sites upstream (left panel), downstream (right panel), and within the body (middle panel) of differentially retained and expressed non-retained introns. The line graphs (left/right) show the distance-dependent fold difference in Ctcf peaks near short/GC-rich (solid line) and other (dashed line) differentially retained and expressed non-retained introns. The stacked bar on top of the line graph (left panel) shows the preferential genomic region of the enriched Ctcf binding sites of the area highlighted in grey. The insets show the cumulative number of Ctcf peaks with increasing distance from all the differentially retained (orange) and non-retained introns (grey). Data points represent 1,000 iterations of x (where x = number of retained introns in the relevant group) randomly selected non-retained introns with varying distance to the splice sites. The stacked bar graph in the middle shows the percentage (y-axis) of introns harbouring Ctcf binding sites. Significance is denoted by * (p-value < 0.05, Fisher's exact test; ns – not significant). (B) Proposed model of *CTCF* haploinsufficiency-mediated IR changes. In contrast to the normal situation (green box) where splicing proceeds constitutively, in the *CTCF* haploinsufficient liver (red box) various factors (e.g. RNA Pol II elongation rate, CTCF expression, CTCF binding proximal to splice sites and the methylation status of these CTCF binding sites) contribute to increased IR. Black circles indicate methylated sites altered by CTCF haploinsufficiency.

fact that IR coupled with NMD can reduce protein synthesis [36], our data suggest that *Ctcf* haploinsufficiency may drive IR-mediated regulation of splicing factor expression. While the fate and role of these dysregulated splicing factors remain to be determined, it would be reasonable to assume that they may contribute to increased IR in liver.

Discussion

IR is a widespread and conserved mechanism of post-transcriptional gene expression regulation affecting over 80% of all protein coding genes [35]. For instance, we have previously shown that IR regulates differentiation across diverse haematopoietic lineages [36,38,41,44]. Moreover, the majority of cancer types have abundant IR events that affect their transcriptomes, including genes involved in RNA splicing [45,46]. Therefore, identifying the potential modulators of AS, particularly IR, is essential to fully comprehend the aetiology and molecular pathophysiology of cancer.

Here, we demonstrate that *Ctcf* haploinsufficiency, which has previously been shown to induce spontaneous tumour formation in mice and genome-wide hypermethylation [17], also perturbs the AS landscape. Upon examining a number of tissues, we observed a specific increase of IR in *Ctcf*$^{+/-}$ liver and kidney. Moreover, we observed a similar tissue-specific effect of *Ctcf* haploinsufficiency on gene expression. Given that CTCF expression and DNA occupancy are variables between tissues [2–5,8,9], these differences may govern how CTCF specifically regulates the transcriptome and AS in tissues.

While the impact of *CTCF* haploinsufficiency on the global AS landscape has not been previously investigated, we utilised this model to examine *in vivo* the global effect of *Ctcf* dose-dependent regulation on AS in specific tissues. Although we used *Ctcf* hemizygous mice, we and others have shown that any sustained decrease in CTCF expression is partially compensated at the protein levels [17,31]. However, with a less than 50% reduction (except spleen) of *Ctcf* mRNA and protein expression, we observed tissue-specific differences in gene expression and AS. These data suggest that more prominent effects might be observed with loss-of-function mutations or the inactivation of CTCF, which are commonly observed in cancer [7,47], or at an even lower CTCF dosage. However, we have previously shown that the enforced genetic ablation of *CTCF* has a negative impact on somatic cell viability [16].

Although the exact mechanism underlying *Ctcf* haploinsufficiency-mediated IR regulation remains to be characterised, several possibilities can be proposed based on our data and previous studies. An optimal RNA Pol II elongation rate is essential for constitutive splicing [48]; therefore, both slow and fast RNA Pol II elongation rates were observed to promote IR [49]. Slowing RNA Pol II elongation mediated by CTCF promoted exon inclusion in B-lymphoma cell lines, providing the first evidence for a role of CTCF in AS. Binding of CTCF downstream of *CD45* exon 5 leads to decelerating RNA Pol II elongation, thus permitting exon 5 inclusion [20]. This mechanism is tightly regulated by cytosine methylation at CTCF binding sites in the DNA by the methylcytosine dioxygenases TET1 and TET2 [21]. In another example, CTCF binding to the *BDNF2a* locus was shown to be essential for its splicing by inhibiting TET1-induced DNA methylation and the subsequent interaction between TET1, MeCP2 and the splicing factor YB1 [22]. In contrast, DNA methylation at CTCF binding sites located upstream or downstream of alternatively spliced exons was associated with inclusion of exons [24]. Knowing that CTCF regulates DNA methylation and preserves methylation-free regions throughout the genome [50–52], we propose that CTCF regulates AS in a locus-dependent and tissue-specific manner.

CTCF has been recognised as a coordinator of chromatin looping and architecture [53–55]. Of immediate relevance is the observation that the RNA Pol II elongation rate can be controlled by chromatin organisation and influence AS decisions [56,57]. A recent study showed that CTCF-mediated chromatin loops formed between promoters and intragenic regions, particularly at upstream sites of alternatively spliced exons, lead to exon inclusion [25]. However, this mechanism still requires further validation. Overall, these studies suggest a role for CTCF as a key modulator of splicing decisions at individual loci via different molecular and epigenetic mechanisms.

Given its global transcriptional regulation and insulation functions [6], altering CTCF expression could perturb various transcriptional signalling pathways and subsequently AS. In this context, reduced CTCF expression was found to be associated with increased exon exclusion, particularly in those exons located 1 kb upstream of CTCF binding sites in BL41 and BJAB B-lymphoma cells [20]. This suggests that reduction of CTCF expression and the distribution of CTCF binding sites can influence pre-mRNA processing decisions. Consistent with that, we showed that enrichment of Ctcf binding sites upstream and downstream of differentially retained introns in liver is associated with increased IR. From these results, we speculate that CTCF may regulate IR through modulating the RNA Pol II elongation rate and methylation status at CTCF target binding sites located proximal to differentially retained introns (Fig. 4B).

Our data support previous observations from our group and others that IR mostly affects introns which are short and GC-rich [36,38,58]. During granulocytic differentiation, some intron-retaining transcripts undergo NMD, triggered by premature termination codons located within the retained introns [36]. This process can lead to an overall reduction of gene expression consequent to IR. Here, we observed a net increase in gene expression in the liver of *Ctcf*$^{+/-}$ mice, suggesting an alternative predominant fate of intron-retaining transcripts.

We observed that there was an enrichment of genes related to RNA splicing and metabolism among the intron-retaining genes. Splicing factors affected by IR particularly in *Ctcf*$^{+/-}$ liver and kidney such as *Srsf1, Esrp2* and *Prpf40b* have been previously reported to be associated with the regulation of AS including IR [35,59,60]. In addition to splicing, genes involved in metabolic processes were affected by IR in *Ctcf* haploinsufficient liver. We previously reported that *Ctcf* haploinsufficient mice exhibit ~14% reduction in body weight during post-natal development [16]. Given that the liver plays an essential role in organismal metabolism, we propose that Ctcf-regulated IR in liver may affect metabolic genes and pathways that contribute to this body weight phenotype. *Ctcf* is a haploinsufficient tumour suppressor that induces spontaneous tumour formation in various mouse tissues including liver [17]. The liver exhibited the most abundant IR events observed in our study. Given the fact that IR

is upregulated in most human cancers including liver cancer [45], a new link between *CTCF* haploinsufficiency and hepatocellular carcinoma causation via IR may be possible.

In this study, we have further defined the role of Ctcf in AS regulation and found Ctcf dosage-dependent effects on tissue-specific gene expression. Our analysis of Ctcf binding sites around retained introns in normal liver confirmed their relevance especially in short and GC-rich introns. To gain a mechanistic understanding of tissue-specific IR in *Ctcf* haploinsufficiency, additional studies are required that focus on Ctcf DNA occupancy and the interplay with the epigenetic regulation of AS. Moreover, causal links between *Ctcf* haploinsufficiency, IR, and increased tumour formation should be further investigated.

Materials and methods

Mouse model, genotyping and organ collection

All mouse experiments were conducted in accordance with the New South Wales Animal Research Act 1985, and the Australian code for the care and use of animals for scientific purposes (8th edition 2013). Animal ethics approval for all mouse experiments was obtained from the Sydney Local Health District Animal Welfare Committee (Protocol number 2016/020). C57BL/6J female mice from Australian BioResources, Australia were used as breeding stock. *Ctcf*$^{+/-}$ mice were originally obtained on a mixed C57BL/6:129SvJ background from the Fred Hutchinson Cancer Research Centre (Seattle, WA, USA) as a kind gift from Galina Filippova. These mice were previously backcrossed onto C57Bl/6J mice for at least 10 generations. *Ctcf*$^{+/-}$ mice contain a hemizygous deletion of the entire *Ctcf* coding region and substituted with an expression cassette containing the 3-phosphoglycerate kinase promoter and Neomycin gene flanked by two loxP sites (Pgk-Neo) [15,16]. Genotyping primers used to distinguish WT and *Ctcf*$^{+/-}$ alleles were Ctcf-WT-5′ (CTCACGCCTGAGATGATCC), Ctcf-WT-3′ (CATGCCATC CTACTGGTGTG) and Ctcf-neo-5′ (TGGGCTCTATGG CTTCTGAG). The resultant amplicons represent the WT allele (519 bp) and PGK-Neo-containing allele (348 bp).

Six healthy female C57BL/6 *Ctcf*$^{+/+}$ and *Ctcf*$^{+/-}$ mice (3 each) were utilised. At 11 weeks of age, two litters, one containing 2xWT and 2x*Ctcf*$^{+/-}$ female mice, the other containing one WT and one *Ctcf*$^{+/-}$ female mouse each were euthanised by carbon dioxide asphyxiation as recommended by the institutional animal welfare guidelines. Organs of interest (brain, kidney, liver, muscle and spleen) were collected and snap-frozen in liquid nitrogen, and stored at −80°C.

Protein extraction and quantitation

To extract protein from mouse organs, a small piece of the organ was minced in ice-cold RIPA protein lysis buffer (50 mM Tris-HCL (pH 8.0), 150 mM NaCl, 0.1% (w/v) sodium dodecyl sulphate (SDS), 0.5% (w/v) sodium deoxycholate, 1% (v/v) NP-40 and 0.02% (w/v) sodium azide in distilled H_2O with the addition of 1X EDTA-free SIGMAFAST protease cocktail tablet (Sigma). To quantify protein in the tissue lysates, the Micro BCA Protein Assay Kit (Thermo Fisher Scientific) was used according to the manufacturer's protocol.

Western blot, antibodies and densitometry

For Western blot analysis, protein extracts were prepared as equal aliquots (10 μg) and boiled in 2X protein loading buffer containing NuPAGE LDS Sample Buffer and 100 mM Dithiothreitol (DTT) for 5 minutes. The samples were then resolved by 4–12% SDS-PAGE using NuPAGE Bis-Tris Protein Gel system (Thermo Fisher Scientific), and transferred onto a PVDF membrane (Merck Millipore) in a semi-dry transfer apparatus. Rabbit polyclonal anti-CTCF (1:2,000; #3418, Cell Signalling) or mouse monoclonal anti-GAPDH (1:5,000 dilution; ab8245; Abcam) was used as a primary antibody, followed by washing and staining with horseradish peroxidase-conjugated donkey anti-rabbit or anti-mouse IgG secondary antibodies (1:5,000 dilution; Merck Millipore), respectively. A Chemidoc Touch (BioRad) was used to visualise the protein bands on these replicate blots, which were subjected to densitometric analysis using ImageJ software.

RNA isolation, purification and quantitation

To isolate RNA from WT and *Ctcf*$^{+/-}$ tissues of interest, a small piece of tissue (<30 mg weight) was homogenised with TRIzol reagent (Invitrogen) followed by isopropanol precipitation at −80°C. Next, contaminating DNA was eliminated by using the TURBO DNA-free kit (Invitrogen). The RNA quantitation and quality were assessed by NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific). Only samples with a 260/280 nm absorbance ratio of 1.9 or above were used. The RNA integrity was evaluated using a 2100 Bioanalyzer with the RNA 6000 Nano kit (Agilent) where samples with an RNA integrity number (RIN) below 7 were excluded.

RT-qPCR

To determine the relative expression level of *Ctcf* mRNA, cDNA was synthesised using SuperScript III Reverse Transcriptase (Invitrogen) followed by RNaseOUT treatment (Invitrogen). Next, qPCR was performed on the cDNA with SYBR Green (Bio-Rad) and *Ctcf*-specific forward (CCACCTGCCAAGAAGAGAAG) and reverse (CGACCTG AATGATGGCTGTT) primers, and subsequently run on a CFX96 real-time PCR machine (BioRad). Primers detecting *Hprt* were used to normalise gene expression: forward (AGTGTTGGATACAGGCCAGAC) and reverse (CGT GATTCAAATCCCTGAAGT). The fold-change was calculated using the $2^{-\Delta\Delta Ct}$ method. The *Ctcf* mRNA expression levels in *Ctcf*$^{+/-}$ mouse tissues were normalised to WT to obtain the relative mRNA expression.

RNA sequencing

A total of 30 RNA samples (five tissues collected from three biological replicates each of WT and *Ctcf*$^{+/-}$ mice), which met the quality control parameters described above, were sent to

Novogene (Beijing, China) for library preparation (strand-specific, poly-A enriched) and RNA-seq (150 bp paired-end reads, over 100 million reads) using the Illumina HiSeq-PE150 platform.

Bioinformatic analysis

Initial quality control of the RNA-seq raw data was conducted using FastQC (bioinformatics.babraham.ac.uk/projects/fastqc) and MultiQC (multiqc.info) to check for sequencing quality prior to analysis [61]. After passing the quality control check, reads were mapped to the mouse reference genome GRCm38/mm10 using STAR software [62]. To examine differential gene expression in WT and *Ctcf*$^{+/-}$ tissue samples, read counts were determined using featureCounts 1.5.0-p3 [63] and then subjected to differential gene expression analysis using the R Bioconductor package DESeq2 1.25.15 [64]. Genes with low read counts (<5) across all samples were removed. Differential AS was analysed using rMATS 4.0.1 [34], while further analysis of differential IR was performed using IRFinder 1.2.5 [35]. Specific IR features such as intron length and GC content were calculated using BEDTools v2.26.0 [65]. GO analysis of differentially expressed genes and intron-retaining genes were performed using the DAVID 6.8 online tool [66].

Validation of retained-intron candidates

Several candidates from the subset of intron-retaining transcripts in *Ctcf*$^{+/-}$ liver were selected for validation by RT-qPCR according to stringent filtering parameters from the main IRFinder output file. These filtering parameters included mean IR-ratio ≥0.1, mean intron coverage >75%, mean intron depth ≥5 reads, mean splice junction depth ≥20 reads, and *p*-value of IR-ratio fold-change between WT and *Ctcf*$^{+/-}$ samples <0.05. Further details of these parameters are provided in the IRFinder manual (github.com/williamritchie/IRFinder/wiki). Selected differentially retained introns were also visually inspected using the Integrative Genomics Viewer software (software.broadinstitute.org/software/igv/). On the other hand, expressed non-retained introns were selected for analyses based on the following criteria: (i) mean IR-ratio = 0 and (ii) mean splice junction depth >50 (as surrogate for expression).

Next, the selected candidates were validated by RT-qPCR using a set of three specific primers (Supplementary **Table 5**). To measure the relative mRNA expression, two primers were used to amplify sequence spanning the exonic region adjacent to the retained intron (Forward) to cross the exon-exon junction flanking the retained intron (Reverse 1). To measure the retained intron in the target spliced isoform, a third primer was designed to anneal within the intronic region of the retained intron (Reverse 2). Finally, the relative mRNA expression was calculated as described above while intron-retaining transcript expression was calculated after normalising its expression to the expression of the intron's flanking exons.

Enrichment of Ctcf binding sites

To examine enrichment of Ctcf binding sites, we obtained publicly available ChIP-seq datasets from C57BL/6 normal mouse liver (E-MTAB-5769) [42] as well as brain, kidney, and spleen (GSE36027) [4]. We could not find ChIP-seq data from muscle. The Ctcf ChIP-seq peaks were then intersected with differentially retained introns and expressed non-retained introns. To identify Ctcf binding sites, the FASTA sequences of the detected Ctcf peaks in the region located up to 3000 nt upstream of the differentially retained introns were first retrieved using BEDTools v2.26.0 [65]. Next, these sequences were scanned for the CTCF binding motif from JASPAR2020 database [67] using the FIMO program from the MEME suite v.4.12.0 [68,69].

Statistical analysis

The data from this study was analysed using different statistical tests relevant to the type of experiment or analysis. Wald test and unpaired two-tailed Student's *t*-test (as indicated in Figure legends) were used to determine statistical significance. P-values <0.05 were considered as statistically significant. All error bars shown in the figures represent standard error of the mean (SEM) from at least three independent experiments. The GraphPad Prism software version 7 as well as R Bioconductor were used to perform the statistical data analyses.

Acknowledgements

The authors acknowledge the Centenary Institute Animal House staff for animal husbandry. We also acknowledge the University of Sydney HPC service for providing HPC resources that have contributed to the research results reported within this paper.

Funding

This work was supported by the National Health and Medical Research Council (Project grants #1128748 and #1128175 to J.E.J.R.) and an anonymous foundation. Financial support was also provided by Tour de Cure (Scott Canner Research Fellowship) to C.G.B. and for research grants to C.G.B. and J.E.J.R; Tour de Rocks project support to C.G.B.; Cancer Council NSW project grants (RG11-12, RG14-09) to J.E.J.R. and C.G.B, (RG20-12 to U.S. and C.G.B.). A.B.A. is supported by a PhD Scholarship from Umm Al-Qura University in Saudi Arabia. U.S. and J.J.L.W. hold Fellowships from the Cancer Institute of New South Wales.

Author contributions

The project was conceived by J.E.J.R., C.G.B., A.D.M. and U.S.; experiments were planned by A.D.M. and A.B.A.; experiments were performed by A.D.M., A.B.A. and J.J.L.W.; RNA-seq data were analysed by A.B.A, U.S and D.V; all other data were analysed and evaluated by A.B.A.; the mice organs were collected by R.N., A.B.A., A.D.M. and M.V.; the manuscript was written by A.B.A., C.G.B., U.S. and J.E.J.R.

Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

ORCID

Adel B Alharbi http://orcid.org/0000-0002-9644-1032

Ulf Schmitz http://orcid.org/0000-0001-5806-4662
Darya Vanichkina http://orcid.org/0000-0002-0406-164X
Justin JL Wong http://orcid.org/0000-0002-7038-5079
Charles G Bailey http://orcid.org/0000-0002-4725-7829
John EJ Rasko http://orcid.org/0000-0003-2975-807X

Data availability

Data have been submitted to GEO under accession number (GSE140532).

References

[1] Filippova GN, Fagerlie S, Klenova EM, et al. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol Cell Biol. 1996;16:2802–2813.
[2] Kim TH, Abdullaev ZK, Smith AD, et al. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell. 2007;128:1231–1245.
[3] Schmidt D, Schwalie PC, Wilson MD, et al. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell. 2012;148:335–348.
[4] Shen Y, Yue F, McCleary DF, et al. A map of the cis-regulatory sequences in the mouse genome. Nature. 2012;488:116–120.
[5] Chen H, Tian Y, Shu W, et al. Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. PLoS One. 2012;7:e41374.
[6] Phillips JE, Corces VG. CTCF: master weaver of the genome. Cell. 2009;137:1194–1211.
[7] Marshall AD, Bailey CG, Rasko JE. CTCF and BORIS in genome regulation and cancer. Curr Opin Genet Dev. 2014;24:8–15.
[8] Wang H, Maurano MT, Qu H, et al. Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome Res. 2012;22:1680–1688.
[9] Cuddapah S, Jothi R, Schones DE, et al. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. Genome Res. 2009;19:24–32.
[10] Uhlen M, Fagerberg L, Hallstrom BM, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015;347:1260419.
[11] Merkin J, Russell C, Chen P, et al. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. Science. 2012;338:1593–1599.
[12] Arzate-Mejia RG, Recillas-Targa F, Corces VG. Developing in 3D: the role of CTCF in cell differentiation. Development. 2018;145 (6):dev137729.
[13] Franco MM, Prickett AR, Oakey RJ. The role of CCCTC-binding factor (CTCF) in genomic imprinting, development, and reproduction. Biol Reprod. 2014;91(125):121–129.
[14] Herold M, Bartkuhn M, Renkawitz R. CTCF: insights into insulator function during development. Development. 2012;139:1045–1057.
[15] Moore JM, Rabaia NA, Smith LE, et al. Loss of maternal CTCF is associated with peri-implantation lethality of Ctcf null embryos. PLoS One. 2012;7:e34915.
[16] Bailey CG, Metierre C, Feng Y, et al. CTCF expression is essential for somatic cell viability and protection against cancer. Int J Mol Sci. 2018;19:3832.
[17] Kemp CJ, Moore JM, Moser R, et al. CTCF haploinsufficiency destabilizes DNA methylation and predisposes to cancer. Cell Rep. 2014;7:1020–1029.
[18] Rasko JE, Klenova EM, Leon J, et al. Cell growth inhibition by the multifunctional multivalent zinc-finger factor CTCF. Cancer Res. 2001;61:6002–6007.
[19] Tiffen JC, Bailey CG, Marshall AD, et al. The cancer-testis antigen BORIS phenocopies the tumor suppressor CTCF in normal and neoplastic cells. Int J Cancer. 2013;133:1603–1613.
[20] Shukla S, Kavak E, Gregory M, et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature. 2011;479:74–79.
[21] Marina RJ, Sturgill D, Bailly MA, et al. TET-catalyzed oxidation of intragenic 5-methylcytosine regulates CTCF-dependent alternative splicing. Embo J. 2016;35:335–355.
[22] Zheng Z, Ambigapathy G, Keifer J. MeCP2 regulates Tet1-catalyzed demethylation, CTCF binding, and learning-dependent alternative splicing of the BDNF gene in Turtle. Elife. 2017;6:e25384.
[23] Agirre E, Bellora N, Allo M, et al. A chromatin code for alternative splicing involving a putative association between CTCF and HP1alpha proteins. BMC Biol. 2015;13:31.
[24] Wan J, Oliver VF, Zhu H, et al. Integrative analysis of tissue-specific methylation and alternative splicing identifies conserved transcription factor binding motifs. Nucleic Acids Res. 2013;41:8503–8514.
[25] Ruiz-Velasco M, Kumar M, Lai MC, et al. CTCF-mediated chromatin loops between promoter and gene body regulate alternative splicing across individuals. Cell Syst. 2017;5:628–637.
[26] Pan Q, Shai O, Lee LJ, et al. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet. 2008;40:1413–1415.
[27] Wang ET, Sandberg R, Luo S, et al. Alternative isoform regulation in human tissue transcriptomes. Nature. 2008;456:470–476.
[28] Kim E, Goren A, Ast G. Insights into the connection between cancer and alternative splicing. Trends Genet. 2008;24:7–10.
[29] Pagani F, Baralle FE. Genomic variants in exons and introns: identifying the splicing spoilers. Nat Rev Genet. 2004;5:389–396.
[30] Tazi J, Bakkour N, Stamm S. Alternative splicing and disease. Biochim Biophys Acta. 2009;1792:14–26.
[31] Aitken SJ, Ibarra-Soria X, Kentepozidou E, et al. CTCF maintains regulatory homeostasis of cancer pathways. Genome Biol. 2018;19:106.
[32] Gao Y, Tan MS, Wang HF, et al. ZCWPW1 is associated with late-onset Alzheimer's disease in Han Chinese: a replication study and meta-analyses. Oncotarget. 2016;7:20305–20311.
[33] Kikuchi M, Hara N, Hasegawa M, et al. Enhancer variants associated with Alzheimer's disease affect gene expression via chromatin looping. BMC Med Genomics. 2019;12:128.
[34] Shen S, Park JW, Lu ZX, et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc Natl Acad Sci U S A. 2014;111:E5593–E5601.
[35] Middleton R, Gao D, Thomas A, et al. IRFinder: assessing the impact of intron retention on mammalian gene expression. Genome Biol. 2017;18:51.
[36] Wong JJ, Ritchie W, Ebner OA, et al. Orchestrated intron retention regulates normal granulocyte differentiation. Cell. 2013;154:583–595.
[37] Wong JJ, Gao D, Nguyen TV, et al. Intron retention is regulated by altered MeCP2-mediated splicing factor recruitment. Nat Commun. 2017;8:15134.
[38] Schmitz U, Pinello N, Jia F, et al. Intron retention enhances gene regulatory complexity in vertebrates. Genome Biol. 2017;18:216.
[39] Wong JJ, Au AY, Ritchie W, et al. Intron retention in mRNA: no longer nonsense: known and putative roles of intron retention in normal and disease biology. Bioessays. 2016;38:41–49.
[40] Monteuuis G, Wong JJL, Bailey CG, et al. The changing paradigm of intron retention: regulation, ramifications and recipes. Nucleic Acids Res. 2019;47:11497–11513.
[41] Edwards CR, Ritchie W, Wong JJ, et al. A dynamic intron retention program in the mammalian megakaryocyte and erythrocyte lineages. Blood. 2016;127:e24–e34.
[42] Thybert D, Roller M, Navarro FCP, et al. Repeat associated mechanisms of genome evolution and function revealed by the Mus caroli and Mus pahari genomes. Genome Res. 2018;28:448–459.

[43] Wiehle L, Thorn GJ, Raddatz G, et al. DNA (de)methylation in embryonic stem cells controls CTCF-dependent chromatin boundaries. Genome Res. 2019;29:750–761.
[44] Green ID, Pinello N, Song R, et al. Macrophage development and activation involve coordinated intron retention in key inflammatory regulators. Nucleic Acids Res. 2020;48(12):6513–6529.
[45] Dvinge H, Bradley RK. Widespread intron retention diversifies most cancer transcriptomes. Genome Med. 2015;7:45.
[46] Wong ACH, Rasko JEJ, Wong JJ. We skip to work: alternative splicing in normal and malignant myelopoiesis. Leukemia. 2018;32:1081–1093.
[47] Bailey MH, Tokheim C, Porta-Pardo E, et al. Comprehensive characterization of cancer driver genes and mutations. Cell. 2018;173:371–385.
[48] Saldi T, Cortazar MA, Sheridan RM, et al. Coupling of RNA polymerase II transcription elongation with Pre-mRNA splicing. J Mol Biol. 2016;428:2623–2635.
[49] Fong N, Kim H, Zhou Y, et al. Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. Genes Dev. 2014;28:2663–2676.
[50] Mukhopadhyay R, Yu W, Whitehead J, et al. The binding sites for the chromatin insulator protein CTCF map to DNA methylation-free domains genome-wide. Genome Res. 2004;14:1594–1602.
[51] Pant V, Mariano P, Kanduri C, et al. The nucleotides responsible for the direct physical contact between the chromatin insulator protein CTCF and the H19 imprinting control region manifest parent of origin-specific long-distance insulation and methylation-free domains. Genes Dev. 2003;17:586–590.
[52] Schoenherr CJ, Levorse JM, Tilghman SM. CTCF maintains differential methylation at the Igf2/H19 locus. Nat Genet. 2003;33:66–69.
[53] Handoko L, Xu H, Li G, et al. CTCF-mediated functional chromatin interactome in pluripotent cells. Nat Genet. 2011;43:630–638.
[54] Splinter E, Heath H, Kooren J, et al. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev. 2006;20:2349–2354.
[55] Redolfi J, Zhan Y, Valdes-Quezada C, et al. DamC reveals principles of chromatin folding in vivo without crosslinking and ligation. Nat Struct Mol Biol. 2019;26:471–480.
[56] Schor IE, Rascovan N, Pelisch F, et al. Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. Proc Natl Acad Sci U S A. 2009;106:4325–4330.
[57] Lorincz MC, Dickerson DR, Schmitt M, et al. Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. Nat Struct Mol Biol. 2004;11:1068–1075.
[58] Braunschweig U, Barbosa-Morais NL, Pan Q, et al. Widespread intron retention in mammals functionally tunes transcriptomes. Genome Res. 2014;24:1774–1786.
[59] Yang Y, Park JW, Bebee TW, et al. Determination of a comprehensive alternative splicing regulatory network and combinatorial regulation by key factors during the epithelial-to-mesenchymal transition. Mol Cell Biol. 2016;36:1704–1719.
[60] Lorenzini PA, Chew RSE, Tan CW, et al. Human PRPF40B regulates hundreds of alternative splicing targets and represses a hypoxia expression signature. RNA. 2019;25:905–920.
[61] Ewels P, Magnusson M, Lundin S, et al. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32:3047–3048.
[62] Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.
[63] Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–930.
[64] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
[65] Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–842.
[66] Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4:44–57.
[67] Fornes O, Castro-Mondragon JA, Khan A, et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 2020;48:D87–d92.
[68] Bailey TL, Boden M, Buske FA, et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009;37:W202–208.
[69] Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. Bioinformatics. 2011;27:1017–1018.