Report for Pathological Society Small Grant Award: SGS 2015/04/04

Title: Investigating the interplay between *Ctcf* haploinsufficiency and DNA stability using proteomic approaches

Name: Sarah J Aitken
Address: (1) Cancer Research UK Cambridge Institute, University of Cambridge
(2) Addenbrooke's Hospital, Cambridge

Outputs (including meeting abstracts, oral presentations, original papers, review articles) from the study in which the Pathological Society has been acknowledged

Conference poster presentation
Chromatin and Epigenetics Keystone Symposia, Whistler, British Columbia, Canada.
Modeling the genome-wide and tumor suppressor effects of *Ctcf* haploinsufficiency
Sarah J Aitken[1,2], Christine Feig[1], Duncan T Odom[1]
[1]Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK.
[2]Department of Histopathology, Addenbrooke's Hospital, Cambridge, UK.

Investigating the interplay between *Ctcf* haploinsufficiency and DNA stability using proteomic approaches

Background and aims

CCCTC-binding factor (CTCF) is a master regulator of chromatin architecture and genome organization. CTCF uses differential zinc finger domain combinations to bind chromatin, allowing it to act as a transcriptional activator, repressor or insulator. CTCF binding is enriched at the borders of topologically associating domains (1), and its deletion disrupts global domain structure (2). Homozygous *Ctcf* knockout mice are non-viable, and *Ctcf* germline heterozygous mice are susceptible to spontaneous multi-lineage cancers (3). Furthermore, *CTCF* is mutated in many human lymphoid and epithelial malignancies (4). *Ctcf* is, therefore, implicated as a tumor suppressor gene; however, the underlying mechanisms remain unknown.

Project aim: To explore which chromatin organisation machineries are directly dependent on proper *Ctcf* expression by quantifying global protein changes between wild-type and *Ctcf* heterozygous mice.

Results

A *Ctcf* haploinsufficient mouse model was used to examine the genome- and proteome-wide impact of heterozygosity on transcription and translation. Female *Pgk-Cre* positive transgenic C57BL/6 mice were crossed with male mice carrying a heterozygous *Ctcf* reporter transgene flanked by loxP sites. Embryonic expression of the PGK-1 promoter results in early and uniform expression of the Cre site-specific recombinase, thus resulting in total excision of *Ctcf* in 50% of embryos (the remaining 50% harbouring the un-flanked *Ctcf* allele). Mouse embryonic fibroblast (MEF) cultures were established for *in vitro* studies using embryos from this breeding regimen, harvested at E13.5. Genotyping by PCR identified *Ctcf* heterozygous ($Ctcf^{+/-}$, HET) and wild-type ($Ctcf^{+/+}$, WT) lines.

1. *Ctcf* haploinsufficiency reduces the transcription and translation of Ctcf

qPCR and western blot analyses were first performed to quantify the effect of *Ctcf* haploinsufficiency on its own transcription and translation. RNA was extracted from passage 4 (P4) MEF cultures using QIAzol Lysis Reagent (Qiagen), cDNA was synthesised using the High-Capacity RNA-to-cDNA kit (Thermo Fisher Scientific), and qPCR was performed with TaqMan probes (Thermo Fisher Scientific) against *Ctcf* and *Gapdh* according to manufacturers' instructions (n=6 per genotype). *Ctcf* expression (normalised to *Gapdh*) was 36.8% reduced in $Ctcf^{+/-}$ MEFs relative to wild-type MEFs (Fig. 1A).

Protein was extracted from P4 MEFs using RIPA lysis buffer and quantified using a Direct Detect Infrared Spectrometer (Merck Millipore). 20 µg total protein was used for western blotting according to standard protocols with Ctcf anti-rabbit (1:1000) and β-actin anti-mouse primary antibodies (Cell Signalling Technology, 1:5000) and fluorescent-conjugated infra-red secondary antibodies (LI-COR Odyssey; goat anti-mouse antibody (1:20,000) and goat anti-rabbit antibody (1:5000)). The Odyssey CLx Imaging System was used for visualisation and quantification. Ctcf protein levels (normalised to β -actin) were 26.9% reduced in in $Ctcf^{+/-}$ MEFs compared with wild-type MEFs (Fig. 1B,C).

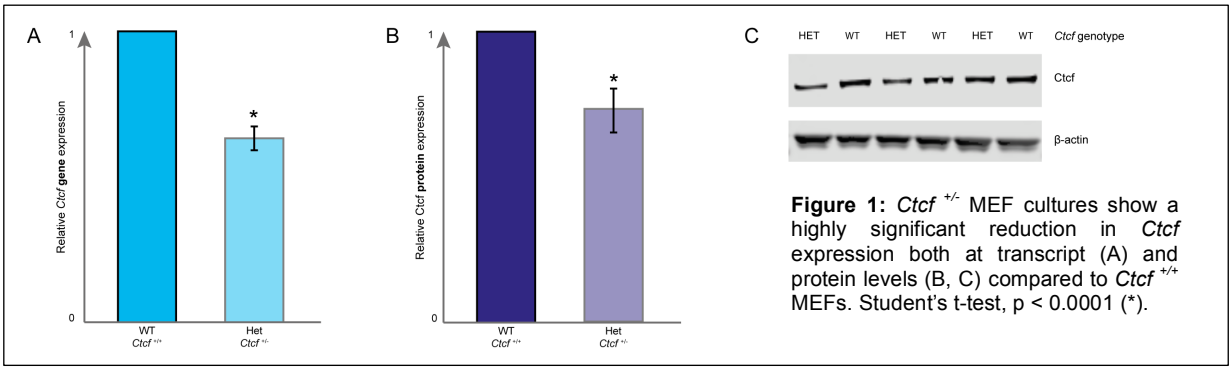

Figure 1: *Ctcf* $^{+/-}$ MEF cultures show a highly significant reduction in *Ctcf* expression both at transcript (A) and protein levels (B, C) compared to *Ctcf* $^{+/+}$ MEFs. Student's t-test, $p < 0.0001$ (*).

2. A reduced genomic concentration of *Ctcf* causes genome-wide transcriptomic and proteomic dysregulation

We next quantified the genome-wide effects of *Ctcf* haploinsufficiency using RNA sequencing (RNA-seq) and quantitative mass spectrometry. Total RNA was extracted as described above (n=6 per genotype) and DNase treated using the Turbo DNA-free kit (Thermo Fisher). Strand-specific libraries were prepared using the TruSeq Stranded Total RNA with Ribo-Zero Gold kit (Illumina) and sequenced with 150bp paired-end reads on a HiSeq 4000 (Illumina). Reads were aligned to the mm10 genome, and differential expression analysis was performed using edgeR (5).

One hundred and thirteen genes were differentially expressed between *Ctcf*$^{+/-}$ and *Ctcf*$^{+/+}$ MEFs (FDR <5%): 69 were significantly under expressed in *Ctcf*$^{+/-}$ MEFs, including *Ctcf* itself (Fig. 2A). KEGG pathway analysis (6) showed enrichment for pathways involved in important cellular processes including cell adhesion and various cancer pathways such as MAPK and Wnt signalling (Fig. 2B).

For quantitative proteome analysis, protein was extracted from MEFs using triethylammonium bicarbonate buffer with 0.1% SDS followed by protease digestion and Tandem Mass Tag (TMT) labelling (10-plex), high pH reverse-phase chromatography for fractionation, and liquid chromatography–MS (Q-Exactive, Thermo Scientific). Peptides were identified and quantified using Proteome Discoverer (Thermo Fisher Scientific), which includes the Mascot search engine and SEQUEST data analysis software. Five biological replicates of each genotype were used (number of replicates limited by the 10-plex TMT labelling protocol).

8,318 proteins were identified in all ten samples (FDR <1%). Eleven peptides in all samples mapped uniquely to Ctcf, ten which ten showed reduced expression in *Ctcf*$^{+/-}$ samples compared with wild-type controls. Protein expression was quantified by summing normalised intensity values for all peptides for a given protein. *Limma* (7) was applied to the quantile-normalised protein data to assess differential protein expression. Unsupervised clustering of the top one hundred differentially expressed proteins revealed two distinct clusters according to genotype (Fig. 2C).

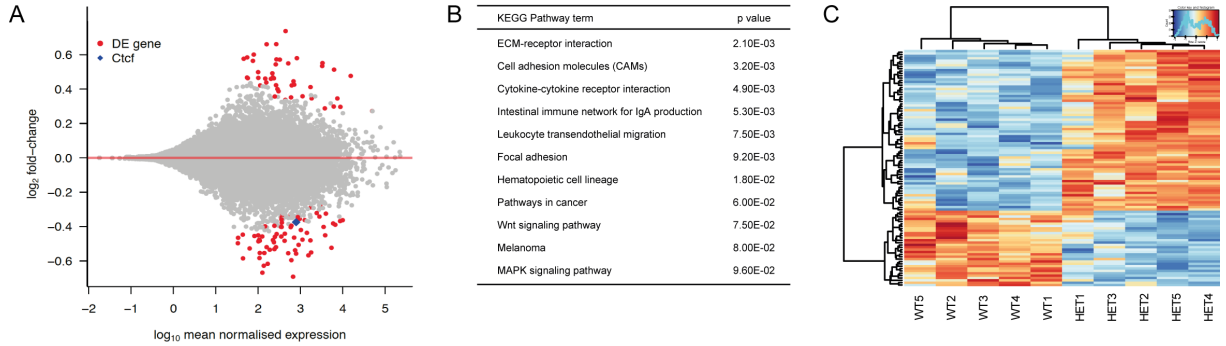

Figure 2: *Ctcf* haploinsufficiency results in differential expression of 113 genes (A) compared with wild-type samples (n = 6 per genotype) representing biological pathways (B) important in cell fate and cancer. (C) A heatmap showing the top 100 differentially expressed proteins identified using TMT proteomics, comparing *Ctcf*$^{+/-}$ and *Ctcf*$^{+/-}$ MEFs (n=5 per genotype).

Conclusions

This *Ctcf* haploinsufficient mouse model shows transcriptional and translational down-regulation of Ctcf. Halving the genomic concentration of *Ctcf* results in a 36.8% reduction in transcription of *Ctcf* itself and a 26.9% reduction in Ctcf protein. There is only partial compensation for haploinsufficiency, thus making this a suitable model for further study. Initial analysis of genome- and proteome-wide changes shows that *Ctcf* haploinsufficiency results in significant changes in biologically important signaling pathways, including those involved in carcinogenesis. Our next steps include technical replication of the TMT experiment and further in-depth analysis of these data. Specifically, we are examining how Ctcf loss alters chromatin structure to regulate these gene and protein networks by chromatin immunoprecipitation with high-throughput sequencing (ChIP-seq) and chromosome conformation capture techniques (HiC).

How closely have the original aims been met

This preliminary work *in vitro* has been essential to pave the way for analysis of liver tissues in a *Ctcf* conditional knockout mouse model of hepatocellular carcinoma, which is now underway. This grant has allowed me to develop and validate this model, start to explore the biological implications of *Ctcf* loss, and help to understand how it acts as a tumour suppressor gene.

References

1. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012 May 17;485(7398):376–80.
2. Zuin J, Dixon JR, van der Reijden MIJA, Ye Z, Kolovos P, Brouwer RWW, et al. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc Natl Acad Sci U S A. 2014 Jan 21;111(3):996–1001.
3. Kemp CJ, Moore JM, Moser R, Bernard B, Teater M, Smith LE, et al. CTCF haploinsufficiency destabilizes DNA methylation and predisposes to cancer. Cell Rep. 2014 May 22;7(4):1020–9.
4. ICGC Data Portal. CTCF [Internet]. 2014 [cited 2014 Dec 7]. Available from: https://dcc.icgc.org/genes/ENSG00000102974
5. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010 Jan;26(1):139–40.
6. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000 Jan 1;28(1):27–30.
7. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015 Apr 20;43(7):e47.