

RESEARCH

Open Access

# The male germ cell gene regulator CTCFL is functionally different from CTCF and binds CTCF-like consensus sites in a nucleosome composition-dependent manner

Frank Sleutels<sup>1\*</sup>, Widia Soochit<sup>1</sup>, Marek Bartkuhn<sup>2</sup>, Helen Heath<sup>1</sup>, Sven Dienstbach<sup>2</sup>, Philipp Bergmaier<sup>2</sup>, Vedran Franke<sup>3</sup>, Manuel Rosa-Garrido<sup>4,5</sup>, Suzanne van de Nobelen<sup>1</sup>, Lisa Caesar<sup>1</sup>, Michael van der Reijden<sup>1</sup>, Jan Christian Bryne<sup>3</sup>, Wilfred van IJcken<sup>6</sup>, J Anton Grootegoed<sup>7</sup>, M Dolores Delgado<sup>4</sup>, Boris Lenhard<sup>3</sup>, Rainer Renkawitz<sup>2</sup>, Frank Grosveld<sup>1,8,9</sup> and Niels Galjart<sup>1,8,9\*</sup>

## Abstract

**Background:** CTCF is a highly conserved and essential zinc finger protein expressed in virtually all cell types. In conjunction with cohesin, it organizes chromatin into loops, thereby regulating gene expression and epigenetic events. The function of CTCFL or BORIS, the testis-specific paralog of CTCF, is less clear.

**Results:** Using immunohistochemistry on testis sections and fluorescence-based microscopy on intact live seminiferous tubules, we show that CTCFL is only transiently present during spermatogenesis, prior to the onset of meiosis, when the protein co-localizes in nuclei with ubiquitously expressed CTCF. CTCFL distribution overlaps completely with that of Stra8, a retinoic acid-inducible protein essential for the propagation of meiosis. We find that absence of CTCFL in mice causes sub-fertility because of a partially penetrant testicular atrophy. CTCFL deficiency affects the expression of a number of testis-specific genes, including Gal3st1 and Prss50. Combined, these data indicate that CTCFL has a unique role in spermatogenesis. Genome-wide RNA expression studies in ES cells expressing a V5- and GFP-tagged form of CTCFL show that genes that are downregulated in CTCFL-deficient testis are upregulated in ES cells. These data indicate that CTCFL is a male germ cell gene regulator. Furthermore, genome-wide DNA-binding analysis shows that CTCFL binds a consensus sequence that is very similar to that of CTCF. However, only ~3,700 out of the ~5,700 CTCFL- and ~31,000 CTCF-binding sites overlap. CTCFL binds promoters with loosely assembled nucleosomes, whereas CTCF favors consensus sites surrounded by phased nucleosomes. Finally, an ES cell-based rescue assay shows that CTCFL is functionally different from CTCF.

**Conclusions:** Our data suggest that nucleosome composition specifies the genome-wide binding of CTCFL and CTCF. We propose that the transient expression of CTCFL in spermatogonia and preleptotene spermatocytes serves to occupy a subset of promoters and maintain the expression of male germ cell genes.

**Keywords:** CTCF, CTCFL, Gametogenesis, Genome-wide binding, Nucleosome

\* Correspondence: f.sleutels@erasmusmc.nl; n.galjart@erasmusmc.nl

<sup>1</sup>Department of Cell Biology Erasmus Medical Center, Rotterdam, The Netherlands

<sup>8</sup>Cancer Genomics Center, NCI, Rotterdam, The Netherlands

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

## Background

Three-dimensional folding of the eukaryotic genome occurs in a highly organized manner so as to compact chromatin while allowing temporal and spatial expression of genes. The genome contains regulatory elements, such as promoters, enhancers, locus control regions, insulators and enhancer blockers, that can orchestrate chromatin folding and gene activity over short and long distances, both in cis and in trans [1]. CTCF is a key coordinator of three-dimensional chromatin structure, allowing loop formation and specific chromatin compositions [2,3]. Gene activity is controlled in a positive or negative manner depending on the regulatory sequences present in the loops that are formed. The importance of CTCF in chromatin organization is emphasized by its evolutionary conservation, its ubiquitous expression, and its essential role in virtually all cells and tissues examined [4,5]. Hence, CTCF has been termed the “master weaver” of the genome [3].

The genome-wide binding by CTCF has been studied by different groups (see, for example, [6-10]). This has revealed ~35,000 CTCF-binding sites in the mammalian genome, of which more than 70% are shared between cell types. A relatively long consensus-binding motif for CTCF has been determined, which displays variability when compared to sites of transcription factors like KLF4, SOX2 and MYC [7]. The majority of CTCF binding-sites are found near genes, and ~8% is in the vicinity of transcription start sites (TSSs). Arrays of positioned (or “phased”) nucleosomes are found surrounding the nucleosome-free CTCF-binding sites [11-13], suggesting that CTCF binding promotes the ordered positioning of histones in its vicinity. CTCF has also been proposed to regulate the positioning of variant histones, such as H2A.Z [6,14]. Interestingly, the cohesin complex binds at the same position as CTCF in a CTCF-dependent manner. Together with CTCF, cohesin is essential for a proper three-dimensional chromatin structure and correct gene regulation [15-17].

CTCF-dependent loop formation is of crucial importance at imprinted loci. A well-studied example is the imprinted *Igf2-H19* locus, in which *Igf2* is expressed from the paternal and *H19* from the maternal allele [18]. The imprinting control region (ICR) located in between the *Igf2* and *H19* genes is methylated on the paternal allele, preventing CTCF binding. As a result the enhancer downstream of the *H19* gene can interact with the *Igf2* promoter and drive expression of this gene. On the non-methylated maternally derived ICR, CTCF does bind, thereby preventing enhancer-*Igf2* interaction, resulting in a chromatin loop that allows enhancer-*H19* association and *H19* expression. By binding the ICR, CTCF therefore acts as a regulator of imprinted sites.

The CTCF-like (CTCFL) protein, or Brother Of the Regulator of Imprinted Sites (BORIS) [19], has a central

domain of 11 zinc fingers (ZFs) that is very similar to that of CTCF and that is essential for DNA binding. The N- and C-terminal domains of CTCF and CTCFL are not homologous. CTCFL is less conserved across species, and the protein arose later in evolution, as it is detected in amniotes only [20]. Furthermore, expression of CTCFL is restricted to testis, several types of cancers and a number of cell lines [21-23].

Studies of CTCF and CTCFL protein distribution in the testis have yielded contradictory results. Initially, a mutually exclusive expression pattern of CTCFL and CTCF was described [19], with CTCF being present in round spermatids (i.e. after meiosis) and CTCFL in primary spermatocytes (i.e. during meiotic prophase).

Surprisingly, CTCFL was reported to be more abundant in the spermatocyte cytoplasm than in the nucleus. This led to the hypothesis that during germ cell development, CTCFL substitutes for the absence of CTCF and might be involved in reprogramming of DNA methylation in the male germ line. CTCFL was later reported to be present in gonocytes during embryonic development and, after birth, in spermatogonia, whereas CTCF was reported to localize to the supporting Sertoli cells [24]. In the same study CTCFL, together with the protein methyltransferase PRMT7, was suggested to regulate DNA methylation of imprinted genes in the male germline. However, defects in imprinting often result in embryonic phenotypes [25], whereas *Ctcf* knockout mice were shown to display a phenotype only in the testis [26]. Recently, enrichment of *Ctcf* mRNA in round spermatids was reported, adding perplexity to the localization and expression of CTCFL [26,27].

While the whole genome DNA-binding profile for CTCF has been elucidated, this has not been done for CTCFL. It therefore remains unclear how CTCFL binding relates to that of CTCF. In addition, it is unknown how these proteins are related functionally and mechanistically. To address these issues, we examined CTCFL function and localization with respect to CTCF, and identified the genome-wide binding sites of CTCFL and CTCF. We show that CTCF and CTCFL are functionally different proteins that co-localize within the nuclei of pre-meiotic germ cells. CTCFL acts as a male germ cell gene regulator, preferably binding near promoters with active chromatin marks. Interestingly, CTCF and CTCFL bind a highly similar DNA motif; nevertheless, only two-third of the ~5,700 CTCFL-binding sites are bound by CTCF. Conversely, the vast majority of CTCF sites are not bound by CTCFL. We find that nucleosome composition specifies CTCF and CTCFL binding. In contrast to CTCF, CTCFL associates with relatively “open” chromatin, and we propose that CTCFL promotes the maintenance of the epigenetic state of a subset of gene promoters and hence gene expression during spermatogenesis.

## Results

### CTCFL and CTCF co-localize transiently in pre-meiotic male germ cells

To resolve the localization of CTCFL and CTCF in testis, and to address CTCFL function, we generated *Ctcf* knockout and GFP-CTCF- and GFP-CTCFL-expressing knockin mice. To obtain information about the organization of the *Ctcf* gene, we mapped its 5' end and examined *Ctcf* expression (Figure 1A, B). We next generated three separate alleles using homologous recombination in ES cells: a *Ctcf* knockout allele (*Ctcf<sup>del</sup>*), in which exons 1–8 of the *Ctcf* gene are deleted, and *Ctcf* and *Ctcf* knockin alleles (*Ctcf<sup>gfp</sup>* and *Ctcf<sup>gfp</sup>*, respectively), to express GFP-CTCFL and GFP-CTCF instead of CTCFL and CTCF, respectively (Figure 1C-I).

Mice were generated, and the distribution of CTCFL was investigated by immunocytochemistry in sections of seminiferous tubules from wild-type and *Ctcf* knockout mice. CTCFL was present in wild-type testis in cells lining the basal lamina (Figure 2A, B). Not all cells lining the lamina were CTCFL-positive, and in some tubules no CTCFL-positive cells were detected. Importantly, no signal was detected on sections derived from CTCFL-deficient mice (Figure 2C, see also Figure 2F), showing that the

CTCFL staining in wild-type sections is specific. The localization of the CTCFL-positive cells in the basal compartment of the seminiferous tubules indicates that these cells are spermatogonia or preleptotene spermatocytes, as only upon progression in meiotic prophase do spermatocytes become disconnected from the basal lamina and move through the Sertoli cell barrier into the adluminal compartment of the seminiferous tubules.

The localization of CTCFL appeared reminiscent of STRA8 (STimulated by Retinoic Acid), which is expressed transiently from B spermatogonia to preleptotene spermatocytes and is essential for retinoic acid-induced commitment to meiosis [29-32]. An absolute overlap between STRA8 and CTCFL was confirmed using dual-color immunofluorescence (Figure 2D, E). Immunofluorescent staining experiments did not reveal an obvious change in the number of STRA8-positive tubules in CTCFL-deficient testis (Figure 2E, G, and data not shown). Thus, absence of the CTCFL signal in *Ctcf* knockout sections is not due to the disappearance of a cell type.

To confirm CTCFL localization and compare its distribution to that of CTCF, we next analyzed expression and localization of the two proteins *ex vivo*. We isolated intact seminiferous tubules from the testes of *Ctcf<sup>gfp</sup>* and *Ctcf<sup>gfp</sup>* male mice, which were injected with Hoechst via the rete testis to stain nuclei of cells at the adluminal compartment of the tubule. We then visualized GFP-CTCF(L) and Hoechst concomitantly using a multiphoton confocal laser scanning microscope setup [33]. Three-dimensional reconstruction of

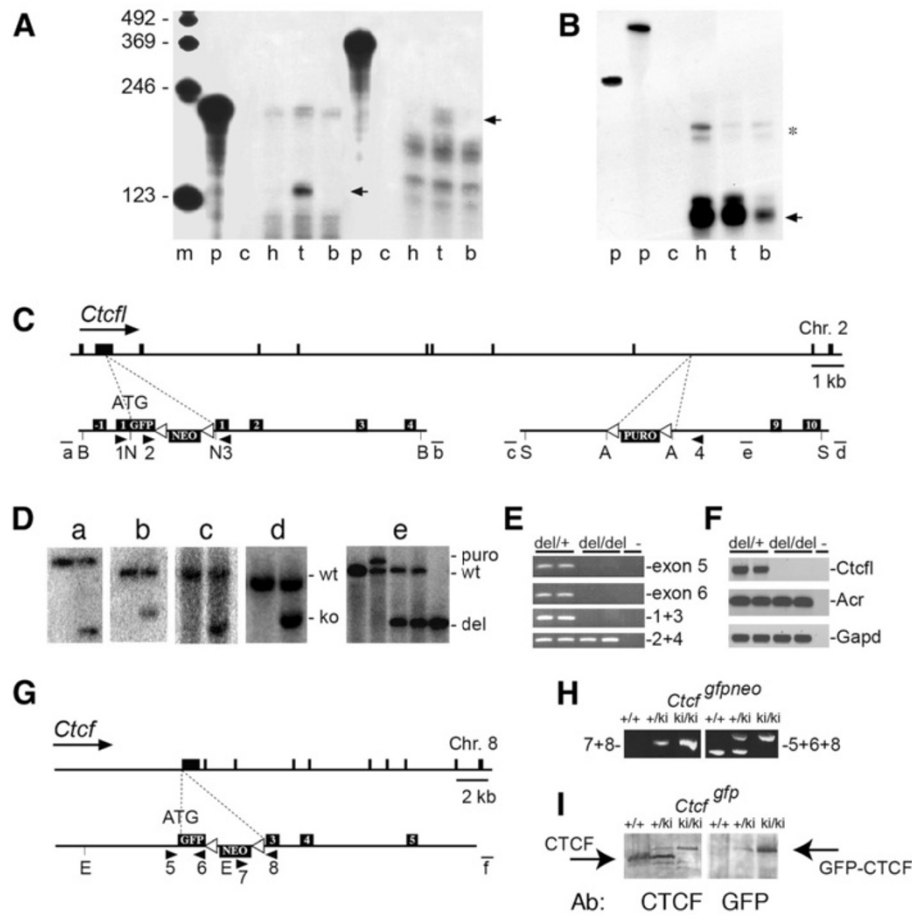
images taken longitudinally through the seminiferous tubules yielded an organizational view of the tubule, and the position and type of the GFP-positive cells (Figure 2H-P). GFP-CTCFL was detected in the nucleus of clustered cells representing a minor fraction of the total testis tubule (Figure 2H-m, and Additional file 1: Movie S1). These cells stained negative for Hoechst, and since the lumenally injected Hoechst does not cross the Sertoli cell barrier, the GFP-CTCFL-positive cells must reside on the basal side of this barrier. Sertoli cells, which form the tight junctions of the Sertoli cell barrier, were Hoechst-positive (Figure 2I, J, L, M and Additional file 1: Movie S1). Primary spermatocytes pass this barrier in the preleptotene and leptotene stage [34]. Based on Hoechst staining, morphology, size and location, we conclude that the GFP-CTCFL positive cells represent spermatogonia and preleptotene spermatocytes. The *ex vivo* GFP/Hoechst results obtained in live tissue are consistent with our data obtained in fixed paraffin-embedded sections of the testis stained with the CTCFL antibodies (Figure 2A-C). Together with the STRA8 colocalization data, they strongly suggest that in the adult testis CTCFL is transiently expressed in late spermatogonia and preleptotene germ cells.

In contrast to GFP-CTCFL, GFP-CTCF was present in the nucleus of all cell types of the seminiferous tubule, including all germ cells prior to spermiogenesis (Figure 2N-P, and data not shown). GFP-CTCF was also expressed in round spermatids, albeit at lower levels. This is consistent with a primary role for CTCF in cells with histone-based chromatin. Thus, live imaging in seminiferous tubules shows that CTCF and CTCFL are co-expressed within late spermatogonia and preleptotene spermatocytes. Measurement of GFP fluorescence intensities indicate that in cells where both proteins are expressed, the level of CTCF is somewhat higher than that of CTCFL.

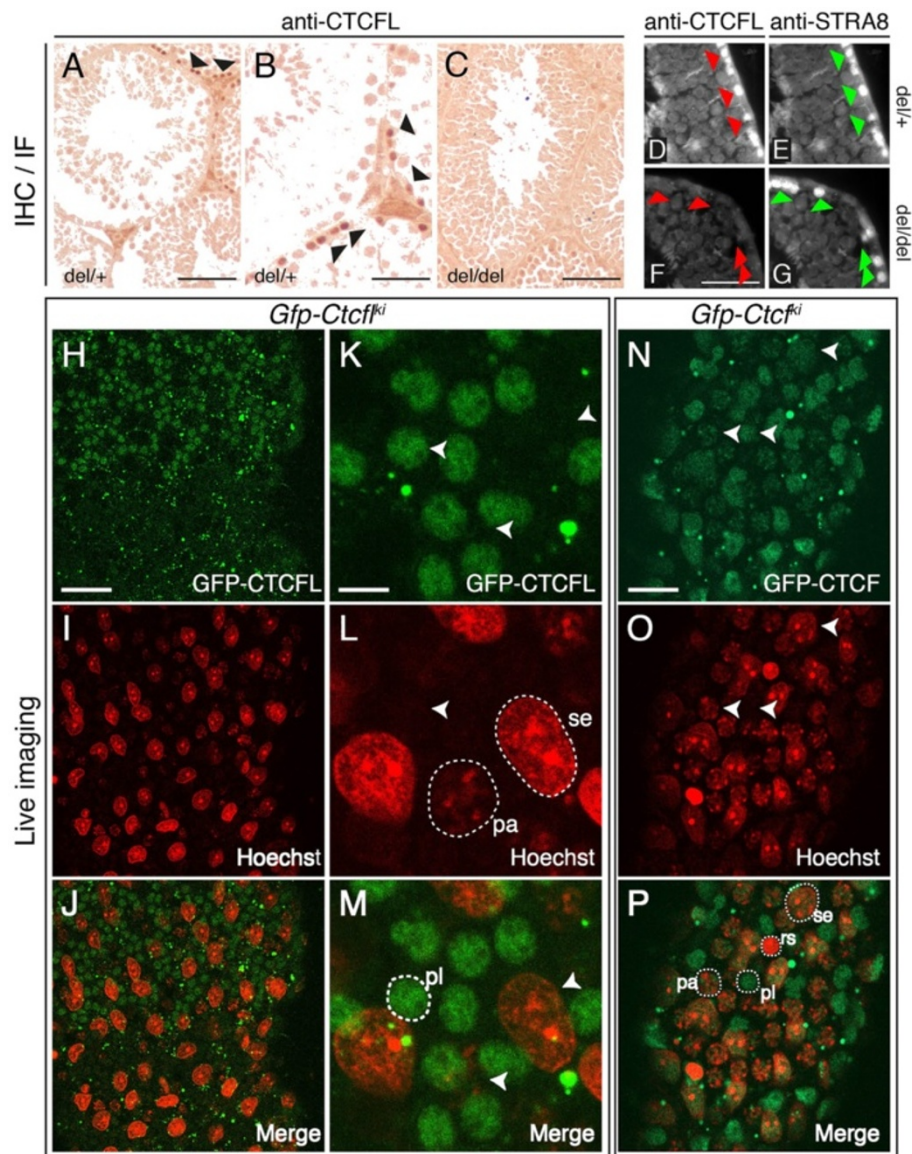
### CTCFL is important for spermatogenesis

To study the role of CTCFL in the male germ line, we analyzed *Ctcf<sup>del/+</sup>* and *Ctcf<sup>del/del</sup>* mice. These mice demonstrated no gross phenotypic defects and appeared normal. Heterozygous and homozygous *Ctcf<sup>del</sup>* females showed normal fertility and yielded offspring with expected ratios (data not shown), consistent with a role for CTCFL in spermatogenesis only. Heterozygous *Ctcf<sup>del/+</sup>* males generated offspring, and demonstrated normal fertility (Table 1). However, homozygous *Ctcf<sup>del</sup>* male littermates generated offspring in only half (14 out of 27) of the breedings (Table 1). Breeding with *Ctcf<sup>del/del</sup>* males yielded significantly ( $p = 0.01$ ; chi test) fewer litters than *Ctcf<sup>del/+</sup>* males, but not a different litter size ( $p = 0.11$ ; *t*-test). These data indicate that CTCFL is important for male fertility.

To further investigate the CTCFL deficiency, we weighed testes from 90-day-old *Ctcf<sup>del/del</sup>* and *Ctcf<sup>del/+</sup>* mice and plotted weight distributions. We found that, on average, the



**Figure 1 Ctcf and Ctcf expression and targeting.** **A, B** RNase protection analysis of Ctcf and Ctcf. For Ctcf (**A**) RNase protection analysis (RPA) was performed on polyA purified mRNA with probes covering parts of Ctcf exon 8 and 9 (left, small fragment) or a 5' end RACE product (right, large fragment). For Ctcf (**B**) the RPA was performed on total RNA with probes protecting Ctcf exon 2. The positions of the respective protected fragments are indicated with arrows. Ctcf mRNA can only be detected in polyA purified mRNA from testis (t), whereas Ctcf is identified in total RNA from all three tissues tested. M, marker; p, input probe; c, tRNA control; h, heart; t, testis; b, brain. Apur exon 3 is used as loading control and marked by an asterisk [28]. This analysis identifies the first exon containing the ATG translation initiation codon in Ctcf and shows that Ctcf is expressed in testis. **C** Schematic overview of the modified Ctcf alleles and targeting constructs. The Ctcf locus is shown on scale, with the constructs (not on scale) used for homologous recombination in ES cells underneath. Targeting at the 5' end of Ctcf yielded the Ctcf<sup>gfp-neo</sup> allele. Cre-mediated excision of the LoxP-embedded neomycin resistance gene yielded the Ctcf<sup>gfp</sup> allele (not shown). The 3' end targeting was performed on the Ctcf<sup>gfp-neo</sup> allele, and yielded the Ctcf<sup>gfp-neo-puro</sup> allele. Cre-mediated excision of the sequence in between the outermost LoxP sites yielded the Ctcf<sup>del</sup> allele, in which exons 1–8 of the Ctcf gene are deleted (not shown). A major difference between the Ctcf<sup>del</sup> allele described here and the Ctcf knockout published earlier [26] is that in the Ctcf<sup>del</sup> allele the GFP coding sequence is fused in frame with the CTCF coding sequence. Black boxes represent exons, GFP tag, neomycin and puromycin cassettes. Probes a, b, c, d and e are indicated by lines. Oligos 1, 2, 3 and 4 are represented by arrowheads. White triangles are LoxP sites. B = BglII; N = NcoI; S = SpeI; A = AvrII. **D** DNA blot showing Ctcf targeting. Probes a and b were used on DNA blots from ES cells for identification of the Ctcf<sup>gfp-neo</sup> allele and probes c and d for the Ctcf<sup>puro</sup> allele. Probe e identifies the Ctcf<sup>del</sup> allele from Ctcf<sup>gfp-neo-puro</sup> mice that were crossed to a chicken Actin-Cre transgene. Probe a, HindIII digest (wt 8.9 kb, ko 5.7 kb); probe b, EcoRI digest (wt 14 kb; ko 11 kb); probe c, BamHI digest (wt 16.1 kb; ko 6.8 kb); probe d, BamHI digest (wt 16.1 kb; ko 11.1 kb). **E** Absence of Ctcf DNA in the Ctcf<sup>del</sup> allele. PCR on tail DNA indicates that Ctcf<sup>del/del</sup> mice are deleted for exons 1–8 (top three panels) and are positive for GFP (oligos 2 and 4). **F** Absence of Ctcf RNA in Ctcf mutant mice. PCR on cDNA derived from testis mRNA shows that Ctcf is absent from Ctcf<sup>del/del</sup> mice. Acrosin and Gapd function as positive controls. **G** Schematic overview of the Ctcf allele and targeting strategy for the Ctcf<sup>gfp-neo</sup> allele. The Ctcf locus is shown on scale, with the construct (not on scale) used for homologous recombination in ES cells underneath. Cre-mediated excision of the LoxP-embedded neomycin resistance gene yielded the Ctcf<sup>gfp</sup> (or Ctcf<sup>ki</sup>) allele (not shown). Black boxes represent exons, GFP tag and neomycin cassette. Oligos 5, 6, 7 and 8 are represented by arrowheads. White triangles are LoxP sites. E = EcoRI. **H** PCR confirming Ctcf<sup>gfp-neo</sup> allele. Identification of the Ctcf<sup>gfp-neo</sup> (or Ctcf<sup>ki</sup>) allele by PCR with oligos 7 and 8 or oligos 5, 6 and 8 (see panel G). **I** Western blot confirming GFP-CTCF expression from the Ctcf<sup>gfp</sup> allele. We isolated MEFs from E13.5 day wild-type (+/+), heterozygous Ctcf<sup>gfp/+</sup> (or Ctcf<sup>ki/+</sup>) or homozygous Ctcf<sup>gfp/gfp</sup> (or Ctcf<sup>ki/ki</sup>) embryos, and identified the GFP-CTCF fusion protein by Western blot of MEF extracts using anti-CTCF or anti-GFP antibodies. Note the increased size of the GFP-CTCF protein compared to the CTCF protein due to the GFP tag.



**Figure 2 Expression of CTCFL and CTCF in the testis.** **A-C** Immunohistochemical staining of testis sections. Paraffin-embedded sections from day 90 testes from heterozygous (*del/+*) and homozygous (*del/del*) *Ctcf* mutant mice were stained with anti-CTCFL, followed by diaminobenzidine (*DAB*) coloring. Some of the CTCFL-positive cells are indicated with *black arrowheads*. Scale bars **A, C**: 100  $\mu$ m, **B**: 50  $\mu$ m. **D-G** Immunofluorescence staining of testis sections. Sections as described in **A-C** were stained with CTCFL (**D** and **F**) or STRA8 (**E** and **G**) antibodies. STRA8-positive cells in panels **E** and **G** are indicated with *green arrowheads*; the same cells are indicated with *red arrowheads* in the sections stained with anti-CTCFL antibodies (panels **D** and **F**). In *Ctcf* mutant mice, STRA8 distribution is not changed. Scale bar is 50  $\mu$ m. **H-P** Ex vivo confocal and multiphoton imaging of intact seminiferous tubules. Testis tubules were dissected from GFP-CTCFL- (**H-M**) or GFP-CTCF- (**N-P**) expressing mice, exposed to Hoechst at the adluminal side of the seminiferous tubule, and analyzed with a confocal/multiphoton microscope (GFP-CTCFL and GFP-CTCF, *green*; Hoechst, *red*). Panel **H-J** (see also Movie S1) shows a low magnification view of GFP-CTCFL distribution. Notice the presence of GFP-CTCFL-positive cells in the upper half of the tubule and their absence in the bottom half, indicating a transient population of cells. In (**K-M**) a high-magnification view of the same GFP-CTCFL-positive cells is shown. Notice the non-homogenous distribution of GFP-CTCFL in the nucleus. In (**N-P**) GFP-CTCF staining is shown. For clarity, some of the cell types are encircled, and their position is indicated in the other panels using *white arrowheads*. *pl* = preleptone spermatocyte; *rs* = round spermatid; *pa* = pachytene spermatocyte; *se* = Sertoli cell. Bars, **H-J**: 70  $\mu$ m, **K-M**: 10  $\mu$ m, **N-P**: 25  $\mu$ m.

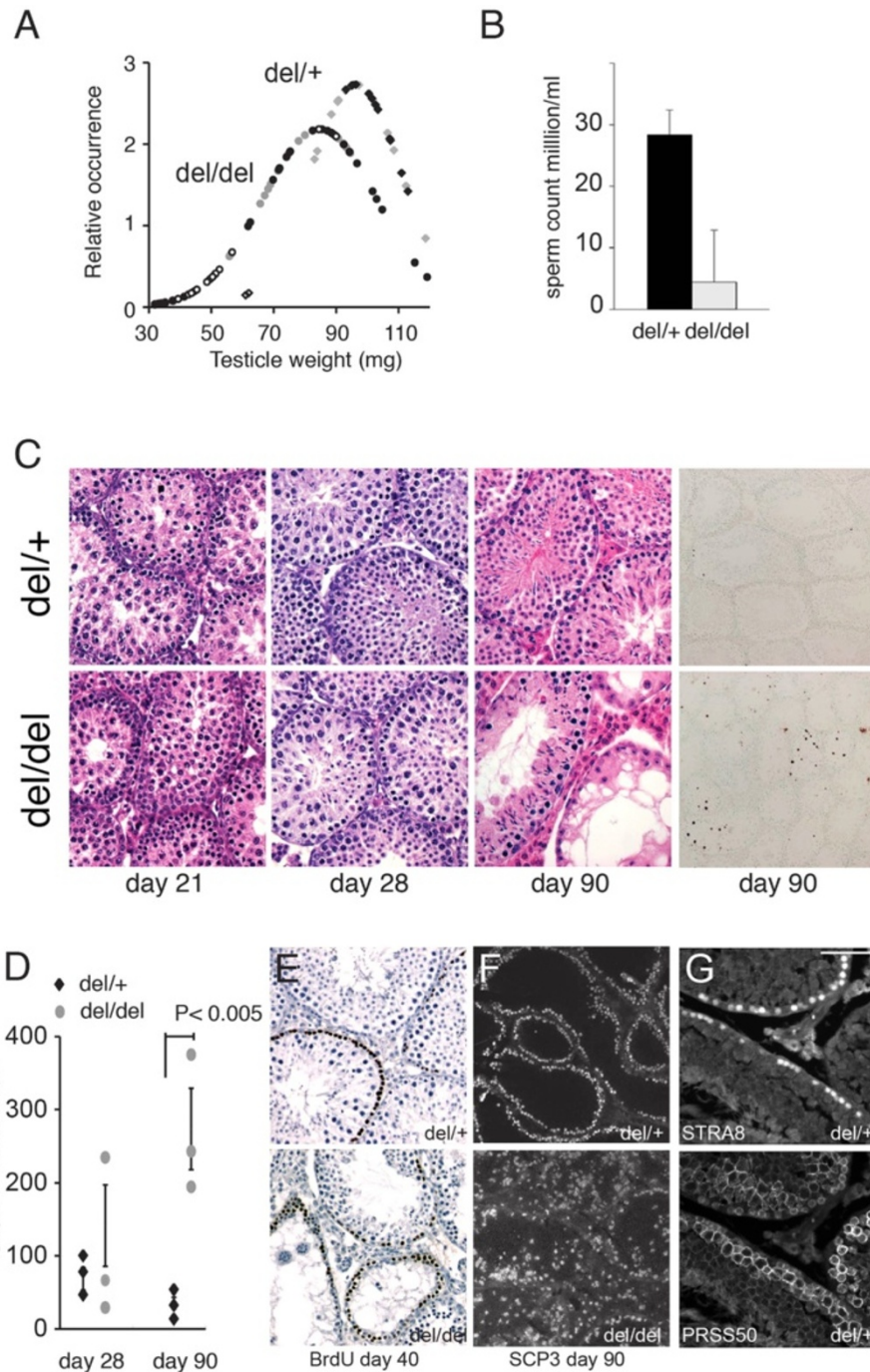
CTCFL-deficient testes weighed significantly less compared to testes from heterozygous littermates (Figure 3A). In addition, we found that lower testes weights coincided with infertile males (Figure 3A). The weight distribution shows

that there are also normal testes in the *Ctcf<sup>del/del</sup>* population. Still, on average, the epididymides from homozygous *Ctcf<sup>del</sup>* mice contained only 15% of sperm compared to heterozygous littermates (Figure 3B).

**Table 1 Sub-fertility of CTCFL mutant mice**

Genotype	<i>Ctcf<sup>del/+</sup></i>	<i>Ctcf<sup>del/del</sup></i>
Number (percentage) of breedings w/o offspring	3/60 (5%)	14/27 (51.9%)
Average number of offspring per litter ( $\pm$ SD)	7.7 $\pm$ 2.6	6.4 $\pm$ 2.8

From day 28 onwards *Ctcf* mutant mice displayed loss of germ cells by apoptosis and an increasing level of atrophy that increased with age (Figure 3C, D). Mitotic spermatogonia, staining positive for BrdU incorporation, were still often observed (Figure 3E), whereas SCP3, a marker for spermatocytes, revealed severe tubule disorganization (Figure 3F). In fact, the level of atrophy and disorganization between individual mice and between individual



**Figure 3** (See legend on next page.)

(See figure on previous page.)

**Figure 3 CTCFL is important for spermatogenesis.** **A** Testicular weight distribution. The testicular weight of *Ctcf1* heterozygous (*Ctcf1*<sup>del/+</sup>; diamonds) and homozygous (*Ctcf1*<sup>del/del</sup>; circles) mice was measured and plotted as a normalized probability distribution (i.e., the surface under the curve represents a total probability of 1). Testes of knockout mice are significantly smaller ( $p < 0.0005$ , *t*-test). *White symbols* represent infertile males, *black symbols* are fertile males, and *grey symbols* correspond to males not tested for fertility. **B** *Ctcf1* mutant mice display reduced fertility. Epididymal sperm count from *Ctcf1* heterozygous (*Ctcf1*<sup>del/+</sup>; black bar) and homozygous (*Ctcf1*<sup>del/del</sup>; grey bar) mice. Standard deviation is plotted ( $p = 0.0002$ ,  $n = 4$ ). **C** Testis histology. In the *left three panels* a timed series of HE-stained testicle sections is shown (postnatal day 21, 28, 90), while in the *right hand panel* an apoptosis assay (TUNEL staining) of testicle sections at day 90 is shown. Note that in CTCFL-deficient testes some seminiferous tubules appear normal, whereas others (which can be adjacent to the normal ones) have lost most of their meiotic and post-meiotic germ cells, leaving only mitotic spermatogonia (that stain positive for BrdU incorporation, see panel **E**) and Sertoli cells. Yet other tubules contain disorganized spermatocytes, and some of them even elongated sperm. Thus, there is no absolute block in differentiation or progression of germ cell development, but the incomplete penetrance of the infertility phenotype is however directly linked to the testicle weight (panel **A**) and to the degenerative level of the seminiferous tubules. **D** Apoptosis plot. Number of TUNEL-positive apoptotic cells per 100 seminiferous tubules counted at day 28 and day 90. Standard deviation of three animals per genotype and time point is indicated. **E** DNA synthesis marked by a 1-h pulse of BrdU in day 40 testicles reveals that mitotic spermatogonia are still present in degenerated tubules. Counterstaining with hematoxylin. **F** SCP3 staining in spermatocytes of day 90 testicles as a marker for tubule organization. **G** PRSS50 co-localizes only partially with STRA8. Immunofluorescence staining with a STRA8 antibody (*top panel*) or PRSS50 antibody (*bottom panel*) of adult testicle sections shows that PRSS50 and STRA8 expression overlaps only partially. Scale bars are 50  $\mu$ m.

seminiferous tubules within mice was variable, and normal seminiferous tubules could even be adjacent to abnormal ones. Thus, the penetrance of the atrophic testes and sterility phenotype in CTCFL deficiency is incomplete and differs considerably per mouse.

Next we performed a microarray analysis on day 23 testis mRNA (a time point that precedes the start of apoptosis and degeneration in the testes of CTCFL-deficient mice) on *Ctcf1*<sup>del/del</sup> and *Ctcf1*<sup>del/+</sup> littermates. This revealed several affected genes in *Ctcf1*<sup>del/del</sup> testes (Figure 4A). The *Prss50* and *Gal3st1* genes were most downregulated (~1.5 fold), which matches results from another study [26]. Real-time RT-PCR verified results from the microarray (Figure 4B).

GAL3ST1 is crucial for spermatogenesis as mutant mice are infertile because of an arrest at the end of the meiotic prophase [35]. PRSS50 (Testis Specific Protease) has an exclusively testicular expression pattern, and is detected both in CTCFL-positive cells as well as in later stages of spermatogenesis [36], including STRA8/CTCFL-negative pachytene spermatocytes (Figure 3G). Since a CTCFL-deficiency affects differentiation of cells subsequent to the preleptotene stage, the reduction in *Prss50* and *Gal3st1* mRNA may be the result of a reduction in the number of cells going through meiosis or a reduction in *Prss50* and *Gal3st1* mRNA per cell. We therefore investigated PRSS50 expression in sections of wild-type and CTCFL-mutant mice and noted reduced protein levels per cell (Figure 4C).

#### CTCFL regulates testis-specific gene expression

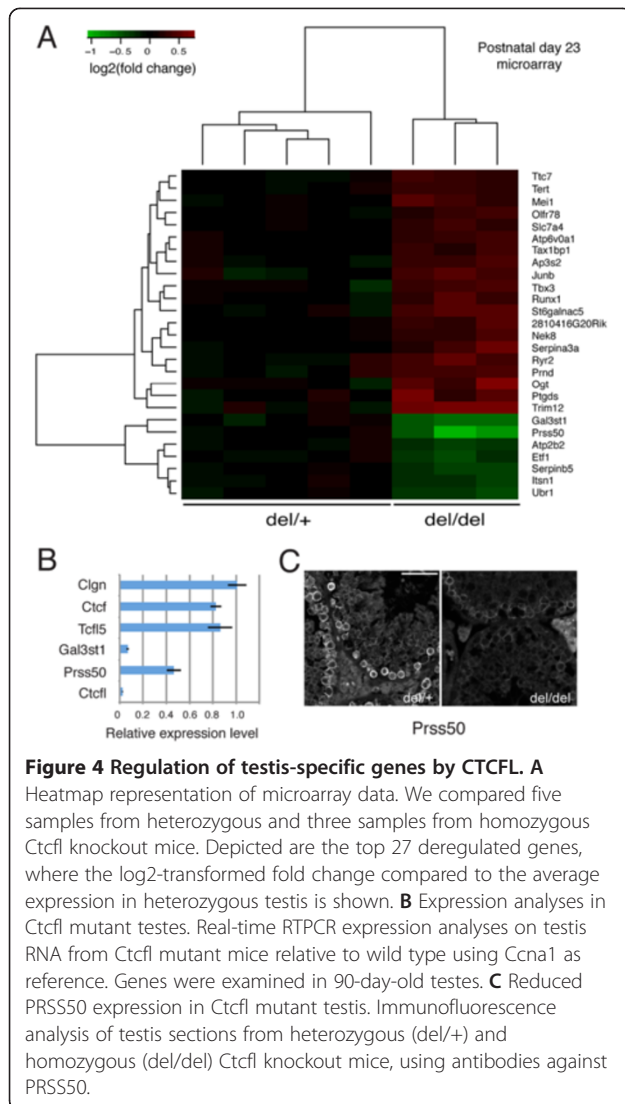
The whole genome DNA-binding profile for CTCF has been elucidated in several cell systems (see, for example, [6-9]). We sought to compare the DNA-binding profiles for CTCF and CTCFL in the same cell type. Since CTCF is ubiquitously expressed in the testis, whereas the presence of CTCFL is highly restricted, genome-wide DNA-

binding patterns derived from whole or partially purified testis preparations cannot be compared (see Discussion). We therefore generated ES cells, a cell type closely related to germ cells [37-40], in which expression of a V5- and GFP-tagged CTCFL protein could be induced (Figure 5A), thereby mimicking the CTCFL-positive germ cells that express both CTCFL and CTCF. Advantages of this system are furthermore the unlimited source of cells and the possibility to sort for GFP-positive cells that express the fusion protein to obtain a pure population of cells. In addition, the V5 tag permits stringent and exclusive immunoprecipitation of CTCFL. Thus, with this system genome-wide RNA (micro-arrays) and DNA-binding studies (ChIP-Seq) were carried out (Figures 5B).

Comparison of the expression of all genes on the microarray to expression of genes bound by CTCF or CTCFL revealed that the CTCFL-bound genes were, on average, more abundantly expressed (Figure 5C). The same held true when CTCFL-bound genes were compared to random gene sets (not shown). These data indicate that CTCFL associates with active genes. Several genes upregulated in CTCFL-GFP-V5-induced ES cells were also detected in the list of genes downregulated in CTCFL-deficient testes. Real-time RT-PCR confirmed that *Gal3st1*, *Prss50* and even *Stra8* expression were upregulated in CTCFL-induced ES cells (Figure 5D). Thus, CTCFL can act on male-specific germ cell genes in ES cells, and two of the most downregulated genes in CTCFL-deficient testis are upregulated in CTCFL-GFP-V5-expressing ES cells. These data underscore the notion that ES cells resemble germ cells and indicate that CTCFL acts as a male germ cell gene regulator.

#### Genome-wide binding of CTCFL and CTCF

To determine the genome-wide binding pattern of CTCFL, we used GFP-sorted CTCFLV5-GFP-induced ES cells, which express both CTCFL and CTCF. The V5



antibody was used for ChIP of CTCFL, with non-induced ES cells as control. Normal ES cells and a rabbit polyclonal antibody to CTCF [4] were used for ChIP-sequencing of CTCF. ChIP-sequencing revealed 5707 CTCFL and 37691 CTCF-binding sites (Figure 5E). To validate our data, we compared the number and position of CTCF sites determined by us with published data from the same cell type [7] and found a very high overlap (Figure 5F). Sorting the CTCFL-binding sites on the number of unique sequence reads yielded a list of genes that was headed by *Stra8* and *Prss50* (Table 2), two genes that are upregulated in CTCFL-inducible ES cells. Thus, the most prominent CTCFL sites locate at genes that are important for germ cells.

Interestingly, only 64% (3677) of CTCFL sites overlap with those of CTCF; conversely, only ~10% of CTCF sites are bound by CTCFL (Figure 5E). Despite their partial overlap, CTCFL and CTCF bind almost identical consensus

**Table 2 Top ten CTCFL-binding sites in induced ES cells**

Chr	Position	Gene	CTCF bound*	CTCF bound	
1	6	34,872,000	<i>Stra8</i>	134	Yes
2	9	110,760,000	<i>Tsp50</i>	113	Yes
3	14	103,450,500	<i>Irg1</i>	105	Yes
4	9	106,114,000	<i>Twf2</i>	101	Yes
5	9	50,260,500	-	100	Yes
6	5	125,061,000	<i>Tctn2</i>	97	Yes
7	8	107,058,500	<i>Nae1</i>	95	Yes
8	2	29,475,000	<i>Rapgef1</i>	94	Yes
9	9	108,838,500	<i>Uqcrc1</i>	93	Yes
10	12	112,970,500	<i>Bag5</i>	92	Yes

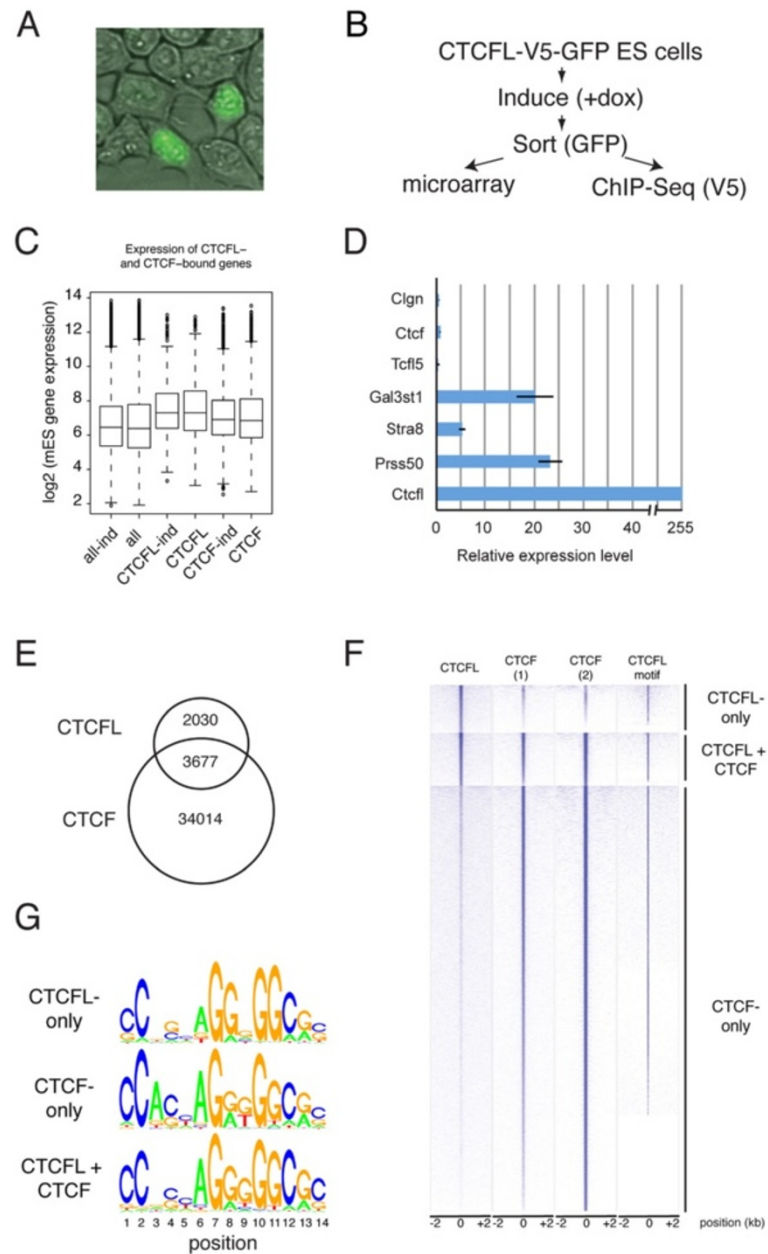
\*Sites are ranked based on the number of ChIP-sequence reads, filtered for duplicates.

sequences (Figure 5G). The most notable differences in the DNA-binding motif are the lower prevalence of a C at positions 1 and 2, the absence of A at position 3 and a lower prevalence for A at position 6, as well as a higher prevalence of G at positions 8 and 11, for the CTCFL motif relative to the one of CTCF. Whether subtle motif differences relate to differences in numbers of binding sites or to effects mediated through CTCFL and/or CTCF are questions currently under investigation. We also noted that, similar to CTCF [6,9,41], not all binding sites for CTCFL contain a consensus motif (Figure 5F).

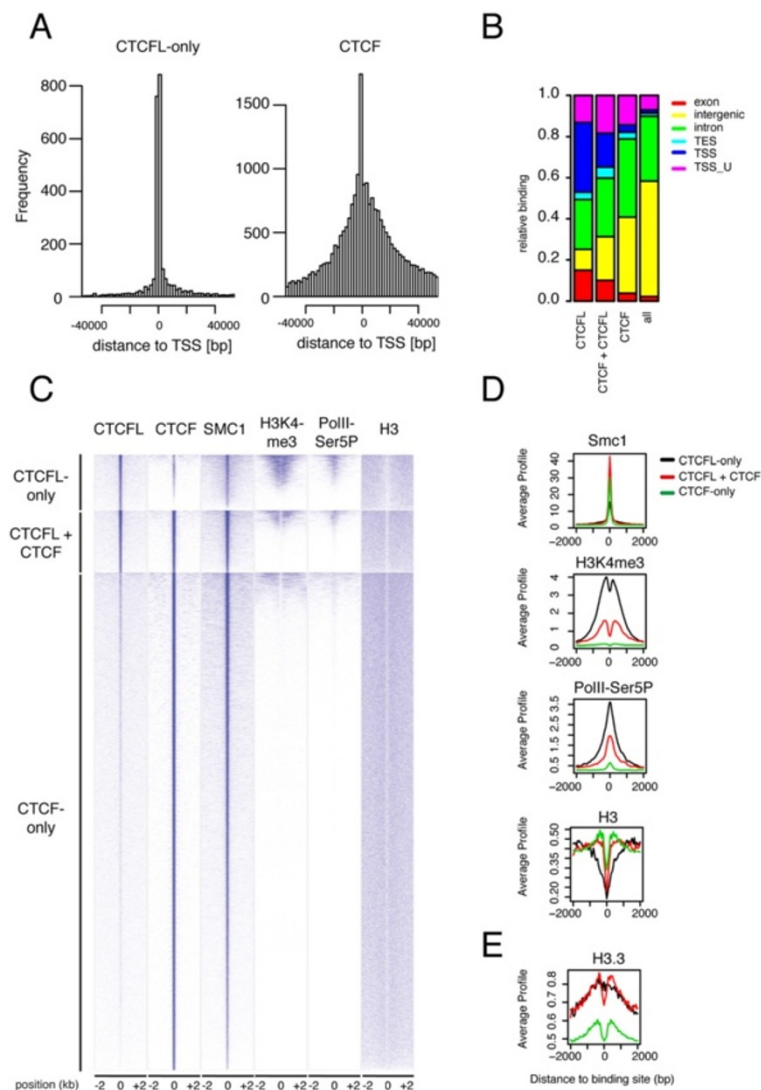
#### Nucleosome occupancy specifies binding of CTCFL versus CTCF

Further analysis of the genome-wide binding of CTCFL(L) revealed that CTCFL binds almost exclusively to CTCF consensus sites near promoter areas, in contrast to CTCF (Figure 6A, B). We next split CTCFL(L)-binding sites into three groups, i.e., CTCFL-only sites, CTCFL + CTCF sites and CTCF-only sites, and compared CTCFL(L) binding to published data sets of transcription factors and other chromatin constituents. Binding sites are shown as heatmaps, which represent individual ChIP-Seq profiles from -2 kb to +2 kb relative to the center (peak maximum) of the analyzed peaks (Figure 6C) and as cumulative profiles (Figure 6D, E), which represent average ChIP-Seq profiles. Sites were sorted for binding strength within the three subsets. This comparative analysis revealed, for example, that CTCFL colocalizes with cohesin at CTCF consensus sites that are not occupied by CTCF (Figure 6C, D). In addition, CTCFL is enriched at transcriptionally active promoters, which are marked by H3K4me3 and PolII phosphorylation on serine 5 (Figure 6C, D). By contrast, CTCF-only sites are not associated with these marks. These data confirm the observation that CTCFL associates with transcriptionally active genes (see Figure 5C).





**Figure 5 Genome-wide analysis of CTCFL expression in ES cells.** **A** Inducible expression of CTCFL-V5-GFP in ES cells. Notice the nuclear localization of CTCFL-V5-GFP in cells expressing the protein. **B** Flow chart of experiments. ES cells with a Tet-on inducible expression of a CTCFLV5-GFP transgene were sorted for GFP and used for microarray and ChIP-Seq analyses. **C** CTCFL expression and DNA binding are associated with elevated gene expression levels. We plotted gene expression levels, as determined by microarray analysis of induced (*ind*) or non-induced ES cells, for all genes (*all*), or those bound by CTCF, or CTCFL, to the respective promoter region (−2 k to +1 kb around TSS). Differences are highly significant (*p*-value CTCF-ind versus CTCFL-ind:  $5.1 \times 10^{-14}$ ; *p*-value CTCF versus CTCFL:  $5.9 \times 10^{-13}$ ). **D** Transcript analyses in ES cells expressing CTCFL-V5-GFP. Real-time RT-PCR expression analyses of CTCFL-V5-GFP-induced and GFP-sorted ES cells, relative to non-induced ES cells, for the indicated genes, referenced to *Cdk2* expression. **E** Venn diagram of DNA-binding sites for CTCFL and CTCF. **F** Clustered heatmap representation of three classes of CTCF/CTCFL-binding sites. Shown are the binding profiles of CTCFL and CTCF (1: our own data; 2: [7]) across all CTCF/CTCFL-binding sites identified in mES cells. Sites are grouped into CTCFL-only, CTCF-only, and combined CTCFL and CTCF sites. Within the three classes, data sets were sorted decreasingly from top to bottom for average binding across the interval from 2 kb to +2 kb around the identified binding peak center positions. Additionally the occurrences of predicted CTCFL motifs within these intervals are plotted. **G** Motif comparison of CTCF and CTCFL. DNA-binding motif for CTCFL-only (top panel), CTCF + CTCFL (middle panel) and CTCF-only binding sites (bottom panel).



**Figure 6 Characterization of CTCFL and CTCF binding.** **A** A large fraction of CTCFL-binding sites is located close to promoters. We determined for each CTCFL-only-binding site the distance to the nearest transcriptional start site (TSS) and plotted the frequencies of binding sites in the depicted window from  $-40$  kb to  $+40$  kb around the center of CTCFL-binding sites. CTCF is plotted as comparison. **B** Comparison of the genomic distribution of CTCF- and CTCFL-binding sites. Sites are separated into CTCF-only, CTCFL-only and (CTCF + CTCF). The entire genome is also plotted (*all*). The binding location is separated into exon, intron, intergenic, transcription start site (TSS) and transcriptional end sites (TES), and plotted as frequencies of total (Y ax). **C** Clustered heatmap representation of the three different classes of CTCF/CTCF-binding sites with respect to chromatin context. We compared CTCF and CTCFL binding to published ChIP-sequencing data sets for the cohesin complex subunit SMC1, H3K4me3, a phosphorylated form (serine 5) of RNA PolII, (PolII-Ser5P) and histone H3 [8,42,43]. **D** Cumulative profiles across the three different classes of CTCF/CTCF-binding sites with respect to chromatin context. The average ChIP-sequencing profiles are shown for the same data sets as in (C). **E** Cumulative profiles across the three different classes of CTCF/CTCF-binding sites with respect to H3.3.

When we compared binding of CTCFL(L) to that of histone H3 we noted that CTCFL preferentially binds large H3-depleted areas (Figure 6C, D). By contrast, CTCF is enriched on sites that display H3 phasing around the CTCF-binding site (Figure 6D). These sites, in turn, do not attract CTCFL. Shared CTCFL/CTCF sites associate with “intermediate” H3-free regions (Figure 6C, D). As the H3-binding site analysis was performed in ES cells that do not express CTCFL, we conclude that the H3 depletion in

these cells is not caused by CTCFL, but that H3-depleted regions appear to attract CTCFL.

It has been observed that many “H3-free” regions in the genome in actual fact do contain histones, but that these are loosely assembled and are lost upon DNA extraction with high salt [44]. The variant histone H3.3 has been shown to occupy these areas, often together with another variant histone, H2A.Z [45]. We therefore compared CTCFL binding to that of H3.3 (for which data

are available in mouse ES cells [46]) and found that these two proteins colocalize (Figure 6E), whereas CTCF does not associate with H3.3-enriched regions. We conclude that in addition to nucleotide sequence, nucleosome occupancy and composition specify the genome-wide binding of CTCFL and, surprisingly, of CTCF.

#### Competition between CTCFL and CTCF on distinct sites

ChIP-sequencing and direct ChIP experiments showed that CTCF and CTCFL bind the same site within the *Stra8* and *Prss50* promoters, but not in the *Gal3st1* promoter (Figure 7A). We therefore tested the idea that these two proteins compete for binding on selected sites. Using *in vitro* band shift assays, we confirmed that CTCFL and CTCF bind the *Stra8* and *Prss50* promoters (Figure 7B, C). When proteins were added together we did not observe a higher band, indicating that CTCF and CTCFL do not interact to bind a probe simultaneously (Figure 7C). Instead, with increasing amounts of CTCFL, the amount of bound CTCF diminished (Figure 7C), suggesting that CTCFL and CTCF compete for binding sites *in vitro*.

To examine whether competition occurs *in vivo*, we used ChIP analysis on CTCFL-induced and -non-induced ES cells using a selected number of sites. In the presence of CTCFL, the amount of bound CTCF was reduced for both *Stra8* and *Prss50* (Figure 7D). CTCFL induction had no effect on CTCF in the CTCF-only-binding site within the *Chr10* locus (Figure 7D). However, for the shared CTCFL/CTCF-binding site at *Vps18* we also saw no effect on CTCF binding (Figure 7D). These ChIP results indicate that CTCFL can compete with CTCF, but only at specific sites. To test this hypothesis on a genome-wide level, we transiently transfected GFP-CTCFL into ES cells and examined differences in CTCF binding in ES cells expressing GFP-CTCFL compared to cells not expressing this protein. As shown in Figure 7E (*left panel*), in the presence of CTCFL, binding of CTCF was reduced on ~1,100 sites, whereas binding on ~100 sites was increased. Binding of CTCF to the affected sites was significantly reduced as compared to all CTCF-binding sites (Figure 7E, *right panel*). Among sites displaying reduced CTCF binding were the *Prss50* and *Stra8* promoters, but not *Vps18* and chromosome 10 binding sites (not shown). These results are consistent with the ChIP data (Figure 7D). We conclude that in ES cells, CTCFL and CTCF compete on distinct sites.

To estimate the physiological relevance of our genome-wide ES cell data we performed ChIP using antibodies against CTCFL and CTCF on selected sites in cells isolated from wild-type testis by elutriation. Results show preferential binding of CTCFL to the *Stra8* and *Prss50* promoters, and preferential binding of CTCF to *Vps18* (Figure 7F), indicating that the differential binding pattern of CTCF and CTCFL observed in ES cells is present in testis as well.

#### CTCFL is functionally different from CTCF

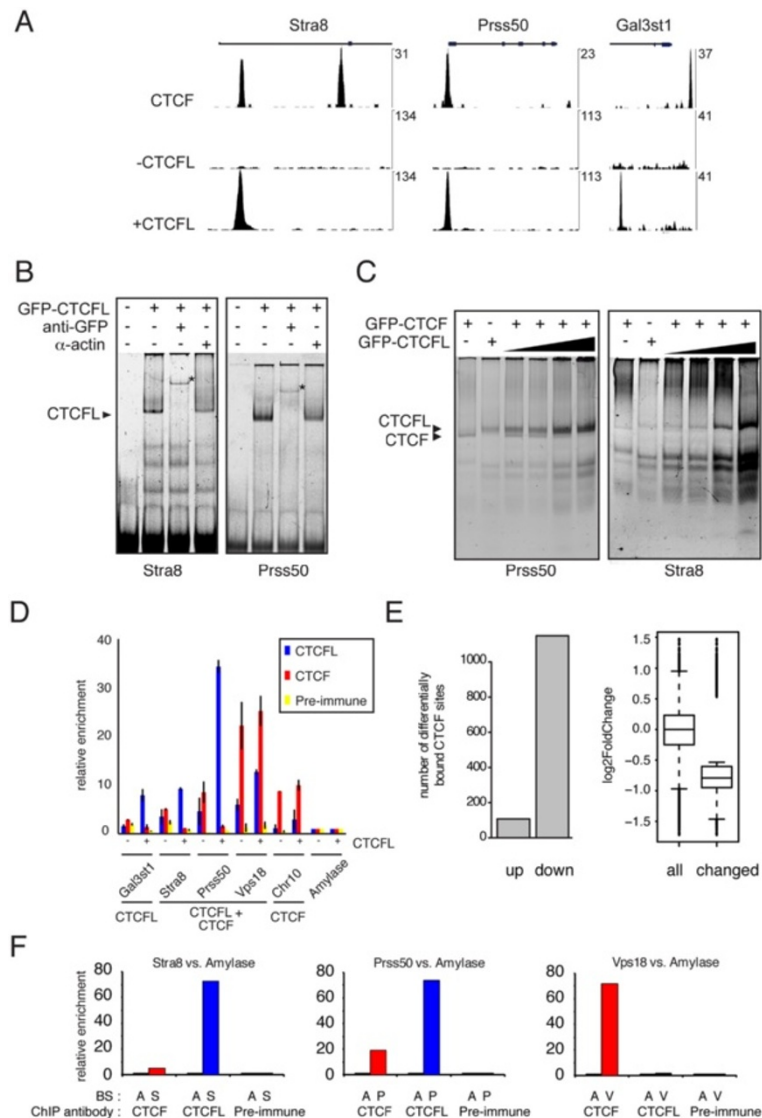
The co-expression of CTCFL and CTCF in late spermatogonia and preleptotene spermatocytes combined with their differential genome-wide binding patterns raises the question whether CTCFL and CTCF are functionally redundant, complementary or antagonistic. To test whether CTCFL can functionally substitute for CTCF, we designed an ES cell rescue assay (Figure 8A). ES cell lines were derived from mice carrying the conditional *Ctcf* knockout allele (*Ctcf*<sup>lox/lox</sup> [4]). *Ctcf* is deleted upon lentivirus-mediated Cre recombination, and these cells fail to form colonies because CTCF-deficient cells (*Ctcf*<sup>del/del</sup>) do not survive. A rescue of cell death by concurrent introduction of CTCFL would show that the two proteins compensate for each other.

Using this strategy we co-expressed Cre with GFP-tagged mouse CTCF, YFP-tagged chicken CTCF or GFP-tagged mouse CTCFL in *Ctcf*<sup>lox/lox</sup> ES cells (Figure 8A). Resulting colonies were analyzed on the DNA level for Cre-mediated CTCF deletion of *Ctcf*<sup>lox/lox</sup> into *Ctcf*<sup>del/del</sup> (Figure 8B) and on the protein level for expression of endogenous or exogenous protein (Figure 8C). A few surviving colonies transduced with Cre-only were observed (Table 3), but these had not performed the Cre-mediated CTCF deletion completely and thus still expressed endogenous CTCF (Figure 8B, C). However, nearly all colonies transduced with N- or C-terminally tagged mouse CTCF, or with C-terminally tagged chicken CTCF, had deleted endogenous *Ctcf*<sup>lox/lox</sup>, and expressed fluorescently tagged exogenously introduced protein (Figure 8B, C, Table 3). Thus GFP-tagged mouse CTCF and even chicken CTCF, which is 96% identical at the amino acid level to mouse, can functionally substitute for endogenous CTCF.

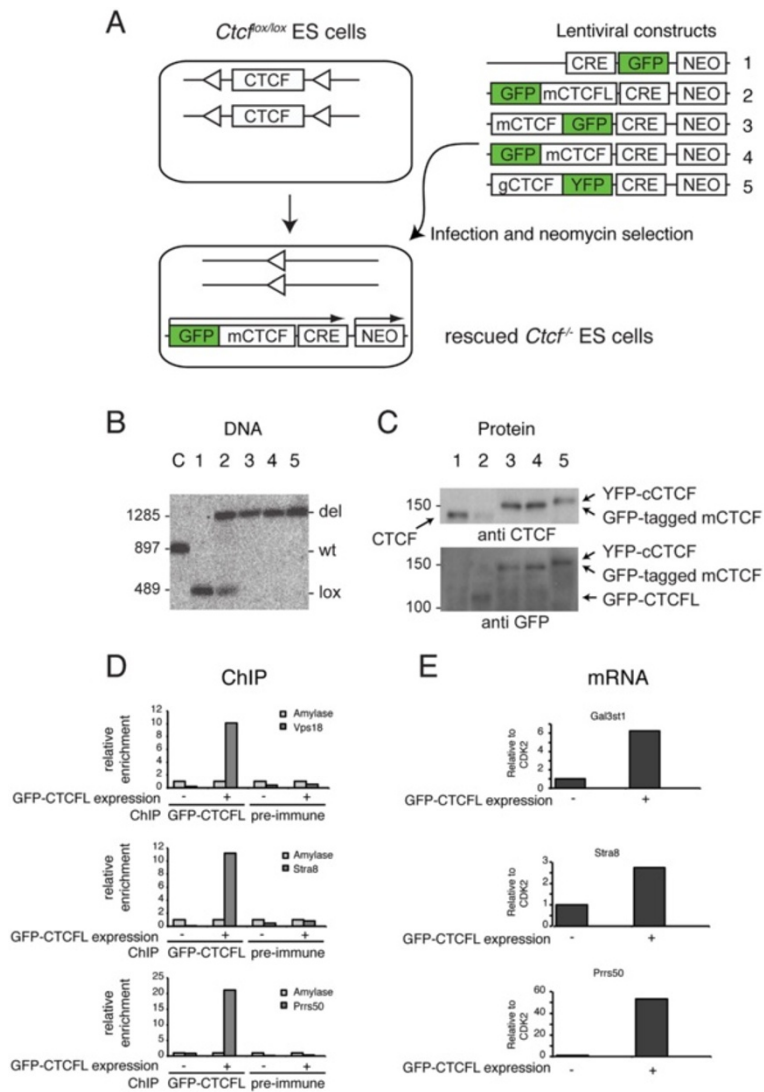
Strikingly, rescue experiments with GFP-tagged mouse CTCFL yielded no ES cells in which both endogenous *Ctcf*<sup>lox/lox</sup> alleles were deleted and wild-type protein was replaced (Figure 8B, C, Table 3). These data indicate that either CTCFL and CTCF are not interchangeable or that GFP-CTCFL is not a functional protein. To demonstrate that GFP-CTCFL is functional, we transiently transfected the protein into ES cells and examined DNA binding of GFP-CTCFL on selected sites and the induction of expression of testis-specific genes. ChIP experiments showed that GFP-CTCFL binds the three selected sites (Figure 8D) and that *Gal3st1*, *Stra8* and *Prss50* expression is induced in ES cells expressing this fusion protein (Figure 8E). These data demonstrate that GFP-CTCFL is functional.

#### Discussion

We have used a combination of approaches and technologies to unravel the physiological function of the testis-specific paralog of CTCF, called CTCFL or BORIS. We find that CTCFL is only expressed in late spermatogonia and preleptotene spermatocytes, and that CTCFL-deficient mice



**Figure 7 Characterization of CTCF and CTCFL binding.** **A** Examples of CTCF- and CTCFL-binding site location. The genomic location of CTCF (upper part) and CTCFL (middle and bottom parts) binding sites in the absence (-CTCFL, middle) or presence (+CTCFL, bottom) of CTCFL, within the Stra8, Prss50 and Gal3st1 genes. The vertical axes show the number of unique sequence reads. **B** CTCFL binds to Stra8 and Prss50. Band shift analyses of GFP-CTCFL on Stra8 and Prss50 fragments. GFP-CTCFL binding can be super shifted (marked with asterisks) with anti-GFP, but not with an Actin antibody. Band shifts were performed under excess probe conditions. **C** In vitro effect of CTCFL on CTCF binding. Band shift analyses with GFP-CTCF and/or GFP-CTCFL on Prss50- and Stra8-binding sites. GFP-CTCFL is added in increasing amounts (1-, 2-, 5- and 10-fold compared to GFP-CTCF). To allow competition, the band shift was performed under probe-limiting concentrations. **D** Cellular effect of CTCFL on CTCF binding. ChIP analyses with CTCFL (blue), CTCF (red) and pre-immune (yellow) antisera in ES cells that were either non-transfected (-) or transiently transfected CTCFL-V5-GFP (+). According to ChIP-sequencing data, Prss50, Stra8 and Vps18 bind both CTCF and CTCFL, whereas Gal3st1 only binds CTCFL, and Chr10 only binds CTCF. A CTCF- and CTCFL-negative site within the Amylase gene is used as reference and set to 1. Error bars represent standard deviations of biological replicates. **E** Competition between CTCF and CTCFL in ES cells. Genome-wide binding of CTCF was compared to that of CTCFL by ChIP-Seq using non-transfected ES cells and ES cells transiently transfected with GFP-CTCFL. The left hand panel shows the effect of CTCFL binding on shared CTCFL/CTCF sites that showed >1.5 fold difference in CTCF binding. The effect is categorized into sites with increased (up) or decreased (down) CTCF binding. The right hand panel shows a more general effect of CTCFL binding on CTCF binding. Here, we examined the change in CTCF binding in all shared CTCFL(L)-binding sites (all) compared to those shared sites that were significantly changed in CTCF binding (changed). The effect on CTCF binding is plotted as log<sub>2</sub>-fold difference. **F** In vivo CTCFL(L) binding. ChIP was performed using anti-CTCF (red) or anti-CTCFL (blue) antibodies, or pre-immune serum, on the indicated sites (A: Amylase, S: Stra8, P: Prss50, V: Vps18) in nuclei from dissociated seminiferous tubules, partly purified by elutriation. Relative enrichment is shown compared to Amylase.



**Figure 8 CTCFL is functionally different from CTCF** **A** Strategy for the rescue of CTCF-depleted ES cells. *Ctf<sup>lox/lox</sup>* ES cells were infected with lentivirus containing the Cre recombinase and/or fluorescently tagged CTCF(L) proteins. After infection neomycin-resistant colonies were picked and analyzed. *m* = mouse, *g* = chicken. **B** Analysis of *Ctf<sup>lox/lox</sup>* deletion. After infection with CRE-containing constructs, *Ctf<sup>lox/lox</sup>* ES cells were scored for the status of the conditional *Ctf* alleles by DNA blot. The position of wild-type (wt), deleted (*Ctf<sup>del</sup>*, or del) and conditional (*Ctf<sup>lox</sup>*, or lox) loci in control ES cells (**C**), non-treated *Ctf<sup>lox/lox</sup>* ES cells (**1**) and lentivirally transduced *Ctf<sup>lox/lox</sup>* ES cells (**2–5**, see panel **A** for numbering of constructs) is indicated. Cells are considered rescued when both conditional CTCF alleles have been deleted. **C** Analysis of CTCF protein expression. Neomycin-resistant colonies were grown and analyzed by Western blot for CTCF (*upper panel*) or GFP (*lower panel*) expression. Note that rescued cells are negative for endogenous CTCF. **D, E** GFP-CTCF is a functional protein. ES cells were transiently transfected and harvested after 1 day. ChIP (DNA, **D**) and RT-PCR (mRNA, **E**) analyses revealed that GFP-CTCF binds Vps18, Stra8 and Prss50 promoters (**D**) and is able to induce expression of Gal3st1, Stra8 and Prss50 (**E**).

have defects in spermatogenesis. CTCFL and CTCF are functionally different proteins. CTCFL therefore has a unique role in the adult testis. It has been proposed that CTCFL is involved in genomic imprinting of the *Igf2-H19* locus and other sites [19,24]. However, imprint-related mutations often have embryonic phenotypes [25]. We did not observe this in *Ctcf<sup>del/del</sup>* mice, and despite their reduced fertility *Ctcf<sup>del/del</sup>* mice could be bred through multiple generations. Furthermore, we have not been able

to detect DNA methylation aberrations in specific loci in *Ctcf<sup>del/del</sup>* mutant mice and in CTCFL-over-expressing cells (data not shown). This makes a role for CTCFL in DNA methylation-dependent genomic imprinting unlikely. The combined microarray data from CTCFL-deficient testis and CTCFL-expressing ES cells, and the preference of CTCFL for promoters instead suggest a function as a transcriptional regulator, required for the proper expression of a subset of male germ cell genes.

**Table 3 Rescue of CTCF-deficient ES cells by exogenously introduced GFP-CTCF(L)**

Species	Construct	Deletion in ES cells**	Partial deletion in ES cells**	No deletion in ES cells**	Functional CTCF substitution
N.a.*	GFP-CRE	0% (0/65)	28% (18/65)	72% (47/65)	No
Mouse	GFPCTCF-L	0% (0/18)	50% (9/18)	50% (9/18)	No
Mouse	CTCFGFP-CRE	83% (19/23)	0% (0/23)	17% (4/23)	Yes
Mouse	GFPCTCF-CRE	95% (40/42)	2% (1/42)	2% (1/42)	Yes
Chicken	CTCFYFP-CRE	90% (37/41)	2% (1/41)	7% (3/41)	Yes

\*: Not applicable.

\*\* : Percentage and number of clones (between brackets) in which the conditional CTCF alleles were deleted, partially deleted or not deleted are shown.

The most prominent CTCFL-binding sites in ES cells are on the promoters of the testis-specific *Stra8* and *Prss50* genes. The expression of these genes, and of *Gal3st1*, is upregulated in ES cells expressing CTCFL. Conversely, expression of *Prss50* and *Gal3st1* is downregulated in germ cells lacking CTCFL, at all ages examined, whereas *Stra8* expression is affected at some but not all ages (data not shown). We speculate that the combined transcriptional deregulation of genes causes the testicular degeneration and reduction in fertility in *Ctcf* knockout mice. Note that the expression of these genes is not completely hampered, which explains why the testicular phenotype in the knockouts is milder than the fully sterile phenotype described, for example, for *STRA8*- and *GAL3ST1*-deficient mice [29,30,35].

The phenotype of the *Ctcf<sup>del/del</sup>* mice reported here only partly matches a recent report on another strain of CTCFL-deficient mice, in which exons 1 to 8 of *Ctcf* were also deleted [26]. For example, the effect of a *Ctcf* deletion on the average testicular size and on *Gal3st1* and *Prss50* expression is similar. However, our analysis also reveals a reduction in fertility in the *Ctcf<sup>del/del</sup>* mice not noted previously [26]. In addition, the fact that some *Ctcf<sup>del/del</sup>* mice have normal testis size and others have a combination of normal and abnormal seminiferous tubules was also not described. This is relevant, as this incomplete penetrance of the *Ctcf* phenotype, even within a single testis, suggests that a stochastic mechanism determines whether CTCFL-deficient tubules degenerate or not. Finally, CTCFL was proposed to be present in round spermatids and to function during meiosis based on mRNA expression data [26]. By contrast, our data show that CTCFL is expressed earlier, just prior to the onset of meiosis, and we conclude that CTCFL protein expression precedes the developmental germ cell stages that show the major phenotypes in *Ctcf* knockout mice. We propose that in the absence of CTCFL, epigenetic marks controlled by this protein gradually break

down in a stochastic manner. Spermatogonia and primary spermatocytes exist in syncytia, in which each cell is connected with the other cells at the same step of development via intercellular bridges. Only in syncytia where the expression of CTCFL-controlled genes has been affected beyond a specific threshold will degeneration become apparent.

Neither CTCFL nor CTCF is saturating all consensus-binding sites present in the genome, and thus the DNA sequence is not the sole determinant of CTCF(L) binding. DNA methylation and hydroxymethylation are not a decisive aspect, as comparisons of DNA (hydroxy)methylation data sets to our CTCF(L)-binding sites does not provide an explanation for why CTCFL and CTCF occupy different binding sites (data not shown) [47]. Instead, the data suggest that binding of CTCFL and of the “master weaver” CTCF is specified by nucleosome occupancy and composition. We find that CTCFL prefers CTCF consensus sites in promoters that are embedded in regions that appear to be nucleosome-free. By contrast, CTCF is enriched on distinct sites, which are devoid of histone H3 on the binding site itself, but which are surrounded by ordered, or “phased,” nucleosomes. This preference of CTCF has already been described [11-13].

It has recently been shown that unstable nucleosomes are lost when histones are prepared with conventional conditions; thus, regions containing these histones appear as nucleosome-free in the analysis, but are in reality not free [45]. Nucleosomes containing the variant histone H3.3 are quite unstable, and those containing both H3.3 and H2A.Z even less [44]. Since we find a correlation between CTCFL binding and H3.3 occupancy in ES cells, H3.3 and H3.3/H2A.Z might be determinant factors able to attract CTCFL and evict CTCF. It is important to realize that in ES cells H3.3-enriched genomic regions do not require CTCFL to be set up, yet the protein prefers such areas after its induction. A similar situation may exist in testis, i.e., specific H3.3/H2A.Z-containing regions might be set up during early phases of spermatogenesis; upon its expression, CTCFL “lands” on these regions, possibly evicting CTCF from some promoters. Notably, during male meiosis, and thus subsequent to CTCFL expression, H3.3 is incorporated into unsynapsed chromatin, which is transcriptionally inactive [48]. The function of CTCFL might be to ensure the expression status of genes by distinguishing specific promoter-associated H3.3 domains from whole chromosome domains that also contain H3.3. Through its interaction with SET1A [49], CTCFL might enhance H3K4 trimethylation at a subset of its binding sites.

The cohesin complex has a role in chromosome segregation, DNA-damage repair and gene regulation [50]. Although cohesin does not have a typical DNA-binding motif, it was shown to bind primarily to CTCF consensus sites [16,17,51]. Moreover, the SA2 subunit of cohesin directly interacts with the C-terminus of CTCF [52].

Cohesin's role in gene regulation therefore seems tied to that of CTCF. Recent studies revealed that also in ES cells cohesin binding largely overlaps with that of CTCF; however, there are ~2,000 cohesin sites with a CTCF motif that do not bind CTCF, while ~270 other cohesin sites do not have a CTCF consensus site [10]. Our data suggest that CTCFL binds these ~2,000 cohesin sites in CTCFL-GFP-V5-expressing ES cells.

However, in normal ES cells CTCFL is not expressed, raising the questions how a specific nucleosome composition and occupancy can be built around CTCF consensus sites that appear not to be occupied by CTCF, and how cohesin can stably bind these very same sites. We hypothesize that these sites might be bound by a modified form of CTCF, such as poly(ADP-ribosyl)ated CTCF [53]. This protein would not be able to bind DNA tightly and could be replaced very efficiently by CTCFL. Perhaps another molecular function of CTCFL in the testes is to interfere with and/or change the dynamics of CTCF and cohesin-mediated chromatin looping.

We observed competition between CTCF and CTCFL in ES cells, but only on a small subset of all CTCF-binding sites. Nucleosome occupancy and composition, CTCFL(L) expression levels and posttranslational modifications on CTCFL(L) could all determine whether competition between the proteins occurs on a given site. Our data reveal that CTCF and CTCFL co-localize within the nuclei of late spermatogonia and preleptotene spermatocytes, and the proteins might therefore also compete *in vivo*. ChIP experiments in testis extracts indeed reveal preferential binding of CTCFL at the *Stra8* and *Prss50* promoters and exclusive binding of CTCF to the *Vps18* site. These data are consistent with binding profiles in ES cells. If competition on the *Stra8* and *Prss50* genes does occur *in vivo*, then CTCFL could be a gene activator by preventing the binding of CTCF. In *Ctcf* knockout mice binding of CTCF to these genes might actually diminish their expression. However, CTCF is ubiquitously expressed in the testis, whereas CTCFL is only transiently present in spermatogonia and preleptotene germ cells. One would expect to see significant binding of CTCF to the *Stra8* and *Prss50* sites in the testicular extracts that we used, since most cells in these extracts contain CTCF and not CTCFL. The questions why CTCF is not highly enriched on the *Stra8* and *Prss50* promoters in testis, and whether these proteins compete *in vivo* can only be answered once there are tools available to isolate CTCFL-positive and -negative cell populations from testis so that genome-wide analyses can be performed on purified testicular fractions.

In human germ cell tumors, CTCFL is specifically upregulated in spermatocytic seminomas, which are benign testicular tumors originating from a spermatogonium or primary spermatocyte [54]. This fits with our observed cellular localization of CTCFL and could potentially point

to an oncogenic role for CTCFL in these tumors. In fact, CTCFL belongs to the group of cancer testis antigens (CTAs), genes that are normally expressed in testis yet aberrantly expressed in a variety of cancers. One model holds that competition between CTCF and CTCFL plays a role in tumorigenesis, i.e., aberrant CTCFL expression would displace CTCF, and affect DNA methylation and the expression of other CTAs, including the *NY-ESO-1* and *MAGE-A1* genes [22,23], and even other important genes, such as the *TERT* gene, which encodes telomerase [55]. However, while there might be a relationship between DNA demethylation and the expression of CTAs [56], recent reports have shown that expression of CTCFL alone is not sufficient to induce expression of CTAs [27,57]. Furthermore, our data in CTCFL-deficient testis indicate that, if anything, CTCFL represses the *Tert* gene instead of activating it. To address a potential role of CTCFL in cancer, a correlation analysis of CTCFL binding, nucleosome occupancy and composition, and CTA expression in different types of cancers might be more revealing.

## Conclusions

The three-dimensional folding of the eukaryotic genome serves to compact DNA while allowing gene expression. CTCF has been termed the "master weaver" of the genome, since this protein is a key coordinator of chromatin loop formation. In this study we have analyzed the physiological function and DNA-binding profile of CTCFL, a protein that is highly similar to CTCF but that is only expressed in the male germ line. Using a combination of cell biological, biochemical and bioinformatics approaches, we show that CTCF and CTCFL are functionally different proteins that bind to similar sites in the genome, but whose binding does not overlap completely. Our data suggest that nucleosome composition specifies the genome-wide binding of both CTCFL and CTCF. We show that CTCFL is only transiently expressed, in spermatogonia and preleptotene spermatocytes, prior to male meiosis. We propose that during its expression CTCFL occupies a subset of promoters and thereby maintains the expression of selected male germ cell genes.

## Methods

### RACE PCR and RNase protection assay

Human CTCFL was shown to consist of 23 isoforms with variations in N- and C-termini and zinc finger modules with different DNA-binding and transcriptional characteristics [58]. To analyze the genomic organization of the murine *Ctcf* gene, we cloned the 5' end of the *Ctcf* cDNA by a rapid amplification of cDNA ends-polymerase chain reaction (RACE-PCR) procedure, using first choice RACE ready testicular cDNA from Balb/c mice (Ambion) and nested oligos (see Table 4 for sequence). Compared to the

published murine *Ctcf* sequence [19], the RACE-PCR-derived first *Ctcf* exon is smaller and lacks an upstream ATG, and it is preceded by an intron of 489 bp and an additional exon of 130 bp (see also panel 1C). The sequence of the *Ctcf* 5' end product has been submitted to Genbank (accession no.: EU154995). Our cDNA structure matches the HAVANA/VEGA curated sequences in Ensembl, Build 36 [59].

RNase protection assay (RPA) was performed according to the manufacturer's instructions (RPAIL, Ambion). For *Ctcf* the RPA was performed on poly A + purified RNA with probes spanning *Ctcf* exons 8 and 9 (protecting 146 bp), spanning bp 1–220 of the *Ctcf* race PCR product (up to oligo 3, see figure S1C) or, alternatively, spanning bp131–220 (protecting 89 bp), to detect the existence of another start site [19]. For *Ctcf* the RPA was performed on total RNA with a probe protecting 99 bp of *Ctcf* exon 2. The RPA with the 5' end RACE confirms that *Ctcf* mRNA contains the additional upstream exon as identified by the RACE PCR. We found no evidence for alternative splicing in murine *Ctcf*.

#### Mouse models

To generate the *Ctcf* and *Ctcf* knockin alleles we inserted a *Gfp*-encoding cDNA, followed by a *Loxp*-flanked neomycin selection cassette, in the *Ctcf* and *Ctcf* exons, respectively, that contain the ATG start codons. Insertion of GFP immediately downstream of the translational start sites yielded *Ctcf<sup>gfp-neo</sup>* and *Ctcf<sup>gfp-neo</sup>* ES cells. Homology arms were generated by cloning from the RPCI21 129 PAC library (Geneservice). Constructs were sequenced, electroporated into isogenic ES cells (129/IB10) and neomycin- (or, later on, puromycin-) selected, analyzed by Southern blot and PCR, and injected into C57/Bl6 blastocysts.

Mice generated from *Ctcf<sup>gfp-neo</sup>* and *Ctcf<sup>gfp-neo</sup>* ES cells were crossed to transgenic mice expressing Cre to delete the *Loxp*-flanked neomycin cassette. This yielded *Ctcf<sup>gfp</sup>* mice in which the GFP is fused in frame to CTCFL and *Ctcf<sup>gfp</sup>* mice where GFP is fused to CTCF. These mice are phenotypically normal and fertile (data not shown).

The *Ctcf<sup>gfp-neo</sup>* ES cells were retargeted with a *Loxp*-flanked puromycin cassette downstream of exon 8. Mice were generated using the *Ctcf<sup>gfp-neo-puro</sup>* ES cells. The *Ctcf<sup>del</sup>* mice were subsequently generated by crossing *Ctcf<sup>gfp-neo-puro</sup>* mice to Cre-expressing mice. This resulted in the in vivo deletion of *Ctcf* exons 1–8 and both selection cassettes as these were in between the *Loxp* sites.

Mice were maintained on a C57/Bl6 background at the Erasmus MC animal care facility under specific pathogen-free conditions. Animal experiments were reviewed and approved by the Erasmus University committee of animal experiments.

#### Cell culture, transfection and infection

The *Ctcf<sup>lox/lox</sup>* ES cells were isolated de novo from CTCF conditional mice [4] and grown on plastic in the presence of LIF. Lentiviral constructs were generated with *Ctcf* and *Ctcf* cDNAs driven from a CAG promoter (CMV early enhancer/chicken  $\beta$  actin), followed by an IRES sequence and the Cre recombinase. Expression of a neomycin selection cassette was driven by a PGK promoter. Lentivirus particles were produced as described (Addgene). ES cells were infected in suspension for 4 h, plated and selected with G418 the next day. Clones were analyzed by Southern blot for the status of the *Ctcf<sup>lox/lox</sup>* conditional allele [4] and by Western blot using GFP (Abcam 32146) or CTCF antibody (BD Bioscience).

For the inducible CTCFL expressing ES cells, ROSA26-rtTA ES cells were Lipofectamine transfected (Invitrogen) with a TRE-mCTCFL-V5-GFP-neomycin construct and selected with G418. Clones were analyzed for the induction and expression of CTCFL-V5-GFP by flow cytometry for GFP (FACS Aria, BD Biosciences), and by Western blot and immunofluorescence using rat monoclonal anti-CTCFL antibodies raised against mouse CTCFL (AA 1–113 and AA 569–635) and V5 antibody (Sigma, V8012). Transient transfections of mCTCFL-V5-GFP and of GFP-mCTCFL in ES cells were done with Lipofectamine 2000 (Invitrogen).

#### Chromatin immunoprecipitation (ChIP)

\ChIP was performed as described [14] or according to the Magnify system procedure (Invitrogen). Briefly, preparation of cross-linked chromatin ( $2 \times 10^7$  cells treated with 1% formaldehyde for 10 min at room temperature), sonication of chromatin to yield fragments up to 800 bp, and immunoprecipitation with V5-Agarose beads (Sigma, A7345) or with polyclonal CTCF(L) antibodies [4] were performed as described in the Upstate protocol (<http://www.upstate.com>). Ct values from real-time PCR were normalized to input measurements, and enrichment was calculated relative to the Amylase gene. For oligos used, see Table 4. ChIP was performed on nuclei derived from induced or transiently transfected ES cells (see above) or from seminiferous tubules in which multiple testicular cell populations were first dissociated by enzymatic digestion of seminiferous tubules and subsequently isolated by elutriation.

#### ChIP-sequencing and analysis

For ChIP-sequencing a DNA library was prepared from the ChIPped DNA according to the Illumina protocol ([www.illumina.com](http://www.illumina.com)). Briefly, 10 ng of end-repaired DNA was ligated to adapters, size selected on gel ( $200 \pm 25$ -bp range) and PCR amplified using Phusion polymerase as follows: 30 s at 98 °C, 18 cycles of (10 s at 98 °C, 30 s at 65 °C, 30 s at 72 °C) and 5 min at 72 °C final extension.



**Table 4 Oligos**

<b>Oligos used for RACE PCR</b>		
<b>Name</b>	<b>Forward</b>	<b>Backward</b>
RACE	GGACACTCGTATTGGGCACATTC	CACAGGGAGCACTGAGGGCTTC
<b>Oligos used in real-time PCR for ChIP</b>		
<b>Name</b>	<b>Forward</b>	<b>Backward</b>
Gal3st1	TCCTGGGTGAGGTCAGGAAG	GGAAGTCCGAGTAGCTTCAATG
Stra8	TCCTAGAGAAGGGGTGTTACC	AGCTGACCACCACACGTTTTTC
Prss50	AGAGGAGGGTAGGGGTATCGAC	TCGCCTCAGTAATTTCTAAGC
Vps18	CTGCTGCAGTTCCTCATGTTG	GTGTGACAGATGGAGGAGCAC
Chr10	AAGGTTGGTAGCTCTGCTTGGACTGCTCG	AATGTCACAAGCAAAGAAAAGCACGCAAAT
Amylase	AATTCTCCTTGTACGGCTTCGTG	TAGCAATGATGTGCACAGCTGAA
<b>Oligos used in real-time PCR for RNA expression</b>		
<b>Name</b>	<b>Forward</b>	<b>Backward</b>
Clgn	TGCTTCTTACTCTTCTTCCG	GAAGCCAGGTGAAGCTGAGGTC
Ctcf	CCACCTGCCAAGAAGAGAAGG	GCACCTGTATTCTGATCTTCGAC
Tcf5	ACGAGATAGGAGGCGCAGAATC	GTTGTTGCTTTATCTGTCTCCG
Gal3st1 testis-specific form	GCTACTCGGAGTTCGGAAA	GACTTGCAGGGCTCTTTGG
Gal3st1	ACTGTATCCCAACATGGCCTTC	ATATCTCGCCGAGGTTGACAC
Stra8	GGCAGTTTACTCCCAGTCTGATA	CAACTTATCCAGGCTTCTTCTCT
Prss50	GACAGTTCTCTGCACTGTGAC	CACATTTCTGTCTGTTCCAGGATA
Ctcf	GCTCTGGCTGTGCACCTTACG	CCCCTGTGCCACCATCATC
Ccna1	GAGTATGCAGAGGAGATTCATCG	TCATGTAGTGAGCCTTGGGTCTG
lpcat2	AGCACCCAGTGAGGAAGAGA	TTCGTAGGTGTGATCCGTC
itfg3	ACGAGGTGTCTTCTGCCTGT	GTTCCCACTAAAGCTGCTGG
dio2	TGCAGATCCTGCCAGCTTTT	CACACTGGAATTGGGAGCAT
hgf	GATGAGTGTCCAAACAGGTG	GGTCAAATTCATGGCCAAAC
akr1c18	CCAGGCCATTCTAAGCAAGA	TCAGGGAATTTTCCAAGCTG
<b>Oligos used in EMSA</b>		
<b>Name</b>	<b>Forward</b>	<b>Backward</b>
Stra8	GGATCTGTGCTGTGTGCTCCTCGACTCCT AGAGCATGCATGTGTGGTCACTTGCGACTC CTAGAGGA	CCTCTAGGAGTCGCAAGTGACCCACACATGCATGC  TCTAGGAGTCGAGGAGGACACACAGCACAGATCCT
Prss50	ATCTAGGGGGGCCACGCAGGCTGGGCACC AGCGCACCATGAGGCGCCCGATGGAGGGCG CCATTGTGGA	CCACAATGGCGCCCTCCATCGGGCGCCTCATGGT GCGCTGGTGCCAGCCTGCGTGGCGCCCTAGATG
<b>Oligos used for generating transgenic mice and cells</b>		
<b>Name</b>		
Oligo 1	5-TCTTTTCCATCAGGGGTCGTCAC-3	
Oligo 2	5-GAGAAGCGGATCACATGGTCTG-3	
Oligo 3	5-GCACCGTTTGCAGGGTCAGGATC-3	
Oligo 4	5-TCCAAATCACAGCGCCACCTACAG-3	
Oligo 5	5-GGTTCTTAGAGATAGGTTTCTCTG-3	
Oligo 6	5-GGTGTTCTGCTGGTAGTGGTC-3	
Oligo 7	5-CGGCATCAGAGCAGCCGATTG-3	
Oligo 8	5-GTTATGATCTGGGTATCGTCCACTG-3	

Cluster generation was performed using the Illumina Cluster Reagents preparation, and the library was sequenced on the Illumina Genome Analyzer Iix platform to generate 36-bp reads. Images were recorded and analyzed by the Illumina Genome Analyzer Pipeline (GAP) and processed using the IPAR (Integrated Primary Analysis Reporting Software) and the GAP. The resultant sequences were mapped against NCBI Build 37.1 of the mouse genome using the ELAND alignment software (Illumina).

Published data sets generated for mouse ES cells were downloaded from NCBI's gene expression omnibus (GEO). We used the following data sets: H3: GSM587479, CTCF: GSM288351 [7], Smc1: GSM560341 and GSM560342 [8], H3K4me3: GSM594581 [42], PolIIser5p: GSM515662 [43] and H3.3-HA: GSM423355 [46]. Reads were converted to the fastq format and aligned to a precompiled mm9 reference index with BOWTIE [60]. In case multiple sequencing lanes were available, fastq files were merged before alignment. Unambiguously mapped and unique reads were kept for subsequent generation of binding profiles and calling of peaks using MACS with an  $\text{fdr} < 0.05$  [61]. Downstream analysis was done in R/BioConductor (<http://www.bioconductor.org>), partly according to published strategies [62].

For comparative ChIP-Seq analysis mapped reads were transformed to continuous binding profiles. Those were used to collect data in 4-kb windows spanning CTCF and CTCFL binding sites. The binding sites were grouped into three classes based on intersection analysis: sites bound by CTCF only, CTCFL only, or both CTCF and CTCFL. The binding data were binned across binding sites in 50-bp windows, and the mean was calculated at each position in order to generate cumulative binding profiles. Alternatively the complete data were plotted in heatmaps. The identified CTCFL-binding motif was used to scan the mm9 genome using the Patser tool [63] and plotted as a heatmap after the motif data had been binned, as explained for the binding profiles.

### RNA analyses

Expression analyses by real-time PCR were performed as follows: total RNA was isolated with RNA-Bee (Tel-Test Inc.). RNA was reverse transcribed (RT) with a combination of random and oligo-dT primers by Superscript reverse transcriptase (Invitrogen), and real-time RT-PCR was performed with a Sybrgreen platform on a Bio-Rad CFX Cycler. For oligos used, see Supplemental Information.

For testis and ES cell microarray analysis, the purity and quality of the isolated RNA were assessed by RNA 6000 Nano assay on a 2100 Bioanalyzer (Agilent Technologies). Then 5  $\mu\text{g}$  testes RNA was used for the production of cRNA. Labeled cRNA was hybridized to the GeneChip Mouse Genome 430 2.0 array oligonucleotide microarray (Affymetrix) according to manufacturer's

recommendations; 300 ng ES cell RNA was used for production of end-labeled biotinylated ssDNA. Labeled ssDNA was hybridized to the Mouse Gene 1.0ST array (Affymetrix) according to manufacturer's recommendations. Measured intensity values were analyzed using the Gene Expression Console (Affymetrix) and normalized by quantile normalization.

Scanned microarray data were processed using R/Bioconductor using standard procedures. Normalization and background correction were done by RMA. Differentially expressed genes were determined using the limma package within R [64]. For visualization the mean expression was determined across the heterozygous samples, which was then subtracted from the expression levels for the individual samples. For the analysis of the association between gene expression and CTCFL/CTCF binding, RefSeq genes were downloaded from the UCSC genome browser homepage. For each gene represented on the MoGene 1.0 ST array, the nearest CTCF or CTCFL site was calculated. Genes with a binding site within an interval from  $-2$  kb to  $+1$  kb around the transcriptional start sites were determined as bound. Log<sub>2</sub>-transformed expression values derived from Affymetrix analysis of mES cells was then plotted for the identified genes.

All Chip-seq and Microarray datasets are available at NCBI Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) under accessions: GSE34091, GSE34092, GSE34093 and GSE34094.

### Electrophoretic mobility shift analysis (EMSA) or band shift analysis

Nuclear extracts were obtained from mock-transfected HEK 293 T cells and HEK 293 T transfected with pEGFP, pEGFP-mCTCF or pEGFP-mCTCFL. After 24 h, cells were harvested, washed with cold PBS, resuspended in buffer 1 [10 mM HEPES; 10 mM KCl; 0.25 mM EDTA pH 8, 0.125 M EGTA-K pH 8, 0.5 mM Spermidin, 0.1%, NP40, 1 mM DTT, protease inhibitor cocktail set I (Calbiochem)] and incubated for 10 min on ice. Cells were then centrifuged 5 min at 1,500 rpm. The supernatant was removed, and the pellet was resuspended in 50  $\mu\text{l}$  of buffer 2 [20 mM HEPES; 0.4 M NaCl; 0.25 mM EDTA, 1.5 mM MgCl<sub>2</sub>, 0.5 mM DTT, protease inhibitor cocktail set I (Calbiochem)] and incubated 1 h at 4 °C. Samples were centrifuged for 30 min at 15,000 rpm, and the supernatant (nuclear extract) was frozen at  $-70$  °C until used.

Radiolabeled probes were generated by PCR of genomic DNA (for oligos used, see Table 4). In all cases the PCR was performed in a final volume of 50  $\mu\text{l}$  containing 3  $\mu\text{l}$  of [ $\alpha$ -<sup>32</sup>P]dCTP (Hartman Analytic), 20 ng of genomic DNA (from K562 cells), 0.2 mM each dNTP, 0.5  $\mu\text{M}$  of each primer and 1U of DFS-Taq DNA polymerase (BIORON). The PCR fragments were purified using the Wizard SV Gel and PCR CleanUp System (Promega).

The EMSA reaction was performed by mixing 10 µg of the nuclear extract with 6 µl of EMSA buffer (1.5 µg of poly-dIdC, 20 mM HEPES pH 7.5, 50 mM KCl, 5% glycerol, 0.175 mM EDTA) in a volume of 19 µl. Mixtures were pre-incubated at 25 °C for 10 min. Then 162 µl of the radiolabeled probe was added to each condition, and the resulting mixture was incubated for 30 additional min at 25 °C. For competition, 10 µg of CTCF nuclear extract, followed by increasing amounts of competing extracts, was added to the binding reaction. Then, the mixtures were pre-incubated as previously described. For supershift experiments, 1 µl of anti-CTCF mouse monoclonal antibody (BD Biosciences) or 1 µl of anti-actin (Santa Cruz, sc-1616), used as a non-specific antibody, was added to the binding reaction prior to the radiolabeled probe. Complexes were analyzed by electrophoresis on a 4% polyacrylamide gel at 160 V for 2 h with 0.5×Tris-borate-EDTA buffer. Gels were fixed using 10% acetic acid for 10 min and then dried for 30 min using a Gel Dryer (Bio-Rad). Radioactive complexes were revealed using a Molecular Imager Fx (Bio-Rad).

#### Pathological analysis of *ctcf* knockout mice

Testis weight was determined immediately after dissection. Weights were measured within the tunica albuginea, excluding the cauda epididymis. Sperm analysis and counts were performed as described [33]. The epididymis was dissected and transferred into a small conical glass grinder, and homogenized by hand. The total number of sperm present in the epididymis was counted using a Neubauer hemocytometer and a phase contrast microscope (magnification 400×). At least 200 sperm in two different samples were counted. Fertility of mice was determined by breeding the mice to multiple mates and scoring the number of offspring.

#### Immunofluorescence and immunohistochemistry techniques

For BrdU incorporation, mice were injected intraperitoneally with 1.2 µg BrdU. One hour after injection, testes were dissected, fixed in 4% paraformaldehyde or Bouin, paraffin embedded and sectioned. For H/E staining, Bouin or PFA fixed testes were fixed overnight at 4 °C, washed and dehydrated using ethanol and xylene, and embedded in paraffin. Sections of 10 µm were cut, mounted on SuperFrost Plus slides (Menzel-gläser), rehydrated and stained with H/E.

For immunofluorescence analyses, Bouin or PFA fixed and paraffin embedded testes were sectioned, treated with the microwave (three times for 5 min, 750 W) in 10 mM NaCitrate buffer (pH 6.0) to expose antigens and stained using standard procedures. Antibodies used: Rat monoclonal anti-CTCF (see above), PRSS50 (Abcam 49405) and STRA8 (Abcam 49405).

#### Live imaging in seminiferous tubules

Imaging of testis tubules was performed as described [33]. Briefly, testis were injected through the rete testis with Hoechst 33342 and Trypan blue (Sigma) in 3–5 µl of PBS, 1 h prior to testis dissection, to allow spreading of the vital DNA stain throughout the adluminal compartment of the testis tubules and uptake by nuclei on the adluminal side of the Sertoli cell barrier. Trypan blue served as a marker for injected tubules. Individual seminiferous tubules were isolated from testes using a collagenase and hyaluronidase method, and Trypan blue positive tubules were transferred into a drop of PBS + with 0.2% BSA in a live-cell chamber overlaid with PBS-saturated mineral oil. The testis tubules were examined at 33 °C, using a Zeiss LSM510NLO confocal/multiphoton setup, to allow simultaneous acquisition of phase-contrast, GFP and Hoechst images.

#### Additional file

**Additional file 1: Movie S1.** Live imaging of GFP-CTCF. Ex vivo confocal imaging of a live seminiferous tubule derived from a *Ctcf<sup>flp</sup>* knockin mouse. Images were acquired throughout tubules using a combined multiphoton (Hoechst) and confocal laser (GFP) scanning microscope setup. Images were assembled for 3D reconstruction afterwards. The GFP-CTCF fusion protein is shown in green. The DNA stain Hoechst, which was injected at the adluminal site of the testis tubule, is shown in red. Hoechst-positive cells represent Sertoli cells, leptotene stage IX and later stage spermatocytes, and spermatids, all of which are negative for GFP-CTCF. Notice the presence of the GFP-CTCF-positive cells on one side (basal lamina) of the tubule.

#### Abbreviations

BrdU: bromodeoxyuridine; ChIP: chromatin immunoprecipitation; DAB: diaminobenzidine; dCTP: deoxycytidine triphosphate; dNTP: deoxynucleoside triphosphates; DTT: dithiothreitol; EDTA: ethylenediaminetetraacetic acid; EMSA: electrophoretic mobility shift analysis; H/E: hematoxylin/eosin; HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; PCR: polymerase chain reaction; PBS: phosphate-buffered saline; PFA: paraformaldehyde; RMA: robust multi-array average; RPA: RNase protection analysis; TSS: transcription start site.

#### Competing interests

The authors declare that they have no competing interests.

#### Acknowledgements

This work was supported by the Earth and Life Sciences (ALW) and Medical Sciences (ZonMw) divisions of the Netherlands Organization for Scientific Research (NWO), the Dutch Cancer Society (KWF), the Dutch Cancer Genomics Centre (CGC) and Centre for Biomedical Genetics (CBG), an EC Integrated Project (EuTRACC), and the Spanish Fondo Investigaciones Sanitarias.

#### Author details

<sup>1</sup>Department of Cell Biology Erasmus Medical Center, Rotterdam, The Netherlands. <sup>2</sup>Institut für Genetik, Justus-Liebig-Universität, Giessen, Heinrich-Buff-Ring 58-62, 35392, Giessen, Germany. <sup>3</sup>Computational Biology Unit, Bergen Center for Computational Science, University of Bergen, Thormøhlensgate 55, N-5008, Bergen, Norway. <sup>4</sup>Department of Molecular Biology, Instituto de Biomedicina y Biotecnología de Cantabria, IBBTEC Universidad de Cantabria-CSIC-IDICAN, Santander, Spain. <sup>5</sup>Cell Cycle, Stem Cell Fate and Cancer Laboratory, Fundación Marqués de Valdecilla (IFIMAV), Santander, Spain. <sup>6</sup>Center for Biomics Erasmus Medical Center, Rotterdam, The Netherlands. <sup>7</sup>Department of Reproduction and Development Erasmus

Medical Center, Rotterdam, The Netherlands. <sup>8</sup>Cancer Genomics Center, NGL, Rotterdam, The Netherlands. <sup>9</sup>Center for Biomedical Genetics, Rotterdam, The Netherlands.

#### Authors' contributions

FS, WS, MB, HH, SD, PB, VF, MR-G, SvdN, LC, MvdR, JCB, WvIJ, BL, MDD, RR, and NG contributed to acquisition of data. FS, WS, MB, SD, PB, MR-G, WvIJ, AG, MDD, RR, FG, and NG designed the experiments, and analyzed and interpreted the data. FS, WS, MB, RR, and NG drafted the manuscript which was approved by all authors.

Received: 23 April 2012 Accepted: 18 June 2012

Published: 18 June 2012

#### References

1. Bartkuhn M, Renkawitz R: Long range chromatin interactions involved in gene regulation. *Biochimica et biophysica acta* 2008, **1783**:2161–2166.
2. Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, Lee CW, Ye C, Ping JL, Mulawadi F, et al: CTCF-mediated functional chromatin interactome in pluripotent cells. *Nature Genetics* 2011, **43**:630–638.
3. Phillips JE, Corces VG: CTCF: master weaver of the genome. *Cell* 2009, **137**:1194–1211.
4. Heath H, Ribeiro De Almeida C, Sleutels F, Dingjan G, Van De Nobelen S, Jonkers I, Ling KW, Gribnau J, Renkawitz R, Grosveld F, et al: CTCF regulates cell cycle progression of alphabeta T cells in the thymus. *Embo J* 2008, **27**:2839–2850.
5. Ribeiro De Almeida C, Heath H, Krpic S, Dingjan GM, Van Hamburg JP, Bergen I, Van De Nobelen S, Sleutels F, Grosveld F, Galjart N, Hendriks RW: Critical role for the transcription regulator CCCTC-binding factor in the control of Th2 cytokine expression. *J Immunol* 2009, **182**:999–1010.
6. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K: High-resolution profiling of histone methylations in the human genome. *Cell* 2007, **129**:823–837.
7. Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J, et al: Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 2008, **133**:1106–1117.
8. Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, et al: Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 2010, **467**:430–435.
9. Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenkov VV, Ren B: Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 2007, **128**:1231–1245.
10. Nitzsche A, Paszkowski-Rogacz M, Matarese F, Janssen-Megens EM, Hubner NC, Schulz H, de Vries I, Ding L, Huebner N, Mann M, et al: RAD21 cooperates with pluripotency transcription factors in the maintenance of embryonic stem cell identity. *PLoS one* 2011, **6**:e19470.
11. Cuddapah S, Jothi R, Schones DE, Roh TY, Cui K, Zhao K: Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome research* 2009, **19**:24–32.
12. Fu Y, Sinha M, Peterson CL, Weng Z: The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS genetics* 2008, **4**:e1000138.
13. Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sidow A: Determinants of nucleosome organization in primary human cells. *Nature* 2011, **474**:516–520.
14. van de Nobelen S, Rosa-Garrido M, Leers J, Heath H, Soochit W, Joosen L, Jonkers I, Demmers J, van der Reijden M, Torrano V, et al: CTCF regulates the local epigenetic state of ribosomal DNA repeats. *Epigenetics & chromatin* 2010, **3**:19.
15. Nativo R, Wendt KS, Ito Y, Huddleston JE, Uribe-Lewis S, Woodfine K, Krueger C, Reik W, Peters JM, Murrell A: Cohesin is required for higher-order chromatin conformation at the imprinted *Igf2-H19* locus. *PLoS Genet* 2009, **5**:e1000739.
16. Parelho V, Hadjur S, Spivakov M, Leleu M, Sauer S, Gregson HC, Jarmuz A, Canzonetta C, Webster Z, Nesterova T, et al: Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 2008, **132**:422–433.
17. Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T, et al: Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 2008, **451**:796–801.
18. Kurukuti S, Tiwari VK, Tavosoidana G, Pugacheva E, Murrell A, Zhao Z, Lobanenkov V, Reik W, Ohlsson R: CTCF binding at the *H19* imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to *Igf2*. *Proc Natl Acad Sci U S A* 2006, **103**:10684–10689.
19. Loukinov DI, Pugacheva E, Vatolin S, Pack SD, Moon H, Chernukhin I, Mannan P, Larsson E, Kanduri C, Vostrov AA, et al: BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. *Proc Natl Acad Sci U S A* 2002, **99**:6806–6811.
20. Hore TA, Deakin JE, Marshall Graves JA: The evolution of epigenetic regulators CTCF and BORIS/CTCF in amniotes. *PLoS Genet* 2008, **4**:e1000169.
21. D'Arcy V, Abdullaev ZK, Pore N, Docquier F, Torrano V, Chernukhin I, Smart M, Farrar D, Metodiev M, Fernandez N, et al: The potential of BORIS detected in the leukocytes of breast cancer patients as an early marker of tumorigenesis. *Clin Cancer Res* 2006, **12**:5978–5986.
22. Vatolin S, Abdullaev Z, Pack SD, Flanagan PT, Custer M, Loukinov DI, Pugacheva E, Hong JA, Morse H 3rd, Schrupp DS, et al: Conditional expression of the CTCF-paralogous transcriptional factor BORIS in normal cells results in demethylation and derepression of *MAGE-A1* and reactivation of other cancer-testis genes. *Cancer Res* 2005, **65**:7751–7762.
23. Hong JA, Kang Y, Abdullaev Z, Flanagan PT, Pack SD, Fischette MR, Adnani MT, Loukinov DI, Vatolin S, Risinger JI, et al: Reciprocal binding of CTCF and BORIS to the NY-ESO-1 promoter coincides with derepression of this cancer-testis gene in lung cancer cells. *Cancer Res* 2005, **65**:7763–7774.
24. Jelinic P, Stehle JC, Shaw P: The testis-specific factor CTCFL cooperates with the protein methyltransferase PRMT7 in *H19* imprinting control region methylation. *PLoS Biol* 2006, **4**:e355.
25. Tomizawa S, Sasaki H: Genomic imprinting and its relevance to congenital disease, infertility, molar pregnancy and induced pluripotent stem cell. *J Hum Genet* 2012, **57**:84–91.
26. Suzuki T, Kosaka-Suzuki N, Pack S, Shin DM, Yoon J, Abdullaev Z, Pugacheva E, Morse HC 3rd, Loukinov D, Lobanenkov V: Expression of a testis-specific form of Gal3st1 (CST), a gene essential for spermatogenesis, is regulated by the CTCF paralogous gene BORIS. *Mol Cell Biol* 2010, **30**:2473–2484.
27. Kosaka-Suzuki N, Suzuki T, Pugacheva EM, Vostrov AA, Morse HC 3rd, Loukinov D, Lobanenkov V: Transcription factor BORIS (Brother of the Regulator of Imprinted Sites) directly induces expression of a cancer-testis antigen, TSP50, through regulated binding of BORIS to the promoter. *J Biol Chem* 2011, **286**:27378–27388.
28. Sleutels F, Zwart R, Barlow DP: The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* 2002, **415**:810–813.
29. Anderson EL, Baltus AE, Roepers-Gajadien HL, Hassold TJ, de Rooij DG, van Pelt AM, Page DC: *Stra8* and its inducer, coincides with, regulate meiotic initiation in both spermatogenesis and oogenesis in mice. *Proc Natl Acad Sci U S A* 2008, **105**:14976–14980.
30. Mark M, Jacobs H, Oulad-Abdelghani M, Dennefeld C, Feret B, Vernet N, Codreanu CA, Chambon P, Ghyselinck NB: *STRA8*-deficient spermatocytes initiate, but fail to complete, meiosis and undergo premature chromosome condensation. *J Cell Sci* 2008, **121**:3233–3242.
31. Oulad-Abdelghani M, Bouillet P, Decimo D, Gansmuller A, Heyberger S, Dolle P, Bronner S, Lutz Y, Chambon P: Characterization of a premeiotic germ cell-specific cytoplasmic protein encoded by *Stra8*, a novel retinoic acid-responsive gene. *The Journal of cell biology* 1996, **135**:469–477.
32. Zhou Q, Nie R, Li Y, Friel P, Mitchell D, Hess RA, Small C, Griswold MD: Expression of stimulated by retinoic acid gene 8 (*Stra8*) in spermatogenic cells induced by retinoic acid: an in vivo study in vitamin A-sufficient postnatal murine testes. *Biol Reprod* 2008, **79**:35–42.
33. Akhmanova A, Mausset-Bonnefont AL, van Cappellen W, Keijzer N, Hoogenraad CC, Stepanova T, Drabek K, van der Wees J, Mommaas M, Onderwater J, et al: The microtubule plus-end-tracking protein CLIP-170 associates with the spermatid manchette and is essential for spermatogenesis. *Genes Dev* 2005, **19**:2501–2515.
34. Cheng CY, Mruk DD: Cell junction dynamics in the testis: Sertoli-germ cell interactions and male contraceptive development. *Physiol Rev* 2002, **82**:825–874.

35. Honke K, Hirahara Y, Dupree J, Suzuki K, Popko B, Fukushima K, Fukushima J, Nagasawa T, Yoshida N, Wada Y, Taniguchi N: **Paranodal junction formation and spermatogenesis require sulfoglycolipids.** *Proc Natl Acad Sci U S A* 2002, **99**:4227–4232.
36. Xu HP, Yuan L, Shan J, Feng H: **Localization and expression of TSP50 protein in human and rodent testes.** *Urology* 2004, **64**:826–832.
37. Geijsen N, Horoschak M, Kim K, Gribnau J, Eggan K, Daley GQ: **Derivation of embryonic germ cells and male gametes from embryonic stem cells.** *Nature* 2004, **427**:148–154.
38. Kanatsu-Shinohara M, Shinohara T: **Culture and genetic modification of mouse germline stem cells.** *Ann N Y Acad Sci* 2007, **1120**:59–71.
39. Sharova LV, Sharov AA, Piao Y, Shaik N, Sullivan T, Stewart CL, Hogan BL, Ko MS: **Global gene expression profiling reveals similarities and differences among mouse pluripotent stem cells of different origins and strains.** *Dev Biol* 2007, **307**:446–459.
40. Zwaka TP, Thomson JA: **A germ cell origin of embryonic stem cells?** *Development* 2005, **132**:227–233.
41. Chen Q, Lin L, Smith S, Huang J, Berger SL, Zhou J: **CTCF-dependent chromatin boundary element between the latency-associated transcript and ICP0 promoters in the herpes simplex virus type 1 genome.** *Journal of virology* 2007, **81**:5192–5201.
42. Creighton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al: **Histone H3K27ac separates active from poised enhancers and predicts developmental state.** *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**:21931–21936.
43. Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, Sharp PA, Young RA: **c-Myc regulates transcriptional pause release.** *Cell* 2010, **141**:432–445.
44. Jin C, Felsenfeld G: **Nucleosome stability mediated by histone variants H3.3 and H2A.Z.** *Genes & development* 2007, **21**:1519–1529.
45. Jin C, Zang C, Wei G, Cui K, Peng W, Zhao K, Felsenfeld G: **H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions.** *Nature Genetics* 2009, **41**:941–945.
46. Goldberg AD, Banaszynski LA, Noh KM, Lewis PW, Elsaesser SJ, Stadler S, Dewell S, Law M, Guo X, Li X, et al: **Distinct factors control histone variant H3.3 localization at specific genomic regions.** *Cell* 2010, **140**:678–691.
47. Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PA, Rappilber J, Helin K: **TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity.** *Nature* 2011, **473**:343–348.
48. Van der Heijden GW, Derijck AA, Posfai E, Giele M, Pelczar P, Ramos L, Wansink DG, Van der Vlag J, Peters AH, De Boer P: **Chromosome-wide nucleosome replacement and H3.3 incorporation during mammalian meiotic sex chromosome inactivation.** *Nature Genetics* 2007, **39**:251–258.
49. Nguyen P, Bar-Sela G, Sun L, Bisht KS, Cui H, Kohn E, Feinberg AP, Gius D: **BAT3 and SET1A form a complex with CTCF/BORIS to modulate H3K4 histone dimethylation and gene expression.** *Mol Cell Biol* 2008, **28**:6720–6729.
50. Peters JM, Tedeschi A, Schmitz J: **The cohesin complex and its roles in chromosome biology.** *Genes & development* 2008, **22**:3089–3114.
51. Rubio ED, Reiss DJ, Welch PL, Distechi CM, Filippova GN, Baliga NS, Aebersold R, Ranish JA, Krumm A: **CTCF physically links cohesin to chromatin.** *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**:8309–8314.
52. Xiao T, Wallace J, Felsenfeld G: **Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity.** *Molecular and cellular biology* 2011, **31**:2174–2183.
53. Yu W, Ginja V, Pant V, Chernukhin I, Whitehead J, Docquier F, Farrar D, Tavosidana G, Mukhopadhyay R, Kanduri C, et al: **Poly(ADP-ribosylation) regulates CTCF-dependent chromatin insulation.** *Nat Genet* 2004, **36**:1105–1110.
54. Looijenga LH, Stoop H, Hersmus R, Gillis AJ, Wolter Oosterhuis J: **Genomic and expression profiling of human spermatocytic seminomas: pathogenetic implications.** *Int J Androl* 2007, **30**:328–335. discussion 335–326.
55. Renaud S, Loukinov D, Alberti L, Vostrov A, Kwon YW, Bosman FT, Lobanenkov V, Benhattar J: **BORIS/CTCF-mediated transcriptional regulation of the hTERT telomerase gene in testicular and ovarian tumor cells.** *Nucleic Acids Res* 2011, **39**:862–873.
56. Woloszynska-Read A, Zhang W, Yu J, Link PA, Mhaweche-Faucegglia P, Collamat G, Akers SN, Ostler KR, Godley LA, Odunsi K, Karpf AR: **Coordinated cancer germline antigen promoter and global DNA hypomethylation in ovarian cancer: association with the BORIS/CTCF expression ratio and advanced stage.** *Clinical cancer research: an official journal of the American Association for Cancer Research* 2011, **17**:2170–2180.
57. Woloszynska-Read A, James SR, Song C, Jin B, Odunsi K, Karpf AR: **BORIS/CTCF expression is insufficient for cancer-germline antigen gene expression and DNA hypomethylation in ovarian cell lines.** *Cancer Immun* 2010, **10**:6.
58. Pugacheva EM, Suzuki T, Pack SD, Kosaka-Suzuki N, Yoon J, Vostrov AA, Barsov E, Strunnikov AV, Morse HC 3rd, Loukinov D, Lobanenkov V: **The structural complexity of the human BORIS gene in gametogenesis and cancer.** *PLoS one* 2010, **5**:e13872.
59. Wilming LG, Gilbert JG, Howe K, Trevanion S, Hubbard T, Harrow JL: **The vertebrate genome annotation (Vega) database.** *Nucleic Acids Res* 2008, **36**:D753–760.
60. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol* 2009, **10**:R25.
61. Zhang Y, Liu T, Meyer CA, Eickhout J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS: **Model-based analysis of CHIP-Seq (MACS).** *Genome Biol* 2008, **9**:R137.
62. Soler E, Andrieu-Soler C, Boer E, Bryne JC, Thongjuea S, Rijkers E, Demmers J, Ijcken W, Grosveld F: **A systems approach to analyze transcription factors in mammalian cells.** *Methods* 2011, **53**:151–162.
63. Hertz GZ, Stormo GD: **Identifying DNA and protein patterns with statistically significant alignments of multiple sequences.** *Bioinformatics* 1999, **15**:563–577.
64. Smyth GK: **Linear models and empirical Bayes methods for assessing differential expression in microarray experiments.** *Stat Appl Genet Mol Biol* 2004, **3**. Article3 <http://www.degruyter.com/view/j/sagmb.2004.3.1/sagmb.2004.3.1.1027/sagmb.2004.3.1.1027.xml>.

doi:10.1186/1756-8935-5-8

**Cite this article as:** Sleutels et al.: The male germ cell gene regulator CTCFL is functionally different from CTCF and binds CTCF-like consensus sites in a nucleosome composition-dependent manner. *Epigenetics & Chromatin* 2012 **5**:8.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

