Genome-wide and parental allele-specific analysis of CTCF and cohesin DNA binding in mouse brain reveals a tissue-specific binding pattern and an association with imprinted differentially methylated regions

Adam R. Prickett,[1] Nikolaos Barkas,[1] Ruth B. McCole,[1] Siobhan Hughes, Samuele M. Amante, Reiner Schulz, and Rebecca J. Oakey[2]

Department of Medical & Molecular Genetics, King's College London, Guy's Hospital, London, SE1 9RT, United Kingdom

DNA binding factors are essential for regulating gene expression. CTCF and cohesin are DNA binding factors with central roles in chromatin organization and gene expression. We determined the sites of CTCF and cohesin binding to DNA in mouse brain, genome wide and in an allele-specific manner with high read-depth ChIP-seq. By comparing our results with existing data for mouse liver and embryonic stem (ES) cells, we investigated the tissue specificity of CTCF binding sites. ES cells have fewer unique CTCF binding sites occupied than liver and brain, consistent with a ground-state pattern of CTCF binding that is elaborated during differentiation. CTCF binding sites without the canonical consensus motif were highly tissue specific. In brain, a third of CTCF and cohesin binding sites coincide, consistent with the potential for many interactions between cohesin and CTCF but also many instances of independent action. In the context of genomic imprinting, CTCF and/or cohesin bind to a majority but not all differentially methylated regions, with preferential binding to the unmethylated parental allele. Whether the parental allele-specific methylation was established in the parental germlines or post-fertilization in the embryo is not a determinant in CTCF or cohesin binding. These findings link CTCF and cohesin with the control regions of a subset of imprinted genes, supporting the notion that imprinting control is mechanistically diverse.

[Supplemental material is available for this article.]

DNA sequences that control transcription are frequently located in the noncoding portion of the mammalian genome (The ENCODE Project Consortium 2012). These elements can act over long distances (Noonan and McCallion 2010). The identification of these control elements is important for elucidating human genetic disease since genome-wide association studies regularly point to noncoding regions as candidates in disease etiology (Manolio 2010). One of the proteins that contributes to the regulation of gene expression across the genome is CTCF (CCCTC-binding factor), a protein with 11 zinc fingers (Filippova et al. 1996) and multiple regulatory functions (Ohlsson 2001; Gaszner and Felsenfeld 2006). CTCF can act as an insulator by blocking interactions between enhancers and promoters (Bell et al. 1999), it can directly regulate chromosomal interactions (Yusufzai and Felsenfeld 2004; Hadjur et al. 2009), and it can act as an enhancer of transcription (Kuzmin et al. 2005). CTCF binds regions of DNA with high sequence specificity and is sensitive to DNA methylation, having a lower binding affinity for methylated DNA (Mukhopadhyay et al. 2004). The canonical consensus binding motif of CTCF and the sites of CTCF binding are evolutionarily conserved between mammals and birds (Martin et al. 2011; Schmidt et al. 2012). In vitro assays

have shown that CTCF can use different combinations of its zinc fingers to bind to distinct DNA sequences (Filippova et al. 1996). CTCF interacts with a variety of other factors. In particular, the cohesin complex, best known for its role in mediating sister-chromatid cohesion during cell division, has been found to frequently colocalize with CTCF during interphase (Parelho et al. 2008; Rubio et al. 2008; Wendt et al. 2008; Xiao et al. 2011) with consequences for gene expression. At specific loci, cohesin is required for cell-type-specific long-range chromosomal interactions in cis during cellular differentiation (Hadjur et al. 2009).

Genomic imprinting refers to the parental allele-specific transcription of a subset of genes in mammals and flowering plants (Reik and Walter 2001; da Rocha et al. 2008). Roughly 140 transcripts are known to be imprinted in mammals (Schulz et al. 2008). Imprinting is controlled by epigenetic modifications that differ between the two parental genomes, including differences in DNA methylation (Li et al. 1993). Imprinted genes can occur in large, coordinately regulated clusters exemplified by the *Gnas* locus (Peters and Williamson 2007); they can form small domains such as the *Mcts2/H13* locus (Wood et al. 2008) that are comprised of only two genes (McCole and Oakey 2008), or they can be singletons like *Impact* (Hagiwara et al. 1997). In all cases, their parental allele-specific expression is ultimately due to an imprinting control region (ICR), a region of DNA that is differentially methylated between the parental alleles. The parental allele-specific methylation of a differentially methylated region (DMR) is in most cases the consequence of the sex-specific epigenetic reprogramming of the

parental germ cells (Edwards and Ferguson-Smith 2007; Bartolomei 2009). In addition, DMRs are actively protected from post-fertilization epigenetic reprogramming (Quenneville et al. 2012) and, thus, persist into adulthood. In some cases, the parental allele-specific methylation of a DMR is set up post-fertilization during early embryogenesis (somatic DMRs) (Kobayashi et al. 2012). DMRs with a direct germline origin are referred to as germline DMRs (gDMRs). Disruption of imprinted gene expression after deletion of a DMR is considered evidence that the latter functions as an ICR. There are 22 well-established gDMRs in mouse, of which 19 are maternally methylated and three are paternally methylated. Many mechanisms exist to "translate" allele-specific methylation into differential gene expression, including differential protein binding (Lewis and Reik 2006).

CTCF has long been associated with genomic imprinting due to its selective binding of the unmethylated maternal allele of the *Igf2/H19* ICR resulting in parent-of-origin-specific expression of *Igf2* and *H19* in mouse and human (Bell and Felsenfeld 2000; Hark et al. 2000; Kanduri et al. 2000; Fedoriw et al. 2004; Szabo et al. 2004). CTCF has been studied at several other imprinted loci, and it binds the unmethylated allele at the gDMRs of *Rasgrf1*, *Peg13*, *Kcnq1ot1* (Yoon et al. 2005; Fitzpatrick et al. 2007; Singh et al. 2011), and *Grb10* (Hikichi et al. 2003; Mukhopadhyay et al. 2004). CTCF-mediated regulation is postulated to be one of two major control mechanisms operating at ICRs (Lewis and Reik 2006; Kim et al. 2009). Cohesin also has been linked to imprinting through its association with CTCF at the *H19/Igf2* and *Kcnq1ot1* DMRs (Stedman et al. 2008; Lin et al. 2011), and a role for cohesin in the allele-specific organization of higher-order chromatin has been proposed (Nativio et al. 2009). Here we present the first comprehensive analysis of allele-specific CTCF and cohesin binding at all known DMRs in a single tissue, providing an unbiased assessment of the extent to which CTCF and cohesin are involved in imprinting control.

Genome-wide ChIP-seq in mouse ES cells (Chen et al. 2008; Kagey et al. 2010) and human cells (Kim et al. 2007b) has shown that CTCF and cohesin bind tens of thousands of discrete sites across the genome, and CTCF binding is enriched in and near genes, consistent with a role in the control of gene expression. Mouse embryonic stem (ES) cell data identify CTCF and cohesin binding at the gDMRs of *Peg13*, *Zim2* (*Peg3*), *Peg10*, *Grb10*, and *Mest* but not at the *H19/Igf2* ICR, even though CTCF is known to be important for imprinting regulation at this domain. Imprinting is dispensable in ES cells, where loss of imprinting frequently occurs without affecting viability in culture (Kim et al. 2007a; Rugg-Gunn et al. 2007; Frost et al. 2011). The same is true for *Dnmt1*$^{-/-}$, *Dnmt3a*$^{-/-}$, *Dnmt3b*$^{-/-}$ triple knockout mouse ES cells that consequently lack all DNA methylation imprints but yet are viable (Tsumura et al. 2006). In contrast, a differentiated tissue where imprinting plays an important role is the brain (Davies et al. 2007), and this is supported by multiple lines of evidence. Firstly, the human imprinting disorders Prader-Willi syndrome and Angelman syndrome present with behavioral and neurodevelopmental phenotypes (Cassidy et al. 2000; Lossie et al. 2001; Williams et al. 2006); secondly, of the ~140 imprinted gene transcripts in the mouse, more than 50 are expressed in brain (Wilkins 2008); thirdly, the disruption of certain mouse imprinted genes, including *Peg3* (Li et al. 1999), *Mest* (Lefebvre et al. 1998), *Nesp55* (Plagge et al. 2005), and *Grb10* (Garfield et al. 2011), results in behavioral phenotypes; finally, genome-wide allele-specific studies of transcription in mouse brain suggest that this tissue is a focus for imprinted gene expression (Gregg et al. 2010a,b; DeVeale et al. 2012).

Our analyses of CTCF and cohesin binding in mouse brain are based on ChIP-seq data of high quality and an order of magnitude higher read depth than existing data. The use of reciprocal inter-subspecies hybrid mice enabled independent interrogation of the parental alleles in terms of CTCF and cohesin binding in unprecedented detail. We examined postnatal day 21 (P21) mouse brain, a time point in development shortly after the growth spurt in neurogenesis that occurs in the first 2 wk of postnatal development (Lyck et al. 2007). In the adult mouse brain, ~56% of cells are neurons and 44% are nonneuronal cells (Fu et al. 2012). Neurons and the principle type of nonneuronal cells, the macroglia, both derive from the neuroepithelium. These data are representative of adult rather than immature brain cell types and are unaffected by long-term aging effects.

Results

We demonstrate that in mouse brain, CTCF and cohesin each bind to ~50,000 sites in the genome, with ~27,000 sites bound by both factors, indicative of CTCF and cohesin acting throughout the genome both in concert as well as independently. Genes are enriched for CTCF binding sites, while intergenic regions are depleted. The binding sites are highly enriched for the canonical consensus binding motif. CTCF binding sites are relatively hypomethylated, both in the CpG and non-CpG sequence context. Parental allele-specific CTCF binding is rare, with most sites at or near imprinted loci. However, a majority but not all DMRs are bound by CTCF (or cohesin), and the binding is not necessarily allele specific. The *Magel2/Peg12* imprinted locus is unique in the genome, comprising a cluster of eight allele-specific CTCF binding sites. Comprehensive expression profiling in mouse brain of genes near allele-specific CTCF binding sites not previously associated with imprinting did not reveal novel imprinted genes. No allele-specific cohesin binding sites of genome-wide significance were found, although at allele-specific CTCF binding sites, there is a trend for cohesin to bind the same allele.

Deep ChIP-seq for CTCF and cohesin to detect parental allele-specific binding

Sites of CTCF and cohesin binding to DNA were determined genome wide in whole P21 mouse brain by chromatin immunoprecipitation (ChIP) using antibodies specific to CTCF and the RAD21 cohesin subunit followed by high-throughput sequencing (ChIP-seq). The mice were the offspring of crosses between C57BL/6 (Bl6) females and *Mus musculus castaneus* (cast) males (B × C), and vice versa (C × B) (Fig. 1A). We generated 235 million and 231 million high-quality and uniquely mapping sequence reads for CTCF and cohesin, respectively (Fig. 1A). The percentage of reads representing clonal duplication was below 6.2% for all samples (Supplemental Fig. S1). Duplicate reads were excluded from further analysis, and regions of CTCF and RAD21 binding were identified using USeq and assigned to either the Bl6 or cast genome based on known SNPs (Fig. 1B; Supplemental Fig. S2; Supplemental Table S1; Keane et al. 2011; Yalcin et al. 2011).

A systematic read mapping bias toward the reference Bl6 genome was observed, consistent with a Bl6 allele read in a polymorphic region being more likely to align for both CTCF and cohesin. However, our use of reciprocal crosses prevented parental allele-specific binding being confounded: There was no overall bias toward either of the parental alleles when the reads generated from both reciprocal crosses were considered together (Fig. 1C,D).
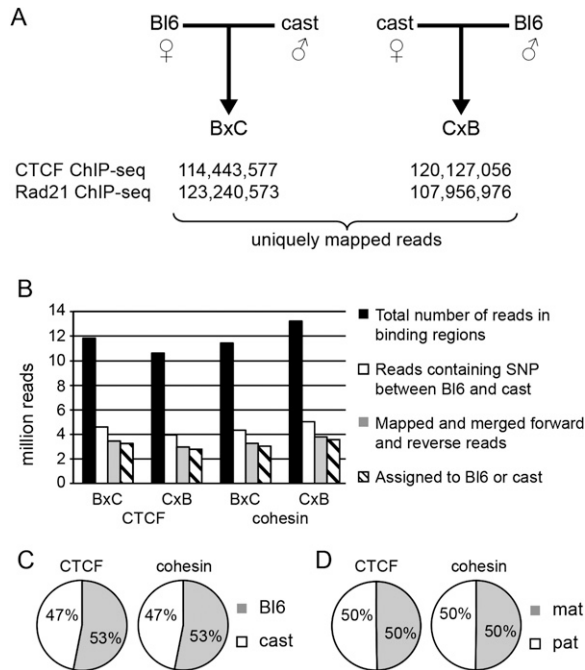
Figure 1. (*A*) ChIP-seq was performed for CTCF and cohesin (RAD21) on P21 brain in B × C and C × B F$_1$ hybrid animals. The experimental design and number of uniquely mapped reads taken forward for further analysis are shown. (*B*) Regions of CTCF and RAD21 binding were identified using the Useq, and regions identified with a FDR of <13 were considered significant and were tested for parent-of-origin-specific binding. (Black bar) The number of reads for each experiment that fell at, or within ±500 bp of a binding region; (white bar) indicates reads in binding regions that aligned over a SNP between C57BL/6 (Bl6) and *Mus m. castaneus* (cast); (gray bar) the number of reads after the paired reads are considered together and the best-quality read is used to map the read to Bl6 or cast. (Hatched bar) The final number of reads assigned. There was a consistent bias toward the reference sequence (*C*); however, this effect was eliminated after we combined B × C and C × B reads (*D*).

CTCF and cohesin binding in mouse brain

Genome wide, we detected 49,358 CTCF and 52,938 cohesin binding sites with a high degree of statistical confidence. Of these, 27,241 sites were bound by both CTCF and cohesin, accounting for 55.3% of the CTCF and 51.5% of the cohesin binding sites, respectively (Fig. 2). This is consistent with previous studies that show both independent and coordinated roles for these factors (Wendt et al. 2008; Lin et al. 2011).

CTCF binds to regions containing the canonical consensus motif

CTCF binds to a specific DNA sequence motif in ES cells (Chen et al. 2008) and liver (Schmidt et al. 2012). To search for CTCF binding motifs in brain, we applied the MEME de novo motif-finding tool to the sequences of all CTCF binding regions in P21 mouse brain (Fig. 3A). The most significant motif ($P = 2.9 \times 10^{-199}$) is highly similar to the published CTCF binding motif (Chen et al. 2008; Schmidt et al. 2010, 2012). To ensure the consistency of the comparison, we repeated the MEME analysis using identical parameters on CTCF binding regions previously identified using ChIP-seq in ES cells and liver (Chen et al. 2008; Schmidt et al. 2012). Again, the canonical motif was identified as the most significant motif in both ES cells ($P = 7.4 \times 10^{-924}$) and liver ($P = 1.4 \times 10^{-367}$).

All three motifs display a high degree of sequence homology, particularly at the core 12-bp sequence at the center of the identified motifs (Fig. 3A).

CTCF binding sites are hypomethylated

The preference of CTCF to bind unmethylated DNA was confirmed by assessing the level of cytosine methylation at CTCF binding sites in brain. Using genome-wide bisulfite-sequencing (BS-seq) data for adult mouse brain (Xie et al. 2012), we compared the overall genome-wide level of methylation at cytosine residues, separately for CpG dinucleotides and non-CpG cytosines, with the portion of the genome corresponding to regions of CTCF binding. We found that methylation at CpG dinucleotides appears to have a greater influence on CTCF binding than non-CpG methylation. Genome wide, 60.8% of CpGs are methylated in the mouse brain, in contrast to 51.9% of CpGs in regions of CTCF binding. Non-CpG methylation also is less frequent in regions of CTCF binding (2.1%) compared with the genome-wide level (2.5%) (Fig. 3B). These differences are statistically significant (χ^2 test, $P < 1 \times 10^{-6}$).

CTCF preferentially binds near genes

We explored the genome-wide location of both CTCF binding regions and parent-of-origin-specific CTCF binding regions using the *cis*-regulatory element annotation (CEAS) tool (Shin et al. 2009). CTCF binding is particularly enriched in regions up to ±3 kb upstream of and downstream from genes, but is depleted in intergenic regions (Fig. 3C). This is consistent with the insulator function of CTCF and, more generally, its involvement in controlling gene expression. When we limited our analysis to parent-of-origin-specific CTCF binding sites, we found the results to be similar. However, intronic regions appeared to be slightly underrepresented and intergenic regions slightly overrepresented relative to the distribution of all CTCF binding sites (Fig. 3C). Given the small number of parent-of-origin-specific CTCF binding sites, these differences are likely due to chance.

Noncanonical CTCF binding sites are tissue specific

We compared the locations of CTCF binding sites in P21 brain with those reported for mouse ES cells and liver (Chen et al. 2008;
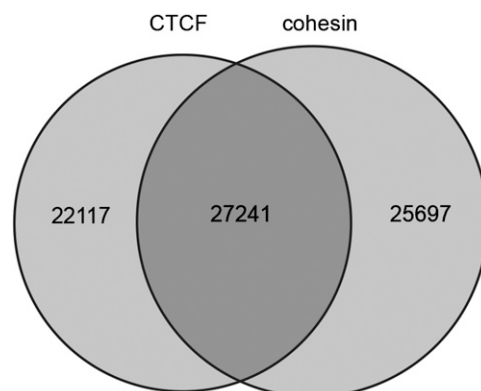
Figure 2. Overlap of the 49,358 CTCF and 52,938 cohesin binding regions in mouse brain. This demonstrates that just over half of CTCF (55%) and cohesin (51%) binding sites are shared, suggesting both independent and combinatorial functions for CTCF and cohesin in the 3-wk mouse brain.
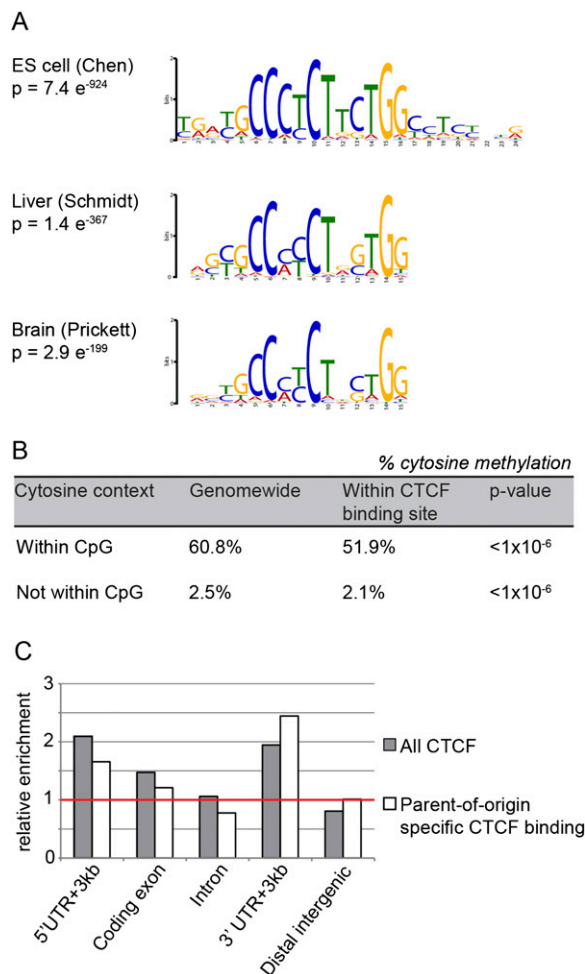
Figure 3. CTCF binding analysis. (A) MEME motif finder was executed on the CTCF binding locations identified by ChIP-seq in brain and compared with motifs identified using previously published ES cells and liver binding locations. Each data set found the canonical motif with high degrees of certainty. (B) The level of cytosine methylation within CTCF binding sites in the brain was compared with that across the whole genome using data from Xie et al. (2012). In both CpG and non-CpG context cytosine methylation, cytosines within CTCF binding sites are methylated less than those outside CTCF binding sites (χ^2 contingency table tests, $P < 0.001$ for CpG and non-CpG context), confirming that CTCF prefers to bind to unmethylated DNA. (C) Genomic locations of CTCF binding are normalized to the proportion of the genome that constitutes each location (represented by the red line). This was considered for all CTCF peaks called with an FDR < 13 and separately for the 116 regions where CTCF binding was seen on one parental allele only (regions identified with a $P < 0.001$). CTCF is significantly enriched at genic regions, but depleted in distal intergenic regions. Parent-of-origin-specific CTCF binding locations are similar but show that binding is depleted in introns but not in intergenic regions.

Schmidt et al. 2012). The incidence of overlap between sites reported in different studies increases with increasing the peak size used for the comparison. Beyond a certain peak size, increases in overlap are mostly due to chance. Therefore, we iteratively increased peak size and compared the incidence of overlap between sites in ES cells, liver, and brain with randomized site locations. Beyond a peak size of 1 kb, increases in the incidence of overlaps were likely due to chance (Supplemental Fig. S3). For a common peak size of 1 kb, 32.0% of all binding sites were shared between ES

cells, brain, and liver, suggesting that they are invariant during differentiation and regardless of cell type (Fig. 4A). Only 1893 binding sites were occupied exclusively in ES cells (5.1% of ES-cell binding sites), suggesting that most CTCF binding sites in ES cells represent a ground state that is added to during differentiation, with few binding sites being characteristic of pluripotency per se. In differentiated tissues, 29.1% of brain and 31.2% of liver binding sites are unique to the respective tissue. These analyses were repeated using alternative CTCF ChIP-seq data from ES cells and liver (Shen et al. 2012), producing similar results, even though significantly fewer binding sites were identified in these studies because of limited read depth and quality (Supplemental Fig. S4). We hypothesize that the canonical consensus CTCF binding motif may be at the core of binding sites that are largely invariant with respect to cell type, concordant with other findings (Essien et al. 2009). We restricted the above overlap analysis to CTCF binding sites that lack the canonical binding motif. There was a large reduction in the number of binding sites shared between tissue types (Fig. 4A), with most binding sites now being tissue specific: 84.2% of binding sites in brain that lack the canonical motif were brain specific, and similarly for ES cells (81.2%) and liver (82.9%) (Fig. 4B). These results suggest that CTCF binding to tissue-specific sites may involve other consensus motifs recognized by cofactors or tissue-specific conformations of the 11 zinc finger domains of CTCF itself.

Parent-of-origin-specific CTCF and/or cohesin binding is limited to specific DMRs

We systematically investigated the binding of CTCF and cohesin at or near 22 known well-characterized mouse gDMRs (Table 1) associated with imprinted gene expression (most of which are classified as ICRs). Of the 22 gDMRs (Table 1), 19 have a CTCF and/or cohesin binding site in brain within 2.5 kb. Of these sites, 12 are bound by both CTCF and cohesin, three by CTCF alone and four by cohesin alone. gDMRs with both CTCF and cohesin binding
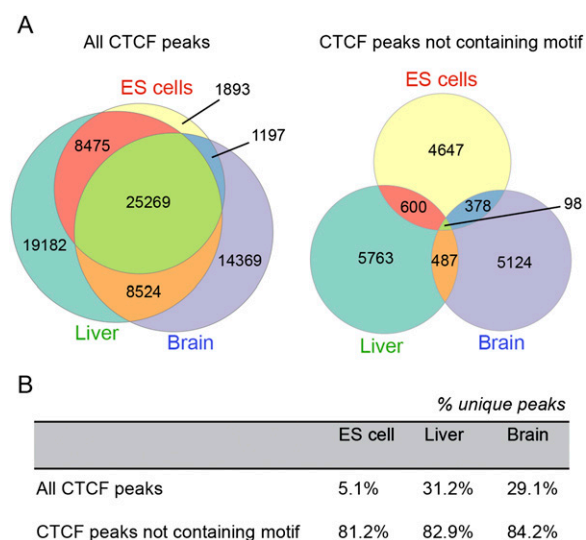
Figure 4. (A) Proportional Venn diagrams comparing coincidence of CTCF binding sites between ES cells, liver, and brain demonstrate significant overlap of CTCF binding in these tissues. Coincident binding was also considered after the removal of binding regions containing the consensus CTCF motif; overlap of CTCF binding in the absence of the consensus motif was much lower than when all binding sites were considered. (B) The percentages of shared peaks for each tissue type for all peaks and for nonmotif peaks.

Table 1. Gene names for imprinted regions are listed with corresponding gDMR positions

| gDMR information (Wamidex) | | | CTCF and cohesin binding by ChIP-seq | | | | | | Previous evidence |
| | | | CTCF | | | Cohesin | | | |
gDMR name	gDMR position (Wamidex)	Methylated allele	CTCF binding?	Binding allele	Allele specific P-value	Cohesin binding?	Binding allele	Allele binding P-value	Reference
CTCF and cohesin precisely colocalized at gDMR									
Grb10	chr11: 11,925,485–11,925,790	Maternal	Yes	N/A	N/A	Yes	N/A	N/A	Hikichi et al. 2003
H19/Igf2	chr7: 149,766,168–149,768,424	Paternal	Yes	Maternal	1.11×10^{-74}	Yes	Maternal	1.08×10^{-4}	For example, Hark et al. 2000
Inpp5f_v2	chr7: 135,831,788–135,832,156	Maternal	Yes	Bi-allelic	N/A	Yes	Bi-allelic	N/A	
Mcts2	chr2: 152,512,491–152,513,011	Maternal	Yes	N/A	N/A	Yes	N/A	N/A	
Mest	chr6: 30,686,709–30,687,273	Maternal	Yes	Paternal	9.00×10^{-15}	Yes	Paternal	0.0414	
Nnat	chr2: 157,385,786–157,387,398	Maternal	Yes	N/A	N/A	Yes	N/A	N/A	
Peg13	chr15: 72,636,765–72,642,079	Maternal	Yes	Paternal	1.35×10^{-57}	Yes	Paternal	5.50×10^{-3}	Singh et al. 2011
Plagl1	chr10: 12,810,276–12,810,604	Maternal	Yes	Bi-allelic	N/A	Yes	Bi-allelic	N/A	
CTCF and cohesin not precisely colocalized at gDMR									
Cdh15	chr8: 125,387,861–125,390,344	Maternal	Yes	Paternal	0.0463	Yes	N/A	N/A	
Nespas	chr2: 174,121,208–174,126,482	Maternal	Yes	N/A	N/A	Yes	N/A	N/A	
Zrsr1	chr11: 22,871,842–22,872,319	Maternal	Yes	N/A	N/A	Yes	Bi-allelic	N/A	
Zim2 (Peg3)	chr7: 6,680,287–6,684,827	Maternal	Yes	Paternal	1.16×10^{-30}	Yes	Paternal	0.049	
CTCF binding only									
Peg10	chr6: 4,697,209–4,697,507	Maternal	Yes	N/A	N/A	No	N/A	N/A	
Meg3/Dlk1	chr12: 110,761,563–110,768,989	Paternal	Yes	N/A	N/A	No	N/A	N/A	
Impact	chr18: 13,130,706–13,132,250	Maternal	Yes	N/A	N/A	No	N/A	N/A	
Cohesin binding only									
Igf2r-air	chr17: 12,934,163–12,935,573	Maternal	No	N/A	N/A	Yes	N/A	N/A	
Gnas-exon1A	chr2: 174,153,279–174,153,502	Maternal	No	N/A	N/A	Yes	N/A	N/A	
Kcnq1ot1	chr7: 150,481,060–150,481,397	Maternal	No	N/A	N/A	Yes	N/A	N/A	Fitzpatrick et al. 2007
Snurf/Snrpn	chr7: 67,149,878–67,150,301	Maternal	No	N/A	N/A	Yes	N/A	N/A	
No binding									
Nap1l5	chr6: 58,856,690–58,857,056	Maternal	No	N/A	N/A	No	N/A	N/A	
Rasgrf1	chr9: 89,774,406–89,774,691	Paternal	No	N/A	N/A	No	N/A	N/A	Yoon et al. 2005
Slc38a4	chr15: 96,885,270–96,886,284	Maternal	No	N/A	N/A	No	N/A	N/A	

The methylated allele is shown and CTCF binding is indicated according to the criteria used in this study. The parental allele specificity is shown, but for some regions, these data did not reach significance, or "no data" was registered, which was due either to the absence of an SNP or insufficient sequence reads over the SNP. *P*-values are given for the allele-specific binding, where $P < 0.05$. (Gray shading) Genome-wide significant parent-of-origin-specific binding. References to previous binding are indicated where known.

sites within 2 kb formed two categories: those where CTCF and cohesin colocalized precisely at the gDMR (eight regions), and a further four regions where binding occurred near but not over the gDMR and CTCF and cohesin each bound distinct sites (Table 1). Where the two factors are precisely colocalized on the DNA, cohesin binding is probably linked mechanistically to CTCF, while this is less likely at gDMRs where binding sites do not coincide.

For gDMRs where genome-wide significant ($P < 1 \times 10^{-6}$) parent-of-origin-specific binding events for CTCF (Table 1) were detected (*H19/Igf2*, *Mest*, *Peg13*, and *Zim2* [*Peg3*]), binding occurred as expected on the unmethylated allele (Table 1). The *Mest* and *Zim2* gDMRs were not previously known to bind CTCF in a parent-of-origin-specific manner. In addition, the 95% confidence intervals (Supplemental Fig. S5; Supplemental Table S2) for parent-of-origin-specific binding showed a trend toward preferential binding of CTCF to the unmethylated alleles of the *Grb10*, *Mcts2*, *Cdh15*, *Nespos*, *Zrsr1*, *Peg10*, and *Meg3/Dlk1* gDMRs. We considered CTCF binding to be completely biallelic if the 95% confidence interval for the maternal-over-paternal read ratio was between 0.35 and 0.65 and spanned 0.5. This was the case for the *Inpp5f_v2* and *Plagl1* gDMRs. At *H19/Igf2* and *Peg13*, parent-of-origin-specific binding of cohesin was detected but was not genome-wide significant after Bonferroni multiple testing correction (Table 1). The overall pattern of the 95% confidence intervals for the ratio of maternal-to-paternal reads for CTCF and cohesin suggests that in comparison to CTCF, cohesin binding is less biased toward the unmethylated parental allele (Supplemental Fig. S5; Supplemental Table S2). This is consistent with increased recruitment of cohesin to sites bound by CTCF.

Parent-of-origin-specific CTCF and cohesin binding at gDMRs could only be tested where a Bl6-cast SNP is within the bound region (Table 1). In addition, CTCF and cohesin peaks did not always overlap perfectly so that for some gDMRs, a SNP was informative for one factor but not the other. Another limitation for the detection of parent-of-origin-specific binding arose when a SNP was located at the periphery of the respectively bound region where fewer reads align and the statistical power of the binomial test was diminished. For CTCF, these limitations applied in particular to the *Grb10*, *Mcts2*, *Nnat*, *Nespos*, *Zrsr1*, *Impact*, and *Peg10* gDMRs. For cohesin, the above limitations applied to half of the cohesin-bound gDMRs (Supplemental Table S2). The results for CTCF support the notion that it plays a central role in imprinting control at several loci. This is in contrast to cohesin, in particular, four gDMRs (*Gnas-exon1A*, *Igf2r-air*, *Kcnq1ot1*, and *Snurf/Snrpn*) were bound by cohesin but not by CTCF; here binding was not parental allele specific. Cohesin binding independently of CTCF is not unprecedented (Schmidt et al. 2012), and there is evidence that it is more generally involved in transcriptional activation (Kagey et al. 2010).

Genome-wide, parental allele-specific binding of CTCF and cohesin is rare and mostly restricted to imprinted loci

CTCF binding efficiency to methylated DNA is reduced compared with unmethylated DNA (Mukhopadhyay et al. 2004) explaining parental allele-specific binding at the *H19/Igf2* and other gDMRs. If CTCF and cohesin are exerting a key regulatory role at several imprinted loci, then genome wide, other occurrences of parental allele-specific CTCF and/or cohesin binding may identify novel DMRs and imprinted genes. Four known ICRs—*H19/Igf2*, *Peg13*, *Zim2* (*Peg3*), and *Mest*—met the genome-wide significance threshold for parental allele-specific CTCF binding providing proof of principle.

Only an additional 17 regions reached genome-wide significance (Fig. 5; Supplemental Fig. S3; Table 2). Eight of these sites clustered in a 250-kb region on chromosome 7 at the *Peg12/Magel2* imprinted domain (Fig. 6). A further four sites were within 6 Mb of other known imprinted regions. Two more are 30 kb apart on chromosome 14 (Fig. 5). Many chromosomes were devoid of parental allele-specific CTCF binding, and no cohesin binding regions were detected at genome-wide significance (Supplemental Table S3). Of all 21 genome-wide significant parental allele-specific CTCF binding sites, six were on the maternal and 15 on the paternal allele. We tested all gene transcripts at or near these sites not previously reported as imprinted for parental allele-specific expression in mouse brain. Many showed a complex organization of transcripts (Supplemental Fig. S6), but none were imprinted (Supplemental Table S4).

Eight sites of parental allele-specific CTCF binding at the *Peg12/Magel2* imprinted domain (Fig. 6) bound CTCF on the paternal allele, indicating maternal methylation. We assayed methylation of the CpG island at the promoter of *Magel2*, which is in close proximity to two CTCF sites and maternally methylated (Supplemental Fig. S7). This is confirmation of the parental allele-specific methylation of the region recently reported (Xie et al. 2012). The *Magel2* DMR is likely somatic and established post-fertilization that is supported by genome-wide methylation data in oocytes (Smallwood et al. 2011). In addition, *Dnmt3L*$^{-/+}$ 8.5 days postcoitum (dpc) embryos are unchanged at the *Magel2* promoter relative to wild type and are unmethylated (Proudhon et al. 2012). This suggests that the maternal allele-specific methylation at the *Magel2* promoter, and presumably the other sites of paternal allele-specific CTCF binding in the domain, is established post-implantation and/or is brain specific. The regulation of the imprinted domain comprising *Ndn*, *Magel2*, *Mkrn3*, and *Peg12* deserves further study since the human orthologs of *Ndn*, *Magel2*, and *Mkrn3* are in the region associated with Prader-Willi syndrome (Lee and Wevrick 2000), with patients displaying a notable range of neurological symptoms. Given this extensive investigation of parental allele-specific CTCF binding, we predict that there are few additional DMRs in the adult mouse brain bound by CTCF.

Validation of CTCF and cohesin binding at specific loci

Quantitative assays for *H19/Igf2*, *Peg10*, *Nap1l5*, *Nnat*, and *Grb10* DMR validated that the ChIP-seq data (Supplemental Fig. S8A,B)
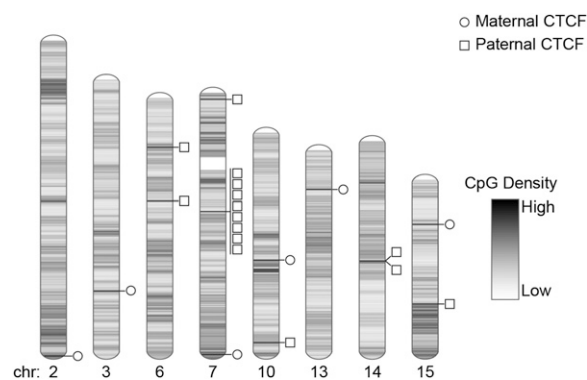
Figure 5. Chromosomal location of genome-wide significant parent-of-origin-specific CTCF binding regions. Where CTCF is bound on the maternally inherited allele, this is illustrated with a circle; where CTCF is bound on the paternally inherited allele, this is illustrated with a square. Only chromosomes where parent-of-origin-specific binding was seen are shown. CpG density is indicated.

Table 2. 21 regions were identified where CTCF binds to one allele in a parent-of-origin-specific manner

Location (mm9 reference)	Binding allele	P-value	Ratio of binding:nonbinding	Nearest genes	Notes
chr7: 149,764,416–149,768,874	Maternal	1.11×10^{-74}	10.3	H19, Igf2	Known imprinted gene region
chr15: 72,638,890–72,641,957	Paternal	1.35×10^{-57}	12.6	Peg13, Trappc9	Known imprinted gene region
chr7: 6,678,325–6,681,689	Paternal	1.17×10^{-30}	5.5	Zim2 (Peg3)	Known imprinted gene region
chr6: 30,686,300–30,688,046	Paternal	9.10×10^{-15}	9.4	Mest	Known imprinted gene region
chr7: 69,543,049–39,547,037	Paternal	1.35×10^{-12}	3.4	Magel2	Known imprinted gene region
chr7: 69,580,613–69,582,990	Paternal	2.05×10^{-10}	5.6	Magel2	Known imprinted gene region
chr7: 69,323,407–69,325,218	Paternal	3.91×10^{-10}	6.6	Magel2	Known imprinted gene region
chr10: 74,395,653–74,404,537	Maternal	1.28×10^{-9}	1.5	Rtdr1, Gnaz	No known imprinted genes within 20 Mb
chr7: 69,526,343–69,528,366	Paternal	3.10×10^{-9}	4.1	Magel2	Known imprinted gene region
chr7: 69,372,124–69,373,922	Paternal	3.26×10^{-9}	8.0	Ndn, Magel2	Known imprinted gene region
chr2: 180,079,574–180,091,367	Maternal	5.84×10^{-9}	1.5	Gata5, Gm14318	6 Mb from Gnas locus
chr7: 69,608,918–69,610,897	Paternal	6.90×10^{-9}	5.6	Peg12, Mkrn3	Known imprinted gene regions
chr14: 69,941,084–69,946,555	Paternal	1.68×10^{-8}	1.5	Gm16677, Entpd4, Loxl2	5 Mb from Htr2a imprinted locus
chr7: 69,353,580–69,355,185	Paternal	2.15×10^{-8}	5.1	Ndn, Magel2	Known imprinted gene regions
chr7: 69,519,941–69,521,489	Paternal	3.38×10^{-8}	5.3	Magel2	Known imprinted gene regions
chr14: 69,994,239–70,003,685	Paternal	3.98×10^{-8}	1.5	Entpd4, AK086749, Loxl2	5 Mb from Htr2a imprinted locus
chr10: 120,737,183–120,739,873	Paternal	4.75×10^{-8}	2.7	Tbc1d30	No known imprinted genes within 20 Mb
chr3: 121,236,161–121,244,419	Maternal	6.85×10^{-8}	1.7	A530020G20Rik, Slc44a3	No known imprinted genes on chromosome 3
chr15: 27,817,069–27,819,622	Maternal	6.95×10^{-8}	2.3	Trio	No known imprinted genes within 20 Mb
chr13: 25,098,042–25,100,314	Maternal	9.98×10^{-8}	2.1	Mrs2, Gpld1	No known imprinted genes within 20 Mb
chr6: 60,631,333–60,634,328	Paternal	2.47×10^{-7}	2.2	Snca	1.6 Mb from Herc3

After correction for multiple testing, regions are ranked in order of statistical significance (P-value). Twelve regions are associated with known imprinted genes, of which eight are associated with the Peg12/Magel2 imprinted locus. Four further regions occur within close proximity of an imprinted locus. All novel candidates were tested for imprinting (Supplemental Table S4).

results were in agreement (Table 1) with the exception of CTCF binding at *Nnat* and cohesin binding at *Peg10*. Both are borderline cases. Using qPCR to detect CTCF binding at the *Nnat* DMR resulted in $P = 0.08$, just above our cutoff for binding, At *Peg10*, RAD21 binding was detected by qPCR, but no peak was identified by ChIP-seq. When the stringency of the ChIP-seq peak detection is relaxed, two RAD21 binding regions ~1 kb either side of the qPCR regions are detected (Supplemental Fig. S8C).

Validation of parental allele-specific binding

To validate allele-specific binding we pyrosequenced ChIP'd mouse brain from reciprocal crosses. We selected three representative DMRs, based on the ChIP-seq results: *Inpp5f_v2*, where biallelic CTCF binding was detected; *Mest*, where we detected paternal allele-specific binding; and *Peg10*, where the CTCF binding site did not meet the significance threshold for allele-specific

binding but where the 95% confidence interval was suggestive of CTCF binding on the paternal allele. These results agreed with our ChIP-seq data (Supplemental Fig. S9): *Inpp5f_v2* does not deviate from the expected 50:50 allelic ratio ($P = 0.3214$), *Mest* shows paternal binding ($P = 0.0017$), and *Peg10* shows a bias toward enrichment of the paternal allele ($P = 0.0813$).

CTCF and cohesin binding at somatic DMRs

A set of 23 known somatic and novel putative somatic DMR coordinates has recently been defined by whole-genome bisulfite sequencing (BS-seq) in mouse brain (Xie et al. 2012). We evaluated CTCF and cohesin binding in the somatic DMRs identified in this study (Supplemental Table S5). We found 13 instances of CTCF binding, two of which were parental allele specific ($P < 1 \times 10^{-6}$) and 14 instances of cohesin binding, none of which were parental allele specific. All parental allele-specific binding involved the
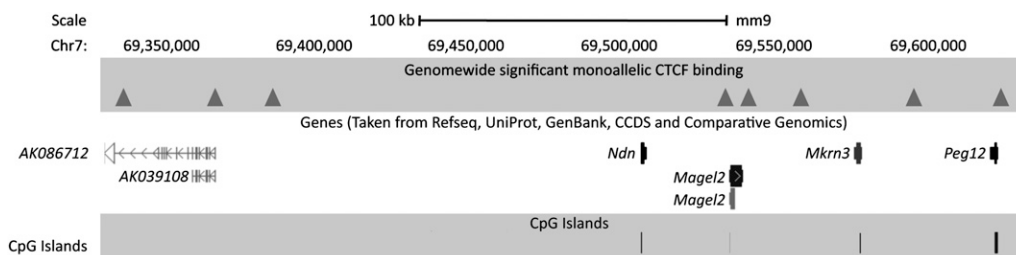
Figure 6. Multiple parent-of-origin-specific CTCF binding sites are observed on the paternal allele at the *Magel2/Peg12* locus. (Triangles) Paternally bound CTCF binding sites. Genes and CpG islands are indicated. This region represents a unique example in the mouse genome of CTCF bound only on the paternal allele at eight regions in close proximity. This figure was adapted from the UCSC Genome Browser.

unmethylated allele. Overall, the results for somatic DMRs (Supplemental Table S6) are in close agreement with the results for gDMRs (Table 1) so that the origin of a DMR, germline versus somatic, is not a determinant of CTCF and/or cohesin involvement in the regulation of imprinting.

Discussion

CTCF and cohesin act synergistically and independently in mouse brain and show tissue-specific distributions compared with undifferentiated cells

Several studies have examined the colocalization of CTCF with cohesin (Chen et al. 2008; Parelho et al. 2008; Rubio et al. 2008; Kagey et al. 2010; Schmidt et al. 2012), and CTCF physically associates with cohesin via the Stag1 (Scc3/SA1) subunit in human cells (Wendt et al. 2008). Here 55% of CTCF binding sites overlap with cohesin binding, and the remaining sites binding independently (Fig. 2), indicating that CTCF fulfils a role independent from as well as in combination with cohesin in brain. This supports the idea that different functions may be a result of the context of CTCF binding (Gaszner and Felsenfeld 2006), and it is possible that coordinate binding of cohesin may influence CTCF. Cohesin is involved in tissue-specific transcriptional control (Faure et al. 2012) and associated with the Mediator complex, which has a role in transcriptional activation (Taatjes 2012). Studies have shown a link between cohesin, the Mediator complex, transcription, and chromatin looping (Kagey et al. 2010). We report that 51% of cohesin sites in brain are not coincident with CTCF, consistent with CTCF not being required for the loading of cohesin onto DNA (Rubio et al. 2008).

This comparison of CTCF binding in ES cells, liver, and brain reveals more unique CTCF binding sites in differentiated cells than in ES cells, suggesting tissue-specific CTCF binding in the specification and/or maintenance of differentiated tissue. We observe a significant overlap in CTCF binding between tissues (Fig. 4A), consistent with studies reporting highly conserved CTCF binding between cell types (Kim et al. 2007b).

CTCF binding

In brain, CTCF binds to unmethylated regions of DNA, usually to the canonical CTCF motif (Chen et al. 2008; Schmidt et al. 2010, 2012). CTCF binding outside this motif is much more tissue specific, and there is little overlap between tissues (Fig. 4A). This is consistent with evidence from CTCF knockout mouse studies, which exhibit embryonic lethality prior to implantation (Splinter et al. 2006). We found that the canonical consensus binding motif is most frequent at CTCF binding sites shared between ES cells, brain, and liver; thus, it is associated with invariant binding during differentiation. CTCF binding appears overrepresented just up or downstream from gene bodies with a paucity in distal intergenic regions, unsurprising given the known role of CTCF in gene expression (Bell et al. 1999; Cuddapah et al. 2009). CTCF mediates long-range chromosomal interactions genome wide in *cis* and in *trans* in ES cells (Handoko et al. 2011), and we provide additional evidence genome wide that CTCF is an insulator at or near gene coding regions by binding to noncoding DNA.

Cytosine methylation at CTCF binding regions

Cytosine methylation in both a CpG and non-CpG context is reduced in regions of CTCF binding compared with the level observed

genome wide, consistent with published data (Mukhopadhyay et al. 2004). Interestingly the canonical motif lacks CpG dinucleotides, suggesting that methylation of DNA in the motif does not preclude CTCF binding, but surrounding methylation is important. The canonical motif may not function alone, but in concert with another region of DNA ~20 bp downstream, suggesting that CTCF interaction with DNA is not limited to the 20-mer motifs (Schmidt et al. 2012).

Parent-of-origin-specific CTCF and cohesin binding

CTCF and cohesin bind at numerous imprinting control regions and other DMRs as previously detected, but not systematically tested. The presence of CTCF and cohesin together at 12 imprinting associated gDMRs in brain (Table 1) is consistent with a regulatory role for these proteins at imprinted loci. Studies using 3C and 4C have shown that several imprinted domains are physically clustered (Sandhu et al. 2009), in part because CTCF (Botta et al. 2010) and cohesin (Murrell et al. 2004; Nativio et al. 2009) form loops that contribute to three-dimensional (3D) nuclear architecture (Phillips and Corces 2009). The CTCF and cohesin binding (Table 1) supports the idea of three types of imprinting mechanisms: CTCF dependent, CTCF/cohesin mediated, and CTCF/cohesin independent.

CTCF and cohesin bind at somatic DMRs, suggesting a role for them here. Parental allele-specific binding of CTCF together with cohesin regulates allele-specific expression in somatic cells (Lin et al. 2011), while cohesin binding alone may be involved in the transcriptional regulation of imprinted gene expression generally. CTCF and cohesin are likely to have distinct functions in different cell types at a subset of targets (Lin et al. 2011), and findings in mouse brain support the idea that these proteins play a role in imprinting at some loci and at others they act more generally.

These data provide a resource for interrogating the roles of CTCF and cohesin and point to a role at more imprinted loci than was previously appreciated, although further functional studies would be needed to confirm this. Three gDMRs do not bind either factor, illustrating the heterogeneous nature of gDMRs as a group of regulatory regions. For example, the four imprinted retrogene/host gene pairs *Mcts2/H13*, *Nap1l5/Herc3*, *Inpp5f/Inpp5f_v2*, and *Zrsr1/Commd1* share several sequence-based and genomic context–related features (Wood et al. 2007). Since within this group, *Mcts2* binds both CTCF and cohesin together, *Nap1l5* binds neither CTCF nor cohesin, *Inpp5f_v2* binds CTCF on both parental alleles equally, and *Zrsr1* binds both CTCF and cohesin, but they do not colocalize, this suggests no consistent mechanism for imprinting control despite the other shared features.

CTCF binding profiles vary between different tissues. We show that many CTCF binding sites are shared between ES cells and differentiated tissues and that this type of invariant CTCF binding is associated with the canonical CTCF motif. CTCF binding in the absence of the canonical motif is associated with tissue-specific CTCF binding.

Methods

Chromatin immunoprecipitation

Chromatin from whole tissue was isolated, sonicated, and immunoprecipitated for ChIP-seq library preparation according to Supplemental Methods 1.

Next-generation sequencing

Library preparation

DNA enriched through ChIP was quantified using the Qubit (Invitrogen) and Quant-iT dsDNA high-sensitivity assay kit (Invitrogen:Q32854) and was sized using the Agilent Bioanalyzer with a High Sensitivity DNA Bioanalyzer kit (5067-4626). DNA was fragmented to a size appropriate for the library preparation step using the Covaris S220, samples were sheared over two cycles: 5% duty cycle, 3 intensity, 200 cycles per burst, and time of 65 sec. DNA from ChIPs performed on chromatin extracted from two mice was pooled.

ChIP-seq libraries were prepared using the Illumina ChIP-seq library preparation kit (IP-102-1001) and the NEBNext ChIP-seq library preparation kit (E6240). Libraries were sized and quantified using an Agilent Bioanalyzer and a High Sensitivity Kit (5067-4626).

ChIP–seq data analysis

Sequence reads were aligned to the mouse reference genome (mm9) using Novoalign (v. 2.01.13; http://www.novocraft.com/). USeq (Nix et al. 2008) was used to identify mean peak shift separately for CTCF and cohesin reads using only the first of each pair-end matched read. Peaks were identified using peak shifts and window sizes of 138 bp and 144 bp for CTCF and cohesin, respectively, and a False Discovery Rate (FDR) of 95% (Supplemental Table S2A). A subset of peaks was obtained to a false discovery ratio of 5% (Phred-scaled FDR 13) expanded by 500 bp upstream and downstream and overlapping peaks merged prior to further analysis. Refer to Supplemental Table S2A for the number of raw reads that pass a quality control map in a CTCF or cohesin binding region.

Parental allele–specific binding analysis

Parental allele-specific binding was assessed by binomial testing, using a custom bioinformatics pipeline. For performance reasons, only reads of interest, which overlapped the previously identified CTCF or RAD21 binding sites and a SNP between the two parental strains, were extracted from the SAM files and used for subsequent analysis.

Individual reads were assigned to one of the parental alleles using a custom Perl script, using the SAMtools Perl library. Each read was mapped as either derived from the reference sequence (Bl6) or from the cast allele on the basis of a SNP between the parental strains. If more than one SNP was present, the SNP with the best quality of read sequence was used. Reads were only considered for subsequent analysis if the Phred-scaled alignment mapping quality exceeded 50 and the base call quality at the SNP used for mapping of the read exceeded 20.

Paired reads were mapped to parental strains separately and merged. Because paired reads are not independent data points, when they were in disagreement (<1%) the read pair was assigned on the basis of the best SNP in either of the two reads.

Assigned reads were converted to maternally or paternally derived, and data from both B × C and C × B reciprocal crosses were merged for the CTCF and RAD21 data sets independently. Counts of maternal and paternal reads were obtained on a per-region basis using MySQL. Binding regions were only tested for parent-of-origin-specific expression if three or more reads could be mapped.

Parental allele-specific binding was assessed using a two-sided binomial test (implemented in R) of the maternal-versus-paternal allelic read counts. Regions were sorted by *P*-value score using MySQL. The genome-wide significance of *P*-values was assessed by means of Bonferroni correction. UCSC BED tracks were prepared at different cutoffs with maternal/paternal annotation.

Peak intersections

All subsequent bioinformatic analyses were performed on expanded regions unless otherwise specified. CTCF peak intersections between ES cell, brain, and liver data were performed using an optimized peak size of 1 kb for all data sets. ES cell data were converted to mm9 using the UCSC liftOver tool. Peak overlap counts were obtained used the BEDTools intersectBed command. For each intersection, counts of the intersecting peaks were calculated in both possible ways, and the peaks count reported was the mean of the two measurements. For intersections of more than two data sets, only one of the possible configurations of intersections was examined; the same configuration was used for all analyses.

Identification of non–motif–containing peaks

Peaks that did not contain the CTCF motif were identified using the FIMO tool from the MEME suite (Grant et al. 2011). Peak sequences were obtained using the UCSC Genome Browser table tool in FASTA format with repeat sequences masked. The CTCF motif identified from the brain data set was used throughout, and the threshold for detection was set to 10^{-3}. Custom UNIX shell scripts were used to extract the coordinates of the peaks from the FIMO output.

Motif finding

Motif finding was performed using MEME (Bailey et al. 2009) on binding regions using default MEME parameters. For the brain data, the best subwindow coordinates were used.

Genomic distribution of peaks

The CEAS tool (Shin et al. 2009) was used to assess the genomic distribution of unexpanded CTCF binding regions, and unexpanded parent-of-origin-specific CTCF binding sites were detected with a $P < 0.001$. Relative abundance was normalized to the proportion of the genome represented by each genomic region. For the CEAS analysis, parent-of-origin-specific CTCF expanded regions were assigned back to their original constituent unexpanded peaks using bedmap (Neph et al. 2012).

Quantitative PCR validation of CTCF and cohesin ChIP–seq

These assays are detailed in Supplemental Methods 1.

Validation of parent–of–origin–specific binding using pyrosequencing

Chromatin was extracted from four biological replicates, two B × C and two C × B P21 brains. Pyrosequencing validated CTCF binding at three regions (Supplemental Table S6). ChIP was performed as for ChIP-seq. Maternal-to-paternal proportions were assigned using SNPs between Bl6 and cast. Allelic proportions were normalized to input DNA, which represents a 50:50 ratio of maternal-to-paternal reads. Using the normalized maternal proportion, a two-sided *t*-test against a 0.5 null proportion was performed.

Testing for imprinted expression

Transcripts were tested for allele-specific expression using PCR followed by Sanger sequencing (using the primers in Supplemental Table S7).

Bisulfite mutagenesis

Genomic DNA from B × C and C × B intercross mouse brain tissue was converted using the Zymo EZ DNA Methylation-Direct Kit (D5020). Amplified regions of interest were ligated into pGEM-T Easy (Promega:A1360), transformed into competent *Escherichia coli*, and sequenced. Primers were designed with MethPrimer (For:GTGTTTGTTGAGAGTTGTTGAGAGA; Rev: ACCAAACAACC ATAAAAACCTACAA) (Li et al. 2002).

Data access

Primary sequencing data have been deposited in the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE35140.

Acknowledgments

Author contributions: A.R.P. carried out the ChIP-seq experimental work, made the majority of the data acquisition, performed some of the analyses, and drafted part of the paper. N.B. performed the majority of the bioinformatic analyses, made the tables and figures, and revised some of the paper. R.B.M. contributed to the conception and initial experimental design and to initial technique development and some bioinformatics and drafted part of the paper. S.H. performed the bisulfite analysis. S.M.M. performed the allele-specific assays. R.S. contributed to the conception and design of the experiments and bioinformatic analyses and wrote part of the paper. R.J.O. contributed to the conception and design of the experiments and wrote part of the paper. All authors contributed to the final version of the paper.

References

Bailey TL, Bodén M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res* **37:** W202–W208.

Bartolomei M. 2009. Genomic imprinting: Employing and avoiding epigenetic processes. *Genes Dev* **23:** 2124–2133.

Bell A, Felsenfeld G. 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405:** 482–485.

Bell A, West A, Felsenfeld G. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98:** 387–396.

Botta M, Haider S, Leung IX, Lio P, Mozziconacci J. 2010. Intra- and inter-chromosomal interactions correlate with CTCF binding genome wide. *Mol Syst Biol* **6:** 426.

Cassidy S, Dykens E, Williams C. 2000. Prader-Willi and Angelman syndromes: Sister imprinted disorders. *Am J Med Genet* **97:** 136–146.

Chen X, Xu H, Yuan P, Fang F, Huss M, Vega V, Wong E, Orlov Y, Zhang W, Jiang J, et al. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133:** 1106–1117.

Cuddapah S, Jothi R, Schones DE, Roh TY, Cui K, Zhao K. 2009. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res* **19:** 24–32.

da Rocha S, Edwards C, Ito M, Ogata T, Ferguson-Smith A. 2008. Genomic imprinting at the mammalian Dlk1–Dio3 domain. *Trends Genet* **24:** 306–316.

Davies W, Isles A, Humby T, Wilkinson L. 2007. What are imprinted genes doing in the brain? *Epigenetics* **2:** 201–206.

DeVeale B, van der Kooy D, Babak T. 2012. Critical evaluation of imprinted gene expression by RNA-seq: A new perspective. *PLoS Genet* **8:** e1002600.

Edwards C, Ferguson-Smith A. 2007. Mechanisms regulating imprinted genes in clusters. *Curr Opin Cell Biol* **19:** 281–289.

The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489:** 57–74.

Essien K, Vigneau S, Apreleva S, Singh LN, Bartolomei MS, Hannenhalli S. 2009. CTCF binding site classes exhibit distinct evolutionary, genomic, epigenomic and transcriptomic features. *Genome Biol* **10:** R131.

Faure AJ, Schmidt D, Watt S, Schwalie PC, Wilson MD, Xu H, Ramsay RG, Odom DT, Flicek P. 2012. Cohesin regulates tissue-specific expression by stabilizing highly occupied *cis*-regulatory modules. *Genome Res* **22:** 2163–2175.

Fedoriw A, Stein P, Svoboda P, Schultz R, Bartolomei M. 2004. Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting. *Science* **303:** 238–240.

Filippova GN, Fagerlie S, Klenova EM, Myers C, Dehner Y, Goodwin G, Neiman PE, Collins SJ, Lobanenkov VV. 1996. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-*myc* oncogenes. *Mol Cell Biol* **16:** 2802–2813.

Fitzpatrick G, Pugacheva E, Shin J-Y, Abdullaev Z, Yang Y, Khatod K, Lobanenkov V, Higgins M. 2007. Allele-specific binding of CTCF to the multipartite imprinting control region KvDMR1. *Mol Cell Biol* **27:** 2636–2647.

Frost J, Monk D, Moschidou D, Guillot PV, Stanier P, Minger SL, Fisk NM, Moore HD, Moore GE. 2011. The effects of culture on genomic imprinting profiles in human embryonic and fetal mesenchymal stem cells. *Epigenetics* **6:** 52–62.

Fu Y, Rusznak Z, Herculano-Houzel S, Watson C, Paxinos G. 2012. Cellular composition characterizing postnatal development and maturation of the mouse brain and spinal cord. *Brain Struct Funct* doi: 10.1007/s00429-012-0462-x.

Garfield AS, Cowley M, Smith FM, Moorwood K, Stewart-Cox JE, Gilroy K, Baker S, Xia J, Dalley JW, Hurst LD, et al. 2011. Distinct physiological and behavioural functions for parental alleles of imprinted Grb10. *Nature* **469:** 534–538.

Gaszner M, Felsenfeld G. 2006. Insulators: Exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet* **7:** 703–713.

Grant CE, Bailey TL, Noble WS. 2011. FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **27:** 1017–1018.

Gregg C, Zhang J, Butler J, Haig D, Dulac C. 2010a. Sex-specific parent-of-origin allelic expression in the mouse brain. *Science* **329:** 682–685.

Gregg C, Zhang J, Weissbourd B, Luo S, Schroth G, Haig D, Dulac C. 2010b. High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science* **329:** 643–648.

Hadjur S, Williams L, Ryan N, Cobb B, Sexton T, Fraser P, Fisher A, Merkenschlager M. 2009. Cohesins form chromosomal *cis*-interactions at the developmentally regulated IFNG locus. *Nature* **460:** 410–413.

Hagiwara Y, Hirai M, Nishiyama K, Kanazawa I, Ueda T, Sakaki Y, Ito T. 1997. Screening for imprinted genes by allelic message display: Identification of a paternally expressed gene impact on mouse chromosome 18. *Proc Natl Acad Sci* **94:** 9249–9254.

Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, Lee CW, Ye C, Ping JL, Mulawadi F, et al. 2011. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat Genet* **43:** 630–638.

Hark A, Schoenherr C, Katz D, Ingram R, Levorse J, Tilghman S. 2000. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* **405:** 486–489.

Hikichi T, Kohda T, Kaneko-Ishino T, Ishino F. 2003. Imprinting regulation of the murine Meg1/Grb10 and human GRB10 genes; roles of brain-specific promoters and mouse-specific CTCF-binding sites. *Nucleic Acids Res* **31:** 1398–1406.

Kagey M, Newman J, Bilodeau S, Zhan Y, Orlando D, van Berkum N, Ebmeier C, Goossens J, Rahl P, Levine S, et al. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467:** 430–435.

Kanduri C, Pant V, Loukinov D, Pugacheva E, Qi C-F, Wolffe A, Ohlsson R, Lobanenkov V. 2000. Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Curr Biol* **10:** 853–856.

Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, Heger A, Agam A, Slater G, Goodson M, et al. 2011. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477:** 289–294.

Kim KP, Thurston A, Mummery C, Ward-van Oostwaard D, Priddle H, Allegrucci C, Denning C, Young L. 2007a. Gene-specific vulnerability to imprinting variability in human embryonic stem cell lines. *Genome Res* **17:** 1731–1742.

Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenkov VV, Ren B. 2007b. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128:** 1231–1245.

Kim JK, Samaranayake M, Pradhan S. 2009. Epigenetic mechanisms in mammals. *Cell Mol Life Sci* **66:** 596–612.

Kobayashi H, Sakurai T, Sato S, Nakabayashi K, Hata K, Kono T. 2012. Imprinted DNA methylation reprogramming during early mouse embryogenesis at the Gpr1–Zdbf2 locus is linked to long cis-intergenic transcription. *FEBS Lett* **586:** 827–833.

Kuzmin I, Geil L, Gibson L, Cavinato T, Loukinov D, Lobanenkov V, Lerman MI. 2005. Transcriptional regulator CTCF controls human interleukin 1 receptor-associated kinase 2 promoter. *J Mol Biol* **346:** 411–422.

Lee S, Wevrick R. 2000. Identification of novel imprinted transcripts in the Prader-Willi syndrome and Angelman syndrome deletion region: Further evidence for regional imprinting control. *Am J Hum Genet* **66:** 848–858.

Lefebvre L, Viville S, Barton S, Ishino F, Keverne E, Surani A. 1998. Abnormal maternal behaviour and growth retardation associated with loss of the imprinted gene *Mest. Nat Genet* **20:** 163–169.

Lewis A, Reik W. 2006. How imprinting centres work. *Cytogenet Genome Res* **113:** 81–89.

Li LC, Dahiya R. 2002. MethPrimer: Designing primers for methylation PCRs. *Bioinformatics* **18:** 1427–1431.

Li E, Beard C, Jaenisch R. 1993. Role for DNA methylation in genomic imprinting. *Nature* **366:** 362–365.

Li LL, Keverne EB, Aparicio SA, Ishino F, Barton SC, Surani MA. 1999. Regulation of maternal behavior and offspring growth by paternally expressed Peg3. *Science* **284:** 330–334.

Lin S, Ferguson-Smith A, Schultz R, Bartolomei M. 2011. Nonallelic transcriptional roles of CTCF and cohesins at imprinted loci. *Mol Cell Biol* **31:** 3094–3104.

Lossie AC, Whitney MM, Amidon D, Dong HJ, Chen P, Theriaque D, Hutson A, Nicholls RD, Zori RT, Williams CA, et al. 2001. Distinct phenotypes distinguish the molecular classes of Angelman syndrome. *J Med Genet* **38:** 834–845.

Lyck L, Kroigard T, Finsen B. 2007. Unbiased cell quantification reveals a continued increase in the number of neocortical neurons during early post-natal development in mice. *Eur J Neurosci* **26:** 1749–1764.

Manolio TA. 2010. Genomewide association studies and assessment of the risk of disease. *N Engl J Med* **363:** 166–176.

Martin D, Pantoja C, Fernandez Minan A, Valdes-Quezada C, Molto E, Matesanz F, Bogdanovic O, de la Calle-Mustienes E, Dominguez O, Taher L, et al. 2011. Genome-wide CTCF distribution in vertebrates defines equivalent sites that aid the identification of disease-associated genes. *Nat Struct Mol Biol* **18:** 708–714.

McCole R, Oakey R. 2008. Unwitting hosts fall victim to imprinting. *Epigenetics* **3:** 258–260.

Mukhopadhyay R, Yu W, Whitehead J, Xu J, Lezcano M, Pack S, Kanduri C, Kanduri M, Ginjala V, Vostrov A, et al. 2004. The binding sites for the chromatin insulator protein CTCF map to DNA methylation-free domains genome-wide. *Genome Res* **14:** 1594–1602.

Murrell A, Heeson S, Reik W. 2004. Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. *Nat Genet* **36:** 889–893.

Nativio R, Wendt K, Ito Y, Huddleston J, Uribe-Lewis S, Woodfine K, Krueger C, Reik W, Peters J-M, Murrell A. 2009. Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. *PLoS Genet* **5:** e1000739.

Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK, Rynes E, Maurano MT, Vierstra J, Thomas S, et al. 2012. BEDOPS: High-performance genomic feature operations. *Bioinformatics* **28:** 1919–1920.

Nix D, Courdy S, Boucher K. 2008. Empirical methods for controlling false positives and estimating confidence in ChIP-seq peaks. *BMC Bioinformatics* **9:** 523.

Noonan JP, McCallion AS. 2010. Genomics of long-range regulatory elements. *Annu Rev Genomics Hum Genet* **11:** 1–23.

Ohlsson R. 2001. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet* **17:** 520–527.

Parelho V, Hadjur S, Spivakov M, Leleu M, Sauer S, Gregson H, Jarmuz A, Canzonetta C, Webster Z, Nesterova T, et al. 2008. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132:** 422–433.

Peters J, Williamson C. 2007. Control of imprinting at the Gnas cluster. *Epigenetics* **2:** 207–213.

Phillips JE, Corces VG. 2009. CTCF: Master weaver of the genome. *Cell* **137:** 1194–1211.

Plagge A, Isles A, Gordon E, Humby T, Dean W, Gritsch S, Fischer-Colbrie R, Wilkinson L, Kelsey G. 2005. Imprinted Nesp55 influences behavioral reactivity to novel environments. *Mol Cell Biol* **25:** 3019–3026.

Proudhon C, Duffie R, Ajjan S, Cowley M, Iranzo J, Carbajosa G, Saadeh H, Holland ML, Oakey RJ, Rakyan VK, et al. 2012. Protection against de novo methylation is instrumental in maintaining parent-of-origin methylation inherited from the gametes. *Mol Cell* **47:** 909–920.

Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S, Baglivo I, Pedone PV, Grimaldi G, Riccio A, et al. 2012. In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol Cell* **44:** 361–372.

Reik W, Walter J. 2001. Genomic imprinting: Parental influence on the genome. *Nat Rev Genet* **2:** 21–32.

Rubio E, Reiss D, Welcsh P, Disteche C, Filippova G, Baliga N, Aebersold R, Ranish J, Krumm A. 2008. CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci* **105:** 8309–8314.

Rugg-Gunn PJ, Ferguson-Smith AC, Pedersen RA. 2007. Status of genomic imprinting in human embryonic stem cells as revealed by a large cohort of independently derived and maintained lines. *Hum Mol Genet* **16:** R243–R251.

Sandhu KS, Shi C, Sjolinder M, Zhao Z, Gondor A, Liu L, Tiwari VK, Guibert S, Emilsson L, Imreh MP, et al. 2009. Nonallelic transvection of multiple imprinted loci is organized by the H19 imprinting control region during germline development. *Genes Dev* **23:** 2598–2603.

Schmidt D, Schwalie P, Ross-Innes C, Hurtado A, Brown G, Carroll J, Flicek P, Odom D. 2010. A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res* **20:** 578–588.

Schmidt D, Schwalie PC, Wilson MD, Ballester B, Goncalves A, Kutter C, Brown GD, Marshall A, Flicek P, Odom DT. 2012. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148:** 335–348.

Schulz R, Woodfine K, Menheniott TR, Bourc'his D, Bestor T, Oakey RJ. 2008. WAMIDEX: A web atlas of murine genomic imprinting and differential expression. *Epigenetics* **3:** 89–96.

Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov VV, et al. 2012. A map of the cis-regulatory sequences in the mouse genome. *Nature* **488:** 116–120.

Shin H, Liu T, Manrai AK, Liu XS. 2009. CEAS: Cis-regulatory element annotation system. *Bioinformatics* **25:** 2605–2606.

Singh P, Wu X, Lee D-H, Li A, Rauch T, Pfeifer G, Mann J, Szabo P. 2011. Chromosome-wide analysis of parental allele-specific chromatin and DNA methylation. *Mol Cell Biol* **31:** 1757–1770.

Smallwood SA, Tomizawa S, Krueger F, Ruf N, Carli N, Segonds-Pichon A, Sato S, Hata K, Andrews SR, Kelsey G. 2011. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat Genet* **43:** 811–814.

Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, Grosveld F, Galjart N, de Laat W. 2006. CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus. *Genes Dev* **20:** 2349–2354.

Stedman W, Kang H, Lin S, Kissil J, Bartolomei M, Lieberman P. 2008. Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. *EMBO J* **27:** 654–666.

Szabo P, Pfeifer G, Mann J. 2004. Parent-of-origin-specific binding of nuclear hormone receptor complexes in the H19-Igf2 imprinting control region. *Mol Cell Biol* **24:** 4858–4868.

Taatjes DJ. 2012. The human Mediator complex: A versatile, genome-wide regulator of transcription. *Trends Biochem Sci* **35:** 315–322.

Tsumura A, Hayakawa T, Kumaki Y, Takebayashi S, Sakaue M, Matsuoka C, Shimotohno K, Ishikawa F, Li E, Ueda HR, et al. 2006. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11:** 805–814.

Wendt K, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T, et al. 2008. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451:** 796–801.

Wilkins J. 2008. *Genomic imprinting*. Landes Bioscience/Springer, Austin, TX.

Williams C, Beaudet A, Clayton-Smith J, Knoll J, Kyllerman M, Laan L, Magenis E, Moncla A, Schinzel A, Summers J, et al. 2006. Angelman syndrome 2005: Updated consensus for diagnostic criteria. *Am J Med Genet* **140A:** 413–418.

Wood AJ, Roberts RG, Monk D, Moore GE, Schulz R, Oakey RJ. 2007. A screen for retrotransposed imprinted genes reveals an association between X chromosome homology and maternal germ-line methylation. *PLoS Genet* **3:** e20.

Wood AJ, Schulz R, Woodfine K, Koltowska K, Beechey CV, Peters J, Bourc'his D, Oakey RJ. 2008. Regulation of alternative polyadenylation by genomic imprinting. *Genes Dev* **22:** 1141–1146.

Xiao T, Wallace J, Felsenfeld G. 2011. 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. *Mol Cell Biol* **31:** 2174–2183.

Xie W, Barr CL, Kim A, Yue F, Young Lee A, Eubanks J, Dempster EL, Ren B. 2012. Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* **148:** 816–831.

Yalcin B, Wong K, Agam A, Goodson M, Keane TM, Gan X, Nellaker C, Goodstadt L, Nicod J, Bhomra A, et al. 2011. Sequence-based characterization of structural variation in the mouse genome. *Nature* **477:** 326–329.

Yoon B, Herman H, Hu B, Park Y, Lindroth A, Bell A, West A, Chang Y, Stablewski A, Piel J, et al. 2005. Rasgrf1 imprinting is regulated by a CTCF-dependent methylation-sensitive enhancer blocker. *Mol Cell Biol* **25:** 11184–11190.

Yusufzai TM, Felsenfeld G. 2004. The 5′-HS4 chicken β-globin insulator is a CTCF-dependent nuclear matrix-associated element. *Proc Natl Acad Sci* **101:** 8620–8624.