SCIENTIFIC REPORTS

Received: 07 July 2016 Accepted: 19 December 2016 Published: 01 February 2017

OPEN Testis-specific transcriptional regulators selectively occupy **BORIS-bound CTCF target regions** in mouse male germ cells

Samuel Rivero-Hinojosa^{1,*,†}, Sungyun Kang^{2,*}, Victor V. Lobanenkov¹ & Gabriel E. Zentner²

Despite sharing the same sequence specificity in vitro and in vivo, CCCTC-binding factor (CTCF) and its paralog brother of the regulator of imprinted sites (BORIS) are simultaneously expressed in germ cells. Recently, ChIP-seq analysis revealed two classes of CTCF/BORIS-bound regions: single CTCF target sites (1xCTSes) that are bound by CTCF alone (CTCF-only) or double CTCF target sites (2xCTSes) simultaneously bound by CTCF and BORIS (CTCF&BORIS) or BORIS alone (BORIS-only) in germ cells and in BORIS-positive somatic cancer cells. BORIS-bound regions (CTCF&BORIS and BORIS-only sites) are, on average, enriched for RNA polymerase II (RNAPII) binding and histone retention in mature spermatozoa relative to CTCF-only sites, but little else is known about them. We show that subsets of CTCF&BORIS and BORIS-only sites are occupied by several testis-specific transcriptional regulators (TSTRs) and associated with highly expressed germ cell-specific genes and histone retention in mature spermatozoa. We also demonstrate a physical interaction between BORIS and one of the analyzed TSTRs, TATA-binding protein (TBP)-associated factor 7-like (TAF7L). Our data suggest that CTCF and BORIS cooperate with additional TSTRs to regulate gene expression in developing male gametes and histone retention in mature spermatozoa, potentially priming certain regions of the genome for rapid activation following fertilization.

CTCF is a ubiquitously expressed chromatin-binding factor that contains 11 zinc fingers (ZFs) and is conserved from *Drosophila* to human^{1,2}. CTCF regulates transcription through promoter binding^{3,4} as well as methylation-sensitive insulator activity of the H19 imprinting control region, where it blocks enhancer looping to potential target promoters^{2,5-7} and can directly interact with the largest subunit of RNA polymerase II (RNAPII) to tether it to certain CTCF target sites (CTSes) in promoters and enhancers8. CTCF has also been implicated in numerous other chromatin-based processes, including RNAPII pausing9, regulation of ribosomal RNA transcription¹⁰⁻¹², X chromosome inactivation^{13,14}, and the formation of higher-order chromatin structures¹⁵⁻¹⁷. In addition to CTCF, mammals also express the paralogous protein brother of the regulator of imprinted sites (BORIS, also known as CTCFL)¹⁸. While CTCF is ubiquitously expressed¹⁸ and essential for development¹⁹, BORIS expression is limited to the adult testis¹⁸, where it is involved in spermatogenesis. BORIS-null mice are subfertile with reduced testis size and delayed gamete production^{20,21}. Various tumors and cancer cell lines also express BORIS, and so BORIS is considered to be a cancer-testis antigen²²⁻²⁸.

The existence of conditions under which both CTCF and BORIS, which have nearly identical DNA binding domains and thus bind the same DNA sequences^{20,21,25,29}, are co-expressed led to speculation that they might compete for a finite number of binding sites^{18,21}. However, mapping of CTCF and BORIS binding by ChIP-seq in human cancer and mouse germ cells revealed the existence of distinct classes of binding sites: 1xCTSes, containing a single motif and bound by CTCF alone (CTCF-only sites), and 2xCTSes, containing two or more clustered motifs and simultaneously bound by CTCF and BORIS (CTCF&BORIS sites) or BORIS alone (BORIS-only sites)30.

¹Molecular Pathology Section, Laboratory of Immunogenetics, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, MD 20852, USA. ²Department of Biology, Indiana University, Bloomington, IN 47405, USA. [†]Present address: Center for Cancer and Immunology Research, Children's National Medical Center, Washington, DC 20010, USA. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to G.E.Z. (email: gzentner@indiana.edu)

Despite the presence of multiple CTCF motifs within 2xCTSes, they display single peaks in ChIP-seq experiments, indicating that CTCF motifs generally lie too close together to be resolved by ChIP-seq^{30,31}. Indeed, single motifs are usually reported for CTCF ChIP-seq peaks³², suggesting that 2xCTSes remain generally unexamined in genome-wide CTCF mapping studies. At CTCF&BORIS and BORIS-only sites, robust enrichment of RNA RNAPII, H3K4me3, and H3K27ac was observed, suggesting a function in active transcription. Furthermore, CTCF&BORIS and BORIS-only sites were associated with histone retention in mature spermatozoa, which has been proposed to be a consequence of active transcription in spermatids³³ and is associated with genes expressed in early zygotic development³³⁻³⁷. These features contrast with 1xCTSes, which are not bound by BORIS, are generally located intergenically, and devoid of chromatin features associated with transcriptional activity³⁰. Combined with the observation that BORIS-only sites do not overlap binding sites for cohesin, involved in long-range chromatin interactions and transcriptional insulation^{38,39}, these data suggest that CTCF-only sites function in chromatin architecture, while BORIS-bound sites serve as transcriptional regulatory platforms. Indeed, prior to the genome-wide characterization of 2xCTSes, closely spaced CTCF motifs were described in several important regulatory regions including alternative human BORIS promoters⁴⁰, the testis-specific TSP50 promoter⁴¹, and the mouse KvDMR1 imprinting region⁴², suggesting that 2xCTSes are involved in important transcriptional regulatory mechanisms.

To better understand the potential functions of 2xCTSes, we undertook an integrative genomic analysis, incorporating ChIP-seq, RNA-seq, and MNase-seq data. We find that subsets of CTCF&BORIS and BORIS-only sites in mouse round spermatids are also occupied by a number of testis-specifc transcriptional regulators (TSTRs). These TSTR-occupied CTCF&BORIS (TSTR-CTCF&BORIS) and BORIS-only sites (TSTR-BORIS-only sites) are associated with genes highly expressed in testis and involved in male germ cell development. We also find that TSTR-associated BORIS-bound sites, but not their counterparts not bound by TSTRs, are associated with histone retention in mature sperm. Together, these observations suggest a functional link between CTCF and BORIS and other testis-specific TSTRs in the regulation of germ cell transcription. In support of this idea, we find that BORIS physically interacts with TATA-binding protein (TBP)-associated factor 7-like (TAF7L) in both cultured cells and human testis. Our analyses provide a refined view of BORIS-bound sites, identifying subsets of sites potentially involved in testis transcription and suggesting novel avenues for the exploration of functional connections between CTCF, BORIS, and several TSTRs in the regulation of male gametogenesis.

Results

TSTRs occupy a subset of CTCF&BORIS sites. To facilitate analysis of 2xCTSes, we first re-identified BORIS&CTCF sites in mouse round spermatids, making use of previously published CTCF and BORIS ChIP-seq data³⁰. We identified 35,308 CTCF peaks, 8,680 (24.6%) of which were also occupied by BORIS (Fig. 1A), comparable to the ~8,500 round spermatid CTCF&BORIS sites identified in a previous study³⁰. As we identified CTCF&BORIS sites using CTCF peaks as a reference, we first focused our analysis on a comparison of CTCF&BORIS to CTCF-only sites. Consistent with previously described properties of CTCF&BORIS sites³⁰, we found that our CTCF&BORIS sites were, on average, more highly enriched than CTCF-only sites for both RNAPII binding in testis as well as nucleosomes in mature sperm (Fig. 1B). Consistent with previous work⁴³, nucleosomes flanking CTCF-only and CTCF&BORIS sites were strongly phased.

Given that RNAPII is strongly enriched at CTCF&BORIS compared to CTCF-only sites³⁰ (Fig. 1B), we hypothesized that CTCF&BORIS sites might also be bound by testis-specific transcriptional regulators (TSTRs) in addition to CTCF and BORIS. To test this idea, we analyzed ChIP-seq datasets for a number of TSTRs: A-MYB, a regulator of male meiosis^{44,45}; BRD4, a bromodomain-containing factor that promotes elongation by RNAPII⁴⁶; DMRT1, essential for testis development⁴⁷; DMRT6, involved in spermatogonial differentiation⁴⁸; TAF7L, a paralog of the general transcription factor TAF7 that is essential for spermiogenesis^{49,50}; RFX2, essential for spermatid elongation⁵¹; and TRIM33, a poorly characterized transcriptional regulator (TR) containing a ubiquitin ligase domain, a bromodomain, and a plant homeodomain (PHD) finger responsible for silencing retrotransposons in the male germ line⁵².

We determined enrichment of each factor at CTCF&BORIS sites and then performed clustering, which revealed a striking partitioning of CTCF&BORIS sites into two classes. While the majority of CTCF&BORIS sites (7,514/8,680, 86.6%) displayed robust enrichment of CTCF and BORIS only, a substantial fraction (1,166/8,680, 13.4%) was also enriched for TSTRs (Fig. 2A, Supplementary Dataset 1). We refer this second class of sites as TSTR-associated CTCF&BORIS sites (TSTR-CTCF&BORIS sites). When datasets were scaled to identical read numbers and analyzed in aggregate, the most robustly enriched TSTRs were A-MYB, TAF7L, and DMRT6 (Fig. 2B). More modest enrichment was observed for BRD4, RFX2, and TRIM33, while little enrichment of DMRT1 was seen (Fig. 2B). Little to no enrichment of any TSTR was observed at CTCF-only sites (Fig. 2B). Of the identified TSTR-CTCF&BORIS sites, 693/1,166 (59.4%) were located between positions –1000 and +100 relative to the TSS, with the remainder located in untranslated regions, introns/exons, transcription termination sites, and intergenic regions. This observation suggests that TSTR-CTCF&BORIS sites are not simply due to a tendency for CTCF&BORIS sites to be located proximal to TSSs.

We next examined specific CTCF&BORIS sites through genome browser visualization. Analysis of a 40kb segment of chromosome 11 revealed two TSTR-CTCF&BORIS, one non-TSTR-CTCF&BORIS, and two CTCF-only sites (Fig. 3). As expected from our average analysis (Fig. 2B), no TSTR enrichment was observed at the non-TSTR-CTCF&BORIS site or the CTCF-only sites (Fig. 3). We also noted different combinations of TSTRs at the TSTR-CTCF&BORIS sites. The site located at the promoter of *Dnajc7* displayed robust binding of BRD4, RFX2, TAF7L, and TRIM33 with lower occupancy of A-MYB, DMRT1, and DMRT6, while the site in promoter of the nearby *Nkiras2* gene was robustly bound by BRD4 and TAF7L, with little or no binding of A-MYB, DMRT1, DMRT6, RFX2, or TRIM33 (Fig. 3). Visualization of the 2xCTS-containing CTCF&BORIS and



Distance from CTCF peak midpoint (kb)

Figure 1. *De novo* identification and characterization of CTCF&BORIS sites in mouse round spermatids. (A) Heatmaps showing enrichment of CTCF and BORIS at 35,308 CTCF ChIP-seq peaks in mouse round spermatids. (B) Average plots of (top) RNAPII enrichment at CTCF-only and CTCF&BORIS sites in mouse testis and (bottom) nucleosome enrichment at CTCF-only and CTCF&BORIS sites in mouse spermatozoa.

BORIS-only regions as single peaks regardless of TSTR occupancy highlights the previously described inability of ChIP-seq to resolve these closely spaced binding events in every analyzed mammalian cell type³⁰.

TSTR-CTCF&BORIS sites are associated with highly expressed male germ cell-specific genes. Given that TSTR-CTCF&BORIS sites are associated with TSTRs (Fig. 2), we anticipated that they might be associated with genes highly expressed in testis. We therefore analyzed the expression of genes associated with CTCF-only, non-TSTR-CTCF&BORIS, and TSTR-CTCF&BORIS sites using previously published testis RNA-seq data⁵⁰. We found that genes associated with TSTR-CTCF&BORIS sites were expressed at a significantly higher level than those associated with either CTCF-only ($p = 3.36 \times 10^{-66}$ by *t*-test) or non-TSTR-CTCF&BORIS sites ($p = 1.77 \times 10^{-73}$ by *t*-test) (Fig. 4A). This is consistent with previous work suggesting association of BORIS&CTCF sites with active promoters and enhancers³⁰. We next performed gene ontology (GO) analysis



Figure 2. Testis-specific transcriptional regulators occupy a subset of CTCF&BORIS sites. (A) Heatmaps showing enrichment of CTCF, BORIS, A-MYB, BRD4, DMRT1, DMRT6, RFX2, TAF7L, and TRIM33 at 8,680 CTCF&BORIS sites. (B) Average plots of TR enrichment at CTCF-only, non-STTR-CTCF&BORIS, and TSTR-CTCF&BORIS sites.

of genes associated with each category of site, focusing on germ cell-related annotations. This revealed significant association of eight distinct germ cell-related annotations with TSTR-CTCF&BORIS sites, compared with three for non-TSTR-CTCF&BORIS sites and one for CTCF-only sites (Fig. 4B). These observations support the idea that TSTR-CTCF&BORIS sites are involved in particular transcriptional programs specialized for both







Figure 4. TSTR-CTCF&BORIS sites are associated with highly transcribed male germ cell-specific genes. (A) Boxplots of log₂-transformed testis RNA-seq RPKM values for the closest gene to each CTCF-only, non-TSTR-CTCF&BORIS, and TSTR-CTCF&BORIS site. (B) Significant male or non-sex-specific germ cell-related GO terms for the nearest gene to each TSTR-CTCF&BORIS, non-TSTR-CTCF&BORIS, and CTCF-only site.

meiotic and post-meiotic developmental stages of spermatogenesis that support histone retention in mature sperm and thus facilitate transgenerational bookmarking of future promoters and enhancers on paternal DNA for post-fertilization zygotic genome activation (ZGA)⁵³.

TSTR-CTCF&BORIS sites are associated with histone retention in mature sperm. Previous work³⁰ and our own analysis (Fig. 1B) indicate that CTCF&BORIS sites are, on average, enriched in regions that retain histones in mature spermatozoa after the large-scale replacement of histones for protamines has occurred. Given that previous work has suggested that robust transcriptional activity in spermatids is linked to histone retention³³ and that TSTR-CTCF&BORIS sites are associated with highly expressed genes (Fig. 4A), we wondered if TSTR-CTCF&BORIS sites might also be preferentially associated with regions of histone retention in sperm. We therefore used MNase-seq data to assess the presence of nucleosomes at TSTR-CTCF&BORIS, sites were associated with robust nucleosome retention in spermatozoa (Fig. 5). As nucleosome retention in sperm has been proposed to be a consequence of transcription-coupled deposition of histone H3.3 in spermatids³³, we also assessed spermatozoal H3.3 enrichment. Consistent with the robust expression of TSTR-CTCF&BORIS site-associated genes, H3.3 was highly enriched at TSTR-CTCF&BORIS relative to CTCF-only and non-TSTR-CTCF&BORIS sites (Fig. 5).

A subset of BORIS-only sites are occupied by TSTRs. Thus far, we have shown that a fraction of CTCF&BORIS sites is also occupied by a number of TSTRs, and that these TSTR-CTCF&BORIS sites are linked to robust gene expression and retention of histones in sperm relative to non-TSTR-CTCF&BORIS and CTCF-only sites. However, 2xCTSes may also be occupied by BORIS alone (BORIS-only sites), and these BORIS-only sites display chromatin characteristics similar to those of CTCF&BORIS sites³⁰. We therefore asked





if some BORIS-only sites were also occupied by TSTRs. We called peaks on the BORIS ChIP-seq dataset and identified 1,693/8,644 (19.6%) sites not also bound by CTCF (Fig. 6A). We then quantified enrichment of A-MYB, BRD4, DMRT1, DMRT6, RFX2, TAF7L, and TRIM33 around BORIS-only sites and performed clustering. We found that 748/1,693 (44.2%) of BORIS-only peaks were also occupied by TSTRs (Fig. 6B). We note that the proportion of such TSTR-BORIS-only sites in BORIS-only sites is much higher than that of TSTR-CTCF&BORIS in CTCF&BORIS sites (44.2% vs. 13.4%). Of the identified TSTR-BORIS-only sites, 455/748 (60.8%) were located at the promoter/TSS regions, suggesting that, similar to TSTR-CTCF&BORIS sites, TSTR-BORIS-only sites are not simply the result of a locational bias. Genes associated with TSTR-BORIS-only sites were more highly expressed than those associated with non-TSTR-BORIS-only sites ($p = 8.48 \times 10^{-52}$ by *t*-test) (Fig. 6C), and genes associated with TSTR-BORIS-only sites, TSTR-BORIS of performance of the test of the more strongly linked to male germ cell biological process annotations (Fig. 6D). Like TSTR-CTCF&BORIS sites, TSTR-BORIS-only sites showed strong retention of nucleosomes bearing histone H3.3 in mature spermatozoa (Fig. 6E).

BORIS and TAF7L physically interact *in vitro* **and** *in vivo*. The genomic colocalization of BORIS with several TSTRs with or without concomitant CTCF binding led us to investigate potential physical interactions between these proteins. We focused on TAF7L, a paralog of the TFIID component TAF7 that is essential for spermiogenesis⁴⁹, as it is localized to the nucleus in round spermatids⁵⁴, from which the CTCF and BORIS ChIP-seq data were generated³⁰. We expressed HA-tagged BORIS, GFP-tagged CTCF, and myc-tagged TAF7L in HEK293T cells and performed co-IPs. BORIS but not CTCF was efficiently co-IPed with TAF7L, though we were unable to detect reciprocal co-IP of TAF7L with BORIS despite repeated attempts (Fig. 7A). To confirm the observation of BORIS/TAF7L physical interaction in cultured cells *in vivo*, we performed an *in situ* proximity ligation assay (ISPLA) on frozen human testis. Numerous fluorescent foci, indicative of spatial proximity, were observed (Fig. 7B), further suggesting a physical interaction between BORIS and TAF7L, though we were unable to determine the cell type in which these interactions occurred. To investigate a functional connection between TAF7L and BORIS in gene regulation, we first examined the effect of TAF7L loss on the expression of *Prss50*, a well-characterized BORIS target gene²⁹. Interestingly, *Taf7l^{-/Y}* testes displayed a ~2.7-fold increase in *Prss50* compared to wild type (Fig. 7C), perhaps suggesting that TAF7L acts to restrain transcriptional activation by BORIS.

We also examined potential physical interactions between CTCF and BORIS and two additional TSTRs, BRD4 and RFX2. We detected weak co-IP of CTCF with BRD4, though the reciprocal result was not observed (Fig. 8A). Similarly, we detected a small amount of RFX2 in a CTCF IP, though no CTCF was observed in an RFX2 IP (Fig. 8B). We note that the relative weakness or absence of these interactions does not necessarily preclude their taking place in germ cells, as cofactors and/or post-translational modifications necessary for their interaction may not be present in HEK293T cells.

Discussion

CTCF and its paralog BORIS occupy thousands of sites, both individually and in combination, throughout the genome in mouse male germ cells. Sites bound by BORIS, both with and without CTCF, are not well understood. Here, we integrated ChIP-seq, RNA-seq, and MNase-seq data into a genomic characterization of CTCF&BORIS and BORIS-only sites. We found that substantial fractions of CTCF&BORIS and BORIS-only sites are also occupied by TRs with known roles in male gametogenesis and associated with genes highly expressed in testis as well as sites of histone retention in spermatozoa.

Our results place BORIS and CTCF in the context of a larger germ cell-specific transcriptional network, suggesting that they cooperate with a number of testis-specific TSTRs potentially capable of physically interacting with CTCF and BORIS through protein-protein interactions with either the common central 11-ZF region or their unique C- and N-termini. At first approximation, it may appear that BORIS is a relatively minor component of this transcriptional network, as only a handful of genes have been identified as dysregulated in *BORIS*-null testes^{20,21}, with only two such genes (*Prss50, Gal3st1*) reported to be affected in both studies. However, it is known that CTCF can partially compensate for BORIS loss in mouse ES cells²¹ and that BORIS-only 2xCTSes are occupied by CTCF homodimers in BORIS-negative cells³⁰. Additionally, a recent study showed that while conditional



Figure 6. A subset of BORIS-only sites is associated with TSTRs. (A) Heatmaps showing enrichment of CTCF and BORIS at 8,644 BORIS ChIP-seq peaks in mouse round spermatids. (B) Heatmaps showing enrichment of CTCF, BORIS, A-MYB, BRD4, DMRT1, DMRT6, RFX2, TAF7L, and TRIM33 at 1,693 BORIS-only sites. (C) Boxplots of log₂-transformed testis RNA-seq RPKM values for the closest gene to each non-TSTR-BORIS-only and TSTR-BORIS-only site. (D) Significant male and non-sex-specific germ cell-related biological process GO terms for the nearest gene to each TSTR-BORIS-only site (non-TSTR-BORIS-only sites displayed no enrichment of male or non-sex-specific germ cell-related GO terms). (E) Average plot of nucleosome and histone H3.3 enrichment in spermatozoa around TSTR-BORIS-only and non-TSTR-BORIS-only sites.

knockout of Ctcf in male germ cells decreased the expression of 2,549 genes, only a small fraction of these genes displayed CTCF binding to their promoter or a nearby putative enhancer⁵⁵, suggesting that CTCF alone directly regulates the expression of relatively few genes. We note the possibility that homodimerization of CTCF or CTCF/ BORIS heterodimerization might be responsible for the formation of some TSTR-bound regions, as occupancy of TSTRs at CTCF-only sites, occupied by a CTCF monomer, is not observed. In this model, in BORIS-null testes, CTCF homodimers occupy 2xCTSes normally bound by CTCF/BORIS heterodimers or BORIS homodimers, allowing the formation of TSTR-bound regions and limiting the transcriptional effects of BORIS loss. However, because not all 2xCTSes are TSTR-occupied, there must be additional factors beyond the dimerization of CTCF and/or BORIS that promote TSTR binding at these sites. Moreover, the effect of conditional Ctcf knockout on male mouse fertility is more severe than that of *BORIS* deletion^{20,21,55}. Future studies comparing the transcriptional effects of CTCF and BORIS loss alone and in combination will be informative in understanding potential redundancy and specialization between these factors. One interesting direction to pursue in this regard may be *in vivo* analysis of the regulation of the TFIIA α/β -like factor (ALF) gene, encoding a germ cell-specific component of TFIIA⁵⁶, the promoter of which is differentially regulated by CTCF and BORIS in cultured cells⁵⁷. It may also be of interest to analyze chromatin interactions in germ cells expressing various levels of CTCF and BORIS. Fusion of the CTCF central 11-ZF domain to a different N-terminus, thereby creating a CTCF decoy with altered protein-protein interaction capabilities, causes disruptions in chromatin looping and imprinting at the human IGF2 locus⁵⁸. Given that the central 11-ZF regions of CTCF and BORIS are highly conserved but their N- and C-termini very divergent¹⁸, BORIS could potentially serve as a decoy for CTCF, altering chromatin interactions. Indeed, BORIS-only regions do not overlap with peaks for cohesin, thought to be generally required for long-range chromatin interactions³⁹, in human cancer cells³⁰, suggesting that expression of BORIS can alter the chromatin interaction landscape of a cell. Beyond analysis of redundancy and specialization between CTCF and BORIS, comparison of the transcriptional profiles of germ cells with single and combined BORIS and CTCF







Figure 8. Analysis of physical interactions of CTCF and BORIS with BRD4 and RFX2. (A) Western blots detecting GFP-CTCF, HA-BORIS, and myc-BRD4 following co-IP with the indicated antibodies in HEK293T cells. (B) Same as (A), but with myc-RFX2.

deficiency to those harboring mutations in other TSTRs associated with CTCF&BORIS and BORIS-only sites, such as TAF7L, shown here to physically interact with BORIS, may be a productive direction for understanding the transcriptional roles of CTCF and BORIS in male germ cells.

We note that while the CTCF and BORIS ChIP-seq data were generated using round spermatids, the TSTR ChIP-seq datasets, with the exception of BRD4, were generated using whole testis. It is therefore possible that our analyses were confounded by the presence of additional cell types. However, there is evidence that several of the analyzed factors function in round spermatids. TAF7L physically interacts with BORIS but not CTCF, suggesting a germ cell-exclusive role for this interaction, and is localized to the nucleus of round spermatids⁵⁴, providing further evidence that genomic colocalization with CTCF and BORIS does occur in these cells. There is also evidence that the other TSTRs analyzed here function in round spermatids. A-MYB can be detected in early round spermatids⁵⁹, and RFX2, itself present in round spermatids⁶⁰ and required for their elongation⁵¹, has been proposed to be an amplifier of A-MYB regulation⁵⁹. TRIM33 is also expressed in round spermatids⁶¹. Additionally, DMRT1, which was not found at TSTR-occupied regions, is not detected in spermatids. However, ChIP-seq mapping of additional TSTRs in round spermatids is necessary to conclusively address this.

Our data also bear on the proposed role of BORIS-bound regions as markers for histone retention in mature spermatozoa³⁰. We found that, of the identified CTCF&BORIS and BORIS-only sites, only those regions that were associated with TSTRs were strongly linked to protamine-refractory, histone-retaining regions in mature sperm. It has been suggested that ongoing transcription and concomitant deposition of the H3.3 histone variant in spermatids is a trigger for histone retention during the histone-to-protamine transition in the haploid phase of spermatogenesis³³. Indeed, we found that TSTR-CTCF&BORIS and TSTR-BORIS-only sites were associated both with highly expressed genes in testis and H3.3 occupancy in sperm, suggesting that retention of histones at BORIS binding sites in sperm is dependent on cooperation with TSTRs to generate high levels of transcription. TSTR-marked BORIS-bound sites in spermatids could therefore contribute to histone retention in sperm by promoting high levels of transcription in spermatids and also organizing sperm chromatin. However, it is unclear if BORIS is expressed in spermatozoa, so these regions might be bound by CTCF homodimers. CTCF protein is present in sperm⁶³, nuclease-sensitive regions in human spermatozoal chromatin are enriched in CTCF binding motifs³⁵, and there is potentially widespread occupancy of CTCF in mouse spermatozoa^{64,65}, suggesting a role for CTCF in organizing sperm chromatin. Indeed, CTCF-deficient spermatozoa display defects in chromatin condensation⁵⁵. Regardless, ChIP-seq profiling of CTCF and BORIS in spermatozoa will be necessary to determine the contribution of TSTR-associated BORIS binding sites to sperm chromatin organization.

Our results provide a refined view of the nature of BORIS-bound sites in the mouse male germ cell genome, indicating that subsets of these sites are potentially involved in promoting testis transcription through the synergistic action of CTCF and BORIS with a number of TSTRs. We find that the presence of a CTCF&BORIS or BORIS-only site is in and of itself insufficient for histone retention in spermatozoa, but suggest that synergy between CTCF and/or BORIS and additional TSTRs provides a high level of transcription that leads to histone H3.3 deposition and subsequent retention of histones after the histone-to-protamine transition. This suggests that TSTR-occupied BORIS-bound regions function in the maintenance of transgenerational bookmarks at promoters and enhancers to be activated during ZGA following fertilization.

Methods

Datasets. All datasets were obtained from the SRA. ChIP-seq: A-MYB, whole testis⁴⁴ (SRX244353); BORIS, round spermatids³⁰ (SRX1091833); CTCF, round spermatids³⁰ (SRX1091832); BRD4, round spermatids⁴⁶ (SRX719833); DMRT1, whole testis⁴⁷ (SRX838549); DMRT6, whole testis⁴⁸ (SRX681861); (SRX112976); Histone H3.3, mature spermatozoa³³ (SRX207534); input, round spermatids³⁰ (SRX1091834); RFX2, whole testis⁵¹ (SRX1011026); RNAPII, whole testis⁵⁰ (SRX349391); TAF7L, whole testis⁵⁰ (SRX349390); TRIM33, whole testis⁵² (SRX1019997); MNase-seq: mature spermatozoa³³ (SRX207533); RNA-seq: whole testis⁵⁰ (SRX349392).

Data analysis. Alignment. ChIP-seq and MNase-seq data were aligned to the mm10 mouse genome build using Bowtie2⁶⁶ with default parameters. RNA-seq data were aligned to the mm10 mouse genome build with STAR⁶⁷ version 2.4.1 d using the corresponding refGene GTF file from the UCSC Table Browser.

Peak calling. Tag directories were created from CTCF, BORIS, and input aligned SAM files with HOMER⁶⁸ (http://homer.salk.edu). Peaks were called with HOMER findPeaks using –style factor and round spermatid input as control. CTCF and BORIS peaks are listed in Additional File 1, Tabs 1 and 2.

Cluster analysis and heatmaps. Aligned SAM files were converted to BAM format with SAMTools⁶⁹ and input into SeqMiner⁷⁰ for clustering. We used linear *k*-means clustering with k = 3 of the CTCF and BORIS datasets around CTCF peaks to identify CTCF-only and CTCF&BORIS sites (Supplementary Dataset 1, Tabs 3 and 4). We then used ranked *k*-means clustering with k = 4 of all datasets around CTCF&BORIS sites to identify CTCF&BORIS sites, supplementary Dataset 1, Tabs 5 and 6). We used linear *k*-means clustering with k = 2 of the CTCF and BORIS datasets around BORIS peaks to identify BORIS-only sites (Supplementary Dataset 1, Tabs 7). We then used ranked *k*-means clustering with k = 2 of all datasets around BORIS peaks to identify BORIS-only sites bound or not bound by additional transcriptional regulators (TSTR-BORIS-only sites to identify BORIS-only sites bound or not bound by additional transcriptional regulators (TSTR-BORIS-only sites to identify BORIS-only sites bound or not bound by additional transcriptional regulators (TSTR-BORIS-only sites to identify BORIS-only sites bound or not bound by additional transcriptional regulators (TSTR-BORIS-only sites to identify BORIS-only sites bound or not bound by additional transcriptional regulators (TSTR-BORIS-only and non-TSTR-BORIS-only, Supplementary Dataset 1, Tabs 8 and 9). All heatmaps were generated with SeqMiner using a contrast value of 50.

Average plots. The annotatePeaks.pl feature of HOMER was used to generate average plots around CTCF-only and CTCF&BORIS sites with 1 bp resolution. Default normalization (scaling of tag directories to 10 million reads) was used.

Genome browser visualization. HOMER makeUCSCfile was used to generate bedgraphs normalized to 10 million reads. Bedgraphs were converted to bigwig format and visualized in IGV^{71} .

Transcript quantification. HOMER analyzeRepeats.pl was used to quantify RNA-seq reads in exons, allowing a maximum of 3 tags per position (-pc 3) and using gene expression matrix normalization (-normMatrix 1e7).

Gene ontology analysis. GO analysis was performed with HOMER annotatePeaks.pl. The $-\ln(p$ -value) of each biological process annotation related to male germ cell development was reported.

Co-immunoprecipitation (co-IP). HEK293T cells were grown in DMEM + 10% FBS and transfected with pCI-Neo-GFP-CTCF and pCI-Neo-HA-BORIS along with pCI-Neo-myc-TAF7L, pReceiver-myc-BRD4 (Genecopoeia), or pReceiver-myc-RFX2 (Genecopoeia). For co-IP analysis, 1.5 mg of total protein was incubated with anti-Myc antibody (Sigma), anti-HA antibody (Roche), or anti-GFP antibody (Abcam ab290) overnight at 4°C with gentle agitation, followed by incubation with 50µl of a 1:1 mixture of Protein A and Protein G Dynabeads (Life Technologies) for 1 h at room temperature. The immunoprecipitates were collected using a magnetic rack and washed and analyzed by SDS-PAGE and western blotting.

In situ proximity ligation assay (ISPLA). Frozen human testis slides were fixed in 4% paraformaldehyde for 10 min at 37 °C. Slides were then permeabilized for 2 h with 0.2% Triton X-100 in PBS then blocked in 3% bovine serum albumin (Sigma) in a humid chamber for 1 h at 37 °C and incubated overnight at 4 °C with mouse and rabbit antibodies: custom mouse monoclonal anti-BORIS and rabbit anti-TAF7L (Santa-Cruz, S-24) in blocking solution. After washing, the slides were incubated with Duolink PLA Rabbit MINUS and PLA Mouse PLUS probes. Ligation and detection were performed using the Duolink reagents kit (Olink Bioscience, Sigma) according to the manufacturer's protocol. Fluorescence was detected using a Zeiss Plan Apochromat microscope with a 63x/oil objective. Images were acquired with an Axiocam MRm camera and the Zeiss LSM Image Browser. Presence or absence of foci in specific testis cell types was not assessed.

References

- 1. Ohlsson, R., Renkawitz, R. & Lobanenkov, V. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet.* **17**, 520–527 (2001).
- 2. Moon, H. *et al.* CTCF is conserved from Drosophila to humans and confers enhancer blocking of the Fab-8 insulator. *EMBO reports* **6**, 165–170 (2005).
- Filippova, G. N. *et al.* An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol. Cell. Biol.* 16, 2802–2813 (1996).
- Yang, Y., Quitschke, W., Vostrov, A. A. & Brewer, G. J. CTCF Is Essential for Up-Regulating Expression from the Amyloid Precursor Protein Promoter During Differentiation of Primary Hippocampal Neurons. *J. Neurochem.* 73, 2286–2298 (1999).
 Kanduri, C. *et al.* Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and
- Kanduri, C. *et al.* 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 (2000).
- Szabó, P. E., Tang, S.-H. E., Rentsendorj, A., Pfeifer, G. P. & Mann, J. R. Maternal-specific footprints at putative CTCF sites in the H19 imprinting control region give evidence for insulator function. *Curr. Biol.* 10, 607–610 (2000).
- 7. Kurukuti, S. *et al.* CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proceedings of the National Academy of Sciences* **103**, 10684–10689 (2006).
- Chernukhin, I. et al. CTCF Interacts with and Recruits the Largest Subunit of RNA Polymerase II to CTCF Target Sites Genome-Wide. Mol. Cell. Biol. 27, 1631–1648 (2007).
- 9. Shukla, S. et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature 479, 74-79 (2011).
- van de Nobelen, S. *et al.* CTCF regulates the local epigenetic state of ribosomal DNA repeats. *Epigenetics & Chromatin* 3, 19 (2010).
 Guerrero, P. A. & Maggert, K. A. The CCCTC-Binding Factor (CTCF) of *Drosophila* Contributes to the Regulation of the Ribosomal
- DNA and Nucleolar Stability. *PLoS ONE* **6**, e16401 (2011). 12. Huang, K. *et al.* Ribosomal RNA Gene Transcription Mediated by the Master Genome Regulator Protein CCCTC-binding Factor
- (CTCF) Is Negatively Regulated by the Condensin Complex. J. Biol. Chem. **288**, 26067–26077 (2013). 13. Chao, W., Huynh, K. D., Spencer, R. J., Davidow, L. S. & Lee, J. T. CTCF, a Candidate Trans-Acting Factor for X-Inactivation Choice.
- Science 295, 345–347 (2002).
- Filippova, G. N. et al. Boundaries between Chromosomal Domains of X Inactivation and Escape Bind CTCF and Lack CpG Methylation during Early Development. Dev. Cell 8, 31–42 (2005).
- 15. Handoko, L. et al. CTCF-mediated functional chromatin interactome in pluripotent cells. Nat. Genet. 43, 630–638 (2011).
- Guo, Y. *et al.* CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* 162, 900–910 (2015).
- 17. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. USA* **112**, E6456–E6465 (2015).
- Loukinov, D. I. 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. Proceedings of the National Academy of Sciences 99, 6806–6811 (2002).
- 19. Moore, J. M. *et al.* Loss of Maternal CTCF Is Associated with Peri-Implantation Lethality of *Ctcf* Null Embryos. *PLoS ONE* 7, e34915 (2012).
- Suzuki, T. et al. 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. 30 (2010).
- Sleutels, F. 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 5, 1–21 (2012).
- D'Arcy, V. et al. The Potential of BORIS Detected in the Leukocytes of Breast Cancer Patients as an Early Marker of Tumorigenesis. American Association for Cancer Research 12, 5978–5986 (2006).
- Risinger, J. I. et al. Global Expression Analysis of Cancer/Testis Genes in Uterine Cancers Reveals a High Incidence of BORIS Expression. American Association for Cancer Research 13, 1713–1719 (2007).

- 24. D'Arcy, V. *et al.* BORIS, a paralogue of the transcription factor, CTCF, is aberrantly expressed in breast tumours. *Br. J. Cancer* **98**, 571–579 (2008).
- 25. Pugacheva, E. M. et al. The structural complexity of the human BORIS gene in gametogenesis and cancer. PloS one 5 (2010).
- 26. Martin-Kleiner, I. BORIS in human cancers A review. Eur. J. Cancer 48, 929–935 (2012).
- 27. Marshall, A. D., Bailey, C. G. & Rasko, J. E. J. CTCF and BORIS in genome regulation and cancer. Curr. Opin. Genet. Dev. 24, 8–15 (2014).
- 28. Cheema, Z. et al. Expression of the cancer-testis antigen BORIS correlates with prostate cancer. The Prostate 74, 164–176 (2014).
- Kosaka-Suzuki, N. et al. 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. 286 (2011).
- Pugacheva, E. M. et al. Comparative analyses of CTCF and BORIS occupancies uncover two distinct classes of CTCF binding genomic regions. Genome Biology 16, 1–24 (2015).
- Rowley, M. J. & Corces, V. G. The three-dimensional genome: principles and roles of long-distance interactions. *Curr. Opin. Cell Biol.* 40, 8–14 (2016).
- Wang, J. et al. Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. Nucleic Acids Res. 41, D171–D176 (2013).
- Erkek, S. et al. Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. Nat. Struct. Mol. Biol. 20, 868–875 (2013).
- 34. Gardiner-Garden, M., Ballesteros, M., Gordon, M. & Tam, P. P. L. Histone- and Protamine-DNA Association: Conservation of Different Patterns within the β-Globin Domain in Human Sperm. *Mol. Cell. Biol.* 18, 3350–3356 (1998).
- Arpanahi, A. et al. Endonuclease-sensitive regions of human spermatozoal chromatin are highly enriched in promoter and CTCF binding sequences. Genome Res. 19, 1338–1349 (2009).
- 36. Hammoud, S. S. et al. Distinctive chromatin in human sperm packages genes for embryo development. Nature 460, 473-478 (2009).
- Brykczynska, U. et al. Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. Nat. Struct. Mol. Biol. 17, 679–687 (2010).
- 38. Wendt, K. S. et al. Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature 451 (2008).
- Sofueva, S. & Hadjur, S. Cohesin-mediated chromatin interactions—into the third dimension of gene regulation. Briefings in Functional Genomics (2012).
- 40. Renaud, S. *et al.* Expression of the CTCF-paralogous cancer-testis gene, brother of the regulator of imprinted sites (BORIS), is regulated by three alternative promoters modulated by CpG methylation and by CTCF and p53 transcription factors. *Nucleic Acids Res.* **35**, 7372–7388 (2007).
- Kosaka-Suzuki, N. *et al.* 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. *The Journal of Biological Chemistry* 286, 27378–27388 (2011).
- 42. Fitzpatrick, G. V. et al. Allele-Specific Binding of CTCF to the Multipartite Imprinting Control Region KvDMR1. Mol. Cell. Biol. 27, 2636–2647 (2007).
- 43. Fu, Y., Sinha, M., Peterson, C. L. & Weng, Z. The Insulator Binding Protein CTCF Positions 20 Nucleosomes around Its Binding Sites across the Human Genome. *PLoS Genet.* **4**, e1000138 (2008).
- Li, Xin Z. et al. An Ancient Transcription Factor Initiates the Burst of piRNA Production during Early Meiosis in Mouse Testes. Mol. Cell 50, 67–81 (2013).
- Bolcun-Filas, E. et al. A-MYB (MYBL1) transcription factor is a master regulator of male meiosis. Development 138, 3319–3330 (2011).
- Bryant, J. M. et al. Characterization of BRD4 during Mammalian Postmeiotic Sperm Development. Mol. Cell. Biol. 35, 1433–1448 (2015).
- Murphy, M. W. et al. An ancient protein-DNA interaction underlying metazoan sex determination. Nat. Struct. Mol. Biol. 22, 442–451 (2015).
- Zhang, T., Murphy, M. W., Gearhart, M. D., Bardwell, V. J. & Zarkower, D. The mammalian Doublesex homolog DMRT6 coordinates the transition between mitotic and meiotic developmental programs during spermatogenesis. *Development* 141, 3662–3671 (2014).
 Cheng, Y. *et al.* Abnormal Sperm in Mice Lacking the Taf7l Gene. *Mol. Cell. Biol.* 27, 2582–2589 (2007).
- 50. Zhou, H. *et al.* Taf71 cooperates with Trf2 to regulate spermiogenesis. *Proceedings of the National Academy of Sciences* **110**, 16886–16891 (2013).
- 51. Kistler, W. S. et al. RFX2 Is a Major Transcriptional Regulator of Spermiogenesis. PLoS Genet. 11, e1005368 (2015).
- Isbel, L. et al. Trim33 Binds and Silences a Class of Young Endogenous Retroviruses in the Mouse Testis; a Novel Component of the Arms Race between Retrotransposons and the Host Genome. PLoS Genet. 11, e1005693 (2015).
- 53. Gill, M. E., Erkek, S. & Peters, A. H. F. M. Parental epigenetic control of embryogenesis: a balance between inheritance and reprogramming? *Curr. Opin. Cell Biol.* 24, 387–396 (2012).
- Pointud, J.-C. et al. The intracellular localisation of TAF7L, a paralogue of transcription factor TFIID subunit TAF7, is developmentally regulated during male germ-cell differentiation. J. Cell Sci. 116, 1847–1858 (2003).
- 55. Hernández-Hernández, A., Lilienthal, I., Fukuda, N., Galjart, N. & Höög, C. CTCF contributes in a critical way to spermatogenesis and male fertility. *Scientific Reports* 6, 28355 (2016).
- Upadhyaya, A. B., Lee, S. H. & DeJong, J. Identification of a General Transcription Factor TFIIAα/β Homolog Selectively Expressed in Testis. J. Biol. Chem. 274, 18040–18048 (1999).
- 57. Kim, M. et al. Regulatory Factor Interactions and Somatic Silencing of the Germ Cell-specific ALF Gene. J. Biol. Chem. 281, 34288-34298 (2006).
- Zhang, H. *et al.* Interruption of intrachromosomal looping by CCCTC binding factor decoy proteins abrogates genomic imprinting of human insulin-like growth factor II. *The Journal of Cell Biology* 193, 475–487 (2011).
- Horvath, G. C., Kistler, M. K. & Kistler, W. S. RFX2 is a candidate downstream amplifier of A-MYB regulation in mouse spermatogenesis. BMC Dev. Biol. 9, 1–15 (2009).
- Horvath, G. C., Kistler, W. S. & Kistler, M. K. RFX2 Is a Potential Transcriptional Regulatory Factor for Histone H1t and Other Genes Expressed During the Meiotic Phase of Spermatogenesis. *Biol. Reprod.* 71, 1551–1559 (2004).
- 61. Yan, K.-P. *et al.* Molecular cloning, genomic structure, and expression analysis of the mouse transcriptional intermediary factor 1 gamma gene. *Gene* **334**, 3–13 (2004).
- Matson, C. K. et al. The Mammalian Doublesex Homolog DMRT1 Is a Transcriptional Gatekeeper that Controls the Mitosis versus Meiosis Decision in Male Germ Cells. Dev. Cell 19, 612–624 (2010).
- Tang, J.-B. & Chen, Y.-H. Identification of a tyrosine-phosphorylated CCCTC-binding nuclear factor in capacitated mouse spermatozoa. Proteomics 6, 4800–4807 (2006).
- 64. Carone, B. R. et al. High resolution mapping of chromatin packaging in mouse ES cells and sperm. Dev. Cell 30, 11–22 (2014).
- 65. Johnson, G. D., Jodar, M., Pique-Regi, R. & Krawetz, S. A. Nuclease Footprints in Sperm Project Past and Future Chromatin Regulatory Events. *Scientific Reports* **6**, 25864 (2016).
- 66. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357-359 (2012).
- 67. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15-21 (2013).

- Heinz, S. et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol. Cell 38, 576–589 (2010).
- 69. Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079 (2009).
- 70. Ye, T. et al. seqMINER: an integrated ChIP-seq data interpretation platform. Nucleic Acids Res. 39, e35-e35 (2011).
- 71. Robinson, J. T. et al. Integrative Genomics Viewer. Nat. Biotechnol. 29, 24–26 (2011).

Acknowledgements

We thank Drs Dmitry Loukinov, Elena Pugacheva, and Sue Robinson for invaluable practical assistance during the portion of this work carried out in the Lobanenkov lab and members of the Lobanenkov and Zentner labs for helpful comments on the manuscript. This work was supported by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases, National Institutes of Health [to VVL] and Indiana University startup funds [to GEZ].

Author Contributions

S.R.-H., V.V.L., and G.E.Z. designed the study; S.R.-H. and S.K. performed experiments; S.R.-H. and G.E.Z. analyzed data; V.V.L. and G.E.Z. wrote the paper. All co-authors participated in editing of manuscript drafts and have approved the final manuscript.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Rivero-Hinojosa, S. *et al.* Testis-specific transcriptional regulators selectively occupy BORIS-bound CTCF target regions in mouse male germ cells. *Sci. Rep.* 7, 41279; doi: 10.1038/srep41279 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© The Author(s) 2017