


RESEARCH

Open Access



The cancer-associated CTCFL/BORIS protein targets multiple classes of genomic repeats, with a distinct binding and functional preference for humanoid-specific SVA transposable elements

Elena M. Pugacheva², Evgeny Teplyakov¹, Qiongfang Wu¹, Jingjing Li¹, Cheng Chen¹, Chengcheng Meng¹, Jian Liu¹, Susan Robinson², Dmitry Loukinov², Abdelhalim Boukaba¹, Andrew Paul Hutchins³, Victor Lobanenkov² and Alexander Strunnikov^{1*} 

Abstract

Background: A common aberration in cancer is the activation of germline-specific proteins. The DNA-binding proteins among them could generate novel chromatin states, not found in normal cells. The germline-specific transcription factor BORIS/CTCF, a paralog of chromatin architecture protein CTCF, is often erroneously activated in cancers and rewires the epigenome for the germline-like transcription program. Another common feature of malignancies is the changed expression and epigenetic states of genomic repeats, which could alter the transcription of neighboring genes and cause somatic mutations upon transposition. The role of BORIS in transposable elements and other repeats has never been assessed.

Results: The investigation of BORIS and CTCF binding to DNA repeats in the K562 cancer cells dependent on BORIS for self-renewal by ChIP-chip and ChIP-seq revealed three classes of occupancy by these proteins: elements cohabited by BORIS and CTCF, CTCF-only bound, or BORIS-only bound. The CTCF-only enrichment is characteristic for evolutionary old and inactive repeat classes, while BORIS and CTCF co-binding predominately occurs at uncharacterized tandem repeats. These repeats form staggered cluster binding sites, which are a prerequisite for CTCF and BORIS co-binding. At the same time, BORIS preferentially occupies a specific subset of the evolutionary young, transcribed, and mobile genomic repeat family, SVA. Unlike CTCF, BORIS prominently binds to the VNTR region of the SVA repeats in vivo. This suggests a role of BORIS in SVA expression regulation. RNA-seq analysis indicates that BORIS largely serves as a repressor of SVA expression, alongside DNA and histone methylation, with the exception of promoter capture by SVA.

Conclusions: Thus, BORIS directly binds to, and regulates SVA repeats, which are essentially movable CpG islands, via clusters of BORIS binding sites. This finding uncovers a new function of the global germline-specific transcriptional regulator BORIS in regulating and repressing the newest class of transposable elements that are actively transposed in human genome when activated. This function of BORIS in cancer cells is likely a reflection of its roles in the germline.

*Correspondence: alex@gibh.org

¹ Molecular Epigenetics Laboratory, Guangzhou Institutes of Biomedicine and Health, Guangzhou 510530, Guangdong, China
Full list of author information is available at the end of the article

Background

Transposable elements (TEs) play active roles in normal genome evolution in humans [1] and in primates in general [2], as well as in sporadic genome rearrangement [3–5] including deleterious events associated with pathology [6–12]. Multiple polymorphisms and intron evolution in normal human populations are largely facilitated by TE insertions [13, 14]. A substantial and distinct role of satellite repeats was also recently demonstrated for double-strand breaks (DSBs) incidence upon replication stress [15]. Active families of TEs (L1, Alu, and SVA) account for a large number of germline mutations [16]. In cancer, insertions of mobile element and the recombination between them have been identified as causes of many cancers [12, 17, 18], with some repeats shown to become aberrantly expressed [17, 19] to acquire a potential to change the regulation of neighboring genes [17, 20, 21] and to destabilize chromosomes [7, 22]. The effect of repeated DNA in the origins and progression of cancer and tumor cell physiology could be two-pronged: the induced change of expression in neighboring or targeted genes [22–24] and the structural destabilization of the epigenetic landscape of chromosomes [2, 25]. These two effects are interrelated, as epigenetic changes in the repeats open chromosomal domains for both aberrant changes in gene expression and elevated somatic recombination. Some elements were also shown to act as *bona fide* enhancers [26].

The presence of a strong epigenetic component in such repeats and TE-mediated genome regulation and instability is well established [20, 27–30]. In cancer cells, there is likely a higher epigenetic impact of TEs, compared to the norm [12], as promoters of expressed mobile elements become hypomethylated and their transcription elevated [22, 31, 32].

The array of epigenetic changes leading to repeat deregulation in cancer cannot be understood without molecular analysis of repeats' chromatin. This brings to light the role of CTCF and its paralog CTCFL/BORIS in these processes. In addition to serving as a *bona fide* transcription factor, CTCF reads the epigenetic marks [33–36] and plays a key role in the formation of topologically associated domains (TADs) in chromatin [37–39], in remodeling chromatin structure [40], and in the formation of chromatin boundaries [29, 41]. CTCF was also shown to have multiple binding sites embedded in TEs [42, 43]. CTCF target sites (CTSs) are also important for telomere repeat stability [44, 45]. Furthermore, the fact that CTCF control of gene expression and recombination requires physical contacts between different CTSs via looping [46–49] indicates that CTCF sites in repeats are not inert in the chromatin architecture, as indeed was demonstrated at some instances [50–53].

Taking into account the important role of CTCF in regulating TE expression and epigenetic maintenance, it is possible that the aberrant activation of its germline paralog CTCFL/BORIS in cancer has an impact on repeat physiology and genome stability. BORIS is a cancer testis (CT) gene [54], and its ectopic expression could be lethal or inhibitory for somatic cells because BORIS, being a germline transcription factor, activates gene expression of germline-specific genes on its own or in cooperation with CTCF [55]. Nevertheless, some cancer cells undergo adaptation/addiction to BORIS activation and incorporate the BORIS protein into their physiology [55, 56]. BORIS also interferes with a variety of other CTCF-specific functions in somatic cells, such as in the organization of chromatin loops that are alternative to the chromatin configuration of normal cells [55]. The ultimate molecular and physiological role of BORIS in cancer is still poorly understood, however, beyond the association with stemness [56], phenocopying of germline-specific gene expression pattern, and the corresponding 3D chromatin organization [55]. In particular, it is not clear how some cancer cells became dependent on BORIS for their proliferation, making BORIS a potential anticancer target [57, 58].

While many genomic repeats are heavily methylated and BORIS has a probable role in DNA demethylation [57, 59–61], the role of BORIS in repeat biology has not been studied. Incidentally, even the most comprehensive genome-wide studies on CTCF tended to ignore the possible simultaneous presence of BORIS in cells studied, be it cancer or embryonic stem cells [48, 50, 62–64]. In this present study, we attempted to assess the specific pattern of BORIS recognition of genomic repeats in cancer cells and to link it to TE expression. As a result, we uncovered a surprising association of BORIS with one of the evolutionary youngest families of actively transcribed and mobile repeats in human genome, the SVA family of TEs. Follow-up analysis of the modulation of BORIS expression revealed that it predominately acts as one of the mechanisms repressing the expression of these elements.

Results

BORIS expression in K562 forms a specific pattern of repeat binding

We have previously shown that tandem repeats (TRs) in a human cancer cell line may serve as foci for multiple DNA damage events induced upon the resolution of mitotic chromosome bridges [65]. In that study, custom repeat microarray ChIP-chip was used to validate some of the enrichments identified in the preceding ChIP-seq analysis. The need for a two-method validation procedure stems from the fact that at present there is no unbiased way to align short next-generation sequencing

(NGS) reads to massively repeated DNA, while microarray analysis has well documented limitations of its own. Here, we employed a similar two-step approach in reverse; the repeats' enrichment by DNA-binding proteins was first assessed by ChIP-chip and then validated by ChIP-seq. We used mainly the established cancer cell line K562 as a model for the coexistence of CTCF and BORIS stably expressed at a relatively the same level, as assayed by RT-PCR [55], to assess genome repeat occupancy by these two proteins. K562 retains a set of properties characteristic for cancer stem cells, e.g., the ability to initiate tumors in graft models, and the propensity to differentiate in response to exogenous stimuli [66]. As CTCF and BORIS have essentially the same composition of the DNA-binding domain, including the number of ZF and their spacing, as well as residues involved in DNA contacts (Fig. 1a), they show the virtually identical DNA-binding specificity in vitro, albeit not in native chromatin [55]. Therefore, it was important to use a cell line where two proteins are expressed in equivalent amounts, such as K562. Unlike most established cancer cell lines or primary non-germline tumor cells, where the expression of BORIS is low, with only a minor subset of cells characterized by high BORIS expression [56], K562 expresses high level of BORIS largely localized to the nuclei (Fig. 1b). BORIS was also confirmed to be incorporated into transcription regulation in K562 and to be required for its self-renewal [55].

For the initial analysis, by ChIP-chip, anti-CTCF and anti-BORIS immunoprecipitations were conducted and microarray hybridization was performed as described in Methods. The plot of normalized ChIP-chip fluorescence intensities showed indications of distinct binding patterns for BORIS and CTCF on highly enriched tiles (Fig. 1c). Significance analysis of microarray (SAM) indicated that over 40,000 tiles were enriched differentially by CTCF and BORIS, but provided little clue about the occupancy of the rest of the repeats. The principal component analysis (PCA) of arrays hybridized to CTCF and BORIS ChIP samples confirmed the presence of differentially bound genomic repeats (Fig. 1d). The PCA also revealed the three expected scenarios of occupancy: binding by BORIS only, by CTCF only, and BORIS and CTCF co-binding being by far the largest group (Fig. 1e). As CTCF and BORIS have essentially the same DNA-binding specificities in vitro, the differences in occupancy observed in vivo must be largely driven by the epigenetic factors.

Prior to proceeding further with analyses of repeat binding sequences, we conducted a validation of ChIP-chip data using an alternative high-throughput procedure, ChIP-seq, as conventional qPCR validation methods are not applicable or scalable to the TRs

genome-wide. We set out to validate the three identified subsets: first, repeats preferentially enriched by CTCF (Fig. 1e), second, repeats preferentially enriched by BORIS (Fig. 1e), and, third, repeats equally enriched by both CTCF and BORIS (Fig. 1e, a subset of the middle group). Based on detailed PCA analysis, an additional cutoff across the three groups was applied to make uniform criteria for selecting the representative subsets for validation. For co-bound repeats we chose the 4× enrichment for both proteins in all three ChIP-chip replicates, while for the Z5 groups we used 4× enrichment for one protein, with no enrichment for the other, also in all three replicates. Drawing the threshold at such a relatively high level also significantly reduced repeat redundancy in the TR dataset. For the ChIP-seq validation, we considered a ChIP-chip-positive repeat validated, if any tile from that repeat was reproducibly enriched at least twofold in ChIP-seq datasets with 95 % DNA match. Thus, all the repeats discussed below are repeats identified by ChIP-chip and validated by ChIP-seq.

Co-binding of BORIS and CTCF is characteristic for the simple tandem repeats

The simultaneous binding of BORIS and CTCF genome-wide in cancer cell lines was shown to reset, at least partially, the functions of CTSs in transcriptional regulation in accordance with germline-like program [55]. Thus, from the standpoint of cancer biology, it was important to characterize repeats bound by both CTCF and BORIS (CTCF and BORIS repeats, Additional file 1: Table S2), especially as they outnumbered other classes (Fig. 1e). The 171 distinct repeats in the CTCF and BORIS class were mostly represented by uncharacterized simple repeats, which can also be classified as VNTRs, and a small fraction of TEs, with the telomeric satellite *TAR1* notably dominating the rest of the group (Fig. 2a; Additional file 1: Table S2). It has to be appreciated that there is no certain way to determine whether both CTCF and BORIS co-bind the given individual repeat sequence, due to the multiple copies of repeats present and the propensity of CTCF and BORIS to induce interchromatin contacts [49, 55]. Nevertheless, the presence of cluster CTS is a strong indication of co-binding [55]. While this group included simple repeats long enough to harbor a single CTS, a more peculiar repeat type dominated this group. Namely, while a conventional 20-nucleotide GC-rich signature sequence was readily derived for the group as a whole, consistent with the CTCF-binding motif generated for the whole genome (Fig. 2b, c), a longer consensus, which is more in line with the span of the actual CTCF binding [67], showed that a duplication of a shorter binding signature (denoted CTS') is present in these repeats (Fig. 2d). Thus, while an individual repeat

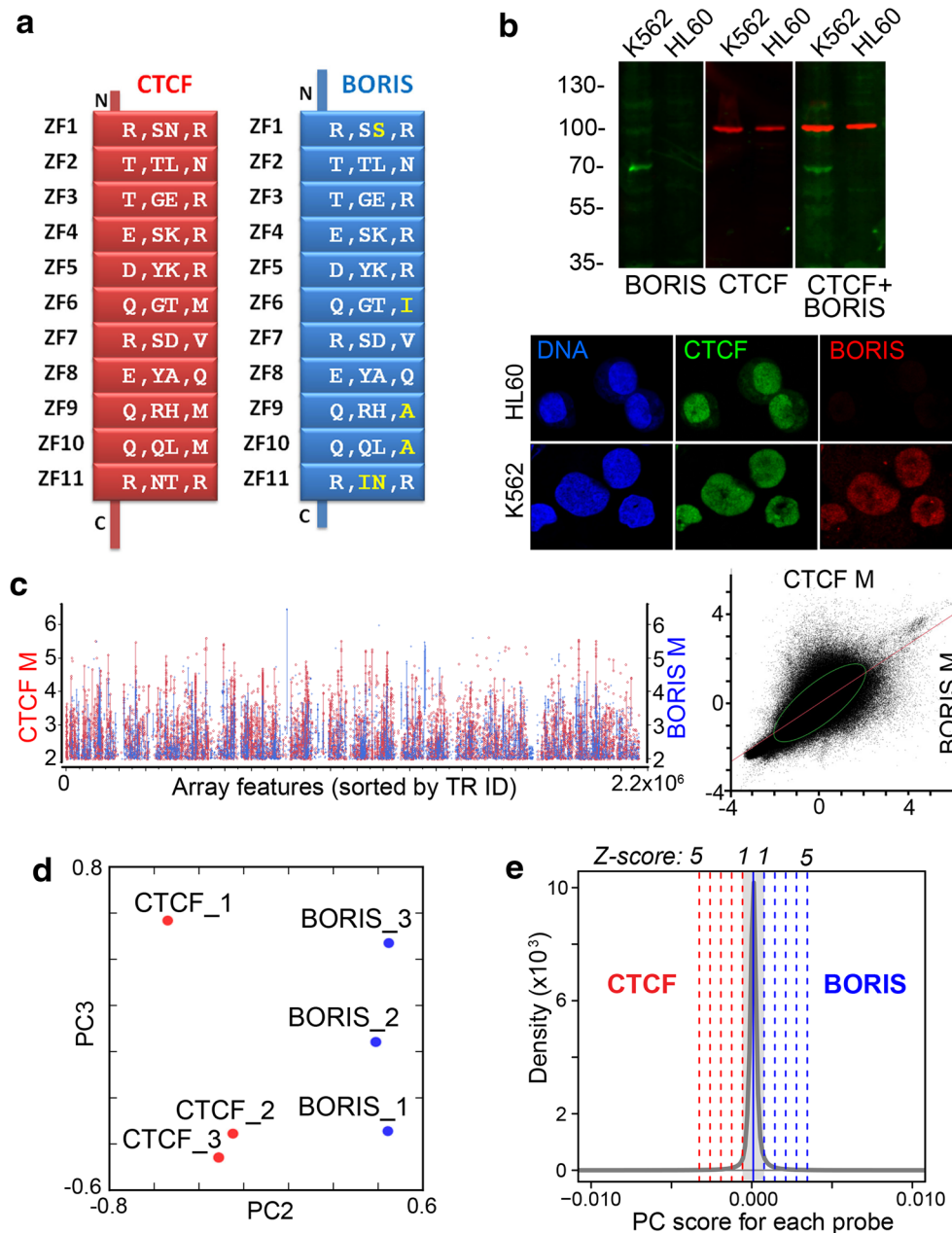
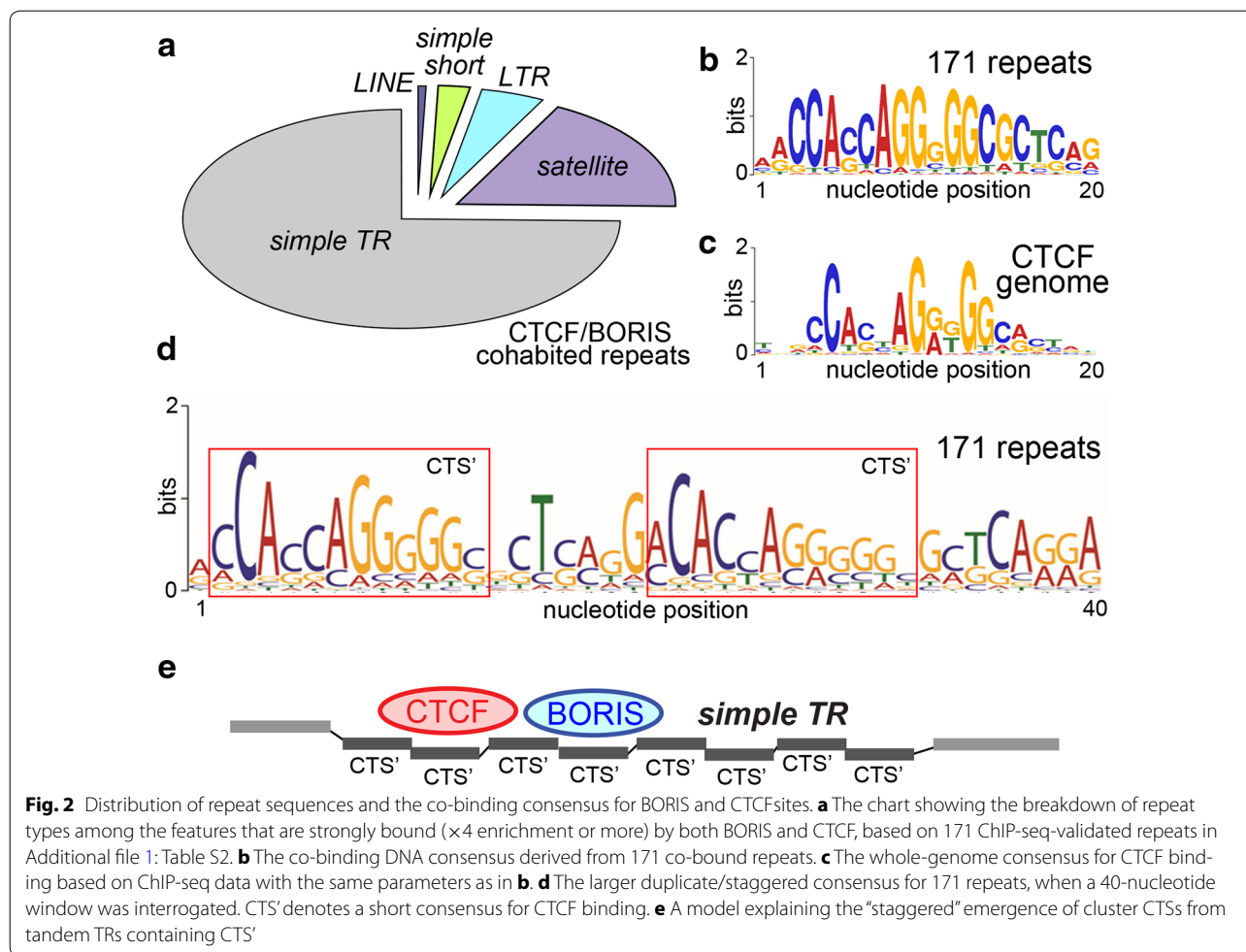


Fig. 1 BORIS expression in K562 establishes a definitive pattern of repeat binding. **a** A schematic of CTCF and BORIS proteins with the four amino acids residues essential for DNA recognition by each zinc finger (ZFs) showing. The minor differences, indicated in yellow, do not affect the DNA-binding specificity in vitro neither the consensus derived from the genome-wide binding study [55]. **b** LI-COR image of immunoblotting for BORIS and CTCF proteins in whole-cell protein extracts of K562 (BORIS positive) and HL60 (BORIS negative) cancer cell lines of myeloid origin. Below the immunofluorescent and DNA staining of the two corresponding cell lines. **c** The left panel shows the enrichment ratio (M) for CTCF and BORIS across all the tiles of the TR microarray. Dots represent microarray tiles enriched ≥ 4 by either CTCF (red) or BORIS (blue) with lines connecting different tiles belonging to the same repeat. SAM showed that 42,715 tiles were differentially occupied with FDR ranging from 0.103 to 0.245, with 0.75 correlations between CTCF and BORIS arrays. Both measures indicate that a minority of the repeats were differentially bound by the two proteins. The right panel shows the linear fit of BORIS M ratios by CTCF M ratios, with the fit line and 95 % bivariate normal ellipse displayed. **d** Principal component analysis of ChIP-chip data from K562. PCA was performed with singular value decomposition (SVD), and the first principal component describing the trend of the data was excluded from the analysis. PC2 (42 %) explains the difference between CTCF and BORIS experiments, and PC3 (16 %) explains variance between the replicates. **e** Smoothed histogram for probe loading kernel density (Y axis) estimate for all probes along the PC2 axis. The mean is indicated, as well as the number of standard deviations from the mean (Z-score). Most probes show no significant loading on PC2, and no significant difference between CTCF and BORIS. Only a fraction of probes show a significant contribution to the PC2 axis



unit does not enclose a bona fide cluster CTS, the tandem arrangement of this class sets a potentially multiple/staggered binding mode for CTCF and BORIS at these elements potentially generating a cluster site, if the tandem structure is long enough (Fig. 2e). Therefore, we can hypothesize that co-binding of CTCF and BORIS to the same site, as in this group of repeats, is facilitated when two binding regions are juxtaposed in cis, as happens in the rest of the genome [55]. The fact that multiple uncharacterized simple repeats were found in this class indicates that these elements should have a regulatory function in the epigenome mediated by dual binding by CTCF and BORIS.

Analysis of CTCF and BORIS co-binding at repeated DNA would have been incomplete without assessing the least characterized region of human epigenome—the chromatin of nucleolar organizer (NOR, or rDNA repeats). The bona fide human genomic rDNA has a very complex structure with multiple intervening sequences [68], and the NOR sequence from any human chromosome still remains to be determined. Therefore, human

rDNA was not represented at TRF database and was not present on our microarrays. While we did not validate rDNA binding by CTCF and BORIS in ChIP-chip, it is known that the repeat unit contains a strong hotspot for CTCF binding facilitating CTCF's interaction with PolII transcription machinery [69]. We used a "consensus" human rDNA repeat, as in [65], to align ChIP-seq reads and assess the potential differences between CTCF and BORIS binding (Additional file 2: Figure S2B). Comparing BORIS and CTCF binding showed that CTCF has a single binding site upstream of rDNA PolII promoter, consistent with published data in mice [70]. At the same time, BORIS appeared to have some enrichment at additional sites (Additional file 2: Figure S2A). These locations, however, corresponded to low-complexity regions (Additional file 3: Table S1), which were also present elsewhere in the genome. Unlike the established CTCF binding site, the two selected BORIS sequences that appeared to be enriched in ChIP-seq were not confirmed to bind BORIS by EMSA in vitro (Additional file 2: Figure S2C). Thus, one may assume that such sites likely represent an

artifact of short reads' alignment to tandemly repeated DNA, and the additional such sites were not tested. The presence of BORIS at the main Pol I regulatory site in rDNA, however, indicates that BORIS might be involved in ribosome biogenesis in cancer cells by virtue of co-regulating the rDNA transcription with CTCF.

CTCF-only enrichment is found in older repeat classes

The CTCF-only binding sites have a still unknown function in the genome, possibly unrelated to transcription [55]. PCA results in Fig. 1e enabled us to separate the CTCF-bound repeats that were refractory to BORIS intrusion (Fig. 3a). Thirty-eight individual CTCF-only repeats in this group were validated by ChIP-seq (Additional file 2: Table S2). This set includes major known types of repeats with long evolutionary history, while evolutionary young and simple TRs were largely absent. This agrees well with the studies, indicating that some CTCF-only binding sites in repeats are conserved in evolution [67]. Two examples of ChIP-seq analysis for repeats in this class, a TR of two Alu elements (Fig. 3b) and a run of divergent centromeric alpha-satellites (Fig. 3c), showed a robust enrichment by CTCF as compared to BORIS. As the enrichment of alpha-satellites by CTCF did not appear to be very strong, it is possible that a substantial fraction of alphoid elements in the K562 genome are not occupied by CTCF. Combined with the fact that CTCF binding does not appear to be correlating with CENP-B box presence (Fig. 3c), this may even indicate that only non-centromeric alpha-satellites are bound by CTCF. The absence of strong BORIS binding to this group of repeats agrees well with the underrepresentation of clustered CTS consensus in this repeat group (not shown).

A movable and evolutionary youngest class of TEs is specifically enriched in BORIS binding

The BORIS-only repeats, where BORIS binds without the equivalent presence of CTCF, are the most revealing with respect to BORIS biology in cancer cells, as they are directly involved in the transcriptional regulation of the non-repeated part of the genome [55]. Remarkably, in this group, the only 10 TRF classes that were validated fell within a single repeat type: the SVA family (Fig. 4a; Additional file 1: Table S2). The SVA repeats are a hominid-specific family, which is still currently mobile in the human genome owing to L1 activity [71, 72]. Overall, ChIP-seq analysis indicates that as much as 70 % of SVA elements could be occupied by BORIS in K562 (Fig. 4b). When this preference for SVA repeats was dissected for individual genomic repeat sequences, it became apparent that the enrichment by BORIS peaked in the central part of the element composed of the GC-rich VNTRs

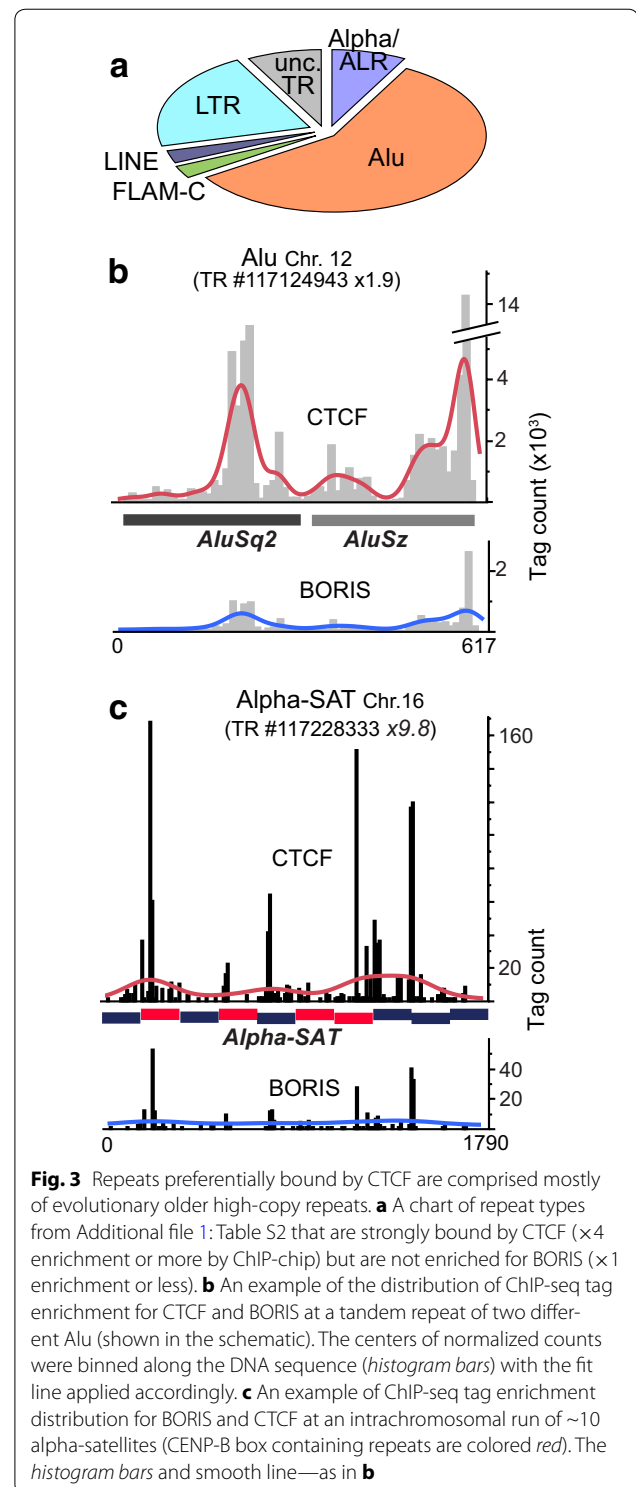
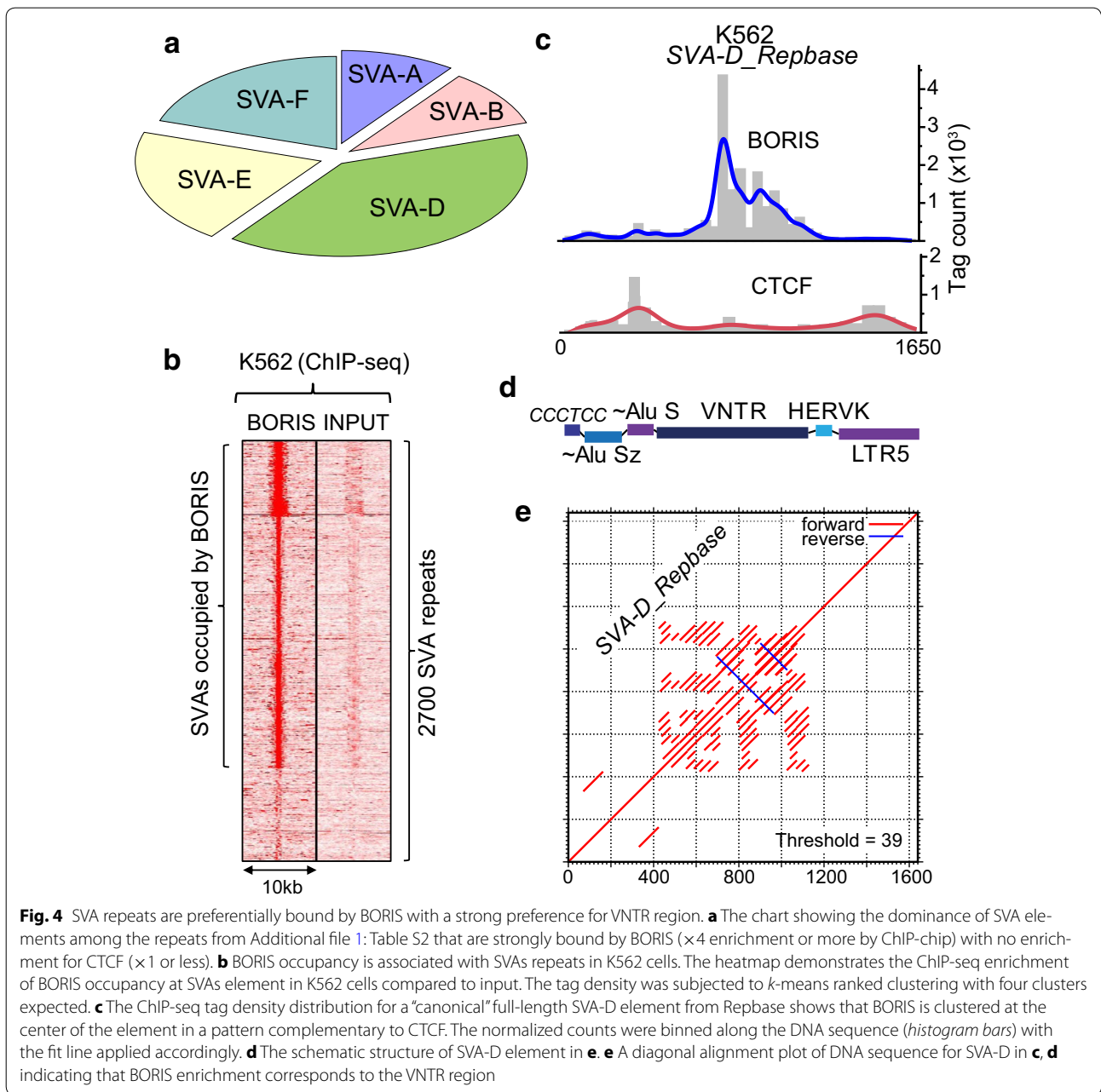


Fig. 3 Repeats preferentially bound by CTCF are comprised mostly of evolutionary older high-copy repeats. **a** A chart of repeat types from Additional file 1: Table S2 that are strongly bound by CTCF ($\times 4$ enrichment or more by ChIP-chip) but are not enriched for BORIS ($\times 1$ enrichment or less). **b** An example of the distribution of ChIP-seq tag enrichment for CTCF and BORIS at a tandem repeat of two different Alu (shown in the schematic). The centers of normalized counts were binned along the DNA sequence (*histogram bars*) with the fit line applied accordingly. **c** An example of ChIP-seq tag enrichment distribution for BORIS and CTCF at an intrachromosomal run of ~ 10 alpha-satellites (CENP-B box containing repeats are colored red). The *histogram bars* and smooth line—as in **b**

(Fig. 4c–e). VNTRs in SVA are GC-rich sequences with unknown molecular function. The patterns of CTCF and BORIS occupancy at SVA elements were distinct (Fig. 4c), unlike in other elements analyzed in Fig. 3. This might indicate the exceptional specialization of the



VNTRs for BORIS binding in cancer cells. In order to exclude the possibility that SVA enrichment by BORIS is a specific property of K562, myeloid cells, or the female epigenome in general, we conducted ChIP-seq analysis of an unrelated cancer cell line with aberrantly activated BORIS, Delta-47 cells [55]. Although the difference between BORIS and CTCF enrichment was not as dramatic as in K562, the preference of BORIS was evident (Additional file 4: Figure S1A), notwithstanding the lower level of BORIS in Delta-47 [55]. Considering that the SVA’s VNTRs are dynamic in number and composition

themselves [73], the finding of a global regulator BORIS bound to a mobile and extremely variable repeat class could be indicative of an additional germline-specific function of BORIS.

In order to map the locations of BORIS binding sites in SVA elements with higher precision, we designed nine probes corresponding together to a full-size SVA-D element (Fig. 5a) and analyzed them by EMSA with BORIS and CTCF proteins produced by in vitro translation. EMSA assay showed that the weak binding found in the AluS part can be attributed to a short unique sequence

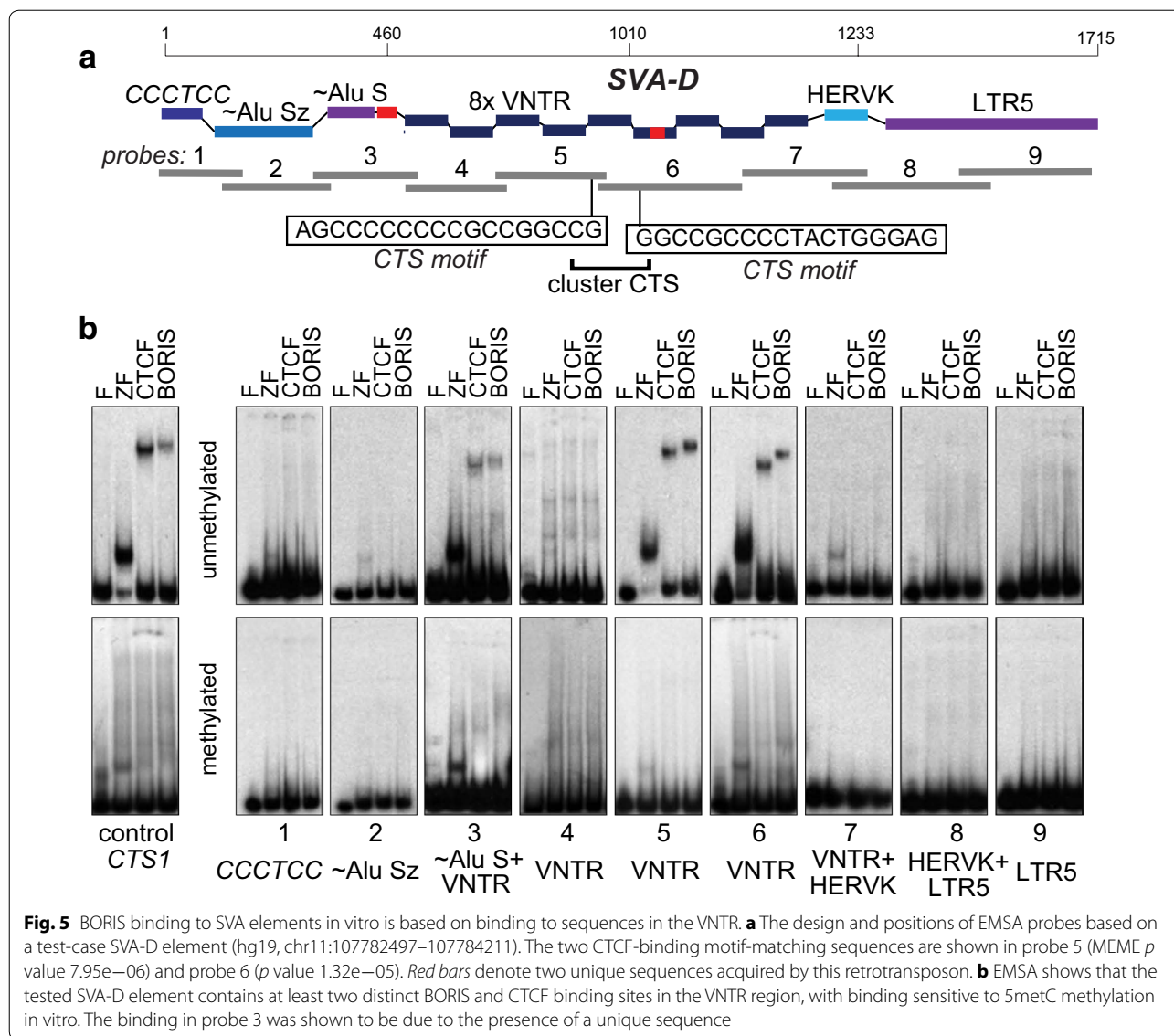


Fig. 5 BORIS binding to SVA elements in vitro is based on binding to sequences in the VNTR. **a** The design and positions of EMSA probes based on a test-case SVA-D element (hg19, chr11:107782497–107784211). The two CTCF-binding motif-matching sequences are shown in probe 5 (MEME p value $7.95e-06$) and probe 6 (p value $1.32e-05$). Red bars denote two unique sequences acquired by this retrotransposon. **b** EMSA shows that the tested SVA-D element contains at least two distinct BORIS and CTCF binding sites in the VNTR region, with binding sensitive to 5methylcytosine (5mC) methylation in vitro. The binding in probe 3 was shown to be due to the presence of a unique sequence

there (Fig. 5b). The central core of VNTR region, represented by two probes (5 and 6) in an EMSA, showed reproducible binding to both BORIS and CTCF proteins (Fig. 5b). Based on the EMSA data and CTCF motif analysis (Fig. 5b), these two VNTR sites juxtaposed to each other together form a cluster CTS, which is required for BORIS-only binding [55]. The 83-bp unique sequence embedded in the probe 6 in Fig. 5 was by itself unable to bind either protein (not shown). Not surprisingly, no discernible difference was detected between CTCF and BORIS in binding in vitro (Fig. 5b). This indicates that the BORIS' preference for SVA binding observed in chromatin (ChIP data) is likely determined by epigenetic factors. As CTCF is known to have both DNA methylation-sensitive and methylation-insensitive binding sites, we verified

whether BORIS is able to bind VNTRs when CpGs are methylated. EMSA analysis with methylated probes (Additional file 4: Figure S1B) showed that both CTCF and BORIS binding were abolished by full CpG methylation (Fig. 5b). This likely indicates that the preference of these sites for BORIS binding in chromatin, even if partially controlled by DNA methylation, must be fine-tuned with respect to specific CpGs methylation.

What could be BORIS activity at SVA elements? Our previous results on the genome-wide consequences of modulation of BORIS expression indicated that BORIS could serve as an activator as well as repressor [55]. The distinct preference of the aberrantly expressed BORIS for SVA elements may potentially indicate that BORIS has some regulatory activity at these elements in germline

and/or in cancer cells. As there is little doubt that SVAs mobilization is detrimental to genome stability, because they are under a strong repression in primates [73–76], a possible BORIS involvement in the regulation of SVA transcription must be biologically important. Indeed, the transcription is required for SVA transposition, and it could also have a regulatory role in the expression of neighboring genes.

BORIS acts as a transcriptional co-repressor of a significant proportion of SVAs in K562 cells

While the transcription unit of SVAs is not well characterized [76, 77], the Alu-derived sequences are the chief drivers of transposition in SVA [78]. Thus, SVAs contain sequences potentially transcribed by both RNA Pol III and Pol II, either of which can drive retrotransposition [79]. At the same time, based on structural considerations, it is unlikely that SVA elements are actually transcribed by Pol III [77]. We tested whether there was a difference in the occupancy of RNA Pol III factors at SVA elements between the publicly available ChIP-seq datasets for BORIS-positive K562 and BORIS-negative NHEK. Incidentally, we found no notable enrichment at any SVA elements for POLR3G, BDP1, BRF1, BRF2, or RPC155 (data not shown).

Next, we focused on the RNA Pol II transcription of SVAs and first took advantage of CAGE datasets available for K562 (BORIS positive) and NHEK (BORIS negative). The CAGE reads were aligned to the genome, and the extended areas corresponding to SVA elements were analyzed separately. However, the levels of SVA transcription were low, and SVA transcription in BORIS-positive K562 cells was mostly well correlated with the BORIS-negative NHEK cells (Pearson correlation 0.98). At the same time, RNA-seq data available for human testis suggest that some SVA elements could be highly expressed; however, the two full-length (FL) SVA elements with highest expression in human testis showed no ChIP-seq enrichment for BORIS at the VNTRs (Fig. 6a). The extension of analysis in Fig. 6a to 59 additional SVA elements with various degrees of BORIS occupancy showed only marginal levels of expression without any correlation with BORIS presence at the VNTRs (not shown). Thus, it is highly unlikely that BORIS bound to VNTRs serves as a transcription activator of SVA transcription in K562 cells.

At this point, one may hypothesize that the affinity of BORIS to VNTRs of SVA elements demonstrated in K562 is a reflection of its role in germline pertaining to these elements and that this role is likely a repressive one. Indeed, we recently showed that despite BORIS previously perceived as an activator, BORIS upregulation was linked to the repression of some genes and,

vice versa, BORIS downregulation has resulted in some gene being activated [55]. Therefore, we investigated the K562 cells with downregulated BORIS. As SVA elements might be rapidly repressed by some other mechanism in the absence of BORIS, we could not rely on BORIS KO data [55], as the points of comparison there were separated by a long period of time. Instead, we experimented with the downregulation of BORIS expression in K562 cells for a short period of time using inducible shRNA. This approach enabled us to assess immediate downstream effects of BORIS downregulation. We constructed K562 cell lines with two alternative inducible anti-BORIS shRNA constructs stably integrated into the genome and conducted RNA-seq experiments after BORIS KD for 48 h. Neither the degree of BORIS depletion nor the time span of the experiment was sufficient to induce the differentiation, as was described for BORIS KO [55]. While genome-wide expression of genes responding to BORIS KD was almost evenly divided between up- and downregulation of transcription (data not shown), SVA elements longer than 1 kb were notably activated (Fig. 6d). In order to address whether any SVA were actually downregulated upon BORIS KD, we isolated the subclass of SVA elements that were already expressed in K562 and compared their expression to BORIS KD cells. As shown in Additional file 5: Figure S3A, the 70 SVA elements that were expressed did not significantly change their expression upon the downregulation of BORIS.

In order to understand better the nature of SVA activation, we treated control K562 cells with 5-AzaCyd (5-Aza-2'-deoxycytidine), an inhibitor of DNA methylation [27, 80–82], and DZNep (3-deazaneplanocin A), which indirectly suppresses EZH2 that catalyzes histone H3 lysine 27 methylation [83, 84]. Both drugs result in the removal of inhibitory epigenetic marks from DNA and chromatin, respectively. RNA-seq analysis of K562 cells treated with these DNA methylation or H3K27me3 inhibitors indicated that SVA elements that were already active were upregulated slightly (Additional file 5: Figure S3B, S3C), while the group as a whole was preferentially activated. The 5-AzaCyd effect was similar to BORIS KD, and the DZNep effect was more pronounced (Fig. 6d). Thus, we next asked whether these treatments could be preferentially affecting the same subset of SVA elements as BORIS KD or a distinct one. Using the DZNep treatment as an example, Fig. 6e, f, we showed that BORIS KD largely acted concordantly with DZNep (correlation 0.77) to activate SVA transcription of the elements that were silent in the control. It was also evident that the BORIS KD-dependent activation was not specific to any particular subclass of SVA repeats (Fig. 6g), indicative of a common pathway.

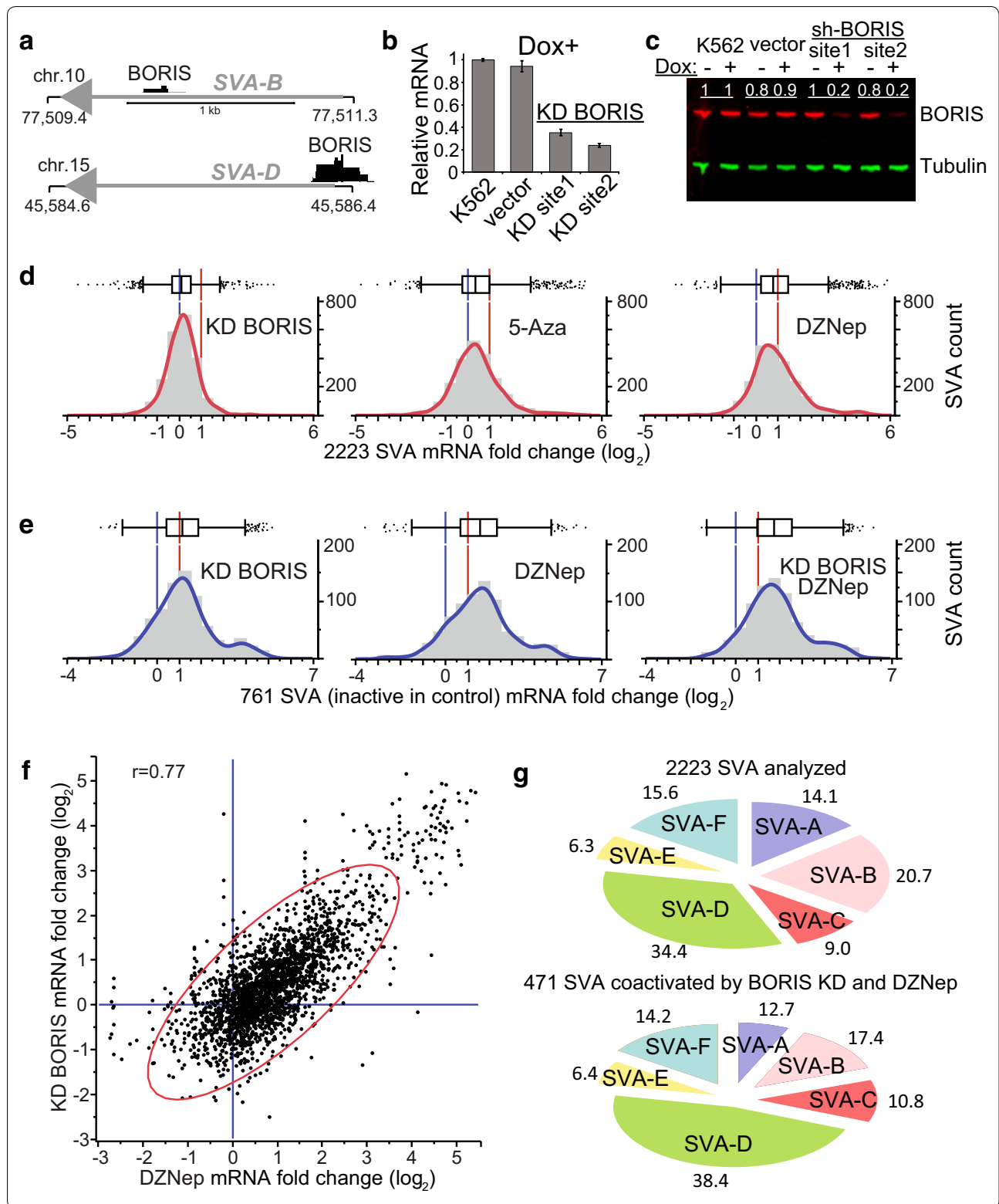


Fig. 6 Downregulation of BORIS and epigenetic remodeling show concordant activation of SVA transcription in K562 cells. **a** The two SVA elements with highest expression in testis with the position of BORIS ChIP-seq peaks in K562. The absence of strong BORIS binding indicates that BORIS is unlikely to act as SVA activator in testis. Genomic coordinates are in kb. The SVA-D shown is intergenic, while the SVA-B is antisense intronic. **b** RT-qPCR shows the downregulation of BORIS in K562 clones with stable integration of Tet-On inducible anti-BORIS shRNA constructs (site 1 and site 2), 48 h after shRNA induction. Un-infected K562 cells and a clone with the integrated empty vector were controls. Only the experiments in the presence of doxycycline (Dox+) are shown. **c** Immunoblotting with anti-BORIS mAbs demonstrates fourfold–fivefold depletion of BORIS protein in shRNA clones. The quantification of relative BORIS amount (*white numbers*) was performed using LiCor software and alpha-tubulin as a reference. **d** RNA-seq analysis of differential expression of 2223 SVA elements longer than 1 Kb mapped in the human genome versus K562 infected with the empty vector (SVAs that were constitutively silent were not included). Shown are the distributions of ratios of RNA-seq difference in: BORIS KD K562 cells (paired two-tailed *t* test *p* value <0.001), K562 treated with 5-AzadCyd (5Aza) (paired two-tailed *t* test *p* value <0.001), and K562 treated with DZNep (paired two-tailed *t* test *p* value = 0.001). The RNA-seq reads enrichments for SVA elements were normalized to the total number of reads in each individual experiment. The mean and standard deviation diagrams are shown on *top* of each graph. The graphs demonstrate the overall increase in the shift toward higher SVA expression from BORIS KD to 5Aza and especially DZNep treatments. *Vertical blue lines* correspond to the unchanged expression over control, *red*—to twofold increase. **e** The RNA-seq analysis of BORIS KD, DZNep treatment, and the combination of both on the transcription of SVA elements that were apparently silent in control experiments (i.e., <10 normalized counts with over twofold increase by any treatment) shows a reproducible compound effect of BORIS KD and DZNep treatment on SVA activation (for the latter, the whole 2223 SVA sample's paired two-tailed *t* test *p* value <0.001). The corresponding means of the distributions with standard errors of the mean: KD 1.03 ± 0.03 , DZNep 1.37 ± 0.03 , KD and DZNep 1.55 ± 0.02 . **f** *Dot-plot* of RNA-seq normalized counts of BORIS KD versus DZNep treatment of K562, expressed as fold enrichment over the empty vector control. Only the SVA elements that were silent in the control are included. The *blue lines* correspond to cutoffs with no change in expression. **g** SVA elements that show concordant activation by BORIS KD and DZNep treatment do not belong to a preferential SVA class. The *pie diagrams* show the breakdown of SVA classes among all 2223 elements included in the analysis and among 471 elements co-activated by both treatments

A distinct type of BORIS function at the SVA-F1 TEs

The prevalent repressive role of BORIS on SVAs does not exclude the possibility that under certain conditions it could actually serve as an SVA activator. One such case could be the *MAST2*/SVA-F exon trap [85–87]. The capturing of *MAST2* sequence by SVA-F resulted in the formation of a novel family (SVA-F1), represented by 81 members in the hg19 human genome assembly [85, 88]. The 5' flanking region of SVA-F1 family is the result of a fusion between the first exon of *MAST2*, a gene expressed in testes, with the SVA-F repeat. Thus, it is conceivable that in testis BORIS acts as an activator of SVA-F1. This is possible as the binding of BORIS to SVA-A through SVA-F is within the VNTR region, but for SVA-F1 BORIS preferentially binds within the 5' flanking region of the SVAs, upstream of the hexamer repeat region (Fig. 7a–c). It is worth noting that the first exon of *MAST2* is not just occupied by BORIS in K562 cells but is also aberrantly expressed in cancer cells together with BORIS expression (Additional file 6: Figure S4A). Thus, BORIS binding outside of SVA elements may serve as an external promoter for SVA-F1 expression. The numbers of nucleotides captured from the *MAST2* exon by SVA-F1 vary from 36 to 382, with potentially four BORIS binding sites incorporated into 382 bp-promoter sequence (Additional file 6: Figure S4B). That may create a possibility for multiple TSSs starting from any of four BORIS binding sites. It may also explain the presence of *MAST2* SVA-F1 sequences of varying length. Indeed, the common feature of nearly all SVA-F1 transduced sequences is the presence of at least one BORIS binding site. In agreement with multiple BORIS binding sites in the transduced sequence the BORIS occupancy significantly correlates

with the length of transduced sequence (Additional file 6: Figure S4B). While SVA-F1 sequences are strongly expressed in testis, they remain methylated in other instances of substantial hypomethylation of the genome [89]. Their expression is also quite low in BORIS-positive cell lines (Fig. 7d). Neither did the KO of BORIS in K562 cells change the overall expression of SVA-F1 (Fig. 7e). Nevertheless, the ectopic BORIS expression in BORIS-negative cells appears to have a slight activating effect on SVA-F1 (Fig. 7f). We also analyzed the putative promoter-trapping events similar to the *MAST2* case throughout human genome and identified several putative locations of such occurrences. For example, we found that *NDUFV2*, *FDX1*, *PHKA1*, *WDR33*, *RHOT1*, *ZNF488*, *ZNF487*, *PHLPP2*, *TOM1L2*, *ARL4A*, and *MPPE1* promoters were trapped by SVA repeats and used for SVA expression in K562 cells (Additional file 7: Figure S5; Additional file 8: Table S3). One of the common features of all these promoters is the presence of BORIS binding sites inside the trapped sequences, occupied by BORIS in K562 cells and transcribed in BORIS-positive cells (Additional file 6: Figure S4; Additional file 7: Figure S5). Based on such data, one would be compelled to conclude that the capture of BORIS binding sites by SVAs is beneficial for their transcription. The trapping of BORIS binding sites within the promoter region of SVA repeats may also be indicative of an existing pathway for non-random SVA integration.

In conclusion, it appears that BORIS acts as a co-repressor of SVA transcription in K562 cells, alongside DNA methylation and heterochromatinization. It is therefore likely that BORIS plays a similar role in the germline, with the exception of promoter-trapping events.

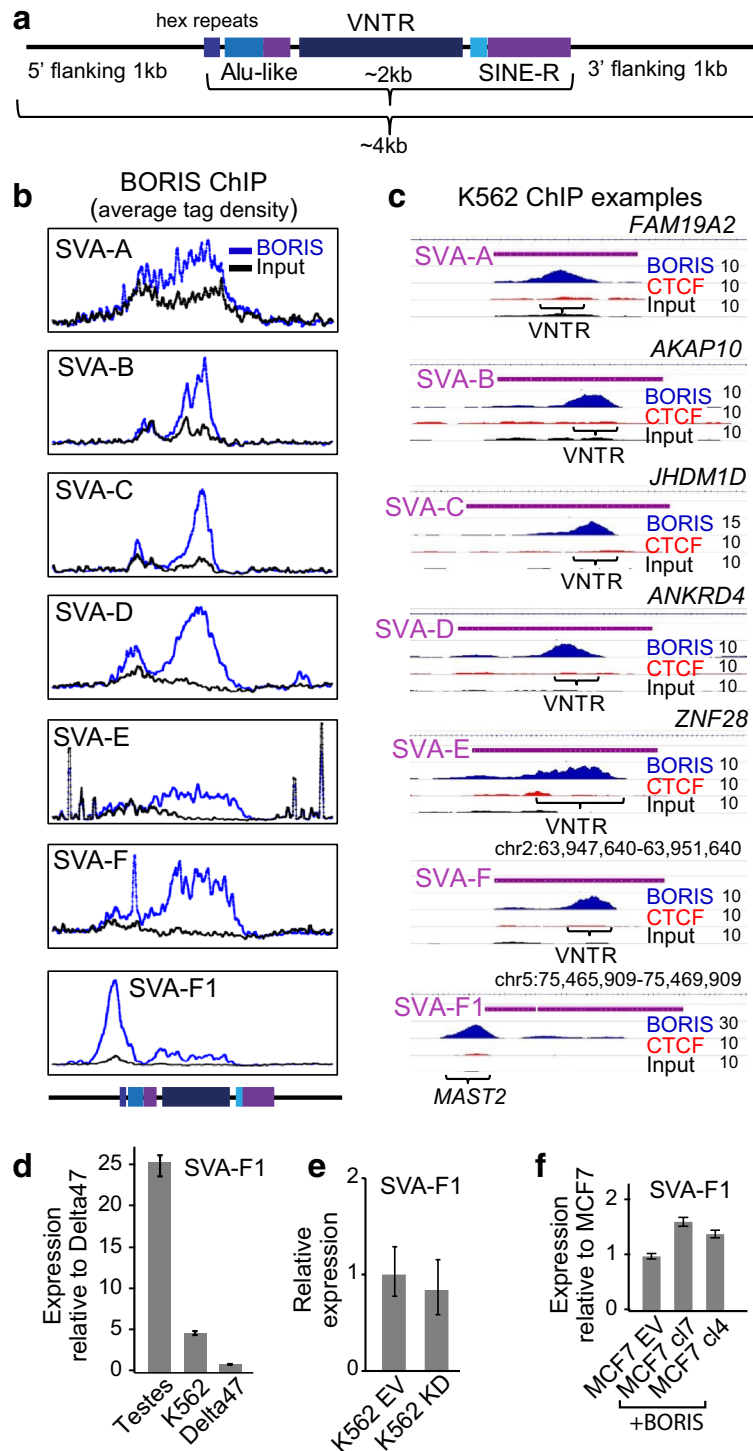


Fig. 7 BORIS is enriched at the 5'-transduced sequence at the SVA-F1 family repeats. **a** A schematic of tested ~4 kb sequence encompassing SVAs elements. The sequences were used to plot the average tag density of BORIS ChIP-seq in K562 cells. **b** BORIS is predominantly associated with VNTR repeats of SVAs, with the exception of SVA-F1 where BORIS is bound at the 5'-transduced sequence. The average tag density (tags/ten million) is shown for BORIS versus the input of ChIP-Seq data in K562 cells. **c** The individual genomic examples of BORIS and CTCF occupancy corresponding to **b**. Data were normalized to the number of mapped reads and the number of SVA elements. **d** SVA-F1 expression is upregulated in testis. qRT-PCR data on total testis mRNA and two BORIS-positive cancer cell lines. **e** KD of BORIS in K562 does not change the SVA-F1 expression. qRT-PCR data of BORIS KD by shRNA relatively to the control empty vector (EV); SVA-F1 primers were from [88]. **f** The induced BORIS expression in MCF7 cells slightly upregulates SVA-F1 expression. qRT-PCR for two clones expressing ectopic BORIS [55] is compared to MCF7 transfected with the vector only (EV)

These findings indicate a potential biological role of BORIS as a regulator of active TEs in human genome.

Discussion

The “explosive” chromosome instability is confirmed to be one of the defining features of cancer genome [90, 91]. This notion has sparked multiple attempts to find either a unifying mechanism or a set of concurrent mechanisms for this process [92, 93]. The early onset of chromosome instability in cancer and pre-cancer cells strongly indicates the epigenetic roots of the destabilization. In this context, the roles of chromatin states of genomic repeats in cancer are of significant interest because they directly bridge the epigenetic landscape with a potential to destabilize genome via transposition and/or recombination. TEs that can pose a danger to genome integrity tend to be silenced for recombination and retrotransposition by epigenetic mechanisms [17, 73, 94]. Here, we found evidence of BORIS involvement in the co-regulation of TEs. The established role of BORIS as a transcriptional regulator in cancer [55, 95] and as activator of testis-specific genes [70, 96, 97] might also be applicable to the states of genomic repeats in cancer cells. Nevertheless, the role of BORIS with respect to genomic repeats was previously totally unknown, despite the significant recent progress in understanding the transposition as the primary venue of genome evolution pertaining to the distribution of CTCF binding sites [67].

In this study, we established that BORIS, upon its activation at a relatively high level in cancer cells, has a substantial capacity to occupy the same sites in the repeated elements as CTCF (Fig. 1e). We can presume, with a high level of certainty, that it is a manifestation of the BORIS' co-functions with CTCF in the normal germline [55, 70]. While co-binding is generally expected due to the DNA-binding properties of the two proteins *in vitro*, the recent discovery of cluster sites being a prerequisite for CTCF and BORIS co-binding or binding of BORIS alone [55] suggests that a significant fraction of such repeats have cluster site configuration. Indeed, the assessment of DNA consensus characteristic for BORIS and CTCF co-bound repeat sites (Fig. 2c) showed no significant deviation from the basic unit of CTCF consensus derived from the genome-wide binding studies (Fig. 2b), but revealed the presence of a staggered arrangement (Fig. 2d), which potentially enables such TR locations to become super-cluster sites with ample co-binding capacity. The characterization of repeats that are co-occupied by CTCF and BORIS showed that the bulk of co-binding seems to be associated with the low-copy simple TRs (Fig. 2a). These elements have a relatively narrow length distribution, most are longer than 50 nt, indicating that they are under selection, possibly by the requirement to bind CTCF or BORIS. While expansion of short TRs is known

to cause disease in a number of studied cases [98, 99], their genome-wide biological role is obscure. Thus, it is likely that BORIS and CTCF co-binding there uncovered a putative regulatory role for these elements in germline and/or cancer transcription.

The few repeat types that show a significant bias toward CTCF-only binding are rather enigmatic, as the function of CTCF-only sites genome wide is not well characterized [55]. The most notable case here is the centromeric repeats, where recombination is highly undesirable [100], but the transcription was nevertheless found to be of paramount importance for normal kinetochore formation [101]. While CTCF's binding at alpha-satellites and its involvement in centromeric transcription were not studied, the interaction between CTCF and some centromeric proteins has been invoked at ectopic sites [102].

The most distinctive result generated by this study is the high preference of SVA repeats for BORIS binding, as compared to binding by CTCF in K562 (Fig. 4). Unfortunately, in the absence of ChIP data for BORIS from human testis one cannot be absolutely sure that it is also the situation in normal testis. The functions for SVA that are described so far are attributed to the disruption/features of insertion sites rather than to the transcription originating within the insertion [103, 104]; yet the finding of BORIS binding hints at the regulatory role of SVA VNTRs themselves. The presence of several BORIS binding sites within the VNTR repeats (Figs. 4c, f, 5), which are actually required for SVA transposition [78], indicates that the BORIS protein and SVA elements may have even undergone co-evolution, as has been recently suggested for other ZF proteins [73]. Thus, one may expect the SVA elements to play a notable regulatory role in germline development and genome evolution in primates. In that regard, the recent studies on gibbon genome [2, 105] provided some invaluable insight into the new level of plasticity that SVA-like elements LAVA infused into primate genomes. At present, one cannot conclude whether SVA TEs merely represent a genetic load or actually have a physiological role in germline. Despite human SVAs being associated with at least some chromosomal breaks [106], we could probably exclude the direct contribution of SVA elements into the meiotic recombination, as DSB maps of human meiosis [107] did not correspond to SVA locations (not shown).

By applying RNA-seq analyses to the K562 cells, we found a strong evidence of a substantial fraction of SVA elements being transcriptionally activated upon BORIS KD (Fig. 6d–f). This was a strong indication that BORIS acted as a repressor of SVA transcription for that repeat group. This conclusion is further reinforced by the finding that this repressive activity is additive with DNA methylation and with the formation of repressive chromatin structure (Fig. 6e, f). Therefore, we could conclude

that BORIS participates in the repression of SVA elements that are located in the heterochromatin-like regions of epigenome. This BORIS-mediated tier of SVA repression could have an exceptional significance in male germline, where the rounds of DNA demethylation [108] could potentially open SVA retrotransposons for a transient activation leading to germline mutations, as it has been found in pluripotent cells [109].

The addition of BORIS to cancer cells' chromatin constitutes a potent epimutation, as it could introduce a substantial change into CTCF's functions [36]. Some of these changes were recently documented, particularly with respect of recapitulating the germline pattern of gene regulation [55]. With respect to the genomic repeats, the associated rewiring of epigenetic regulatory network, which is normally embodied by CTCF alone in somatic cells, may greatly alter the functional role of inserted repeats themselves, e.g., their expression and transposition, as well as their propensity to regulate neighboring genes and chromatin domains.

Conclusions

As a result of this study, by employing ChIP-chip and ChIP-seq approaches, we characterized CTCF and BORIS binding patterns of genomic repeat binding upon aberrant BORIS expression in the K562 cancer cell line, which is dependent on BORIS for proliferation. This study showed that, while CTCF-only enrichment is found in most known repeat classes, BORIS and CTCF bind together predominately to the uncharacterized simple TRs, which likely form compound cluster binding sites. We discovered that the SVA elements, a presently active family of TEs in human genome with a strong mutagenic potential and a role in transcription regulation, are specifically enriched in BORIS, with binding concentrated at the VNTR region. Furthermore, RNA-seq analysis of BORIS KD in K562 showed that BORIS acts to repress multiple SVA, alongside the transcriptionally repressive histone modification and DNA methylation. These findings uncovered a novel function of BORIS in controlling the levels of TE transcription in cancer cells and likely in the germline.

Methods

Cell culture, transfection, and lentiviral infection

K562, Delta-47, and HL60 cell lines were grown in IMDM (Hyclone) supplemented with 10 or 20 % Tet-approved-FBS. HEK293T/17 cell line was grown in DMEM (Hyclone) supplemented with 10 % FBS. Transfection was done according to manufacturer's instructions using X-tremeGENE 9 DNA Transfection Reagent (Roche). To package lentivirus, HEK293T/17 cells were cotransfected with the vector Tet-pLKO-Neo (Addgene) or anti-BORIS shRNA derivatives and two packaged

plasmids psPAX2 and Pmd2.G. Lentivirus stocks were collected 72 h post-transfection and used to infect K562 at 40–50 % confluence using 500 μ l lentivirus stock and 8 μ g/ml polybrene (Sigma). The media were then changed 12 h after infection to include 600 μ g/ml G418, and the cells were selected for G418 resistance for at least 4 weeks. The resistant clones were selected in 96-well plates and analyzed by RT-qPCR and immunoblotting. The stable clones were induced by 200 ng/ml doxycycline to activate the Tet-On promoter.

The tiling repeat microarray

The design for this custom array [65] was conducted at Roche/Nimblegen using tiling approach. As a source for the design, we used a catalogue of human TRs generated by TR finder [110, 111]. The version of TRF algorithm used for the design of the array generated 947,696 distinct repeat instances based on the human genome. The tentative estimate of redundancy conducted by applying the most stringent versions of TRF suggests that the repeat dataset had about 40 % sequence redundancy. The repeats were broken into 50-base tiles using the following rules: Tiles were picked based on the predicted hybridization normalization; when the repeat was shorter than 50 nucleotides, it was extended in tandem fashion. Our tiling approach has generated some additional redundancy within tiles themselves because long homogeneous repeats produced a number of identical tiles. The redundancies within the array did not interfere with microarray data analysis, as the primary hybridization signal was recorded for each tile independently of any other. The final array design contained 2,166,672 features, including two control sets: 29,161 random sequence tiles and 181 tiles from the rDNA locus of *Saccharomyces cerevisiae*.

ChIP-chip and ChIP-seq

For the ChIP-chip and ChIP-seq, anti-CTCF and anti-BORIS ChIP were conducted from at least 50 million cells growing asynchronously. ChIP-seq preparation and analysis were done essentially as described in [55]. The specificity of ChIP reactions was validated by qPCR for known targets: the *TSP50* and *CST* promoters for BORIS, and the MYC promoter sites for CTCF as in [96, 97].

For ChIP, cells growing asynchronously were cross-linked (10 min, 1 % formaldehyde, 23 °C) quenched for 10 min by 200 mM glycine, washed three times with PBS, and then resuspended in chromatin buffer (150 mM NaCl, 1 % Triton X100, 0.1 % SDS, 20 mM Tris-HCl pH8.0, and 2 mM EDTA). DNA was sheared using Covaris S220, so that most fragments were in the 300- to 500-bp range. Chromatin was immunoprecipitated overnight with magnetic beads (DiaMag, Diagenode, Inc.) loaded with anti-CTCF or anti-BORIS antibodies as described in [55]. The

immunoprecipitate was washed, cross-links reversed, protein component was digested with proteinase K, and DNA was extracted using phenol/chloroform/isoamyl alcohol. DNA concentration was measured by Qubit (Life Technologies) and/or Nanodrop (Thermo Scientific) fluorimeters. For ChIP-chip, the immunoprecipitated DNA was amplified using the Phi29 strand-displacement procedure (GE Bioscience) following the concatemerization of precipitated DNA fragments via ligation to double-strand adaptors containing BamHI overhangs and internal SapI sites. Both amplified and non-amplified samples showed essentially the same relative enrichment for known sites of CTCF and BORIS binding. Following the amplification, adapters were removed by SapI digestion and agarose gel purification. Input DNA was used as a hybridization reference for the hybridization of amplified ChIP DNA to a set of custom TR arrays (Roche-Nimblegen). Raw intensities for each channel were centered against the mean of control features set, including random oligonucleotides and yeast rDNA. Then, Lowess smoothing was applied to two-channel data to generate corrected M values that were used in subsequent analyses. The Lowess normalization, SAM, and PCA calculations were done using publicly available R scripts. For downstream analysis of ChIP-seq data, the Illumina reads (50 bp) were aligned to human repeat subgenome generated by TRF [111] using BLAT [112] (allowing 95 % identity) and/or Bowtie [113] (with parameters `-v 2 --best --strata --tryhard`). seqMINER [114] was used to analyze and plot CAGE expression data from published datasets. Motif Elicitation (MEME) software [115] was used to derive consensus sequences from genomic repeats with parameters (`-mod oops -revcomp -w 20`) to identify motifs on both DNA strands.

Analysis of public high-throughput genomic data

ENCODE/RIKEN data (GSE34448) for K562 and NHEK cell lines were used in this study. The DSB maps of human meiosis were derived from [107].

Immunoblotting

Protein extracts were prepared by lysing cells SDS-PAGE sample buffer after washing with PBS supplemented with 1× protease inhibitor cocktail (Roche Applied Science). Protein samples were separated by SDS-PAGE, transferred to a PVDF membrane, and incubated with the appropriate primary antibodies, followed by detection using LiCor secondary antibodies fused to fluorochromes. Photoluminescent images were captured by scanning and processed for quantification using LiCor workstation.

Immunofluorescent cell staining

K562 and HL60 cells were spun down in Cytospin centrifuge (Thermo Scientific) onto poly-Lysine-coated

coverslips and fixed with 4 % paraformaldehyde for 10 min, followed by cold methanol for 10 min. Cells were permeabilized with 0.1 % Triton X-100/PBS for 10 min and then blocked with BSA for 30 min, after which they were incubated with primary antibodies. After washes, the anti-rabbit or anti-mouse secondary antibodies conjugated to either Alexa Fluor 647 or Alexa Fluor 488 were applied. Cells were mounted for microscopy in mounting media containing DAPI and images captured using either confocal (Zeiss) or wide-field (Olympus) inverted microscopes.

Electrophoretic mobility shift assay (EMSA)

To map CTCF and BORIS binding sites in SVA repeats, the SVA subfamily D repeat (chr11: 107,782,495–107,784,189, GRCh37/hg19) was covered with nine overlapping DNA probes either amplified by PCR or synthesized as oligonucleotides (Additional file 3: Table S1). PCR amplified products were cloned into the pCR2.1 TOPO vector (Invitrogen), and the sequence was confirmed by DNA sequencing. DNA fragments were labeled with [γ -³²P] ATP at the 5' ends by T4 polynucleotide kinase per Invitrogen protocol. Labeled DNA fragments were gel purified, and equal amount of each fragment was used for EMSAs. FL human CTCF, 11ZF domain of CTCF, and FL human BORIS were synthesized from pCITE expression vectors (EMD Millipore), using the reticulocyte lysate-coupled in vitro transcription-translation system (TNT, Promega). Binding reactions for EMSA were for 1 h at 23 °C with 4 μ l of in vitro synthesized DNA-binding proteins in binding buffer [25 mM HEPES pH7.6, 100 mM KCl, 2 mM MgCl₂, 10 % glycerol, 0.5 μ g poly(dIdC) × poly(dIdC)]. DNA-protein complexes were resolved on 5 % non-denaturing polyacrylamide gels in 0.5× Tris-borate-EDTA buffer. *Gal3ST1* promoter fragment was used in EMSA as a positive control for both CTCF and BORIS binding [97]. To test methylation sensitivity of protein binding, all labeled probes used in EMSA were methylated using SssI methyltransferase (New England BioLabs) by the following protocol: 200 ng of each oligonucleotide was combined with 2.7 μ l of NEBuffer 2, 3 μ l (12 U) of SssI methylase and 1 μ l of *S*-adenosylmethionine (32 mM). After 3 h of incubation at 37 °C, 0.5 μ l of NEBuffer 2, 3 μ l (12 U) of SssI methylase, and 1 μ l of *S*-adenosylmethionine (32 mM) were added, and the reaction incubated for an additional 3 h at 37 °C. The completion of methylation was assessed by digesting them with the methylation-sensitive enzyme AciI (Additional file 2: Figure S2B).

RT-PCR and quantitative PCR

Total RNA was prepared using Trizol (Invitrogen). cDNA was prepared using the Primescript™ RT Reagent Kit

with genomic DNA Eraser (perfect real time) (TaKaRa) according to the manufacturer's protocol. Quantitative PCR (qPCR) was performed using SYBR Premix Ex Taq™ (TaKaRa) and the Mx30005P QPCR System (Agilent).

RNA-seq analysis

For the RNA-seq experiments, inducible BORIS knock down (KD) and control cell lines were created by infecting K562 cells with 3 different Tet-on lentivirus constructs: empty vector pLKO-Tet-ON-neo [116], and two alternative anti-BORIS shRNA constructs. Several stable clones of each infected cell line were selected using 600 µg/ml G418. BORIS KD vectors were constructed to express the following shRNA templates: GGAAATACCACGATGCAAATT (Site 1) and GGTGTGAAATGCTCCTCAACA (Site 2). For lentivirus vectors construction, the annealed oligonucleotides were inserted into the pLKO-Tet-On-neo vector between AgeI and EcoRI restriction sites. After 72-h induction by doxycycline, BORIS mRNA was reproducibly showing 2.5-fold to threefold reduction, while BORIS protein levels were robustly decreased over fivefold (Fig. 6c, d). For RNA analysis, these K562-inducible stable shRNA cells were plated in 10-cm plates at 40–50 % confluence in DMEM media and left to grow in the presence of doxycycline (200 ng/ml) for 96 h. For the 5-aza-deoxycytidine and DZNep experiments, cells were identically pretreated with doxycycline, harvested, and re-plated at 50–60 % confluence to grow 48 h in the presence of either 500 nM 5-aza-2'-deoxycytidine, 1 µM DZNep or DMSO. The degree of genomic DNA demethylation was assessed using DNA IP with anti-5-methylcytosine mAb MABE146, clone 33D3 (EMD Millipore), and qPCR against known targets. The effectiveness of DZNep treatment was assessed by immunoblotting against the EZH2 protein with D2C9 rabbit mAb (Cell Signaling Technology). The cells were then collected, frozen, and outsourced for Illumina sequencing to RiboBio (Guangzhou). The amount of RNA submitted for each individual run was on average 85 µg (Nanodrop). The quality of RNA was assessed by the Agilent 2200 TapeStation. About 20 million reads were obtained for each individual experiment. Four biological replicates were produced and analyzed for each set of experimental conditions. The results of all RNA-seq experiments were analyzed for consistency and reproducibility using Cufflinks 2.0.0 [117] following reads alignment to the human reference genome (hg38) using TopHat2, with the default parameter setting. Upon that validation, for SVA alignment to RNA-seq data, a sub-genome file of 2223 SVA elements was assembled from elements mapped in hg38 that were longer than 1 kb, i.e., to ensure that VNTRs were included. The SVA elements were aligned to RNA-seq reads with Bowtie (-v0), and read counts per each element were normalized according to total read numbers in each experiment. Then,

fold-enrichment ratios relative to the averaged normalized reads in the empty vector experiments were calculated.

Additional files

Additional file 1: Table S2. Classes of repeats co-occupied or differentially occupied by CTCF and BORIS.

Additional file 2: Figure S2. BORIS binds at the same regulatory site as CTCF in rDNA. (A) The distribution of Chip-seq tags across the consensus rDNA repeat is shown for the input, CTCF ChIP-seq and BORIS ChIP-seq. The sites chosen for EMSA are indicated with brackets. (B) The corresponding structure of "canonical" rDNA repeat. The long arrow corresponds to the Pol I transcript; the short arrow—noncoding RNA; NTS—non-transcribed spacers. (C) EMSA of the chosen rDNA sites confirming that the known CTCF site in PolI promoter is co-occupied by CTCF in BORIS, while sampling of BORIS-only putative sites shows that there is no BORIS binding to these sites in vitro. (D) The assessment of CpG methylation for probe #3 used in (C) using the diagnostic digestion with *Aci* I endonuclease, before and after methylation.

Additional file 3: Table S1. EMSA oligonucleotides.

Additional file 4: Figure S1. Extended analyses of SVA-D. (A) BORIS binding peaks at SVA VNTRs in Delta-47 cells. The ChIP-seq tag density distribution for the full-length SVA-D element from Repbase indicates that BORIS retains preference for VNTR region event in this cell line completely unrelated to K562 and with a substantially lower BORIS expression level. The normalized counts were binned along the DNA sequence (histogram bars) with the smoothing line added. (B) The assessment if CpG methylation of oligonucleotides used in EMSA. DNA fragments were digested by the methylation-sensitive endonuclease *Aci* I before and after methylation.

Additional file 5: Figure S3. The expression of SVA elements that are transcribed in K562 cells is not affected by BORIS dosage. RNA-seq differential ratio distribution for 75 SVA elements, which were apparently transcriptionally active in the untreated K562 cells (i.e., over 10 normalized counts in empty vector control). Only elements longer than 1 Kb were included in analysis. Shown are the graphs for BORIS KD K562 cells, K562 treated with DZNep and the combination of both treatments.

Additional file 6: Figure S4. SVA elements capture BORIS binding sites from unique gene promoters. (A) BORIS ChIP-seq and deep CAGE (Cap Analysis of Gene Expression)-seq (ENCODE data) coverage tracks for the *MAST2* gene (upper track) and for the SVA-F1 element (lower track) in K562 cells. BORIS occupancy at the *MAST2* first exon sequence coincided with the multiple transcription start sites (TSS) for *MAST2* and SVA-F1 family expression in K562 cells. The black arrows show the direction of transcription based on CAGEs enrichment on plus strand. The red double-headed arrows show the *MAST2* sequence captured by SVA-F1 family from the *MAST2* gene. (B) ChIP-seq enrichment of BORIS occupancy depends on the number of BORIS binding sites in the transduced sequences. The top panel is the schematic representation of SVA-F1 elements with different numbers of BORIS binding sites depending on the length of 5'-transduced sequence. The bottom panel is the plot showing the average tag density of BORIS ChIP-Seq across the transduced sequences of different length.

Additional file 7: Figure S5. Examples of BORIS binding at promoters trapped by SVA elements. The gene tracks represent cases of BORIS binding sites within genes' promoters trapped by the indicated SVA element. The red double-headed arrows show the sequences trapped by SVAs and occupied by BORIS in K562 cells. BORIS ChIP-seq coverage and the CAGE tracks are shown for K562 cells. Expression from either minus or plus strands is shown by blue and red CAGE tracks, respectively. The particular examples are: BORIS binding site as the part of *FDX1* promoter trapped by SVA-D, *RHOT1* trapped by SVA-A, *NDUFV2*—by SVA-D, *WRD33*—SVA-F, *PHKA1* and *MMPE1*—by SVA-D.

Additional file 8: Table S3. BORIS binding sites in the promoters of unique genes captured by SVA elements.

Abbreviations

CAGE: cap analysis of gene expression; ChIP: chromatin immunoprecipitation; ChIP-chip: microarray analysis of ChIP; ChIP-seq: NGS analysis of ChIP; CT: cancer testis (genes); CTS: CTCF target sites; DSB: double-strand breaks; DZNeP: 3-deazaneplanocin A; EMSA: electrophoretic mobility shift assay; KD: knock down; KO: knock out; NGS: next-generation sequencing; NOR: nucleolar organizer; PCA: principal component analysis; rDNA: ribosomal RNA genes; RT-qPCR: reverse transcription and quantitative polymerase chain reaction; SVA: SINE, VNTR, and Alu (transposable element); SVD: singular value decomposition; TAD: topologically associated domains; TE: transposable elements; TR: tandem repeat; TRF: tandem repeat finder; TSSs: transcription start sites; VNTR: variable number tandem repeat; ZF: zinc finger; 5-Aza-2'-deoxycytidine.

Authors' contributions

AS, EMP, DL and VL conceived and designed the experiments; EMP, QFW, JLL, CC, CCM, JL, and AB performed experiments; AS, EMP, ET, JL, and APH conducted data analysis; SR and DL contributed reagents and tools; and AS, EMP, APH, DL and VL wrote the paper. All authors read and approved the final manuscript.

Author details

¹ Molecular Epigenetics Laboratory, Guangzhou Institutes of Biomedicine and Health, Guangzhou 510530, Guangdong, China. ² Laboratory of Immunogenetics, NIH, NIAID, Rockville, MD 20852, USA. ³ Department of Biology, Southern University of Science and Technology of China, Shenzhen 518055, Guangdong, China.

Acknowledgements

Authors would like to acknowledge the Drug Discovery Center of the Guangzhou Institutes of Biomedicine and Health for logistical support. It was funded by the Guangzhou sciences and technology Grant 201508020131.

Competing interests

The authors declare that they have no competing interests.

Availability of supporting data

NGS data were deposited to the Gene Expression Omnibus (GEO) repository with the accession number GSE70764. The TRF microarray design and the ChIP-chip datasets were deposited at the GEO with accession number GSE84326.

Funding

This work was supported by the PRC government's "1000 Talents Program" grant to AS, the Guangdong provincial government's "Guangdong High Talent" award to AS, and the Intramural Program of the National Institute of Allergy and Infectious Diseases for VL.

Received: 31 May 2016 Accepted: 18 August 2016

Published online: 31 August 2016

References

- Alkan C, Ventura M, Archidiacono N, Rocchi M, Sahinalp SC, Eichler EE. Organization and evolution of primate centromeric DNA from whole-genome shotgun sequence data. *PLoS Comput Biol*. 2007;3:1807–18.
- Carbone L, Harris RA, Gnerre S, Veeramah KR, Lorente-Galdos B, Huddleston J, Meyer TJ, Herrero J, Roos C, Aken B, et al. Gibbon genome and the fast karyotype evolution of small apes. *Nature*. 2014;513:195–201.
- Schumann GG, Gogvadze EV, Osanai-Futahashi M, Kuroki A, Munk C, Fujiwara H, Ivics Z, Buzdin AA. Unique functions of repetitive transcripts. *Int Rev Cell Mol Biol*. 2010;285:115–88.
- Huang CR, Burns KH, Boeke JD. Active transposition in genomes. *Annu Rev Genet*. 2012;46:651–75.
- Hutchins AP, Pei D. Transposable elements at the center of the crossroads between embryogenesis, embryonic stem cells, reprogramming, and long non-coding RNAs. *Sci Bull*. 2015;60:1722–33.
- Sen SK, Han K, Wang J, Lee J, Wang H, Callinan PA, Dyer M, Cordaux R, Liang P, Batzer MA. Human genomic deletions mediated by recombination between Alu elements. *Am J Hum Genet*. 2006;79:41–53.
- Han K, Lee J, Meyer TJ, Remedios P, Goodwin L, Batzer MA. L1 recombination-associated deletions generate human genomic variation. *Proc Natl Acad Sci USA*. 2008;105:19366–71.
- Cordaux R. The human genome in the LINE of fire. *Proc Natl Acad Sci USA*. 2008;105:19033–4.
- Schneider AM, Duffield AS, Symer DE, Burns KH. Roles of retrotransposons in benign and malignant hematologic disease. *Cellscience*. 2009;6:121–45.
- Gray LT, Fong KK, Pavelitz T, Weiner AM. Tethering of the conserved piggyBac transposase fusion protein CSB-PGBD3 to chromosomal AP-1 proteins regulates expression of nearby genes in humans. *PLoS Genet*. 2012;8:e1002972.
- Gasior SL, Wakeman TP, Xu B, Deininger PL. The human LINE-1 retrotransposon creates DNA double-strand breaks. *J Mol Biol*. 2006;357:1383–93.
- Rodic N, Steranka JP, Makohon-Moore A, Moyer A, Shen P, Sharma R, Kohutek ZA, Huang CR, Ahn D, Mita P, et al. Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma. *Nat Med*. 2015;21:1060–4.
- Stewart C, Kural D, Stromberg MP, Walker JA, Konkeli MK, Stutz AM, Urban AE, Grubert F, Lam HY, Lee WP, et al. A comprehensive map of mobile element insertion polymorphisms in humans. *PLoS Genet*. 2011;7:e1002236.
- Wang D, Su Y, Wang X, Lei H, Yu J. Transposon-derived and satellite-derived repetitive sequences play distinct functional roles in mammalian intron size expansion. *Evol Bioinform*. 2012;8:301–19.
- Crosetto N, Mitra A, Silva MJ, Bienko M, Dojer N, Wang Q, Karaca E, Chiarle R, Skrzypczak M, Ginalski K, et al. Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat Methods*. 2013;10:361–5.
- Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, Brennan PM, Rizzu P, Smith S, Fell M, et al. Somatic retrotransposition alters the genetic landscape of the human brain. *Nature*. 2011;479:534–7.
- Ting DT, Lipson D, Paul S, Brannigan BW, Akhavanfar S, Coffman EJ, Contino G, Deshpande V, Iafate AJ, Letovsky S, et al. Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. *Science*. 2011;331:593–6.
- Chenais B. Transposable elements and human cancer: a causal relationship? *Biochim Biophys Acta*. 2013;1835:28–35.
- Goodier JL. Retrotransposition in tumors and brains. *Mob DNA*. 2014;5:11.
- Estecio MR, Gallegos J, Dekmezian M, Lu Y, Liang S, Issa JP. SINE retrotransposons cause epigenetic reprogramming of adjacent gene promoters. *Mol Cancer Res*. 2012;10:1332–42.
- Babatz TD, Burns KH. Functional impact of the human mobilome. *Curr Opin Genet Dev*. 2013;23:264–70.
- Lee E, Iskov R, Yang L, Gokcumen O, Haseley P, Luquette LJ 3rd, Lohr JG, Harris CC, Ding L, Wilson RK, et al. Landscape of somatic retrotransposition in human cancers. *Science*. 2012;337:967–71.
- Soriano P, Gridley T, Jaenisch R. Retroviruses and insertional mutagenesis in mice: proviral integration at the *Mov 34* locus leads to early embryonic death. *Genes Dev*. 1987;1:366–75.
- Kim DS, Kim TH, Huh JW, Kim IC, Kim SW, Park HS, Kim HS. LINE FUSION GENES: a database of LINE expression in human genes. *BMC Genomics*. 2006;7:139.
- Hancks DC, Kazazian HH Jr. Active human retrotransposons: variation and disease. *Curr Opin Genet Dev*. 2012;22:191–203.
- Nakanishi A, Kobayashi N, Suzuki-Hirano A, Nishihara H, Sasaki T, Hirakawa M, Sumiyama K, Shimogori T, Okada N. A SINE-derived element constitutes a unique modular enhancer for mammalian diencephalic *Fgf8*. *PLoS One*. 2012;7:e43785.
- Jaenisch R, Schnieke A, Harbers K. Treatment of mice with 5-azacytidine efficiently activates silent retroviral genomes in different tissues. *Proc Natl Acad Sci USA*. 1985;82:1451–5.
- Byun HM, Heo K, Mitchell KJ, Yang AS. Mono-allelic retrotransposon insertion addresses epigenetic transcriptional repression in human genome. *J Biomed Sci*. 2012;19:13.

29. Rebollo R, Miceli-Royer K, Zhang Y, Farivar S, Gagnier L, Mager DL. Epigenetic interplay between mouse endogenous retroviruses and host genes. *Genome Biol.* 2012;13:R89.
30. Casa V, Gabellini D. A repetitive elements perspective in Polycomb epigenetics. *Front Genet.* 2012;3:199.
31. Belancio VP, Roy-Engel AM, Deininger PL. All y'all need to know 'bout retroelements in cancer. *Semin Cancer Biol.* 2010;20:200–10.
32. Rebollo R, Horak B, Hubert B, Vieira C. Jumping genes and epigenetics: towards new species. *Gene.* 2010;454:1–7.
33. Maurano MT, Wang H, John S, Shafer A, Canfield T, Lee K, Stamatoyannopoulos JA. Role of DNA Methylation in modulating transcription factor occupancy. *Cell Rep.* 2015;12:1184–95.
34. Landan G, Cohen NM, Mukamel Z, Bar A, Molchadsky A, Brosh R, Horn-Saban S, Zalcenstein DA, Goldfinger N, Zundelovich A, et al. Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nat Genet.* 2012;44:1207–14.
35. Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, Lee K, Canfield T, Weaver M, Sandstrom R, et al. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* 2012;22:1680–8.
36. Ong CT, Corces VG. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet.* 2014;15:234–46.
37. Gomez-Marin C, Tena JJ, Acemel RD, Lopez-Mayorga M, Naranjo S, de la Calle-Mustienes E, Maeso I, Beccari L, Aneas I, Vielmas E, et al. Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. *Proc Natl Acad Sci USA.* 2015;112:7542–7.
38. Lupianez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene–enhancer interactions. *Cell.* 2015;161:1012–25.
39. Ji X, Dadon DB, Powell BE, Fan ZP, Borges-Rivera D, Shachar S, Weintraub AS, Hnisz D, Pegoraro G, Lee TI, et al. 3D chromosome regulatory landscape of human pluripotent cells. *Cell Stem Cell.* 2016;18:262–75.
40. Weth O, Paprotka C, Gunther K, Schulte A, Baierl M, Leers J, Galjart N, Renkawitz R. CTCF induces histone variant incorporation, erases the H3K27me3 histone mark and opens chromatin. *Nucleic Acids Res.* 2014;42:11941–51.
41. Liu M, Maurano MT, Wang H, Qi H, Song CZ, Navas PA, Emery DW, Stamatoyannopoulos JA, Stamatoyannopoulos G. Genomic discovery of potent chromatin insulators for human gene therapy. *Nat Biotechnol.* 2015;33:198–203.
42. Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew JL, Ruan Y, Wei CL, Ng HH, Liu ET. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res.* 2008;18:1752–62.
43. Schwalie PC, Ward MC, Cain CE, Faure AJ, Gilad Y, Odom DT, Flicek P. Co-binding by YY1 identifies the transcriptionally active, highly conserved set of CTCF-bound regions in primate genomes. *Genome Biol.* 2013;14:R148.
44. Deng Z, Wang Z, Stong N, Plasschaert R, Moczan A, Chen HS, Hu S, Wikramasinghe P, Davuluri RV, Bartolomei MS, et al. A role for CTCF and cohesin in subtelomere chromatin organization, TERRA transcription, and telomere end protection. *EMBO J.* 2012;31:4165–78.
45. Stong N, Deng Z, Gupta R, Hu S, Paul S, Weiner AK, Eichler EE, Graves T, Fronick CC, Courtney L, et al. Subtelomeric CTCF and cohesin binding site organization using improved subtelomere assemblies and a novel annotation pipeline. *Genome Res.* 2014;24:1039–50.
46. Holwerda S, de Laat W. Chromatin loops, gene positioning, and gene expression. *Front Genet.* 2012;3:217.
47. Shih HY, Verma-Gaur J, Torkamani A, Feeny AJ, Galjart N, Krangel MS. Tcra gene recombination is supported by a Tcra enhancer- and CTCF-dependent chromatin hub. *Proc Natl Acad Sci USA.* 2012;109:E3493–502.
48. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature.* 2012;489:109–13.
49. Phillips-Cremis JE, Sauria ME, Sanyal A, Gerasimova TI, Lajoie BR, Bell JS, Ong CT, Hookway TA, Guo C, Sun Y, et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell.* 2013;153:1281–95.
50. Horakova AH, Calabrese JM, McLaughlin CR, Tremblay DC, Magnuson T, Chadwick BP. The mouse DXZ4 homology retains Ctfc binding and proximity to Pls3 despite substantial organizational differences compared to the primate macrosatellite. *Genome Biol.* 2012;13:R70.
51. Ottaviani A, Schluth-Bolard C, Gilson E, Magdinier F. D4Z4 as a prototype of CTCF and lamins-dependent insulator in human cells. *Nucleus.* 2010;1:30–6.
52. Horakova AH, Moseley SC, McLaughlin CR, Tremblay DC, Chadwick BP. The macrosatellite DXZ4 mediates CTCF-dependent long-range intrachromosomal interactions on the human inactive X chromosome. *Hum Mol Genet.* 2012;21:4367–77.
53. Arnold R, Maueler W, Bassili G, Lutz M, Burke L, Epplen TJ, Renkawitz R. The insulator protein CTCF represses transcription on binding to the (gt)(22)(ga)(15) microsatellite in intron 2 of the HLA-DRB1(*)0401 gene. *Gene.* 2000;253:209–14.
54. Wang C, Gu Y, Zhang K, Xie K, Zhu M, Dai N, Jiang Y, Guo X, Liu M, Dai J, et al. Systematic identification of genes with a cancer-testis expression pattern in 19 cancer types. *Nat Commun.* 2016;7:10499.
55. Pugacheva EM, Rivero-Hinojosa S, Espinoza CA, Mendez-Catala CF, Kang S, Suzuki T, Kosaka-Suzuki N, Robinson S, Nagarajan V, Ye Z, et al. Comparative analyses of CTCF and BORIS occupancies uncover two distinct classes of CTCF binding genomic regions. *Genome Biol.* 2015;16:161.
56. Alberti L, Losi L, Leyvraz S, Benhattar J. Different effects of BORIS/CTCF on stemness gene expression, sphere formation and cell survival in epithelial cancer stem cells. *PLoS One.* 2015;10:e0132977.
57. Vatolin S, Abdullaev Z, Pack SD, Flanagan PT, Custer M, Loukinov DI, Pugacheva E, Hong JA, Morse H III, Schrupp DS, et al. Conditional expression of the CTCF-paralogous transcriptional factor BORIS in normal cells results in demethylation and derepression of MAGE-A1 and reactivation of other cancer-testis genes. *Cancer Res.* 2005;65:7751–62.
58. Dougherty CJ, Ichim TE, Liu L, Reznik G, Min WP, Ghochikyan A, Agadjanyan MG, Reznik BN. Selective apoptosis of breast cancer cells by siRNA targeting of BORIS. *Biochem Biophys Res Commun.* 2008;370:109–12.
59. Bhan S, Negi SS, Shao C, Glazer CA, Chuang A, Gaykalova DA, Sun W, Sidransky D, Ha PK, Califano JA. BORIS binding to the promoters of cancer testis antigens, MAGEA2, MAGEA3, and MAGEA4, is associated with their transcriptional activation in lung cancer. *Clin Cancer Res.* 2011;17:4267–76.
60. Messerschmidt DM, Knowles BB, Solter D. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes Dev.* 2014;28:812–28.
61. Ehrlich M, Lacey M. DNA hypomethylation and hemimethylation in cancer. *Adv Exp Med Biol.* 2013;754:31–56.
62. Neph S, Stergachis AB, Reynolds A, Sandstrom R, Borenstein E, Stamatoyannopoulos JA. Circuitry and dynamics of human transcription factor regulatory networks. *Cell.* 2012;150:1274–86.
63. Teif VB, Vainshtein Y, Caudron-Herger M, Malm JP, Marth C, Hofer T, Rippe K. Genome-wide nucleosome positioning during embryonic stem cell development. *Nat Struct Mol Biol.* 2012;19:1185–92.
64. Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, Lee CW, Ye C, Ping JL, Mulawadi F, et al. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat Genet.* 2011;43:630–8.
65. Samoshkin A, Dulev S, Loukinov D, Rosenfeld JA, Strunnikov AV. Condensin dysfunction in human cells induces nonrandom chromosomal breaks in anaphase, with distinct patterns for both unique and repeated genomic regions. *Chromosoma.* 2012;121:191–9.
66. Tsiftoglou AS, Pappas IS, Vizirianakis IS. Mechanisms involved in the induced differentiation of leukemia cells. *Pharmacol Ther.* 2003;100:257–90.
67. Schmidt D, Schwalie PC, Wilson MD, Ballester B, Goncalves A, Kutter C, Brown GD, Marshall A, Flicek P, Odom DT. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell.* 2012;148:335–48.
68. Caburet S, Conti C, Schurra C, Lebofsky R, Edelstein SJ, Bensimon A. Human ribosomal RNA gene arrays display a broad range of palindromic structures. *Genome Res.* 2005;15:1079–85.
69. van de Nobelen S, Rosa-Garrido M, Leers J, Heath H, Soochit W, Joosen I, Jonkers I, Demmers J, van der Reijden M, Torrano V, et al. CTCF regulates the local epigenetic state of ribosomal DNA repeats. *Epigenet Chromatin.* 2010;3:19.

70. Sleutels F, Soochit W, Bartkuhn M, Heath H, Dienstbach S, Bergmaier P, Franke V, Rosa-Garrido M, van de Nobelen S, Caesar L, 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. *Epigenet Chromatin*. 2012;5:8.
71. Raiz J, Damert A, Chira S, Held U, Klawitter S, Hamdorf M, Lower J, Stratling WH, Lower R, Schumann GG. The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. *Nucleic Acids Res*. 2012;40:1666–83.
72. Hancks DC, Goodier JL, Mandal PK, Cheung LE, Kazazian HH Jr. Retrotransposition of marked SVA elements by human L1s in cultured cells. *Hum Mol Genet*. 2011;20:3386–400.
73. Jacobs FM, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, Paten B, Salama SR, Haussler D. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature*. 2014;516:242–5.
74. Zhao K, Du J, Han X, Goodier JL, Li P, Zhou X, Wei W, Evans SL, Li L, Zhang W, et al. Modulation of LINE-1 and Alu/SVA retrotransposition by Aicardi-Goutieres syndrome-related SAMHD1. *Cell Rep*. 2013;4:1108–15.
75. Rowe HM, Friedli M, Offner S, Verp S, Mesnard D, Marquis J, Aktas T, Trono D. De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. *Development*. 2013;140:519–29.
76. Quinn JP, Bubb VJ. SVA retrotransposons as modulators of gene expression. *Mob Genet Elem*. 2014;4:e32102.
77. Hancks DC, Kazazian HH Jr. SVA retrotransposons: evolution and genetic instability. *Semin Cancer Biol*. 2010;20:234–45.
78. Hancks DC, Mandal PK, Cheung LE, Kazazian HH Jr. The minimal active human SVA retrotransposon requires only the 5'-hexamer and Alu-like domains. *Mol Cell Biol*. 2012;32:4718–26.
79. Kroutter EN, Belancio VP, Wagstaff BJ, Roy-Engel AM. The RNA polymerase dictates ORF1 requirement and timing of LINE and SINE retrotransposition. *PLoS Genet*. 2009;5:e1000458.
80. Jones PA, Taylor SM. Cellular differentiation, cytidine analogs and DNA methylation. *Cell*. 1980;20:85–93.
81. Jones PA. Effects of 5-azacytidine and its 2'-deoxyderivative on cell differentiation and DNA methylation. *Pharmacol Ther*. 1985;28:17–27.
82. Stresmann C, Lyko F. Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine. *Int J Cancer*. 2008;123:8–13.
83. Miranda TB, Cortez CC, Yoo CB, Liang G, Abe M, Kelly TK, Marquez VE, Jones PA. DZNep is a global histone methylation inhibitor that reactivates developmental genes not silenced by DNA methylation. *Mol Cancer Ther*. 2009;8:1579–88.
84. Tan J, Yang X, Zhuang L, Jiang X, Chen W, Lee PL, Karuturi RK, Tan PB, Liu ET, Yu Q. Pharmacologic disruption of Polycomb-repressive complex 2-mediated gene repression selectively induces apoptosis in cancer cells. *Genes Dev*. 2007;21:1050–63.
85. Damert A, Raiz J, Horn AV, Lower J, Wang H, Xing J, Batzer MA, Lower R, Schumann GG. 5'-Transducing SVA retrotransposon groups spread efficiently throughout the human genome. *Genome Res*. 2009;19:1992–2008.
86. Hancks DC, Ewing AD, Chen JE, Tokunaga K, Kazazian HH Jr. Exon-trapping mediated by the human retrotransposon SVA. *Genome Res*. 2009;19:1983–91.
87. Bantysh OB, Buzdin AA. Novel family of human transposable elements formed due to fusion of the first exon of gene MAST2 with retrotransposon SVA. *Biochemistry*. 2009;74:1393–9.
88. Zabolotneva AA, Bantysh O, Suntsova MV, Efimova N, Malakhova GV, Schumann GG, Gayfullin NM, Buzdin AA. Transcriptional regulation of human-specific SVAF(1) retrotransposons by cis-regulatory MAST2 sequences. *Gene*. 2012;505:128–36.
89. Tang WW, Dietmann S, Irie N, Leitch HG, Floros VI, Bradshaw CR, Hackett JA, Chinnery PF, Surani MA. A unique gene regulatory network resets the human germline epigenome for development. *Cell*. 2015;161:1453–67.
90. Moncunill V, Gonzalez S, Bea S, Andrieux LO, Salaverria I, Royo C, Martinez L, Puiggros M, Segura-Wang M, Stutz AM, et al. Comprehensive characterization of complex structural variations in cancer by directly comparing genome sequence reads. *Nat Biotechnol*. 2014;32:1106–12.
91. Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, Pleasance ED, Lau KW, Beare D, Stebbings LA, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell*. 2011;144:27–40.
92. Zhang CZ, Leibowitz ML, Pellman D. Chromothripsis and beyond: rapid genome evolution from complex chromosomal rearrangements. *Genes Dev*. 2013;27:2513–30.
93. Storchova Z, Kloosterman WP. The genomic characteristics and cellular origin of chromothripsis. *Curr Opin Cell Biol*. 2016;40:106–13.
94. Sunami E, de Maat M, Vu A, Turner RR, Hoon DS. LINE-1 hypomethylation during primary colon cancer progression. *PLoS One*. 2011;6:e18884.
95. Alberti L, Renaud S, Losi L, Leyvraz S, Benhattar J. High expression of hTERT and stemness genes in BORIS/CTCF positive cells isolated from embryonic cancer cells. *PLoS One*. 2014;9:e109921.
96. Kosaka-Suzuki N, Suzuki T, Pugacheva EM, Vostrov AA, Morse HC 3rd, Loukinov D, Lobanenkova V. Transcription factor BORIS (brother of the regulator of imprinted sites) directly induces expression of a cancer-testis antigen, TSP50, through regulated binding of BORIS to the promoter. *J Biol Chem*. 2011;286:27378–88.
97. Suzuki T, Kosaka-Suzuki N, Pack S, Shin DM, Yoon J, Abdullaev Z, Pugacheva E, Morse HC 3rd, Loukinov D, Lobanenkova V. Expression of a testis-specific form of Gal3st1 (CST), a gene essential for spermatogenesis, is regulated by the CTCF paralogous gene BORIS. *Mol Cell Biol*. 2010;30:2473–84.
98. Groh M, Silva LM, Gromak N. Mechanisms of transcriptional dysregulation in repeat expansion disorders. *Biochem Soc Trans*. 2014;42:1123–8.
99. Cleary JD, Ranum LP. Repeat associated non-ATG (RAN) translation: new starts in microsatellite expansion disorders. *Curr Opin Genet Dev*. 2014;26C:6–15.
100. Jarmuz-Szymczak M, Janiszewska J, Szyfter K, Shaffer LG. Narrowing the localization of the region breakpoint in most frequent Robertsonian translocations. *Chromosome Res*. 2014;22:517–32.
101. Bergmann JH, Jakubsche JN, Martins NM, Kagansky A, Nakano M, Kimura H, Kelly DA, Turner BM, Masumoto H, Larionov V, Earnshaw WC. Epigenetic engineering: histone H3K9 acetylation is compatible with kinetochore structure and function. *J Cell Sci*. 2012;125:411–21.
102. Lacoste N, Woolfe A, Tachiwana H, Gareau AV, Barth T, Cantaloube S, Kurumizaka H, Imhof A, Almouzni G. Mislocalization of the centromeric histone variant CenH3/CENP-A in human cells depends on the chaperone DAXX. *Mol Cell*. 2014;53:631–44.
103. Kim DS, Hahn Y. Identification of human-specific transcript variants induced by DNA insertions in the human genome. *Bioinformatics*. 2011;27:14–21.
104. Savage AL, Wilm TP, Khursheed K, Shatunov A, Morrison KE, Shaw PJ, Shaw CE, Smith B, Breen G, Al-Chalabi A, et al. An evaluation of a SVA retrotransposon in the FUS promoter as a transcriptional regulator and its association to ALS. *PLoS One*. 2014;9:e90833.
105. O'Neill MJ, O'Neill RJ. Genomics: something to swing about. *Nature*. 2014;513:174–5.
106. Vogt J, Bengesser K, Claes KB, Wimmer K, Mautner VF, van Minkelen R, Legius E, Brems H, Upadhyaya M, Hogel J, et al. SVA retrotransposon insertion-associated deletion represents a novel mutational mechanism underlying large genomic copy number changes with non-recurrent breakpoints. *Genome Biol*. 2014;15:R80.
107. Pratto F, Brick K, Khil P, Smagulova F, Petukhova GV, Camerini-Otero RD. DNA recombination. Recombination initiation maps of individual human genomes. *Science*. 2014;346:1256442.
108. Oakes CC, La Salle S, Smiraglia DJ, Robaire B, Trasler JM. Developmental acquisition of genome-wide DNA methylation occurs prior to meiosis in male germ cells. *Dev Biol*. 2007;307:368–79.
109. Klawitter S, Fuchs NV, Upton KR, Munoz-Lopez M, Shukla R, Wang J, Garcia-Canadas M, Lopez-Ruiz C, Gerhardt DJ, Sebe A, et al. Reprogramming triggers endogenous L1 and Alu retrotransposition in human induced pluripotent stem cells. *Nat Commun*. 2016;7:10286.
110. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 1999;27:573–80.
111. Gelfand Y, Rodriguez A, Benson G. TRDB—the tandem repeats database. *Nucleic Acids Res*. 2007;35:D80–7.

112. Kent WJ. BLAT—the BLAST-like alignment tool. *Genome Res.* 2002;12:656–64.
113. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009;10:R25.
114. Ye T, Krebs AR, Choukallah MA, Keime C, Plewniak F, Davidson I, Tora L. seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Res.* 2011;39:e35.
115. Bailey TL, Williams N, Misleh C, Li WW. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 2006;34:W369–73.
116. Wee S, Wiederschain D, Maira SM, Loo A, Miller C, deBeaumont R, Stegmeier F, Yao YM, Lengauer C. PTEN-deficient cancers depend on PIK3CB. *Proc Natl Acad Sci USA.* 2008;105:13057–62.
117. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc.* 2012;7:562–78.

Submit your next manuscript to BioMed Central
and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

