Transcription-independent TFIIIC-bound sites cluster near heterochromatin boundaries within lamina-associated domains in *C. elegans*

Alexis Stutzman[1,3)], April Liang[2,4)], Vera Beilinson[1)], and Kohta Ikegami[1)]

[1)]Department of Pediatrics, The University of Chicago, Chicago, Illinois, USA; [2)]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey, USA

Current address:

[3)]Curriculum in Genetics and Molecular Biology, the University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[4)]School of Medicine, University of California San Francisco, San Francisco, California, USA

Correspondence:

Kohta Ikegami

Email, ikgmk@uchicago.edu; Tel, (773) 834-4260

KEYWORDS

Transcription Factor III C (TFIIIC), RNA Polymerase III, *C. elegans*, Insulator, Chromatin, Chromosome, lamina-associated domain, nuclear periphery, LEM-2, Histone H3 Lysine 9 methylation

ABSTRACT

BACKGROUND: Chromatin organization is central to precise control of gene expression. In vertebrates, the insulator protein CTCF plays a central role in organizing chromatin into topologically associated domains (TADs). In nematode *C. elegans*, however, a CTCF homolog is absent, and pervasive TAD structures are limited to the dosage-compensated sex chromosome, leaving the principle of *C. elegans* chromatin organization unclear. Transcription Factor III C (TFIIIC) is a basal transcription factor complex for RNA Polymerase III (Pol III), also implicated in chromatin organization. TFIIIC binding without Pol III co-occupancy, referred to as extra-TFIIIC binding, has been implicated in insulating active and inactive chromatin domains in yeasts, flies, and mammalian cells. Whether extra-TFIIIC sites are present and contribute to chromatin organization in *C. elegans* remain unknown.

RESULTS: We identified, in *C. elegans* embryos, 504 TFIIIC-bound sites absent of Pol III and TATA-binding protein co-occupancy characteristic of extra-TFIIIC sites. Extra-TFIIIC sites constituted half of all identified TFIIIC binding sites in the genome. Unlike Pol III-associated TFIIIC sites predominantly localized in the sex chromosome, extra-TFIIIC sites were highly over-represented within autosomes. Extra-TFIIIC sites formed dense clusters in *cis.* The autosomal regions enriched for extra-TFIIIC site clusters presented a high level of heterochromatin-associated histone H3K9 trimethylation (H3K9me3). Furthermore, extra-TFIIIC site clusters were embedded in the lamina-associated domains. Despite the heterochromatin environment of extra-TFIIIC sites, the individual clusters of extra-TFIIIC sites were devoid of and resided near the individual H3K9me3-marked regions.

CONCLUSION: Clusters of extra-TFIIIC sites were pervasive near the outer boundaries of H3K9me3-marked regions in *C. elegans*. Given the reported activity of extra-TFIIIC sites in heterochromatin insulation in yeasts, our observation raised the possibility that TFIIIC may also demarcate heterochromatin in *C. elegans*.

BACKGROUND

Eukaryotic genomes are organized into domains of various chromatin features including actively transcribed regions, transcription factor-bound regions, and transcriptionally repressed regions [1–4]. Demarcation of chromatin domains is central to precise control and memory of gene expression patterns. Several proteins have been proposed to have activity in demarcating chromatin domains by acting as a physical boundary [5,6], generating nucleosome depleted regions [7], mediating long-range chromatin interactions [8,9], or tethering chromatin to the nuclear periphery [10]. Despite intense studies [11–15], how chromatin domains are demarcated remains poorly understood.

The genome of nematode *Carnorhabditis elegans* has served as a model to study chromatin organization [3,16–18]. The highest level of chromatin organization in *C. elegans* is the chromatin feature that distinguishes between the X chromosome and the autosomes. The X chromosome in *C. elegans* hermaphrodites is organized into topologically associated domains (TADs) [19], the architectural unit mediated by long-range chromatin interactions and commonly seen in metazoan genomes [20–22]. The five autosomes, however, lack robust TAD organization [19]. Instead, each autosome can be subdivided into three, megabase-wide domains, the left arm, the right arm, and the center [23]. The center domains display a low recombination rate [24,25], a high density of essential genes [26], and low heterochromatin-associated histone modifications [16,27]. Autosome arms are rich in repetitive elements [23] and heterochromatin-associated histone modifications [16,27], and are associated with the nuclear membrane [18,28,29]. Within these generally euchromatic centers and heterochromatic arms lie kilobase-wide regions of various chromatin states including transcriptionally active and inactive regions [3,4]. While condensins, a highly conserved class of architectural proteins [30], define half of TAD boundaries in the X chromosome [19], their contribution to autosomal chromatin organization is unclear [14,31]. Furthermore, CTCF, another conserved architectural protein central to chromosome organizations in vertebrates, is thought to be lost during the *C. elegans* evolution [32]. How chromatin domains and chromatin states in the *C. elegans* autosomes are demarcated remains an area of active investigation [4,28,33,34].

The transcription factor IIIC complex (TFIIIC) is a general transcription factor required for recruitment of the RNA Polymerase III (Pol III) machinery to diverse classes of small non-coding RNA genes [35]. TFIIIC has also been implicated in chromatin insulation [36,37]. TFIIIC binds DNA sequence elements called the Box-A and Box-B motifs [35]. When participating in Pol III-dependent transcription, TFIIIC binding to Box-A and Box-B motifs results in recruitment of transcription factor IIIB complex (TFIIIB), which then recruits Pol III [35,38]. By yet unknown mechanisms, however, TFIIIC is also known to bind DNA without further recruitment of TBP and Pol III [37]. These so-called "extra-TFIIIC sites," or "ETC," have been identified in various organisms including yeast [39,40], fly [41], mouse [42], and human [43]. In *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, extra-TFIIIC sites exhibit chromatin boundary functions both as heterochromatin barriers and insulators to gene activation [40,44,45]. In addition, extra-TFIIIC sites in these yeast species have been observed at the nuclear periphery, suggesting a contribution to spatial organization of chromosomes [40,46]. In fly, mouse, and human genomes, extra-TFIIIC sites were found in close proximity to architectural proteins including CTCF, condensin, and cohesin [41–43,47]. These studies collectively suggest a conserved role for extra-TFIIIC sites in chromatin insulation and chromosome organization. However, whether extra-TFIIIC sites exist in the *C. elegans* genome is unknown.

In this study, we unveiled extra-TFIIIC sites in the *C. elegans* genome. Extra-TFIIIC sites were highly over-represented within a subset of autosome arms that presented a high level of heterochromatin-associated histone H3K9 trimethylation. Extra-TFIIIC sites formed dense clusters *in cis* and were embedded in the lamina-associated domains. Despite the heterochromatin environment of extra-TFIIIC sites however, the individual clusters of extra-TFIIIC sites were devoid of and resided near the boundaries of H3K9me3-marked regions. Our study thus raised the possibility that, like extra-TFIIIC sites in other organisms, *C. elegans* extra-TFIIIC sites may also have an activity to demarcate chromatin domains.

RESULTS

Half of *C. elegans* TFIIIC binding sites lack Pol III co-occupancy

The TFIIIC complex is a general transcription factor required for the assembly of the RNA Polymerase III (Pol III) machinery at small non-coding RNA genes such as tRNA genes (**Fig. 1A**). Extra-TFIIIC sites are TFIIIC-bound sites lacking Pol III co-occupancy and are implicated in insulating genomic domains and spatially organizing chromosomes [48]. To determine whether the *C. elegans* genome includes extra-TFIIIC sites, we analyzed the ChIP-seq data published in our previous study [49] for TFIIIC subunits TFTC-3 (human TFIIIC63/GTF3C3 ortholog) and TFTC-5 (human TFIIIC102/GTF3C5 ortholog) (**Fig. 1B**), the Pol III catalytic subunit RPC-1 (human POLR3A ortholog; "Pol III" hereafter), and the TFIIIB component TBP-1 (human TBP ortholog; "TBP" hereafter) in mixed-stage embryos of the wild-type N2 strain *C. elegans*. We identified 1,029 high-confidence TFIIIC-bound sites exhibiting strong and consistent enrichment for both TFTC-3 and TFTC-5 (**Fig. 1C**). tRNA genes were strongly enriched for TFTC-3, TFTC-5, Pol III, and TBP as expected (**Fig. 1D**). However, we also observed TFIIIC-bound sites with low or no Pol III and TBP enrichment (**Fig. 1D**). Of the 1,029 TFIIIC sites, 525 TFIIIC sites (51%) showed strong Pol III enrichment ("Pol III-associated TFIIIC sites" hereafter), whereas 504 TFIIIC sites (49%) showed no or very low Pol III and TBP enrichment ("extra-TFIIIC sites" hereafter; **Fig. 1E, F**).

The lack of Pol III and TBP binding in the subset of the TFIIIC sites may represent a premature Pol III preinitiation complex assembled at Pol III-transcribed non-coding RNA genes [50]. Alternatively, they could be unrelated to Pol III transcription and similar to extra-TFIIIC sites reported in other organisms. To distinguish these two possibilities, we examined the presence of transcription start sites (TSSs) of non-coding RNA genes near TFIIIC-bound sites. We first confirmed that almost all Pol III-associated TFIIIC sites (464 of 525 sites, 88%) harbored non-coding RNA gene TSSs within 100 bp, and the vast majority of these genes encoded tRNAs (376 sites, 72%) or snoRNAs (52 sites, 10%) (**Fig. 1G**). In contrast, only 4% of extra-TFIIIC sites (20/504) harbored non-coding RNA gene TSSs within 100 bp (**Fig. 1G**). Thus, these Pol III-independent TFIIIC binding events in the *C. elegans*

genome are unlikely to participate in local Pol III-dependent transcription, a characteristic behavior of extra-TFIIIC sites reported in other organisms [37].

C. elegans extra-TFIIIC sites possess strong Box-A and Box-B motifs

The TFIIIC complex binds the Box-A and Box-B DNA motifs [35] (**Fig. 1A**). However, the majority of extra-TFIIIC sites in yeast and human possess only the Box-B motif [39,40,43]. To determine whether *C. elegans* extra-TFIIIC sites contain Box-A and Box-B motifs, we performed *de novo* DNA motif analyses at TFIIIC sites. Almost all of the 504 extra-TFIIIC sites in *C. elegans* harbored both the Box-A and Box-B motifs (90% with Box-A, $E=1.9 \times 10^{-1568}$; 87% with Box-B, $E=1.5 \times 10^{-1602}$; **Fig. 1H**). The pervasiveness of these motifs in extra-TFIIIC sites was comparable to that in Pol III-associated TFIIIC binding sites (94% with Box-A, $E=1.1 \times 10^{-589}$; and 92% with Box-B, $E=3.5 \times 10^{-1308}$) (**Fig. 1H**).

Because the Box-A and Box-B motifs constitute the gene-internal promoter for tRNA genes (**Fig. 1A**), we hypothesized that extra-TFIIIC sites correspond to genetic elements similar to tRNA genes. In *C. elegans*, a class of interspersed repetitive elements called CeRep3 has been suspected as tRNA pseudogenes [51]. To determine whether extra-TFIIIC sites coincide with repetitive elements, we surveyed the overlap between TFIIIC sites and all annotated repetitive elements (**Fig. 1I**). Strikingly, 44.6% (225 sites) of extra-TFIIIC sites overlapped repetitive elements (permutation-based empirical $P<0.001$), and almost all of the overlapped elements (95.1%; 214 sites) were the CeRep3 class of repetitive elements (**Fig. 1I**). In contrast, although significant, only 8.2% (43 sites) of Pol III-associated TFIIIC sites overlapped repetitive elements of any class (permutation-based empirical $P=0.001$). Therefore, unlike extra-TFIIIC sites in yeast and humans, *C. elegans* extra-TFIIIC sites harbored both the Box-A and Box-B motifs; furthermore, a large fraction of these sites corresponded to a class of putative tRNA pseudogenes CeRep3.

C. elegans extra-TFIIIC sites are not associated with regulatory elements for protein-coding genes

Previous studies in human and *S. cerevisiae* reported that extra-TFIIIC sites are overrepresented near protein-coding gene promoters, proposing a potential role in protein-coding gene regulation [39,43]. To determine whether *C. elegans* extra-TFIIIC sites were located near protein-coding genes, we measured distance to nearest protein-coding gene TSSs. Unlike extra-TFIIIC sites in other organisms, *C. elegans* extra-TFIIIC sites were not located closer to protein-coding gene TSSs than Pol III-associated TFIIIC sites (Mann Whitney *U* test, P=0.01) or than randomly permutated extra-TFIIIC sites (Mann-Whitney *U* test, $P=2x10^{-5}$; **Fig. 2A**). To further investigate the relationship between extra-TFIIIC sites and regulatory elements for protein-coding genes, we examined the chromatin states defined by a combination of histone modifications in early embryos [4]. Consistent with the distance-based analysis, extra-TFIIIC sites were not overrepresented within "promoter" regions (14 sites, 2.8%, permutation-based empirical P=0.4) (**Fig. 2B**). Instead, as expected from the CeRep3 repeat overrepresentation, extra-TFIIIC sites were highly overrepresented among the chromatin states associated with repetitive elements including "Transcription elongation IV: low expression and repeats" (51 sites, 10.5%), "Repeats, intergenic, low expression introns" (85 sites, 17.5%), and "Repeat, RNA pseudogenes, H3K9me2" (192 sites, 39.5%) (permutation-based empirical P<0.001; **Fig. 2B**). Extra-TFIIIC sites were also overrepresented within the "Enhancer II, intergenic" chromatin state (40 sites, 8.2%, permutation-based empirical P<0.001) (**Fig. 2B**). However, we did not observe robust enhancer-related H3K27ac or H3K4me1 signals at the extra-TFIIIC sites (**Fig. 2C**). Therefore, *C. elegans* extra-TFIIIC sites were not associated with promoter or enhancer regions for protein-coding genes.

C. elegans extra-TFIIIC sites are densely clustered in the distal arms of autosomes

The lack of association with local regulatory features led us to hypothesize that extra-TFIIIC sites were related to large-scale organization of chromosomes as in the case of yeasts [40,52]. Because tRNA genes are highly over-represented in the X chromosome [23], we asked whether extra-TFIIIC sites were also distributed unevenly in the genome. In stark contrast to Pol III-associated TFIIIC sites, which

were highly over-represented in the X chromosome (202 of the 525 sites; permutation-based empirical P<0.001), extra-TFIIIC sites were under-represented in the X chromosome (18 of the 504 sites; permutation-based empirical P<0.001; **Fig. 3A**). Extra-TFIIIC sites were instead highly overrepresented in chromosome V (195 of the 504 sites; permutation-based empirical P<0.001; **Fig. 3A**). *C. elegans* autosomes can be subdivided into three domains of similar size (left arm, center, and right arm) based on repetitive element abundance, recombination rates, and chromatin organization [17,23,25]. We found that most extra-TFIIIC sites were located in autosome arms (486 of the 504 sites; **Fig. 3B**). In addition, extra-TFIIIC sites were overrepresented in only one of each autosome's two arms (overrepresented in the right arm of chromosome I; left arm of chromosome II; left arm of chromosome III; left arm of chromosome IV; right arm of chromosome V; permutation-based empirical P<0.001; **Fig. 3C**). Furthermore, within autosomal arms, extra-TFIIIC sites were locally densely clustered (**Fig. 3D**), with a median interval between neighboring extra-TFIIIC sites of 1207 bp (Mann Whitney *U*-test vs. within-arm permutations $P=2x10^{-16}$; **Fig. 3E**). Among the autosome arms, the chromosome V right arm contained the largest number of extra-TFIIIC sites (188 sites) with extensive clusters (**Fig. 3B, D, E**). Thus, *C. elegans* extra-TFIIIC sites were fundamentally different from Pol III-associated TFIIIC sites in their genomic distribution and highly concentrated at specific locations within autosomal arm domains.

C. elegans extra-TFIIIC sites intersperse H3K9me3-marked heterochromatin domains

The autosome arms in *C. elegans* exhibit high levels of H3K9me2 and H3K9me3, histone modifications associated with constitutive heterochromatin [16,17]. Furthermore, in each autosome, H3K9me2 and H3K9me3 signals are known to be stronger in one arm than the other [16,27]. Because extra-TFIIIC sites have been implicated in heterochromatin insulation [53,54], we hypothesized that extra-TFIIIC sites were located near H3K9me2 or H3K9me3-marked regions. To test this hypothesis, we compared TFIIIC-bound sites with the locations of H3K9me2 and H3K9me3-enriched regions identified in early embryos [3]. Strikingly, the chromosome arms in which extra-TFIIIC sites were overrepresented coincided with the arms that exhibited strong H3K9me2 and H3K9me3 enrichment (**Fig. 4A, B**). Despite this domain-scale co-localization, however, 88% of extra-TFIIIC sites resided in regions not

enriched for H3K9me3 or H3K9me2 at the local level (**Fig. 4C**). In particular, extra-TFIIIC sites were strongly underrepresented in H3K9me3-enriched regions (only 2% in H3K9me3-enriched regions; permutation-based P<0.001; **Fig. 4D**). However, extra-TFIIIC sites were located significantly closer to H3K9me2-enriched regions (median distance 3.1 kb; **Fig. 4E**) and H3K9me3-enriched regions (median distance 12.5 kb; **Fig. 4F**) compared with Pol III-associated TFIIIC sites (H3K9me2, median distance 34.8 kb, Mann Whitney *U*-test $P<2x10^{-16}$; H3K9me3, median distance 50.1 kb, $P<2x10^{-16}$) or extra-TFIIIC sites permutated within autosomal arms (H3K9me2, median distance 11.3 kb, Mann Whitney *U*-test $P=1x10^{-14}$; H3K9me3, median distance 28.9 kb, $P=7x10^{-14}$). Our analysis thus revealed that *C. elegans* extra-TFIIIC sites were located close to, but not overlapped with, H3K9me2 and H3K9me3-enriched regions within autosomal arm domains.

C. elegans extra-TFIIIC sites were located within nuclear membrane-associated domains

In *S. pombe* and *S. cerevisiae*, extra-TFIIIC sites are localized at the nuclear periphery and thought to regulate spatial organization of chromosomes [40,46]. tRNA genes in *S. pombe* are associated with nuclear pores [55]. In *C. elegans*, Pol III-transcribed genes such as tRNA genes are associated with the nuclear pore component NPP-13 [49] (**Fig. 5A**). Owing to the similarity of extra-TFIIIC sites to Pol III-transcribed genes, we hypothesized that extra-TFIIIC sites associated with NPP-13. To test this hypothesis, we compared the location of extra-TFIIIC sites with that of NPP-13-bound sites identified in mixed-stage embryos [49]. As expected, a large fraction of Pol III-associated TFIIIC sites (215 sites, 41%) overlapped NPP-13-bound sites (permutation-based P<0.001; **Fig. 5B**). In contrast, while statistically significant (permutation-based P<0.001), only 6 of the 504 extra-TFIIIC sites (1.2%) overlapped NPP-13-bound sites (**Fig. 5A**). Thus, extra-TFIIIC sites were not likely to be associated with the nuclear pore.

Another mode of chromatin-nuclear envelope interactions in *C. elegans* is mediated by nuclear membrane-anchored, lamin-associated protein LEM-2 [28] (**Fig. 5A**). LEM-2 associates with the specific regions within the arms of chromosomes called “LEM-2 subdomains” interspersed by non-associated “gap” regions of various sizes [28]. We therefore investigated the location of extra-TFIIIC

sites with respect to that of LEM-2 subdomains and gaps (**Fig. 5C**). Strikingly, 441 of the 504 extra-TFIIIC sites (88%) were located within LEM-2 associated subdomains (permutation-based P<0.001, permutation performed within chromosomal domains) (**Fig. 5C, D**). In contrast, Pol III-associated TFIIIC sites were underrepresented within LEM-2 associated domains (permutation-based P<0.001, permutation performed within chromosomal domains). Our results suggest that *C. elegans* extra-TFIIIC sites are localized at the nuclear periphery and intersperse H3K9me3-marked heterochromatin regions (**Fig. 5D**).

DISCUSSION

The TFIIIC complex is responsible for recruiting TBP and Pol III for transcription of small noncoding RNAs, such as tRNAs [35]. In addition to its transcriptional role, the TFIIIC complex has been known to bind genomic locations devoid of Pol III-mediated transcription in yeast [39,40], fly [41], mouse [42], and human [43]. These sites have been termed extra-TFIIIC sites [39,53]. However, whether such extra-TFIIIC sites are present in *C. elegans* had been unknown. Our data demonstrated that half of all TFIIIC-bound sites in *C. elegans* embryos lack RNA Pol III binding, TBP binding, and nearby noncoding RNA genes, revealing pervasive extra-TFIIIC sites in the *C. elegans* genome.

Previous studies have suggested that extra-TFIIIC sites act as genomic insulators by blocking enhancer activity or heterochromatin spreading [40,54,56], or mediating three-dimensional genome interactions [41,56]. Some of the genomic and chromatin features of *C. elegans* extra-TFIIIC sites reported in this paper resemble characteristics of extra-TFIIIC sites participating in chromatin insulation. First, similar to arrays of TFIIIC-bound sites capable of insulating heterochromatin and enhancer activities in *S. pombe* and human cells [40,56], *C. elegans* extra-TFIIIC were densely clustered in *cis*. Second, similar to the observation that some extra-TFIIIC sites are located at the boundaries of heterochromatin [40,54,56], *C. elegans* extra-TFIIIC sites were located close to, but not within, H3K9me3-marked regions. Third, similar to yeast extra-TFIIIC sites localized to the nuclear periphery [40,46], *C. elegans* extra-TFIIIC sites coincided with genomic regions known to be associated with nuclear membrane protein LEM-2 [28]. These observations raise the possibility that *C. elegans* extra-TFIIIC sites may act as a chromatin insulator.

There are also differences between *C. elegans* extra-TFIIIC sites and those in other organisms. First, unlike yeast and human extra-TFIIIC sites that only possess the Box-B motif [39,43], *C. elegans* extra-TFIIIC sites possess both the Box-A and Box-B motifs. Second, unlike human, fly, and mouse extra-TFIIIC sites that are located near regulatory elements for RNA Polymerase II (Pol II) transcription [41–43], *C. elegans* extra-TFIIIC sites were not located near gene promoters or enhancers or associated with chromatin features of Pol II-dependent regulatory regions. Third, unlike human and mouse extra-TFIIIC sites that are located near CTCF binding sites [42,43], *C. elegans* extra-TFIIIC sites

are unrelated to CTCF binding because the genome does not encode CTCF [32]. Our study could thus offer an opportunity for a comparative analysis of extra-TFIIIC functions across eukaryotic species.

The molecular mechanisms underlying the chromatin organization of *C. elegans* autosomes remain poorly understood. Unlike the X chromosome organized into topological associated domains (TADs), the autosomes do not present strong and pervasive TADs [19]. The condensin binding sites that could create TAD boundaries in the X chromosome did not do so in the autosomes [14]. Several mechanisms for autosome chromatin organization have been proposed. These mechanisms include the antagonism between H3K36 methyltransferase MES-4 and H3K27 methyltransferase RPC2 that defines active vs. repressed chromatin boundaries [4,34]; small chromatin loops emanating from the nuclear periphery that allow active transcription within heterochromatin domains [28]; and active retention of histone acetylase to euchromatin that prevents heterochromatin relocalization [33]. Our data that extra-TFIIIC sites are highly overrepresented in autosome arms and cluster in *cis* near the boundaries of H3K9me3-marked regions warrant future investigation of whether TFIIIC proteins participate in chromatin insulation in *C. elegans* autosomes.

How strategies to demarcate chromatin domains have evolved in eukaryotes remain unclear. In vertebrates, CTCF has a central role in defining TAD boundaries by mediating DNA loops and is essential for development [11,57,58]. In *D. melanogaster*, CTCF is essential but does not appear to define TAD boundaries, and instead acts as a barrier insulator [59–61]. In the non-bilaterian metazoans, some bilaterian animals (such as *C. elegans*), plants, and fungi, CTCF orthologs are absent [32,62]. In contrast to CTCF, the TFIIIC proteins are conserved across eukaryotes [63] and extra-TFIIIC sites have been reported in human, [43], mouse [42], fly [41], *C elegans* (this study), and yeast [39,40]. Whether extra-TFIIIC is an evolutionary conserved mechanism for demarcating chromatin domains in eukaryotes, including those lacking a CTCF ortholog, will be an interesting subject of future studies.

CONCLUSIONS

We identified TFIIIC-bound sites not participating RNA Polymerase III-dependent transcription in the *C. elegans* genome. These "extra-TFIIIC" sites were highly over-represented in the arm domains of the autosomes interacting with the nuclear lamina. Extra-TFIIIC sites formed dense clusters in *cis* near the outer boundaries of individual H3K9me3-marked heterochromatin regions. These genomic features of *C. elegans* extra-TFIIIC sites resemble extra-TFIIIC sites reported in other organisms that have activities in insulating heterochromatin. Our study warrants future investigation of whether TFIIIC proteins participate in heterochromatin insulation in *C. elegans*.

METHODS

ChIP-seq dataset

ChIP-seq of TFTC-3, TFTC-5, RPC-1, and TBP-1 were performed in chromatin extracts of the mixed-stage N2-strain embryos in duplicates and have been reported in our previous publication [49]. These data sets are available at Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) with accession numbers GSE28772 (TFTC-3 ChIP and input); GSE28773 (TFTC-5 ChIP and input); GSE28774 (RPC-1 ChIP and input); GSE42714 (TBP-1 ChIP and input). ChIP-seq datasets for H3K4me3, H3K4me1, H3K27ac, H3K9me2, and H3K9me3, performed in chromatin extracts of early-stage N2-strain embryos [3], were downloaded from ENCODE website (https://www.encodeproject.org/comparative/chromatin/).

Reference genome

The ce10 reference sequence was used throughout. The chromosomal domains (left arm, center, and the right arm) defined by recombination rates [24] was used.

Gene annotation

The genomic coordinates and the type of *C. elegans* transcripts were downloaded from the WS264 annotation of WormMine. The WS264 genomic coordinates were transformed to the ce10 genomic coordinates using the liftOver function (version 343) with the default mapping parameter using the *ce11ToCe10.over.chain* chain file downloaded from the UCSC genome browser.

TFIIIC site definition

MACS2 identified 1,658 TFTC-3-enriched sites. Of those, sites that had the TFTC-3 fold-enrichment (FE) score greater than 5, harbored TFTC-5-binding sites within 100 bp, and were located in the nuclear chromosomes were considered “high-confidence” TFIIIC-bound sites (1,029 TFIIIC sites). The “center” of each TFIIIC site was defined by the position of the base with the largest TFTC-3 FE score. Of the 1,029 TFIIIC sites, those with the maximum Pol III (RPC-1) FE greater than 20 within +/–250 bp of the site center were defined as “Pol III-associated TFIIIC” sites (525) and the remaining sites were

defined as “extra-TFIIIC” sites (504). The genomic coordinates for Pol III-associated TFIIIC sites and extra-TFIIIC sites are listed in **Table S1**.

Heatmap

For the heatmaps around TFIIIC sites, a set of 20-bp windows with a 10-bp offset that covered a 2-kb region centered around the center of TFIIIC sites was generated for each site. For each window, the mean of fold-enrichment score was computed from replicate-combined input-normalized fold enrichment bedgraph files. The signals were visualized using the ggplot2’s geom_raster function (version 2.2.1) in R.

Non-coding RNA genes

The genomic location and classification of non-coding RNA genes was described in the *Gene Annotation* section. For each TFIIIC site extended +/–100 bp from the site center, whether the region contained the transcription start site (TSS) of non-coding RNA genes was assessed using the Bedtools *intersect* function [64] (version 2.26.0).

DNA motif analysis

To find DNA motifs *de novo*, 150-bp sequences centered around the center of the TFIIIC-bound sites were analyzed by MEME (v4.11.3) [65] with the following parameters: minimum motif size, 6 bp; maximum motif size, 12 bp; and the expected motif occurrence of zero or one per sequence (-mod zoops) and with the 1st-order Markov model (i.e. the dinucleotide frequency) derived from the ce10 genome sequence as the background.

Genomic intersection and permutation

Unless otherwise noted, the overlap between the 1-bp center of each of the TFIIIC sites and genomic features of interest (with size ≥ 1 bp) was assessed using the Bedtools *intersect* function [64] (version 2.26.0). To estimate the probability of observing the overlap frequency by chance given the frequency,

location, and size of the features of interest and TFIIIC sites, the TFIIIC sites were permutated using the Bedtools *shuffle* function (version 2.26.0). The TFIIIC sites were shuffled across the genome, or within the chromosomes, or within chromosomal domains in which they reside, as described in each analysis section. After each permutation, the permutated set of TFIIIC sites were assessed for the overlap with the features of interest. This permutation was repeated 2,000 times to assess the frequency at which the number of intersections for the permutated set of the TFIIIC sites was greater or less than the number of intersections for the observed TFIIIC sites. If none of the 2,000 permutations resulted in the number of overlaps greater or less than the observed number of overlaps, the observed degree of overlaps was considered overrepresented or underrepresented, respectively, with the empirical P-value cutoff of 0.001. The mean number of overlaps after 2,000 permutations was computed for visualization.

Repetitive element analysis

The ce10 genomic coordinate and classification of repetitive elements, compiled as the "RepeatMasker" feature, were downloaded from the UCSC genome browser. The intersection between repetitive elements and TFIIIC sites was assessed as described in the *Genomic intersection and permutation* section. The permutation of TFIIIC sites was performed within the chromosomes in which they resided.

Protein-coding gene distance

The genomic location of protein-coding RNA genes was described in the *Gene Annotation* section. For each TFIIIC site, the absolute distance between the center of the TFIIIC site and the closest TSS of a protein-coding gene was obtained using the Bedtools *closest* function [64] (version 2.26.0). To assess the probability of observing such distance distribution by chance given the frequency and location of the TSSs and TFIIIC sites, the TFIIIC sites were permutated once using the Bedtools *shuffle* function (version 2.26.0) such that the TFIIIC sites were shuffled within the chromosomes, and the distance between the permutated TFIIIC sites and closest protein-coding gene TSS was obtained. Mann-

Whitney U test, provided by the *wilcox.test* function in R, was used to assess the difference of the distribution of the TFIIIC-TSS distances between groups.

Chromatin state analysis

The chromatin state annotations for autosomes and the X chromosome are reported previously [4]. The intersection between chromatin state annotations and TFIIIC sites was assessed as described in the *Genomic intersection and permutation* section. The permutation of TFIIIC sites was performed within the chromosomal domains (see *Reference genome)* to account for the difference of the chromatin state representation among different chromosomal domains.

Chromosomal distribution of TFIIIC sites

The number of TFIIIC sites in each chromosome was assessed as described in the *Genomic intersection and permutation* section. The permutation of TFIIIC sites was performed across the genome. The number of TFIIIC sites in each chromosomal domain (see *Reference genome)* was assessed as described in the *Genomic intersection and permutation* section. The permutation of TFIIIC sites was performed within chromosomes.

Extra-TFIIIC site interval

For each chromosomal domain, the genomic distance between every pair of two neighboring extra-TFIIIC sites (center-to-center distance) was computed in R. To estimate the degree of closeness between extra-TFIIIC sites only explained by the frequency of extra-TFIIIC sites within chromosomal domains, the extra-TFIIIC sites were permutated 2,000 times within chromosomal domains as described in the *Genomic intersection and permutation* section. In each of the 2,000 permutations, the genomic distance between two neighboring permutated extra-TFIIIC sites was computed, and the mean of the distances was computed. The distribution of the 2,000 means (by 2,000 permutations) was compared with the distribution of observed distribution of TFIIIC interval sizes by Mann-Whitney *U* test.

Analysis of H3K9me2 and H3K9me3 regions

To define H3K9me2-enriched and H3K9me3-enriched regions, the genome was segmented into 1-kb windows, and the mean fold-enrichment score of H3K9me2 and H3K9me3 (see *ChIP-seq dataset*) was computed for each window using the Bedtools *map* function (version 2.26.0). Windows with the mean fold-enrichment score greater than 2.5 (1.5x standard deviation above mean for H3K9me2; and 1.3x standard deviation above mean for H3K9me3) were considered enriched for H3K9me2 or H3K9me3 and merged if located without a gap. This yielded 2,967 H3K9me2-enriched regions (mean size, 2.1 kb) and 1,331 H3K9me3-enriched regions (mean size, 4.9 kb).

The intersection between H3K9me2-enriched or H3K9me3-enriched regions and TFIIIC sites was assessed as described in the *Genomic intersection and permutation* section. The permutation of TFIIIC sites was performed within the chromosomal domains.

For each TFIIIC site, the absolute distance between the center of the TFIIIC site and the closest H3K9me2-enriched and H3K9me3-enriched regions was obtained using the Bedtools *closest* function [64] (version 2.26.0). To assess the probability of observing such distance distribution by chance given the frequency, location, and size of H3K9me2-enriched and H3K9me3-enriched regions and TFIIIC sites, the TFIIIC sites were permutated once using the Bedtools *shuffle* function (version 2.26.0). For the analysis of the distance to H3K9me2-enriched regions, this permutation was performed within the chromosomal domains. For the analysis of the distance to H3K9me3-enriched regions, permutation was performed with the chromosomal domains but excluding the H3K9me3-enriched regions themselves because the TFIIIC sites were strongly underrepresented in the H3K9me3-enriched regions.

TFIIIC sites in NPP-13-binding sites

The 223 NPP-13-binding sites identified in mixed-stage N2-stage embryos are previously reported [49]. The genomic coordinates were converted to the ce10 genomic coordinates using the UCSCtools *liftOver* function (version 343). The intersection between NPP-13-binding sites and TFIIIC sites was

assessed as described in the *Genomic intersection and permutation* section. The permutation of TFIIIC sites was performed within the chromosomal domains.

TFIIIC sites in LEM-2 subdomain and gaps

The LEM-2 subdomains and gaps between LEM-2 subdomains identified in mixed-stage N2-stage embryos are previously reported [28]. The genomic coordinates were converted to the ce10 genomic coordinates using the UCSCtools *liftOver* function (version 343). The intersection between LEM-2 subdomains or gaps of variable size classes and TFIIIC sites was assessed as described in the *Genomic intersection and permutation* section. The permutation of TFIIIC sites was performed within the chromosomal domains.

LIST OF ABBREVIATIONS

C. elegans: *Caenorhabditis elegans*

D. melanogaster: Drosophila melanogaster

S. cerevisiae: Saccharomyces cerevisiae

S. pombe: Schizosaccharomyces pombe

ChIP-seq: chromatin immunoprecipitation followed by sequencing

Pol II: RNA polymerase II

Pol III: RNA polymerase III

TBP: TATA-binding protein

TFIIIB: Transcription factor III B

TFIIIC: Transcription factor III C

TSS: Transcription start site(s)

DECLARATIONS

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

ChIP-seq datasets are available at Gene Expression Omnibus with accession numbers GSE28772, GSE28773, GSE28774, GSE42714.

COMPETING INTERESTS

The authors declare no competing interests.

FUNDING

K.I., A.S., and V.B. are supported by National Institutes of Health grant R21 AG054770-01A1. A.S. was supported in part by National Institutes of Health grant R25 GM5533619. A.L. was supported by the Princeton University Program in Quantitative and Computational Biology and the Lewis-Sigler Richard Fisher `57 Fund.

AUTHOR CONTRIBUTIONS

K.I. conceived the study. K.I., A.S., A.L., and V.B. analyzed the data. K.I., A.S., A.L., and V.B. wrote the manuscript.

ACKNOWLEDGEMENTS

We thank Jason D. Lieb, Sevinc Ercan, and Sebastian Pott for discussion.

REFERENCES

1. Filion GJ, van Bemmel JG, Braunschweig U, Talhout W, Kind J, Ward LD, et al. Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell. 2010;143:212–24.

2. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. Nature Publishing Group; 2011;473:43–9.

3. Ho JWK, Jung YL, Liu T, Alver BH, Lee S, Ikegami K, et al. Comparative analysis of metazoan chromatin organization. Nature. Nature Publishing Group; 2014;512:449–52.

4. Evans KJ, Huang N, Stempor P, Chesney MA, Down TA, Ahringer J. Stable Caenorhabditis elegans chromatin domains separate broadly expressed and developmentally regulated genes. Proc Natl Acad Sci U S A. 2016;113:E7020–9.

5. Narendra V, Rocha PP, An D, Raviram R, Skok JA, Mazzoni EO, et al. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science. 2015;347:1017–21.

6. Sun F-L, Elgin SCR. Putting Boundaries on Silence. Cell. 1999;99:459–62.

7. Lhoumaud P, Hennion M, Gamot A, Cuddapah S, Queille S, Liang J, et al. Insulators recruit histone methyltransferase dMes4 to regulate chromatin of flanking genes. EMBO J. 2014;33:1599–613.

8. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014;159:1665–80.

9. Gerasimova TI, Byrd K, Corces VG. A Chromatin Insulator Determines the Nuclear Localization of DNA. Mol Cell. 2000;6:1025–35.

10. van Bemmel JG, Pagie L, Braunschweig U, Brugman W, Meuleman W, Kerkhoven RM, et al. The insulator protein SU(HW) fine-tunes nuclear lamina interactions of the Drosophila genome. PLoS One. 2010;5:e15013.

11. Nora EP, Goloborodko A, Valton A-L, Gibcus JH, Uebersohn A, Abdennur N, et al. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell. Elsevier; 2017;169:930–44.e22.

12. Ghavi-Helm Y, Jankowski A, Meiers S, Viales RR, Korbel JO, Furlong EEM. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. Nat Genet. 2019;51:1272–82.

13. Despang A, Schöpflin R, Franke M, Ali S, Jerković I, Paliou C, et al. Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. Nat Genet. 2019;51:1263–71.

14. Anderson EC, Frankino PA, Higuchi-Sanabria R, Yang Q, Bian Q, Podshivalova K, et al. X Chromosome Domain Architecture Regulates Caenorhabditis elegans Lifespan but Not Dosage Compensation. Dev Cell [Internet]. 2019; Available from: http://dx.doi.org/10.1016/j.devcel.2019.08.004

15. Quinodoz SA, Ollikainen N, Tabak B, Palla A, Schmidt JM, Detmar E, et al. Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. Cell. 2018;174:744–57.e24.

16. Liu T, Rechtsteiner A, Egelhofer TA, Vielle A, Latorre I, Cheung M-S, et al. Broad chromosomal domains of histone modification patterns in C. elegans. Genome Res. 2011;21:227–36.

17. Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, et al. Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science. American Association for the Advancement of Science; 2010;330:1775–87.

18. Towbin BD, González-Aguilera C, Sack R, Gaidatzis D, Kalck V, Meister P, et al. Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. Cell. Elsevier Ltd; 2012;150:934–47.

19. Crane E, Bian Q, McCord RP, Lajoie BR, Wheeler BS, Ralston EJ, et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. Nature. 2015;523:240–4.

20. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. Nature Publishing Group; 2012;485:376–80.

21. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. Nature Publishing Group; 2012;485:381–5.

22. Sexton T, Cavalli G. The role of chromosome domains in shaping the functional genome. Cell. 2015;160:1049–59.

23. C. elegans Sequencing Consortium. Genome sequence of the nematode C. elegans: a platform for investigating biology. Science. 1998;282:2012–8.

24. Rockman MV, Kruglyak L. Recombinational landscape and population genomics of Caenorhabditis elegans. PLoS Genet. 2009;5:e1000419.

25. Barnes TM, Kohara Y, Codson +. A., Hekimi S. Meiotic Recombination, Noncoding DNA and Genomic Organization in Caenorhabditk elegans. Genetics. 1995;141:159.

26. Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, et al. Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. Nature. 2003;421:231–7.

27. Gu SG, Fire A. Partitioning the C. elegans genome by nucleosome modification, occupancy, and positioning. Chromosoma. 2010;119:73–87.

28. Ikegami K, Egelhofer TA, Strome S, Lieb JD. Caenorhabditis elegans chromosome arms are anchored to the nuclear membrane via discontinuous association with LEM-2. Genome Biol. 2010;11:R120.

29. González-Aguilera C, Ikegami K, Ayuso C, de Luis A, Íñiguez M, Cabello J, et al. Genome-wide analysis links emerin to neuromuscular junction activity in Caenorhabditis elegans. Genome Biol. 2014;15:R21.

30. Hudson DF, Marshall KM, Earnshaw WC. Condensin: Architect of mitotic chromosomes. Chromosome Res. 2009;17:131–44.

31. Kranz A-L, Jiao C-Y, Winterkorn LH, Albritton SE, Kramer M, Ercan S. Genome-wide analysis of condensin binding in Caenorhabditis elegans. Genome Biol. 2013;14:R112.

32. Heger P, Marin B, Schierenberg E. Loss of the insulator protein CTCF during nematode evolution.

BMC Mol Biol. BioMed Central Ltd; 2009;10:84.

33. Cabianca DS, Muñoz-Jiménez C, Kalck V, Gaidatzis D, Padeken J, Seeber A, et al. Active chromatin marks drive spatial sequestration of heterochromatin in C. elegans nuclei. Nature. 2019;569:734–9.

34. Gaydos LJ, Rechtsteiner A, Egelhofer TA, Carroll CR, Strome S. Antagonism between MES-4 and Polycomb repressive complex 2 promotes appropriate gene expression in C. elegans germ cells. Cell Rep. 2012;2:1169–77.

35. Schramm L, Hernandez N. Recruitment of RNA polymerase III to its target promoters. Genes Dev. Cold Spring Harbor Lab; 2002;16:2593–620.

36. Noma K-I, Kamakaka RT. The human Pol III transcriptome and gene information flow. Nat Struct Mol Biol. Nature Publishing Group; 2010;17:539–41.

37. Donze D. Extra-transcriptional functions of RNA Polymerase III complexes: TFIIIC as a potential global chromatin bookmark. Gene. 2012;493:169–75.

38. Ramsay EP, Vannini A. Structural rearrangements of the RNA polymerase III machinery during tRNA transcription initiation. Biochim Biophys Acta Gene Regul Mech. 2018;1861:285–94.

39. Moqtaderi Z, Struhl K. Genome-wide occupancy profile of the RNA polymerase III machinery in Saccharomyces cerevisiae reveals loci with incomplete transcription complexes. Mol Cell Biol. American Society for Microbiology (ASM); 2004;24:4118–27.

40. Noma K-I, Cam HP, Maraia RJ, Grewal SIS. A role for TFIIIC transcription factor complex in genome organization. Cell. Elsevier Ltd; 2006;125:859–72.

41. Van Bortle K, Nichols MH, Li L, Ong C-T, Takenaka N, Qin ZS, et al. Insulator function and topological domain border strength scale with architectural protein occupancy. Genome Biol. BioMed Central; 2014;15:R82.

42. Carrière L, Graziani S, Alibert O, Ghavi-Helm Y, Boussouar F, Humbertclaude H, et al. Genomic binding of Pol III transcription machinery and relationship with TFIIS transcription factor distribution in mouse embryonic stem cells. Nucleic Acids Res. 2012;40:270–83.

43. Moqtaderi Z, Wang J, Raha D, White RJ, Snyder M, Weng Z, et al. Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells. Nat Struct Mol Biol. Nature Publishing Group; 2010;17:635–40.

44. Simms TA, Dugas SL, Gremillion JC, Ibos ME, Dandurand MN, Toliver TT, et al. TFIIIC binding sites function as both heterochromatin barriers and chromatin insulators in Saccharomyces cerevisiae. Eukaryot Cell. 2008;7:2078–86.

45. Valenzuela L, Dhillon N, Kamakaka RT. Transcription independent insulation at TFIIIC-dependent insulators. Genetics. 2009;183:131–48.

46. Hiraga S-I, Botsios S, Donze D, Donaldson AD. TFIIIC localizes budding yeast ETC sites to the nuclear periphery. Mol Biol Cell. American Society for Cell Biology; 2012;23:2741–54.

47. Yuen KC, Slaughter BD, Gerton JL. Condensin II is anchored by TFIIIC and H3K4me3 in the mammalian genome and supports the expression of active dense gene clusters. Sci Adv. 2017;3:e1700191.

48. Van Bortle K, Corces VG. tDNA insulators and the emerging role of TFIIIC in genome organization. Transcription. 2012;3:277–84.

49. Ikegami K, Lieb JD. Integral nuclear pore proteins bind to Pol III-transcribed genes and are required for Pol III transcript processing in C. elegans. Mol Cell. 2013;51:840–9.

50. Kassavetis GA, Braun BR, Nguyen LH, Geiduschek EP. S. cerevisiae TFIIIB is the transcription initiation factor proper of RNA polymerase III, while TFIIIA and TFIIIC are assembly factors. Cell. 1990;60:235–45.

51. Stricklin SL, Griffiths-Jones S, Eddy SR. C. elegans noncoding RNA genes. WormBook; 2005.

52. Iwasaki O, Tanaka A, Tanizawa H, Grewal SIS, Noma K-I. Centromeric localization of dispersed Pol III genes in fission yeast. Mol Biol Cell. 2010;21:254–65.

53. Kirkland JG, Raab JR, Kamakaka RT. TFIIIC bound DNA elements in nuclear organization and insulation. Biochim Biophys Acta. 2013;1829:418–24.

54. Donze D, Adams CR, Rine J, Kamakaka RT. The boundaries of the silenced HMR domain in Saccharomyces cerevisiae. Genes Dev. 1999;13:698–708.

55. Ruben GJ, Kirkland JG, MacDonough T, Chen M, Dubey RN, Gartenberg MR, et al. Nucleoporin mediated nuclear positioning and silencing of HMR. PLoS One. 2011;6:e21923.

56. Raab JR, Chiu J, Zhu J, Katzman S, Kurukuti S, Wade PA, et al. Human tRNA genes function as chromatin insulators. EMBO J. 2012;31:330–50.

57. Fedoriw AM, Stein P, Svoboda P, Schultz RM, Bartolomei MS. Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting. Science. 2004;303:238–40.

58. Carmona-Aldana F, Zampedri C, Suaste-Olmos F, Murillo-de-Ozores A, Guerrero G, Arzate-Mejía R, et al. CTCF knockout reveals an essential role for this protein during the zebrafish development. Mech Dev. 2018;154:51–9.

59. Schwartz YB, Cavalli G. Three-Dimensional Genome Organization and Function in Drosophila. Genetics. 2017;205:5–24.

60. Gerasimova TI, Lei EP, Bushey AM, Corces VG. Coordinated control of dCTCF and gypsy chromatin insulators in Drosophila. Mol Cell. 2007;28:761–72.

61. Rowley MJ, Nichols MH, Lyu X, Ando-Kuri M, Rivera ISM, Hermetz K, et al. Evolutionarily Conserved Principles Predict 3D Chromatin Organization. Mol Cell. 2017;67:837–52.e7.

62. Heger P, Marin B, Bartkuhn M, Schierenberg E, Wiehe T. The chromatin insulator CTCF and the emergence of metazoan diversity. Proc Natl Acad Sci U S A. 2012;109:17507–12.

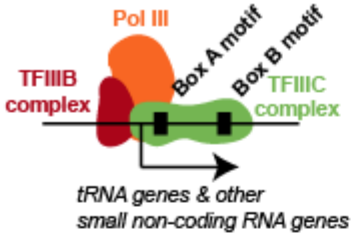
63. Matsutani S. Evolution of the B-Block Binding Subunit of TFIIIC That Binds to the Internal Promoter for RNA Polymerase III. Int J Evol Biol. 2014;2014:609865.

64. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.

65. Machanick P, Bailey TL. MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics. 2011;27:1696–7.

Figure 1

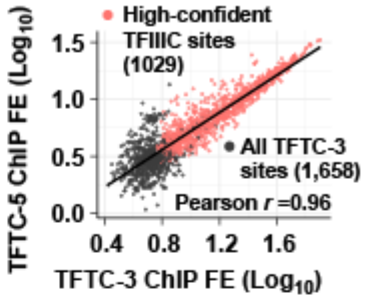
A *Pol III transcriptional machinery*



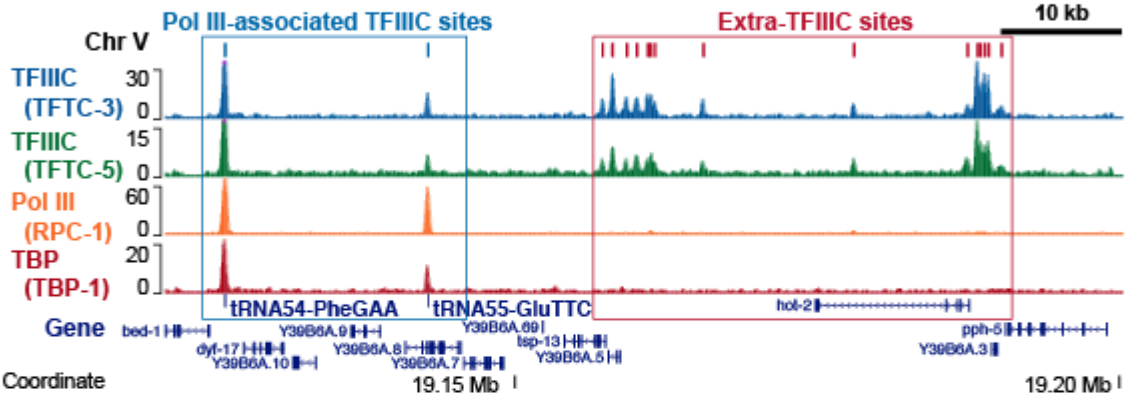
B *C. elegans TFIIIC components*

C. elegans	*S. cerevisiae*	*S. pombe*	*Human*
TFTC-5	Tfc1/t95	Sfc1	TFIIIC63/GTF3C5
TFTC-1	Tfc3/t138	Sfc3	TFIIIC220/GTF3C1
TFTC-3	Tfc4/t131	Sfc4	TFIIIC102/GTF3C3
Unknown	Tfc6/t91	Sfc6	TFIIIC110/GTF3C2
Unknown	Tfc7/t55	Sfc7	TFIIIC35/GTF3C6
Unknown	Tfc8/t60	Sfc9	TFIIIC90/GTF3C4

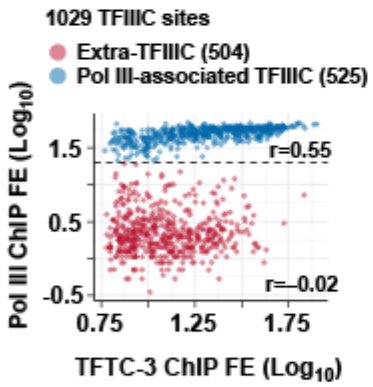
C *TFTC-3 and TFTC-5 ChIP-seq*



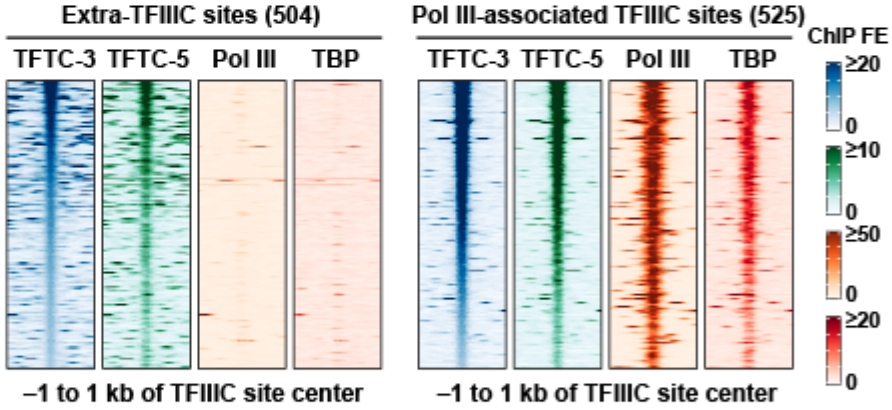
D *Pol III-associated- and extra-TFIIIC sites*



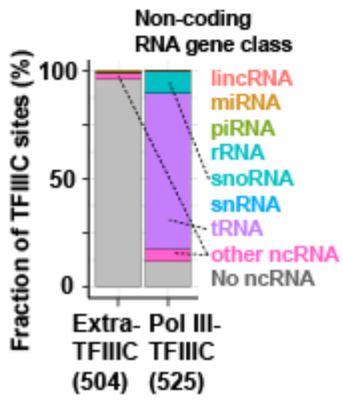
E *Pol III binding levels at TFIIIC sites*



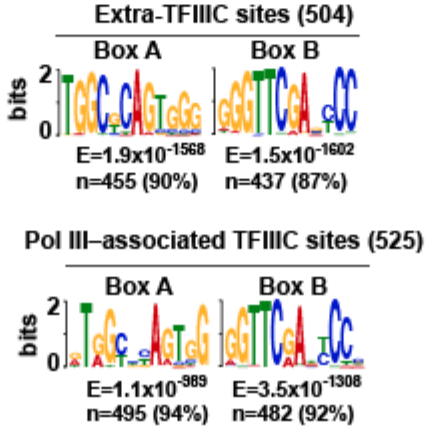
F *TFIIIC, Pol III, TBP signals at TFIIIC sites*



G *Non-coding RNA gene +/–100 bp of TFIIIC sites*



H *Motifs +/–75 bp of TFIIIC sites*



I *Repetitive elements at TFIIIC sites*

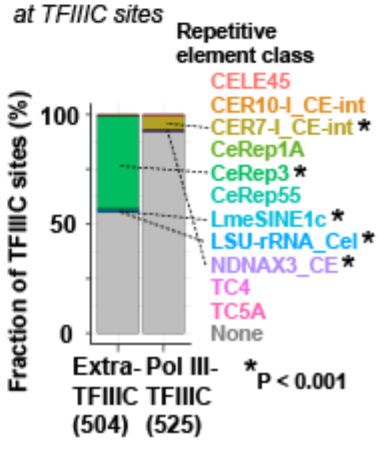


FIGURE 1. Identification of extra-TFIIIC sites in the *C. elegans* genome

(A) A schematic of the RNA polymerase III transcriptional machinery. **(B)** *C. elegans* TFIIIC complex proteins and their yeast and human orthologs for reference. **(C)** Correlation between TFTC-3 and TFTC-5 ChIP-seq fold enrichment (FE) scores at the 1,658 TFTC-3-binding sites. The 1,029 high-confident TFIIIC sites (see **Methods**) are indicated. **(D)** A representative genomic region showing extra-TFIIIC sites and Pol III-associated TFIIIC sites. **(E)** TFTC-3 and Pol III (RPC-1) FE scores at the 1,029 TFIIIC sites. Extra- and Pol III-associated TFIIIC sites are indicated (see **Methods**). r, Pearson correlation coefficient within the TFIIIC subclasses. **(F)** TFTC-3, TFTC-5, Pol III (RPC-1), TBP (TBP-1) ChIP-seq FE signals at extra- and Pol III-associated TFIIIC sites. **(G)** Fraction of TFIIIC sites harboring the transcription start site of non-coding RNA genes within +/–100 bp of the TFIIIC site center. Dotted lines indicate non-coding RNA gene classes found in ≥10 TFIIIC sites. Pol III-TFIIIC, Pol III-associated TFIIIC sites. **(H)** DNA sequence motifs found in +/–75 bp of TFIIIC site centers. **(I)** Fraction of TFIIIC sites overlapping repetitive elements. P, empirical P-value based on 2,000 permutations of the TFIIIC sites. Extra-TFIIIC and Pol III-associated were individually permutated within chromosomes. Dotted lines indicate repetitive element classes with $P < 0.001$.

Figure 2

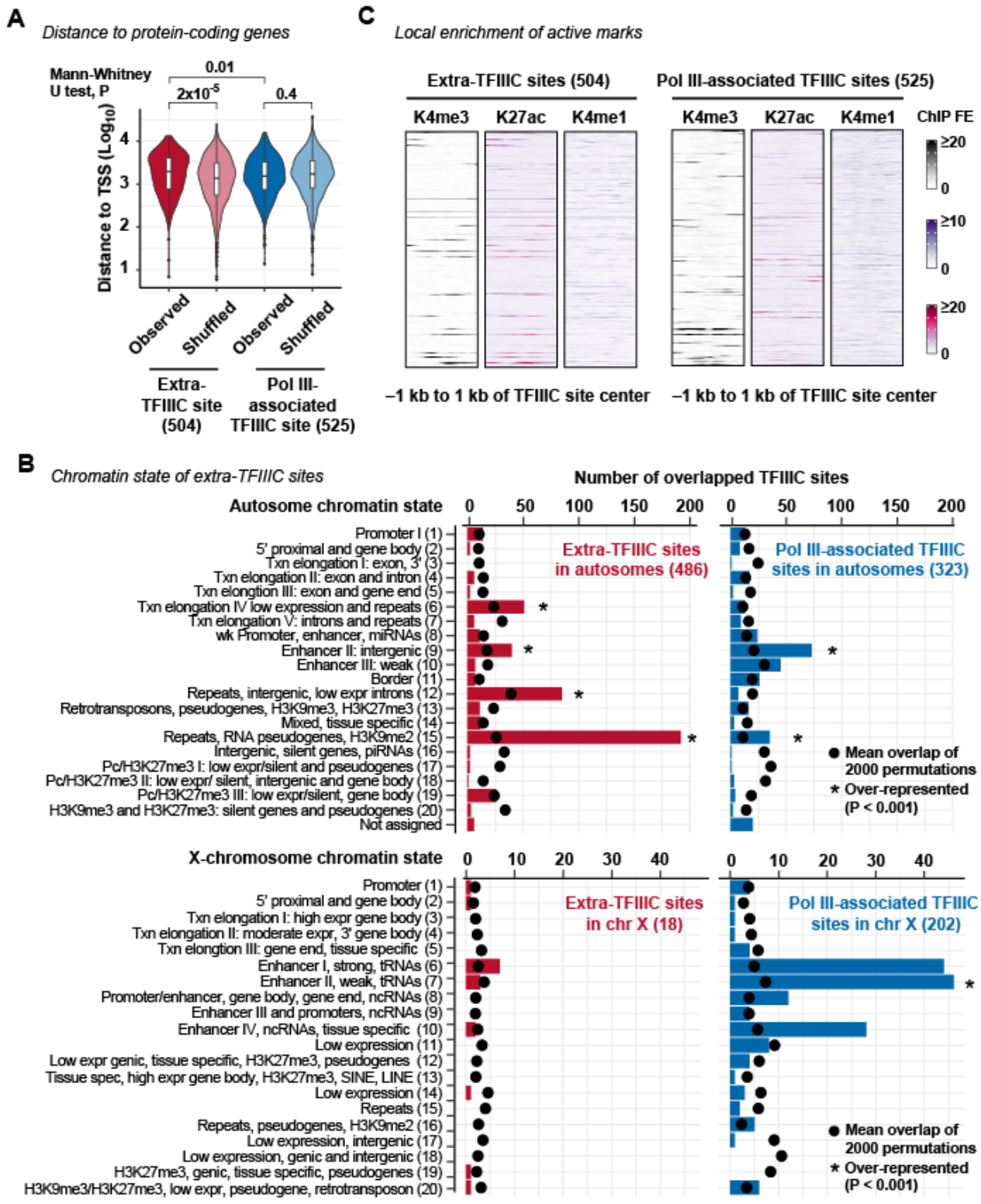


FIGURE 2. Extra-TFIIIC sites do not coincide with protein-coding gene promoters or enhancers
(A) Distance between TFIIIC site center and the transcription start site of protein-coding genes. Observed, the observed distance. Shuffled, the distance of a permutated set of TFIIIC sites. **(B)** Chromatin state annotation of TFIIIC sites. The chromatin state annotation reported by Evans et al. (2016) was used. To obtain mean overlap of permutations, extra- and Pol III-associated TFIIIC sites were individually permutated 2,000 times within

chromosomal domains. P, empirical P-value based on 2,000 permutations of the TFIIIC sites. **(C)** H3K4me3, H3K27ac, and H3K4me1 fold-enrichment scores (FE) at TFIIIC sites.

Figure 3

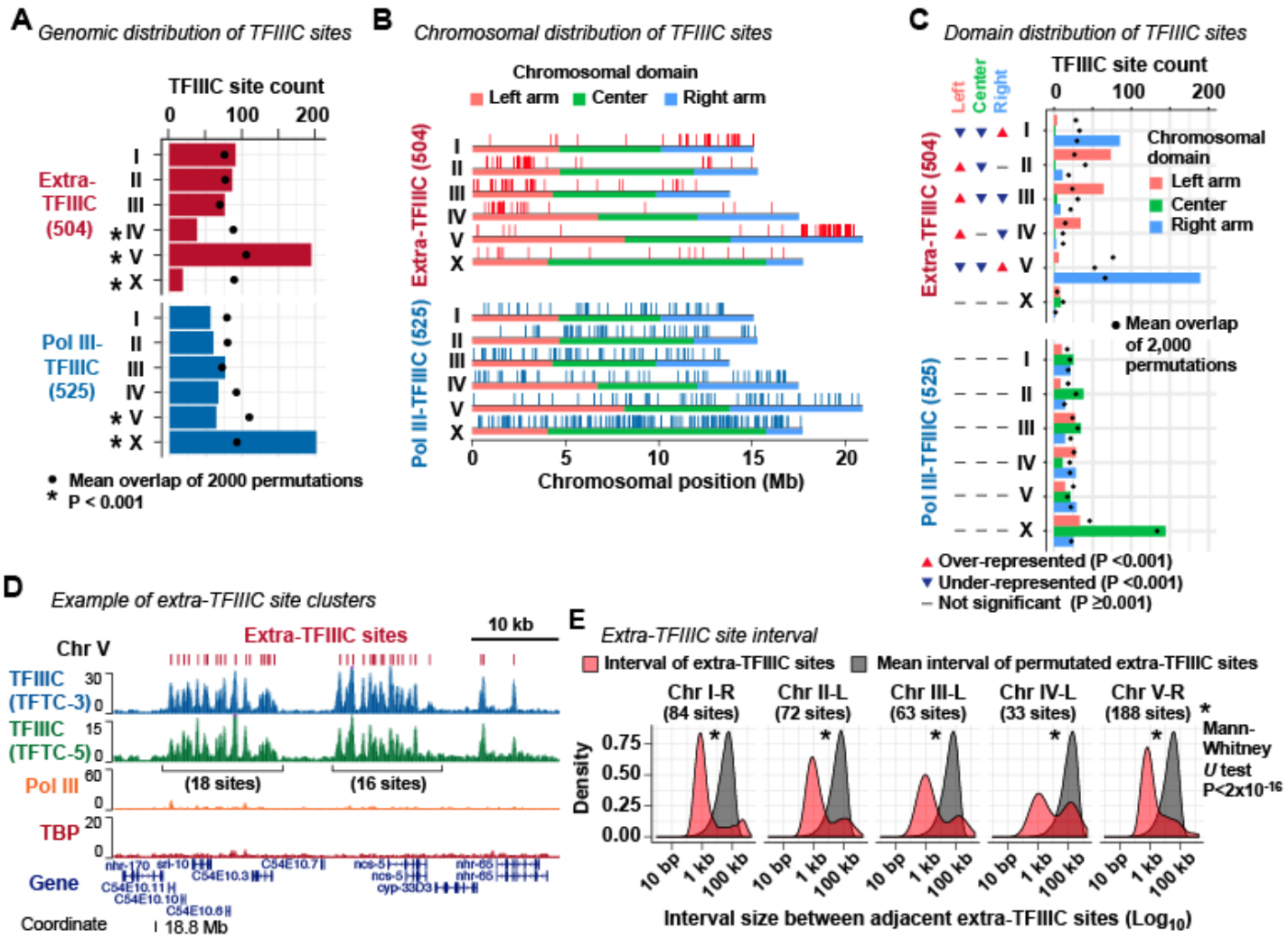


FIGURE 3. Extra-TFIIIC sites are clustered in autosomal arms

(A) Distribution of TFIIIC sites across chromosomes. To obtain mean overlap of permutated TFIIIC sites, extra- and Pol III-associated TFIIIC sites were individually permutated 2,000 times across the genome. P, empirical P-value based on 2,000 permutations of the TFIIIC sites. **(B)** Distribution of TFIIIC sites within chromosomes. Horizontal color bars indicate chromosomal domains defined by Rockman and Kruglyak (2009). **(C)** Fraction of TFIIIC sites within chromosomal domains. To obtain mean overlap of permutated TFIIIC sites, extra- and Pol III-associated TFIIIC sites were individually permutated 2,000 times within chromosomes. P, empirical P-value based on 2,000 permutations of the TFIIIC sites. **(D)** A representative genomic region including clusters of extra-TFIIIC sites. **(E)** Red, distribution of the interval size between two adjacent extra-TFIIIC sites. Gray, distribution of the mean of the interval sizes of permutated TFIIIC sites (2,000 means representing 2,000 permutations). Permutations were performed within chromosomal domains.

Figure 4

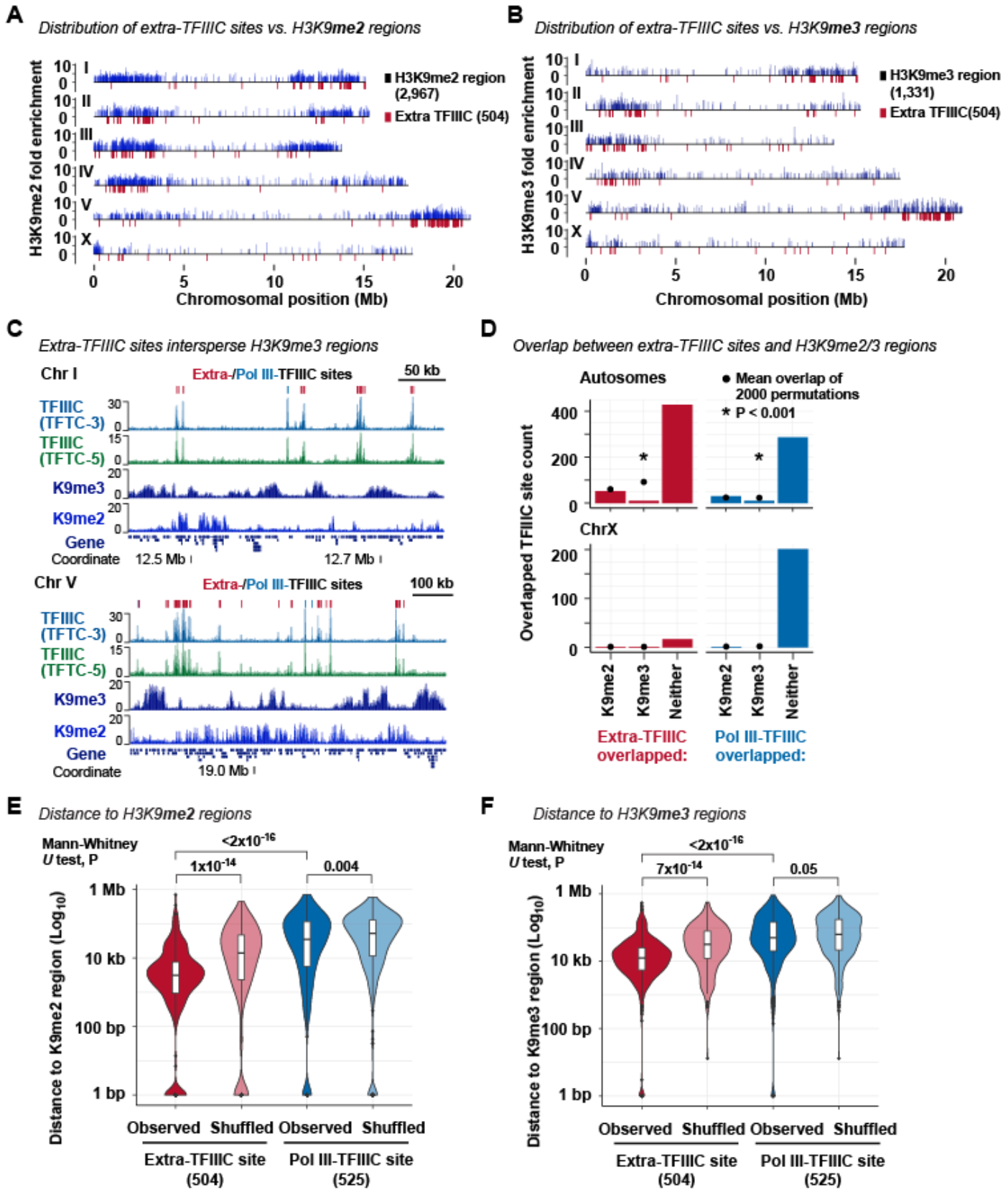


FIGURE 4. Extra-TFIIIC sites are located adjacent to H3K9me3-enriched regions

(A) Distribution of TFIIIC sites along with that of H3K9me2-enriched regions. **(B)** Distribution of TFIIIC sites along with that of H3K9me3-enriched regions. **(C)** Representative genomic regions showing extra-TFIIIC sites

interspersing H3K9me3-enriched regions. **(D)** Fraction of TFIIIC sites overlapping H3K9me2/H3K9me3-enriched regions. Because the H3K9me2/H3K9me3-enriched regions are infrequent in the X chromosome, the X chromosome is plotted separately. To obtain mean overlap of permutated TFIIIC sites, extra- and Pol III-associated TFIIIC sites were individually permutated 2,000 times within chromosomal domains. P, empirical P-value based on 2,000 permutations of the TFIIIC sites. **(E)** Distance between TFIIIC site center and H3K9me2-enriched regions. Observed, the observed distance. Shuffled, distance between a permutated set of TFIIIC sites and H3K9me2-enriched regions. Permutation was performed within chromosomal domains. **(F)** Same as (E), except that distance to H3K9me3-enriched regions is plotted and that permutation was performed excluding H3K9me3-enriched regions to reflect the lack of observed TFIIIC sites within H3K9me3 regions (D).

Figure 5

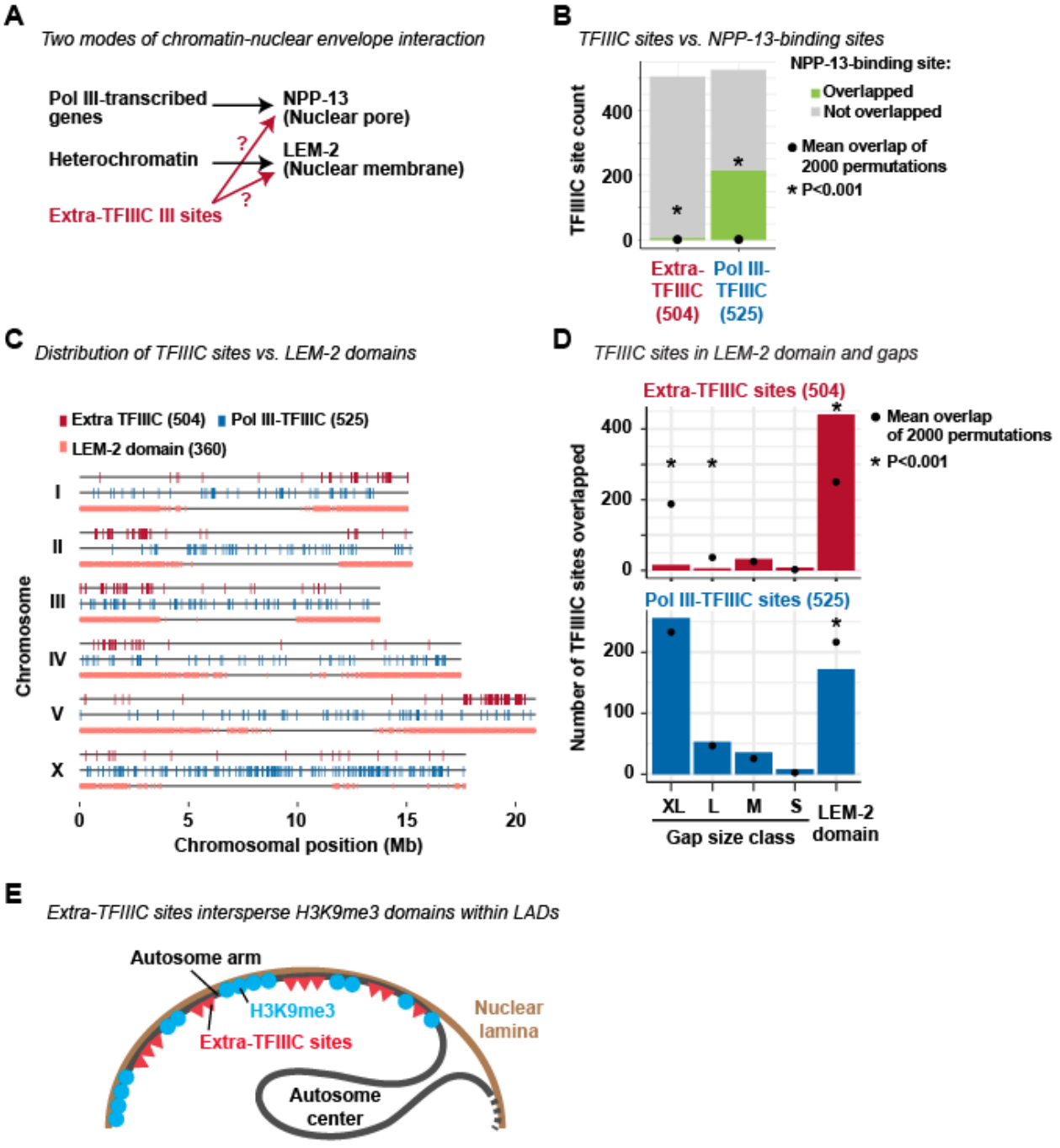


FIGURE 5. Extra-TFIIIC sites reside in LEM-2-associated domains

(A) Two hypothetical scenarios in which extra-TFIIIC sites associate with the nuclear periphery. **(B)** Fraction of TFIIIC sites overlapping NPP-13-binding sites defined by Ikegami and Lieb (2013). To obtain mean overlap of permutated TFIIIC sites, extra- and Pol III-associated TFIIIC sites were individually permutated 2,000 times within chromosomal domains. P, empirical P-value based on 2,000 permutations of the TFIIIC sites. **(C)** Distribution of TFIIIC sites along with that of LEM-2 domains defined by Ikegami et al. (2010). **(D)** Fraction of TFIIIC sites overlapping LEM-2 domains and gaps of various sizes between LEM-2 domains. To obtain mean overlap of permutated TFIIIC sites, extra- and Pol III-associated TFIIIC sites were individually permutated 2,000 times within chromosomal domains. P, empirical P-value based on 2,000 permutations of the TFIIIC

sites. **(E)** Model for chromosomal organization where extra-TFIIIC sites are clustered in the autosome arms corresponding LEM-2 domains and intersperse H3K9me3-enriched regions.