

Widespread 3D genome reorganization precedes programmed DNA rearrangement in *Oxytricha trifallax*

Danylo J. Villano¹, Manasa Prahlad^{1,2}, Ankush Singhal³, Karissa Y.

Sanbonmatsu^{3,4} and Laura F. Landweber¹

¹Departments of Biochemistry & Molecular Biophysics and Biological Sciences, Columbia University,
New York, NY 10032, USA

²Department of Neurobiology & Committee on Genetics, Genomics, and Systems Biology, University of Chicago,
Chicago, IL 60637, USA

³Theoretical Biology and Biophysics, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM
87545, USA.

⁴New Mexico Consortium, Los Alamos, NM 87544

Abstract

Genome organization recapitulates function, yet ciliates like *Oxytricha trifallax* possess highly-specialized germline genomes, which are largely transcriptionally silent. During post-zygotic development, *Oxytricha*'s germline undergoes large-scale genome editing, rearranging precursor genome elements into a transcriptionally-active genome with thousands of gene-sized nanochromosomes. Transgenerationally-inherited RNAs, derived from the parental somatic genome, program the retention and reordering of germline fragments. Retained and eliminated DNA must be distinguished and processed separately, but the role of chromatin organization in this process is unknown. We developed tools for studying *Oxytricha* nuclei and apply them to map the 3D organization of precursor and developmental states using Hi-C. We find that the precursor conformation primes the germline for development, while a massive spatial reorganization during development differentiates retained from eliminated regions before DNA rearrangement. Further experiments suggest a role for RNA-DNA interactions and chromatin remodeling in this process, implying a critical role for 3D architecture in programmed genome rearrangement.

Keywords: Genome editing; genome evolution; ciliate; Hi-C; chromatin organization

Introduction

Genome function depends heavily on chromatin organization. 3D structural features often recapitulate transcriptional or chromatin state and can follow highly dynamic patterns during the cell cycle.¹⁻⁶ Many features persist not just from one generation to the next but also across evolutionary time scales.⁷⁻⁹ The nuclear landscape surrounding DNA can be as important to proper genome function as the DNA sequence itself.

A hierarchy of 3D chromatin features has been identified, but the relative importance and universality of these organizing principles are unclear.¹⁰ Eukaryotic 3D genome structures typically share a dependence on cohesin,¹¹ and organize around compartments or compartmental domains based on chromatin state.^{9,12,13} However, varied patterns of chromosome folding have emerged across evolution.¹⁴ At a smaller scale, topologically-associating domains (TADs)¹⁵ may enclose conserved genes and persist across cell-type differentiation and speciation^{16,17} and their disruption may lead to cell dysfunction and disease.^{3,18} But the prominence of TADs outside metazoa is an open question,¹⁹⁻²¹ and their size and structure can range considerably, even from cell-to-cell.^{22,23} The factors that shape these structures also vary outside mammalian genomes, in which TAD boundaries often form through the blocking of cohesin-mediated loop extrusion by CCCTC-binding factor (CTCF).²⁴⁻²⁶ Preferential binding of CTCF provides a basis upon which natural selection can act,²⁷ but in organisms that lack CTCF,²⁸⁻³¹ or in those like *Drosophila* where it plays a less central role, a wide range of chromatin-associated proteins mediate structural features.^{32,33} Compartmental partitioning correlates with transcriptional state but remains even in the absence of transcription;³⁴ compartmental domains may form via phase separation, derived from biophysical properties of chromatin.³⁵ These previous studies suggest a wide, diverse adaptability in the relationship between genome structure and genome function.

The genome dynamics of the ciliate *Oxytricha trifallax* create unusual demands on genome architecture. *Oxytricha* is a binucleate, unicellular protist with genome functions split between two distinct types of nuclei.^{36,37} All gene expression during vegetative (asexual) growth derives from a somatic genome in a large macronucleus (MAC), comprising over 18,000 highly polyploid, kilobase-scale “nanochromosomes,” most bearing a single gene, two telomeres, and very little non-coding DNA.^{38,39} During sexual reproduction, however, no somatic DNA is inherited from parent to daughter cell. Instead, *Oxytricha* possesses a diploid, transcriptionally-silent germline genome with megabase-scale chromosomes in a micronucleus (MIC), which, at the onset of its sexual lifecycle, undergoes meiosis before fusing with another haploid MIC of a conjugating cell.⁴⁰ A copy of the resulting zygotic MIC develops into a new MAC over 72 hours, while the parental MAC degrades.

Though other transcriptionally silent genomes do demonstrate higher-order, developmentally dynamic 3D patterns,⁴¹ the *Oxytricha* germline genome must act as the substrate for an intricate program of developmental genome rearrangement that discards roughly 90% of DNA, retaining and joining a set of *macronuclear-destined sequences* (MDSs) that together construct a new somatic genome.⁴⁰ MDSs that are syntenic in the MAC are interrupted by *internally eliminated sequences* (IESs) in the germline, and frequently out of order (“scrambled”) or tens of kilobases apart. During development, hundreds of thousands of MDSs must be recognized, marked for retention, protected from degradation, then recombined in the correct order before amplification to high copy number. Rearrangement occurs between short direct repeats at MDS ends and is guided by two classes non-coding RNA (ncRNA) — 27nt piRNAs and long “template RNAs,” respectively.⁴²⁻⁴⁴ These ncRNAs derive from the parental

MAC and provide transgenerational memory of which sequences to retain and the order in which to link them.

How 3D organization of chromatin facilitates rearrangement of the *Oxytricha* germline is unknown. It is as critical that some regions be degraded as that adjacent sequences be protected, and that syntenic MDSs be in close physical proximity. The complex linear arrangement of MDSs in the germline suggests a key role for chromatin organization in *Oxytricha* genome dynamics, and that understanding 3D architecture may illuminate the process of rearrangement. Studies in other ciliates suggest that histone modifications and chromatin state mediate the distinction between retained and eliminated sequences,^{45,46} and in a distantly-related ciliate with modest levels of rearrangement, *Tetrahymena thermophila*, the boundaries of retained germline sequences form TAD-like structures.⁴⁷ Only a minority of *Tetrahymena* sequences are eliminated, germline MDSs are in closer proximity and rarely disordered, and *Tetrahymena* has far fewer germline or somatic chromosomes than *Oxytricha*.⁴⁸ The role of germline chromatin organization in *Oxytricha*'s complex rearrangement has been unknown, in either the initial precursor state or the developing nuclei (called “anlagen”).

Here we present the higher-order 3D genome structure in the germline of *Oxytricha trifallax* in both its precursor and developmental states. We scaffold the previously fragmented germline genome assembly using 3D genomics and long-read sequencing. Through Hi-C, we identify TAD-like structures in the precursor germline that suggest the genome is poised for rearrangement even prior to development. We then present the wholesale reorganization of the genome at the onset of sexual development, which appears to facilitate rearrangement through spatial separation of eliminated and retained sequences. Additional experiments demonstrate that retained sequences increase in both chromatin accessibility and RNA-DNA interactions during

development. These results inform a model in which dynamic chromatin organization plays a critical role in mediating genome rearrangement.

Results

Separation of *Oxytricha* nuclei via FACS based on size and DNA content

In order to study the conformation of germline chromatin in the absence of MAC contamination, we first developed approaches to purify different types of *Oxytricha* nuclei. Previous methods for enriching *Oxytricha* germline DNA do not preserve nuclear integrity or provide sufficient purity.⁴⁹ Fluorescence-activated cell sorting (FACS) has recently been used to isolate nuclei of other ciliate species based on nuclear proteins.^{50,51} However, the size and DNA content of *Oxytricha* nuclei differ so vastly that sorting based on DAPI staining permits scalable purification of intact *Oxytricha* nuclei. Suspensions of formaldehyde-fixed asexually growing *Oxytricha* nuclei of strain JRB310 were sorted by FACS on the basis of particle size and DAPI signal. Clusters of putative large, DNA-rich MACs and small, comparatively DNA-poor MICs are readily distinguishable (Figures 1A and S1A). These populations were sorted, and their DNA assessed via qPCR for somatic-specific and germline-specific sequences, which confirmed the identity of each population and demonstrated the lack of MAC contamination in the MIC sample (Figure 1B). The expected difference in molecular weight of DNA extracted from each population was also confirmed (Figures S1B and S1C).

Suspensions of post-zygotic, developing *Oxytricha* nuclei were prepared and sorted in the same manner. Nuclei were harvested at 24 hours post-mixing of mating-compatible cells (24h), a timepoint early in development when germline DNA in the anlagen has been polytenized to approximately 32X⁵² but most DNA rearrangements have not yet begun,⁵³ and anlagen have

expanded to nearly the size of the old MAC.⁴² Meanwhile, degradation of the old MAC has not yet begun, leading to the observation of all three types of nuclei, using the same size and DAPI signal criteria as before (Figures 1C and S1D). These developmental nuclei were tested via qPCR using the same MIC- and MAC-specific sequences as before, confirming the presence of germline-specific DNA in the population of anlagen (Figure 1D).

3D and long-read scaffolding improves contiguity of the reference germline genome

Previous studies estimated 75-120 megabase-scale chromosomes in the *Oxytricha* germline genome.^{52,54} However, the published JRB310 germline reference genome, based on sequencing of size-selected DNA from MICs enriched via sucrose gradient, has more than 25,000 contigs, all less than 500 kb, covering a total of 496 Mb of sequence.⁴⁰ The previous inability to generate larger, chromosome-sized contigs derived from both the impurity of sucrose gradient-enriched MIC DNA⁴⁹ and the repetitiveness of the germline, which consists of over 30% repetitive sequences alongside extensive MDS paralogy.^{40,55} In order to evaluate 3D germline features, a reference of much higher contiguity than was previously available was needed.

We took a combined approach to scaffold the *Oxytricha* germline into larger contigs, using 3D genomics and long-read sequencing. Approximately 1 μ g genomic DNA was extracted from two biological replicates of FACS-sorted vegetative MICs, which were then sequenced via Oxford Nanopore (Figure S2).⁵⁶ In parallel, *in situ* Hi-C libraries were built (Figure S3) from two biological replicates of over 1,000,000 FACS-sorted vegetative MICs,⁵⁷ which were then sequenced via Illumina, yielding >500M paired-end reads. When Hi-C reads were mapped to the reference germline genome using Juicer,⁵⁸ the highly fragmented state of the assembly was clear,

as numerous small contigs interacted with one another at very high frequency, suggesting they were in fact located on the same chromosome (Figure 2A). To improve the assembly, we first used LRScaf to bridge original contigs using germline Nanopore data (Figure S4A).⁵⁹ Next, Hi-C reads were used to scaffold the product of LRScaf using 3D-DNA, based on the principle that intrachromosomal contacts are far more likely than interchromosomal contacts.⁶⁰ Output from 3D-DNA was visually inspected using Juicebox⁶¹ (Figure S4B), and obvious errors produced by Hi-C-based scaffolding were manually corrected in that program. 3D-DNA leaves scars of ambiguous bases between scaffolded contigs, so another round of LRScaf was performed for correction.

The final scaffolded assembly has substantially improved contiguity and 12 Mb additional sequence (Figure 2B). The initial 25,720 contigs is reduced nearly 5-fold. Despite the total contig count remaining over 5,000, most of these are very small and unable to be processed by 3D-DNA. 81% of the total genome is present on 123 contigs larger than 1 Mb, while 90% of the genome is contained within 358 contigs larger than 50 kb (Figures 2C and S4C). Two megabase-scale scaffolds contain telomeric repeats at both ends (Table 1). Aligning known MDSs from the somatic genome to the scaffolded germline genome via BLAST⁶² produced a similar mappability profile as when aligned to the original assembly (Figure S5A). Upon remapping vegetative germline Hi-C data to the scaffolded assembly, the number of intercontig interactions was reduced more than 40%, reflecting that most were actually long-range (>20 kb) intrachromosomal interactions. There remains an excess of interchromosomal contacts genome-wide, consistent with the current level of scaffolding compared to the estimated ~75 chromosomes in the MIC. If we consider only contacts with at least one read pair mapping to the thirteen scaffolds >5 Mb, then interchromosomal contacts are reduced more than 69% relative to

the original fragmented assembly (Figure S5B). This more complete understanding of the 1D architecture of the *Oxytricha* germline now permits analysis of its 3D architecture.

TAD-like structures in the precursor germline cluster retained sequences

In order to learn how MDSs are organized in vegetative micronuclei, germline Hi-C data mapping to 123 scaffolds longer than 1 Mb (covering 81% of total germline sequence) were extracted from the contact matrix and analyzed with HiCExplorer (see Methods). After KR matrix balancing,⁶³ from a genome-wide view there was little interchromosomal structure (Figures 3A and S6A), with scaffolds resembling classical chromosome territories⁶⁴ and low compartmentalization (Figure S6B). Long-range loops could not be reliably detected in this largely heterochromatic genome,³⁶ due in part to the resolution of the data. Calling TADs at resolutions of 5, 10, and 100 kb, however, revealed the presence of dozens of TAD-like structures (Figure 3B).

Ranging in size from 100 kb to 3.2 Mb, these domains were found to be highly enriched in MDSs (Figure 3C; $p < 10^{-14}$, Student's *t*-test), which tend to occur in dense tracts in the germline (Figure S7), such that terminal MDSs of different nanochromosomes often overlap.⁴⁰ The TAD-like structures have a median MDS content of 26%, compared to 11% for the germline as a whole (Figure S6C). Only 7/74 TAD-like structures had less than the genome-wide average MDS content. After splitting the entire germline genome into 100 kb regions, 46% of the $\geq 90^{\text{th}}$ -percentile most MDS-rich bins at least partially overlapped with the TAD-like structures, including 20 of the 30 richest bins (Figure S6D). In total, over 25% of MDSs were contained within these subchromosomal 3D features (Figure 3D).

TADs are preferentially self-associating regions of chromatin; sequences within TADs are more likely to interact with one another than expected based purely on their 1D distance. In transcriptionally active genomes, TADs permit genes to make contact with distant enhancers.⁶⁵ The *Oxytricha* germline has no need of enhancers, but analogously, its TAD-like structures similarly cluster sequences that benefit from proximity. Co-localization of MDSs reduces the nuclear space in which rearrangement machinery needs to find its substrates, and may facilitate specific interactions between distant MDSs that ultimately will be linked. The 3D co-localization of *Oxytricha* MDSs approximates the linear clustering of MDSs throughout the genome, and suggests that germline chromatin organization anticipates the spatial needs of genome rearrangement, even in the absence of development.

Chromatin dramatically reorganizes during development, distinguishing retained and eliminated regions

We then investigated how 3D conformation changes with the onset of development to permit wholesale genome rearrangement. Nuclear development requires approximately 72 hours to complete. At 24h, the genome is on the cusp of rearrangement. ncRNAs that program genome rearrangement are abundant in the anlagen,^{42,66} and germline DNA has been polytenized to ~32X,⁵² but massive DNA editing has not yet occurred.³⁷ Two biological replicate Hi-C libraries (Figure S8) were prepared from > 250,000 FACS-purified 24h anlagen. Illumina sequencing produced 570 million paired-end reads, which were mapped to the scaffolded reference genome with Juicer,⁵⁷ then corrected and analyzed with HiCEXplorer.^{31,67,68}

24 hour Hi-C contacts across all scaffolds ≥ 1 Mb revealed a striking shift in 3D genome architecture (Figure S9A and S9B). Among the largest scaffolds, *interchromosomal* contact

frequency tripled compared to the precursor, vegetative state, along with an enormous (>5-fold) reduction in the frequency of long-range *intrachromosomal* contacts > 20kb (Figures S9C and S9D). With the loss of precursor chromosome territory-like organization, the structure of vegetative TAD-like structures was also largely absent, and fewer than 11% could be called at 24h (Figure S10). We instead detected the emergence of a dramatic 3D architectural bifurcation of the genome. Whereas there were no apparent compartments in the vegetative germline, the signature checkerboard pattern reflecting compartmentalization⁶⁹ arises in the 24h anlagen (Figure 4A), confirmed by compartmentalization analysis (Figure 4B).⁷⁰ The 3D profile of the two compartments is starkly different. While interchromosomal interactions are enriched genome-wide at 24h, they are 44% more frequent in compartment A vs. compartment B bins (based on the sign of the first eigenvector) (Figure 4C). Inter- and intrachromosomal interactions are roughly equal in compartment B, but in compartment A, which is highly enriched in MDSs, nearly 90% of contacts are *interchromosomal*. Though the genome is roughly equally divided between each compartment (51% vs. 49%), retained segments are more than twice as likely to be located in compartment A than compartment B (Figure 4D). Unlike in the vegetative germline, MDS content correlates well with PCA1, and over 80% of the 30 most MDS-enriched 100 kb bins are found in compartment A (Figures S11A and S11B). DNA fluorescence *in situ* hybridization (Figure S12)^{71,72} and multi-chromosomal 3D reconstructed models (using a modified version of the 4DHiC method,⁷³ allowing for multiple chromosome interactions⁷⁴) of the experimental Hi-C data (Figure S13)⁷⁵⁻⁷⁷ corroborate nonuniform subnuclear localization of developmental compartments and increased interchromosomal mixing during development relative to the precursor vegetative state.

We then investigated whether this radically altered 3D architecture persisted later into genome rearrangement. At 48h, massive DNA editing is underway and rearrangement intermediates are abundant, as somatic sequences are rebuilt before the final stage of DNA amplification. Hi-C was performed on ~80,000 sorted 48h anlagen. When sequences were aligned and analyzed as before, despite lower coverage than the previous libraries (Table S1), the 48h anlagen present marked changes (Figures S14A and S14B). A checkerboard pattern can still be discerned when compared to the vegetative germline (Figure S14C), but weaker than at 24h (Figure S14D). The frequency of interchromosomal contacts (Figure S9C), the compartmentalization signal (Figures S15A and S15B) and the correlation of PCA1 with MDS content (Figure S11C) were all intermediate between the precursor germline and the 24h anlagen. Although the 48h late developmental genome shares with the 24h early developmental genome a dearth of long-range *intrachromosomal* contacts relative to the germline, and although the developmental samples correlate well (Figure S15C), the 48h developmental genome shows a striking increase in short range (< 20 kb) contact frequency, twice that of the precursor germline (Figure S9C), largely at the expense of interchromosomal contacts (Figure S15D). The ongoing deletion of germline-limited regions at 48h complicates this analysis, but this late-stage focus on short-range contacts agrees with the timeline of rearrangement. Somatic syntenic sequences tend to cluster in the germline,⁴⁰ and by 48h the ncRNA-guided joining of consecutive MDSs is underway. Earlier, at 24h, the reorganization of the germline genome into compartments drives MDSs non-specifically across the genome into a shared nuclear space, with MDSs marked for retention and protected from degradation.⁴² As genome rearrangement progresses during development, the 3D architecture transitions from bulk reorganization to promoting smaller-scale interactions.

Retained and eliminated sequences are distinguished by chromatin state

The widespread shift in germline 3D genome architecture during development must ultimately distinguish retained vs. eliminated sequences, allowing the former to interact with each other and insulating the latter. Such chromatin distinctions are a hallmark of differences between compartments in metazoan genomes,²⁶ and have been observed during development in other ciliates.^{45,78} Since disparities in chromatin accessibility are a primary functional difference between chromatin states,⁷⁹ we measured chromatin accessibility in the *Oxytricha* germline by performing ATAC-seq⁸⁰ on sorted nuclei from vegetative and 24h developmental cells (see Methods).

Focusing on features of intermediate size, we observed early in development that retained regions, i.e. MDSs, were far more accessible than deleted regions (IESs; Figure 5A; $p < 10^{-10}$, Student's *t*-test). This relationship was present in the precursor germline as well, as IESs were nearly absent in the germline ATAC-seq data, but MDSs are much more accessible during development. ATAC-seq corroborated the developmental bifurcation of the genome, with 100 kb bins in compartment A averaging nearly twice the accessibility of bins in compartment B at 24h (Figure 5B; $p < 10^{-13}$, Student's *t*-test). These regions had little difference in the precursor MIC. Peak calling in ATAC-seq coverage displays MDS enrichment in both vegetative and developmental samples, though moreso in the latter (Figure 5C). Curiously, telomere-bearing elements (TBEs) — semi-domesticated transposable elements that are transiently expressed from the developing anlagen at 24h before self-elimination^{81,82} and represent approximately 15% of the MIC genome⁴⁰ — are also highly enriched in anlagen peaks and depleted in the vegetative data. Averaging the coverage over thousands of germline features reveals the precision of these

differences in accessibility. In both states, accessibility increases sharply at MDS boundaries (Figure 5D); conversely, accessibility drops precisely within IESs (Figure 5E). Likewise, TBE coverage spikes at either end of the transposon during development (Figure 5F), which may reflect its active transcription.⁸² These results suggest a clear contrast between the chromatin state of retained vs. eliminated sequences in the *Oxytricha* germline, accompanying their 3D separation.

RNA mediates long-range DNA interactions during development

As in other ciliates, sequence retention and elimination in *Oxytricha* development is regulated by transgenerationally-inherited ncRNA.⁸³ piRNAs and template RNAs, both of which derive from genome-wide transcription of the parent's somatic genome, peak in expression prior to the 24h timepoint at which we observe compartmentalization.^{42,44} The specific mechanism by which these ncRNAs lead to retention is unknown, but their precise targeting of retained sequences makes them strong candidates to contribute to the developmental spatial separation between MDSs and eliminated regions. Several recent studies have demonstrated the influence of ncRNA on 3D genome architecture, through either binding of nascent transcripts or RNA-DNA interactions.⁸⁴⁻⁸⁶ In light of the massive processing of all MDSs and the lack of widespread transcription from the anlagen,⁶⁶ we conjectured that ncRNA may be mediating germline development through direct RNA-DNA interactions.

In order to test whether RNA may be directly binding MDSs during development and thereby supporting chromatin reorganization, we measured RNA-DNA hybrid abundance using the S9.6 antibody (Figure 6A).⁸⁷ We observed increased levels of RNA-DNA hybrid formation during *Oxytricha* development, relative to the precursor, vegetative germline, peaking at 24h and

persisting through the end of genome rearrangement. To map the locations of these hybrids, DNA-RNA immunoprecipitation sequencing (DRIP-seq)⁸⁸ was performed on whole cell DNA from precursor, vegetative cells and developing *Oxytricha*. The resulting reads were mapped to the somatic genome, which provides the source of developmental ncRNA.^{42,43} Before development, hybrids in asexually dividing cells were found to be enriched at both the 5' and 3' ends of single-gene nanochromosomes (Figure 6B). Because transcription start sites (TSS) and termination sites (TTS) are very close to nanochromosome ends,⁵⁴ these hybrids are likely transcription-associated, as transcriptional R-loops tend to form at the TSS and TTS.⁸⁹ However, DRIP-seq at the 24h timepoint shows elevated coverage across all MDSs, regardless of distance from nanochromosome ends (Figure 6B). This ubiquitous hybridization of RNA across retained MDS sequences during development suggests direct hybridization of MAC-derived ncRNA to the developing, anlagen DNA.

To test whether RNA provides scaffolds to assist the developmental compartmentalization of the *Oxytricha* germline genome, we treated 24h developing cells with a robust RNase treatment to digest the RNA strand in RNA-DNA hybrids, as well as all single- and double-stranded RNA, prior to fixation and FACS, alongside control, permeabilized cells that were not RNase treated. Hi-C on RNase-treated, sorted anlagen still displayed a qualitatively similar 3D checkerboard pattern relative to the control (Figure 6C), indicating that compartmentalization was not strongly affected by the treatment (Figures S16A and S16B). This suggests that by 24h, RNA is no longer critical for the overall maintenance of this chromatin superstructure. Though interchromosomal interactions are still elevated after RNase treatment, relative to the precursor germline, long-range intrachromosomal interactions were reduced more than twofold relative to the control (Figure S16C). Since template RNAs derive from somatic

nanochromosomes, and most MDSs that form the same nanochromosome are syntenic in the germline, we hypothesized that RNA could be mediating specific interactions between precursor MDSs that contribute to the same MAC chromosome. Binning the Hi-C data at 10 kb, we identified the 400 most distant pairs of MDSs that are located on the same contig in both the somatic and germline genomes, since enrichment of interactions between these genomic bins are most easily measurable. Comparing RNase-treated anlagen to the control, we found an average of almost twofold reduction in the frequency of these distant intrachromosomal interactions, and over 75% of cases decreased with RNase treatment (Figure 6D). This result, along with the increased level of RNA-DNA hybrids, supports a model in which RNA may directly scaffold rearranging DNA⁴⁴ to facilitate the colocalization of MDSs that recombine to form a mature nanochromosome during development, exploiting the broader context of 3D chromatin architecture.

Discussion

Oxytricha trifallax offers a direct yet unconventional relationship between genome structure and function. We show that dramatic chromatin reorganization precedes massive genome rearrangement. The dynamic architecture of its genome, with transcriptionally-silent and often scrambled, precursor gene segments representing the functional units of DNA, foreshadows its extensive remodeling. MDS-rich 3D domains in the precursor germline micronucleus poise the genome for development, and then extensive compartmentalization spatially separates predominantly retained from eliminated regions early in development, followed by a coalescence of short-range interactions late in development, when new macronuclear chromosomes form.

In the distantly-related ciliate *Tetrahymena*, with considerably less genome rearrangement, the boundaries of germline TAD-like structures coincide with chromosome breakage sites during development,⁴⁷ and in mammalian systems, the boundaries of a majority of TADs are demarcated by CTCF-binding sites.⁹⁰⁻⁹² *Drosophila* has several DNA-binding insulator proteins that together fill a similar role,³¹ but studies in representative plants, fungi, and protists with higher order chromatin structure have yet to identify functionally similar architectural proteins.^{30,93} *Oxytricha*, with orders of magnitude more breakage sites than *Tetrahymena*, also lacks a CTCF homolog; however, other master regulators of 3D genome architecture may play a prominent role in its genome dynamics.⁹⁴

Compartmentalization of *Oxytricha*'s germline genome during development appears to self-organize around the location of retained sequences, as MDSs are driven into shared nuclear spaces that exclude MDS-depleted regions. This reduction in complexity would encourage productive searching for and joining of consecutive MDSs. It is also unknown whether interactions occur between multiple or single copies of polytenized germline chromosomes. The ATAC-seq data suggest that compartmentalization could depend on homotypic chromatin state interactions. Histone or DNA modifications can be especially dynamic during meiosis and development⁹⁵ and limited histone data are available for the ciliate *Stylonychia*,^{45,96} but little is known about their role in *Oxytricha* development.

The role of ncRNA in *Oxytricha* genome reorganization is particularly intriguing, as two classes of ncRNA provide a code for sequence retention.⁴²⁻⁴⁴ How ncRNA specifically transmits the information it carries to program genome editing is a major question in *Oxytricha* biology. Like bridge RNAs,⁹⁷ *Oxytricha*'s long template RNAs may provide a scaffolding agent for the rearranging genome, directly mediating thousands of long-range DNA-DNA interactions, while

piRNAs may act indirectly to guide changes in MDS chromatin state. *Oxytricha* stands out among the main ciliate model systems in that its largest class of small RNAs mark retained rather than eliminated sequences, although MDSs occupy a smaller fraction of the *Oxytricha* genome, and in each species the most abundant class of small RNAs typically targets the minority class of sequences for DNA retention or deletion.⁴² For example, in *Tetrahymena*, scnRNA targeting of eliminated sequences leads to histone H3 lysine-9 methylation,⁴⁶ which can cause heterochromatin formation in an RNA-guided manner.⁹⁸ In *Oxytricha*, piRNA targeting of retained regions could lead to a chromatin mark that prevents heterochromatinization and promotes chromatin accessibility and compartmentalization, ultimately ensuring retention.

Hi-C, itself, has limitations, starting with the depth of DNA sequencing, the complexity of DNA recovered after sorting nuclei, and the inability to uniquely map short reads across a complex germline genome. Future work may improve the yield of FACS-based nuclear sorting, incorporate multimapping reads through probabilistic mapping, or obtain higher resolution through variations like Micro-C.⁹³ Currently, we must draw inferences on how MDSs are organized without the ability to analyze their specific pairwise interactions or call loop contacts, instead relying on binning the genome and analyzing the enrichment of MDSs in specific regions. Future work may confirm whether all or merely a large majority of MDSs are driven into dense colocalization early in development.

Plasticity is also a biological feature of *Oxytricha* development, sometimes permitting combinatorial rearrangement of MDSs to produce new genes.⁹⁹ Polytenization of germline DNA may provide more opportunity for successful rearrangements, as well as room for innovation. Moreover, the compartmental distinction between retained vs. eliminated sequences does not account for the most granular requirements of genome rearrangement. Long stretches of

eliminated sequence can be insulated and segregated from retained sequences by compartmentalization early in development, but how are the shortest IESs identified and removed if they are no larger than introns and interspersed between MDSs in compartment A? Multiple waves of DNA elimination might use different molecular mechanisms for depletion of megabase-long stretches of germline-restricted DNA versus precise excision of small IESs. While a full understanding of *Oxytricha* genome rearrangement is still coming to light, our work demonstrates a tight-knit relationship between genome structure and function, and how adaptable and flexible that relationship can be in the face of complex DNA dynamics.

Acknowledgements

We thank Sheela George and Ryan Tran for laboratory support. We thank David Dai, Eoghan Harrington, John Beaulaurier, Lynn Ly, and Sissel Juul at Oxford Nanopore Technologies in New York for providing sequencing and advice. We thank Michael Kissner at the Columbia Stem Cell Initiative Flow Cytometry Institutional Core Facility for providing FACS training and advice. We also thank Brian Beliveau, Bill Jack, Stavros Lomvardas, Lorraine Symington, and all current and past Landweber lab members for discussions about genome architecture and *Oxytricha* biology. This study used the Confocal and Specialized Microscopy Shared Resource of the Herbert Irving Comprehensive Cancer Center at Columbia University, funded in part through NIH/NCI Cancer Center Support Grant P30CA013696. This study was supported by NIH R35-GM122555 and NSF 1764366 to LFL. DJV also received support from 5T32GM008798. KS and AS are grateful to generous support from Los Alamos National Laboratory Institutional Computing for allocations on LANL Chicoma, and LANL Venado. KS and AS thank the generous support from NSF (MCB 2337393), DOE LANL LDRD

(20210082DR and 20210134ER) and the U.S. Department of Energy, Office of Science, through the Biological and Environmental Research (BER) and the Advanced Scientific Computing Research (ASCR) programs under contract number 89233218CNA000001 to Los Alamos National Laboratory (Triad National Security, LLC).

Author Contributions

DJV designed, performed, and analyzed most of the experiments. MP contributed to experiments on RNA-DNA hybrids. AS and KS designed and performed coarse-grained modeling of Hi-C data. DJV and LFL conceived the study and wrote the paper with input from other authors.

References

- 1) Ghavi-Helm Y, Klein FA, Pakozdi T, Ciglar L, Noordermeer D, Huber W, Furlong EE. Enhancer loops appear stable during development and are associated with paused polymerase. *Nature*. 2014 Aug 7;512(7512):96-100. doi: 10.1038/nature13417. Epub 2014 Jul 2. PMID: 25043061.
- 2) Hsieh TS, Cattoglio C, Slobodyanyuk E, Hansen AS, Rando OJ, Tjian R, Darzacq X. Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Mol Cell*. 2020 May 7;78(3):539-553.e8. doi: 10.1016/j.molcel.2020.03.002. Epub 2020 Mar 25. PMID: 32213323; PMCID: PMC7703524.
- 3) Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, Santos-Simarro F, Gilbert-Dussardier B, Wittler L, Borschiwer M, Haas SA, Osterwalder M, Franke M, Timmermann B, Hecht J, Spielmann M, Visel A, Mundlos S. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*. 2015 May 21;161(5):1012-1025. doi: 10.1016/j.cell.2015.04.004. Epub 2015 May 7. PMID: 25959774; PMCID: PMC4791538.
- 4) Mach P, Kos PI, Zhan Y, Cramard, J, Gaudin S, Tünnermann J, Marchi E, Eglinger J, Zuin J, Kryzhanovska M, Smallwood S, Gelman L, Roth G, Nora EP, Tiana G, Giorgetti L. Cohesin and CTCF control the dynamics of chromosome folding. *Nat Genet* 54, 1907–1918 (2022). <https://doi.org/10.1038/s41588-022-01232-7>.
- 5) Nagano T, Lubling Y, Várnai C, Dudley C, Leung W, Baran Y, Mendelson Cohen N, Wingett S, Fraser P, Tanay A. Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature*. 2017 Jul 5;547(7661):61-67. doi: 10.1038/nature23001. PMID: 28682332; PMCID: PMC5567812.
- 6) Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* 489, 109–113 (2012).
- 7) Fudenberg G, Pollard KS. Chromatin features constrain structural variation across evolutionary timescales. *Proc Natl Acad Sci U S A*. 2019 Feb 5;116(6):2175-2180. doi: 10.1073/pnas.1808631116. Epub 2019 Jan 18. PMID: 30659153; PMCID: PMC6369792.
- 8) Marchal C, Sima, J, Gilbert, DM. Control of DNA replication timing in the 3D genome. *Nat Rev Mol Cell Biol* 20, 721–737 (2019). <https://doi.org/10.1038/s41580-019-0162-y>.
- 9) Rowley MJ, Nichols MH, Lyu X, Ando-Kuri M, Rivera ISM, Hermetz K, Wang P, Ruan Y, Corces VG. Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol Cell*. 2017 Sep 7;67(5):837-852.e7. doi: 10.1016/j.molcel.2017.07.022. Epub 2017 Aug 17. PMID: 28826674; PMCID: PMC5591081.
- 10) McCord RP, Kaplan N, Giorgetti L. Chromosome Conformation Capture and Beyond: Toward an Integrative View of Chromosome Structure and Function. *Mol Cell*. 2020 Feb

20;77(4):688-708. doi: 10.1016/j.molcel.2019.12.021. Epub 2020 Jan 27. PMID: 32001106; PMCID: PMC7134573.

11) Sanborn AL, Rao SS, Huang SC, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J, Geeting KP, Gnirke A, Melnikov A, McKenna D, Stamenova EK, Lander ES, Aiden EL. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A*. 2015 Nov 24;112(47):E6456-65. doi: 10.1073/pnas.1518552112. Epub 2015 Oct 23. PMID: 26499245; PMCID: PMC4664323.

12) Falk M, Feodorova Y, Naumova N, Imakaev M, Lajoie BR, Leonhardt H, Joffe B, Dekker J, Fudenberg G, Solovei I, Mirny LA. Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature*. 2019 Jun;570(7761):395-399. doi: 10.1038/s41586-019-1275-3. Epub 2019 Jun 5. Erratum in: *Nature*. 2019 Aug;572(7771):E22. PMID: 31168090; PMCID: PMC7206897.

13) Wang S, Su JH, Beliveau BJ, Bintu B, Moffitt JR, Wu CT, Zhuang X. Spatial organization of chromatin domains and compartments in single chromosomes. *Science*. 2016 Aug 5;353(6299):598-602. doi: 10.1126/science.aaf8084. Epub 2016 Jul 21. PMID: 27445307; PMCID: PMC4991974.

14) Hoencamp C, Dudchenko O, Elbatsh AMO, Brahmachari S, Raaijmakers JA, van Schaik T, Sedeño Cacciatore Á, Contessoto VG, van Heesbeen RGHP, van den Broek B, Mhaskar AN, Teunissen H, St Hilaire BG, Weisz D, Omer AD, Pham M, Colaric Z, Yang Z, Rao SSP, Mitra N, Lui C, Yao W, Khan R, Moroz LL, Kohn A, St Leger J, Mena A, Holcroft K, Gambetta MC, Lim F, Farley E, Stein N, Haddad A, Chauss D, Mutlu AS, Wang MC, Young ND, Hildebrandt E, Cheng HH, Knight CJ, Burnham TLU, Hovel KA, Beel AJ, Mattei PJ, Kornberg RD, Warren WC, Cary G, Gómez-Skarmeta JL, Hinman V, Lindblad-Toh K, Di Palma F, Maeshima K, Multani AS, Pathak S, Nel-Themaat L, Behringer RR, Kaur P, Medema RH, van Steensel B, de Wit E, Onuchic JN, Di Pierro M, Lieberman Aiden E, Rowland BD. 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. *Science*. 2021 May 28;372(6545):984-989. doi: 10.1126/science.abe2218. PMID: 34045355; PMCID: PMC8172041.

15) Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, Gribnau J, Barillot E, Blüthgen N, Dekker J, Heard E. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*. 2012 Apr 11;485(7398):381-5. doi: 10.1038/nature11049. PMID: 22495304; PMCID: PMC3555144.

16) Harmston N, Ing-Simmons E, Tan G, Perry M, Merckenschlager M, Lenhard B. Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. *Nat Commun*. 2017 Sep 5;8(1):441. doi: 10.1038/s41467-017-00524-5. PMID: 28874668; PMCID: PMC5585340.

- 17) Krefting J, Andrade-Navarro MA, Ibn-Salem J. Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC Biol.* 2018 Aug 7;16(1):87. doi: 10.1186/s12915-018-0556-x. PMID: 30086749; PMCID: PMC6091198.
- 18) Franke M, Ibrahim DM, Andrey G, Schwarzer W, Heinrich V, Schöpflin R, Kraft K, Kempfer R, Jerković I, Chan WL, Spielmann M, Timmermann B, Wittler L, Kurth I, Cambiaso P, Zuffardi O, Houge G, Lambie L, Brancati F, Pombo A, Vingron M, Spitz F, Mundlos S. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature.* 2016 Oct 13;538(7624):265-269. doi: 10.1038/nature19800. Epub 2016 Oct 5. PMID: 27706140.
- 19) Eser U, Chandler-Brown D, Ay F, Straight AF, Duan Z, Noble WS, Skotheim JM. Form and function of topologically associating genomic domains in budding yeast. *Proc Natl Acad Sci U S A.* 2017 Apr 11;114(15):E3061-E3070. doi: 10.1073/pnas.1612256114. Epub 2017 Mar 27. Erratum in: *Proc Natl Acad Sci U S A.* 2017 Oct 10;114(41):E8801. PMID: 28348222; PMCID: PMC5393236.
- 20) Wang M, Wang P, Lin M, Ye Z, Li G, Tu L, Shen C, Li J, Yang Q, Zhang X. Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. *Nat Plants.* 2018 Feb;4(2):90-97. doi: 10.1038/s41477-017-0096-3. Epub 2018 Jan 29. PMID: 29379149.
- 21) Liu C, Wang C, Wang G, Becker C, Zaidem M, Weigel D. Genome-wide analysis of chromatin packing in *Arabidopsis thaliana* at single-gene resolution. *Genome Res.* 2016 Aug;26(8):1057-68. doi: 10.1101/gr.204032.116. Epub 2016 May 25. PMID: 27225844; PMCID: PMC4971768.
- 22) Bonev B, Cavalli G. Organization and function of the 3D genome. *Nat Rev Genet.* 2016 Oct 14;17(11):661-678. doi: 10.1038/nrg.2016.112. Erratum in: *Nat Rev Genet.* 2016 Dec;17(12):772. PMID: 27739532.
- 23) Stevens TJ, Lando D, Basu S, Atkinson LP, Cao Y, Lee SF, Leeb M, Wohlfahrt KJ, Boucher W, O'Shaughnessy-Kirwan A, Cramard J, Faure AJ, Ralser M, Blanco E, Morey L, Sansó M, Palayret MGS, Lehner B, Di Croce L, Wutz A, Hendrich B, Klenerman D, Laue ED. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature.* 2017 Apr 6;544(7648):59-64. doi: 10.1038/nature21429. Epub 2017 Mar 13. PMID: 28289288; PMCID: PMC5385134.
- 24) Guo Y, Xu Q, Canzio D, Shou J, Li J, Gorkin DU, Jung I, Wu H, Zhai Y, Tang Y, Lu Y, Wu Y, Jia Z, Li W, Zhang MQ, Ren B, Krainer AR, Maniatis T, Wu Q. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell.* 2015 Aug 13;162(4):900-10. doi: 10.1016/j.cell.2015.07.038. PMID: 26276636; PMCID: PMC4642453.
- 25) Kentepozidou E, Aitken SJ, Feig C, Stefflova K, Ibarra-Soria X, Odom DT, Roller M, Flicek P. Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains. *Genome Biol.* 2020 Jan 7;21(1):5. doi: 10.1186/s13059-019-1894-x. PMID: 31910870; PMCID: PMC6945661.

- 26) Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014 Dec 18;159(7):1665-80. doi: 10.1016/j.cell.2014.11.021. Epub 2014 Dec 11. Erratum in: *Cell*. 2015 Jul 30;162(3):687-8. PMID: 25497547; PMCID: PMC5635824.
- 27) Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DT, Tanay A, Hadjur S. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep*. 2015 Mar 3;10(8):1297-309. doi: 10.1016/j.celrep.2015.02.004. Epub 2015 Feb 26. PMID: 25732821; PMCID: PMC4542312.
- 28) Gaiti F, Calcino AD, Tanurdžić M, Degnan BM. Origin and evolution of the metazoan non-coding regulatory genome. *Dev Biol*. 2017 Jul 15;427(2):193-202. doi: 10.1016/j.ydbio.2016.11.013. Epub 2016 Nov 20. PMID: 27880868.
- 29) 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 Oct 23;109(43):17507-12. doi: 10.1073/pnas.1111941109. Epub 2012 Oct 8. PMID: 23045651; PMCID: PMC3491479.
- 30) Ouyang W, Xiong D, Li G, Li X. Unraveling the 3D Genome Architecture in Plants: Present and Future. *Mol Plant*. 2020 Dec 7;13(12):1676-1693. doi: 10.1016/j.molp.2020.10.002. Epub 2020 Oct 13. PMID: 33065269.
- 31) Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, Habermann B, Akhtar A, Manke T. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun*. 2018 Jan 15;9(1):189. doi: 10.1038/s41467-017-02525-w. PMID: 29335486; PMCID: PMC5768762.
- 32) Hou C, Li L, Qin ZS, Corces VG. Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Mol Cell*. 2012 Nov 9;48(3):471-84. doi: 10.1016/j.molcel.2012.08.031. Epub 2012 Oct 4. PMID: 23041285; PMCID: PMC3496039.
- 33) Wang Q, Sun Q, Czajkowsky DM, Shao Z. Sub-kb Hi-C in *D. melanogaster* reveals conserved characteristics of TADs between insect and mammalian cells. *Nat Commun*. 2018 Jan 15;9(1):188. doi: 10.1038/s41467-017-02526-9. PMID: 29335463; PMCID: PMC5768742.
- 34) Hug CB, Grimaldi AG, Kruse K, Vaquerizas JM. Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell*. 2017 Apr 6;169(2):216-228.e19. doi: 10.1016/j.cell.2017.03.024. PMID: 28388407.
- 35) Strom AR, Emelyanov AV, Mir M, Fyodorov DV, Darzacq X, Karpen GH. Phase separation drives heterochromatin domain formation. *Nature*. 2017 Jul 13;547(7662):241-245. doi: 10.1038/nature22989. Epub 2017 Jun 21. PMID: 28636597; PMCID: PMC6022742.

- 36) Prescott DM. The DNA of ciliated protozoa. *Microbiol Rev.* 1994 Jun;58(2):233-67. doi: 10.1128/mr.58.2.233-267.1994. PMID: 8078435; PMCID: PMC372963.
- 37) Yerlici VT, Landweber LF. Programmed Genome Rearrangements in the Ciliate *Oxytricha*. *Microbiol Spectr.* 2014 Dec;2(6):10.1128/microbiolspec.MDNA3-0025-2014. doi: 10.1128/microbiolspec.MDNA3-0025-2014. PMID: 26104449; PMCID: PMC4480583.
- 38) Swart EC, Bracht JR, Magrini V, Minx P, Chen X, Zhou Y, Khurana JS, Goldman AD, Nowacki M, Schotanus K, Jung S, Fulton RS, Ly A, McGrath S, Haub K, Wiggins JL, Storton D, Matese JC, Parsons L, Chang WJ, Bowen MS, Stover NA, Jones TA, Eddy SR, Herrick GA, Doak TG, Wilson RK, Mardis ER, Landweber LF. The *Oxytricha trifallax* macronuclear genome: a complex eukaryotic genome with 16,000 tiny chromosomes. *PLoS Biol.* 2013;11(1):e1001473. doi: 10.1371/journal.pbio.1001473. Epub 2013 Jan 29. PMID: 23382650; PMCID: PMC3558436.
- 39) Lindblad KA, Pathmanathan JS, Moreira S, Bracht JR, Sebra RP, Hutton ER, Landweber LF. Capture of complete ciliate chromosomes in single sequencing reads reveals widespread chromosome isoforms. *BMC Genomics.* 2019 Dec 30;20(1):1037. doi: 10.1186/s12864-019-6189-9. PMID: 31888453; PMCID: PMC6937825.
- 40) Chen X, Bracht JR, Goldman AD, Dolzhenko E, Clay DM, Swart EC, Perlman DH, Doak TG, Stuart A, Amemiya CT, Sebra RP, Landweber LF. The architecture of a scrambled genome reveals massive levels of genomic rearrangement during development. *Cell.* 2014 Aug 28;158(5):1187-1198. doi: 10.1016/j.cell.2014.07.034. PMID: 25171416; PMCID: PMC4199391.
- 41) Zheng H, Xie W. The role of 3D genome organization in development and cell differentiation. *Nat Rev Mol Cell Biol.* 2019 Sep;20(9):535-550. doi: 10.1038/s41580-019-0132-4. PMID: 31197269.
- 42) Fang W, Wang X, Bracht JR, Nowacki M, Landweber LF. Piwi-interacting RNAs protect DNA against loss during *Oxytricha* genome rearrangement. *Cell.* 2012 Dec 7;151(6):1243-55. doi: 10.1016/j.cell.2012.10.045. PMID: 23217708; PMCID: PMC3678556.
- 43) Lindblad KA, Bracht JR, Williams AE, Landweber LF. Thousands of RNA-cached copies of whole chromosomes are present in the ciliate *Oxytricha* during development. *RNA.* 2017 Aug;23(8):1200-1208. doi: 10.1261/rna.058511.116. Epub 2017 Apr 27. PMID: 28450531; PMCID: PMC5513065.
- 44) Nowacki M, Vijayan V, Zhou Y, Schotanus K, Doak TG, Landweber LF. RNA-mediated epigenetic programming of a genome-rearrangement pathway. *Nature.* 2008 Jan 10;451(7175):153-8. doi: 10.1038/nature06452. Epub 2007 Nov 28. PMID: 18046331; PMCID: PMC2647009.
- 45) Postberg J, Jönsson F, Weil PP, Bulic A, Juranek SA, Lipps HJ. 27nt-RNAs guide histone variant deposition via 'RNA-induced DNA replication interference' and thus transmit parental

genome partitioning in *Stylonychia*. *Epigenetics Chromatin*. 2018 Jun 12;11(1):31. doi: 10.1186/s13072-018-0201-5. PMID: 29895326; PMCID: PMC5996456.

46) Taverna SD, Coyne RS, Allis CD. Methylation of histone h3 at lysine 9 targets programmed DNA elimination in *Tetrahymena*. *Cell*. 2002 Sep 20;110(6):701-11. doi: 10.1016/s0092-8674(02)00941-8. PMID: 12297044.

47) Luo Z, Hu T, Jiang H, Wang R, Xu Q, Zhang S, Cao J, Song X. Rearrangement of macronucleus chromosomes correspond to TAD-like structures of micronucleus chromosomes in *Tetrahymena thermophila*. *Genome Res*. 2020 Mar;30(3):406-414. doi: 10.1101/gr.241687.118. Epub 2020 Mar 12. PMID: 32165395; PMCID: PMC7111529.

48) Hamilton EP, Kapusta A, Huvos PE, Bidwell SL, Zafar N, Tang H, Hadjithomas M, Krishnakumar V, Badger JH, Caler EV, Russ C, Zeng Q, Fan L, Levin JZ, Shea T, Young SK, Hegarty R, Daza R, Gujja S, Wortman JR, Birren BW, Nusbaum C, Thomas J, Carey CM, Pritham EJ, Feschotte C, Noto T, Mochizuki K, Papazyan R, Taverna SD, Dear PH, Cassidy-Hanley DM, Xiong J, Miao W, Orias E, Coyne RS. Structure of the germline genome of *Tetrahymena thermophila* and relationship to the massively rearranged somatic genome. *Elife*. 2016 Nov 28;5:e19090. doi: 10.7554/eLife.19090. PMID: 27892853; PMCID: PMC5182062.

49) Lu MW, Beh LY, Yerlici VT, Fang W, Kulej K, Garcia BA, Landweber LF. Exploration of the Nuclear Proteomes in the Ciliate *Oxytricha trifallax*. *Microorganisms*. 2023 Jan 30;11(2):343. doi: 10.3390/microorganisms11020343. PMID: 36838311; PMCID: PMC9958989.

50) Guérin F, Arnaiz O, Boggetto N, Denby Wilkes C, Meyer E, Sperling L, Duharcourt S. Flow cytometry sorting of nuclei enables the first global characterization of *Paramecium* germline DNA and transposable elements. *BMC Genomics*. 2017 Apr 26;18(1):327. doi: 10.1186/s12864-017-3713-7. PMID: 28446146; PMCID: PMC5405496.

51) Zangarelli C, Arnaiz O, Bourge M, Gorrichon K, Jaszczyszyn Y, Mathy N, Escoriza L, Bétermier M, Régnier V. Developmental timing of programmed DNA elimination in *Paramecium tetraurelia* recapitulates germline transposon evolutionary dynamics. *Genome Res*. 2022 Nov-Dec;32(11-12):2028-2042. doi: 10.1101/gr.277027.122. Epub 2022 Nov 23. PMID: 36418061; PMCID: PMC9808624.

52) Spear BB, Lauth MR. Polytene chromosomes of *Oxytricha*: Biochemical and morphological changes during macronuclear development in a ciliated protozoan. *Chromosoma*. 1976 Jan 27;54(1):1-13. doi: 10.1007/BF00331828. PMID: 813980.

53) Möllenbeck M, Zhou Y, Cavalcanti AR, Jönsson F, Higgins BP, Chang WJ, Juranek S, Doak TG, Rozenberg G, Lipps HJ, Landweber LF. The pathway to detangle a scrambled gene. *PLoS One*. 2008 Jun 4;3(6):e2330. doi: 10.1371/journal.pone.0002330. PMID: 18523559; PMCID: PMC2394655.

54) Beh LY, Debelouchina GT, Clay DM, Thompson RE, Lindblad KA, Hutton ER, Bracht JR, Sebra RP, Muir TW, Landweber LF. Identification of a DNA N6-Adenine Methyltransferase Complex and Its Impact on Chromatin Organization. *Cell*. 2019 Jun 13;177(7):1781-1796.e25. doi: 10.1016/j.cell.2019.04.028. Epub 2019 May 16. PMID: 31104845; PMCID: PMC6570567.

55) Feng Y, Neme R, Beh LY, Chen X, Braun J, Lu MW, Landweber LF. Comparative genomics reveals insight into the evolutionary origin of massively scrambled genomes. *Elife*. 2022 Nov 24;11:e82979. doi: 10.7554/eLife.82979. PMID: 36421078; PMCID: PMC9797194.

56) Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol*. 2016 Nov 25;17(1):239. doi: 10.1186/s13059-016-1103-0. Erratum in: *Genome Biol*. 2016 Dec 13;17(1):256. PMID: 27887629; PMCID: PMC5124260.

57) Belaghzal H, Dekker J, Gibcus JH. Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. *Methods*. 2017 Jul 1;123:56-65. doi: 10.1016/j.ymeth.2017.04.004. Epub 2017 Apr 18. PMID: 28435001; PMCID: PMC5522765.

58) Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst*. 2016 Jul;3(1):95-8. doi: 10.1016/j.cels.2016.07.002. PMID: 27467249; PMCID: PMC5846465.

59) Qin M, Wu S, Li A, Zhao F, Feng H, Ding L, Ruan J. LRScaf: improving draft genomes using long noisy reads. *BMC Genomics*. 2019 Dec 9;20(1):955. doi: 10.1186/s12864-019-6337-2. PMID: 31818249; PMCID: PMC6902338.

60) Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, Aiden EL. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*. 2017 Apr 7;356(6333):92-95. doi: 10.1126/science.aal3327. Epub 2017 Mar 23. PMID: 28336562; PMCID: PMC5635820.

61) Robinson JT, Turner D, Durand NC, Thorvaldsdóttir H, Mesirov JP, Aiden EL. Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. *Cell Syst*. 2018 Feb 28;6(2):256-258.e1. doi: 10.1016/j.cels.2018.01.001. Epub 2018 Feb 7. PMID: 29428417; PMCID: PMC6047755.

62) Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009 Dec 15;10:421. doi: 10.1186/1471-2105-10-421. PMID: 20003500; PMCID: PMC2803857.

63) Knight PA, Ruiz D. A fast algorithm for matrix balancing. *IMA J. Numer. Anal.* 2013 Jul; 33(3): 1029-47. doi: 10.1093/imanum/drs019.

- 64) Cremer T, Cremer M. Chromosome territories. *Cold Spring Harb Perspect Biol.* 2010 Mar;2(3):a003889. doi: 10.1101/cshperspect.a003889. PMID: 20300217; PMCID: PMC2829961.
- 65) Ringel AR, Szabo Q, Chiariello AM, Chudzik K, Schöpflin R, Rothe P, Mattei AL, Zehnder T, Harnett D, Laupert V, Bianco S, Hetzel S, Glaser J, Phan MHQ, Schindler M, Ibrahim DM, Paliou C, Esposito A, Prada-Medina CA, Haas SA, Giere P, Vingron M, Wittler L, Meissner A, Nicodemi M, Cavalli G, Bantignies F, Mundlos S, Robson MI. Repression and 3D-restructuring resolves regulatory conflicts in evolutionarily rearranged genomes. *Cell.* 2022 Sep 29;185(20):3689-3704.e21. doi: 10.1016/j.cell.2022.09.006. PMID: 36179666; PMCID: PMC9567273.
- 66) Khurana JS, Wang X, Chen X, Perlman DH, Landweber LF. Transcription-independent functions of an RNA polymerase II subunit, Rpb2, during genome rearrangement in the ciliate, *Oxytricha trifallax*. *Genetics.* 2014 Jul;197(3):839-49. doi: 10.1534/genetics.114.163279. Epub 2014 May 1. PMID: 24793090; PMCID: PMC4096365.
- 67) Wolff J, Bhardwaj V, Nothjunge S, Richard G, Renschler G, Gilsbach R, Manke T, Backofen R, Ramírez F, Grüning BA. Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* 2018 Jul 2;46(W1):W11-W16. doi: 10.1093/nar/gky504. PMID: 29901812; PMCID: PMC6031062.
- 68) Wolff J, Rabbani L, Gilsbach R, Richard G, Manke T, Backofen R, Grüning BA. Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* 2020 Jul 2;48(W1):W177-W184. doi: 10.1093/nar/gkaa220. PMID: 32301980; PMCID: PMC7319437.
- 69) Harris HL, Gu H, Olshansky M, Wang A, Farabella I, Eliaz Y, Kalluchi A, Krishna A, Jacobs M, Cauer G, Pham M, Rao SSP, Dudchenko O, Omer A, Mohajeri K, Kim S, Nichols MH, Davis ES, Gkountaroulis D, Udupa D, Aiden AP, Corces VG, Phanstiel DH, Noble WS, Nir G, Di Pierro M, Seo JS, Talkowski ME, Aiden EL, Rowley MJ. Chromatin alternates between A and B compartments at kilobase scale for subgenomic organization. *Nat Commun.* 2023 Jun 6;14(1):3303. doi: 10.1038/s41467-023-38429-1. PMID: 37280210; PMCID: PMC10244318.
- 70) Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, Fonseca NA, Huber W, Haering CH, Mirny L, Spitz F. Two independent modes of chromatin organization revealed by cohesin removal. *Nature.* 2017 Nov 2;551(7678):51-56. doi: 10.1038/nature24281. Epub 2017 Sep 27. PMID: 29094699; PMCID: PMC5687303.
- 71) Beliveau BJ, Apostolopoulos N, Wu CT. Visualizing genomes with Oligopaint FISH probes. *Curr Protoc Mol Biol.* 2014 Jan 6;105:Unit 14.23.. doi: 10.1002/0471142727.mb1423s105. PMID: 24510436; PMCID: PMC3928790.
- 72) Beliveau BJ, Kishi JY, Nir G, Sasaki HM, Saka SK, Nguyen SC, Wu CT, Yin P. OligoMiner provides a rapid, flexible environment for the design of genome-scale oligonucleotide in situ

hybridization probes. *Proc Natl Acad Sci U S A*. 2018 Mar 6;115(10):E2183-E2192. doi: 10.1073/pnas.1714530115. Epub 2018 Feb 20. PMID: 29463736; PMCID: PMC5877937.

73) Lappala A, Wang CY, Kriz A, Michalk H, Tan K, Lee JT, Sanbonmatsu KY. Four-dimensional chromosome reconstruction elucidates the spatiotemporal reorganization of the mammalian X chromosome. *Proc Natl Acad Sci U S A*. 2021 Oct 19;118(42):e2107092118. doi: 10.1073/pnas.2107092118. PMID: 34645712; PMCID: PMC8594586.

74) Singhal A, Roth C, Micheva-Viteva SN, Venu V, Lappala A, Lee JT, Starkenburg SR, Steadman CR, Sanbonmatsu KY. Human Coronavirus Infection Reorganizes Spatial Genomic Architecture in Permissive Lung Cells. *Res Sq [Preprint]*. 2024 Mar 13:rs.3.rs-3979539. doi: 10.21203/rs.3.rs-3979539/v1. PMID: 38559036; PMCID: PMC10980144.

75) Thompson AP, Aktulga HM, Berger R, Bolintineanu DS, Brown WM, Crozier PS, in 't Veld PJ, Kohlmeyer A, Moore SG, Nguyen TD, Shan R, Stevens MJ, Tranchida J, Trott C, Plimpton SJ. LAMMPS—a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput Phys Commun*. 2022 Feb; 271, 108171. doi: 10.1016/j.cpc.2021.108171.

76) Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J Comput Chem*. 2011 Jul 30;32(10):2319-27. doi: 10.1002/jcc.21787. Epub 2011 Apr 15. PMID: 21500218; PMCID: PMC3144279.

77) Jakupovic E, Beckstein O. MPI-parallel Molecular Dynamics Trajectory Analysis with the H5MD Format in the MDAnalysis Python Package. in *Proceedings of the 20th Python in Science Conference*, pp. 40–48, 2021. doi:10.25080/majora-1b6fd038-005.

78) Singh A, Maurer-Alcalá XX, Solberg T, Häußermann L, Gisler S, Ignarski M, Swart EC, Nowacki M. Chromatin remodeling is required for sRNA-guided DNA elimination in *Paramecium*. *EMBO J*. 2022 Nov 17;41(22):e111839. doi: 10.15252/embj.2022111839. Epub 2022 Oct 11. PMID: 36221862; PMCID: PMC9670198.

79) Tsompana M, Buck MJ. Chromatin accessibility: a window into the genome. *Epigenetics Chromatin*. 2014 Nov 20;7(1):33. doi: 10.1186/1756-8935-7-33. PMID: 25473421; PMCID: PMC4253006.

80) Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol*. 2015 Jan 5;109:21.29.1-21.29.9. doi: 10.1002/0471142727.mb2129s109. PMID: 25559105; PMCID: PMC4374986.

81) Nowacki M, Higgins BP, Maquilan GM, Swart EC, Doak TG, Landweber LF. A functional role for transposases in a large eukaryotic genome. *Science*. 2009 May 15;324(5929):935-8. doi: 10.1126/science.1170023. Epub 2009 Apr 16. PMID: 19372392; PMCID: PMC3491810.

82) Chen X, Landweber LF. Phylogenomic analysis reveals genome-wide purifying selection on TBE transposons in the ciliate *Oxytricha*. *Mobile DNA*. 2016 Jan 25;7:2. doi: 10.1186/s13100-016-0057-9. PMID: 26811739; PMCID: PMC4724952.

83) Allen SE, Nowacki M. Roles of Noncoding RNAs in Ciliate Genome Architecture. *J Mol Biol*. 2020 Jul 10;432(15):4186-4198. doi: 10.1016/j.jmb.2019.12.042. Epub 2020 Jan 10. PMID: 31926952; PMCID: PMC7374600.

84) Barutcu AR, Blencowe BJ, Rinn JL. Differential contribution of steady-state RNA and active transcription in chromatin organization. *EMBO Rep*. 2019 Oct 4;20(10):e48068. doi: 10.15252/embr.201948068. Epub 2019 Aug 26. PMID: 31448565; PMCID: PMC6776903.

85) Creamer KM, Kolpa HJ, Lawrence JB. Nascent RNA scaffolds contribute to chromosome territory architecture and counter chromatin compaction. *Mol Cell*. 2021 Sep 2;81(17):3509-3525.e5. doi: 10.1016/j.molcel.2021.07.004. Epub 2021 Jul 27. PMID: 34320406; PMCID: PMC8419111.

86) Quinodoz SA, Jachowicz JW, Bhat P, Ollikainen N, Banerjee AK, Goronzy IN, Blanco MR, Chovanec P, Chow A, Markaki Y, Thai J, Plath K, Guttman M. RNA promotes the formation of spatial compartments in the nucleus. *Cell*. 2021 Nov 11;184(23):5775-5790.e30. doi: 10.1016/j.cell.2021.10.014. Epub 2021 Nov 4. PMID: 34739832; PMCID: PMC9115877.

87) Bou-Nader C, Bothra A, Garboczi DN, Leppla SH, Zhang J. Structural basis of R-loop recognition by the S9.6 monoclonal antibody. *Nat Commun*. 2022 Mar 28;13(1):1641. doi: 10.1038/s41467-022-29187-7. PMID: 35347133; PMCID: PMC8960830.

88) Wahba L, Costantino L, Tan FJ, Zimmer A, Koshland D. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev*. 2016 Jun 1;30(11):1327-38. doi: 10.1101/gad.280834.116. PMID: 27298336; PMCID: PMC4911931.

89) Aguilera A, García-Muse T. R loops: from transcription byproducts to threats to genome stability. *Mol Cell*. 2012 Apr 27;46(2):115-24. doi: 10.1016/j.molcel.2012.04.009. PMID: 22541554.

90) Barrington C, Georgopoulou D, Pezic D, Varsally W, Herrero J, Hadjur S. Enhancer accessibility and CTCF occupancy underlie asymmetric TAD architecture and cell type specific genome topology. *Nat Commun*. 2019 Jul 2;10(1):2908. doi: 10.1038/s41467-019-10725-9. PMID: 31266948; PMCID: PMC6606583.

91) Downen JM, Fan ZP, Hnisz D, Ren G, Abraham BJ, Zhang LN, Weintraub AS, Schujiers J, Lee TI, Zhao K, Young RA. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell*. 2014 Oct 9;159(2):374-387. doi: 10.1016/j.cell.2014.09.030. PMID: 25303531; PMCID: PMC4197132.

92) Phillips JE, Corces VG. CTCF: master weaver of the genome. *Cell*. 2009 Jun 26;137(7):1194-211. doi: 10.1016/j.cell.2009.06.001. PMID: 19563753; PMCID: PMC3040116.

- 93) Hsieh TH, Weiner A, Lajoie B, Dekker J, Friedman N, Rando OJ. Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell*. 2015 Jul 2;162(1):108-19. doi: 10.1016/j.cell.2015.05.048. Epub 2015 Jun 25. PMID: 26119342; PMCID: PMC4509605.
- 94) Vietri Rudan M, Hadjur S. Genetic Tailors: CTCF and Cohesin Shape the Genome During Evolution. *Trends Genet*. 2015 Nov;31(11):651-660. doi: 10.1016/j.tig.2015.09.004. Epub 2015 Oct 1. PMID: 26439501.
- 95) Lam KG, Brick K, Cheng G, Pratto F, Camerini-Otero RD. Cell-type-specific genomics reveals histone modification dynamics in mammalian meiosis. *Nat Commun*. 2019 Aug 23;10(1):3821. doi: 10.1038/s41467-019-11820-7. PMID: 31444359; PMCID: PMC6707301.
- 96) Postberg J, Heyse K, Cremer M, Cremer T, Lipps HJ. Spatial and temporal plasticity of chromatin during programmed DNA-reorganization in *Styloynchia* macronuclear development. *Epigenetics Chromatin*. 2008 Oct 27;1(1):3. doi: 10.1186/1756-8935-1-3. PMID: 19014664; PMCID: PMC2603335.
- 97) Durrant MG, Perry NT, Pai JJ, Jangid AR, Athukoralage JS, Hiraizumi M, McSpedon JP, Pawluk A, Nishimasu H, Konermann S, Hsu PD. Bridge RNAs direct programmable recombination of target and donor DNA. *Nature*. 2024 Jun;630(8018):984-993. doi: 10.1038/s41586-024-07552-4. Epub 2024 Jun 26. PMID: 38926615; PMCID: PMC11208160.
- 98) Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science*. 2002 Sep 13;297(5588):1833-7. doi: 10.1126/science.1074973. Epub 2002 Aug 22. PMID: 12193640.
- 99) Chen X, Jung S, Beh LY, Eddy SR, Landweber LF. Combinatorial DNA Rearrangement Facilitates the Origin of New Genes in Ciliates. *Genome Biol Evol*. 2015 Sep 2;7(10):2859-70. doi: 10.1093/gbe/evv172. PMID: 26338187; PMCID: PMC4684698.
- 100) Hartono SR, Malapert A, Legros P, Bernard P, Chédin F, Vanoosthuyse V. The Affinity of the S9.6 Antibody for Double-Stranded RNAs Impacts the Accurate Mapping of R-Loops in Fission Yeast. *J Mol Biol*. 2018 Feb 2;430(3):272-284. doi: 10.1016/j.jmb.2017.12.016. Epub 2017 Dec 28. PMID: 29289567; PMCID: PMC5987549.
- 101) Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018 Sep 15;34(18):3094-3100. doi: 10.1093/bioinformatics/bty191. PMID: 29750242; PMCID: PMC6137996.
- 102) Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- 103) Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014 Aug 1;30(15):2114-20. doi: 10.1093/bioinformatics/btu170. Epub 2014 Apr 1. PMID: 24695404; PMCID: PMC4103590.

104) Chen X, Shen Y, Draper W, Buenrostro JD, Litzenger U, Cho SW, Satpathy AT, Carter AC, Ghosh RP, East-Seletsky A, Doudna JA, Greenleaf WJ, Liphardt JT, Chang HY. ATAC-seq reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat Methods*. 2016 Dec;13(12):1013-1020. doi: 10.1038/nmeth.4031. Epub 2016 Oct 17. PMID: 27749837; PMCID: PMC5509561.

105) 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(3):R25. doi: 10.1186/gb-2009-10-3-r25. Epub 2009 Mar 4. PMID: 19261174; PMCID: PMC2690996.

106) Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9(9):R137. doi: 10.1186/gb-2008-9-9-r137. Epub 2008 Sep 17. PMID: 18798982; PMCID: PMC2592715.

107) Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016 Jul 8;44(W1):W160-5. doi: 10.1093/nar/gkw257. Epub 2016 Apr 13. PMID: 27079975; PMCID: PMC4987876.

108) Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. Fiji: an open-source platform for biological-image analysis. *Nat Methods*. 2012 Jun 28;9(7):676-82. doi: 10.1038/nmeth.2019. PMID: 22743772; PMCID: PMC3855844.

109) Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2. 2013. [q-bio.GN].

Figures and Legends

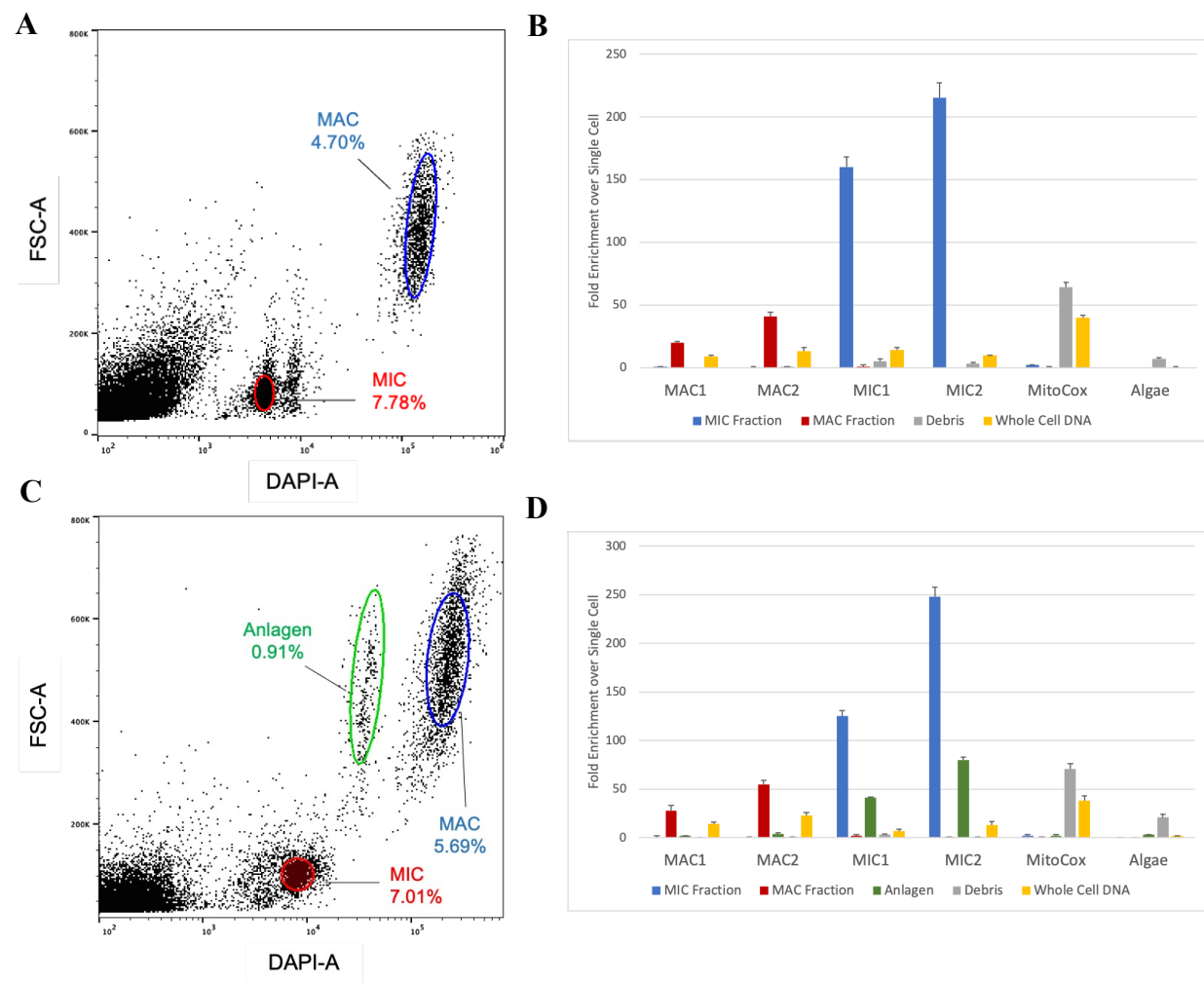


Figure 1. Isolation of *Oxytricha* nuclei via FACS.

(A) Selection of vegetative JRB310 nuclei based on DAPI and forward scatter (“FSC”) signal.

(B) qPCR of sorted vegetative populations. “MAC1” and “MAC2” target MAC-specific loci that span scrambled MDS-MDS junctions. “MIC1” and “MIC2” target high copy number MIC-specific loci within transposable elements. “MitoCox” targets the mitochondrial COX1 locus. “Algae” targets *Chlamydomonas reinhardtii*-specific sequence. “Debris” sample derives from dense, undefined cloud of particles to the left of MIC population in (A).

(C) Selection of developmental nuclei at 24h post-mixing of JRB310 and JRB510 cells.

(D) qPCR of sorted developmental populations, as in (B).

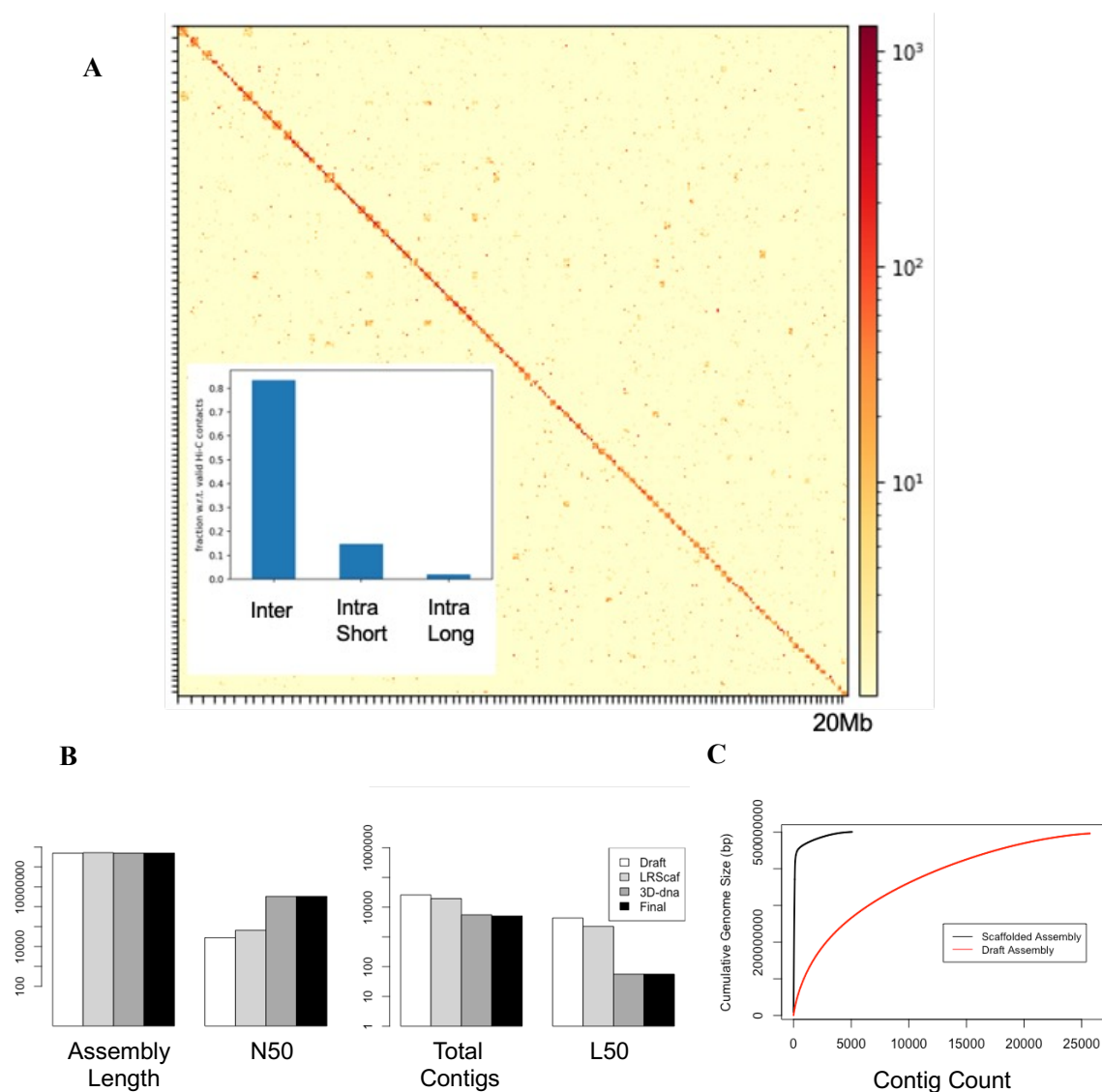


Figure 2. Hi-C and long-read data produce a highly contiguous scaffolded genome assembly.

(A) Hi-C map of vegetative germline Hi-C data (log1p) mapped to the original assembly's 100 longest contigs, totaling ~4% of the genome and <0.5% of contigs. Hash marks indicate contig boundaries. High interaction frequency between small contigs indicates likely synteny. Inset shows frequency of different interaction types: Interchromosomal, Intrachromosomal Short ($\leq 20\text{kb}$), and Intrachromosomal Long ($>20\text{kb}$).

(B) Assembly statistics after different stages of scaffolding. “Draft” indicates original germline reference genome.⁴⁰

(C) Cumulative distribution of contig sizes in original and scaffolded assemblies.

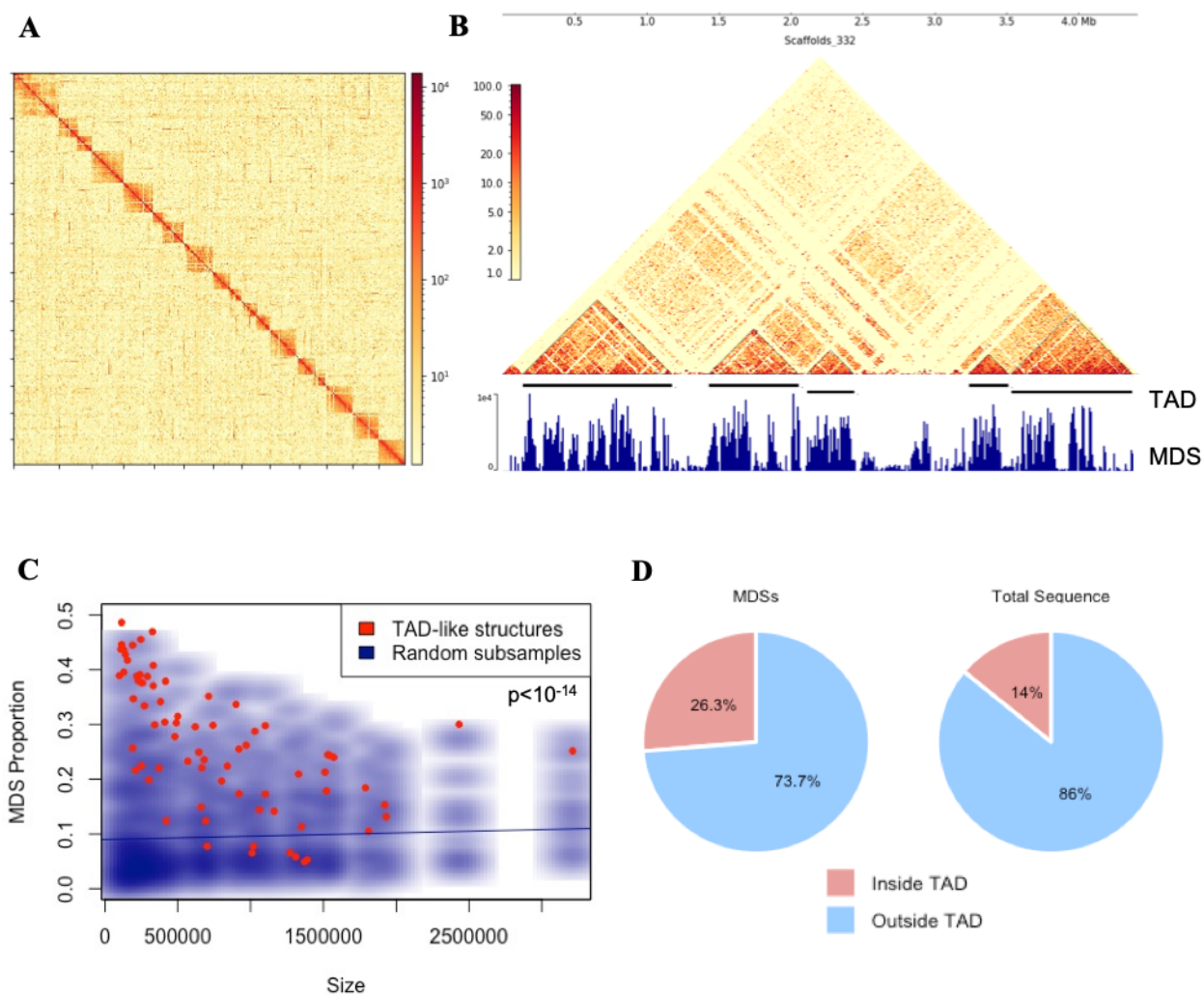


Figure 3. TAD-like structures form around dense tracts of retained sequences.

(A) Germline Hi-C contacts (log1p) from 13 scaffolds larger than 5 Mb, totaling 78 Mb.

(B) Germline Hi-C contacts (log1p) from representative scaffold bearing several TAD-like structures. “MDS” track shows total MDS sequence within 10 kb bins.

(C) Distribution of TAD-like structure lengths and MDS proportion compared to smoothed scatter plot of 25 iterations of randomly subsampled regions of the genome with the same size distribution as the TAD-like structures. Line is linear model of the random genomic subsamples. p -value calculated by Student’s t -test on MDS proportion of TADs vs. average of subsample iterations.

(D) Percent of total MDS sequence and genomic sequence contained within TAD-like structures.

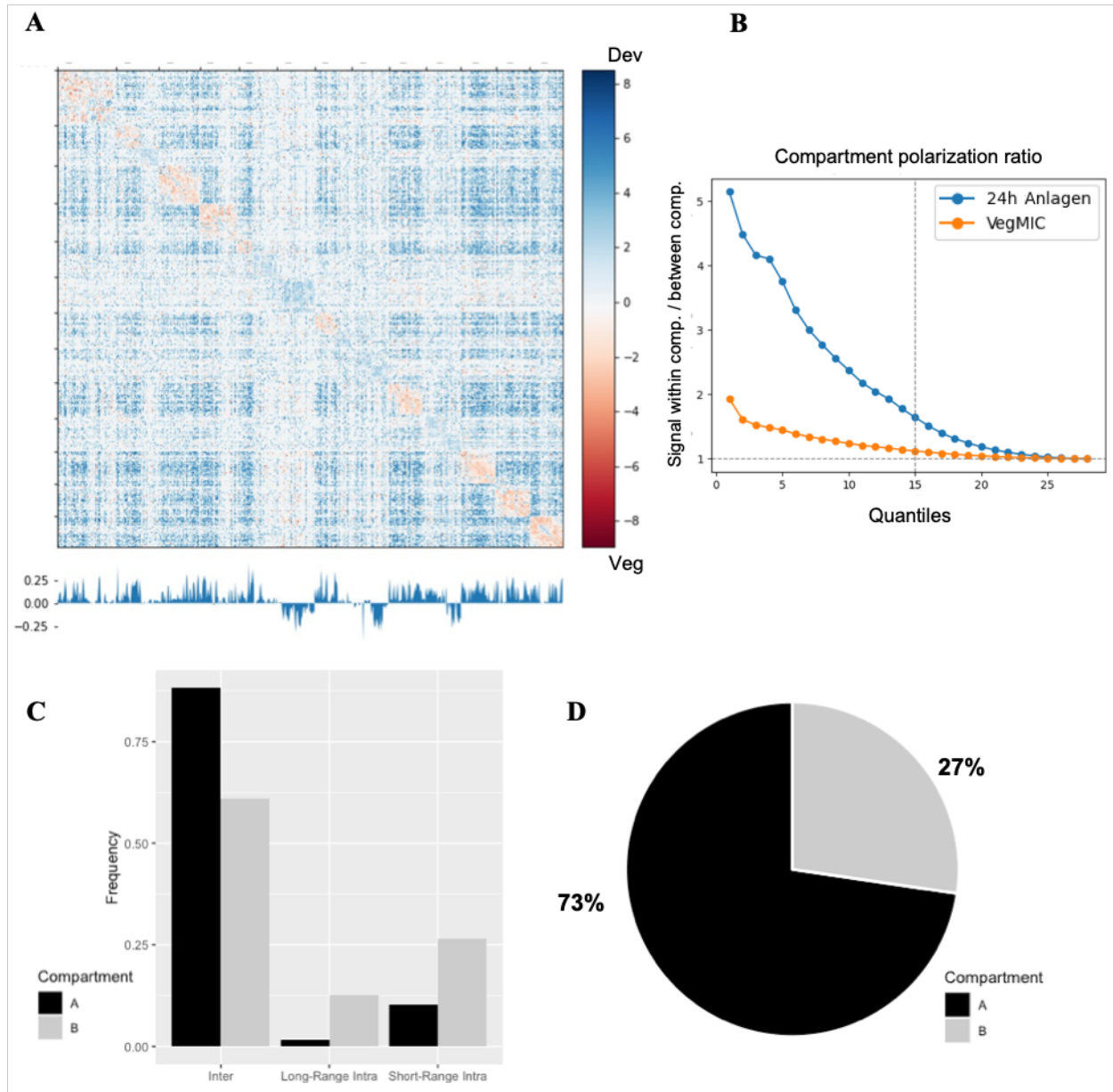


Figure 4. The developing genome compartmentalizes to separate retained and eliminated regions.

(A) 24h anlagen Hi-C data across scaffolds ≥ 5 Mb. Plotted values are \log_2 ratio over vegetative germline (input matrices KR-corrected and normalized by coverage). Track below is PC1 value indicating compartment.

(B) Compartmentalization signal of the 24h developmental genome relative to vegetative germline. 24h PC1 values used for division into quantiles.

(C) Frequency of different interaction types within compartment A or B bins at 24h.

(D) Percentage of MDS sequence in compartment A or B at 24h.

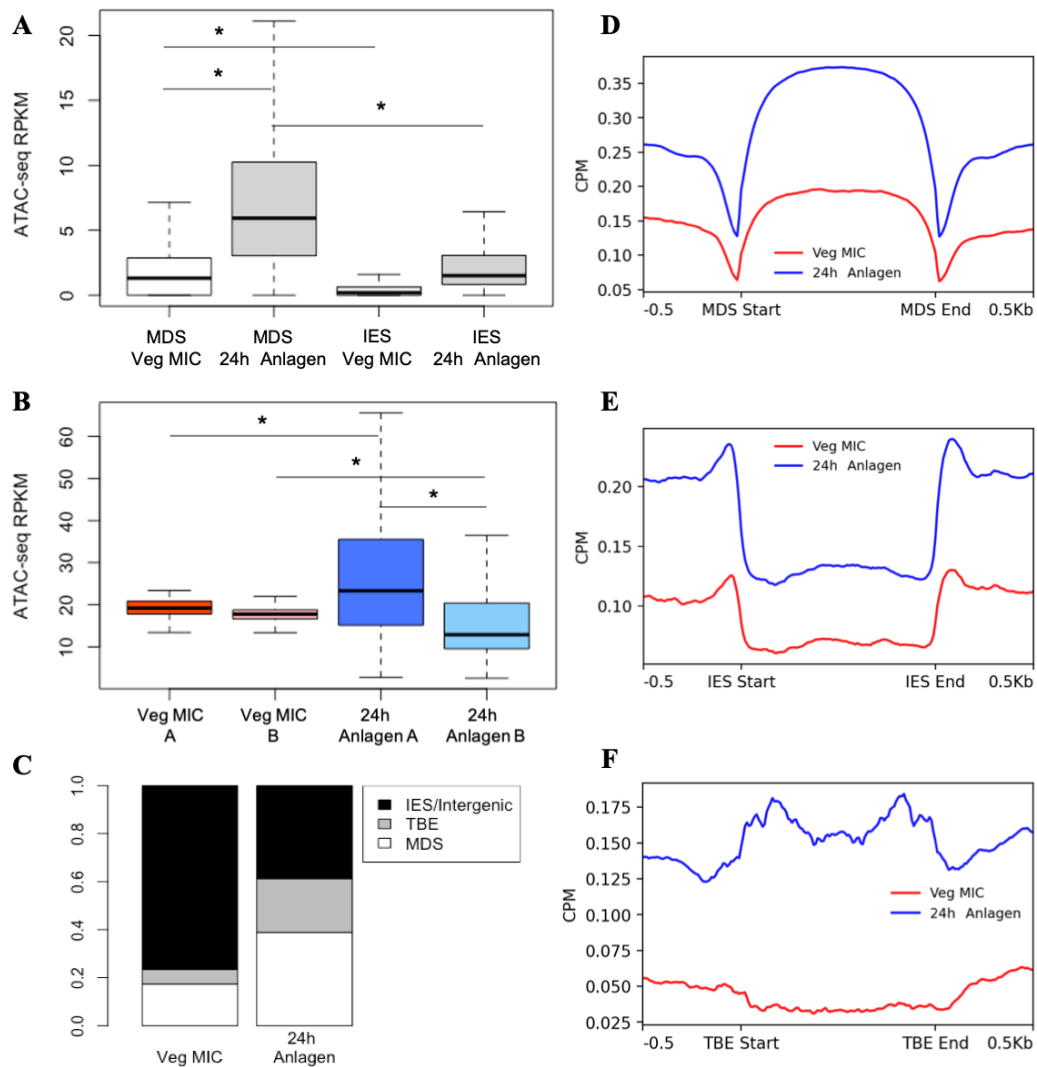


Figure 5. Retained sequences are more accessible at 24h than eliminated sequences.

(A) ATAC-seq coverage over MDSs and IESs in vegetative germline and 24h anlagen. Germline features were restricted to those of similar size, between 500bp-1,000 bp ($> 10,000$ features for each type). * indicates p -value $< 10^{-10}$, Student's t -test.

(B) Vegetative and developmental ATAC-seq coverage within 100 kb regions identified as compartment A or B at 24h. * indicates p -value $< 10^{-13}$, Student's t -test.

(C) Proportion of vegetative and 24h ATAC-seq peaks mapping within MDSs, TBEs, or other eliminated sequences.

(D) ATAC-seq coverage profile of size-selected MDSs in each ATAC-seq dataset.

(E) ATAC-seq coverage profile of size-selected IESs in each ATAC-seq dataset.

(F) ATAC-seq coverage profile of size-selected TBE transposons (3-4 kb) in each ATAC-seq dataset.

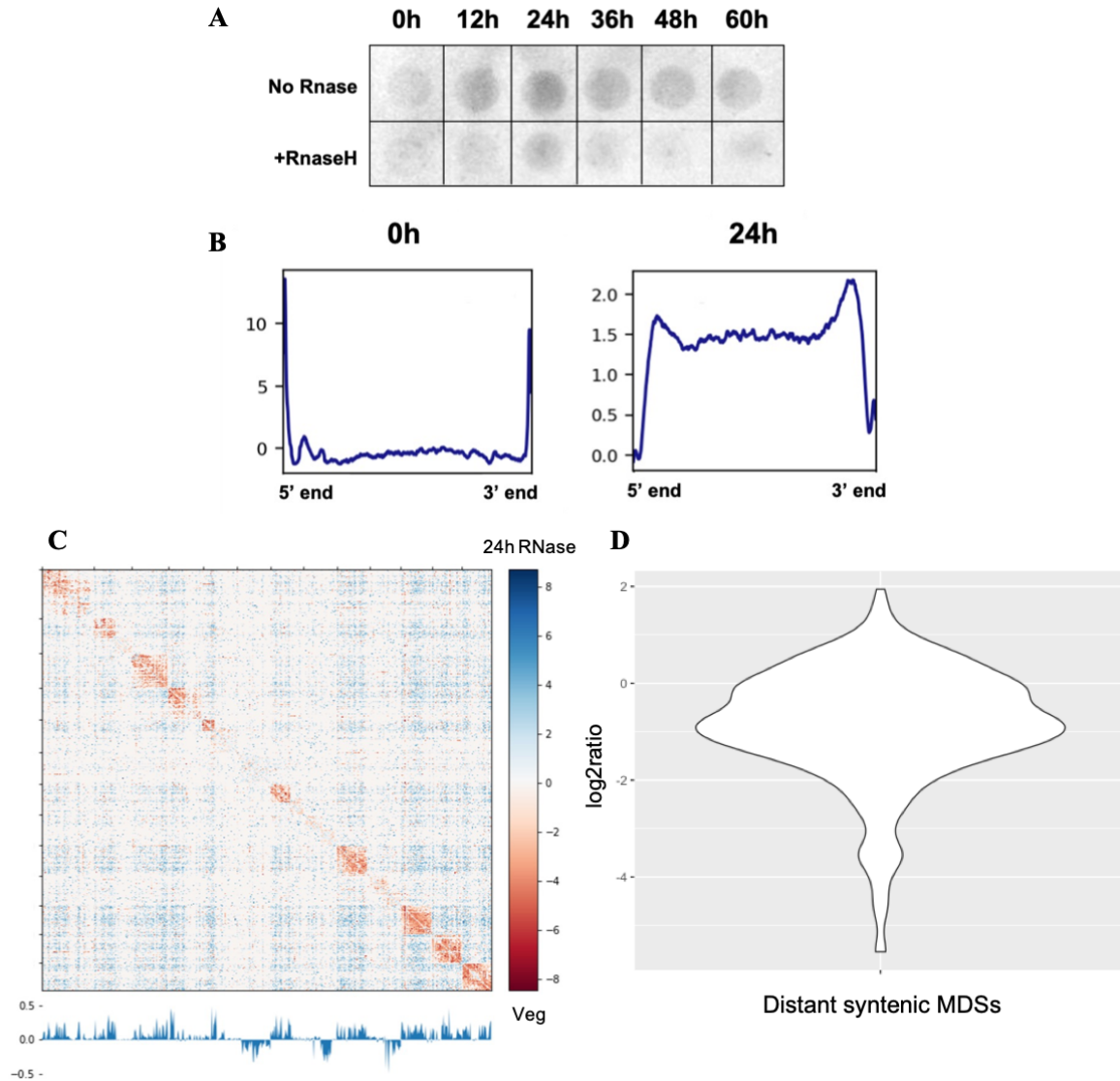


Figure 6. RNA-DNA interactions influence developmental 3D organization.

(A) Dot blot of RNA-DNA hybrids across developmental timecourse visualized with S9.6 antibody, including RNase H control. Remaining signal in RNase H control is most likely double-stranded RNA, which has some antibody affinity¹⁰⁰ and are ncRNA precursors in developmental.⁴³ Membrane loaded with 130ng genomic DNA per sample.

(B) DRIP-seq coverage across 5,000 nanochromosomes (all 1.5-3 kb, bearing single genes) in vegetative and developmental states.

(C) \log_2 ratios of RNase-treated 24h anlagen Hi-C vs. vegetative germline across scaffolds > 5 Mb. Lower track is PCA1 value of RNase-treated 24h anlagen in 100 kb bins.

(D) Violin plot of \log_2 ratios of contact frequency between 10 kb bins bearing distant, syntenic MDSs, RNase-treated 24h anlagen/24h control.

Tables

		Original Draft Assembly⁴⁰	Final Scaffolded Assembly
Total length		496,291,018 bp	500,711,191 bp
	GC%	28.4%	28.4%
Total contigs		25,720	5084
	> 1Mb	0	123
	Largest Contig	0.381 Mb	8.96 Mb
	N50	27,807	3,349,992
	L50	4333	56
MDS occurrences		752,900	708,192
	MDS sequence length	74.9 Mb	56.4 Mb
	MDS%	15.1%	11.2%
Somatic MDSs mapping	0 times	0	2960
	1 time	164,062	169,943
	2 times	65,188	71,550
	> 2 times	68,790	53,587
Telomere-bearing element (TBE) transposons		34,721	36,506
Telomeric repeats	Short (<10 repeats)	149	64
	Long (>10 repeats)	0	27
	Two-telomere contigs	0	2 (1.5 Mb and 3.9 Mb)

Table 1. Germline micronuclear genome scaffolding statistics.

Methods

***Oxytricha* culturing and mating**

Vegetative (asexually growing) *Oxytricha trifallax* strains JRB310 and JRB510 were cultured as previously described.³⁸ Cells were maintained at approximately 10,000 cells/ml in Pringsheim media (0.11mM Na₂HPO₄, 0.08mM MgSO₄, 0.85mM Ca(NO₃)₂, 0.35mM KCl, pH 7.0), filtered daily through cheesecloth, and fed daily with *Chlamydomonas reinhardtii*. Cells were also fed every other day with *Klebsiella pneumoniae* liquid cultures. Developmental samples were prepared by mating starved JRB310 and JRB510 at a concentration of 5,000 cells/ml, with timepoints measured from the moment parental cells were mixed (0h). Matings were confirmed to have at least 70% efficiency by counting conjugating pairs.

FACS sample preparation and sorting

Nuclear suspensions for sorting via FACS were collected by filtering cultures through cheesecloth and concentrating on a 10µm Nitex mesh filter before being pelleted by centrifugation for 2 minutes at 200xg. Cells were washed once in Pringsheim media, then collected again by centrifugation. Cells were then fixed in a 1% formaldehyde (v/v) solution in Pringsheim at room temperature with gentle shaking for 10 minutes. Glycine was then added to a final concentration of 125mM to quench fixation, followed by incubation at 4°C for 15 minutes with gentle shaking. Fixed cells were collected by centrifuging for 2 minutes at 300xg. Cells were then lysed by addition of 50ml of lysis buffer (20mM Tris-HCl (pH 6.8), 3% sucrose (w/v), 0.2% Triton X-100 (v/v), 0.01% spermidine-trihydrochloride (w/v)) and incubation on ice for 5 minutes. Nuclei were collected by centrifugation for 5 minutes at 4,000xg, and then washed in 50ml of FACS buffer (1x PBS, 1mM EDTA, 1% BSA) and collected again by centrifuging for 5 minutes at 4,000xg. Nuclei were resuspended in 1ml FACS buffer per million cells in original

culture, and DAPI was added to samples at a 1:1,000 dilution. Samples were stored at 4°C until sorted.

FACS was performed on BD FACSAria II (Becton Dickinson) and MA900 (Sony) machines. FSC threshold was set at 5%, and sensor gain settings were typically as follows: FSC: 5, BSC/SSC: 23%, DAPI: 34%. Sensor settings were adjusted from these benchmarks depending on the signal of individual samples. Gating was first performed on FSC-A x FSC-H to exclude doublets and then on FSC-A x SSC-A to enrich nuclei from cellular debris. Populations of nuclei were then selected on the basis of DAPI-A x FSC-A signal. Sorted nuclei for subsequent sequencing experiments were flash frozen in liquid nitrogen and stored at -80°C.

To confirm the identity of sorted nuclei, DNA was isolated through phenol-chloroform extraction. The difference in the molecular weight of DNA extracted from putative MICs and MACs was measured by BioAnalyzer (Agilent Technologies) and a 1.5% agarose gel. The presence of MAC- and MIC-specific sequences, as well as mitochondrial and algae-derived sequences, in DNA derived from sorted nuclei was then measured via qPCR with Power SYBR Green Master Mix (Thermo Fisher Scientific) on a Bio-Rad CFX384 machine using the following primers:

MIC1:

F: GGGAAGTAGCCTGGTGTATTAAG R: ATGCTCAACCTCTATCCTTTGG

MIC2:

F: GTAACCTGGAGTACCAAGAACC R: TGCCCTTTAGAAGTAGTCGTAATC

MAC1:

F: GCAGAGACAAGGCAGTACAA R: GAACTCCTACTCTGGCTTGATTAT

MAC2:

F: GAAGGGCTTTATGGCTAGACTC R: GCATTCCAGCAGTCATCAAAC

MitoCox:

F: CTCGATGTCGGCTCATCATATC R: ACGTTCTAAACCCAGCTCAC

Algae:

F: CCAATGAGGTCAACGAGATTTCC R: ATCACAAATCCGTCCTTACCAG

Five nanograms of DNA were used per reaction, derived from primary populations of *Oxytricha* FACS-derived nuclei, non-specific sorted “debris” particles, whole cell DNA, and lysates of single vegetative *Oxytricha* cells. Enrichment of targets was calculated as $2^{(Cq[\text{Sample}] - Cq[\text{Single Cell}])}$ with a minimum enrichment score of 0.

Hi-C library preparation

Hi-C libraries were built from sorted nuclei largely as in Belaghzal et al.,⁵⁷ starting from the post-cell lysis stage. 1-2 million sorted micronuclei were used as input for vegetative samples, while 250,000-500,000 anlagen were used as input for 24h developmental samples and 80,000 for 48h developmental Hi-C. Fixed, sorted nuclei were pelleted for 5 minutes at 4,000xg and resuspended in 342µl NEBuffer3.1 (New England Biolabs). 38µl 1% sodium dodecyl sulfate was added and samples were incubated for 10 minutes at 65°C. 43µl 10% Triton X-100 was then added. DNA was then digested with 400U DpnII (New England Biolabs) overnight at 37°C in a ThermoMixer with gentle shaking. The next day, the following reagents were added and samples were incubated at 23°C for 4h in a ThermoMixer with intermittent shaking: 1.5µl each of 10mM dCTP, dGTP, and dTTP (Thermo Fisher Scientific), 37.5µl 0.4mM biotin-14-dATP (Thermo Fisher Scientific), 6µl 10x NEBuffer 3.1, 2µl Dnase-free water (Thermo Fisher Scientific), and 10µl 5U/µl DNA polymerase I Klenow (New England Biolabs). Blunt end ligation was then performed at 16°C for 4h in a ThermoMixer with intermittent shaking with the following reagents: 24µl water, 240µl 5x ligation buffer (Invitrogen), 120µl 10% Triton X-100, 10µl 10mg/ml BSA, and 50µl T4 DNA ligase (Invitrogen). Crosslinking was reversed with 50µl

10mg/ml proteinase K and a 2 hour incubation at 65°C, followed by addition of another 50µl proteinase K and overnight incubation at 65°C. The following day, DNA was purified by phenol-chloroform extraction, followed by incubated with 1µl 1mg/ml RnaseA (Thermo Fisher) for 30 minutes at 37°C. Biotin was removed from unligated ends by treating DNA with 5µl 10x NEBuffer 2.1 (New England Biolabs), 0.125µl 10mM dATP, 0.125µl 10mM dGTP, and 5µl 3,000 U/ml T4 DNA polymerase (New England Biolabs) in a total reaction volume of 50µl and incubating at 20°C for 4h followed by 75°C for 20 minutes. DNA was then washed and collected using a 30kD Amicon column (Millipore).

Illumina libraries were then prepared from Hi-C DNA using the NEBNext Ultra II kit (New England Biolabs). DNA was first sheared to 200-300bp using a Covaris M220 ultrasonicator (Covaris). After end preparation, and prior to adapter ligation and PCR, biotin-containing fragments were pulled down with MyOne Streptavidin C1 beads (Thermo Fisher Scientific). Streptavidin beads were twice washed with 400µl Tween wash buffer (5 mM Tris-HCl, pH 8.0; 0.5 mM EDTA, 1 M NaCl, 0.05% Tween 20), rocked for 3 minutes at room temperature, then reclaimed on a magnetic rack. 400µl 2x binding buffer (10 mM Tris-HCl, pH 8.0; 1 mM EDTA, 2 M NaCl) was added, along with the end prepped DNA and TLE buffer (10 mM Tris-HCl; 0.1 mM EDTA, pH 8.0) up to 800µl total. DNA was bound to beads for 15 minutes at room temperature with rotation, after which beads were reclaimed, washed first with 400µl binding buffer, then with 400µl TLE, before being resuspended in 20 µl TLE. The remaining steps of NEBNext Ultra II library prep were performed on-bead. After 3-8 cycles of indexing PCR, final libraries were purified using Ampure XP beads (Beckman Coulter).

Hi-C libraries were analyzed via Bioanalyzer and sequenced on NextSeq 500 instruments (Illumina) at the Columbia Genome Center. All Hi-C datasets were comprised of 2x75 and

2x150 paired-end reads. A summary of all Hi-C datasets is available in Table S1. Aliquots of the vegetative germline Hi-C library were set aside for quality control via forward-forward PCR. Three pairs of primers spanning DpnII GATC motifs and pointing in the same direction were designed, such that no PCR product is expected in the unprocessed germline genome. PCR was performed on unprocessed germline DNA and the vegetative germline Hi-C library to check for amplification exclusively from Hi-C chimeric DNA.

Forward A1: TGTCTCCATCTTGGGAATGAGA

Forward A2: GAGCATCAAAGATGTACCTCAAAG

Forward B1: ATCTGACAGCTTCTGTACGATAAC

Forward B2: ACTGGCTGTGGATTGATTTAAC

Forward C1: GTTGTTTGGCTGCAAGATGAG

Forward C2: TTTCCATAAATAAAGATTCAGCTCTCT

RNase-treated Hi-C

For the RNase treated 24h Hi-C library, cells were pelleted prior to fixation by centrifugation for 2 minutes at 200xg. Cells were resuspended in 50µl 0.5x PBS with 0.02% Tween 20 and incubated on ice for 10 minutes. RNA was digested with 10µl 700U/ml RNase A (Thermo Fisher Scientific) for 30 minutes at room temperature in a Thermomixer at 950 rpm. Under these low salt conditions (70 mM NaCl), RNase A cleaves single- and double-stranded RNA as well the RNA strand in RNA-DNA hybrids. For the control sample, 10µl 0.5x PBS was added rather than RNase A before incubation. Post digestion, samples were brought up to 50ml with Pringsheim media, pelleted, then washed and pelleted again in 50ml Pringsheim media before proceeding with the standard fixation, lysis, and sorting protocol as above. 10% of these

samples were set aside for total RNA isolation using TRIzol (Thermo Fisher Scientific) and quantification of RNA by Qubit (Thermo Fisher Scientific).

***Oxytricha* germline genome scaffolding**

For Nanopore sequencing, DNA from sorted JRB310 *Oxytricha* micronuclei was isolated via phenol-chloroform extraction. The high molecular weight of the MIC DNA was confirmed on a 1% agarose gel. Two biological replicates of 450-550ng were prepared, and library preparation and long-read sequencing of these were performed at Oxford Nanopore Technologies New York lab. The combined Nanopore MIC dataset consisted of 13.8 Gb of long-read data. Reads with a GC content outside of the expected range for *Oxytricha* DNA (20-45%) were discarded. The remaining reads were used to scaffold the original draft germline genome assembly⁴⁰ with LRScf.⁵⁹ Nanopore reads were first mapped to the draft assembly using Minimap2,¹⁰¹ and then LRScf was run with the following settings: maximum overhang length 600, maximum overhang ratio 0.3, minimum overlap ratio 0.3. The LRScf output was then fed into the pipeline for 3D-DNA.⁶⁰ JRB310 vegetative MIC Hi-C reads were mapped to the LRScf output using Juicer (v1.6).⁵⁸ The subsequent merged_nodups.txt file was provided as input to 3D-DNA with the following settings, given the abundance of very short contigs in the draft assembly: size threshold 1000, polisher input size threshold 10000, splitter input size threshold 20000. The resulting .hic matrix and assembly were reviewed in Juicebox Assembly Tools⁶¹ to manually correct any suspected mis-joins. After the 3D-DNA assembly pipeline was completed, a final round of LRScf was performed with the same settings as above. All resulting contigs less than 10 kb were checked against the nr database using BLAST,⁶² and contigs with exclusively prokaryotic hits in the top 10 reported results were removed from the scaffolded assembly. MDS

and TBE annotations from the initial fragmented assembly were mapped onto the scaffolded assembly via BLAST, requiring 98% identity to the queries.

Hi-C data processing

Hi-C Illumina reads were assessed using FastQC¹⁰² and were trimmed using Trimmomatic.¹⁰³ Reads were then mapped to the scaffolded genome assembly using the Juicer (v1.6) pipeline.⁵⁸ For data visualization and subsequent analyses, Juicer output was converted to .h5 matrices at resolutions of 1000, 5000, 10000, and 100000 using HiCExplorer (v3.7.1).^{31,67,68} PCR duplicate removal, KR matrix balancing⁶³ and normalization between samples was performed within HiCExplorer, as were PCA, TAD, and log2 analyses at various resolutions. Raw matrix counts were exported for analysis of interactions between specific bins. Compartments were assigned on the basis of the sign of the first eigenvector following PCA of observed/expected Hi-C matrices. TAD calling was performed using the hicFindTADs function at resolutions of 5000, 10000, and 100000, using f.d.r. correction for multiple testing, minDepth of 3x[resolution], maxDepth 10x[resolution], step equal to [resolution], and delta threshold of 0.05. Overlapping TAD calls from multiple resolutions were merged using the hicMergeDomains function.

ATAC-seq

ATAC-seq (Assay for Transposase-Accessible Chromatin sequencing) was performed on fixed, sorted nuclei following previously published protocols.¹⁰⁴ For vegetative samples, at least 500,000 fixed and sorted micronuclei were used as input for each sample. For developmental samples, at least 100,000 24h anlagen were used as input. Three replicate libraries of each were prepared with varying amounts of Tn5 transposase (Illumina). Sorted nuclei were pelleted for 5 minutes at 4,000xg and resuspended in 50µl lysis buffer (10mM Tris-HCl, pH 7.4; 10mM NaCl;

3mM MgCl₂; 0.05% Igepal CA-630), then centrifuged again at 4,000xg for 5 minutes. Nuclei pellets were resuspended in 50µl 1x Tagment DNA Buffer (Illumina) along with, for each type of input nucleus, 1µl, 2.5µl, or 5µl Tn5 transposase (TDE1 enzyme, Illumina). Transposase reactions were incubated for 30 minutes at 37°C. After transposition, a reverse-crosslinking solution was added up to 200µl, with a final concentration of 50mM Tris-HCl, 1mM EDTA, 1% SDS, 0.2M NaCl, and 5 ng/ml proteinase K, and samples were incubated overnight at 65°C in a Thermomixer at 1,000 rpm. The next day, DNA was extracted via NucleoSpin Mini kit (Macherey-Nagel), and Illumina libraries were prepared with NEBNext Ultra II. Libraries were sequenced on a NextSeq 500 at the Columbia Genome Center, generating between 10 million and 13.5 million 2x75 paired-end reads per sample. The quality of reads was assessed using FastQC¹⁰² and reads were trimmed using Trimmomatic.¹⁰³ Reads were mapped to the scaffolded germline assembly using Bowtie2¹⁰⁵ in “very-sensitive” mode. For peak and coverage analyses, replicates treated with different quantities of Tn5 were merged. Narrow peaks were called from mapped ATAC-seq reads using MACS2.¹⁰⁶ Average coverage along MDSs, IESs, and TBEs was assessed by identifying features of a narrow size range (500-1,000bp for MDSs and IESs, 3-4 kb for TBEs), then building coverage matrices and plotting coverage profiles using Deeptools (v3.5.1),¹⁰⁷ with 500 bp upstream and downstream of features included.

DNA fluorescence *in situ* hybridization

Potential representative 40 kb regions for compartments A and B were identified on the basis of PC1 scores, and probes ranging from 35-40 nucleotides were designed to cover these regions using OligoMiner⁷² and BLAST.⁶² Three regions on which to perform FISH from each compartment were then selected on the basis of available unique sequences, in order to have sufficient specific probe coverage, with a target of 8 probes per kilobase. OligoFISH probe pools

were then ordered from and manufactured by Daicel Arbor Biosciences. For each compartment, one probe pool was labeled with Alexa 488, a second pool was labeled with ATTO 550, and a third probe pool was labeled with ATTO 647N. Probe pools were stored at -20°C at a concentration of 10 pmol/μl in TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0).

Oxytricha cells on which FISH was performed were prepared by mating JRB310 and JRB510 mating types, harvesting at 24h, and fixing for 10 minutes in a 1% formaldehyde (v/v) solution in Pringsheim at room temperature with gentle shaking. Glycine was then added to a final concentration of 125mM to quench fixation, followed by incubation at 4°C for 15 minutes with gentle shaking. Cells were pelleted for 3 minutes at 300xg, washed once with 1xPBS, then pelleted again and stored in 1xPBS at 4°C.

A protocol for 3D-FISH in *Oxytricha* was developed based on previous recommendations for other systems.⁷¹ 24h developmental *Oxytricha* cells were affixed overnight at 4°C to slides coated in poly-L-lysine (Sigma-Aldrich). Cells were washed with 20μl 1X PBS for 1 minute at room temperature. Cells were then permeabilized by treatment with 0.5% Triton X-100 (v/v) in 1X PBS for 30 minutes at room temperature, followed by washing with 1X PBS for 2 minutes at room temperature, and then treatment with 0.1N HCl for 5 minutes at room temperature. Cells were then washed twice with 2X saline-sodium citrate buffer (SSC) (Thermo Fisher Scientific) for two minutes each at room temperature. Pre-hybridization was performed by equilibrating cells for 5 minutes at room temperature in 40% formamide in 2X SSC, followed by treating with 40% formamide in 2X SSC for 20 minutes at 60°C. Meanwhile, pools of probes in hybridization buffer were prepared. For cells probed for compartment A, 1.5 pmol of each compartment A probe pool were used; for cells probed for compartment B, 1.5 pmol of each compartment B probe pool were used. Probes were added to 20μl hybridization buffer consisting of final

concentrations of 40% formamide (v/v), 2X SSC, 0.5U/μl RNase A (Thermo Fisher Scientific), and 10% dextran. The mixture of probes in hybridization buffer was pre-heated at 70°C for 5 minutes and then added to equilibrated cells. Slides were transferred to a humid container, then denatured for 2.5 minutes at 78°C before being incubated at 42°C overnight.

After overnight hybridization, slides were mounted with VECTASHIELD Antifade Mounting Medium with DAPI (Vector Laboratories). Slides were imaged with a Nikon Ti2 AXR MP confocal microscope at 100X with 0.5μm Z steps. Images were processed and single Z-slices were selected using Fiji.¹⁰⁸

3D reconstructions of scaffolds guided by the empirically measured pairwise *cis*-interactions, captured by Hi-C experiments

The 3D reconstructions were produced using a modified 4DHiC method,⁷³ which enables interactions between multiple chromosomes. Here, an extended, disordered beads-on-springs polymer chain is collapsed using additional constraints defined by the empirically determined Hi-C contact map, using contacts exceeding a threshold and separate levels of constraints are used for intra- and inter-chromosomal contacts.⁷⁴ Specifically, all simulations were conducted using the Large-Scale Atomic/Molecular Massively Parallel Simulator (LAMMPS).⁷⁵ Each scaffold consists of several coarse-grained (CG) beads, with each bead representing 20 kb resolution. The total number of CG beads varies depending on the length of the individual scaffold. For example, Scaffold 528 (5.06 Mb) was represented by a polymer with 253 coarse-grained (CG) beads. The initial structure is modeled as a connected chain, which was generated through a self-avoiding random walk algorithm. The pair interactions between the CG beads were simulated using the FENE and Lennard-Jones potentials. Experimental Hi-C constraints were applied as harmonic constraints to simulate cross-linking contacts. To minimize noise from

low-intensity signals, the frequency of Hi-C contacts was set to be greater than 10 for all intra- and 20 for all inter- scaffold interaction, respectively. The simulation temperature was maintained at a constant value using the Langevin thermostat, set at 1 in reduced units, with a damping coefficient of $1 \tau^{-1}$. A timestep of 0.01τ was utilized, where τ represents the reduced (Lennard-Jones) time. The polymer, subjected to harmonic constraints, was analyzed using Brownian dynamics under implicit solvent conditions. To compare the similarities between the experimental and simulated Hi-C maps, the Pearson correlation coefficient was calculated. The built-in LAMMPS routines were employed to calculate the radius of gyration (R_g). To understand the spatial organization of compartments in 3D, the experimental A and B compartment PC_1 scores were mapped onto simulated 3D models.^{76,77}

DNA-RNA immunoprecipitation

DNA-RNA hybrids were captured from whole cell DNA using the S9.6 antibody (Sigma-Aldrich). DNA was purified via phenol-chloroform extraction from a mating timecourse, with samples taken every 12 hours. Half the extracted DNA was digested with $15 \mu\text{l}$ $5\text{U}/\mu\text{l}$ RNaseH in 1x Reaction Buffer (Thermo Fisher Scientific) for 2h at 37°C and then purified again by phenol-chloroform. 130ng DNA was loaded onto a nitrocellulose membrane for each timepoint and each condition (+/- RNaseH). The membrane was blocked for 2.5h while rocking in TBS-T (0.05% Tween-20, 20 mm Tris-HCl, 150 mm NaCl, pH 7.5) with 5% BSA (w/v). The membrane was then incubated with the S9.6 antibody for 1 hour at a 1:2,000 dilution in TBS-T with 0.1% BSA (w/v), then washed three times in TBS-T for 5 minutes each, followed by incubation with secondary goat anti-mouse horseradish peroxidase-conjugated secondary antibody diluted 1:1,000 in TBS-T with 5% BSA for 30 minutes. The membrane was washed for 15 minutes in TBS-T and then for 5 minutes in TBS (20 mm Tris-HCl, 150 mm NaCl, pH 7.5), after which it

was placed within a plastic sheet, Pierce ECL substrate (Thermo Fisher Scientific) mixture was added, and the membrane was kept out of light for 1 minute before being imaged via chemiluminescence on an Amersham Imager 600 (GE).

For DRIP-seq library prep, genomic DNA from approximately 500,000 cells each was extracted from a mating timecourse every 12h via phenol-chloroform up until 60h. Immunoprecipitation generally followed the protocol in Wahba et al.⁸⁸ Half of each sample was used for RNaseH-treated controls, which were incubated with 10 μ l 5U/ μ l RNaseH in 1x Reaction Buffer (Thermo Fisher Scientific) overnight at 37°C before being repurified by NucleoSpin (Macherey-Nagel). DNA from RNaseH-treated and -untreated samples was then sonicated in an S220 Covaris machine to 350bp. For each sample, 5 μ g of S9.6 antibody was bound to 50 μ l Protein A Dynabeads (Thermo Fisher Scientific) for 1 hour at 4°C. Beads were recovered on a magnetic rack, and sonicated DNA samples were then incubated with S9.6-bound Dynabeads in 100 μ l 1X FA (1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, 50mM HEPES, 150mM NaCl, 1mM EDTA) for 2h rotating at 4°C. Beads were then recovered and serially washed once with 1X FA, once with 1X FA with 0.5M NaCl, once with lithium chloride detergent solution (0.25M LiCl, 0.5% NP-40, 0.5% sodium deoxycholate, 1mM EDTA, and 10mM Tris-HCl), and twice with TE buffer. DNA was eluted from the beads with 300 μ l elution buffer (1% SDS, 0.1M NaHCO₃). After elution, 5 μ l RNase Cocktail (Thermo Fisher Scientific) in 50 μ l 0.5X PBS was added to the beads, followed by rotation at room temperature for 15 minutes, after which that supernatant was reclaimed and then added to the initial eluate. Samples were then treated with 1 μ l 20mg/ml proteinase K (Takara Bio) for 1 hour at 42°C in a Thermomixer at 1000 rpm. DNA was then purified via phenol-chloroform extraction, and Illumina libraries were prepared with NEBNext Ultra II. Libraries were sequenced on a NextSeq

500 at the Columbia Genome Center, yielding at least 4.5 million 1x75 single-end reads per sample. The quality of reads was assessed using FastQC¹⁰² and were trimmed using Trimmomatic,¹⁰³ before being mapped to the *Oxytricha* somatic reference genome³⁸ using BWA MEM.¹⁰⁹ 5,000 somatic contigs were identified between 1.5 kb and 3 kb and bearing single genes on the same strand, and DRIP samples at 0h and 24h were normalized to their RNaseH-digested controls using the difference operation in Deeptools bamCompare, after which the resulting matrices were averaged over the selected somatic contigs and plotted in Deeptools.¹⁰⁷

Quantification and statistical analysis

All statistical tests were performed in R (v4.1.2) and described in figure legends.