



OPEN

Transcription-dependent domain-scale three-dimensional genome organization in the dinoflagellate *Breviolum minutum*

Georgi K. Marinov^{1,9}✉, Alexandro E. Trevino^{2,3,9}, Tingting Xiang^{4,5,9}, Anshul Kundaje^{6,1,6}, Arthur R. Grossman⁴ and William J. Greenleaf^{1,2,7,8}✉

Dinoflagellate chromosomes represent a unique evolutionary experiment, as they exist in a permanently condensed, liquid crystalline state; are not packaged by histones; and contain genes organized into tandem gene arrays, with minimal transcriptional regulation. We analyze the three-dimensional genome of *Breviolum minutum*, and find large topological domains (dinoflagellate topologically associating domains, which we term ‘dinoTADs’) without chromatin loops, which are demarcated by convergent gene array boundaries. Transcriptional inhibition disrupts dinoTADs, implicating transcription-induced supercoiling as the primary topological force in dinoflagellates.

The three-dimensional (3D) genome architecture of cells has functional consequences for gene regulation, organismal development, DNA replication and mutational processes. Topologically associating domains (TADs) and compartments on the sub-megabase scale are conserved architectural features of eukaryote genomes, defined by increased intradomain contact frequencies and interdomain contact insulation¹. Mechanisms known to drive the folding of such domains include constraints on cohesin-mediated loop extrusion—imposed in part by CTCF in vertebrates—and self-associations between similar chromatin states². Other mechanisms, including insulation of domains by polymerases or specific boundary proteins, have also been proposed to play roles in genome architecture³. However, the extent to which genome function may influence genome folding, for example, through transcriptional activity, is poorly understood. There has also been little exploration of 3D organization across eukaryotes, even though major deviations from conventional norms are known to exist, presenting natural experiments that may reveal deeper underlying organizational principles masked in other lineages.

Dinoflagellates are the most radical such departure. They are a diverse, widespread clade playing major roles in aquatic ecosystems, for example, as symbionts of corals, providing the metabolic basis for reef ecosystems. Dinoflagellates possess numerous highly divergent molecular features⁴, including, uniquely among eukaryotes, the loss of nucleosomal packaging of chromatin. Histones are extremely conserved across eukaryotes, were present in their current form already in the last eukaryotic common ancestor⁵, and they and their post-translational modifications are pivotal to all biochemical processes involving chromatin.

Dinoflagellates are the sole known exception. Their chromosomes exist in a liquid crystalline state and are permanently condensed throughout the cell cycle, and, although highly divergent histone genes are retained in their genomes⁶, a combination of virus-derived nucleoproteins and bacterial-derived histone-like proteins have taken over as the main packaging components⁷. Dinoflagellate genomes are often huge (up to ≥ 200 gigabases), genes are organized into tandem gene arrays, individual messenger RNAs are generated through *trans*-splicing and transcriptional regulation is largely absent⁷. These fascinating features simultaneously pose intriguing questions regarding the adaptation of transcriptional and regulatory mechanisms to the absence of nucleosomes, and provide a unique opportunity to explore the biophysical forces underlying genomic organization in the context of a large eukaryotic genome nearly devoid of nucleosomes.

To explore these questions, we applied chromosome conformation capture using Hi-C on the coral symbiont *Breviolum minutum*. We generated multiple libraries under standard growth conditions and for cells grown at elevated temperature, obtaining ~ 150 – 220 million Hi-C contacts for each (Supplementary Table 1). We pooled these libraries to generate a chromosome-level scaffolding of the previously fragmented *B. minutum* assembly⁸. We identified 91 major pseudochromosomes (≥ 500 kilobases (kb)), encompassing $\sim 94\%$ of the total sequence (Fig. 1a,b and Supplementary Fig. 1a), the longest being ~ 11 megabases (Mb) in size, with a median length of 6.7 Mb (Supplementary Fig. 1a). At 1-Mb resolution, they exhibit a bipartite (occasionally tripartite) structure (Supplementary Fig. 2).

Additional untreated libraries (Supplementary Table 1) were combined to generate an even-higher-resolution map (~ 1.05 billion contacts), which was used to examine fine-scale features of topological organization. High-resolution (5 kb) maps revealed well-defined (comparably so with those observed in mammals) topological domains, ≤ 200 kb to ≥ 2 Mb in size (Fig. 1b,e and Supplementary Figs. 3–12). In mammals, most TAD boundaries are demarcated by CTCF sites blocking loop extrusion, reflected in Hi-C maps by chromatin loops and ‘stripes’. We observed no loop or stripe features in *B. minutum* (Fig. 1c), suggesting a different mechanism for the formation of dinoflagellate TADs, which we term ‘dinoTADs’. Omitting the denaturation step in the Hi-C protocol, which should better preserve protein–protein contacts, strongly accentuated

¹Department of Genetics, Stanford University, Stanford, CA, USA. ²Center for Personal Dynamic Regulomes, Stanford University, Stanford, CA, USA.

³Department of Bioengineering, Stanford University, Stanford, CA, USA. ⁴Department of Plant Biology, Carnegie Institution for Science, Stanford, CA, USA.

⁵Department of Biological Sciences, University of North Carolina at Charlotte, Charlotte, NC, USA. ⁶Department of Computer Science, Stanford University, Stanford, CA, USA. ⁷Department of Applied Physics, Stanford University, Stanford, CA, USA. ⁸Chan Zuckerberg Biohub, San Francisco, CA, USA.

⁹These authors contributed equally: Georgi K. Marinov, Alexandro E. Trevino, Tingting Xiang. ✉e-mail: marinovg@stanford.edu; wjg@stanford.edu

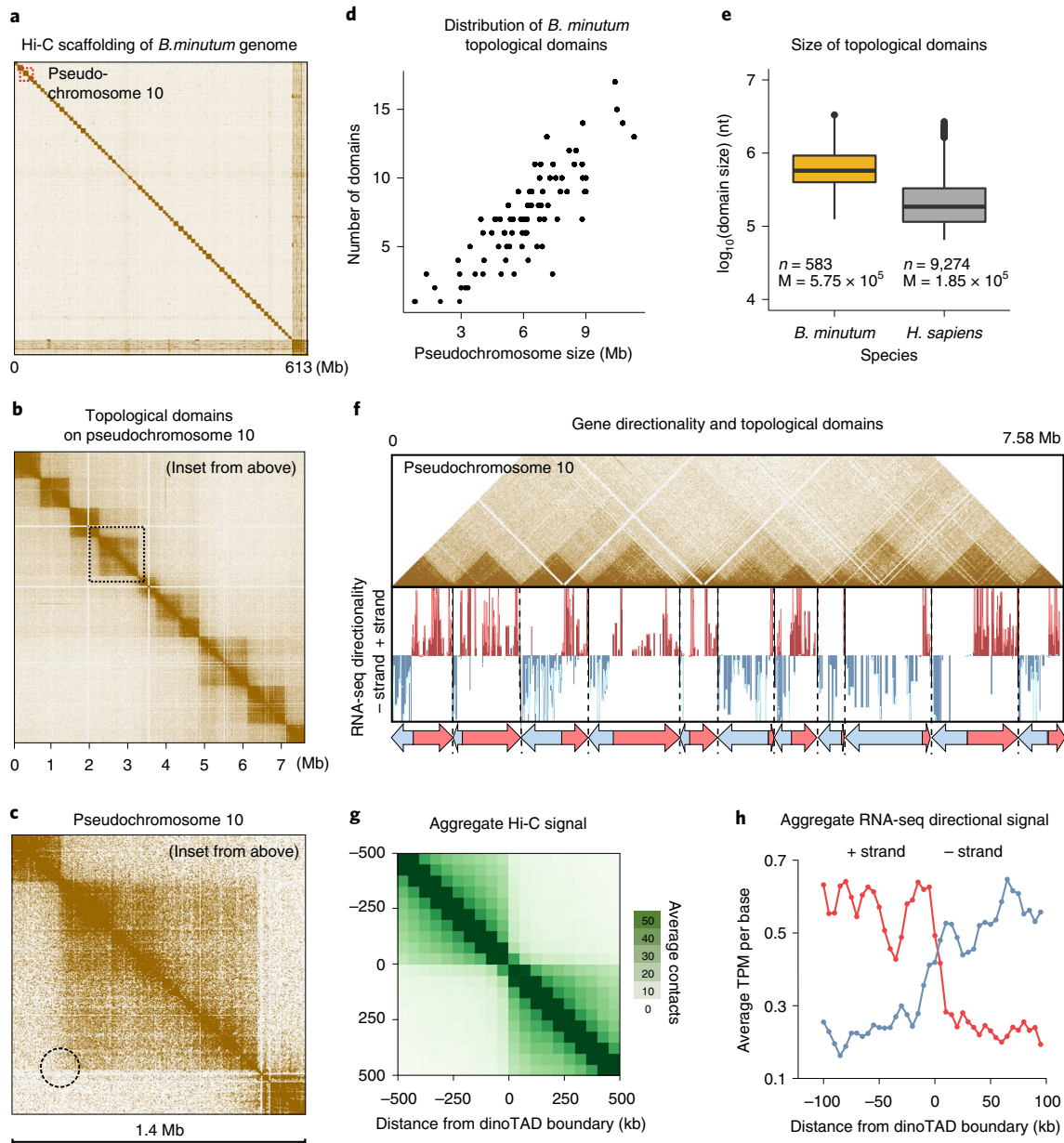


Fig. 1 | *B. minutum* genome is physically partitioned into dinoTADs defined by tandem gene arrays. **a**, Hi-C scaffolding of the *B. minutum* draft genome assembly. **b**, Inset from **a**. KR-normalized 5-kb resolution Hi-C map for pseudochromosome 10. **c**, Inset from **b**. Hi-C loops and stripes are not observed in dinoTADs (dotted circle notes where a loop would be). **d**, Scaling of chromosome size with dinoTAD number. **e**, Comparison of human and *B. minutum* topological domain sizes. Box plots show the 25th, 50th and 75th percentiles; whiskers show the 5–95% intervals; $n = 583$ for dinoTADs; $n = 9,274$ for human TADs. **f**, Hi-C map (5-kb resolution) for pseudochromosome 10 together with forward- and reverse-strand transcript levels and gene arrays. **g**, Average Hi-C contacts across dinoTAD boundaries (dinoTADs were called at a 50-kb resolution using HiCEplorer²²; see the Methods for details). **h**, Average forward- and reverse-strand RNA-seq levels across dinoTAD boundaries. *H. sapiens*, *Homo sapiens*; KR, KR normalization; M, mean (bp); nt, nucleotides; TPM, transcripts per million.

dinoTADs, but still did not reveal signs of loops or loop extrusion domains (Supplementary Fig. 14). Detected dinoTAD number correlated with chromosome size (Fig. 1d), and observed dinoTADs were considerably larger than mammalian TADs (Fig. 1e).

We next compared Hi-C maps with available annotation features. Remarkably, we found that each dinoTAD corresponded to a pair of divergent gene arrays (Fig. 1f), and dinoTAD boundaries coincided with convergence between gene arrays (Fig. 1g,h).

Numerous models for dinoflagellate chromosome organization have been suggested since the 1960s, primarily based on electron

microscopy. These include proposals that chromosomes are organized as ‘toroidal chromonemas’⁹, as ‘stacks of disks’¹⁰, as ‘cored pineapples’¹¹ or around ‘central core fibers’¹². Most of these models imply specific topological constraints maintaining the proposed shapes and are not directly reconcilable with our Hi-C observations.

Instead, the correspondence between dinoTADs and gene arrays suggested a role for transcription in their formation. Although TADs form independently of transcription in animal cells, transcription-induced self-interacting domains have been previously demonstrated in bacteria¹³, and similar mechanisms have been

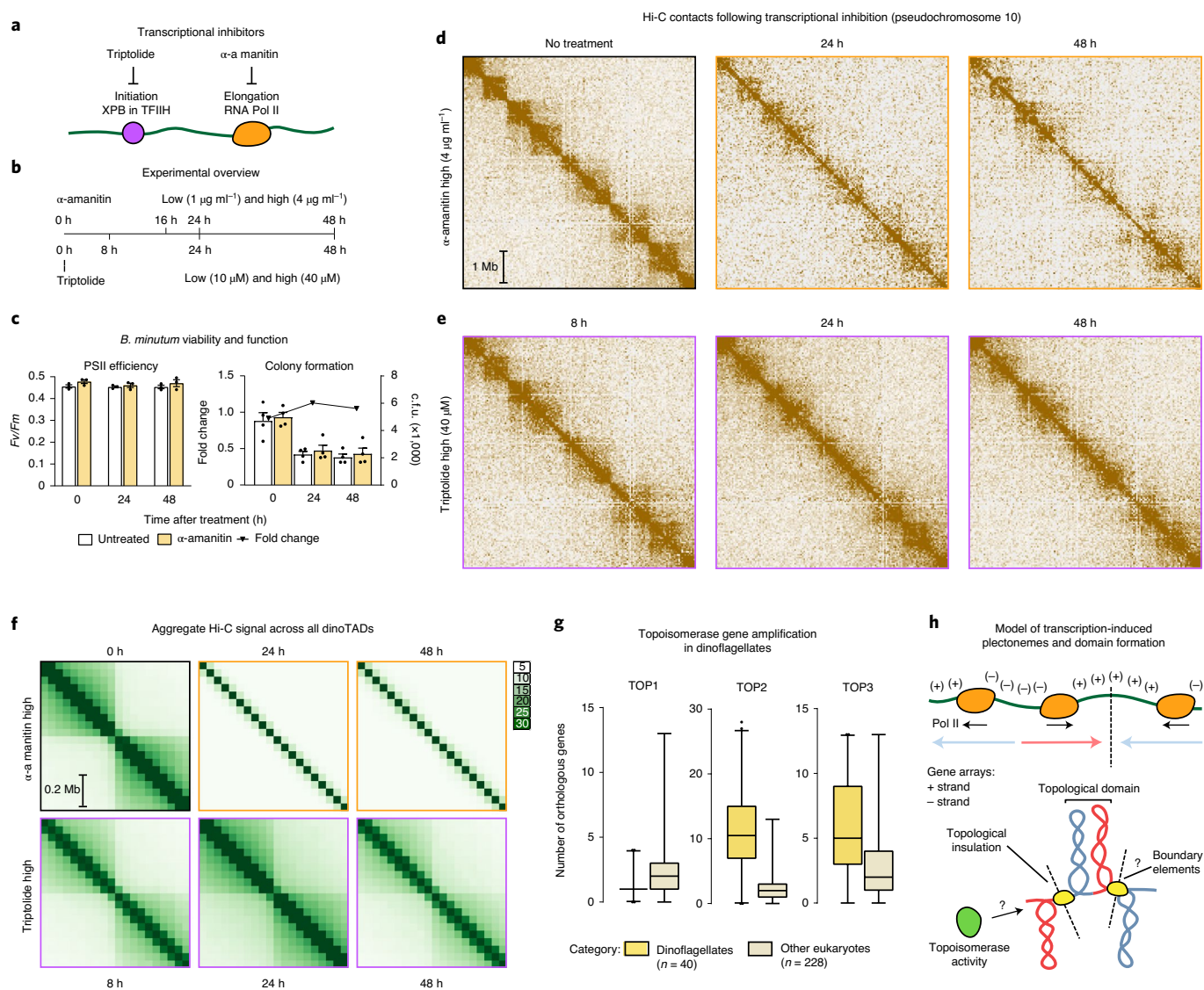


Fig. 2 | Decompaction of dinoTADs upon application of transcriptional inhibitors and the transcription-induced supercoiling model for their formation.

Shown is pseudochromosome 10 as in Fig. 1. **a**, Outline of transcription inhibition experiments. **b**, Outline of transcription inhibition time course. **c**, Comparison of cell function, measured by PSII photosynthetic efficiency (left; $n=3$ biological replicates for each condition), and cell viability, measured by colony formation (right; $n=4$ biological replicates for each condition), between α -amanitin-treated and untreated cells. Treatment with α -amanitin does not affect PSII activity (Student t -test, $P=0.75304979$ for 0 h, $P=0.442327976$ for 24 h, $P=0.23349803$ for 48 h). Error bars show mean \pm s.d. **d**, KR-normalized Hi-C maps (50-kb resolution) show marked loss of dinoTADs after α -amanitin treatment. **e**, Hi-C maps show reduction of insulation at dinoTAD boundaries after triptolide treatment. **f**, Metaplots of Hi-C signal around domain boundaries (50-kb resolution). **g**, Amplification of *TOP2* and *TOP3* topoisomerases in dinoflagellates (based on MMETSP²³ transcriptome assemblies). Box plots show the 25th, 50th and 75th percentiles; whiskers show the 5–95% intervals. The dinoflagellate ($n=41$) and nondinoflagellate ($n=243$) species shown are the ones from Supplementary Table 2. **h**, Transcription-induced supercoiling as driver of dinoflagellate chromatin folding. Transcribing polymerases introduce negative/positive DNA supercoiling behind/ahead of the transcription machinery. Interactions within supercoiled domains could explain the physical association of divergently oriented arrays. Topological insulation could be driven by supercoiling-related effects, or by specific boundary elements. c.f.u., colony-forming units.

proposed to explain some topological features in fission yeast¹⁴. We also note that a handful of models of dinoflagellate chromosome structure have suggested the presence of coil/plectoneme-like features^{15,16}, but without relating them to gene arrays and transcription. This model is also supported by the observation that frequently each dinoTAD can be divided into more diffuse ‘sub-dinoTADs’ corresponding to the two individual gene arrays in a pair (Fig. 1c and Supplementary Figs. 3–12), which could be the result of torsion generated in each direction of transcription.

The model makes a clear prediction—inhibition of transcription should result in dinoTADs decompaction. To test this

relationship, we first compared Hi-C maps for cells grown at 34 °C versus 27 °C, as heat stress could result in general transcription reduction¹⁷. We observed mild decompaction of dinoTADs at 34 °C, although domains remained intact (Supplementary Figs. 19–21).

We next carried out chemical transcription inhibition experiments. Since transcription inhibition conditions for *B. minutum* are not well established, we chose two inhibitors—triptolide and α -amanitin—with distinct mechanisms of action, and assayed multiple time points and doses (Fig. 2a,b). Amanitin directly inhibits RNA polymerase II and is slow acting, while triptolide quickly blocks initiation by targeting the TFIIH XPB subunit¹⁸. While dinoflagellate RNA

polymerase II has been reported to be sensitive to α -amanitin, it is possible that the sensitivity is somewhat partial¹⁹; in addition, the *B. minutum* XPB homolog is highly divergent⁸, and thus a moderate inhibition effect is not unexpected. We therefore carried out several experiments to directly estimate the extent of transcription inhibition. Direct metabolic labeling approaches²⁰ were unsuccessful, as it appears that Symbiodiniaceae cells are impermeable to nucleotide and nucleoside analogs such as 4-thiouridine and 4-thiouracil. We were, however, able to qualitatively assess inhibition using the proxy of nascent RNA, as measured by the proportion of unspliced reads in poly(A)⁺ RNA-sequencing (RNA-seq) datasets (Supplementary Fig. 30). We observed more than 50% reduction in unspliced reads in both α -amanitin and triptolide cells after 48 h, suggesting that transcription was indeed inhibited. We also did not observe large-scale changes in the levels of individual transcripts (Supplementary Fig. 31). Finally, even at high doses, α -amanitin treatment did not detectably affect photosynthetic efficiency or cell viability relative to controls (Fig. 2c), excluding cell death as a confounding factor.

Strikingly, α -amanitin treatment resulted in a dose-dependent, progressive dinoTAD decompaction (Fig. 2d,f and Supplementary Figs. 22–25). These effects were observed in both technical and biological replicates (Supplementary Figs. 22–25). We also observed clear dose-dependent blurring of dinoTAD boundaries after triptolide treatment, although broad dinoTAD-like structures remained visible to a greater extent than in α -amanitin-treated cells (Fig. 2e,f and Supplementary Figs. 26–29).

These experiments support a transcription-induced supercoiling model for dinoTAD formation. Torque generated by active polymerases produces positive/negative supercoiling ahead of/behind the transcription bubble. This can alter the twist of the double helix or induce superhelical writhe, which in turn can be accommodated through nucleosome remodeling, local alterations in DNA secondary structure, or formation of structures such as plectonemes²¹, from which we would expect strong Hi-C signals comprising our observed domains.

Although other topological constraints might also be involved, supercoiling-induced plectoneme formation over gene arrays is an intuitive mechanistic explanation for the presence of dinoTADs. An examination of dinoflagellate gene repertoires also corroborates this model, revealing a striking, dinoflagellate-specific expansion of topoisomerase II- and topoisomerase III-like genes (Fig. 1d, Supplementary Fig. 18 and Supplementary Table 2), further suggestive of contending with increased levels of writhed forms of helical twist.

Comparison with self-interacting domains in bacteria or *Schizosaccharomyces pombe* shows much stronger topological insulation for dinoTADs (Supplementary Figs. 15 and 16). Remarkably, no TAD domains are observed in kinetoplastids, the other lineage with long gene arrays and no transcriptional regulation (Supplementary Fig. 17).

These differences can be rationalized by the unusual dinoflagellate properties. First, neither bacteria nor yeast possess comparably long gene arrays and transcription in those species is highly nonuniform; less transcription-induced torsional stress is therefore expected. Nucleosome loss is the second, and most salient, difference. Single mammalian genes as long as dinoTADs are quite common, yet contact domains aligning with gene boundaries is not apparent in mammalian Hi-C maps, nor is it seen in kinetoplastids, which have gene arrays but also conventional chromatin. We therefore hypothesize that plectonemic structures form due to transcription-induced supercoiling in the nucleosome-depleted genomes of dinoflagellates, while, in other eukaryotes, a combination of the wrapping of DNA around nucleosomes, interactions between nucleosomes and accumulation of DNA twist prevents their formation (Fig. 2h).

These results generate a number of open questions. How exactly are boundaries between dinoTADs formed mechanistically?

Specific boundary elements of markedly different chromatin states could exist; alternatively, these boundaries may self-organize purely through torsion-related mechanisms. The roles that dinoflagellates' divergent histone genes play are also not clear. Finally, the relationship between Hi-C features and the 'toroidal chromonemas'⁹ observed by electron microscopy remains unknown. Answers to these questions, together with the dissection of specific roles of different topoisomerase classes, will help to fully elucidate the interplay between packaging proteins, transcription-induced torsional stress and genome folding in dinoflagellates.

These observations also identify transcription-induced torsional stress as a key direction of future studies in eukaryotes generally. The strength of dinoTADs underlines the potency of this fundamental biological process for generating topological structure. The precise manner by which torsion is accommodated as twist and writhe, as well as its consequences for regulatory protein occupancy, transcriptional activity and other chromatin processes, such as the behavior of ATP-dependent chromatin remodelers, are exciting questions remaining to be unraveled.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41588-021-00848-5>.

Received: 30 June 2020; Accepted: 17 March 2021;

Published online: 29 April 2021

References

- Szabo, Q., Bantignies, F. & Cavalli, G. Principles of genome folding into topologically associating domains. *Sci. Adv.* **5**, eaaw1668 (2019).
- Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
- Rowley, M. J. et al. Evolutionarily conserved principles predict 3D chromatin organization. *Mol. Cell* **67**, 837–852 (2017).
- Hackett, J. D., Anderson, D. M., Erdner, D. L. & Bhattacharya, D. Dinoflagellates: a remarkable evolutionary experiment. *Am. J. Bot.* **91**, 1523–1534 (2004).
- Postberg, J., Forcob, S., Chang, W. J. & Lipps, H. J. The evolutionary history of histone H3 suggests a deep eukaryotic root of chromatin modifying mechanisms. *BMC Evol. Biol.* **10**, 259 (2010).
- Marinov, G. K. & Lynch, M. Diversity and divergence of dinoflagellate histone proteins. *G3 (Bethesda)* **6**, 397–422 (2015).
- Janouskovec, J. et al. Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics. *Proc. Natl Acad. Sci. USA* **114**, E171–E180 (2017).
- Shoguchi, E. et al. Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. *Curr. Biol.* **23**, 1399–1408 (2013).
- Oakley, B. R. & Dodge, J. D. Evidence for a double-helically coiled toroidal chromonema in the dinoflagellate chromosome. *Chromosoma* **70**, 277–291 (1979).
- Livolant, F. & Bouligand, Y. New observations on the twisted arrangement of dinoflagellate chromosomes. *Chromosoma* **68**, 21–44 (1978).
- Levi-Setti, R., Gavrilov, K. L. & Rizzo, P. J. Divalent cation distribution in dinoflagellate chromosomes imaged by high-resolution ion probe mass spectrometry. *Eur. J. Cell Biol.* **87**, 963–976 (2008).
- Spector, D. L. & Triemer, R. E. Chromosome structure and mitosis in the dinoflagellates: an ultrastructural approach to an evolutionary problem. *Biosystems* **14**, 289–298 (1981).
- Le, T. B., Imakaev, M. V., Mirny, L. A. & Laub, M. T. High-resolution mapping of the spatial organization of a bacterial chromosome. *Science* **342**, 731–734 (2013).
- Benedetti, F., Racko, D., Dorier, J., Burnier, Y. & Stasiak, A. Transcription-induced supercoiling explains formation of self-interacting chromatin domains in *S. pombe*. *Nucleic Acids Res.* **45**, 9850–9859 (2017).
- Livolant, F. & Bouligand, Y. Double helical arrangement of spread dinoflagellate chromosomes. *Chromosoma* **80**, 97–118 (1980).
- Wong, J. T. Y. Architectural organization of dinoflagellate liquid crystalline chromosomes. *Microorganisms* **7**, 27 (2019).
- Levin, R. A. et al. Sex, scavengers, and chaperones: transcriptome secrets of divergent symbiodinium thermal tolerances. *Mol. Biol. Evol.* **33**, 2201–2215 (2016).

18. Bensaude, O. Inhibiting eukaryotic transcription: which compound to choose? How to evaluate its activity? *Transcription* **2**, 103–108 (2011).
19. Rizzo, P. J. RNA synthesis in isolated nuclei of the dinoflagellate *Cryptocodinium cohnii*. *J. Protozool.* **26**, 290–294 (1979).
20. Herzog, V. A. et al. Thiol-linked alkylation of RNA to assess expression dynamics. *Nat. Methods* **14**, 1198–1204 (2017).
21. Teves, S. S. & Henikoff, S. DNA torsion as a feedback mediator of transcription and chromatin dynamics. *Nucleus* **5**, 211–218 (2014).
22. Ramirez, F. et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* **9**, 189 (2018).
23. Keeling, P. J. et al. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* **12**, e1001889 (2014).

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

Methods

Except where otherwise stated, computational analyses were carried out using custom-written Python scripts.

***B. minutum* cell culture.** The clonal axenic *B. minutum* strain SSB01 was used in all experiments. Stock cultures were grown as previously described^{24,25} in Daigo's IMK medium for marine microalgae (Wako Pure Chemicals) supplemented with casein hydrolysate (IMK + Cas) at 27°C at a light intensity of 10 μmol photons m⁻² s⁻¹ from Philips ALTO II 25-W bulbs on a 12-h-light/12-h-dark cycle. The medium was prepared in artificial seawater.

Transcription inhibition experiments. For α-amanitin treatment, *B. minutum* cells at a density of ~1 × 10⁶ cells per ml were treated with α-amanitin (Sigma-Aldrich, catalog no. A2263) at concentrations of 1 μg ml⁻¹ ('normal' dose) and 4 μg ml⁻¹ ('high' dose).

Samples were collected at 0, 24 and 48 h after treatment.

For triptolide treatment, *B. minutum* cells at a density of ~1 × 10⁶ cells per ml were treated with triptolide (Sigma-Aldrich, catalog no. T3652) at concentrations of 10 μM ('normal' dose) and 40 μM ('high' dose). Samples were collected at 0, 8, 24 and 48 h after treatment.

Cell viability measurements. *Photosynthetic activity.* Maximum quantum yield of photosystem II, $F_v/F_m = (F_m - F_0)/F_m$, where F_m is maximum fluorescence, F_v is variable fluorescence, and F_0 is minimum fluorescence, was used to indicate photosynthetic function. *B. minutum* cultures (approximately 10⁶ cells per ml) were collected and dark adapted for 5 min, and F_v/F_m was determined using a Dual Pam-100 fluorometer (Heinz Walz).

Colony formation assay. Fresh SSB01 cells were sampled at 0, 24 and 48 h after the treatment with transcription inhibitor α-amanitin. For each condition, cell suspensions were diluted 1:5 and 1:10 before plating 1 μl of each dilution on marine broth (BD) agar plates. Plates were incubated at 27°C at a light intensity of 10 μmol photons m⁻² s⁻¹. Cell numbers on each plate were counted after 3 weeks.

Hi-C experiments. The in situ Hi-C procedure used to map 3D genomic interactions in *B. minutum* was adapted from previous studies² as follows:

B. minutum SSB01 cells were first crosslinked using 37% formaldehyde (Sigma) at a final concentration of 1% for 15 min at room temperature. Formaldehyde was then quenched using 2.5 M glycine at a final concentration of 0.25 M. Cells were subsequently centrifuged at 2,000g for 5 min, washed once in 1× PBS and stored at -80°C.

Cell lysis was initiated by incubation with 250 μl of cold Hi-C lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2% Igepal CA630) on ice for 15 min, followed by centrifugation at 2,500g for 5 min, a wash with 500 μl of cold Hi-C lysis buffer and centrifugation at 2,500g for 5 min. The pellet was then resuspended in 50 μl of 0.5% SDS and incubated at 62°C for 10 min (except for the 'no-denaturation' sample, for which the pellet was resuspended in 50 μl of H₂O). SDS was quenched by adding 145 μl of H₂O and 25 μl of 10% Triton X-100 and incubating at 37°C for 15 min.

Restriction digestion was carried out by adding 25 μl of 10× NEBuffer 2 and 100 U of the MboI restriction enzyme (NEB, R0147) and incubating for ≥2 h at 37°C in a Thermomixer at 900 r.p.m. The reaction was then incubated at 62°C for 20 min to inactivate the restriction enzyme.

Fragment ends were filled in by adding 37.5 μl of 0.4 mM biotin-14-dATP (ThermoFisher Scientific, no. 19524-016); 1.5 μl each of 10 mM dCTP, dGTP and dTTP; and 8 μl of 5 U μl⁻¹ DNA Polymerase I Large (Klenow) Fragment (NEB M0210). The reaction was then incubated at 37°C in a Thermomixer at 900 r.p.m. for 45 min.

Fragment end ligation was carried out by adding 663 μl of H₂O, 120 μl of 10× NEB T4 DNA ligase buffer (NEB B0202), 100 μl of 10% Triton X-100, 12 μl of 10 mg ml⁻¹ BSA (100× BSA, NEB) and 5 μl of 400 U μl⁻¹ T4 DNA ligase (NEB M0202), and incubating at room temperature for ≥4 h with rotation.

Nuclei were then pelleted by centrifugation at 3,500g for 5 min; the pellet was resuspended in 200 μl of chromatin immunoprecipitation elution buffer (1% SDS, 0.1 M NaHCO₃), Proteinase K was added and the mixture was incubated at 65°C overnight to reverse crosslinks.

After addition of 600 μl of 1× TE buffer, DNA was sonicated using a Qsonica S-4000 with a 1/16" tip for 3 min, with 10-s pulses at intensity 3.5, and 20 s of rest between pulses. DNA was then purified using the MinElute PCR Purification Kit (Qiagen no. 28006), with elution in a total volume of 300 μl of 1× EB buffer.

For streptavidin pulldown of biotin-labeled DNA, 150 μl of 10 mg ml⁻¹ Dynabeads MyOne Streptavidin T1 beads (Life Technologies, 65602) were separated on a magnetic stand, then washed with 400 μl of 1× Tween washing buffer (TWB; 5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl, 0.05% Tween 20). The beads were resuspended in 300 μl of 2× binding buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 2 M NaCl), the sonicated DNA was added and the beads were incubated for ≥15 min at room temperature on a rotator. After separation on a magnetic stand, the beads were washed with 600 μl of 1× TWB, and heated at 55°C

in a Thermomixer with shaking for 2 min. After removal of the supernatant on a magnetic stand, the TWB wash and 55°C incubation were repeated.

Final libraries were prepared on beads using the NEBNext Ultra II DNA Library Prep Kit (NEB, no. E7645) as follows. End repair was carried out by resuspending beads in 50 μl of 1× EB buffer, and adding 3 μl of NEB Ultra End Repair Enzyme and 7 μl of NEB Ultra End Repair Buffer, followed by incubation at 20°C for 30 min and then at 65°C for 30 min.

Adapters were ligated to DNA fragments by adding 30 μl of Blunt Ligation mix, 1 μl of Ligation Enhancer and 2.5 μl of NEB Adapter; incubating at 20°C for 20 min; adding 3 μl of USER enzyme; and incubating at 37°C for 15 min.

Beads were then separated on a magnetic stand, and washed with 600 μl of TWB for 2 min at 55°C, 1,000 r.p.m. in a Thermomixer. After separation on a magnetic stand, beads were washed in 100 μl of 0.1× TE buffer, then resuspended in 16 μl of 0.1× TE buffer and heated at 98°C for 10 min.

For PCR, 5-μl samples of each of the i5 and i7 NEB Next sequencing adapters were added together with 25 μl of 2× NEB Ultra PCR Mater Mix. PCR was carried out with a 98°C incubation for 30 s and 12 cycles of 98°C for 10 s, 65°C for 30 s and 72°C for 1 min, followed by incubation at 72°C for 5 min.

Beads were separated on a magnetic stand, and the supernatant was cleaned up using 1× AMPure XP beads.

Libraries were sequenced in a paired-end format on an Illumina NextSeq instrument using NextSeq 500/550 high-output kits (either 2×75 or 2×36 cycles).

Hi-C data processing and assembly scaffolding. As an initial step, Hi-C sequencing reads from all libraries were trimmed of adapter sequences, pooled together and processed against the previously published *B. minutum* assembly⁸ using the Juicer pipeline²⁶ for analyzing Hi-C datasets (v.1.8.9 of Juicer Tools).

The resulting Hi-C matrices were then used as input to the 3D DNA pipeline²⁷ for automated scaffolding with the following parameters: --editor-coarse-resolution 5000 --editor-coarse-region 5000 --polisher-input-size 100000 --polisher-coarse-resolution 1000 --polisher-coarse-region 300000 --splitter-input-size 100000 --splitter-coarse-resolution 5000 --splitter-coarse-region 300000 --sort-output --build-gapped-map -r 10 -i 5000.

Manual correction of obvious assembly and scaffolding errors was then carried out using Juicebox²⁶.

After finalizing the scaffolding, Hi-C reads were reprocessed against the new assembly using the Juicer pipeline. This was done individually for each library as well as together for the pooled set of reads.

Data were extracted from the final read matrices using the Juicer suite of tools for Hi-C data analysis.

Identification of Hi-C domains. Hi-C matrices were first converted to *cool* format using HiCExplorer²² 'hicConvertFormat' with parameters --inputFormat hic --outputFormat5 and default resolutions. Subsequent HiCExplorer commands were carried out at 10-kb, 25-kb and 50-kb resolutions; the 50-kb domains were used for subsequent analysis as they matched visually apparent domain boundaries best. Matrices were normalized using 'hicNormalize' with parameter --normalizesmallest, and corrected using 'hicCorrectMatrixcorrect' with parameters --correctionMethod KR. Hi-C domains were computationally identified using the 'hicFindTADS' from HiCExplorer with parameter --correctForMultipleTestingfdr. The domains derived from the 50-kb resolution analysis were used for subsequent analyses.

RNA-seq experiments. Total RNA was isolated following previously described protocols²⁵.

RNA-seq libraries were generated after selection of polyadenylated RNA using the Nebnext Poly(A) mRNA Magnetic Isolation Module (NEB E7490) and using the NEBNext Ultra II Directional RNA Library Prep (NEB E7765), following manufacturer's instructions.

RNA-seq data analysis. For the analysis of unspliced transcripts, RNA-seq reads were aligned against the original *B. minutum* assembly and annotation using the STAR aligner²⁸ (v.2.5.3a) with the following settings: --limitSjdbInsertNsj 10000000 --outFilterMultimapNmax 50 --outFilterMismatchNmax 999 --outFilterMismatchNoverReadLmax 0.04 --alignIntronMin 10 --alignIntronMax 1000000 --alignMatesGapMax 1000000 --alignSjOverhangMin 8 --alignSJDBoverhang 1 --sjdbScore 1 --twopassMode Basic --twopass1readsN -1. The fraction of intronic reads was estimated from the resulting BAM files.

For the purpose of differential expression analysis, reads were aligned against the transcriptome space using Bowtie²⁹ (v.1.0.1) with the following settings: -e200-a, and quantified using eXpress³⁰ (v.1.5.1). The resulting effective counts were used as input to DESeq2 (ref.³¹) for differential expression analysis. An adjusted *P* value threshold of 0.05 was used to derive lists of significantly differential genes.

External RNA-seq datasets. Approximately 5 × 10⁷ cells were collected by centrifugation at 100g for 5 min at room temperature. Total RNA was extracted and libraries were constructed for RNA-seq using the TruSeq RNA Library Prep Kit V2 (Illumina) according to the manufacturer protocol. All of the raw

sequencing reads are available at the Sequence Read Archive with accession number [SRX7258938](#).

External RNA-seq data analysis. RNA-seq reads were aligned against the corresponding assemblies using the STAR aligner²⁸ (v.2.5.3a) with the following settings: `--limitSjdbInsertNsj 10000000 --outFilterMultimapNmax 50 --outFilterMismatchNmax 999 --outFilterMismatchNoverReadLmax 0.04 --alignIntronMin 10 --alignIntronMax 1000000 --alignMatesGapMax 1000000 --alignSJoverhangMin 8 --alignSJDBoverhang 1 --sjdbScore 1 --twopassModeBasic --twopass1readsN -1`. As available RNA-seq datasets for *B. minutum* are not strand-specific, the strand orientation of the transcriptome was visualized as follows. Aligned reads were first de novo assembled into transcripts and quantified at the transcript level using Stringtie³² (v.1.3.3.b); the orientation of splice junctions serves as a reliable guide for the directionality of these transcripts. Open reading frames were identified for each transcript, and transcripts with open reading frames shorter than 60 amino acids were filtered out of the transcript set. Strand-specific genomic tracks were then generated by assigning to each base pair covered by at least one exon in that set the sum of the TPM (transcript per million transcripts) values of all transcripts it is included in.

External Hi-C datasets. Hi-C datasets for *Trypanosoma brucei* were obtained from GEO accession [GSE118764](#). Hi-C datasets for *S. pombe* were obtained from GEO accession [GSE57316](#). Hi-C datasets for *Caulobacter vibrioides* CB15 were obtained from GEO accession [GSE45966](#).

Sequence analysis. Topoisomerase and other replication-related proteins were identified in annotated Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP) transcriptome assemblies using HMMER3.0 (ref.³³) and the Pfam 27.0 protein domain database³⁴ as previously described⁶.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data associated with this manuscript have been submitted to GEO under accession number [GSE153950](#).

Code availability

Custom code used to process the data is available at <https://github.com/georgimarinov/GeorgiScripts>

References

- Xiang, T., Hambleton, E. A., DeNofrio, J. C., Pringle, J. R. & Grossman, A. R. Isolation of clonal axenic strains of the symbiotic dinoflagellate *Symbiodinium* and their growth and host specificity. *J. Phycol.* **49**, 447–458 (2013).
- Xiang, T., Nelson, W., Rodriguez, J., Tolleter, D. & Grossman, A. R. *Symbiodinium* transcriptome and global responses of cells to immediate changes in light intensity when grown under autotrophic or mixotrophic conditions. *Plant J.* **82**, 67–80 (2015).
- Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
- Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
- Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
- Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
- Roberts, A. & Pachter, L. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nat. Methods* **10**, 71–73 (2013).
- Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
- Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
- Finn, R. D. et al. Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222–D230 (2014).

Acknowledgements

This work was supported by NIH grants (grant nos. P50HG007735, R01HG008140, R01HG00990901, U19AI057266 and UM1HG009442 to W.J.G.; 1UM1HG009436 to W.J.G. and A.K.; 1DP2OD022870-01 and 1U01HG009431 to A.K.), the Rita Allen Foundation (to W.J.G.), the Baxter Foundation Faculty Scholar Grant and the Human Frontiers Science Program grant no. RGY006S (to W.J.G.). W.J.G. is a Chan Zuckerberg Biohub investigator and acknowledges grants no. 2017-174468 and no. 2018-182817 from the Chan Zuckerberg Initiative. Fellowship support was provided by the Stanford School of Medicine Dean's Fellowship (G.K.M.), the Siebel Scholars, the Enhancing Diversity in Graduate Education Program and the Weiland Family Fellowship (A.E.T.). This work is also supported by NSF-IOS EDGE Award no. 1645164 to A.R.G. and Carnegie Venture grant no. 10907 (to T.X. and G.K.M.). We thank A.J. Spakowitz, Z. Shipony, S. Klemm, O. Dudchenko, E. Lieberman Aiden, J.R. Pringle, P. Cleves and members of the Greenleaf, Kundaje, Pringle and Grossman laboratories for helpful discussion and suggestions regarding this work.

Author contributions

G.K.M. performed Hi-C experiments. G.K.M. and A.E.T. analyzed the data. A.E.T. and T.X. designed and carried out transcription inhibition experiments and cell viability experiments. T.X. carried out *B. minutum* culture and heat stress treatment. W.J.G., A.R.G., and A.K. supervised the study. G.K.M., A.E.T. and T.X. interpreted the data and wrote the manuscript with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41588-021-00848-5>.

Correspondence and requests for materials should be addressed to G.K.M. or W.J.G.

Peer review information *Nature Genetics* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Data exclusions

Replication

Randomization

Blinding

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involvement
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals

Wild animals

Field-collected samples

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.