PHi-C: deciphering Hi-C data into polymer dynamics

Soya Shinkai\textsuperscript{1,2,*}, Masaki Nakagawa\textsuperscript{2,3}, Takeshi Sugawara\textsuperscript{2,4}, Yuichi Togashi\textsuperscript{2,5,6}, Hiroshi Ochiai\textsuperscript{5,7}, Ryuichiro Nakato\textsuperscript{8}, Yuichi Taniguchi\textsuperscript{7,9} & Shuichi Onami\textsuperscript{1,*}

\textsuperscript{1}Laboratory for Developmental Dynamics, RIKEN Center for Biosystems Dynamics Research, Kobe, Hyogo, Japan.

\textsuperscript{2}Research Center for the Mathematics on Chromatin Live Dynamics, Hiroshima University, Higashi-hiroshima, Hiroshima, Japan.

\textsuperscript{3}Graduate School of Informatics and Engineering, The University of Electro-Communications, Chofu, Tokyo, Japan.

\textsuperscript{4}Graduate School of Medicine and Faculty of Medicine, The University of Tokyo, Bunkyo-ku, Tokyo, Japan.

\textsuperscript{5}Graduate School of Science, Hiroshima University, Higashi-hiroshima, Hiroshima, Japan.

\textsuperscript{6}Laboratory for Cell Field Structure, RIKEN Center for Biosystems Dynamics Research, Higashi-hiroshima, Hiroshima, Japan.

\textsuperscript{7}PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama, Japan.

\textsuperscript{8}Institute for Quantitative Biosciences, The University of Tokyo, Bunkyo-ku, Tokyo, Japan.

\textsuperscript{9}Laboratory for Cell Systems Control, RIKEN Center for Biosystems Dynamics Research, Suita, Osaka, Japan.

*e-mail: soya.shinkai@riken.jp; sonami@riken.jp
Computational modelling methods for Hi-C data have revealed averaged and static features of the 3D genome in cell nuclei. Here, we describe a 4D simulation method, PHi-C (Polymer dynamics deciphered from Hi-C data), that depicts dynamic 3D genome features through polymer modelling. This method allows for demonstrations of dynamic characteristics of genomic loci and chromosomes, as observed in live-cell imaging experiments, and provides physical insights into Hi-C data.

Genomes consist of one-dimensional DNA sequences and are spatio-temporally organized within the cell nucleus. Contact frequencies in the form of matrix data, measured using genome-wide chromosome conformation capture (Hi-C) technologies, have uncovered three-dimensional (3D) features of average genome organization in a cell population \(^1, 2\). Moreover, live-cell imaging experiments can reveal dynamic chromatin organization in response to biological perturbations within single cells \(^3, 4\). Bridging the gap between these different sets of data derived from population and single cells is a challenge for modelling dynamic genome organization \(^5, 6\).

Several modelling methods have been developed to reconstruct 3D genome structures and predict Hi-C data \(^7, 8\). In addition, there has been development of bioinformatic normalization techniques in Hi-C matrix data processing to reduce experimental biases \(^9–11\). However, the meaning of a contact matrix as quantitative probability data has not been discussed; moreover, a four-dimensional (4D) simulation method to explore dynamic 3D genome organization remains lacking.

Here, we introduce PHi-C, a method that can overcome these challenges by polymer modelling from a mathematical perspective and at low computational cost. PHi-C is a method that
deciphers Hi-C data into polymer dynamics simulations (Fig. 1a, https://github.com/soyashinkai/PHi-C). PHi-C uses Hi-C contact matrix data generated from a hic file through JUICER as input (Supplementary Fig. 1a). PHi-C assumes that a genomic region of interest at an appropriate resolution can be modelled using a polymer network model, in which one monomer corresponds to the genomic bin size of the contact matrix data with attractive and repulsive interaction parameters between all pairs of monomers described as matrix data (Methods, Supplementary Note). Instead of finding optimized 3D conformations, we can utilize the optimization procedure (Supplementary Fig. 1b,c) to obtain optimal interaction parameters of the polymer network model by using an analytical relationship between the parameters and the contact matrix. We can then reconstruct an optimized contact matrix validated by input Hi-C matrix data using Pearson’s correlation r. Finally, we can perform polymer dynamics simulations of the polymer network model equipped with the optimal interaction parameters.

First, we evaluated PHi-C’s theoretical assumption about chromosome contact. Here, we started with a simple polymer model called the bead-spring model, in which the characteristic length b between adjacent beads (or monomers) represents the physical size corresponding to one genomic bin of the contact matrix data. To mathematically define the contact between a pair of monomers, we introduced the contact Gaussian kernel with the contact distance σ (Fig. 1b). The above assumption can be used to derive the theoretical scaling relationship of the contact probability, \( P(s) \sim s^{-d_f/3} \), as a function of genomic distance s in terms of the fractal dimension of polymer organization \( d_f \) (Supplementary Fig. 2a,b). In addition, interestingly, the ratio of the contact distance to the length between adjacent monomers, \( \sigma/b \), makes the shape of the contact probability
rounder at a small genomic distance (Supplementary Fig. 2b). This phenomenon implies that the rounded shape conveys information about the ratio $\sigma/b$. To assess our theoretical framework about contact and how the contact distance varies in Hi-C experiments, we analysed yeast Hi-C data at nucleosome resolution. The fitted value was $\sigma/b = 1.12$, suggesting that contacts mainly occur within a distance corresponding to the size of a nucleosome in this super-resolution Hi-C experiment (Fig. 1c, Supplementary Fig. 2c). Other high-resolution Hi-C data for human GM12878 revealed $\sigma/b = 1.38$, suggesting that cross-linking of Hi-C experiments almost exactly captures chromosome contacts with an appropriate resolution (Supplementary Fig. 2d).

An important step in the optimization procedure is based on analytical matrix transformations between the polymer network model and the contact matrix (Fig. 1d, Methods). The matrix transformations provided us with a low computational cost optimization strategy that can be applied to find optimal interaction parameters of the polymer network model without sampling optimal static 3D polymer conformations (Supplementary Fig. 1b,c). Moreover, we can depict any contact patterns in a moment by using the matrix transformations and perform polymer dynamics simulations by designing interactions in the polymer network model (Fig. 1e–g, Supplementary Videos 1–3). Intra- and inter-domain interactions generate a chequerboard pattern reminiscent of A/B compartments, and the attractive interaction domains form a combined domain (Fig. 1e). Loop interactions show a clear punctate pattern (Fig. 1f). Furthermore, we can depict a topologically associating domain (TAD)-like pattern by only tuning heterogeneous connectivity along the polymer backbone, where less connected regions behave as domain boundaries (Fig. 1g; left). The 3D conformation suggests that the boundary regions are physically elongated with insulat-
ing inter-TAD-like-domain interactions. In addition, the removal of a boundary part causes the adjacent domain fusion (Fig. 1g; right) that reminds us of fusions of TADs \(^8,15\).

To investigate how PHi-C explains 4D features of chromosomes within living cells, we applied this approach to Hi-C data for mouse embryonic stem cells (mESCs)\(^16\). A live-cell imaging experiment showed a marked difference in the movements of Nanog and Oct4 loci in mESCs \(^17\): statistically significant enhancement of Nanog diffusive movement compared with Oct4 diffusive movement was revealed (Supplementary Fig. 3). The optimization step of the PHi-C analysis for chromosomes 6 and 17 provided optimized contact matrices with correlations of more than 97\% between the Hi-C and optimized contact matrices (Fig. 2a). The mean-squared displacement (MSD) curves that were theoretically derived from the optimized data for the Nanog locus on chromosome 6 and the Oct4 locus on chromosome 17 are consistent with the experimental dynamics (Fig. 2b). We also compared the physical sizes of 50.5-Mb genomic regions around the Nanog and Oct4 loci, with the inclusion of several areas that highly interact with each locus, and observed more compact organization of the Nanog region (Fig. 2c). Taking together, these findings indicate that PHi-C analysis can provide new insights into genome organization and dynamics: for example, a region of 50.5 Mb around Nanog adopts a more compact organization than an equivalent region around Oct4, and the Nanog locus on chromosome 6 is more mobile than the Oct4 locus on chromosome 17 (Fig. 2d).

Finally, we used PHi-C to demonstrate the dynamic chromosome condensation process for the highly synchronous entry of DT-40 cells, which revealed a pathway for mitotic chromosome
The optimized contact matrices at five different time points were reconstructed with high correlations (Fig. 2e, Supplementary Fig. 4). Using the optimized interaction parameters in the polymer network model (Supplementary Fig. 5), we conducted 4D simulations starting from a comparatively elongated conformation at 0 min. The polymer conformation dynamically changed into a rod-shaped structure, revealing the condensation state of chromosomes at prometaphase (Fig. 2f, Supplementary Videos 4, 5). We evaluated dynamic changes in polymer conformations in simulations by calculating the characteristic shape lengths of the major and minor axes. As observed by microscopy, rapid and gradual decreases in the minor and major axes within 15 and 60 min indicated thin and thick rod-shaped formations, respectively (Fig. 2g). In addition, the optimized interaction parameters averaged at each genomic separation represent not only strong compaction within 2 Mb during mitosis but also increase of periodicity of long-range attractive interactions from 3 Mb to 12 Mb in prometaphase (Fig. 2h). These physical findings are consistent with a helical organization in rod-shaped chromosomes during prometaphase.

We have shown that PHi-C can decipher Hi-C data into polymer dynamics, based on a mathematical theory of chromosome contacts in the polymer network model. As shown for mESC Hi-C data, PHi-C analysis can bridge the gap between Hi-C data and imaging data with respect to chromatin dynamics in living cells. In addition, PHi-C’s theoretical basis allows for the depiction of any Hi-C pattern by designing the appropriate interaction parameters, which supports a model for TAD formation: physical chromatin stiffness based on certain molecular interactions creates insulation at TAD boundaries. Polymer modelling studies have revealed that chromatin modifications alter the physical properties of chromatin fibres and affect chromosome organiza-
Because PHi-C analysis can extract physical interaction parameters from Hi-C data, it should be elucidated which molecular interactions on chromatin are related to physical parameters. Further comprehensive PHi-C analysis could provide physical insights into molecular interactions on chromosomes.

References

1. Dekker, J., Marti-Renom, M. A. & Mirny, L. A. *Nat. Rev. Genet.* **14**, 390–403 (2013).

2. Bonev, B. & Cavalli, G. *Nat. Rev. Genet.* **17**, 661–678 (2016).

3. Nozaki, T. et al. *Mol. Cell* **67**, 282–293 (2017).

4. Hauer, H. M. et al. *Nat. Struct. Mol. Biol.* **24**, 99–107 (2017).

5. Marti-Renom, M. A. & Mirny, L. A. *PLoS Comput. Biol.* **7**, e1002125 (2011).

6. Fudenberg, G. & Imakaev, M. *Nat. Methods* **14**, 673–678 (2017).

7. Serra, F. et al. *FEBS Lett.* **589**, 2987–2995 (2015).

8. Bianco, S. et al. *Nat. Genet.* **50**, 662–667 (2018).

9. Imakaev, M. et al. *Nat. Methods* **9**, 999–1003 (2012).

10. Rao, S. et al. *Cell* **159**, 1665–1680 (2014).

11. Serra, F. et al. *PLoS Comput. Biol.* **13**, e1005665 (2017).

12. Durand, N. C. et al. *Cell Systems* **3**, 95–98 (2016).
13. Shinkai, S. et al. *PLoS Comput. Biol.* **12**, e1005136 (2016).

14. Ohno, M. et al. *Cell*, **176**, 520–534.e25 (2019).

15. Lupiánez, D. G. et al. *Cell* **161**, 1012–1025 (2015).

16. Bonev, B. et al. *Cell* **171**, 557–572 (2017).

17. Ochiai, H., Sugawara, T. & Yamamoto, T. *Nucleic Acids Res.* **43**, e127 (2015).

18. Gibcus, J. H. et al. *Science* **359**, eaao6135 (2018).

19. Dixon, J. R., Gorkin, D. B. & Ren, B. *Mol. Cell* **62**, 668–680 (2016).

20. Di Pierro, M. et al. *Proc. Natl. Acad. Sci. USA* **114**, 12126–12131 (2017).

Acknowledgments  We thank K. Shiroguchi and K. Kinoshita for helpful comments. We thank M. Marti-Renom and M. Di Stefano for stimulative discussions. We thank I. Hiratani for critical reading of this manuscript. We also thank Nature Research Editing Service for English language editing. This work was supported by JSPS KAKENHI Grant Numbers 16H01408 and 18H04720 (to S.S.).

Author Contributions  S.S., M.N. and T.S. conceived the mathematical concept to define the chromosome contacts. S.S. and S.O. designed the study. S.O. supervised the study. S.S., M.N. and T.S. performed mathematical calculations. S.S. and Y.To. developed the optimization algorithm. S.S. and Y.To. wrote the codes. S.S., R.N. and Y.Ta. analysed the Hi-C data. S.S. and H.O. analysed the movements of genome loci. S.S. wrote the manuscript.

Competing Financial Interests  The authors declare no competing financial interests.
Methods

Overview of PHi-C. PHi-C enables us to decipher Hi-C data into polymer dynamics simulation. PHi-C is based on a theory of the polymer network model and defining contacts between two monomers on the polymer (Supplementary Note). The theoretical framework provides the following matrix transformations (Fig. 1d), where the matrix size is $N \times N$: (i) the normalized interaction matrix \( \tilde{K} = \frac{\sigma^2}{3k_B T} (k_{ij}) = (\tilde{k}_{ij}) \) into the normalized Laplacian matrix \( \tilde{L} = \frac{\sigma^2}{3k_B T} (L_{ij}) \) by \( \tilde{L} = \tilde{D} - \tilde{K} \), where the normalized degree matrix \( \tilde{D} = \text{diag}(\tilde{D}_0, \tilde{D}_1, \cdots, \tilde{D}_{N-1}) \), where \( \tilde{D}_i = \sum_{j=0}^{N-1} \tilde{k}_{ij} \), (ii) \( \tilde{L} \) into the normalized covariant matrix relative to the centre-of-mass \( \tilde{M} = \frac{3k_B T}{\sigma^2} (M_{ij}) = (\tilde{M}_{ij}) \) by \( \tilde{M} = Q \text{diag}(0, \tilde{\lambda}_1^{-1}, \tilde{\lambda}_2^{-1}, \cdots, \tilde{\lambda}_{N-1}^{-1}) Q^T \), where \( 0 = \tilde{\lambda}_0 < \tilde{\lambda}_1 \leq \tilde{\lambda}_2 \leq \cdots \leq \tilde{\lambda}_{N-1} \) are eigenvalues of the matrix \( \tilde{L} \) and \( Q \) is the orthogonal matrix satisfying \( Q^T \tilde{L} Q = \text{diag}(0, \tilde{\lambda}_1, \tilde{\lambda}_2, \cdots, \tilde{\lambda}_{N-1}) \), (iii) \( \tilde{M} \) into the normalized variance matrix \( \tilde{\Sigma}^2 = \left( \Sigma_{ij}^2 / \sigma^2 \right) = \left( \left[ \tilde{M}_{ii} + \tilde{M}_{jj} - 2\tilde{M}_{ij} \right] / 3 \right) \), and (iv) \( \tilde{\Sigma}^2 \) into the contact matrix \( \tilde{C} = (C_{ij}) = \left( 1 + \tilde{\Sigma}^2 \right)^{-3/2} \). The matrix \( \tilde{K} \) describing attractive and repulsive interactions of the polymer network model is optimized so that the difference between an input Hi-C contact matrix \( C_{\text{Hi-C}} \) and the reconstructed contact matrix \( C_{\text{reconstructed}} \), through the above transformations, is minimized. Finally, a 4D simulation of the polymer network model with the optimized matrix \( \tilde{K}_{\text{optimized}} \) is performed. Below, we describe each step in Fig. 1a in detail.

Input data. PHi-C requires \( N \times N \) contact matrix data for a genomic region of interest as an input, which is generated through the JUICER and JUICER TOOLS \(^2\) from public Hi-C data with a normalization option (VC / VC_SQRT / KR). Here, we used the KR normalization \(^2\). In our theoretical framework, the diagonal elements of the contact matrix should satisfy \( C_{ii} = 1 \), so
we additionally normalized the contact matrix such that the shape of the contact probability as a function of genomic distance \( P(s) \) is unaltered, with an interpolation if needed (Supplementary Fig. 1a). Note that the interpolation may result in artefacts for sparse contact matrix data. Finally, we obtained a normalized Hi-C contact matrix \( C_{\text{Hi-C}} \).

**Optimization and validation.** The optimization algorithm is designed to minimize the Frobenius norm \( \| \log_{10} C_{\text{reconstructed}} - \log_{10} C_{\text{Hi-C}} \|_F \) as a cost function, where the contact matrix \( C_{\text{reconstructed}} \) is generated from the normalized interaction matrix \( \tilde{K} \). At every optimization step, an integer pair \((i, j)\) is randomly selected, and the values of \( \tilde{k}_{ij} \) and \( \tilde{k}_{ji} \) are slightly altered. If the alteration decreases the cost function, the matrix \( \tilde{K} \) is updated. A flowchart of the algorithm is presented in Supplementary Fig. 1b.

As our optimization method is based on the random sampling of integer pairs, the amount of calculation in the procedure is proportional to \( O(N^2) \). In our demo codes, the hyper-parameters for the optimization are tuned, and it takes about 13 min to obtain an optimized solution for \( N = 97 \), even on our laptop PC (Intel® CoreTM i7-6600U, dual-core 2.60GHz).

After optimization, an optimized contact matrix \( C_{\text{optimized}} \) is converted from an optimized matrix \( \tilde{K}_{\text{optimized}} \). To assess the compatibility between contact matrices \( C_{\text{optimized}} \) and \( C_{\text{Hi-C}} \) in a logarithmic scale, we used Pearson’s correlation coefficient.

**Polymer dynamics simulation.** We performed 4D simulations of the polymer network model by using the normalized interaction matrix \( \tilde{K} \). First, \( \tilde{K} = (\tilde{k}_{ij}) \) is converted into the normal-
Laplacian matrix $\tilde{L}$. Using the eigendecomposition of the matrix $\tilde{L}$, the normalized eigenvalues $\{\tilde{\lambda}_p\}_{p=0}^{N-1}$ and the orthogonal matrix $Q$ are obtained. For a normalized polymer conformation vector $\tilde{R}_\alpha = \left( \frac{R_{0,\alpha}}{\sigma}, \frac{R_{1,\alpha}}{\sigma}, \ldots, \frac{R_{N-1,\alpha}}{\sigma} \right)^T$, where $R_{i,\alpha}$ stands for the $\alpha$ ($=x, y, z$) coordinate of the $i$-the monomer, the converted vector $\tilde{X}_\alpha = Q^T\tilde{R}_\alpha$ satisfies the variance relationship $\langle \tilde{X}_{p,\alpha}^2 \rangle = 1/(3\tilde{\lambda}_p)$ for $p = 1, 2, \cdots, N - 1$. Therefore, an initial conformation of the converted vector in thermal equilibrium is given: $\tilde{X}_{0,\alpha}|_{t=0} = 0$, so that the centre-of-mass is the origin, and $\tilde{X}_{p,\alpha}|_{t=0}$ is a random variable obeying the normal distribution with mean 0 and variance $1/(3\tilde{\lambda}_p)$ for $p = 1, 2, \cdots, N - 1$. Then, the initial normalized conformation in thermal equilibrium is calculated as $\tilde{R}_\alpha|_{t=0} = Q\tilde{X}_\alpha|_{t=0}$. Finally, to calculate the polymer dynamics, we numerically integrated the stochastic differential equation (SDE) by using Heun’s method \cite{22}: the integral algorithm is defined by first predicting
\begin{equation}
\tilde{R}_\alpha|_{t} = -3 \epsilon \tilde{L} \tilde{R}_\alpha|_{t} + \sqrt{2\epsilon} \xi_\alpha,
\end{equation}
and then correcting
\begin{equation}
\tilde{R}_\alpha|_{t+\epsilon} = -3 \epsilon \tilde{L} \left( \tilde{R}_\alpha|_{t} + \tilde{R}_\alpha|_{t} \right) + \sqrt{2\epsilon} \xi_\alpha,
\end{equation}
where $\epsilon = \frac{k_B T \Delta t}{\gamma \sigma^2}$ and the vector $\xi_\alpha = (\xi_{0,\alpha}, \xi_{1,\alpha}, \cdots, \xi_{N-1,\alpha})^T$ consists of random variables $\{\xi_{i,\alpha}\}_{i=0}^{N-1}$ obeying the normal distribution with mean 0 and variance 1. The parameter $\epsilon$ is non-dimensional and represents a normalized step time, and it determines the accuracy of the SDE integration. $\Delta t$ stands for the step time of the integration in actual time. Here, we set $\epsilon = 0.0001$.

Visualization of polymer conformation. The code for polymer dynamics simulation outputs XYZ and PSF files to visualize the simulated polymer dynamics. Polymer conformations were visualized using VMD \cite{23} by reading these files.
Fitting contact probability. We theoretically derived how the contact probability as a function of genomic distance averaged across the genome, \( P(s) \), behaves in terms of the fractal polymer (Supplementary Note, Supplementary Fig. 2). The function from high-resolution Hi-C data was fitted by

\[
P(s) = \left(1 + \frac{1}{3(\sigma/b)^2} \left(\frac{s}{c}\right)^{2/d_f}\right)^{-3/2}
\]

(3)

for a small genomic region and

\[
P(s) \sim s^{-3/d_f}
\]

(4)

for a large genomic region. Here, the ratio \( \sigma/b \) and the fractal dimension \( d_f \) are the fitted parameters, and \( c \) stands for the genomic size corresponding to the bin size of the Hi-C matrix. We used the nonlinear least-squares Marquardt–Levenberg algorithm on GNUPLOT.

Calculating MSD of genome loci. We re-analysed movements of Nanog and Oct4 loci in mESCs. In each session of live imaging, 3D time-series of a genome locus (Nanog or Oct4) and the nucleus centre-of-mass, \( \{S_{\text{locus}}(t_m)\}_{m=0}^{M-1} \) and \( \{S_{\text{nucleus}}(t_m)\}_{m=0}^{M-1} \), were simultaneously acquired, where the maximum frame number was \( M = 50 \), the time interval was \( \Delta t = 10 \) s, and \( t_m = m \Delta t \) \((m = 0, 1, 2, \cdots, M - 1)\). To eliminate the effect of movement of the nucleus, we dealt with movements of the locus relative to the nucleus centre described by the time-series \( \{S(t_m)\}_{m=0}^{M-1} = \{S_{\text{locus}}(t_m) - S_{\text{nucleus}}(t_m)\}_{m=0}^{M-1} \). Then, the time-averaged mean square displacement (TAMSD) for a time-series \( \{S(t_m)\}_{m=0}^{M-1} \) was calculated as follows:

\[
\text{TAMSD}(t_m) = \frac{1}{M-m} \sum_{i=0}^{M-1-m} [S(t_{m+i}) - S(t_i)]^2.
\]

(5)
Calculating theoretical MSD curve. The optimized matrix $\tilde{K}_{\text{optimized}}$ derives a theoretical MSD curve for the $i$-th monomer in the polymer network model as follows (Supplementary Note):

$$\text{MSD}(t; i)/(\sigma^2) = \frac{6}{N}(\epsilon t/\Delta t) + 2 \sum_{p=1}^{N-1} Q_{ip}^2 \lambda_p^{-1} \left(1 - e^{-3\lambda_p(\epsilon t/\Delta t)}\right).$$  \hspace{1cm} (6)$$

Here, not only is the MSD normalized by $\sigma^2$ in the length scale, but the time-step is also normalized in time, that is, $\text{MSD}/\sigma^2$ and $\epsilon t/\Delta t$ are dimensionless.

Calculating radius of gyration. As we described in polymer dynamics simulation, a normalized polymer conformation $\{\vec{R}_i = (\vec{R}_{i,x}, \vec{R}_{i,y}, \vec{R}_{i,z})\}_{i=0}^{N-1}$ in thermal equilibrium can be sampled on the basis of the optimized matrix $\tilde{K}_{\text{optimized}}$. We calculated the radius of gyration for the 50.5-Mb genomic regions around Nanog and Oct4 loci in mESCs. By using two integers $n_{\text{start}}$ and $n_{\text{end}}$ corresponding to the 50.5-Mb region, the radius of gyration is calculated as

$$\bar{R}_G = \sqrt{\frac{1}{n_{\text{end}} - n_{\text{start}} + 1} \sum_{i=n_{\text{start}}}^{n_{\text{end}}} (\vec{R}_i - \bar{R}_G)^2},$$  \hspace{1cm} (7)$$

where $\bar{R}_G = \frac{1}{n_{\text{end}} - n_{\text{start}} + 1} \sum_{i=n_{\text{start}}}^{n_{\text{end}}} \vec{R}_i$ represents the centre-of-mass of the polymer conformation of the 50.5-Mb genomic region.

Simulating polymer dynamics during chromosome condensation. We applied PHi-C to Hi-C data during mitotic chromosome formation in chicken DT-40 cells. We used the second dataset of chromosome 7 (binned at 100 kb) for the wild type at G2 (0 min), 5, 15, 30 and 60 min. We eliminated the centromere region due to the lack of associated read counts. Through the optimization of PHi-C, we obtained the optimized matrices $\tilde{K}_0$, $\tilde{K}_{500}$, $\tilde{K}_{1500}$, $\tilde{K}_{3000}$ and $\tilde{K}_{6000}$, respectively (Supplementary Fig. 5). To simulate the polymer dynamics during chromosome condensation, we linearly
interpolated the matrices \( \{ \tilde{K}_n \}_{n=0}^{6000} \) at \( \frac{n}{100} \) min as follows: \( \tilde{K}_n = \tilde{K}_0 + (\tilde{K}_{500} - \tilde{K}_0) \times \frac{n}{500} \) for \( 0 \leq n < 500 \), \( \tilde{K}_n = \tilde{K}_{500} + (\tilde{K}_{1500} - \tilde{K}_{500}) \times \frac{n-500}{1000} \) for \( 500 \leq n < 1500 \), \( \tilde{K}_n = \tilde{K}_{1500} + (\tilde{K}_{3000} - \tilde{K}_{1500}) \times \frac{n-1500}{1500} \) for \( 1500 \leq n < 3000 \), and \( \tilde{K}_n = \tilde{K}_{3000} + (\tilde{K}_{6000} - \tilde{K}_{3000}) \times \frac{n-3000}{3000} \) for \( 3000 \leq n \leq 6000 \). By using \( \tilde{K}_0 \), an initial polymer conformation was sampled. Then, the polymer dynamics between \( \frac{n}{100} \) min and \( \frac{n+1}{100} \) min was calculated by 1000 steps of numerical integration with \( \tilde{K}_n \) based on the integral algorithm (equations (1) and (2)).

In the visualization, we fixed the centre-of-mass of polymer conformations to the origin (Fig. 2f, Supplementary Videos 4, 5).

**Calculating shape length of polymer conformation.** To quantify the characteristic shape of a polymer conformation \( \{ \tilde{R}_a \}_{a=x,y,z} \) during chromosome condensation, we evaluated the characteristic shape lengths as an ellipsoidal conformation based on the gyration tensor \( G = \frac{1}{N} (\tilde{R}_a \cdot \tilde{R}_b) \).

We calculated the three eigenvalues \( g_1^2 \leq g_2^2 \leq g_3^2 \) of the tensor \( G \). Then, we adopted \( g_3 \) and \( \sqrt{\frac{g_1^2 + g_2^2}{2}} \) as the characteristic shape lengths of the major and minor axes, respectively (Fig. 2g).

**Code availability.** Python codes of PHi-C are available at [https://github.com/soyashinkai/PHi-C/Codes](https://github.com/soyashinkai/PHi-C/Codes). Scripts to generate data and the figures shown in Fig. 1a can be found at [https://github.com/soyashinkai/PHi-C/Demos](https://github.com/soyashinkai/PHi-C/Demos).

**Data availability.** Published publicly available Hi-C data were used in this study: Ohno et al. (PRJNA427106), Rao et al. (GSE63525), Bonev et al. (GSE96107), and Gibcus et al. (GSE96107).
(GSE102740). Input Hi-C matrix data of PHi-C were generated through the JUICER and JUICER TOOLS 12.

References

21. Knight, P. & Ruiz, D. *IMA J. of Numer. Anal.* **33**, 1029–1047 (2013).

22. Sekimoto, K. *Stochastic Energetics* (Springer-Verlag Berlin Heidelberg, 2010).

23. Humphrey, W., Dalke, A. & Schulten, K. *J. Mol. Graph.* **14**, 33–38 (1996).
**Figure legends**

**Figure 1: Features of PHi-C.** a, Overview of PHi-C procedure. b, (Upper) In the bead-spring model, the probability density of the distance $r_{ij}$ between the $i$-th and $j$-th beads (or monomers) is characterized by only the standard deviation $\Sigma_{ij}$. (Lower) The contact Gaussian kernel function can capture contacts with the contact distance $\sigma$. c, Contact probability (blue) for yeast cells with 160-bp nucleosome-resolution as a function of genomic distance averaged across the genome, and the theoretically fitted curve (orange) at a small genomic distance. d, The polymer network model is characterized by connectivity between all pairs of monomers, expressed by the interaction matrix $k_{ij}$. The matrix $k_{ij}$ is reversibly converted into the contact matrix $C_{ij}$ through matrix transformations. Each matrix has dimensionless values with a normalization factor. e–g, Painting contact patterns for intra- and inter-domain interactions (e), loop interactions (f), and heterogeneous connectivity along the polymer backbone (g). (Upper) Designed interactions in the polymer network model. (Lower) Converted contact matrix, with a snapshot of polymer conformations in the polymer dynamics simulation (Supplementary Videos 1–3). g, TAD-like domains are highlighted by dashed lines. (Right) Removal of a domain-boundary part (yellow) results in domain fusion.
Figure 2: Demonstrations of PHi-C for Hi-C data of mESCs and DT-40 cells. a, PHi-C analysis for chromosomes 6 (left) and 17 (right) of mESCs. (Upper) Contact matrices of the Hi-C experiment (binned at 500 kb), and contact probabilities as a function of genomic distance. (Middle) 4C-like profiles of Nanog (left) and Oct4 (right) loci. High-interaction regions are highlighted (pink). (Lower) Optimized contact matrices by PHi-C, and correlation plots between $\log_{10} C_{\text{Hi-C}}$ and $\log_{10} C_{\text{Optimized}}$. b, Theoretical MSD curves of Nanog and Oct4 loci. c, Probability densities of the gyration radius of $10^5$ conformations for the 50.5-Mb genomic regions around Nanog and Oct4 loci in mESCs. d, Polymer models derived from PHi-C analysis for Nanog and Oct4 loci on chromosomes 6 and 17, respectively. Pink-highlighted regions on the polymer models correspond to the regions in the 4C-like profile of a. e, PHi-C analysis for chromosome 7 of DT-40 cells at G2 (0 min), 5, 15, 30 and 60 min. Contact matrices of the Hi-C experiment with 100-kb-sized bins (upper) and optimized contact matrices with the correlation value (lower). f, Snapshots of polymer conformations in a 4D polymer dynamics simulation. g, Time series of the shape lengths of the major (yellow) and minor (purple) axes for polymer conformations in 100 polymer dynamics simulations starting from the same initial conformation. Thick curves represent the averages. h, Curves of optimized interaction parameters, $\langle k_{ij} \rangle$, averaged at each genomic distance (separation, $|i - j| \times 100$ kb). A triangle indicates a position of a local peak inducing compaction within 2 Mb. Arrows indicate positions of a local peak generating periodicity of attractive interactions around 4, 6 and 10 Mb at 15, 30 and 60 min, respectively.
Figure 1
Figure 2