

Integration of scHi-C and scRNA-seq data defines distinct 3D-regulated and biological-context dependent cell subpopulations

Received: 29 May 2024

Accepted: 6 September 2024

Published online: 27 September 2024

Yufan Zhou¹, Tian Li¹, Lavanya Choppavarapu^{2,3}, Kun Fang^{2,3}, Shili Lin⁴ & Victor X. Jin^{2,3}✉

An integration of 3D chromatin structure and gene expression at single-cell resolution has yet been demonstrated. Here, we develop a computational method, a multiomic data integration (MUDI) algorithm, which integrates scHi-C and scRNA-seq data to precisely define the 3D-regulated and biological-context dependent cell subpopulations or topologically integrated subpopulations (TISPs). We demonstrate its algorithmic utility on the publicly available and newly generated scHi-C and scRNA-seq data. We then test and apply MUDI in a breast cancer cell model system to demonstrate its biological-context dependent utility. We find the newly defined topologically conserved associating domain (CAD) is the characteristic single-cell 3D chromatin structure and better characterizes chromatin domains in single-cell resolution. We further identify 20 TISPs uniquely characterizing 3D-regulated breast cancer cellular states. We reveal two of TISPs are remarkably resemble to high cycling breast cancer persister cells and chromatin modifying enzymes might be functional regulators to drive the alteration of the 3D chromatin structures. Our comprehensive integration of scHi-C and scRNA-seq data in cancer cells at single-cell resolution provides mechanistic insights into 3D-regulated heterogeneity of developing drug-tolerant cancer cells.

Three-dimension (3D) chromatin architecture within a nucleus can be constructed from chromosome conformation capture (3C) related techniques including 3C¹, 4C², 5C³, ChIA-PET⁴, Hi-C⁵, TCC⁶ and in situ Hi-C⁷. These profiling methods have revealed major 3D genomic features, including genomic compartments^{5,8}, topologically associating domains (TADs)⁹ and chromatin loops⁷. Many computational methods have been simultaneously developed to determine these features, including normalizing interacting contact maps^{8,10}, computing A/B compartments^{5,11}, calling TADs^{12,13}, detecting significant interactions^{7,14,15}, enhancing the low sequencing depth data^{16,17}, and visualizing the contact matrices^{18–21}. Further, in order to delineate the heterogeneity of population cells, single-cell Hi-C (scHi-C) protocols have been newly developed to identify 3D chromatin architecture at single-cell resolution^{22–25}. For instance,

the dynamic chromosomal organization of cell cycle²⁶, the organization of zygote chromatin^{27,28}, the nuclear changes of stem cell differentiation²⁹, and single-allele chromatin interactions^{30,31} have been fully examined by scHi-C technique. Meanwhile, new sets of computational methods have been developed for processing scHi-C data to reconstruct single-cell 3D chromatin^{32–34}, to impute the chromosome contact matrices^{35–37}, to identify TAD-like domains³⁸, to classify single cells³⁹, to identify chromatin loops⁴⁰, and to provide toolbox of scHi-C⁴¹. However, none of these methods were designed to algorithmically integrate scHi-C and single-cell (sc)RNA-seq data. Therefore, it is imperative to develop a method for comprehensively integrating single-cell chromatin domains and single-cell gene expression to precisely define 3D-regulated cell subpopulations.

¹Department of Molecular Medicine, University of Texas Health San Antonio, San Antonio, TX, USA. ²Division of Biostatistics, The Medical College of Wisconsin, Milwaukee, WI, USA. ³MCW Cancer Center, The Medical College of Wisconsin, Milwaukee, WI, USA. ⁴Department of Statistics, The Ohio State University, Columbus, OH, USA. ✉e-mail: vjin@mcw.edu

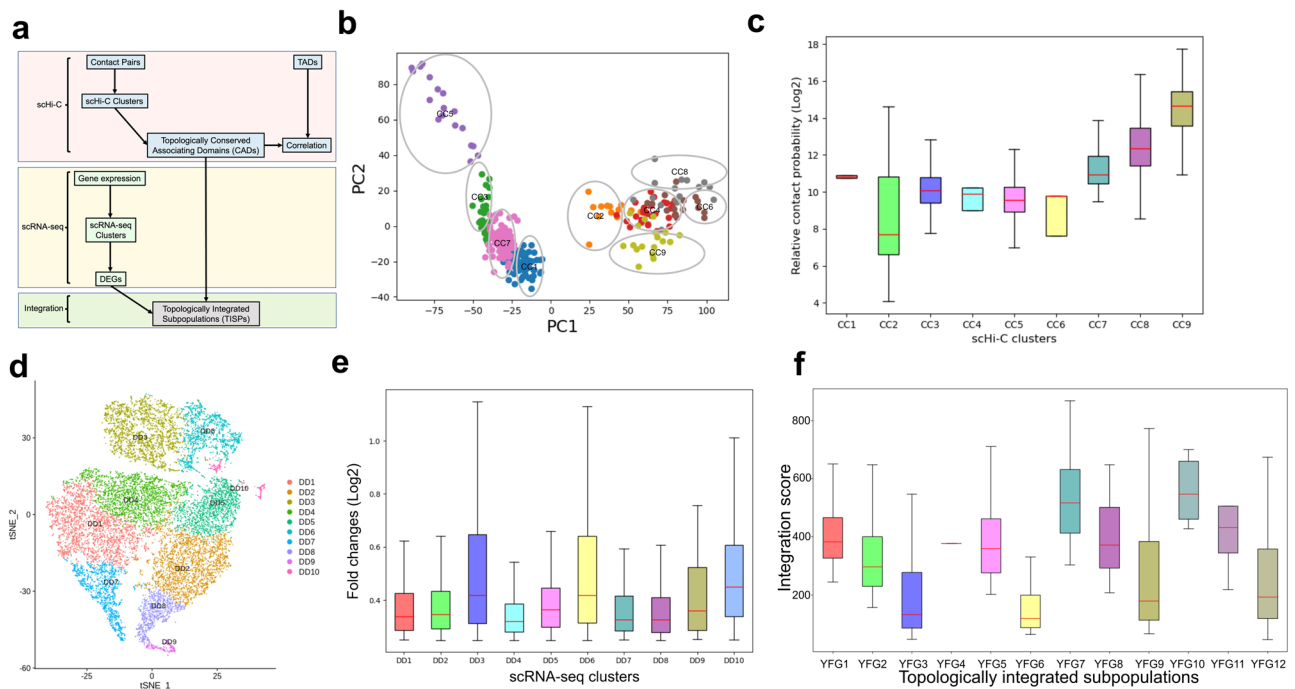


Fig. 1 | Development of a computational method for integrating scHi-C and scRNA-seq data. **a** Flowchart of Multiomic Data Integration (MUDI) algorithm. DEGs differentially expressed genes, TADs topologically associating domains. **b** Nine scHi-C clusters (CC1–CC9) identified from scHi-C data of WTC11/MCF7. **c** Relative contact probability of scHi-C clusters. **d** Ten scRNA-seq clusters

(DD1–DD10) identified from scRNA-seq data of WTC11/MCF7. **e** Fold changes of DEGs of scRNA-seq clusters. **f** Integration scores of 12 topologically integrated subpopulations (TISPs), YFG1–12. Values in box plot of (c–f) from big to small are maxima, the 75th percentile, median, the 25th percentile and minima. Source data are provided as a Source Data file.

Drug-tolerant cancer cells (DTCCs) are a subpopulation of cancer cells that resist the anti-cancer drug treatment and likely cause the patient relapse after therapeutics. DTCCs usually consists of three different groups according to the period of drug treatment⁴². The first group is cancer persister cells survived in the short-term drug shock. The second group is extended persister cells revived and proliferated in the mid-term drug stress. The third group is stable drug-resistant cancer cells survived with clonal selection in the long-term drug treatment. Studies have shown that genetic⁴³ or non-genetic mechanisms^{44,45} were involved in regulating the development of DTCCs. In our recent study, we found that the dynamic changes of 3D chromatin structures might be a non-genetic mechanism driving breast cancer endocrine resistance⁴⁶. However, the patterning and characteristics of 3D chromatin structures in DTCCs at single-cell resolution have not been elucidated.

Here, we develop a computational method, a multiomic data integration (MUDI) algorithm, which integrates scHi-C and scRNA-seq data to precisely define the 3D-regulated and biological-context dependent cell subpopulations or topologically integrated subpopulations (TISPs). We demonstrate its algorithmic utility on the publicly available and newly generated scHi-C and scRNA-seq data. We then apply MUDI in a breast cancer cell model system, including three stages of breast cancer cells, tamoxifen-sensitive breast cancer cells (MCF7), MCF7 cells after being temporally treated with tamoxifen for 1 month (MCF7M1), and MCF7 derived tamoxifen-resistant cells (MCF7TR) after being temporally treated with tamoxifen for 6 months. We identify and characterize distinct 3D-regulated cancer cell subpopulations, and further determine 3D-regulated heterogeneity of developing drug-tolerant cancer cells.

Results

Developing a computational method to integrate scHi-C and scRNA-seq data

To comprehensively integrate scHi-C and scRNA-seq data, we developed a novel computational method, a multiomic data integration

(MUDI) algorithm, to precisely define 3D-regulated cell subpopulations or TISPs (Fig. 1a). We first identified distinct scHi-C clusters from scHi-C data, and scRNA-seq clusters from scRNA-seq data, respectively. We then integrated these two types of clusters by the MUDI algorithm (see Methods: Integration of scHi-C and scRNA-seq data) to precisely define the distinct TISPs (Fig. 1a). Briefly, we first defined topologically conserved associating domains (CADs) representing the conserved 3D chromatin structure of any individual scHi-C cluster. We then integrated CADs with differentially expressed genes (DEGs) of each of scRNA-seq clusters to derive TISPs by implementing an empirical quantitative formula to calculate an integration score of the interaction frequency and the gene expression values. We tested our MUDI on two cell types: pluripotent stem cells WTC11 from 4D Nucleome Project of Bing Ren Lab and breast cancer cells MCF7 generated from this study. From scHi-C data, nine scHi-C clusters (CC1–CC9) were identified with variable relative contact probability (Fig. 1b, c and Supplementary Fig. 1a–c), where CC1/3/5/7 and CC2/4/6/8/9 are majorly composed of WTC11 cells and MCF7 cells, respectively. From scRNA-seq data, ten scRNA-seq clusters (DD1–DD10) were classified with variable fold changes of differentially expressed genes (DEGs) (Fig. 1d, e). DD1/2/4/5/7/8/9 and DD3/6/10 are majorly composed of WTC11 cells and MCF7 cells, respectively. Our MUDI was initially able to identify four TISPs (WMG1–WMG4) with the distinct subpopulation features based on the number (M) of data types (here $M = 2$) and the number of (N) of cell types (here $N = 2$), such that WMG1 is the subpopulation with integration of CC1/3/5/7 and DD1/2/4/5/7/8/9, WMG2 is the subpopulation with the integration of CC1/3/5/7 and DD3/6/10, WMG3 is the subpopulation with integration of CC2/4/6/8/9 and DD1/2/4/5/7/8/9, WMG4 is the subpopulation with integration of CC2/4/6/8/9 and DD3/6/10 (Supplementary Fig. 1d, e). More importantly, the MUDI is further designed to be tailored to a biological-context dependent integration, such that the number of TISPs can be optimized according to a particular biologically meaningful factor on individual studies. Since Yamanaka Factors, MYC, POU5F1, SOX2, KLF4,

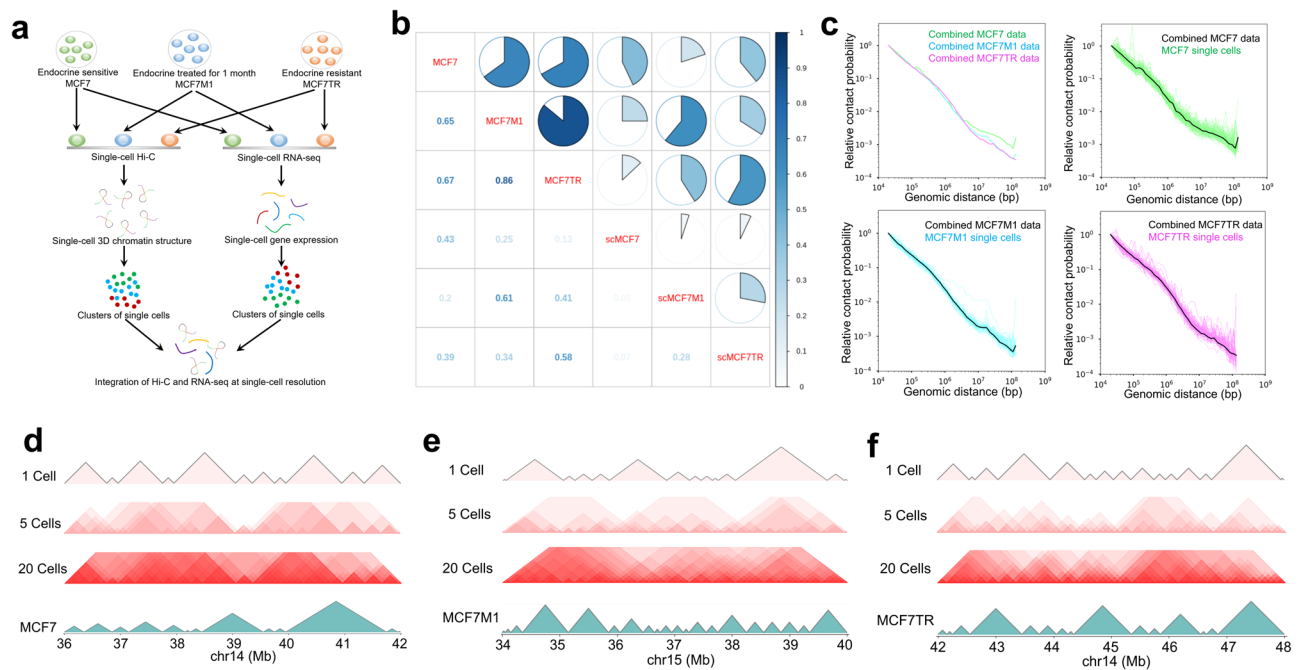


Fig. 2 | Generation of high quality scHi-C and scRNA-seq data in a breast cancer cell model system. a Workflow for the identification of 3D chromatin structures of breast cancer cell lines at single-cell resolution. **b** Pearson correlation coefficients of combined scHi-C data with population data. scMCF7: combined MCF7 scHi-C data. scMCF7M1: combined MCF7M1 scHi-C data. scMCF7TR: combined MCF7TR

scHi-C data. **c** Genomic distance dependent contact probability. The thick lines are combined single cells and the thin lines are individual single cells. Superimposing single-cell TADs with 5 or 20 cells compared to population Hi-C TADs **d** for MCF7, **e** for MCF7M1, and **f** for MCF7TR, respectively. All TADs were generated at the resolution of 100 Kb contact map. Source data are provided as a Source Data file.

were used to characterize the stem cell differentiation, we were able to obtain 12 distinct TISPs (Fig. 1f and Supplementary Fig. 2a), where one of subpopulations YFG1 was enriched with REACTOME developmental biology signaling pathway (Supplementary Fig. 2b, c), suggesting this subpopulation has high stemness and strong chromatin activities.

To further demonstrate the sensitivity and robustness of the MUDI, we have first performed a sub-sampling analysis on WTC11 cells and MCF7 cells (Supplementary Fig. 3a). We found that compared to the whole set of 277 cells, it showed no significant difference of the overlapped CADs in each cluster for the subset of 75% (208) cells and the subset of 50% (138) cells, respectively, but significant difference for the subset of less than 25% (69) cells. Therefore, our MUDI algorithm is sensitive to at least half of cells. We then tested the MUDI on sn-m3c-seq data⁴⁷ and scRNA-seq data⁴⁸ generated from human brain tissues. We first identified scHi-C clusters from human cortex sn-m3c-seq data (Supplementary Fig. 3b) and scRNA-seq clusters from human cortex scRNA-seq data (Supplementary Fig. 3c), respectively. Upon the integration, we identified 24 TISPs for the excitatory neurons (Supplementary Fig. 4a). We not only captured the ground truth TISPs but also identified new transition TISPs (Supplementary Fig. 4b, c). Similarly, we identified 16 TISPs for the inhibitory neurons (Supplementary Fig. 4d–f) including both the ground truth TISPs as well as new transition TISPs. Furthermore, our MUDI was successfully applied in three datasets with significantly different sequencing depths, including (1) sn-m3c-seq data of human prefrontal cortex tissue with an average of 1.2 M contact pairs per cell, (2) scHi-C data of WTC11 cells with an average of 10.5 M contact pairs per cell, and (3) our newly generated scHi-C data of three breast cancer cells with an average of 36.4 M contact pairs per cell (see next four sections). Our MUDI has been able to identify computationally significant and biologically meaningful TISPs, suggesting that our algorithm was much less dependent on the sequencing depth. In summary, we have developed a novel and powerful method, MUDI, to precisely define 3D-regulated and biological-context dependent cell subpopulations.

Generating high quality scHi-C and scRNA-seq data in a breast cancer cell model system

In order to further test and demonstrate the biological-context dependent utility of MUDI, we have generated high quality scHi-C and scRNA-seq data in a breast cancer cell model system, MCF7, MCF7M1 and MCF7TR cells (Fig. 2a), a model system routinely used in the lab⁴⁶. A total of 293 cells (89 MCF7 cells, 91 MCF7M1 cells, 113 MCF7TR cells) were used for scHi-C profiling (Supplementary Fig. 5a) and 22,425 cells (6172 MCF7 cells, 10,156 MCF7M1 cells, 6097 MCF7TR cells) were used for scRNA-seq profiling (Supplementary Fig. 5b). Single-cell chromatin contacts with very high quality were obtained (Supplementary Fig. 5c) upon preprocessing scHi-C data (Supplementary Fig. 5d, e and Supplementary Data 1). The combined scHi-C data showed a significant correlation with population Hi-C data, i.e., correlation coefficient $r = 0.43$ for combined single cells MCF7 to population MCF7, $r = 0.61$ for combined single cells MCF7M1 to population MCF7M1, and $r = 0.58$ for combined single cells MCF7TR to population MCF7TR, respectively. The correlations were weak among combined single cells, i.e., correlation coefficient $r = 0.05$ for combined single cells MCF7 to combined single cells MCF7M1, $r = 0.28$ for combined single cells MCF7M1 to combined single cells MCF7TR, $r = 0.07$ for combined single cells MCF7 to combined single cells MCF7TR, respectively (Fig. 2b). Genomic distance dependent contact probability showed markedly characteristic shapes of combined single cells (Fig. 2c, upper left) and individual single cells (Fig. 2c, upper right, lower left, and lower right panels). We also observed that the single cells had highly variable TADs but with more superimposing of cells, the enriched TADs have more similar features of population TADs (Fig. 2d–f). These results demonstrated a high quality of scHi-C data had been successfully produced in cancer cells. Since single-cell omics-seq data are generally sparse, an optimal resolution is needed for the downstream analysis. Our scHi-C data have a low slope of ratio of read pairs to square of bin numbers until the resolution reaches to 1 Mb (Supplementary Fig. 5f), thus the 1 Mb resolution was used for clustering of scHi-C data.

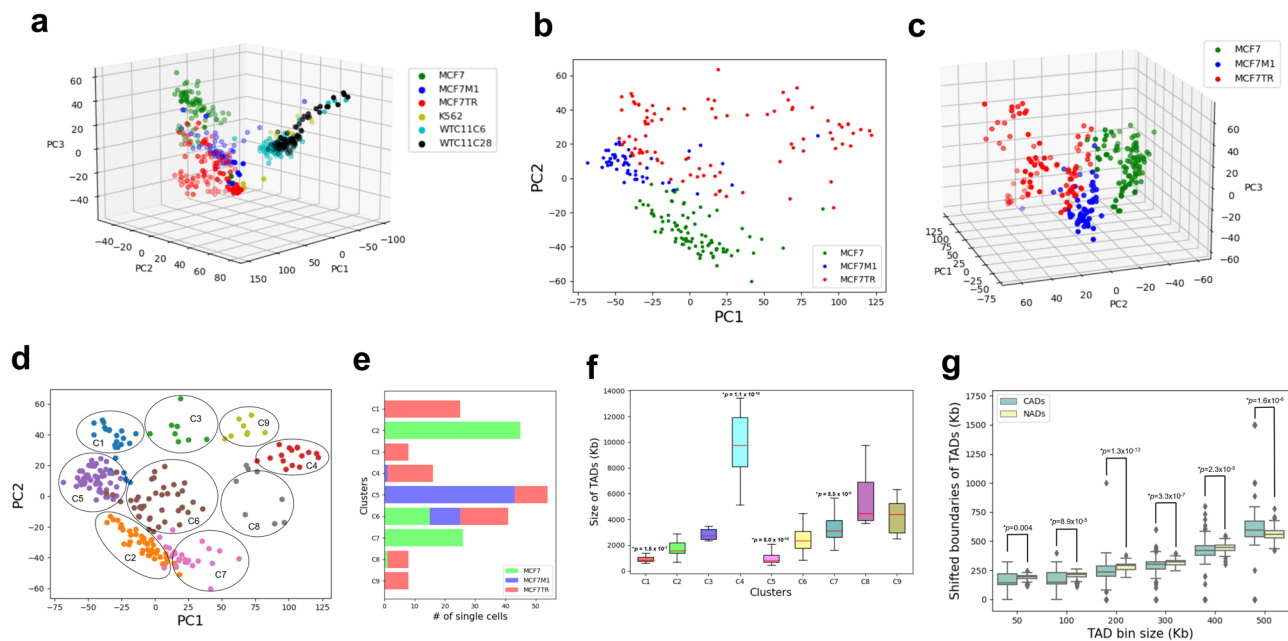


Fig. 3 | Definition of the characteristic single-cell chromatin structure.

a Comparing our scHi-C data with public human scHi-C data. PC1, PC2 and PC3 are first three eigenvectors. **b** 2D view of scHi-C data of breast cancer cells. **c** 3D view of scHi-C data of breast cancer cells. **d** Nine clusters (C1–C9) identified from scHi-C data of breast cancer cells. Each cluster is labeled with oval and assorted colors. **e** Number and the composition of single cells in individual scHi-C clusters. **f** The size

of TADs of clusters. *: Two-sided Wilcoxon rank-sum test. **g** The shifted boundaries of TADs of CADs and NADs when TAD bin size is 50 K, 100 K, 200 K, 300 K, 400 K or 500 K. *: Two-sided Wilcoxon rank-sum test. Values in box plot of (**f**) and (**g**) from big to small are maxima, the 75th percentile, median, the 25th percentile and minima. Source data are provided as a Source Data file.

To exclude the effect of structure variations (SVs), we performed single-cell DNA-seq on three breast cancer cell lines each with a biological replicate: 33 MCF7 cells, 33 MCF7M1 cells and 39 MCF7R cells with a total of 105 cells. We found that (1) there was no clear difference on copy number variations (CNVs) among single cells (Supplementary Fig. 5g), (2) scHi-C contacts in the genomic regions where 10% cells had CNVs had a very low ratio (almost zero) and (3) there was not any significant difference between MCF7 cells and MCF7R cells (Supplementary Fig. 5h). These results illustrated that single-cell level SVs didn't significantly influence the chromatin contacts.

Defining the characteristic single-cell 3D chromatin structure

Before performing scHi-C clustering, we first examined our scHi-C data quality by comparing it with publicly available human scHi-C data. The breast cancer cells from our study were clearly separated from other types of human cells, leukemia cells K562²⁷ and two pluripotent stem cell types, WTC11C6 and WTC11C28 (4D Nucleome Project, Bing Ren Lab) (Fig. 3a and Supplementary Fig. 6a, b). Furthermore, three stages of breast cancer cells, MCF7, MCF7M1 and MCF7R were also distinctly located in different spaces defined by first three eigenvectors (Fig. 3b, c and Supplementary Fig. 6c). This analysis further validated the high quality of our scHi-C data. We then applied scHiCluster³⁶ to identify an optimal nine scHi-C clusters, C1 to C9 (Fig. 3d) since the peak of the Silhouette coefficient is at 9 (Supplementary Fig. 6d). We removed the cells with the contacts lower than 6 in 1 Mb bins to minimize the false positive rate (Supplementary Fig. 7a–d) and thus obtained a good quality of 231 cells (87 MCF7 cells, 54 MCF7M1 cells and 90 MCF7R cells). Of nine clusters, a majority of cells in C2 and C7 were MCF7, a majority of cells in C1, C3, C4, C8, C9 were MCF7R, and the cells in C5 and C6 were miscellaneous of three stages of cells (Fig. 3e). Interestingly, C1 and C5 had the smallest size of TADs and the most numbers of TADs (Fig. 3f and Supplementary Fig. 8a), while MCF7M1 cells had smaller sizes of TADs than MCF7 and MCF7R cells did (Supplementary Fig. 8b, c).

Although Higashi³⁷ was able to increase our scHi-C data to 20 Kb resolution, there was no significant correlation between cell-type specific TADs and cell-type specific gene expression for each of three breast cancer cell types (Supplementary Fig. 9a–d). Therefore, to better characterize chromatin domains in single-cell resolution, we proposed a novel framework for analyzing 3D chromatin domain behavior among single cells and defined a CAD which is the common 1 Mb genomic region shared by all individual cells within any particular scHi-C cluster that has very high chromatin contact probabilities. Indeed, CADs showed lower shifted boundaries of TADs and greater standard deviations than non-conserved associating domains (NADs) (Fig. 3g and Supplementary Fig. 10a). CADs had different characteristics from NADs in each of nine clusters. For example, CADs in C1 showed the highest shifted boundaries in compared to NADs at 100Kb TAD size (Supplementary Figs. 10b–d and 11a–f), and there were the most CADs either in all cells or per cell for C1, C3, C5, and C9 (Supplementary Fig. 12a, b). Our results thus elucidated that the newly defined CAD is the characteristic single-cell 3D chromatin structure useful for functional analysis of scHi-C clusters.

Precisely identifying distinct 3D-regulated cancer cell subpopulations

To precisely identify the 3D-regulated cancer cell subpopulations, we further conducted scRNA-seq data (Supplementary Fig. 13a, b) with the replicates showing a highly identical pattern in MCF7, MCF7M1 and MCF7R cells (Supplementary Fig. 13c). We then identified 13 scRNA-seq clusters, D1–D13 (Fig. 4a), in which a majority of cells in D2, D6, and D11 are MCF7, a majority of cells in D1, D4, D5, D8, D9 and D10 are MCF7M1, a majority of cells in D3, D7, D12, D13 are MCF7R (Fig. 4b). We also identified a gene signature of differentially expressed genes (DEGs) for each of 13 clusters (Fig. 4c and Supplementary Data 2). Interestingly, we found that the cell cycle signaling was among the top enriched pathways from the top 2000 variably expressed genes (Supplementary Figs. 13d and 14a) and the standardized variance of cycling genes is much higher than that of housekeeping genes (Fig. 4d

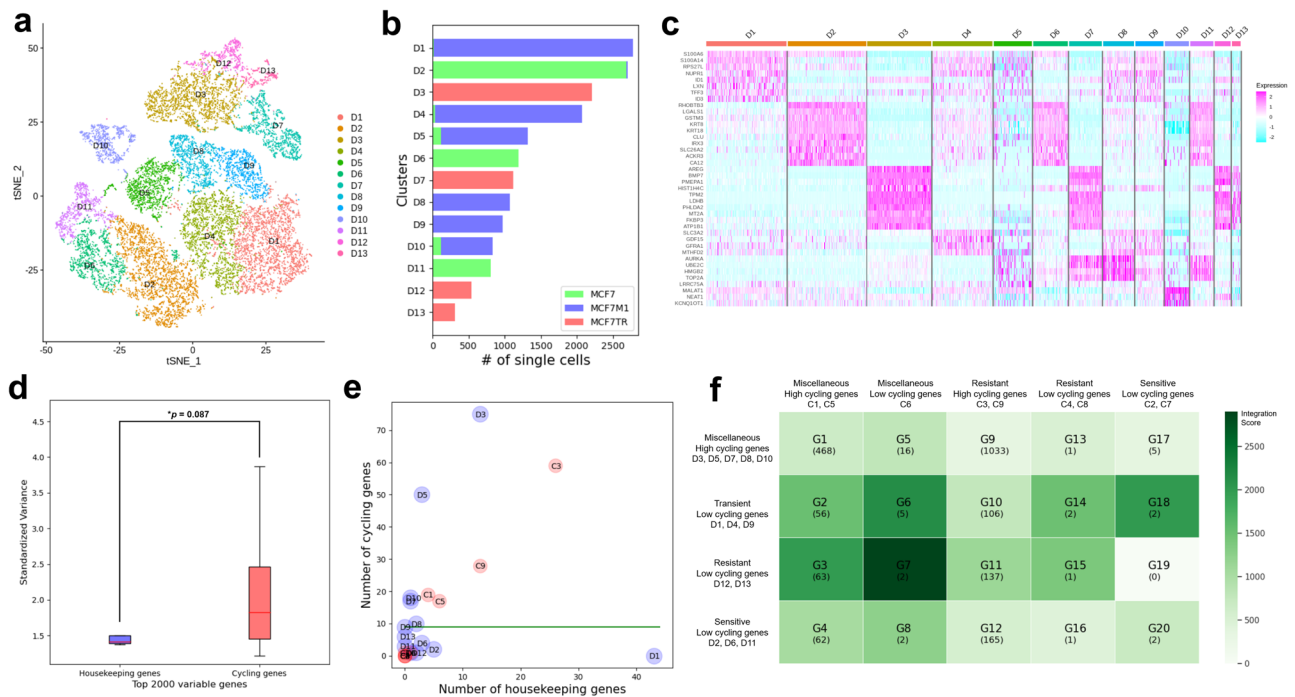


Fig. 4 | Precise identification of 3D-regulated and biological-context dependent cancer cell subpopulations. **a** Thirteen scRNA-seq clusters (D1–D13) identified from scRNA-seq data of breast cancer cells. **b** Number and the composition of single cells in individual scRNA-seq clusters. **c** Gene expression heatmap of DEGs of scRNA-seq clusters. **d** The standardized variance of cycling genes and housekeeping genes in top 2000 variable genes. *: Two-sided Wilcoxon rank-sum test. Values in box plot from big to small are maxima, the 75th percentile, median, the

25th percentile and minima. **e** The distribution of CADs in scHi-C clusters and DEGs in scRNA-seq clusters according to the number of cycling genes and the number of housekeeping genes in each cluster. Green line is the cutoff for high cycling genes and low cycling genes. **f** Twenty topologically integrating subpopulations (TISPs) (G1–G20) dependent on the number of cycling genes and cell compositions of the scHi-C clusters and scRNA-seq clusters.

and Supplementary Data 3). More specifically, there were much more cycling genes within DEGs in D3, D5, D7, D8, D10 as well as within CADs in C1, C3, C5, C9 than other scHi-C or scRNA-seq clusters (Fig. 4e). Remarkably, cycling signaling has been used to characterize cancer persister cells, a rare subpopulation of DTCCs with a reversible property⁴⁵. We thus grouped scHi-C clusters into five categories based on the breast cancer cell stage and the number (high: >9; low: ≤9) of cycling genes within CADs: (1) C1, C5—miscellaneous cells with high cycling genes; (2) C6—miscellaneous cells with low cycling genes; (3) C3, C9—resistant cells with high cycling genes; (4) C4, C8—resistant cells with low cycling genes; (5) C2, C7—sensitive cells with low cycling genes. Miscellaneous cells either with high cycling genes (C1, C5) or with low cycling genes (C6) showed higher contact probabilities than sensitive cells (C2, C7) (Supplementary Fig. 14b, c). On the contrary, resistant cells regardless of with high (C3, C9) or low (C4, C8) cycling genes had lower contact probabilities than sensitive cells (C2, C7) (Supplementary Fig. 14d, e). Although both Categories (1) and (3) have high cycling genes, miscellaneous cells (C1, C5) have more contact probabilities than resistant cells (C3, C9) (Supplementary Fig. 13f). We then computed an integration score within MUDI program to integrate five scHi-C categories with four scRNA-seq categories, and thus precisely defined 20 TISPs, G1–20, each representing a 3D-regulated breast cancer cellular state by an integration score (Fig. 4f).

Characterizing specific topologically integrated subpopulations

We further examined a few of the TISPs related to cycling genes. Despite both G1 and G9 had high cycling genes in both CADs of scHi-C clusters and DEGs of scRNA-seq clusters, G1 had a higher integration score than G9 (Fig. 5a and Supplementary Fig. 15a). In addition, some of G1 and G9 genes were marked with super-enhancers (Supplementary Fig. 15b, c). Interestingly, G1 genes were enriched with a REACTOME chromatin modifying enzyme signaling pathway and these enriched

enzymes had higher integration scores in G1 than those in G9 (Fig. 5b, c). Of 15 enriched genes, ATXN7, ENY2, PRMT6, KDM5B, KMT5A, MBIP, SMARCB1, TADA3 occurred in G1 and G9, BRWD1, CCND1, ELP2, HMG20B, JADE1, KMT2E, MORF4L1 in G9 (Supplementary Fig. 15d). Higher expression of chromatin modifying enzymes in breast cancer patient cohorts showed a lower recurrence-free survival (Fig. 5d and Supplementary Fig. 15e–k). Of these genes, CCND1, ENY2 and KMT5A had epithelial cell-specific *cis*-regulatory elements at their distal regions in luminal breast cancer patient tissue⁴⁹. Together, these results suggest G1 and G9 might resemble to cycling breast cancer persister cells and their 3D chromatin structures might be regulated by chromatin modifying enzymes.

On the other hand, cell subpopulations, G2, G3, G10 and G11, had high cycling genes in CADs of scHi-C clusters but low cycling genes in DEGs of scRNA-seq clusters. REACTOME RNA polymerase II transcription signaling pathway was the top enriched pathway from these four subpopulations (Fig. 5e). Of 21 enriched genes, CEBPB and YEATS4 existed in G2, THOC7 and TXNRD1 in G2 and G10, and COX7A2L, RPS27A, UBE2I, ZNF221 and ZNF223 in G10, while RPRD1A existed in G3, NELFA, PPM1D and SRAFI in G3 and G10, and BNIP3L, BTG2, CNOT6, DYRK2, EAF1, MED1, PABPN1 and TIGAR in G10 (Supplementary Fig. 16a). Higher expression of transcription regulators in breast cancer patient cohorts was correlated with a lower recurrence-free survival (Fig. 5f and Supplementary Figs. 16b–h, 17a–e). Among them, CEBPB, COX7A2L, NELFA, SRSF1, TXNRD1, UBE2I had epithelial cell-specific *cis*-regulatory elements at their distal regions in luminal breast cancer patient tissue⁴⁹. Collectively, these results suggest that these four cell subpopulations might resemble to non-cycling breast cancer persister cells and their 3D chromatin structures might be regulated by transcription regulators.

To further substantiate our findings, we performed an experimental validation for the drug treatment on the two selected genes

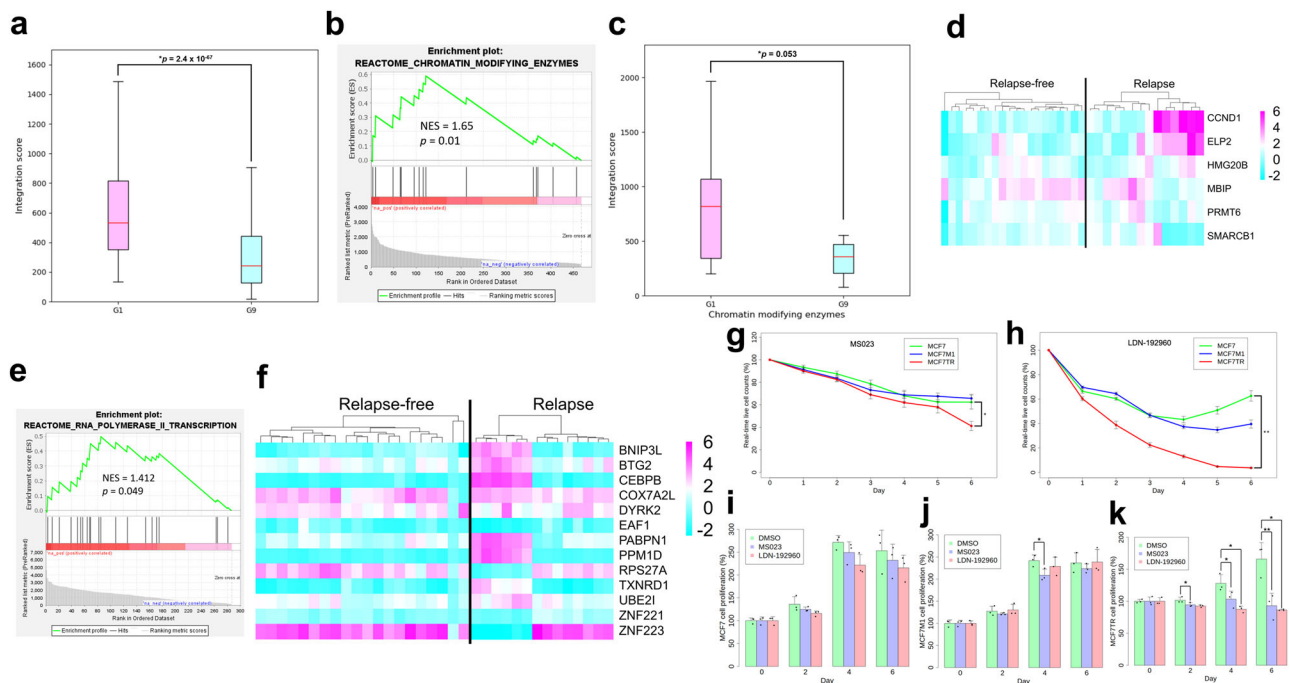


Fig. 5 | Characteristics of TISPs in breast cancer cells. **a** The integration score of G1 and G9. *: Two-sided Wilcoxon rank-sum test. Values in box plot from big to small are maxima, the 75th percentile, median, the 25th percentile and minima. **b** Enrichment of REACTOME chromatin modifying enzymes signaling pathway of G1 genes. NES normalized enrichment score. p value was determined by permutation-based calculation with number of permutations at 1000. **c** Comparison of the integration score between G1 and G9. *: Two-sided Wilcoxon rank-sum test. Values in box plot from big to small are maxima, the 75th percentile, median, the 25th percentile and minima. **d** The expression of chromatin modifying enzymes in relapse-free and relapse breast cancer patient cohort GSE2990. **e** Enrichment of REACTOME RNA polymerase II transcription signaling pathway of the combination of G2, G3, G10 and G11. NES normalized enrichment score. p value was determined by permutation-based calculation with number of permutations at 1000. **f** The expression of transcription regulators in relapse-free and relapse breast

cancer patient cohort GSE2990. **g** Real-time live cell growth curve of PRMT6 inhibitor MS023. Cells treated with DMSO as reference. * $p < 0.05$, two-sided paired Student's t test, p value is 0.0291. **h** Real-time live cell growth curve of DYRK2 inhibitor LDN-192960. Cells treated with DMSO as reference. ** $p < 0.01$, two-sided paired Student's t test, p value is 0.0097. **i** Cell proliferation assay of MS023 and LDN-192960 in MCF7 cells. **j** Cell proliferation assay of MS023 and LDN-192960 in MCF7M1 cells. * $p < 0.05$, two-sided paired Student's t test, p value is 0.0262. **k** Cell proliferation assay of MS023 and LDN-192960 in MCF7TR cells. * $p < 0.05$, ** $p < 0.01$, two-sided paired Student's t test. p value of MS023 vs. DMSO in day 2, 4, 6 are 0.0187, 0.0396, 0.0035 individually. p value of LDN-192960 vs. DMSO in day 2, 4, 6 are 0.0302, 0.0307 individually. Three biological replicates were performed, and data were presented with mean values \pm standard deviation in (g–k). Source data are provided as a Source Data file in (g–k).

identified by our MUDI, PRMT6 and DYRK2. The section of these two genes was purely due to the commercially available inhibitors to them. We treated MS023, an inhibitor to PRMT6, a key regulator in G1 and G9 subpopulations, and LDN-192960, an inhibitor to DYRK2, a key transcriptional regulator in G10. We found both inhibitors showed stronger growth inhibition in MCF7TR cells than that in MCF7 cells (Fig. 5g, h), as well as impeded MCF7TR cells from cell proliferation but not MCF7 (Fig. 5i–k), demonstrating the capability of the inhibitors of these regulators in restoring the drug-sensitivity.

Taken together, we propose a mechanistic model with two distinct 3D-regulated cellular states for the transition of drug-sensitive to tolerant cancer cells: (1) a drug-sensitive cancer cell subpopulation with silenced chromatin modifying enzymes initially shows very lower chromatin interactions (Supplementary Fig. 17a); upon an interim drug treatment, this subpopulation activates the enzymes to trigger higher chromatin interacting activities for the cycling genes, resulting in reversible cancer persister cells (Supplementary Fig. 17b); under a long-term drug treatment, they further reshape the altered 3D chromatin structures render a cycling drug-tolerant cancer cells (Supplementary Fig. 17c); and (2) another drug-sensitive cancer cell subpopulation with silenced transcription regulators initially shows lower chromatin interactions (Supplementary Fig. 17d); upon an interim drug treatment, this subpopulation activates transcription regulators to trigger higher chromatin interacting activities for the non-cycling genes, resulting in reversible cancer persister cells

(Supplementary Fig. 17e); under a long-term drug treatment, they further reshape the altered 3D chromatin structures render a non-cycling drug-tolerant cancer cells (Supplementary Fig. 17f).

Discussion

In this study, we developed a novel computational method, MUDI, to comprehensively integrate scHi-C and scRNA-seq data and to precisely define distinct 3D-regulated and biological-context dependent cell subpopulations or TISPs. In the MUDI, we first defined CADs representing the conserved 3D chromatin structure of any individual scHi-C cluster. We then integrated CADs with DEGs of each of scRNA-seq clusters to derive TISPs by implementing an empirical quantitative formula to calculate an integration score of the interaction frequency and the gene expression values. A high integration score of a TISP indicates it is strongly associated with a set of higher expressed genes with higher chromatin interacting activities. More importantly, the identified TISPs are readily used to interpret biological-context dependent 3D-regulated cell subpopulations according to a particular biologically meaningful factor on individual studies. Furthermore, these 3D-regulated and biological-context dependent cell subpopulations can be used to elucidate a specific biological mechanism.

Remarkably, upon the application of MUDI in three stages of breast cancer cells, we illustrated cycling breast cancer cell subpopulations (miscellaneous or resistant) have distinctive altered 3D

chromatin structures regulated by different regulators. It is reasonable to speculate these cell subpopulations resemble to breast cancer persister cells. Future studies will be focused on functionally examination of breast cancer persister cells. We may apply a Watermelon, a high-complexity expressed barcode lentiviral library⁴⁵ to simultaneously trace each breast cancer Tam-sensitive cell's clonal origin and proliferative state with a short period series of Tam-treatment (0–14 days), then conduct 3D-FISH, 3C/RT-qPCR and Tam-treatment to confirm if cycling persister cells is indeed 3D-regulated and can be re-sensitized.

Interestingly, we found that cell cycle genes highly enriched within CADs were a key factor to stratify the Tam-sensitive cells from 1-month Tam-treated and Tam-resistant cells. Indeed, many studies have demonstrated cell cycle pathway played important roles in breast cancer tamoxifen resistance^{50–54}. For instance, cyclin D1 was essential for the progression of tamoxifen resistance⁵⁰ and inner nuclear membrane protein LEM4 activated cell cycle proteins to render tamoxifen resistance⁵³. Importantly, our data further linked cell cycle signaling with 3D chromatin organization. This finding is pretty novel but not very surprising given that our other recent studies have demonstrated 3D chromatin architecture was associated with endocrine resistance^{46,55–57}.

Furthermore, we identified two key groups of genes, 15 chromatin modifying enzymes and 21 transcriptional regulators, which were not only essential in 3D-regulated breast cancer cellular states, but also predicted a lower recurrence-free survival. Many of these genes have been extensively demonstrated their functional or mechanistic roles in different cancers^{58–72}. For example, Protein arginine methyltransferase PRMT6 was shown to advance the progression in gastric cancer⁶⁰, endometrial cancer⁶¹ and lung cancer⁶². Transcription factor CEBPB stimulated the metabolic reprogramming to increase the occurrence of cancer⁶⁷. Phosphorylation of transcription mediator MED1 increased the drug resistance in prostate cancer⁷⁰.

During the revision, there are three publications^{73–75} in which the authors developed new co-profiling protocols to simultaneously detect single-cell chromatin architecture and gene expression at the same cell. Despite of their experimental advantage, the technical challenges and complex workflows might prevent it to be easily adopted by many labs. In contrast, our MUDI utilizes a novel computational method to integrate scHi-C and scRNA-seq data from either separately on different cells from the same population, or in tandem from each individual cell. More importantly, our method was designed under a clear biological guidance with the following novelties, (1) the first to discover conserved topological domains of each single-cell cluster where these domains represent the chromatin structure signatures of the cluster; (2) the first to define the integration scores of individual genes, and this integration score includes information of both chromatin structure signature and gene signature. Higher integration score means higher gene expression levels and higher chromatin contacts. This definition makes it possible to quantify chromatin events more precisely; (3) the first to integrate non-simultaneous scHi-C and scRNA-seq data and identify integrated subpopulations; (4) the first to investigate single-cell 3D chromatin structure in cancer cells and to demonstrate how to utilize scHi-C and scRNA-seq to understand single-cell cancer 3D chromatin events; (5) the first to confirm that novel therapeutic targets could be discovered by the integration of scHi-C and scRNA-seq data; and (6) the first to demonstrate three omics-seq (scHi-C, scDNA-seq and scRNA-seq) at single-cell resolution on the same biological system. Our comprehensive single-cell sequencing data will benefit the cancer and genome research communities. In addition, our MUDI is able to identify the TISP genes with higher chromatin interactions but non-differentially expressed. As shown in Supplementary Fig. 19a, we identified many CAD genes with non-DEGs in each of nine clusters, including 1946 in C1, 6606 in C3, 1554 in C5 and 3324 in C9, respectively. Upon the MUDI integration, we

obtained 451, 1607, 324 and 802 MUDI genes in C1, C3, C5 and C9, respectively, and further classified them into high or low chromatin interactions for each of four clusters such that H1: C1 high; H2: C1 low; H3: C3 high; H4: C3 low; H5: C5 high; H6: C5 low; H7: C9 high; H8: C9 low (Supplementary Fig. 19b). Since C5 was mainly composed of MCF7M1 and MCF7TR cells, we thus particularly examined this scHi-C cluster and found there were 153 genes in the high group with higher integrated scores, i.e., H5 (Supplementary Fig. 19c). Interestingly, GO/Pathway analyses showed that protein binding, cytosol, protein transport, negative regulation of cell proliferation, endosome organization and metabolism were the top significantly enriched terms, indicating that these genes with higher chromatin interactions but non-differentially expressed between MCF7TR/MCF7M1 and MCF7 are basic protein binding and involved in transportation, not related to many canonical functional signaling pathways. We then examined our MUDI integrated genes with 3083 human genes that could potentially regulate the dynamic nature of chromatin folding screened by HiDRO, named as chromatin regulators (CRs)⁷⁶, and found there were many overlapped genes for each of four clusters (Supplementary Fig. 19e). In particular, of 153 H5 genes, 20 and 5 were among Top 3000 and Top 500 CRs, respectively (Supplementary Fig. 19f). Our results thus strongly demonstrated that our MUDI is able to provide more biological insights than using scRNA-seq or scHi-C only.

Overall, we demonstrated 3D-regulated cancer cell subpopulations were distinctly associated with different functional regulators. Our work might provide mechanistic insights into 3D-regulated heterogeneity of developing drug-tolerant cancer cells, giving a rationale in designing novel therapeutics of treating drug-tolerant cancer.

Methods

MUDI algorithm

After identifying scHi-C clusters by scHiCluster³⁶, and scRNA-seq clusters by Seurat⁷⁷, the CADs of each scHi-C cluster were integrated with DEGs of each scRNA-seq cluster to acquire integration scores. We defined the integration score calculated by individual genes present both in CADs and DEGs as the following:

$$I_g = \frac{F_g E_g}{DR}$$

where I_g is the integration score of a gene. F_g is the relative contact probability (\log_2) of scHi-C data. E_g is expression fold changes (\log_2) of DEGs of scRNA-seq data. D is the ratio of DEGs of scRNA-seq clusters to total DEGs. R is the ratio of scRNA-seq cluster cells to total cells. “ g ” represents genes present in both scHi-C clusters and scRNA-seq clusters. The statistical p value of the difference of integration score was computed by Wilcoxon rank-sum test. We further classified scHi-C clusters into appropriate X scHi-C categories and scRNA-seq clusters into appropriate Y scRNA-seq categories by the biological-contexts, cell types or stages. Finally, product of X and Y is the total number of subpopulations. Each subpopulation has genes with integration score representing the expression level and chromatin interaction probability.

Data processing for scHi-C data

The raw reads of scHi-C were first aligned to human HG19 genome, then filtered by HiC-Pro version 2.11.1⁷⁸ to get the valid pairs. The correlation of combined single cells to population cells was performed at the resolution of 1 Mb with R package HiCRep version 1.11.0⁷⁹. The relative contact probabilities of individual cells were computed by cooltools version 0.4.0⁸⁰ with the compensation of combined single cells. The TADs were called by Insulation Score¹² at 100 Kb resolution if not specifically mentioned. The clustering of single cells was executed by Python package scHiCluster version 0.1.0³⁶. Commonly Associating Domains (CADs) were defined as the common domains in a particular

cluster at the resolution of 1Mb, and non-commonly associating domains (NADs) were those non-common domains in that cluster. The difference of CADs, NADs and TADs was calculated with Wilcoxon rank-sum test. Super-enhancers were called with ChIP-seq data of H3K27ac in tamoxifen-resistant MCF7 cells⁴⁶ by Rank Ordering of Super-Enhancers (ROSE)⁸¹.

Data processing for scRNA-seq data

The raw reads of scRNA-seq were first aligned to human HG19 genome and then feature-barcode matrices were generated with software Cell Ranger developed by 10X Genomics. The gene expression levels were further identified by Seurat version 4.0.3⁷⁷ with the filtering parameters of min.cells at 3 and min.features at 200 on the module of CreateSeuratObject, and percent.mt <30 on the module of subset. The resolution for finding clusters was set to 0.75 on the module of FindClusters. The differentially expressed genes (DEGs) of clusters were defined by the module of FindAllMarkers with the parameters of min.pct at 0.25 and logfc.threshold at 0.25. The difference of standardized variance between housekeeping genes and cycling genes in top 2000 variable genes were computed with Wilcoxon rank-sum test.

Cell lines and reagents

Human breast cancer parental MCF7 cells and tamoxifen-resistant MCF7TR cells were derived from previous study^{46,82–84}. Temporal tamoxifen-resistant MCF7M1 cells were generated from parental MCF7 cells treated with 100 nM tamoxifen metabolite 4-hydroxytamoxifen (4-OHT) (Sigma, Catalog # H7904-5MG) for 1 month (30 days). MCF7, MCF7M1 and MCF7TR cells were cultured in phenol-free RPMI1640 medium (Thermo Fisher Scientific, Catalog # 11835055) supplemented with 10% charcoal stripped fetal bovine serum (FBS) (Sigma, Catalog # F6765-500ML) and 1% Penicillin-Streptomycin (Thermo Fisher Scientific, Catalog # 15140122), while no 4-OHT for MCF7 and MCF7M1 but supplemented with 100 nM 4-OHT for MCF7TR.

In situ Hi-C (population cells) profiling

In situ Hi-C experiments were performed as previously described with minor modifications¹². Two to five million cells were crosslinked with 1% formaldehyde and then lysed with 0.2 Igelal CA630 to get the cell nuclei. The pelleted nuclei were solubilized with 0.5% sodium dodecyl (SDS) and then digested with restriction enzyme HindIII or DpnII. The restriction fragment overhangs were filled with biotin-14-dATP. The crosslinked proximity DNA was ligated with T4 DNA ligase. The crosslinked proteins were degraded by proteinase K. The DNA was pelleted down with ethanol and with sonication. A size of 300–500 bp DNA was selected with AMPure XP beads and then the biotinylated DNA was pulled down with Dynabeads MyOne Streptavidin T1 beads. The ends of sheared DNA were repaired with DNA polymerase I. After the ligation of the adapter, the Hi-C libraries were amplified and purified. The libraries were sequenced on Illumina HiSeq 3000 Sequencer. Each sample was conducted in biological replicates. The sequencing reads were mapped to human HG19 genome with further normalization and filtering by HiC-Pro⁷⁸.

scHi-C profiling

Single-cell Hi-C experiment was performed majorly referring to Flyamer et al.²⁷ with minor revision. Two to four million MCF7 parental cells were fixed for 10 min by resuspending the cell pellet in 5 ml full culture medium supplemented with 1% formaldehyde. The reaction was quenched by addition of 2 M glycine to a final concentration of 125 mM and incubation for 5 min on ice. After washed with phosphate-buffered saline (PBS), cells were resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.5% NP-40, 1% Triton X-100, 1X protease inhibitor cocktail and incubated on ice for at least 45 min. The lysed cell pellet was resuspended in 100 µl of 0.3% SDS in 1X NEBuffer 3 and incubated at 37 °C for 1 h. Then the resuspension was diluted with 330 µl of 1X

NEBuffer 3 and 53 µl of 20% Triton X-100 and incubated at 37 °C for 1 h to quench SDS. The chromatin pellet was further digested with 600U restriction enzyme DpnII (New England BioLabs, Catalog # R0543M) overnight at 37 °C with rotation. On the second day digestion was inactivated by incubation at 65 °C for 20 min. The digested cell nuclei were ligated with 50U T4 DNA ligase for 4 h and then washed with sterile PBS. The sample was stained with two drops of Hoechst 33342 (Thermo Fisher Scientific, Catalog # R37165) for 30 min at 37 °C. Single cells were picked up by FACS sorter and loaded into 96-well PCR plate which each well filled with 5 µl sample buffer from the GenomiPhi V2 DNA amplification kit (previously GE Healthcare currently Cytiva, Catalog # 25660032), covered by 5 µl mineral oil after the sorting, then incubated at 65 °C overnight. The genomic DNA were amplified according to Kumar et al.⁸⁵. The amplified genomic DNA of amounts more than 1 µg were prepared for sequencing with NEBNext Ultra II DNA Library Prep Kit for Illumina (New England BioLabs, Catalog # E7645L).

scRNA-seq profiling

Cells were digested with 0.5% Trypsin-EDTA (Thermo Fisher Scientific, Catalog # 15400054) at the optimal time to avoid cell death and cell aggregation. After centrifugation, the cell pellet was resuspended in PBS (Thermo Fisher Scientific, Catalog # 14190250) at the concentration of 700–1200 cells per µl. If the viability of cells was higher than 90%, cells were then filtered with 40 µm sterile cell strainer (Fisher Scientific, Catalog # 22363547) to get individual cells. The samples of single cells were loaded on 10X Genomics Chromium system to run single-cell RNA-seq protocol according to the technical manual.

scDNA-seq profiling

MCF7, MCF7M1 and MCF7TR cells were collected and sent to BioSkryb Genomics for isolation of single cell and scDNA-seq libraries preparation with the approach of Primary Template-directed Amplification (PTA)⁸⁶. ResolveDNA Whole Genome Amplification Kit (Catalog # 100136, BioSkryb Genomics) was used for amplification of genomic DNA. ResolveDNA Library Preparation Kit (Catalog # 100080, BioSkryb Genomics) was used for the library construction. Libraries of scDNA-seq were sequenced on Illumina NovaSeq 6000 system. Sequencing raw reads were mapped to human HG19 genome and copy number variation was identified by SCCNV version 1.0.2⁸⁷.

Enrichment of signaling pathway

For scRNA-seq data, genes were pre-ranked by standardized variance then enriched by Gene Set Enrichment Analysis (GSEA) version 4.1.0⁸⁸. Kyoto Encyclopedia of Genes and Genomes (KEGG) were used as gene sets database. For integrated scRNA-seq and scHi-C data, genes were pre-ranked by integration score then enriched by GSEA. REACTOME Pathway Database were used as gene sets database.

Recurrence-free survival analysis

Two cohorts of breast cancer patients were used for survival analysis. Cohort GSE2990 was from Sotiriou et al.⁸⁹ and cohort GSE6532 was from Loi et al.⁹⁰. The patients were filtered by having tamoxifen treatment but no radio therapy or no other chemotherapy. The survival analysis was performed by R package Survival version 3.2-11. The patients were stratified by gene expression levels at the top quartile (25%) as high expression vs. the rest (75%) as low expression. The log-rank test was used for calculation of *p* value.

Incucyte real-time live cell imaging

For a real-time live cell imaging of MCF7, MCF7M1 and MCF7TR, cells were seeded in 96-well plates at a density of 1×10^3 cells per well. The cell media was replaced after 24 h and cells were treated with MS023 (10 µM) and LDN (5 µM) and the proliferation is monitored by the analysis of occupied area (% confluence) of cell images over time. As

cells proliferate, the confluence increases. Confluence was an exceptional replacement for proliferation, until cells were densely packed or when large changes in morphology occurred. The graphs from the phase of cell confluence area were recorded from day 0 to day 6 according to the IncuCyte S3 Live-Cell Analysis System (Sartorius) manufacturer's instructions. IncuCyte S3 software version 2020B was used for the analysis.

Cell proliferation assay

Cell viability was measured by CCK-8 (CCK-8, Dojindo, USA) assay following the manufacturer's instructions. In brief, MCF7, MCF7M1 and MCF7TR cells were harvested and plated at a density of 1×10^3 cells per well in 96-well plates (Corning Inc) and cultured in an incubator 5% CO₂ incubator at 37 °C. After 24 h, the culture media was replaced, and the cells are treated with MS023 (10 µM) and LDN (5 µM). At the end of each time point, 10 µL of CCK-8 solution was added to each 96-well plate and the mixture was incubated for 1 h in the incubator at 37 °C. The OD value of each well was measured by BioTek™ ELx800™ Absorbance Microplate Reader at 450 nm. The assay was repeated three times.

Simulation of 3D chromatin structure

Compartments of single cells were called by CscoreTool version 1.1¹¹ at 50Kb resolution with the compensation of combined single cells. The compartments were then annotated as A1 (Cscore ≥ 0 and ≤ 0.2), A2 (Cscore > 0.2), B1 (Cscore < 0 and > -0.2) and B2 (Cscore ≤ -0.2) followed by simulation with chromatin dynamics software OpenMiChroM version 1.0.0⁹¹. The simulated structures were visualized by UCSF Chimera version 1.15⁹².

Reporting summary

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

Data availability

Raw and processed scHi-C data for MCF7, MCF7M1 and MCF7TR cells are deposited in GEO under accession number [GSE194308](#). Raw and processed scRNA-seq data for MCF7, MCF7M1 and MCF7TR cells are deposited in GEO under accession number [GSE195610](#), and raw and processed in situ Hi-C data for MCF7, MCF7M1 and MCF7TR cells are deposited in GEO under accession number [GSE195810](#). Raw and processed scDNA-seq data for MCF7, MCF7M1 and MCF7TR cells are deposited in GEO under accession number [GSE239435](#). WTC11C6 and WTC11C28 scHi-C datasets are publicly available datasets from 4D Nucleome Project Data Portal under accession numbers [4DNESJQ4RXYS](#) and [4DNESF829JOW](#). WTC11 scRNA-seq datasets are publicly available datasets from the ArrayExpress database under accession number [E-MTAB-6268](#)⁹³. Source data are provided with this paper.

Code availability

The source code of MUDI is available at <https://github.com/yufanzhouonline/MUDI>⁹⁴. Source data and source code for figures are provided with this paper at https://github.com/yufanzhouonline/Nat_Communications_2024⁹⁵.

References

- Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
- Simonis, M. et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* **38**, 1348–1354 (2006).
- Dostie, J. et al. Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* **16**, 1299–1309 (2006).
- Fullwood, M. J. et al. An oestrogen-receptor- α -bound human chromatin interactome. *Nature* **462**, 58–64 (2009).
- Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
- Kalhor, R., Tjong, H., Jayatilaka, N., Alber, F. & Chen, L. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol.* **30**, 90–98 (2011).
- 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).
- Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).
- Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
- Servant, N. et al. HiTC: exploration of high-throughput 'C' experiments. *Bioinformatics* **28**, 2843–2844 (2012).
- Zheng, X. & Zheng, Y. CscoreTool: fast Hi-C compartment analysis at high resolution. *Bioinformatics* **34**, 1568–1570 (2018).
- Crane, E. et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240–244 (2015).
- Shin, H. et al. TopDom: an efficient and deterministic method for identifying topological domains in genomes. *Nucleic Acids Res.* **44**, e70 (2016).
- Ay, F., Bailey, T. L. & Noble, W. S. Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Res.* **24**, 999–1011 (2014).
- Zhou, Y. et al. Modeling and analysis of Hi-C data by HiSIF identifies characteristic promoter-distal loops. *Genome Med.* **12**, 69 (2020).
- Zhang, Y. et al. Enhancing Hi-C data resolution with deep convolutional neural network HiCPlus. *Nat. Commun.* **9**, 750 (2018).
- Liu, Q., Lv, H. & Jiang, R. hicGAN infers super resolution Hi-C data with generative adversarial networks. *Bioinformatics* **35**, i99–i107 (2019).
- Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
- Li, D., Hsu, S., Purushotham, D., Sears, R. L. & Wang, T. WashU Epigenome Browser update 2019. *Nucleic Acids Res.* **47**, W158–W165 (2019).
- Wang, Y. et al. The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol.* **19**, 151 (2018).
- Akdemir, K. C. & Chin, L. HiCPlotter integrates genomic data with interaction matrices. *Genome Biol.* **16**, 198 (2015).
- Nagano, T. et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).
- Ramani, V. et al. Massively multiplex single-cell Hi-C. *Nat. Methods* **14**, 263–266 (2017).
- Stevens, T. J. et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59–64 (2017).
- Li, G. et al. Joint profiling of DNA methylation and chromatin architecture in single cells. *Nat. Methods* **16**, 991–993 (2019).
- Nagano, T. et al. Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61–67 (2017).
- Flyamer, I. M. et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* **544**, 110–114 (2017).
- Gassler, J. et al. A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture. *EMBO J.* **36**, 3600–3618 (2017).
- Bonora, G. et al. Single-cell landscape of nuclear configuration and gene expression during stem cell differentiation and X inactivation. *Genome Biol.* **22**, 279 (2021).
- Allahyar, A. et al. Enhancer hubs and loop collisions identified from single-allele topologies. *Nat. Genet.* **50**, 1151–1160 (2018).

31. Oudelaar, A. M. et al. Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains. *Nat. Genet.* **50**, 1744–1751 (2018).
32. Rosenthal, M. et al. Bayesian estimation of three-dimensional chromosomal structure from single-cell Hi-C data. *J. Comput. Biol.* **26**, 1191–1202 (2019).
33. Zhu, H. & Wang, Z. SCL: a lattice-based approach to infer 3D chromosome structures from single-cell Hi-C data. *Bioinformatics* **35**, 3981–3988 (2019).
34. Meng, L., Wang, C., Shi, Y. & Luo, Q. Si-C is a method for inferring super-resolution intact genome structure from single-cell Hi-C data. *Nat. Commun.* **12**, 4369 (2021).
35. Liu, J., Lin, D., Yardimci, G. G. & Noble, W. S. Unsupervised embedding of single-cell Hi-C data. *Bioinformatics* **34**, i96–i104 (2018).
36. Zhou, J. et al. Robust single-cell Hi-C clustering by convolution- and random-walk-based imputation. *Proc. Natl. Acad. Sci. USA* **116**, 14011–14018 (2019).
37. Zhang, R., Zhou, T. & Ma, J. Multiscale and integrative single-cell Hi-C analysis with Higashi. *Nat. Biotechnol.* **40**, 254–261 (2021).
38. Li, X., Zeng, G., Li, A. & Zhang, Z. DeTOKI identifies and characterizes the dynamics of chromatin TAD-like domains in a single cell. *Genome Biol.* **22**, 217 (2021).
39. Wu, H. et al. scHiCStackL: a stacking ensemble learning-based method for single-cell Hi-C classification using cell embedding. *Brief. Bioinform.* **23**, bbab396 (2021).
40. Yu, M. et al. SnapHiC: a computational pipeline to identify chromatin loops from single-cell Hi-C data. *Nat. Methods* **18**, 1056–1059 (2021).
41. Li, X., Feng, F., Pu, H., Leung, W. Y. & Liu, J. scHiCTools: a computational toolbox for analyzing single-cell Hi-C data. *PLoS Comput. Biol.* **17**, e1008978 (2021).
42. Niveditha, D. et al. Drug tolerant cells: an emerging target with unique transcriptomic features. *Cancer Inf.* **18**, 1176935119881633 (2019).
43. Xue, Y. et al. An approach to suppress the evolution of resistance in BRAF(V600E)-mutant cancer. *Nat. Med.* **23**, 929–937 (2017).
44. Shaffer, S. M. et al. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* **546**, 431–435 (2017).
45. Oren, Y. et al. Cycling cancer persister cells arise from lineages with distinct programs. *Nature* **596**, 576–582 (2021).
46. Zhou, Y. et al. Temporal dynamic reorganization of 3D chromatin architecture in hormone-induced breast cancer and endocrine resistance. *Nat. Commun.* **10**, 1522 (2019).
47. Lee, D. S. et al. Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. *Nat. Methods* **16**, 999–1006 (2019).
48. Darmanis, S. et al. A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci. USA* **112**, 7285–7290 (2015).
49. Kumegawa, K. et al. GRHL2 motif is associated with intratumor heterogeneity of cis-regulatory elements in luminal breast cancer. *NPJ Breast Cancer* **8**, 70 (2022).
50. Kilker, R. L. & Planas-Silva, M. D. Cyclin D1 is necessary for tamoxifen-induced cell cycle progression in human breast cancer cells. *Cancer Res.* **66**, 11478–11484 (2006).
51. Ferraiuolo, R. M., Tubman, J., Sinha, I., Hamm, C. & Porter, L. A. The cyclin-like protein, SPY1, regulates the ER α and ERK1/2 pathways promoting tamoxifen resistance. *Oncotarget* **8**, 23337–23352 (2017).
52. Løkkegaard, S. et al. MCM3 upregulation confers endocrine resistance in breast cancer and is a predictive marker of diminished tamoxifen benefit. *NPJ Breast Cancer* **7**, 2 (2021).
53. Gao, A. et al. LEM4 confers tamoxifen resistance to breast cancer cells by activating cyclin D-CDK4/6-Rb and ER α pathway. *Nat. Commun.* **9**, 4180 (2018).
54. Yu, D., Shi, L., Bu, Y. & Li, W. Cell division cycle associated 8 is a key regulator of tamoxifen resistance in breast cancer. *J. Breast Cancer* **22**, 237–247 (2019).
55. Bi, M. et al. Enhancer reprogramming driven by high-order assemblies of transcription factors promotes phenotypic plasticity and breast cancer endocrine resistance. *Nat. Cell Biol.* **22**, 701–715 (2020).
56. Li, J. et al. Hi-C profiling of cancer spheroids identifies 3D-growth-specific chromatin interactions in breast cancer endocrine resistance. *Clin. Epigenetics* **13**, 175 (2021).
57. Yang, Y. et al. The 3D genomic landscape of differential response to EGFR/HER2 inhibition in endocrine-resistant breast cancer cells. *Biochim. Biophys. Acta Gene Regul. Mech. Nov.* **1863**, 194631 (2020).
58. Montero-Conde, C. et al. Transposon mutagenesis identifies chromatin modifiers cooperating with Ras in thyroid tumorigenesis and detects ATXN7 as a cancer gene. *Proc. Natl. Acad. Sci. USA* **114**, E4951–E4960 (2017).
59. Atanassov, B. S. et al. ATXN7L3 and ENY2 coordinate activity of multiple H2B deubiquitinases important for cellular proliferation and tumor growth. *Mol. Cell* **62**, 558–571 (2016).
60. Okuno, K. et al. Asymmetric dimethylation at histone H3 arginine 2 by PRMT6 in gastric cancer progression. *Carcinogenesis* **40**, 15–26 (2019).
61. Jiang, N. et al. PRMT6 promotes endometrial cancer via AKT/mTOR signaling and indicates poor prognosis. *Int. J. Biochem. Cell Biol.* **120**, 105681 (2020).
62. Avasarala, S. et al. PRMT6 promotes lung tumor progression via the alternate activation of tumor-associated macrophages. *Mol. Cancer Res.* **18**, 166–178 (2020).
63. Gallo, M. et al. MLL5 orchestrates a cancer self-renewal state by repressing the histone variant H3.3 and globally reorganizing chromatin. *Cancer Cell* **28**, 715–729 (2015).
64. Takawa, M. et al. Histone lysine methyltransferase SETD8 promotes carcinogenesis by deregulating PCNA expression. *Cancer Res.* **72**, 3217–3227 (2012).
65. Chen, Y. Y. et al. BNIP3L-dependent mitophagy promotes HBx-induced cancer stemness of hepatocellular carcinoma cells via glycolysis metabolism reprogramming. *Cancers* **12**, 655 (2020).
66. Wagener, N. et al. Endogenous BTG2 expression stimulates migration of bladder cancer cells and correlates with poor clinical prognosis for bladder cancer patients. *Br. J. Cancer* **108**, 973–982 (2013).
67. Ackermann, T. et al. C/EBP β -LIP induces cancer-type metabolic reprogramming by regulating the let-7/LIN28B circuit in mice. *Commun. Biol.* **2**, 208 (2019).
68. Ikeda, K. et al. Mitochondrial supercomplex assembly promotes breast and endometrial tumorigenesis by metabolic alterations and enhanced hypoxia tolerance. *Nat. Commun.* **10**, 4108 (2019).
69. Banerjee, S. et al. Inhibition of dual-specificity tyrosine phosphorylation-regulated kinase 2 perturbs 26S proteasome-addicted neoplastic progression. *Proc. Natl. Acad. Sci. USA* **116**, 24881–24891 (2019).
70. Rasool, R. U. et al. CDK7 inhibition suppresses castration-resistant prostate cancer through MED1 inactivation. *Cancer Discov.* **9**, 1538–1555 (2019).
71. Xiang, Y. et al. Comprehensive characterization of alternative polyadenylation in human cancer. *J. Natl. Cancer Inst.* **110**, 379–389 (2018).
72. Canevari, R. A. et al. Identification of novel biomarkers associated with poor patient outcomes in invasive breast carcinoma. *Tumour Biol.* **37**, 13855–13870 (2016).
73. Liu, Z. et al. Linking genome structures to functions by simultaneous single-cell Hi-C and RNA-seq. *Science* **380**, 1070–1076 (2023).

74. Qu, J. et al. Simultaneous profiling of chromatin architecture and transcription in single cells. *Nat. Struct. Mol. Biol.* **30**, 1393–1402 (2023).
75. Zhou, T. et al. GAGE-seq concurrently profiles multiscale 3D genome organization and gene expression in single cells. *Nat. Genet.* <https://doi.org/10.1038/s41588-024-01745-3> (2024).
76. Park, D. S. et al. High-throughput Oligopaint screen identifies druggable 3D genome regulators. *Nature* **620**, 209–217 (2023).
77. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
78. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
79. Yang, T. et al. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res.* **27**, 1939–1949 (2017).
80. Nora, E. P. et al. Molecular basis of CTCF binding polarity in genome folding. *Nat. Commun.* **11**, 5612 (2020).
81. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
82. Massarweh, S. et al. Tamoxifen resistance in breast tumors is driven by growth factor receptor signaling with repression of classic estrogen receptor genomic function. *Cancer Res.* **68**, 826–833 (2008).
83. Feng, Q. et al. An epigenomic approach to therapy for tamoxifen-resistant breast cancer. *Cell Res.* **24**, 809–819 (2014).
84. Morrison, G. et al. Therapeutic potential of the dual EGFR/HER2 inhibitor AZD8931 in circumventing endocrine resistance. *Breast Cancer Res. Treat.* **144**, 263–272 (2014).
85. Kumar, G., Garnova, E., Reagin, M. & Vidali, A. Improved multiple displacement amplification with phi29 DNA polymerase for genotyping of single human cells. *Biotechniques* **44**, 879–890 (2008).
86. Gonzalez-Pena, V. et al. Accurate genomic variant detection in single cells with primary template-directed amplification. *Proc. Natl. Acad. Sci. USA* **118**, e2024176118 (2021).
87. Dong, X., Zhang, L., Hao, X., Wang, T. & Vijg, J. SCCNV: a software tool for identifying copy number variation from single-cell whole-genome sequencing. *Front. Genet.* **11**, 505441 (2020).
88. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
89. Sotiriou, C. et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J. Natl. Cancer Inst.* **98**, 262–272 (2006).
90. Loi, S. et al. Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen. *BMC Genomics* **9**, 239 (2008).
91. Oliveira Junior, A. B., Contessoto, V. G., Mello, M. F. & Onuchic, J. N. A scalable computational approach for simulating complexes of multiple chromosomes. *J. Mol. Biol.* **433**, 166700 (2021).
92. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
93. Friedman, C. E. et al. Single-cell transcriptomic analysis of cardiac differentiation from human PSCs reveals HOPX-dependent cardiomyocyte maturation. *Cell Stem Cell* **23**, 586–598.e8 (2018).
94. Zhou, Y. & Jin, V. X. Integration of scHi-C and scRNA-seq data defines distinct 3D-regulated and biological-context dependent cell subpopulations. MUDI. <https://doi.org/10.5281/zenodo.13329087> (2024).
95. Zhou, Y. & Jin, V. X. Integration of scHi-C and scRNA-seq data defines distinct 3D-regulated and biological-context dependent cell subpopulations. *Nat. Commun.* <https://doi.org/10.5281/zenodo.13329097> (2024).

Acknowledgements

We thank the UTHSA Next Generation Sequencing Facilities for services rendered for production of the Hi-C, scHi-C and scRNA-seq data. We would also like to thank Dr. Bing Ren at University of California at San Diego for sharing us with their human scHi-C data. We are grateful to Dr. Myles Brown of Center for Functional Cancer Epigenetics at Dana-Farber Cancer Institute for reading the manuscript and providing suggestive comments. This project was partially supported by grants from NIH R01GM114142 and U54CA217297.

Author contributions

V.X.J. conceived the project. Y.Z. conducted the experiments and performed the data analysis. T.L. and L.C. assisted in conducting the experiments. K.F. assisted the data analysis. V.X.J. and Y.Z. wrote the manuscript, with all authors including S.L. contributing to writing and providing the feedback.

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/s41467-024-52440-0>.

Correspondence and requests for materials should be addressed to Victor X. Jin.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024