Loss of multilevel 3D genome organization during breast cancer progression

Roberto Rossini, Saleh Oshaghi, Maxim Nekrasov, et al.

Genome Res. 2026 36: 20-37 originally published online October 9, 2025
Access the most recent version at doi:10.1101/gr.280791.125

Loss of multilevel 3D genome organization during breast cancer progression

Roberto Rossini,[1] Saleh Oshaghi,[1] Maxim Nekrasov,[2] Aurélie Bellanger,[3] Renae Domaschenz,[2] Yasmin Dijkwel,[2] Mohamed Abdelhalim,[3] Rahul Agrawal,[3,4] Marit Ledsaak,[3,4] Philippe Collas,[3,5] Ragnhild Eskeland,[3,4] David Tremethick,[2] and Jonas Paulsen[1]

[1]*Department of Biosciences, Faculty of Mathematics and Natural Sciences, University of Oslo, 0316 Oslo, Norway;* [2]*Department of Genome Sciences, The John Curtin School of Medical Research, The Australian National University, Canberra, Australian Capital Territory 2601, Australia;* [3]*Department of Molecular Medicine, Institute of Basic Medical Sciences, Faculty of Medicine, University of Oslo, 0317 Oslo, Norway;* [4]*Center for Cancer Cell Reprogramming, Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, 0379 Oslo, Norway;* [5]*Department of Immunology and Transfusion Medicine, Oslo University Hospital, 0424 Oslo, Norway*

Breast cancer entails intricate alterations in genome organization and expression. However, how three-dimensional (3D) chromatin structure changes in the progression from a normal to a breast cancer malignant state remains unknown. To address this, we have conducted an analysis combining Hi-C data with lamina-associated domains (LADs), epigenomic marks, and gene expression in an in vitro model of breast cancer progression. Our results reveal that although the fundamental properties of topologically associating domains (TADs) are overall maintained, significant changes occur in the organization of compartments and subcompartments. These changes are closely correlated with alterations in the expression of oncogenic genes. We also observe a restructuring of TAD–TAD interactions, coinciding with a loss of spatial compartmentalization and radial positioning of the 3D genome. Notably, we identify a previously unrecognized interchromosomal insertion event, wherein a locus on Chromosome 8 housing the *MYC* oncogene is inserted into a highly active subcompartment on Chromosome 10. This insertion is accompanied by the formation of de novo enhancer contacts and activation of *MYC*, illustrating how structural genomic variants can alter the 3D genome during oncogenesis. In summary, our findings provide evidence for the loss of genome organization at multiple scales during breast cancer progression, revealing novel relationships between genome 3D structure and oncogenic processes.

[Supplemental material is available for this article.]

The mammalian genome is folded into large-scale dynamic three-dimensional (3D) chromatin conformations that provide a framework for regulated gene expression. In the interphase nucleus, chromosomes exist as territories (Cremer and Cremer 2010; Rowley and Corces 2018) within which gene-rich and active A compartments are segregated from more gene-poor and inactive B compartments (Misteli 2004; Lieberman-Aiden et al. 2009; Crosetto and Bienko 2020), both ranging in size from single promoters up to 2 Mbp (Sood and Misteli 2022; Harris et al. 2023). Compartments consist of multiple subcompartments that differ in their chromosomal contact frequencies, enrichment in histone modifications, and chromatin-binding proteins (Spracklin et al. 2023). At a resolution ranging from ~10–800 kbp, topologically associated domains (TADs) emerge as local, highly interacting domains (Dixon et al. 2012; Nora et al. 2012; Rao et al. 2014). TADs are further organized into sub-TADs and support the formation of internal smaller chromatin loops, which includes enhancer–promoter interactions (Berlivet et al. 2013). TADs and sub-TADs seemingly form independently of compartments (Rowley and Corces 2018). It was initially proposed that TADs provide a platform for the coordinated regulation of subsets of genes by enhancers located within the boundaries of the TAD. However, disrupting TAD structure minimally affects gene expression (Nora et al. 2017; Rao et al. 2017; Schwarzer et al. 2017), suggesting that other levels of genome organization have a more important regulatory role.

An intriguing possibility is that long-range TAD–TAD interactions, forming TAD cliques (Paulsen et al. 2019; Li et al. 2022; Pei et al. 2022; Arnould et al. 2023; Zhao et al. 2023), sculpt the 3D genome into distinct functional domains. TAD cliques are enriched in B compartments and can gain or lose TADs during differentiation, generally in correlation with gene repression or activation, respectively (Paulsen et al. 2019). Additionally, the nuclear envelope imposes positional constraints on chromatin by anchoring heterochromatin through lamina-associated domains (LADs) ~0.1–10 Mbp in size (Rønningen et al. 2015; Gonzalez-Sandoval and Gasser 2016; van Steensel and Belmont 2017). Accordingly, genes located at the nuclear periphery are repressed or expressed at lower levels than genes localized toward the nuclear center (Misteli 2004; Crosetto and Bienko 2020). Integrating Hi-C and LAD data allows

Corresponding authors: david.tremethick@anu.edu.au, jonas.paulsen@ibv.uio.no
Article published online before print. Article, supplemental material, and publication date are at https://www.genome.org/cgi/doi/10.1101/gr.280791.125.

36:20–37 Published by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/26; www.genome.org

the generation of 3D structural genome models that accurately recapitulate the radial nuclear positions of TADs (Li et al. 2017; Paulsen et al. 2017, 2018; Boninsegna et al. 2022) and provide an enhanced understanding of the link between chromatin architecture and gene regulation in disease contexts (Paulsen et al. 2017).

Indeed, transcriptional dysregulation has been attributed to alterations in 3D genome organization in diseases, including cancer (Feng and Pauklin 2020; Osman et al. 2022), yet analysis of multiple cancers shows that TAD boundary deletions only rarely change gene expression, emphasizing their rare involvement in gene expression dysregulation in cancer (Akdemir et al. 2023). Even with shifts in the A and B compartments between colon tumors and normal intestinal cells, TADs remain relatively unchanged (Johnstone et al. 2020), suggesting compartment switching significantly impacts gene expression control (Ibrahim and Mundlos 2020; Sood and Misteli 2022).

Breast cancer is the most common cancer in women. These cancers have been classified into five subtypes based on the expression of receptors for human epidermal growth factor, estrogen and progestogen, and clinical features (Sørlie et al. 2001). Among these subtypes, triple-negative breast cancers, which do not express any of these receptors, are the most aggressive (Bobbitt et al. 2023). Few studies have examined the chromatin architectural features of breast cancer cells or tissues (Zhou et al. 2019; Kim et al. 2022; Wang et al. 2022; Dozmorov et al. 2023). Nonetheless, one study reports that triple-negative breast cancer cells display the most severe disruption of the 3D genome, including weakening of TAD borders, loss of 3D chromatin interactions leading to fewer chromatin loops, and dynamic compartmental changes (Kim et al. 2022). However, how 3D chromatin organization is associated with the progression from a normal to a breast cancer malignant state remains unexplored. Here, we rely on an in vitro isogenic breast cancer progression model (Dawson et al. 1996; Santner et al. 2001) to investigate 3D genome organization changes in this process.

Results

TAD properties are conserved across breast cancer stages

We performed duplicate Hi-C experiments at three stages of an in vitro human breast cancer progression model, corresponding to a nonmalignant state (using MCF10A cells; "10A" from here on), a premalignant state (MCF10AT1 cells; "T1"), and a malignant tumorigenic state (MCF10Ca1a cells; "C1"). We obtained an average of more than 560 million pairwise interactions after filtering (Supplemental Table S1), supporting analyses in the range of 1–10 kbp bin resolution (Ay and Noble 2015; Lajoie et al. 2015). Filtering statistics and the relative fraction of intrachromosomal (*cis*) and interchromosomal (*trans*) contacts indicate high-quality libraries (Lajoie et al. 2015), and samples show high reproducibility between replicates (Supplemental Figs. S1, S2; Yang et al. 2017). We identified and masked out regions with translocations and used ICE-normalized interactions, which adjust for bias resulting from copy-number alterations (see Methods) (Kim et al. 2022).

TAD comparisons across samples and replicates (Fig. 1A) show similar TAD numbers (about 4000 TADs) and genomic size (~0.6 Mbp) (Supplemental Fig. S3), TAD border insulation strength ($R^2 = 0.77–1.00$) (Supplemental Fig. S4) and intra-TAD contact frequencies ($R^2 = 0.91–1.00$) (Supplemental Fig. S5), and genomic positions of both TADs (Jaccard index [JI] = 0.76–0.80; called at 50 kb resolution), and sub-TADs (JI = 0.71–0.75; 10 kb resolution) (Supplemental Fig. S6). Inspecting TADs surrounding known breast cancer–related genes (Lee and Muller 2010) reveals minor differences that were not consistently reproducible across replicates (Supplemental Figs. S7–S13). We conclude that TADs are overall similar across the three cell types and are therefore likely not the main oncogenic drivers in this progression model of breast cancer.

Compartments and subcompartments engage in major switching events during breast cancer progression

To map differences in long-range genome contacts in our progression model, we searched for any A/B-compartment switching, a property reported to be implicated in breast cancer (Barutcu et al. 2015). Most of the genome (2124 Mbp; ~78%) remains in the same compartment across stages (Fig. 1B). However, for the fraction that switches between stages, we note 385 Mbp (~14%) in 10A switching to either B (162 Mbp; ~6%) or A (222 Mbp; ~8%) compartments in T1 (Fig. 1B; Supplemental Table S2). Most of the switched compartments remain in their switched state in C1, but a fraction switches back to B (76 Mbp; ~3%) or A (61 Mbp; ~2%). The extent of switching observed in our model is comparable to that found in other cellular contexts involving changes in cell fate (Dixon et al. 2015; Liu et al. 2021; Vilarrasa-Blasi et al. 2021), indicating that compartment switching is a prominent feature of breast cancer progression.

We next used dcHiC to identify subcompartments (Fig. 1C; Chakraborty et al. 2022). We predict four A subcompartments (A0–A3) and four B subcompartments (B0–B3), with A1, A2, and B1 being the most prominent and with A0 and B0 the least prominent for all cell stages (Supplemental Fig. S14; Supplemental Table S3). Overlaying a range of active and repressive histone modifications from public ChIP-seq data sets onto the subcompartments reveals a gradual enrichment of the ratio of heterochromatin to euchromatin features, going from the B3 to A3 subcompartments (Supplemental Fig. S15). From this analysis, B2 and B3 show features of heterochromatin, with enrichment of repressive histone marks (H3K27me3) and lamin B1. Compared with the other subcompartments, A0 and B0 are the least enriched in active and repressive histone marks, respectively, raising the interesting possibility that they could be more easily remodeled into other stronger euchromatic or heterochromatic subcompartments. In contrast, A3 and B3 display the strongest active and inactive states, respectively, which would arguably be more resistant to configuration changes (Supplemental Fig. S15).

Indeed, we find that, first, weaker, intermediate subcompartments (A0, A1, B0, B1) undergo more dynamic switches than do more prominent and stronger subcompartments (A2, A3, B2, B3) (Fig. 1D; Supplemental Figs. S16–S18). However, second, switching between B0 and A0 (B0↔A0) is minimal; whereas, third, B0↔A1 is the most prominent intercompartment switch (Fig. 1D; Supplemental Figs. S16–S18), indicating specific switching between weakly heterochromatic and euchromatic states. A further characterization of paths of subcompartment switching across the three stages shows that, fourth, a similar fraction of 24%–26% of subcompartments ends up in either a more closed (and thus more B-like) subcompartment in C1 or a more open (A-like) subcompartment. Further, reverting back to the initial 10A-subcompartment landscape in C1 rarely happens (Supplemental Fig. S19). Fifth, switches mainly involve a transition between consecutive subcompartment strengths, occurring

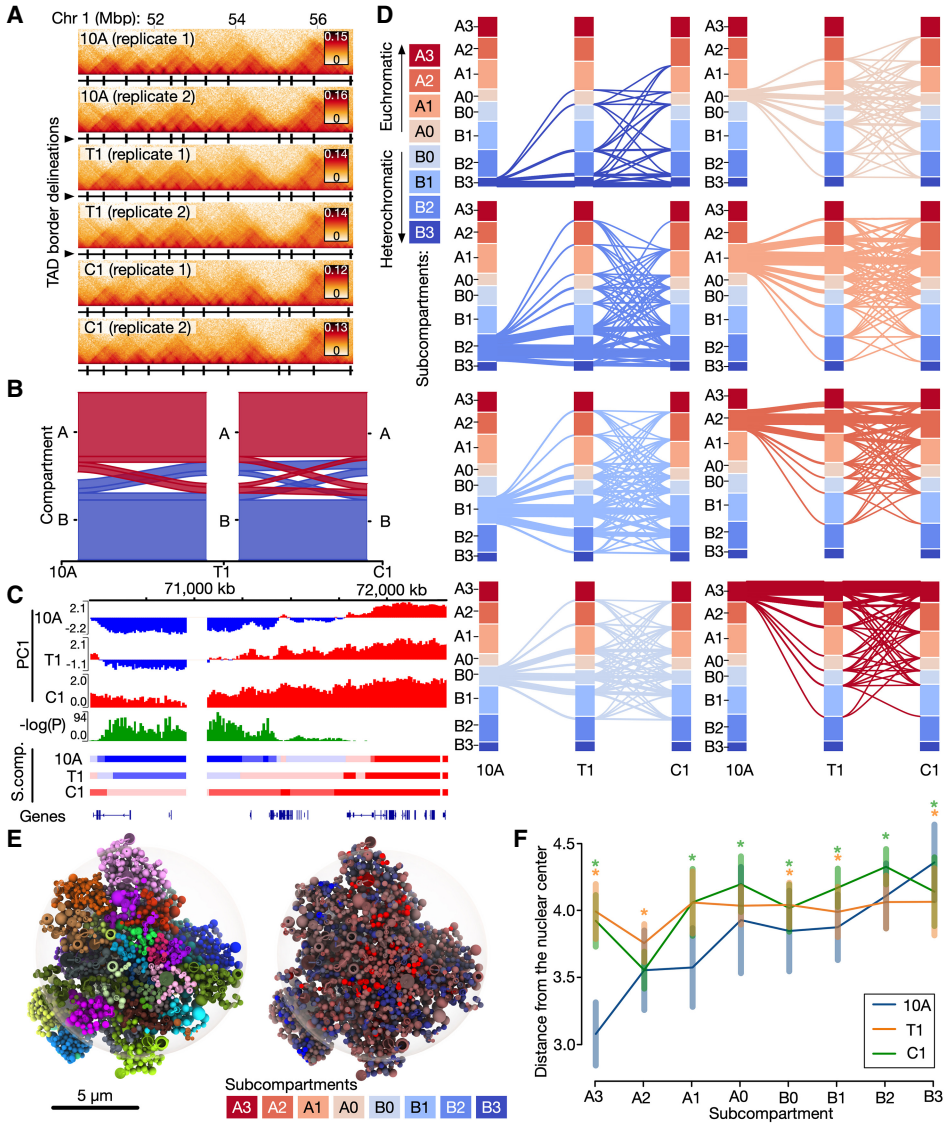

Figure 1. Alterations in genome compartment properties during breast cancer progression stages. (*A*) Example Hi-C data for a region on Chromosome 1. TAD delineations are shown as black vertical lines *below* each track. (*B*) Alluvial plot showing A/B-compartment conservation and switching during progression from 10A (*left*), via T1 (*middle*) and to C1 (*right*). (*C*) Example region on Chromosome 11 showing a statistically significant switch in the first principal component (PC1) in the three stages. Corresponding subcompartments in the three stages (10A, T1, C1) are shown *below*. (*D*) Switching between four A-type subcompartments (A0–A3), and four B-type (B0–B3) called from the Hi-C data (shown in distinct colors). Subcompartment switches across 10A, T1, and C1 shown as separate alluvial plots starting at each of the eight different subcompartments in 10A. (*E*, *left*) Tomographic view of exemplary Chrom3D model from 10A cells with chromosomes colored individually. (*Right*) The same model with regions colored by their subcompartment associations. (*F*) Median distance (with standard deviation) from the nuclear center for each subcompartment in each condition. Asterisks indicate significant differences in T1 and C1 compared with 10A (see Supplemental Table S4).

with similar frequencies toward both more open (e.g., A2→A3) and less open (e.g., B2→B3) subcompartments (Supplemental Fig. S19). Comparing the switching levels between the two replicates shows a high degree of consistency, with coefficients of determination (R^2) ranging from 0.87 to 0.94 ($P<10^{-20}$) (see Supplemental Fig. S20).

To determine implications of subcompartment switching on 3D genome organization, we generated 3D genome models (Paulsen et al. 2017) integrating Hi-C data with nuclear lamina–chromatin interactions (LADs) mapped by ChIP-seq of lamin B1 in each cell type (see Methods). The resulting models (Fig. 1E; Supplemental Figs. S21–S23) display expected genomic features in which more gene-poor chromosomes are located more frequently toward the nuclear periphery (Supplemental Figs. S24–S26). Our 3D modeling of all three cell types reveals major differences between 10A cells versus T1 and C1 cells. As expected, there is a gradual increase in nuclear peripheral localization in subcompartments going from A3 toward B3 in 10A cells. This, however, is less apparent in T1 and C1 cells (Fig. 1F). Moreover, A3 subcompartments significantly shift away from the nuclear center in T1 and C1 cells, compared with 10A (Fig. 1F; Supplemental Table S4). B3 subcompartments appear to be slightly shifted toward the nuclear interior, which concurs with a reduction in LAD coverage in T1 and C1 cells (Supplemental Fig. S27). Taken together,

our data indicate that subcompartment switching characterizing the 10A–T1–C1 cell transition is accompanied by a partial loss of radial disposition of chromatin.

Subcompartment switching reflects changes in gene expression

We next investigated whether the 3D genome structural changes observed were linked to transcriptional changes. Differential gene expression analysis from RNA-seq data between the cell types reveals 3180 differentially expressed (DE) genes between 10A and T1 cells and 8362 DE genes between 10A and C1 cells ($|LFC| > 0.5$; $P < 0.01$) (Supplemental Table S5). Analysis of disease-linked gene set enrichment among DE genes in T1 and C1 reveals "breast carcinoma" as the most enriched disease term, followed by other cancer types. "Organ system benign neoplasm" is the second most associated term for T1, whereas this term is absent for C1, reflecting the expected transcriptomic differences between the transformed (T1) versus the malignant (C1) cell line (Supplemental Fig. S28). Overlaying the RNA-seq data onto Hi-C-based subcompartments in all cell types reveals anticipated associations in which gene expression levels gradually increase from B3 to A3 (Fig. 2A; Supplemental Fig. S29), reinforcing the validity of the subcompartment classification and the RNA-seq data in our model.

Next, we analyzed up- and downregulated DEGs in conjunction with the corresponding subcompartment switches. As expected and as previously reported (Nagai et al. 2019; Liu et al. 2021), this analysis shows that at specific chromosomal regions (Fig. 2B), or at genome-wide levels (Fig. 2C–F), there are correlated patterns of subcompartment switching and differential gene regulation. Inspecting the presence of downregulated genes in T1 relative to 10A reveals that B1(10A)→B2(T1) is the most frequent switch for these genes (Fig. 2C, left). Switching from B0/A0 (10A)→B1(T1) or from A1(10A)→B1/B0/A0(T1) is enriched relative to what is seen for non-DE genes (Fig. 2C, middle) or upregulated genes (Fig. 2C, right).

To quantify whether up- and downregulated genes were associated with expected subcompartment switches toward more open (or A-like), or closed (or B-like) subcompartments, respectively, we analyzed enrichment of DE genes in each subcompartment switch relative to a switch in opposite direction. This analysis revealed a significant enrichment of downregulated genes in switching of B1→B2, B0→B1, A0→B,1 and A1→B0 between 10A and T1 (all $P < 0.05$, binomial test) (Supplemental Table S6). Upregulated genes were significantly enriched in B1→A1, B1→B0, and A1→A3 (all $P < 0.05$) (Fig. 2C; Supplemental Table S6). Comparing 10A and C1, downregulated genes were enriched in A0→B2, A0→B1, B1→B2, and A1→A0 (all $P < 0.05$), whereas upregulated genes were enriched in B3→B1, B1→A1, B1→A2, A0→A2, A1→A2, and A1→A3 (all $P < 0.05$) (Fig. 2E; Supplemental Table S7). This suggests that specific subcompartment switching toward more open or closed states is associated with up- or downregulated DEGs, respectively.

Further, we computed the log-ratio of subdiagonal sums in the upper versus the lower triangular matrices from Figure 2C. When this log-ratio is negative, DE genes are more present in 10A, and when it is positive, DE genes are more present in T1 (Fig. 2D) or C1 (Fig. 2F). We confirm switching trends toward more open subcompartments for upregulated genes, and the opposite for downregulated genes ($P = 4.04 \times 10^{-9}$; Fisher's exact test) (Fig. 2D, right; Supplemental Table S8), whereas no trend is seen for non-DE genes (Fig 2D, middle). When contrasting 10A and C1, similar trends are seen ($P = 6.66 \times 10^{-13}$) (Fig. 2E,F; Supplemental Table S9). Thus, subcompartment switching aligns with changes in gene expression.

Reorganization of TAD–TAD interactions coincides with subcompartment alterations

Beyond their subcompartment organization, we have previously shown that TADs can arrange in densely interconnected long-range contact configurations termed TAD cliques (Paulsen et al. 2019). To identify these cliques, each TAD is represented as a node in a graph, with edges indicating statistically significant Hi-C interactions between pairs of TADs. We define a "TAD clique" as a maximal, fully connected subset of nodes that cannot be expanded without losing full connectivity, which is identified using the Bron–Kerbosch algorithm (Bron and Kerbosch 1973). The size of a TAD clique is the number of nodes in a maximal clique (see Fig. 3A, inset; Supplemental Fig. S30). We computed TAD cliques in 10A, T1, and C1 cells (see Fig. 3A,B) and found that the distribution of maximal clique sizes overall show an expected decrease in their frequency as their size increases (Fig. 3A; Supplemental Fig. S31A). Notably, in 10A TAD cliques of size 3 are 17% less frequent than in T1 ($P = 0.023$) (Supplemental Table S10) and are 24% less frequent than in C1 ($P = 7.3 \times 10^{-8}$) (Supplemental Table S10). Larger TAD cliques (size > 4), on the other hand, seem generally more enriched in 10A than in T1 and C1. T1 generally exhibits fewer TAD cliques than 10A and C1 (Fig. 3A). Therefore, larger clique sizes are more apparent in normal 10A cells. These observed trends and differences are likely not related to structural variants in the T1 and C1 cells, as masking these out reveals similar distributions as in nonmasked data (Supplemental Fig. S31).

We overlapped TAD cliques of different sizes with subcompartments and observed that the association of B-type subcompartments increases with TAD clique size (Supplemental Fig. S32). To characterize TAD clique dynamics, we analyzed alterations in TAD maximal clique sizes across cell types (Fig. 3C,D; Supplemental Figs. S33–S39). A large fraction of TADs do not engage in cliques (Fig. 3C). However, those TADs that do form cliques grow or decrease in size (Fig. 3C,D; Supplemental Figs. S33–S39) across all stages. Thus, like subcompartments, TAD cliques also reconfigure extensively in this model system of breast cancer progression.

Our previous observations do not exclude the possibility that growing TAD cliques involve TADs of various chromatin composition, thus belonging to various subcompartments. To explore this, we used HDBSCAN (McInnes and Healy 2017; McInnes et al. 2017) to cluster each clique based on their TAD-wise enrichment in subcompartments. Clustering was performed targeting clusters with a minimum cluster size of 200 TAD cliques, plus a cluster dedicated to collect outliers (for more details, see Methods). The resulting nine clusters (Fig. 3E) show that individual TAD cliques frequently involve multiple types of subcompartments (Supplemental Fig. S40) and thus represent a level of TAD association at which subcompartments of different types potentially intermix.

To investigate the dynamics of TAD clique clusters, we explored the frequency of each of the nine TAD clique clusters in each cell type (Fig. 3F; Supplemental Table S11). We find that most clusters display a cancer progression stage-specific association, such that a cluster tends to be more enriched in one or two of the three cell types examined here (Fig. 3E,F). Some clusters involving B-type subcompartments (clusters 0 and 2) appear to be significantly more prevalent in T1 and C1 compared with 10A.

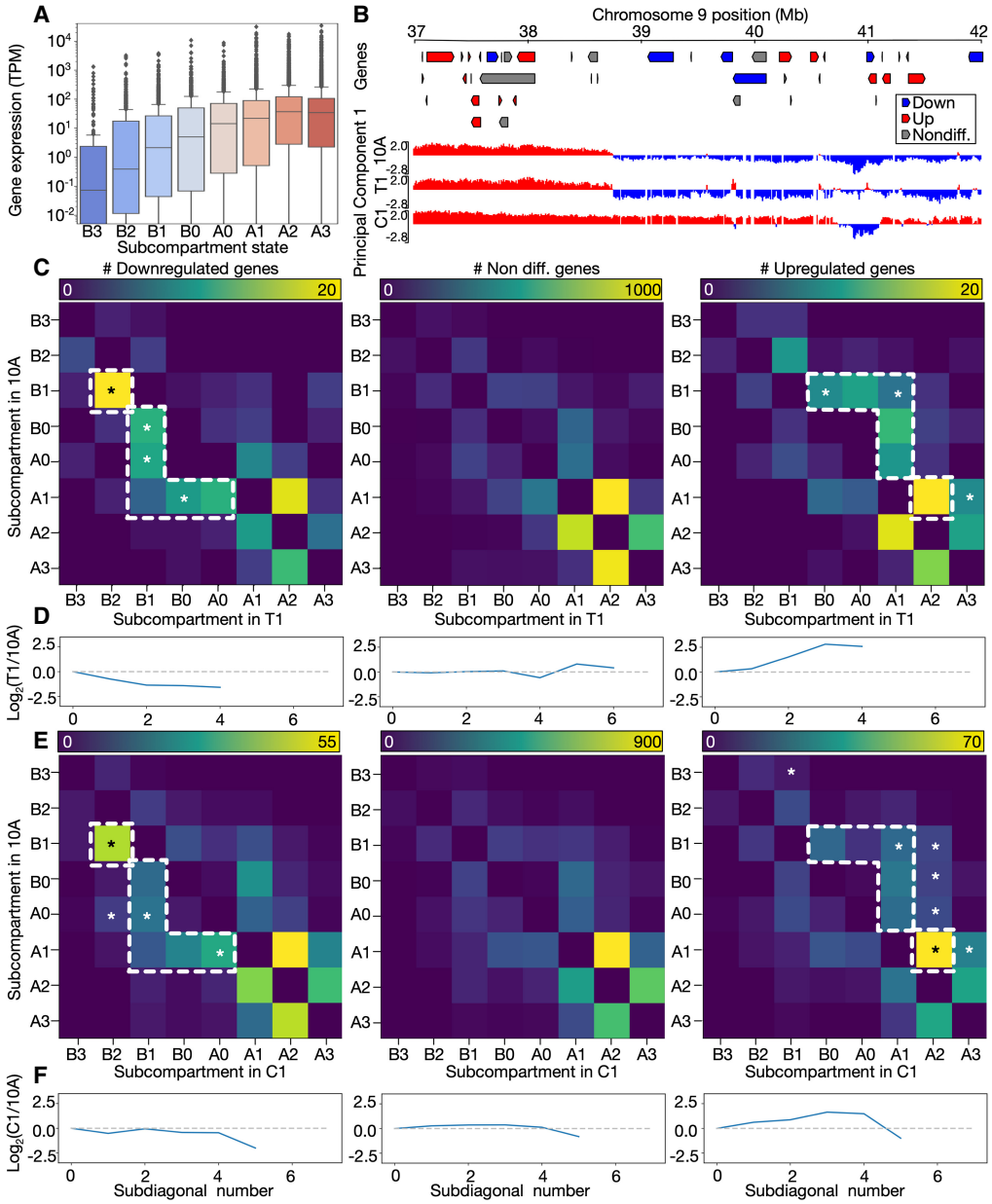

Figure 2. Coordinated changes in subcompartments and gene expression. (*A*) Gene expression levels (TPM) by subcompartment state (protein-coding genes; MCF10A [replicates merged]). Vertical axis in log scale. (*B*) Example of subcompartment switching on Chromosome 9. Up- and downregulated genes in C1 versus 10A indicated in red and blue, respectively. (*C*) Heatmaps showing number of DE downregulated genes (*left*), nondifferential genes (*middle*), and upregulated genes (*right*) in subcompartments switching between 10A (vertical axis) and T1 (horizontal axis). Dotted lines highlight regions with enrichment relative to nondifferential and upregulated genes. Asterisks indicate significant differences (see Supplemental Table S6). (*D*) Log_2-ratio of number of genes in subdiagonal sums in the upper versus lower triangular matrices of the corresponding heatmap from *C*. Ratio not displayed for subdiagonal sums that are equal to zero. (*E*) Heatmaps as in *C* but contrasting 10A with C1. Dotted lines highlight regions with enrichment relative to nondifferential and downregulated genes. Asterisks indicate significant differences (see Supplemental Table S7). (*F*) Log_2-ratio plots as in *D*, but contrasting 10A with C1.

Conversely, some clusters with A-type subcompartments (clusters 1, 5–7) seem to be diminished in T1 but exhibit partial recovery in C1 (Fig. 3F; Supplemental Table S12). These are also clusters enriched in weaker, intermediate subcompartments, which could reflect their increased switching.

In summary, TAD cliques bring together TADs overlapping with various subcompartments, which differ in 10A, T1, and C1 cells. Overall, this signifies the role of subcompartment alterations and intermixing during breast cancer progression in this model system.

Interchromosomal insertion of the *MYC* locus coincides with de novo enhancer contacts and oncogene activation

As expected, the cell types used in our breast cancer progression model harbor structural variations (SVs), including copy-number

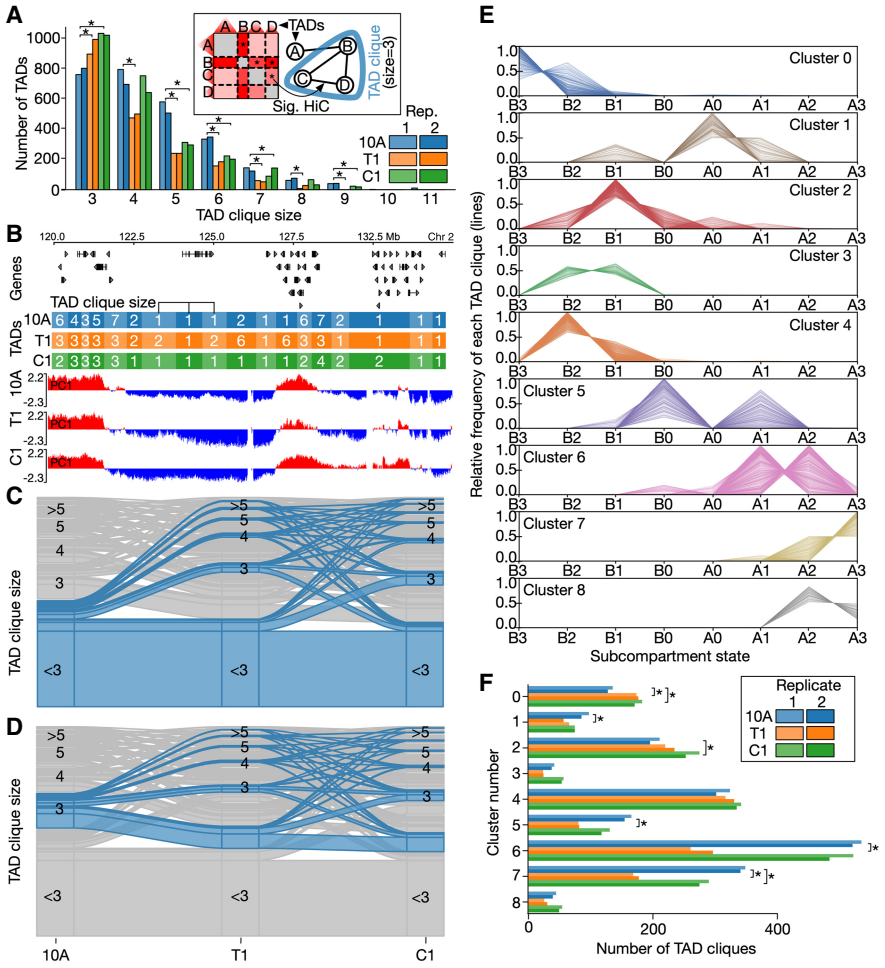

Figure 3. TAD clique dynamics during breast cancer progression stages. (*A*) Absolute number of TADs in cliques with maximal size ranging from three to 11 shown for each of the three breast cancer stages. Asterisks indicate significant differences (see Supplemental Table S10). *Inset* illustrates how "TAD cliques" are identified from the statistically significant Hi-C interactions between TADs. (*B*) Example region on Chromosome 2 showing changes in TAD clique sizes and compartment eigenvalues. (*C*) Alluvial plots highlighting the alluvial path of a nonclique in 10A across the three cancer stages. (*D*) Alluvial path of a TAD clique of size 3 in 10A. (*E*) TAD clique clustering into nine clusters based on their enrichment in subcompartments in the same cell type (vertical axis). Each line represents a TAD clique. (*F*) Numbers of TAD cliques belonging to each of the clusters in *E*. Asterisks indicate significant differences (see Supplemental Table S12). For numbers in the "mixed" cluster, see Supplemental Figure S40.

alterations, insertions, deletions, and translocations. We exploited the ability of Hi-C to detect many of these SVs to call SVs genome-wide (Supplemental Figs. S41–S43). We manually inspected the called variants and confirmed known SVs (Santner et al. 2001), including $t_{(3;9)}$, $t_{(3;5)}$, $t_{(3;17)}$, and $t_{(6;19)}$. In addition, we identified previously undescribed SVs, including $t_{(7;9)}$ and $t_{(10;17)}$. Analyzing the 3D structural consequences of all SVs in these cell lines is computationally infeasible. Nevertheless, to investigate 3D genome consequences of select SVs, we focused on a striking pair of regions on Chromosome 8 (126,330,000–128,235,000 bp; hg38) (Fig. 4A) and Chromosome 10 (71,280,000–73,310,000 bp; hg38) (Fig. 4B). Both regions show a gradual and coordinated increase in copy number, going from T1 to C1 (Fig. 4A,B). A large region at the end of Chromosome 8 is amplified in 10A but not in T1 and C1 (Fig. 4A). The region on Chromosome 8 is well characterized (Visscher et al. 1997; Grisanzio and Freedman 2010) and contains several breast cancer–related genes. These notably include *MYC*, an oncogenic transcription factor with a pivotal role in breast cancer progression (Liao and Dickson 2000), whose amplification and overexpression is a marker of aggressive and invasive breast cancer (Berns et al. 1992; Corzo et al. 2006).

Further inspection of the Hi-C data at the intersection of these two regions in the Hi-C matrix reveals, after ICE balancing, an enrichment of contacts of the region toward the entire Chromosome 10, clearly indicating that the *MYC* locus is inserted in the Chromosome 10 region (Fig. 4C). Enriched contacts with other chromosomes are also seen, owing to the amplification of the locus (Fig. 4C). Zooming in on the specific interchromosomal contacts between the *MYC* locus and the locus on Chromosome 10, we note Hi-C patterns that typically suggest chromatin looping and extrusion of intrachromosomal DNA (Fig. 4D). This again indicates an insertion event. Investigating the region using a different balancing scheme that explicitly models and accounts for copy-number alterations (Servant et al. 2018) reveals similar trends. This indicates that the copy-number alterations themselves are not distorting the Hi-C data in this region (Supplemental Fig. S44). Comparing the contact maps between the three cell types reveals that the insertion is novel in the T1 and C1 cells (Supplemental Fig. S45).

We next examined the chromatin context of this insertion event. Inspection of subcompartments in this region shows that the insertion site on Chromosome 10 is nearly entirely covered by A3 subcompartments (Supplemental Fig. S46A) in all three cell types, indicating that the *MYC* locus is inserted into a highly active genome region on Chromosome 10. Initially, the *MYC* gene resides in an A2 subcompartment on Chromosome 8 in the 10A cell line. However, it switches to a more active or "open" state (A2→A3) in T1 and C1 (Supplemental Fig. S46B).

Next, we assessed the nature of the de novo contacts detected between the *MYC* locus and sites on the host Chromosome 10. To this end, we overlaid positions of MCF10A enhancers predicted from EnhancerAtlas 2.0 (Gao and Qian 2020) onto the Chromosome 8 and 10 amplification units (Fig. 4D; green segments). The four distinct *MYC* interactions detected in the Hi-C map (Fig. 4D, circles) evidently corresponds to (1) two de novo contact points between enhancer elements on Chromosome 10 and the *MYC* promoter and (2) an interaction with enhancer elements ~1 Mbp upstream of the *MYC* promoter (Fig. 4D). Generating a virtual 4C plot (see Methods) (Sexton et al. 2012) anchored at the *MYC* locus confirms that the enrichment of contacts

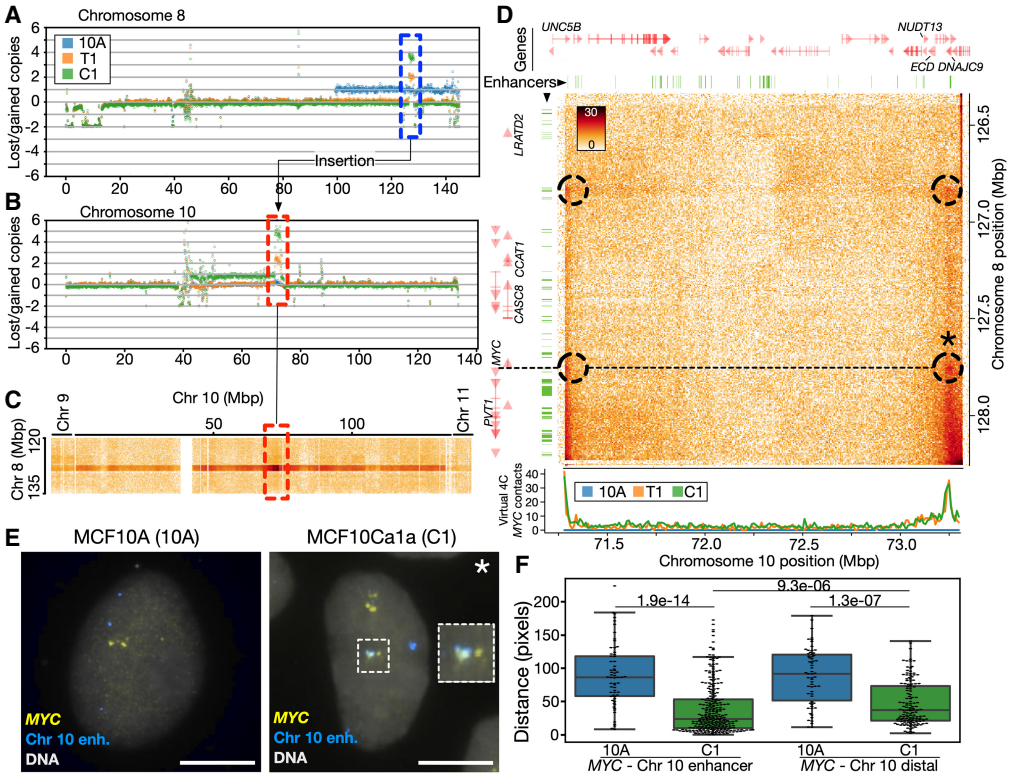

Figure 4. *MYC* locus insertion on Chromosome 10 is accompanied by de novo contacts with a potential enhancer element. (*A*) Lost and gained copies for the entire Chromosome 8 in 10A, T1, and C1 cells. The dotted blue box highlights specific amplification of the Chromosome 10 region. (*B*) Lost and gained copies for the entire Chromosome 10, with a dotted red box highlighting specific amplifications. (*C*) Interchromosomal C1 Hi-C contacts between the region on Chromosome 8 (vertical axis) and the entire Chromosome 10. Highlighted region from *B* indicated in red. End of Chromosome 9 and beginning of Chromosome 11 shown on the *left* and *right* side, respectively. (*D*, *top*) Zoom-in on the C1 Hi-C map of the Chromosome 8/10 amplification unit. Dotted circles indicate enriched "dots" of contacts involving enhancers and *MYC*. Asterisk marks the enhancer interaction validated in FISH in panels *E* and *F*. The dotted line shows the position of the *MYC* gene on Chromosome 8 relative to contacts on Chromosome 10 within the amplification unit. Positions of enhancers from MCF10A indicated as green segments (*top* and *left*). Genes shown in red (*top* and *left*), with names for selected genes. (*Bottom*) Virtual 4C track (see Methods) showing Hi-C contacts from *MYC* toward the Chromosome 10 region for 10A (blue), T1 (yellow), and C1 (green). (*E*) FISH image of the *MYC* gene (yellow) and the Chromosome 10 enhancer (blue; Chr 10 position 73,117,711–73,267,436) expected to interact with *MYC* from the Hi-C data, in 10A and C1 cells. Bars, 7 µm. (*F*) Quantification of probe–probe distances in pixels for the *MYC*–Chr10 enhancer probe pair (*left*; n = 62 [10A] and 269 [C1] observations) and the *MYC*–Chr10 distal probe pair (position 73,794,997–73,992,612; n = 62 [10A] and 140 [C1] observations). N = 28–95 nuclei analyzed. *P*-values are shown (Kolmogorov–Smirnov test).

toward these putative enhancer elements is specific to T1 and C1 (Fig 4D, bottom track).

To validate the de novo interactions (proximity) of the *MYC* gene with elements on Chromosome 10 at the cellular level, we carried out a fluorescence in situ hybridization (FISH) analysis of the relative positions of the *MYC* gene ("*MYC*" probe on Chromosome 8 at position 127,543,567–127,702,694) and either a de novo interacting enhancer on Chromosome 10 ("Chr10 enhancer" probe at position 73,117,711–73,267,436) or another site on Chromosome 10 ~500 kbp downstream ("Chr10 downstream" probe at position 73,794,997–73,992,612; see Methods). The data show that both the "*MYC*–Chr10 enhancer" and the "*MYC*–Chr10 distal" distances are significantly shorter in C1 cells than in 10A cells ($P = 1.9 \times 10^{-14}$ and 1.3×10^{-7} respectively) (Fig. 4E,F; Supplemental Fig. S47A–D), yet the *MYC* enhancer contact is significantly shorter than the distal interaction ($P = 9.3 \times 10^{-6}$). These findings support the view that the *MYC* locus has undergone an insertion into Chromosome 10 in C1 cells. Furthermore, this region harbors enhancer elements that may interact with *MYC*, although resolution limitations of our FISH probes prevent a more detailed characterization.

Furthermore, to provide an element of functional significance to this *MYC* insertion, we examined RNA-seq data for these two regions. The data show an upregulation of breast cancer–related genes in T1 and C1 relative to 10A, including *LRATD2*, *PCAT1*, *CASC19*, *CASC8*, *POU5F1B*, *PVT1*, and *MYC* on Chromosome 8. Several breast cancer–related genes in the Chromosome 10 region are also upregulated, such as *UNC5B*, *CDH23*, *PSAP*, *SPOCK2*, *ASCC1*, *DDIT4*, *NUDT13*, *ECD*, *DNAJC9*, and others (Supplemental Tables S13, S14).

To dissect the functionality of the *MYC* insertion, we designed 5′ and 3′sgRNA pairs to target the de novo putative enhancer interactions (Supplemental Fig. S48A). The *MYC* de novo contacts are proximal to the *DNAJC9* locus on Chromosome 10 (Fig. 4D), at which three putative *cis* regulatory elements (REs) enriched in ENCODE H3K27ac and annotated CREs are found (RE1–3) (Supplemental Fig. S48B). Virtual 4C-track overlays reveal enriched contacts in T1 and C1 cells compared with 10A within this region, with a peak aligning with RE1 (Supplemental Fig. S48B, top). C1 cells were double-transfected with pLentiMultiCRISPR vectors in which Cas9 is coexpressed with puromycin and each sgRNA pair to create six different genomic deletions spanning RE1–3

(Supplemental Fig. S48B) or sgRNAs for mCherry or EGFP as controls. The Cas9 protein was equally expressed from all vectors in C1 cells compared with GAPDH (Supplemental Fig. S48C). After 2 days of puromycin selection for Cas9-expressing C1 cells, we observe a subtle, yet not significant, decrease in *DNAJC9* expression upon Cas9 targeting of the three REs in the locus (Supplemental Fig. S48D). *MYC* expression slightly decreased in C1 cells expressing guide pairs for deletion of RE1.1 and RE1.2 and showed a subtle increase in expression for sgRNAs surrounding RE2.1-2 and RE3.1-2, yet was not statistically significant (Supplemental Fig. S48E). We did not observe any change in expression of *POLR2A*, encoding for the largest subunit of RNA polymerase II on Chromosome 17 upon targeting of Cas9 to RE1–3 or controls (Supplemental Fig. S48F). To evaluate Cas9 editing efficiency, we amplified genomic DNA from puromycin-selected, doubly transfected C1 cells using primers targeting regions surrounding sites recognized by sgRNAs in RE1, -2, or -3 (Supplemental Fig. S49A). PCR products were analyzed by gel electrophoresis, purified, and sequenced (Supplemental Fig. S49B). This evaluates the frequency of indels surrounding each of the sgRNA breakpoints, which was quantified using the TIDE method (Brinkman et al. 2014) and found to be relatively low (0.8–7.6%) (Supplemental Table S19). This may partially explain the subtle effect on *MYC* expression and low Cas9 deletion frequency of RE1–3. However, the genomic complexity and redundancy of the *MYC* locus make definitive conclusions challenging.

Collectively, these observations indicate that the *MYC* locus has not only been amplified but also inserted into a transcriptionally active region of Chromosome 10 in C1 cells, with the formation of de novo *cis* contacts between *MYC* and REs in the host chromosome. Upregulation of *MYC* and of other breast cancer–related genes in malignant C1 cells speculatively suggests functionality of these de novo interactions, yet determination of any involved enhancer elements remains. Our results highlight how structural variants in the genome may cooperate with 3D genomic states during oncogenesis.

Discussion

The intricate 3D organization of the genome plays a pivotal role in gene regulation and cell function. Using a multilayered (epi)genomic approach, we report here for the first time, to our knowledge, significant alterations in higher-order 3D chromatin organization in an in vitro breast cancer progression model.

First, major compartment and subcompartment switching emerges as a prominent dynamic feature that, as expected based on previous studies (Nagai et al. 2019; Liu et al. 2021), correlates with altered expression of genes implicated in breast cancer. Second, we identify a reconfiguration of preferred associations between TADs, or TAD cliques (Paulsen et al. 2019), which bring together new A- and B-type subcompartment associations. Third, spatial subcompartment organization changes in the premalignant and malignant states potentially also contribute to the loss of gene expression control. Collectively, there is a coordinated and progressive reorganization of higher-order genomic structure to an abnormal state. However, casual relationships remain to be determined.

The extent of compartment switching observed in our model (22%) is similar to what is seen during B cell differentiation (28.1%) (Vilarrasa-Blasi et al. 2021), although the latter includes an additional intermediate compartment, inflating the percentage. In contrast, 7.5% of the genome harbors statistically significant compartment switches during differentiation of mouse embryonic stem cells (ESCs) into neuronal progenitors (Chakraborty et al. 2022). A/B-compartment switching varies between 8% and 36% when comparing various ESC-derived cell types (Dixon et al. 2015; Liu et al. 2021). Comparatively, this indicates that the level of switching observed in our oncogenic model is substantial.

Our study did not uncover involvement of substantial TAD border dynamics in the in vitro progression model of breast cancer. However, we cannot entirely rule out its subtle involvement owing to minor differences in consistency across replicates. Further investigation of these differences is needed to uncover any potential functional consequences. Additionally, the potential role of cell heterogeneity in cancer cells cannot be dismissed given that our data represent an overall behavior in bulk cell populations (Zhu et al. 2023). This heterogeneity is expected to primarily affect 3D genome models, but single-cell copy-number variations (CNVs) could also influence estimates of overall copy numbers.

A new observation from our data is a previously unknown insertion of the *MYC* locus (normally localized on Chromosome 8) into highly active A3-subcompartments on Chromosome 10. This is accompanied by de novo interactions correlating not only with elevated *MYC* expression in T1 and C1 cells but also with enhanced expression of other breast cancer–related genes on both chromosomes. Our FISH analysis supports our Hi-C findings of de novo interactions between *MYC* and a region harboring enhancer elements on Chromosome 10 in C1 cells but not in 10A cells. Furthermore, a complementary study has also classified this insertion as an enhancer hijacking event (Wang et al. 2021). Notably, in murine B cells, *MYC* has been shown to relocate to active transcription factories in proximity to *IGH*, a frequent translocation partner (Osborne et al. 2007). However, our FISH results indicate that in 10A cells, *MYC* is not proximal to the Chromosome 10 locus, suggesting *MYC* is not part of a transcription factory with this locus prior to the insertion event. Our analyses did not test for changes in *MYC* location in response to signaling pathway activation as previously described (Osborne et al. 2007) and thus refer to a "baseline" situation in these cells.

Identifying enhancer candidates from Hi-C interactions is challenging owing to the data's low genomic resolution in pinpointing enhancer–promoter contacts. Tools like 4C and Micro-C can provide higher-resolution insights into specific chromatin interactions. However, these techniques do not guarantee that detected contacts correspond to functional interactions, as spatial proximity does not inherently imply biological significance (Splinter et al. 2012; Hsieh et al. 2020).

Focusing on REs1-3 in the *DNAJC9* locus on Chromosome 10 intersecting the central part of the "dot" observed in our Hi-C data (Fig. 4D), we utilized a CRISPR-Cas9 strategy to analyze consequences of deleting these putative enhancer elements in C1 cells. From this, we observed changes in expression of *DNAJC9* and *MYC* compared with the sgRNA controls, yet these changes were not statistically significant. Studies have shown that deleting individual enhancers can have minimal impact owing to redundancy among enhancers and may cause only transient effects (Diao et al. 2016; Osterwalder et al. 2018), possibly explaining why targeting individual *DNAJC9* REs may have a low effect on MYC expression. These effects may, however, also be masked by the large copy numbers of *MYC* and *DNAJC9* loci (Fig. 4A,B). Moreover, the low editing efficiency of individual sgRNAs in pool C1 cells after transient transfection and puromycin selection suggests a low frequency of the RE1–3 deletions (Supplemental Table S19). Therefore, we cannot definitively determine the functional significance of the observed enhancer interactions.

The detailed spatial configuration of this insertion remains unclear, but there are several possibilities. The mirror symmetry and isolated contact map of the Chromosome 10 segment with specific regions of the Chromosome 8 segment (see Fig. 4D), featuring stripes, could suggest that the region exists as circular, extrachromosomal DNA (ecDNA). The formation of ecDNA structures is a common form of oncogenic amplification and often occurs around *MYC* (Hoff et al. 1988; Hung et al. 2022) to promote functional *cis*-regulatory contacts (Morton et al. 2019). However, the focal and limited MYC probe signals (Supplemental Fig. S47A,B) are inconsistent with an ecDNA event (Hung et al. 2021), leaving the structural interpretation unresolved.

Beyond our identified interchromosomal insertion of the *MYC* locus, the intricate aneuploidy patterns commonly observed in cancer pose a challenge that existing standard software and pipelines struggle to effectively address. Promising strides have been made using long-read Hi-C technology (Garg 2023) or statistical approaches (Brunette et al. 2024) to alleviate this issue. Nevertheless, a significant gap remains in the availability of computational pipelines tailored for comprehensive downstream analysis of reordered cancer genome data relative to a karyotypically normal reference genome. This needs to be addressed if further progress in understanding the 3D cancer genome is to be achieved.

In conclusion, our findings provide new insights into the complex genomic structural changes that underlie breast cancer metastatic progression. The interplay between subcompartment switching, LAD-linked subcompartment spatial nuclear reorganization, and TAD clique dynamics represents a multifaceted mechanism by which the 3D genome can be abnormally and dynamically reconfigured during breast cancer development. Targeting these 3D genome alterations could potentially lead to improved or new therapeutic strategies to better treat breast cancer in the future (Park et al. 2023).

Methods

Cells

MCF10A, MCF10AT1, and MCF10Ca1a cell lines were grown in DMEM/nutrient F12 (DMEM/F12) media supplemented with 5% horse serum (MCF10A and MCF10Ca1a) or 2.5% horse serum (MCF10AT1), 14 mM NaHCO3, 10 µg/mL insulin, 2 mM L-glutamine, 20 ng/mL human epidermal growth factor, 500 ng/mL hydrocortisone, and 100 ng/mL cholera toxin. MCF10A cells were obtained from the American Type Culture Collection (CRL-10317). MCF10AT1 and MCF10Ca1a cells were obtained from the Barbara Ann Karmanos Cancer Institute.

Fluorescence in situ hybridization

FISH was done as exactly described by us previously (Paulsen et al. 2017). In short, MCF10A and MCF10Ca1a cells were incubated in a hypotonic buffer, fixed in ice-cold methanol:acetic acid, and dropped on glass slides. BAC FISH probe DNA (BacPac Resource Center) was labeled using a nick translation kit (Roche). The following probes were used:

- "*MYC*" gene probe—clone ID RP11-1136L8, Chr 8, position 127,543,567–127,702,694;
- "Chr10 enh." enhancer probe—clone ID RP11-152N13, Chr 10, position 73,117,711–73,267,436; and
- "Chr10 distal" probe—clone ID RP11-390A15, Chr 10, position 73,794,997–73,992,61.

The *MYC* gene probe was labeled with digoxigenin-11-dUTP (Roche). The Chr10 enhancer and Chr10 distal probes, both positioned to mark sites expected to be proximal to the *MYC* gene on Chromosome 10, were labeled with biotin-16-dUTP (Roche). For each slide, 200 ng of each "MYC gene" + "Chr10 enhancer" probe or of each "MYC gene" + "Chr10 distal" probe was mixed with 30 µg of Cot-1 DNA and 150 µg salmon sperm DNA and precipitated. DNA was dissolved in hybridization mix and preannealed for 1 h. Slides were RNase-treated, washed, dehydrated in ethanol, denatured, and dehydrated again.

Probes were denatured, preannealed, and applied onto cells for overnight hybridization at 37°C. Slides were then washed in 2× SSC and in 0.1× SSC, blocked in 5% skim milk, and incubated with antidigoxigenin (Roche; mouse; 0.4 µg/mL). Slides were washed, incubated with avidin Alexa Fluor 488 conjugate (Invitrogen; 1.7 µg/mL) and Alexa Fluor 594-conjugated anti-mouse (Jackson ImmunoResearch; rabbit; 2.5 µg/mL), washed, and incubated with biotinylated antiavidin D conjugate (Vector; goat; 1.0 µg/mL) and Alexa Fluor 594-conjugated antirabbit (Jackson ImmunoResearch; donkey; 2.5 µg/mL). Slides were washed and incubated 30 min with avidin Alexa Fluor 488 conjugate (Invitrogen; 1.7 µg/mL). Slides were mounted with 0.2 µg/mL DAPI in Dako fluorescent mounting medium. Images were taken under a 100× objective (numerical aperture 1.4) mounted on an IX71 inverted microscope (Olympus) fitted with the DeltaVision wide-field imaging station (GE Healthcare).

Hi-C

Hi-C Libraries were prepared in duplicates for each cell type using the Arima-HiC+ kit (Arima Genomics) according to the manufacturer's instructions. For each cell line (10A, T1, C1) 1 million cells were used per Hi-C reaction. Cells were fixed with 1.2% formaldehyde for 12 min. All subsequent steps were carried out according to the Arima protocol. Resulting libraries were amplified with five PCR cycles and sequenced on an Illumina NovaSeq 6000 instrument in paired-end run with 2 × 101 bp.

RNA-seq

All mRNA-seq experiments were performed in triplicate. Total RNA was isolated using the Qiagen RNeasy kit following the manufacturer's instructions. Stranded mRNA-seq libraries were constructed using an Illumina mRNA prep kit, following the vendor protocol with poly(A) enrichment (Illumina). Libraries were sequenced with 2 × 75 bp paired ends on an Illumina NextSeq 500.

CRISPR-Cas9 RE deletions

Paired sgRNAs (5′- and 3′-sgRNAs) were designed to target both ends of selected candidate REs in the human *DNAJC9* locus to create a deletion. sgRNA with on-target high predicted cleavage and a low number of off-targets was carefully selected using the UCSC Genome Browser (hg38) (Perez et al. 2025) CRISPR target track based on CRISPOR (Haeussler et al. 2016). The sgRNAs were cloned stepwise into a Lenti-multi-CRISPR plasmid that contains human codon-optimized *Streptococcus pyogenes* Cas9 (spCas9) followed by a *porcine teschovirus-1* 2A peptide (P2A) self-cleavage sequence upstream of the puromycin sequence (Addgene plasmid 85402) (Cao et al. 2016). The first gRNA was cloned by insertion of two annealed complementary oligonucleotides into the BsmBI site using the infusion cloning approach (TaKaRa Clontech) (Supplemental Table S20). The plasmid containing one guide was opened with NheI (NEB), and a DNA fragment containing U6 promoter, sgRNA, and scaffold (IDT technologies) was inserted by infusion cloning (Supplemental Table S19). A total of two sgRNA pairs per

three *DNAJC9* locus REs were cloned. sgRNAs for mCherry and EGFP were designed as negative controls and inserted separately as annealed oligonucleotides by infusion cloning into the BsmBI site in two separate Lenti-multi-CRISPR plasmids. All plasmids were fully sequenced by Oxford Nanopore Technologies (Eurofins Genomics) and the sequences manually verified.

A total of 600,000 C1 cells were seeded per six wells in DMEM/F12 supplemented media as described above. After 24 h, three technical replicates of the six wells were transfected with 2.5 µg of Lenti-multi-CRISPR plasmid containing 5′- and 3′-sgRNA RE pair or EGFP sgRNA or Cherry sgRNA using Lipofectamine 3000 (Thermo Fisher Scientific). At 24 h post transfection, each six well was transfected again with 2.5 µg of the same sgRNA-Lenti-multi-CRISPR plasmid using Lipofectamine 3000. After 24 h, the double-transfected C1 cells were selected with 4 µg/mL of puromycin for 48 h. On day five, the medium was changed and the cells left to recover. On day six, the six wells were washed in 1× PBS and either collected in 500 µL TRIzol Reagent (Thermo Fisher Scientific) for total RNA or trypsinized and pelleted for genotyping. Genomic fragments spanning sgRNAs in RE-1, -2, or -3 were amplified by Q5 polymerase (NEB); single bands were purified and Sanger-sequenced (Eurofins) (Supplemental Table S20). The TIDE web tool (version 3.3.0) (Brinkman et al. 2014) was used to determine the efficiency of individual guide RNAs in generating targeted indel mutations in a cell population at 25 nucleotides from the breakpoint (Supplemental Table S19).

To verify Cas9 expression, whole-cell lysates of C1were lysed in 2× SDS loading dye and loaded on a 4%–20% SDS PAA gel (Bio-Rad) after lipofectamine transfection with pLentiMultiCRISPR (Addgene 85402) with sgRNA pairs spanning RE1.1, RE1.2, RE2.1, RE2.2, RE3.1, and RE3.2 and with guide controls against EGFP and mCherry. Whole-cell lysates from C1 cells transfected with a vector with humanized Cas9 (hCas9; Addgene 41815) were loaded as a control (Mali et al. 2013). The proteins were transferred to a PVDF membrane (Millipore) and detected with anti-Cas9 (Cas9-clone 7A9, Sigma-Aldrich MAC133, expected size 160 kDa) and anti-GAPDH (Novus Biologicals, expected size 37 kDa) as a loading control followed by secondary IRDye 800CW Donkey Anti-Mouse IgG (H+L), 0.5 mg Catalog No. 926-32212 and IRDye 680RD Donkey Anti-Goat IgG (H+L), 0.5 mg Catalog No. 926-68074 (LicorBiO) and detection by an OdysseyDLx imager.

Total RNA from the three six wells for each 5′- and 3′-sgRNA pair or control EGFP/mCherry sgRNAs Lenti-multi-CRISPR plasmid-transfected cells was purified using a RNeasy kit with an on-column DNase I treatment according to the manufacturer's protocol (Qiagen). RNA was eluted in 25 µL ddH2O, quantified using NanoDrop, and quality-controlled by agarose gel electrophoresis. A total of three different replicates of double transfections with puromycin treatment conducted at different times was processed. Two milligrams of total RNA was used for generation of cDNA primed with oligo(dT) according to the manufacturer's instructions (AffinityScript cDNA synthesis kit; Agilent 600559).

The endogenous *MYC*, *DNAJC9*, *RPL30*, and *POL2RA* mRNA expression levels were measured by quantitative RT-PCR using EvaGreen dye (Biotium) on a Bio-Rad CFX real-time system. A standard curve was generated for each primer set with thermal cycle conditions for 40 PCR cycles (12 min at 95°C, 15 sec at 95°C, 30 sec at 65°C, 30 sec at 72°C) (Supplemental Table S19). The real-time efficiency (percentage) for each primer pair was calculated according to E = −1 + 10(−1/slope) (Rasmussen 2001) and the converted primer efficiency = E (%)/100 + 1.

Three technical replicates were run per qRT-PCR (n = 3 rounds of CRISPR; n = 9 technical replicates). Relative gene expression was analyzed using Pfaffl's mathematical model (Pfaffl 2001) for 5′- and 3′-sgRNA RE pair or mCherry sgRNA versus EGFP sgRNA-transfected cells normalized to housekeeping gene *RPL30* as a reference:

$$Ratio\ (RQ) = (E_{target\ gene})^{\Delta C(targe\ genet)} / (E_{reference})^{\Delta C(reference)},$$

ΔCt(target) = Ct(Mean target gene in CRISPR control gRNA (EGFP) cells) – Ct(target gene in CRISPR RE or mCherry sgRNA cells),

ΔCt(reference) = Ct(reference gene in CRISPR control gRNA (EGFP) cells) – Ct(reference gene in in CRISPR RE or mCherry sgRNA cells).

The ratio was calculated for the average of technical CP values per replicate, and then these were averaged and STDEV was calculated. The control sample was from pooled cells transfected with *EGFP* gRNA. Samples are RE = pool of cells with guide pair for RE1.1; RE1.2; RE2.1; RE2.2; RE3.1, or RE3.2. A pool of cells with mCherry gRNA acts as a negative CRISPR control. Target genes are *MYC, DNAJC9,* and *POLR2A* (internal control target gene on Chromosome 17, assumed not affected by CRISPR deletions). qRT-PCR data statistical analysis and plots were performed on GraphPad Prism 10.4.1. For statistical evaluations of the determined CP variations and calculated relative expression variations, data were analyzed for significant differences by one-way ANOVA using a Dunnett's multiple comparisons test between mCherry sgRNA and each of the different RE guide pairs performed on GraphPad Prism v10.4.1 (the confidence level was set to 0.05) (Supplemental Table S20).

ChIP-seq of lamin B1 and mapping of LADs

ChIP of lamin B1 was done as previously described by us (Rønningen et al. 2015). In short, cells were cross-linked with 1% formaldehyde; lysed in 50 mM Tris-HCl (pH 7.5), 10 mM EDTA, 1% SDS, and protease inhibitors; and sonicated in a Bioruptor (Diagenode) into ∼200 bp fragments. After sedimentation, the supernatant was diluted 10-fold in the RIPA buffer. Chromatin was incubated overnight at 4°C with antibodies to lamin B1 (10 µg per 10 million cells; Abcam ab16048), coupled to Invitrogen dynabead protein A/G (Thermo Fisher Scientific). ChIP samples were washed four times in ice-cold RIPA buffer, after which cross-links were reversed and the DNA was eluted for 6 h at 68°C. DNA was purified using phenol–chloroform isoamylalcohol and dissolved in H_2O. ChIP-seq libraries were prepared using the Diagenode Microplex library preparation kit v2 and TruSeq LT indexes, and samples were sequenced (single-end) on an Illumina HiSeq 4000.

ChIP sequence reads were mapped to hg38 with Bowtie 2 v2.4.1 (Langmead and Salzberg 2012; https://github.com/BenLangmead/bowtie2) after removing duplicates using Picard MarkDuplicates (http://broadinstitute.github.io/picard/). Reads from both input samples were merged and mapped to hg38, and duplicates were removed as above. To alleviate normalization bias, each pair of mapped ChIP and input read files contained the same read depth after down-sampling reads for each chromosome. Mapped reads were used to call LADs from 10 consecutive runs of enriched domain detector (EDD) (Lund et al. 2014; http://github.com/CollasLab/edd) with auto-estimation of Gap Penalty and BinSize, and mean GapPenalty and BinSize values were used for a final EDD run (Forsberg et al. 2019). Final LADs for each cell type were the union of the three replicates.

Structure of code used for data analyses

The source code used to produce all results presented in this paper (except LADs analysis and downstream analysis of Chrom3D models) is hosted at GitHub (https://github.com/paulsengroup/2022-

mcf10a-cancer-progression) and is archived at Zenodo (https://doi.org/10.5281/zenodo.17121063) and as Supplemental Code.

Most of the computation is organized into Nextflow workflows, with each workflow performing a subset of the data analysis (e.g., the differential expression and comparative analyses are performed by two different workflows).

All workflows mentioned in the method sections are hosted on the GitHub repository inside the folder workflows/. Scripts and Jupyter notebooks are found under the folders bin/ and notebooks/, respectively.

Some of the analysis steps required us patching third-party tools. The patches are described in the section Software Patches, and patch files are available on the GitHub repository under the container/patches/. All our patches were contributed upstream. Most of our patches were accepted by upstream and are already part of a stable release. Figures were produced directly by workflows or using scripts and notebooks under bin/plotting and notebooks/, respectively. The final version of figures was produced by assembling draft images with Inkscape and Omnigraffle.

Data analysis workflows were run on a HPC cluster using Apptainer (Kurtzer et al. 2017).

Hi-C preprocessing

We used a modified version of nf-core/hic v2.0.0 (Servant et al. 2015; Ewels et al. 2016, 2020; da Veiga Leprevost et al. 2017; https://zenodo.org/records/2669513) for quality control, sequence mapping, and filtering using hg38.p14 (International Human Genome Sequencing Consortium 2001). The workflow was run using restriction enzymes for the two-enzyme Arima kit (--restriction_site="^GATC,G^ANTC" and --ligation_site="GATCGATC,GANTGATC,GANTANTC,GATCANTC"). The following optional flags were specified: --skip_maps, --skip_dist_decay, --skip_tads, --skip_compartments, --skip_balancing, --skip_mcool, --split_fastq=false. The .chrom.sizes file given to nf-core/hic was filtered to only retain chromosome sequences using grep "^chr[[:digit:]XY]\+[[:space:]]." Chromosomes were sorted by name using gnu sort -V. The output of nf-core/hic was compressed using workflow robomics/compress-nfcore-hic-output v0.0.1 (https://zenodo.org/records/7949266). Hi-C contact matrices in multiresolution cooler format were generated using a patched version of cooler v0.9.1 (Abdennur and Mirny 2020). We used cooler cload to ingest interactions in .validPair format into a .cool file at 1 kbp resolution. We then converted the .cool file to the .mcool file format using cooler zoomify. Finally, matrices were balanced with several methods using juicer_tools v2.20.00 (VC, KR, SCALE methods using intrachromosomal, interchromosomal, and genome-wide interactions) (Durand et al. 2016), and cooler (ICE method using intrachromosomal, interchromosomal, and genome-wide interactions) (Abdennur and Mirny 2020). Regions overlapping centromeres and assembly gaps for hg38 were masked before balancing (files retrieved from UCSC FTP server on February 24, 2023). Downstream analyses were performed using *cis*-only, ICE balanced matrices unless otherwise specified. For some analyses, replicates for the same condition were merged to generate deeper Hi-C matrices using cooler merge (Abdennur and Mirny 2020). The resulting matrices were then coarsened and balanced using cooler and juicer_tools as outlined above (Durand et al. 2016; Abdennur and Mirny 2020). All the above steps were performed by the running script run_nfcore_hic.sh, which runs workflow postprocess_nfcore_hic.nf after running nf-core/hic. Supplemental Figure S1 was generated by workflow postprocess_nfcore_hic.nf based on the output of nf-core/hic. Supplemental Figure S2 was generated by workflow comparative_analysis_hic.nf by running the Python implementation of HiCRep v0.2.6 (Yang et al. 2017; Lin et al. 2021). Correlation values shown in the plot were computed as the weighted average of the correlation coefficient of each individual chromosome using chromosome sizes as weights. Individual correlation coefficient values are available in Supplemental File S1.

In addition, interactions involving Chromosome 8 and Chromosome 10 were balanced with LOIC using the iced package (v0.5.10) (Servant et al. 2018). LOIC balancing is performed in one of the steps of the comparative_analysis_hic.nf workflow.

RNA-seq preprocessing

We used nf-core/rnaseq v3.12.0 (https://zenodo.org/records/7998767) to perform quality control, trimming, and alignment and to generate the gene expression matrix (Di Tommaso et al. 2017; Ewels et al. 2020). This includes the following tools: featureCount (Liao et al. 2014), GffRead (Pertea and Pertea 2020), MultiQC (Ewels et al. 2016), preseq (Daley and Smith 2013), Qualimap 2 (Okonechnikov et al. 2016), RSeQC (Wang et al. 2012), Salmon (Patro et al. 2017), SortMeRNA (Kopylova et al. 2012), STAR (Dobin et al. 2013), StringTie2 (Kovaka et al. 2019), UCSC Tools (Kent et al. 2010), DESeq2 (Love et al. 2014), dupRadar (Sayols et al. 2016), Tximeta (Love et al. 2020), Bioconda (Grüning et al. 2018), BioContainers (da Veiga Leprevost et al. 2017), and Apptainer (Kurtzer et al. 2017). nf-core/rnaseq was run using hg38.p14 as the reference genome (International Human Genome Sequencing Consortium 2001). The FASTA file used as reference was first filtered using SeqKit v2.5.1 (Shen et al. 2016) to remove mitochondrial chromosomes and unplaced scaffolds by using the following pattern "^chr[XY\d]+$." We used the genome annotation for hg38 from GENCODE v43 (Frankish et al. 2019). The pipeline was launched using the following optional flags: --aligner=star_salmon, --pseudo_aligner=salmon, --extra_salmon_quant_args="--seqBias --gcBias", --gencode=true.

The workflow was run by launching script run_nfcore_rnaseq.sh.

ChIP-seq analyses of public data sets

We used nf-core/chipseq v2.0.0 to perform the quality control, mapping, filtering, and peak calling of publicly available ChIP-seq data sets from MCF10A cells (Di Tommaso et al. 2017; Ewels et al. 2020). This includes the following tools: BWA (Li and Durbin 2009), BEDTools (Quinlan and Hall 2010), BamTools (Barnett et al. 2011), deepTools (Ramírez et al. 2016), featureCounts (Liao et al. 2014), HOMER (Heinz et al. 2010), MultiQC (Ewels et al. 2016), phamtompeakqualtools (Landt et al. 2012), preseq (Daley and Smith 2013), SAMtools (Li et al. 2009), UCSC Tools (Kent et al. 2010), DESeq2 (Love et al. 2014), Bioconda (Grüning et al. 2018), BioContainers (da Veiga Leprevost et al. 2017), and Apptainer (Kurtzer et al. 2017). The pipeline was run using hg38.p14 as reference genome from UCSC (International Human Genome Sequencing Consortium 2001) and the gene annotation from GENCODE v43 (Frankish et al. 2019). BWA was used for mapping, and the effective genome size for MACS2 was computed using unique-kmers.py from khmer v2.1.1 (Döring et al. 2008; Zhang et al. 2008; Li and Durbin 2009; Crusoe et al. 2015). Option --narrow_peak was specified when processing ChIP-seq data sets for CTCF and TP53.

The workflow was launched by executing script run_nfcore_chip.sh.

Calling of CNVs

CNVs were called from Hi-C data using HiNT-CNV v2.2.8 (Wang et al. 2020).

The list of restriction sites for the Arima 2-enzyme kit was generated by the running script bin/compute_restriction_sites_for_hint.py on hg38.14 (International Human Genome Sequencing Consortium 2001).

Reference data and background matrices used by HiNT were downloaded from https://compbio.med.harvard.edu/hint/ and have been archived on Zenodo at the following doi: 10.5281/zenodo.13255302.

Data analysis steps are defined in workflow detect_structural_variants, which was run by the launching script run_detect_structural_variants.sh.

Annotation of SVs

Large SVs such as translocations were manually annotated by us by looking for square regions of enriched interactions and sharp transitions in the *trans* portion of the Hi-C matrix (see Supplemental Tables S15–S17).

TAD analyses

We determined genomic positions of TADs in all samples with HiCExplorer v3.7.2 using default parameters on matrices at 10, 20, 50, and 100 kbp resolution (Ramírez et al. 2018; Wolff et al. 2018, 2020). Workflow tad_analysis.nf was used to generate the draft figures that went into Supplemental Figures S3–S6. The same workflow was also used to run HiCExplorer. Draft figures for Figure 1A were generated using notebook bin/plotting/plot_tads_higlass.ipynb. Supplemental Figure S5 was generated by fetching interactions at 50 kbp resolution for each 10A TAD across the cell types, masking values overlapping the first 150 kbp around the diagonal. Finally, interactions were summed and normalized by TAD size (as the number of pixels overlapping a TAD). To generate Supplemental Figure S6, TADs from two cell types were paired based on the highest overlap (JI of genomic coordinates), and the distribution of pairwise overlaps was plotted. TADs overlapping or around regions masked by matrix balancing were not considered, as the insulation score can be unreliable in these regions.

The analysis is encapsulated by workflow tad_analysis.nf, which was run by launching script run_tad_analysis.sh.

TAD clique analyses

The TAD annotation generated by HiCExplorer was given as input to workflow (https://github.com/robomics/call_tad_cliques) v0.5.1 (doi.org/10.5281/zenodo.12689353) to identify cliques of TADs as previously outlined (Paulsen et al. 2019). TAD cliques were called across all three cell types using TADs from 10A. This is a necessary simplification required us to allow comparing cliques across cell types in downstream analyses. The output of https://github.com/robomics/call_tad_cliques was used as input for workflow postprocess_call_tad_cliques.nf, which generated the draft version of panels B through D of Figure 3 and Supplemental Figures S31, S33–S39.

Workflows were launched by running script run_call_tad_cliques_workflow.sh.

Subcompartment analyses

Subcompartments were determined using a patched version of dcHiC d4eb244 (Chakraborty et al. 2022). A- and B-compartment annotation was generated from subcompartments by aggregating A-subcompartments and B-subcompartments. The dcHiC processing pipeline was wrapped in workflow compartment_analysis.nf, which converts matrices in a .mcool file to a format understood by dcHiC and runs all analysis steps except those for --pcatype=trans, --pcatype=fithic, --pcatype=dloop, and --pcatype=enrich. When appropriate, the reference genome assembly, annotation, and .chrom.sizes for hg38.p14 were provided to dcHiC through the --gfolder option. We fixed the seed used by dcHiC to ensure our results are reproducible by others. Subcompartments were called at the following resolutions: 10 kbp, 20 kbp, 50 kbp, 100 kbp, 200 kbp, and 500 kbp. However, only subcompartments at 10 kbp resolution were used for further data analyses. The same workflow was used to generate the draft figures for panels B and D of Figure 1, as well as Supplemental Figures S16–S19. Figure 1C was generated using the output of the viz step of dcHiC as the starting point.

Supplemental Figures S9–S11 were generated by tracking subcompartment assignments across cell types. Bins were paired based on their genomic coordinates, and then subcompartment switches were classified into five classes: neutral (no switch), reverted (e.g., switch occurred in T1 but was reverted in C1), transition to open/close, and other switches (e.g., partial reversion: A3→A1→A2). Subcompartment switches were further classified using a delta score Δ_{ij}, defined as follows. We quantify the degree of switching between two stages, i and j (Δ_{ij}) by assigning ranks to the subcompartments (B3 = 0…A3 = 7) and calculating the rank difference. A negative rank difference (Δ_{ij}) thus indicates a transition toward B-type subcompartments. Conversely, a positive value signifies a shift toward A-type compartments.

Data analysis steps are defined in workflow compartment_analysis, which was run by executing script run_compartment_analysis.sh.

Virtual 4C

Draft virtual 4C figures (Fig. 4D, bottom; Supplemental Fig. S48B) have been generated using notebook virtual_4c_myc.ipynb.

The two figures were generated by fetching and normalizing interactions for the following regions from the merged Hi-C maps for 10A, T1, and C1:

- Figure 4D, bottom—"Chr 8: 127,736,000–127,737,000" (MYC promoter) and "Chr 10: 71,280,000–73,310,000" (10 kb resolution) and
- Supplemental Figure 48B—"Chr 8: 127,736,000–127,737,000" (MYC promoter) and "Chr 10: 73,240,000–73,270,000" (5 kb resolution).

Differential expression analyses

Differential expression analysis was performed using DESeq2 v1.38.0 and apeglm v1.20.0 (Love et al. 2014; Zhu et al. 2019).

DESeq2 was called with default parameters using the raw count table produced by nf-core/rnaseq as input. Log_2-fold-change shrinkage estimation was computed with the lfcShrink function from DESeq2 using apeglm as shrinkage estimator and the following log_2-fold-change cutoffs: 0.0, 0.1, 0.25, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, and 5.0.

The analysis was automated using workflow diff_expression_analysis.nf, which was launched with run_diff_expression_analysis.sh.

Supplemental Figure S28 was generated by running script run_cluster_profiler_do.py. The script uses clusterProfiler v4.8.1 (Yu et al. 2012; Wu et al. 2021) and the Disease Ontology (DO) database from the DOSE v3.26.1 package (Yu et al. 2015) to perform overrepresentation analysis of DO terms (Schriml et al. 2023). The package enrichplot v1.20.0 (Wu et al. 2021) was used for visualization. The IDs of DE genes were provided as input to clusterProfiler. Genes were considered as DE based if they exhibit an absolute log_2-fold-change value of ≥0.5 and a P-value of ≤0.01 (correcting for

multiple testing). Furthermore, clusterProfiler was run using a *Q*-value of 0.05.

Comparative analyses

All comparative analyses were performed using workflow comparative_analysis.nf.

Supplemental Figure S15 was generated by overlapping subcompartment labels with several different epigenetic markers. First, each ChIP-seq peak was assigned a score by summing the raw ChIP-seq signal over each peak, and then each peak was assigned to a subcompartment. Finally, peaks were grouped by subcompartment label, and the mean signal was computed for each subcompartment.

Figure 2A and Supplemental Figure S29 were generated by overlapping expression levels in TPMs with subcompartment labels overlapping genes from GENCODE v43 for hg38. In case a gene was tagged with multiple subcompartment labels, the gene was assigned the label with the largest coverage. In case of a coverage tie, the gene was discarded (note that this is an extremely rare occurrence).

Figure 2B was generated using pyGenomeTracks (Ramírez et al. 2018; Lopez-Delisle et al. 2021) to visualize up-/downregulated genes overlapping a region involved in subcompartment switching.

Panels C through F of Figure 2 were generated by overlapping subcompartments with DE genes across all three cell types. First, genes were assigned a subcompartment label by overlapping their TSS with the subcompartment annotation generated by dcHiC. Next genes were grouped based on $\log_2$-fold-change and *P*-value in downregulated, non-DE, and upregulated genes (lfc = 2.0; *P*-value = 0.01). Finally, each group of genes was plotted as heatmaps showing subcompartment switches across pairs of cell types (genes not involved in subcompartment switches are not shown). Panels D and F of Figure 2 were generated from the table underlying heatmaps 2C and 2E as follows: compute the sum of genes for each diagonal *i*; starting from diagonal $i = 1$; pair diagonal *i* with diagonal $-i$; and, finally, compute the $\log_2$-ratio between the sum of diagonal *i* and the sum of diagonal $-i$. Positive values indicate a positive correlation between one of the classes of DE genes and switches toward open chromatin, and negative values indicate a positive correlation with subcompartment switches toward close chromatin.

The draft figure for Figure 3A was generated with bin/plotting/plot_tad_cliques.py using TAD cliques, TADs, compartment PCA, and GENCODE v43 gene annotation as input.

The draft figure for Supplemental Figure S32 was generated by overlapping subcompartment states with TADs annotated with the size of the largest clique to which they belonged.

The draft figures for panels F through G of Figure 3 and Supplemental Figure S40 were generated by clustering TADs and TAD cliques using HDBSCAN (McInnes et al. 2017; McInnes and Healy 2017) based on their subcompartment composition.

First, domains (i.e., TADs or TAD cliques) were annotated with their subcompartment state composition by overlapping subcompartment states with the domains using annotate_domains_with_subcompartments.py. This resulted in a count matrix with one row per domain, in which each row counts the number of bins labeled for each subcompartment state. For example, given a domain overlapping three A1 bins, 10 A2 bins and two A3 bins, the entry in the count matrix corresponding to this domain would be 0, 0, 0, 0, 0, 3, 10, 2.

Next, this count matrix was given as input to cluster_domains_by_subcompartment_state.py, which clustered domains using HDBSCAN.flat based on their similarity in subcompartment composition. Clustering was performed using the following settings: n_clusters=9, metric=euclidean, min_cluster_size=200, min_samples=5, cluster_selection_method=leaf. Finally, clusters were visualized using plot_domain_subcompartment_clusters.py.

All comparative analyses are defined in workflow comparative_analysis.nf, which was launched with script run_comparative_analysis.sh.

3D genome modeling

A patched version of Chrom3D v1.0.2 was used to generate 3D genome models for 10A, T1 and C1 (Paulsen et al. 2017). Briefly, this Nextflow workflow runs NCHG (doi.org/10.5281/zenodo.12680450) to identify statistically significant *cis* and *trans* interactions. Translocated regions were excluded when identifying statistically significant trans interactions. For intrachromosomal interactions, we used a log-ratio of 1.2 and adjusted-*P*-value of 0.01 as cutoffs, whereas for interchromosomal interactions a log-ratio of 0.25 and adjusted-*P*-value of 0.01 were used as cutoffs. The workflow then uses statistically significant interactions in Chrom3D simulations as spatially proximal interaction constraints. Furthermore, the workflow matches LADs to the corresponding TAD domains such that these domains can be used as peripheral subnuclear constraints in the simulations. A total of 100 Chrom3D simulations were run for each condition with a nuclear occupancy of 0.15, at TAD resolution and 2 million simulation steps (Paulsen et al. 2017).

ChimeraX was used to visualize the simulated Chrom3D models (Pettersen et al. 2021). The median distance of each chromosome from the nuclear center was calculated for each Chrom3D simulation and visualized using ggplot2 (Villanueva and Chen 2019). The median distance and the standard deviation of each subcompartment were likewise evaluated for each Chrom3D simulation and visualized using ggplot2. Subcompartment distances from the nuclear center were statistically evaluated and compared within conditions using the Wilcoxon rank-sum test in R (R Core Team 2024) with the function wilcox.test.

FISH analysis

Analysis of the FISH data was automated using workflow fish.nf. Draft plots have been generated using notebooks plot_fish_blob_distance_stats.ipynb and plot_fish_blob_radial_positioning.ipynb.

The FISH data analysis consisted of three steps:

- Nuclei segmentation,
- Probe localization, and
- Plot generation and statistical tests.

Nuclei segmentation

Nuclei segmentation was performed using DAPI signal (i.e., the blue channel of the RGB image produced by the microscope).

First, RGB images containing one or more nuclei were converted to grayscale by extracting the blue channel from the images. Grayscale images were then sharpened using unsharp masking. This was done to improve the separation of the profiles of nuclei that were located very close to one another. Next, foreground objects (i.e., nuclei profiles) were highlighted using Otsu thresholding (Otsu 1979). Objects that were significantly smaller than plausible nuclei profiles were masked out. Next, we computed the contour of each individual object using the "marching squares" method. Finally, object contours were used to crop out individual nuclei.

To properly deal with various edge cases (such as mitotic nuclei and overlapping nuclei), nuclei whose contour significantly

deviated from an ellipsoid were flagged so that they could be skipped in subsequent analysis steps. This was achieved by first fitting an ellipse to the contour detected using the first algorithm previously described (Fitzgibbon and Fisher 1995). Next, we computed the overlap coefficient (OC) between the fitted ellipse and the nucleus contour and masked nuclei with an OC of $\leq$0.95. Finally, nuclei that were only partially visible in an image were also flagged.

Probe localization

For each nucleus, we extracted the red and green channels from the RGB image produced by the segmentation step. The resulting grayscale images were processed independently as follows.

Given a probe and a cell line, we can estimate the number of probe signals we expect to find in each nucleus. This information was used to guide the blob (i.e., probe signal) detection using the Laplacian of Gaussian (LoG) method. We performed a sweep of LoG parameters, starting from parameters that were suitable to detect large blobs in an image. If the first run of the LoG algorithm detected N or more blobs, where N is the number of expected probe signals, then the parameter sweep was halted. Otherwise, the coordinates of detected blobs (if any) were stored, and the LoG algorithm was rerun using parameters to detect progressively smaller blobs. Every iteration, the coordinates of newly detected blobs were merged with the coordinates of blobs detected in previous runs of the algorithm. The parameter sweep stops as soon as N or more blobs were detected or when the parameter space was exhausted.

Plot generation and statistical tests

The coordinates of blobs identified in the previous step were finally used to generate Figure 4F and Supplemental Figure S47.

The number of nuclei and observations for each condition is shown in Supplemental Table S18.

Color mapping strategy

We provide colorblind-friendly versions of the raw and processed FISH images. These images were generated using script color_mapping_fish.py. In brief, this script applies a local linear mapping to convert the original colors (blue for DAPI and red/green for the FISH probes) to different colors that can be more easily distinguished by those affected by common forms of colorblindness. The mapping is local because we only map colors for pixels whose original color is perceptually different from black and white.

Software used throughout data analysis code

The following software packages were used throughout the code used for data analysis:

- BEDTools (Quinlan and Hall 2010) to perform common set operations on genomic intervals,
- bedGraphToBigWig (Kent et al. 2010) to convert bedGraph files to bigWig format,
- bioframe (Open2C et al. 2024) to perform common set operations on genomic intervals,
- hictkpy (Rossini and Paulsen 2024) to perform certain input-output operations on .cool files,
- Matplotlib (Hunter 2007) to generate plots,
- NumPy (Harris et al. 2020) to efficiently perform arithmetic operations on vector data,
- OpenCV to manipulate and process microscopy images,
- pandas (McKinney 2010) to read, write and manipulate tabular data using dataframes,
- pyBigWig to read and write bigWig files,
- SAMtools (Danecek et al. 2021) to index FASTA files and perform IO operations on SAM, BAM and CRAM files,
- scikit-image (van der Walt et al. 2014) to manipulate and process microscopy images,
- SciPy (Virtanen et al. 2020) to compute Pearson's correlation values and perform statistical tests, and
- seaborn (Waskom 2021) to generate plots.

Software patches

This section describes the purpose of the patches we applied to Chrom3D, cooler, dcHiC, and HiNT. Patch files are available under containers/patches.

The patch for nf-core/hic v2.0.0 (https://zenodo.org/records/2669513)

Our patch involved updating the pipeline to correctly handle the --restriction_site and --ligation_site parameters.

Our patch was accepted by upstream and is now part of v2.1.0 of the pipeline. For more details, see nf-core/hic/pull/153.

Patches for dcHiC d4eb244 (Chakraborty et al. 2022)

We patched dcHiC as follows:

- Define a default seed and introduce a --seed CLI option to ensure results are reproducible across runs;
- Wrap calls to depmixS4::fit into a try-catch block to attempt model-fitting up to five times in case of spurious failures; and
- Properly handle the possibility that a chromosome does not have any bin overlapping one or more subcompartment type.

Our patches were accepted by upstream but are not yet part of a stable release. For more details, see ay-lab/dcHiC/pull/59, ay-lab/dcHiC/pull/60, ay-lab/dcHiC/pull/62, and containers/patches/dchic.patch.

Patch for cooler v0.9.1 (Abdennur and Mirny 2020)

We patched cooler to ensure that convergence of cooler balance using *cis*-only interactions was correctly reported for all chromosomes instead of just the last one.

When balancing interactions using the --*cis*-only flag, cooler balances interactions for each chromosome independently. Given that there is no guarantee that ICE will converge within the given number of iterations, cooler balance stores an attribute inside balanced cooler files to report whether balancing was successful (i.e., convergence was achieved).

In cooler v0.9.1, there is a bug in the logic that writes this attribute that causes the convergence for the last chromosome balanced to be reported instead of the convergence status for each chromosome. Our patch addresses this bug such that the convergence status of matrices balanced with --*cis*-only can be assessed correctly.

The patch also addressed some minor issues related to detecting pandas' version at runtime and handling of bin sizes represented using unsigned integers.

Our patches were accepted by upstream and are now part of cooler v0.9.2.

For more details, see open2c/cooler/pull/313, open2c/cooler/pull/323, open2c/cooler/pull/324 and containers/patches/cooler.patch.

Patch for Chrom3D v1.0.2 (Paulsen et al. 2017)

We patched Chrom3D to support building the project using CMake and addressing a few compiler warnings and errors that were preventing us from compiling Chrom3D with a modern C+

+ compiler toolchain. For more details, refer to the following two patch files available at robomics/chrom3d-nf (doi.org/10.5281/zenodo.12687842), chrom3d-cmake.patch, and chrom3d-fix-warnings.patch.

Patch for HiNT v2.2.8 (Wang et al. 2020)

We patched HiNT to support processing Hi-C matrices using Arima restriction enzymes. For more details, see containers/patches/hint.patch.

Data access

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession numbers GSE246689, GSE246599, GSE247171, and GSE246947. All raw and processed microscopy data generated in this study have been submitted to Zenodo (https://zenodo.org/records/15167335).

Competing interest statement

The authors declare no competing interests.

Acknowledgments

Molecular graphics and analyses were performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases. HiC and RNA-seq library constructions and Illumina sequencing were done at the Biomolecular Resource Facility (BRF), JCSMR, Australian National University. Some of the analyses were performed on resources provided by Sigma2–the National Infrastructure for High Performance Computing and Data Storage in Norway, with account number NN8041K. We thank Anita L. Sørensen for performing FISH experiments and microscopy imaging. Lenti-multi-CRISPR was a gift from Qin Yan, and hCas9 was a gift from George Church (Addgene plasmids no. 85402 and 41815, respectively; http://n2t.net/addgene:41815; http://n2t.net/addgene:85402). This work was funded by the Norwegian Research Council projects no. 324137 and no. 343102 (J.P.) and the Centres of Excellence funding scheme 262652–CanCell (R.E.), the Norwegian Cancer Society (P.C.), The Norwegian Cancer Society 272510 (R.E.), UNIFOR (P.C. and A.B.), and National Health and Medical Research Council ID 1182759 (D.T.).

Author contributions: J.P., P.C., and D.T. conceived and designed the study; J.P., R.R., and S.O. designed data analyses; R.R. and S.O. analyzed data; R.R. processed Hi-C, RNA-seq, ChIP-seq, and FISH data and did downstream analyses; S.O. performed 3D genome modeling; M.N. did Hi-C experiments; R.D. and Y.D. did cell culture work and RNA-seq; A.B. and M.A. analyzed LADs; R.E. conceived and designed CRISPR-Cas9 RE strategy, verified clones, analyzed data, and generated figures; M.L. performed cloning of Lenti-multi-CRISPR constructs and PCR; R.A. did CRISPR-Cas9 time-course, RNA and genomic extraction, qRT-PCR, PCR, and analysis; J.P., P.C., D.T., and R.E. supervised the work. All authors read and approved the final manuscript.

References

Abdennur N, Mirny LA. 2020. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36:** 311–316. doi:10.1093/bioinformatics/btz540

Akdemir KC, Le VT, Chandran S, Li Y, Verhaak RG, Beroukhim R, Campbell PJ, Chin L, Dixon JR, Futreal PA, et al. 2023. Author correction: disruption of chromatin folding domains by somatic genomic rearrangements in human cancer. *Nat Genet* **55:** 1079. doi:10.1038/s41588-023-01318-w

Arnould C, Rocher V, Saur F, Bader AS, Muzzopappa F, Collins S, Lesage E, Le Bozec B, Puget N, Clouaire T, et al. 2023. Chromatin compartmentalization regulates the response to DNA damage. *Nature* **623:** 183–192. doi:10.1038/s41586-023-06635-y

Ay F, Noble WS. 2015. Analysis methods for studying the 3D architecture of the genome. *Genome Biol* **16:** 183. doi:10.1186/s13059-015-0745-7

Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, Marth GT. 2011. BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* **27:** 1691–1692. doi:10.1093/bioinformatics/btr174

Barutcu AR, Lajoie BR, McCord RP, Tye CE, Hong D, Messier TL, Browne G, van Wijnen AJ, Lian JB, Stein JL, et al. 2015. Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells. *Genome Biol* **16:** 214. doi:10.1186/s13059-015-0768-0

Berlivet S, Paquette D, Dumouchel A, Langlais D, Dostie J, Kmita M. 2013. Clustering of tissue-specific sub-TADs accompanies the regulation of *HoxA* genes in developing limbs. *PLoS Genet* **9:** e1004018. doi:10.1371/journal.pgen.1004018

Berns EMJJ, Klijn JGM, van Putten WLJ, van Staveren IL, Portengen H, Foekens JA. 1992. c-myc amplification is a better prognostic factor than HER2/neu amplification in primary breast cancer. *Cancer Res* **52:** 1107–1113.

Bobbitt JR, Seachrist DD, Keri RA. 2023. Chromatin organization and transcriptional programming of breast cancer cell identity. *Endocrinology* **164:** bqad100. doi:10.1210/endocr/bqad100

Boninsegna L, Yildirim A, Polles G, Zhan Y, Quinodoz SA, Finn EH, Guttman M, Zhou XJ, Alber F. 2022. Integrative genome modeling platform reveals essentiality of rare contact events in 3D genome organizations. *Nat Methods* **19:** 938–949. doi:10.1038/s41592-022-01527-x

Brinkman EK, Chen T, Amendola M, van Steensel B. 2014. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res* **42:** e168. doi:10.1093/nar/gku936

Bron C, Kerbosch J. 1973. Algorithm 457: finding all cliques of an undirected graph. *Commun ACM* **16:** 575–577. doi:10.1145/362342.362367

Brunette GJ, Tourdot RW, Wangsa D, Pellman D, Zhang C-Z. 2024. Haplotype-resolved karyotype construction from Hi-C data using refLinker. bioRxiv doi:10.1101/2024.03.02.583108

Cao J, Wu L, Zhang S-M, Lu M, Cheung WKC, Cai W, Gale M, Xu Q, Yan Q. 2016. An easy and efficient inducible CRISPR/Cas9 platform with improved specificity for multiple gene targeting. *Nucleic Acids Res* **44:** e149. doi:10.1093/nar/gkw660

Chakraborty A, Wang JG, Ay F. 2022. dcHiC detects differential compartments across multiple Hi-C datasets. *Nat Commun* **13:** 6827. doi:10.1038/s41467-022-34626-6

Corzo C, Corominas JM, Tusquets I, Salido M, Bellet M, Fabregat X, Serrano S, Solé F. 2006. The MYC oncogene in breast cancer progression: from benign epithelium to invasive carcinoma. *Cancer Genet Cytogenet* **165:** 151–156. doi:10.1016/j.cancergencyto.2005.08.013

Cremer T, Cremer M. 2010. Chromosome territories. *Cold Spring Harb Perspect Biol* **2:** a003889. doi:10.1101/cshperspect.a003889

Crosetto N, Bienko M. 2020. Radial organization in the mammalian nucleus. *Front Genet* **11:** 33. doi:10.3389/fgene.2020.00033

Crusoe MR, Alameldin HF, Awad S, Boucher E, Caldwell A, Cartwright R, Charbonneau A, Constantinides B, Edvenson G, Fay S, et al. 2015. The khmer software package: enabling efficient nucleotide sequence analysis. *F1000Res* **4:** 900. doi:10.12688/f1000research.6924.1

Daley T, Smith AD. 2013. Predicting the molecular complexity of sequencing libraries. *Nat Methods* **10:** 325–327. doi:10.1038/nmeth.2375

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* **10:** giab008. doi:10.1093/gigascience/giab008

da Veiga Leprevost F, Grüning BA, Alves Aflitos S, Röst HL, Uszkoreit J, Barsnes H, Vaudel M, Moreno P, Gatto L, Weber J, et al. 2017. BioContainers: an open-source and community-driven framework for software standardization. *Bioinformatics* **33:** 2580–2582. doi:10.1093/bioinformatics/btx192

Dawson PJ, Wolman SR, Tait L, Heppner GH, Miller FR. 1996. MCF10AT: a model for the evolution of cancer from proliferative breast disease. *Am J Pathol* **148:** 313–319.

Diao Y, Li B, Meng Z, Jung I, Lee AY, Dixon J, Maliskova L, Guan K-L, Shen Y, Ren B. 2016. A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening. *Genome Res* **26:** 397–405. doi:10.1101/gr.197152.115

Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. 2017. Nextflow enables reproducible computational workflows. *Nat Biotechnol* **35:** 316–319. doi:10.1038/nbt.3820

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485:** 376–380. doi:10.1038/nature11082

Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, Ye Z, Kim A, Rajagopal N, Xie W, et al. 2015. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518:** 331–336. doi:10.1038/nature14222

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29:** 15–21. doi:10.1093/bioinformatics/bts635

Döring A, Weese D, Rausch T, Reinert K. 2008. SeqAn an efficient, generic C++ library for sequence analysis. *BMC Bioinformatics* **9:** 11. doi:10.1186/1471-2105-9-11

Dozmorov MG, Marshall MA, Rashid NS, Grible JM, Valentine A, Olex AL, Murthy K, Chakraborty A, Reyna J, Figueroa DS, et al. 2023. Publisher correction: rewiring of the 3D genome during acquisition of carboplatin resistance in a triple-negative breast cancer patient-derived xenograft. *Sci Rep* **13:** 7522. doi:10.1038/s41598-023-33284-y

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst* **3:** 95–98. doi:10.1016/j.cels.2016.07.002

Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32:** 3047–3048. doi:10.1093/bioinformatics/btw354

Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, Garcia MU, Di Tommaso P, Nahnsen S. 2020. The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol* **38:** 276–278. doi:10.1038/s41587-020-0439-x

Feng Y, Pauklin S. 2020. Revisiting 3D chromatin architecture in cancer development and progression. *Nucleic Acids Res* **48:** 10632–10647. doi:10.1093/nar/gkaa747

Fitzgibbon AW, Fisher RB. 1995. A buyer's guide to conic fitting. In *Proceedings of the British Machine Conference*, pp. 51.1–51.10. BMVA Press, Durham, UK. doi:10.5244/C.9.51. http://www.bmva.org/bmvc/1995/bmvc-95-050.html.

Forsberg F, Brunet A, Ali TML, Collas P. 2019. Interplay of lamin A and lamin B LADs on the radial positioning of chromatin. *Nucleus* **10:** 7–20. doi:10.1080/19491034.2019.1570810

Frankish A, Diekhans M, Ferreira A-M, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J, et al. 2019. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* **47:** D766–D773. doi:10.1093/nar/gky955

Gao T, Qian J. 2020. EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. *Nucleic Acids Res* **48:** D58–D64. doi:10.1093/nar/gkz980

Garg S. 2023. Towards routine chromosome-scale haplotype-resolved reconstruction in cancer genomics. *Nat Commun* **14:** 1358. doi:10.1038/s41467-023-36689-5

Gonzalez-Sandoval A, Gasser SM. 2016. On TADs and LADs: spatial control over gene expression. *Trends Genet* **32:** 485–495. doi:10.1016/j.tig.2016.05.004

Grisanzio C, Freedman ML. 2010. Chromosome 8q24-associated cancers and MYC. *Genes Cancer* **1:** 555–559. doi:10.1177/1947601910381380

Grüning B, Dale R, Sjödin A, Chapman BA, Rowe J, Tomkins-Tinch CH, Valieris R, Köster J, Bioconda Team. 2018. Bioconda: sustainable and comprehensive software distribution for the life sciences. *Nat Methods* **15:** 475–476. doi:10.1038/s41592-018-0046-7

Haeussler M, Schönig K, Eckert H, Eschstruth A, Mianné J, Renaud J-B, Schneider-Maunoury S, Shkumatava A, Teboul L, Kent J, et al. 2016. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol* **17:** 148. doi:10.1186/s13059-016-1012-2

Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, et al. 2020. Array programming with NumPy. *Nature* **585:** 357–362. doi:10.1038/s41586-020-2649-2

Harris HL, Gu H, Olshansky M, Wang A, Farabella I, Eliaz Y, Kalluchi A, Krishna A, Jacobs M, Cauer G, et al. 2023. Chromatin alternates between A and B compartments at kilobase scale for subgenic organization. *Nat Commun* **14:** 3303. doi:10.1038/s41467-023-38429-1

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38:** 576–589. doi:10.1016/j.molcel.2010.05.004

Hoff DDV, Needham-VanDevanter DR, Yucel J, Windle BE, Wahl GM. 1988. Amplified human MYC oncogenes localized to replicating submicroscopic circular DNA molecules. *Proc Natl Acad Sci* **85:** 4804–4808. doi:10.1073/pnas.85.13.4804

Hsieh T-HS, Cattoglio C, Slobodyanyuk E, Hansen AS, Rando OJ, Tjian R, Darzacq X. 2020. Resolving the 3D landscape of transcription-linked mammalian chromatin folding. *Mol Cell* **78:** 539–553.e8. doi:10.1016/j.molcel.2020.03.002

Hung KL, Yost KE, Xie L, Shi Q, Helmsauer K, Luebeck J, Schöpflin R, Lange JT, Chamorro González R, Weiser NE, et al. 2021. ecDNA hubs drive cooperative intermolecular oncogene expression. *Nature* **600:** 731–736. doi:10.1038/s41586-021-04116-8

Hung KL, Luebeck J, Dehkordi SR, Colón CI, Li R, Wong IT-L, Coruh C, Dharanipragada P, Lomeli SH, Weiser NE, et al. 2022. Targeted profiling of human extrachromosomal DNA by CRISPR-CATCH. *Nat Genet* **54:** 1746–1754. doi:10.1038/s41588-022-01190-0

Hunter JD. 2007. Matplotlib: a 2D graphics environment. *Comput Sci Eng* **9:** 90–95. doi:10.1109/MCSE.2007.55

Ibrahim DM, Mundlos S. 2020. Three-dimensional chromatin in disease: What holds us together and what drives us apart? *Curr Opin Cell Biol* **64:** 1–9. doi:10.1016/j.ceb.2020.01.003

International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921. doi:10.1038/35057062

Johnstone SE, Reyes A, Qi Y, Adriaens C, Hegazi E, Pelka K, Chen JH, Zou LS, Drier Y, Hecht V, et al. 2020. Large-scale topological changes restrain malignant progression in colorectal cancer. *Cell* **182:** 1474–1489.e23. doi:10.1016/j.cell.2020.07.030

Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. 2010. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26:** 2204–2207. doi:10.1093/bioinformatics/btq351

Kim T, Han S, Chun Y, Yang H, Min H, Jeon SY, Kim J-I, Moon H-G, Lee D. 2022. Comparative characterization of 3D chromatin organization in triple-negative breast cancers. *Exp Mol Med* **54:** 585–600. doi:10.1038/s12276-022-00768-2

Kopylova E, Noé L, Touzet H. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28:** 3211–3217. doi:10.1093/bioinformatics/bts611

Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. 2019. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol* **20:** 278. doi:10.1186/s13059-019-1910-1

Kurtzer GM, Sochat V, Bauer MW. 2017. Singularity: scientific containers for mobility of compute. *PLoS One* **12:** e0177459. doi:10.1371/journal.pone.0177459

Lajoie BR, Dekker J, Kaplan N. 2015. The hitchhiker's guide to Hi-C analysis: practical guidelines. *Methods* **72:** 65–75. doi:10.1016/j.ymeth.2014.10.031

Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, et al. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* **22:** 1813–1831. doi:10.1101/gr.136184.111

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357–359. doi:10.1038/nmeth.1923

Lee EYHP, Muller WJ. 2010. Oncogenes and tumor suppressor genes. *Cold Spring Harb Perspect Biol* **2:** a003236. doi:10.1101/cshperspect.a003236

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25:** 1754–1760. doi:10.1093/bioinformatics/btp324

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25:** 2078–2079. doi:10.1093/bioinformatics/btp352

Li Q, Tjong H, Li X, Gong K, Zhou XJ, Chiolo I, Alber F. 2017. The three-dimensional genome organization of *Drosophila* melanogaster through data integration. *Genome Biol* **18:** 145. doi:10.1186/s13059-017-1264-5

Li Y, Xu W, Wang Y, Kou J, Zhang J, Hu S, Zhang L, Wang J, Liu J, Liu H, et al. 2022. An improved, chromosome-level genome of the giant panda (*Ailuropoda melanoleuca*). *Genomics* **114:** 110501. doi:10.1016/j.ygeno.2022.110501

Liao DJ, Dickson RB. 2000. c-Myc in breast cancer. *Endocr Relat Cancer* **7:** 143–164. doi:10.1677/erc.0.0070143

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30:** 923–930. doi:10.1093/bioinformatics/btt656

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326:** 289–293. doi:10.1126/science.1181369

Lin D, Sanders J, Noble WS. 2021. HiCRep.py: fast comparison of Hi-C contact matrices in Python. *Bioinformatics* **37:** 2996–2997. doi:10.1093/bioinformatics/btab097

Liu Y, Nanni L, Sungalee S, Zufferey M, Tavernari D, Mina M, Ceri S, Oricchio E, Ciriello G. 2021. Systematic inference and comparison of multi-scale chromatin sub-compartments connects spatial organization to cell phenotypes. *Nat Commun* **12:** 2439. doi:10.1038/s41467-021-22666-3

Lopez-Delisle L, Rabbani L, Wolff J, Bhardwaj V, Backofen R, Grüning B, Ramírez F, Manke T. 2021. pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* **37:** 422–423. doi:10.1093/bioinformatics/btaa692

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15:** 550. doi:10.1186/s13059-014-0550-8

Love MI, Soneson C, Hickey PF, Johnson LK, Pierce NT, Shepherd L, Morgan M, Patro R. 2020. Tximeta: reference sequence checksums for provenance identification in RNA-seq. *PLoS Comput Biol* **16:** e1007664. doi:10.1371/journal.pcbi.1007664

Lund E, Oldenburg AR, Collas P. 2014. Enriched domain detector: a program for detection of wide genomic enrichment domains robust against local variations. *Nucleic Acids Res* **42:** e92. doi:10.1093/nar/gku324

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. 2013. RNA-guided human genome engineering via Cas9. *Science* **339:** 823–826. doi:10.1126/science.1232033

McInnes L, Healy J. 2017. Accelerated Hierarchical Density Based Clustering. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, New Orleans, pp. 33–42. IEEE, Piscataway, NJ.

McInnes L, Healy J, Astels S. 2017. hdbscan: hierarchical density based clustering. *J Open Source Softw* **2:** 205. doi:10.21105/joss.00205

McKinney W. 2010. Data Structures for Statistical Computing in Python. In *Proceedings of the Ninth Python in Science Conference, SciPy*, Austin, TX. https://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf (accessed November 8, 2023).

Misteli T. 2004. Spatial positioning; a new dimension in genome function. *Cell* **119:** 153–156. doi:10.1016/j.cell.2004.09.035

Morton AR, Dogan-Artun N, Faber ZJ, MacLeod G, Bartels CF, Piazza MS, Allan KC, Mack SC, Wang X, Gimple RC, et al. 2019. Functional enhancers shape extrachromosomal oncogene amplifications. *Cell* **179:** 1330–1341.e13. doi:10.1016/j.cell.2019.10.039

Nagai LAE, Park S-J, Nakai K. 2019. Analyzing the 3D chromatin organization coordinating with gene expression regulation in B-cell lymphoma. *BMC Med Genomics* **11:** 127. doi:10.1186/s12920-018-0437-8

Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, et al. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485:** 381–385. doi:10.1038/nature11049

Nora EP, Goloborodko A, Valton A-L, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA, Bruneau BG. 2017. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169:** 930–944.e22. doi:10.1016/j.cell.2017.05.004

Okonechnikov K, Conesa A, García-Alcalde F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32:** 292–294. doi:10.1093/bioinformatics/btv566

Open2C, Abdennur N, Fudenberg G, Flyamer I, Galitsyna AA, Goloborodko A, Imakaev M, Venev SV. 2024. Bioframe: operations on genomic intervals in *Pandas* dataframes. *Bioinformatics* **40:** btae088. doi:10.1093/bioinformatics/btae088

Osborne CS, Chakalova L, Mitchell JA, Horton A, Wood AL, Bolland DJ, Corcoran AE, Fraser P. 2007. Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS Biol* **5:** e192. doi:10.1371/journal.pbio.0050192

Osman N, Shawky A-E-M, Brylinski M. 2022. Exploring the effects of genetic variation on gene regulation in cancer in the context of 3D genome structure. *BMC Genom Data* **23:** 13. doi:10.1186/s12863-021-01021-x

Osterwalder M, Barozzi I, Tissières V, Fukuda-Yuzawa Y, Mannion BJ, Afzal SY, Lee EA, Zhu Y, Plajzer-Frick I, Pickle CS, et al. 2018. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554:** 239–243. doi:10.1038/nature25461

Otsu N. 1979. A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern* **9:** 62–66. doi:10.1109/TSMC.1979.4310076

Park DS, Nguyen SC, Isenhart R, Shah PP, Kim W, Barnett RJ, Chandra A, Luppino JM, Harke J, Wai M, et al. 2023. High-throughput oligopaint screen identifies druggable 3D genome regulators. *Nature* **620:** 209–217. doi:10.1038/s41586-023-06340-w

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14:** 417–419. doi:10.1038/nmeth.4197

Paulsen J, Sekelja M, Oldenburg AR, Barateau A, Briand N, Delbarre E, Shah A, Sørensen AL, Vigouroux C, Buendia B, et al. 2017. Chrom3D: three-dimensional genome modeling from Hi-C and nuclear lamin-genome contacts. *Genome Biol* **18:** 21. doi:10.1186/s13059-016-1146-2

Paulsen J, Liyakat Ali TM, Collas P. 2018. Computational 3D genome modeling using Chrom3D. *Nat Protoc* **13:** 1137–1152. doi:10.1038/nprot.2018.009

Paulsen J, Liyakat Ali TM, Nekrasov M, Delbarre E, Baudement M-O, Kurscheid S, Tremethick D, Collas P. 2019. Long-range interactions between topologically associating domains shape the four-dimensional genome during differentiation. *Nat Genet* **51:** 835–843. doi:10.1038/s41588-019-0392-0

Pei L, Huang X, Liu Z, Tian X, You J, Li J, Fang DD, Lindsey K, Zhu L, Zhang X, et al. 2022. Dynamic 3D genome architecture of cotton fiber reveals subgenome-coordinated chromatin topology for four-staged single-cell differentiation. *Genome Biol* **23:** 45. doi:10.1186/s13059-022-02616-y

Perez G, Barber GP, Benet-Pages A, Casper J, Clawson H, Diekhans M, Fischer C, Gonzalez JN, Hinrichs AS, Lee CM, et al. 2025. The UCSC Genome Browser database: 2025 update. *Nucleic Acids Res* **53:** D1243–D1249. doi:10.1093/nar/gkae974

Pertea G, Pertea M. 2020. GFF utilities: GffRead and GffCompare. *F1000Res* **9:** 304. doi:10.12688/f1000research.23297.2

Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. 2021. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci* **30:** 70–82. doi:10.1002/pro.3943

Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29:** e45. doi:10.1093/nar/29.9.e45

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26:** 841–842. doi:10.1093/bioinformatics/btq033

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44:** W160–W165. doi:10.1093/nar/gkw257

Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, Habermann B, Akhtar A, Manke T. 2018. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun* **9:** 189. doi:10.1038/s41467-017-02525-w

Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159:** 1665–1680. doi:10.1016/j.cell.2014.11.021

Rao SSP, Huang S-C, Glenn St Hilaire B, Engreitz JM, Perez EM, Kieffer-Kwon K-R, Sanborn AL, Johnstone SE, Bascom GD, Bochkov ID, et al. 2017. Cohesin loss eliminates all loop domains. *Cell* **171:** 305–320.e24. doi:10.1016/j.cell.2017.09.026

Rasmussen R. 2001. Quantification on the lightCycler. In *Rapid cycle real-time PCR* (ed. Meuer S, et al.), pp. 21–34. Springer, Berlin.

R Core Team. 2024. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/.

Rønningen T, Shah A, Oldenburg AR, Vekterud K, Delbarre E, Moskaug JØ, Collas P. 2015. Prepatterning of differentiation-driven nuclear lamin A/C-associated chromatin domains by GlcNAcylated histone H2B. *Genome Res* **25:** 1825–1835. doi:10.1101/gr.193748.115

Rossini R, Paulsen J. 2024. hictk: blazing fast toolkit to work with .hic and .cool files. *Bioinformatics* **40:** btae408. doi:10.1093/bioinformatics/btae408

Rowley MJ, Corces VG. 2018. Organizational principles of 3D genome architecture. *Nat Rev Genet* **19:** 789–800. doi:10.1038/s41576-018-0060-8

Santner SJ, Dawson PJ, Tait L, Soule HD, Eliason J, Mohamed AN, Wolman SR, Heppner GH, Miller FR. 2001. Malignant MCF10CA1 cell lines derived from premalignant human breast epithelial MCF10AT cells. *Breast Cancer Res Treat* **65:** 101–110. doi:10.1023/A:1006461422273

Sayols S, Scherzinger D, Klein H. 2016. dupRadar: a Bioconductor package for the assessment of PCR artifacts in RNA-Seq data. *BMC Bioinformatics* **17:** 428. doi:10.1186/s12859-016-1276-2

Schriml LM, Lichenstein R, Bisordi K, Bearer C, Baron JA, Greene C. 2023. Modeling the enigma of complex disease etiology. *J Transl Med* **21:** 148. doi:10.1186/s12967-023-03987-x

Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, Fonseca NA, Huber W, Haering CH, Mirny L, et al. 2017. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551:** 51–56. doi:10.1038/nature24281

Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-J, Vert J-P, Heard E, Dekker J, Barillot E. 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16:** 259. doi:10.1186/s13059-015-0831-x

Servant N, Varoquaux N, Heard E, Barillot E, Vert J-P. 2018. Effective normalization for copy number variation in Hi-C data. *BMC Bioinformatics* **19:** 313. doi:10.1186/s12859-018-2256-5

Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G. 2012. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148:** 458–472. doi:10.1016/j.cell.2012.01.010

Shen W, Le S, Li Y, Hu F. 2016. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* **11:** e0163962. doi:10.1371/journal.pone.0163962

Sood V, Misteli T. 2022. The stochastic nature of genome organization and function. *Curr Opin Genet Dev* **72:** 45–52. doi:10.1016/j.gde.2021.10.004

Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, et al. 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci* **98:** 10869–10874. doi:10.1073/pnas.191367098

Splinter E, de Wit E, van de Werken HJG, Klous P, de Laat W. 2012. Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. *Methods* **58:** 221–230. doi:10.1016/j.ymeth.2012.04.009

Spracklin G, Abdennur N, Imakaev M, Chowdhury N, Pradhan S, Mirny LA, Dekker J. 2023. Diverse silent chromatin states modulate genome compartmentalization and loop extrusion barriers. *Nat Struct Mol Biol* **30:** 38–51. doi:10.1038/s41594-022-00892-7

van der Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD, Yager N, Gouillart E, Yu T, scikit-image contributors. 2014. scikit-image: image processing in Python. *PeerJ* **2:** e453. doi:10.7717/peerj.453

van Steensel B, Belmont AS. 2017. Lamina-associated domains: links with chromosome architecture, heterochromatin, and gene repression. *Cell* **169:** 780–791. doi:10.1016/j.cell.2017.04.022

Vilarrasa-Blasi R, Soler-Vila P, Verdaguer-Dot N, Russiñol N, Di Stefano M, Chapaprieta V, Clot G, Farabella I, Cuscó P, Kulis M, et al. 2021. Dynamics of genome architecture and chromatin function during human B cell differentiation and neoplastic transformation. *Nat Commun* **12:** 651. doi:10.1038/s41467-020-20849-y

Villanueva RAM, Chen ZJ. 2019. ggplot2: elegant graphics for data analysis (2nd ed.). *Measurement* **17:** 160–167. doi:10.1080/15366367.2019.1565254

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, et al. 2020. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nat Methods* **17:** 261–272. doi:10.1038/s41592-019-0686-2

Visscher DW, Wallis T, Awussah S, Mohamed A, Crissman JD. 1997. Evaluation of *MYC* and chromosome 8 copy number in breast carcinoma by interphase cytogenetics. *Genes Chromosomes Cancer* **18:** 1–7. doi:10.1002/(SICI)1098-2264(199701)18:1<1::AID-GCC1>3.0.CO;2-7

Wang L, Wang S, Li W. 2012. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28:** 2184–2185. doi:10.1093/bioinformatics/bts356

Wang S, Lee S, Chu C, Jain D, Kerpedjiev P, Nelson GM, Walsh JM, Alver BH, Park PJ. 2020. HiNT: a computational method for detecting copy number variations and translocations from Hi-C data. *Genome Biol* **21:** 73. doi:10.1186/s13059-020-01986-5

Wang X, Xu J, Zhang B, Hou Y, Song F, Lyu H, Yue F. 2021. Genome-wide detection of enhancer-hijacking events from chromatin interaction data in rearranged genomes. *Nat Methods* **18:** 661–668. doi:10.1038/s41592-021-01164-w

Wang X, Yan J, Ye Z, Zhang Z, Wang S, Hao S, Shen B, Wei G. 2022. Reorganization of 3D chromatin architecture in doxorubicin-resistant breast cancer cells. *Front Cell Dev Biol* **10:** 974750. doi:10.3389/fcell.2022.974750

Waskom M. 2021. Seaborn: statistical data visualization. *J Open Source Softw* **6:** 3021. doi:10.21105/joss.03021

Wolff J, Bhardwaj V, Nothjunge S, Richard G, Renschler G, Gilsbach R, Manke T, Backofen R, Ramírez F, Grüning BA. 2018. Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Res* **46:** W11–W16. doi:10.1093/nar/gky504

Wolff J, Rabbani L, Gilsbach R, Richard G, Manke T, Backofen R, Grüning BA. 2020. Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. *Nucleic Acids Res* **48:** W177–W184. doi:10.1093/nar/gkaa220

Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, et al. 2021. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2:** 100141. doi:10.1016/j.xinn.2021.100141

Yang T, Zhang F, Yardımcı GG, Song F, Hardison RC, Noble WS, Yue F, Li Q. 2017. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res* **27:** 1939–1949. doi:10.1101/gr.220640.117

Yu G, Wang L-G, Han Y, He Q-Y. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16:** 284–287. doi:10.1089/omi.2011.0118

Yu G, Wang L-G, Yan G-R, He Q-Y. 2015. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* **31:** 608–609. doi:10.1093/bioinformatics/btu684

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9:** R137. doi:10.1186/gb-2008-9-9-r137

Zhao Y, Ding Y, He L, Zhou Q, Chen X, Li Y, Alfonsi MV, Wu Z, Sun H, Wang H. 2023. Multiscale 3D genome reorganization during skeletal muscle stem cell lineage progression and aging. *Sci Adv* **9:** eabo1360. doi:10.1126/sciadv.abo1360

Zhou Y, Gerrard DL, Wang J, Li T, Yang Y, Fritz AJ, Rajendran M, Fu X, Stein G, Schiff R, et al. 2019. Temporal dynamic reorganization of 3D chromatin architecture in hormone-induced breast cancer and endocrine resistance. *Nat Commun* **10:** 1522. doi:10.1038/s41467-019-09320-9

Zhu A, Ibrahim JG, Love MI. 2019. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **35:** 2084–2092. doi:10.1093/bioinformatics/bty895

Zhu Q, Zhao X, Zhang Y, Li Y, Liu S, Han J, Sun Z, Wang C, Deng D, Wang S, et al. 2023. Single cell multi-omics reveal intra-cell-line heterogeneity across human cancer cell lines. *Nat Commun* **14:** 8170. doi:10.1038/s41467-023-43991-9

Received April 14, 2025; accepted in revised form October 7, 2025.