Genome reorganization and its functional impact during breast cancer progression

Kathleen S. Metz Reed[1], Andrew Fritz[2], Haley Greenyer[2], Kerstin Heselmeyer-Haddad[3], Seth Frietze[4], Janet Stein[2], Gary Stein[2], Tom Misteli[1]

[1]National Cancer Institute, National Institutes of Health, Bethesda, MD, USA
[2]Department of Biochemistry and University of Vermont Cancer Center, University of Vermont Larner College of Medicine, Burlington, VT, USA
[3]OMICS Technology Facility, Genetics Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA
[4]College of Nursing and Health Sciences, University of Vermont Cancer Center, Burlington, Vermont, USA

Correspondence to: mistelit@mail.nih.gov, Gary.Stein@med.uvm.edu

ABSTRACT

Cancer cells undergo widespread changes in epigenetic patterns that mediate cancer compromised gene expression programs during cancer progression. However, the alterations in higher-order genome organization in which these changes occur and their functional implications are less well understood. To explore how chromatin structure and epigenetic parameters of genome architecture changes during cancer progression at a fine scale and genome-wide, we generated high-resolution Micro-C contact maps in non-malignant, pre-cancerous, and metastatic MCF10 breast cancer epithelial cells. We profiled progression-associated reorganization of chromatin compartments, topologically associated domains (TADs), and chromatin loops, and also identified invariable chromatin features. We found large-scale compartmental shifts occur predominantly in early stages of cancer development, with more fine-scale structural changes in TADs and looping accumulating during the later transition to metastasis. We related these structural features to changes in gene expression, histone marks, and potential enhancers and found a large portion of differentially expressed genes physically connected to distal regulatory elements. While changes in chromatin loops were relatively rare during progression, differential loops were enriched for progression-associated genes, including those involved in proliferation, angiogenesis, and differentiation. Changes in either enhancer-promoter contacts or distal enhancer activity were accompanied by differential gene regulation, suggesting that changes in chromatin contacts are not necessary but can be sufficient for gene regulation. Similar chromatin features and differential gene expression patterns are also present in cancer cell lines and patient tissues. Together, our results demonstrate a functionally relevant connection between gene regulation and genome remodeling in a cancer progression model.

KEY FINDINGS

- In a cell-based model of breast cancer, the cancer genome is reorganized throughout cancer progression at the level of compartments, chromatin domains, and loops
- Compartmental shifts occur in early stages of cancer development, with more fine-scale structural changes accumulating during metastasis
- Chromatin domain boundaries are weakened during cancer progression
- Many progression-regulated genes exhibit changes in distal enhancer histone modifications that are bridged by stable chromatin loops
- Changes in enhancer activity or subtle changes in chromatin contacts can rewire enhancer-promoter connections to facilitate changes in gene expression
- Prominent changes in chromatin loops occur at a small subset of differentially regulated genes during cancer progression

INTRODUCTION

The eukaryotic genome is highly organized in the cell nucleus. Amongst the most prominent structural features are kilobase-sized chromatin loops, medium-scale topologically associating domains (TADs) and higher-order compartments (1–5). How chromatin organization contributes to epigenetic control of gene regulation, including in physiological and pathological settings, such as cancer, remains only partially understood (6–9).

A prominent mechanism to generate higher order chromatin structures is loop extrusion (10–12). During this process, the multi-component cohesin complex is loaded onto chromatin and, using its intrinsic molecular motor activity, extrudes chromatin bidirectionally along the genome to form a loop or a domain until it encounters the major chromatin architectural protein CTCF bound to convergently oriented binding sites. The encounter of cohesin with bound CTCF stalls the extrusion process and generates a chromatin loop or TAD. While loop extrusion has universally been implicated in formation of chromatin loops and domains, formation of chromatin features by other mechanisms have also been observed, especially at a smaller scale (13–16).

A common property of higher order chromatin folding is that the resulting loops, domains and compartments, bring distal genome elements into spatial proximity and into proximity of regulatory elements to their target genes has been implicated in gene regulation (2). For example, it has been suggested that one function of TADs is to facilitate the interaction of regulatory elements, particularly gene enhancers, with their target genes located within the same TAD (17–21). Similarly, long-range interactions via chromatin loops are thought to be essential at some loci to bring gene enhancers into proximity to their target promoters (22–26). However, regulation of many gene loci also appears independent of chromatin organization, and enhancer-promoter proximity often does not correlate with gene activity (27–30). In fact, acute depletion of cohesin revealed genome-wide disruption of chromatin organization but surprisingly limited impact on gene expression (31). A possible explanation for these divergent findings is that chromatin organization may be functionally more relevant to bring about changes in gene activity rather than maintenance of gene expression as suggested by several cohesin depletion studies (30,32,33).

Genome organization is likely relevant for cancer and its progression. Mutations in loop extrusion machinery, such as cohesin and the cohesin processivity factor NIPBL or at CTCF binding sites have been reported in many cancers (6,7,34–36). In addition, many structural variants (SVs) such as deletions, duplications, and translocations, have been documented in various cancer subtypes where SVs and the ensuing reorganization around them can lead to aberrant gene regulation (37–39).

Despite these observations, the full extent of genome reorganization during cancer progression, and its functional consequences, remains largely unknown. Several studies primarily focused on large scale reorganizations have found changes in higher-order chromatin organization such as chromosome clustering and dynamic compartments, some

of which correlated with changes in differentially expressed oncogenes and enhancers (40–48). Analysis of TADs in various cancers have found mixed results, with some studies pointing to increased TADs and gained boundaries and others observing more stable TAD organization or weakened boundaries (42,46,47,49,50). Furthermore, cancer-associated structural variants, such as chromosomal translocations, have been related to altered gene expression, for example via enhancer-hijacking (51,52). However, how local chromatin loops and TADs are restructured during oncogenic reprogramming and how these changes relate to cancer-associated gene expression has not been well documented.

We address the question of how local and global chromatin organization changes during cancer progression, and how they related to cancer gene expression programs, by generating high-resolution Micro-C maps in the well-established MCF10 breast cancer progression model (53). This cancer model consists of three epithelial cell lines that were all originally derived from the same non-cancerous patient (54). MCF10A is an adherent epithelial non-cancerous cell line that spontaneously immortalized from the initial patient sample. Pre-malignant MCF10AT1 cells were derived from MCF10A cells by overexpressing mutant Ha-Ras oncogene followed by long-term passage as mouse xenografts (55). In immunocompromised mice, MCF10AT1 cells form precancerous lesions, and approximately 25% progress to invasive carcinoma over time. The MCF10CA1a cell line is derived from metastatic tumors generated from xenografted MCF10AT1 cells and is considered fully malignant, metastatic, and highly aggressive, forming tumors in 100% of xenografted mice which can metastasize rapidly to the lung (56). The MCF10 progression series represents a spectrum of cells that share a similar genetic background but are increasingly more cancerous and they have been widely used to study the genetic and epigenetic changes that occur during cancer progression and epithelial-mesenchymal transitions (EMT) (41,57–65).

In this study, comparing fine-scale chromatin organization and other epigenetic features in the MCF10 cancer progression model has allowed us to identify changes in genome reorganization and relate them to changes in gene expression, including of cancer progression-associated genes. Our results provide novel insights into the principles of chromatin-mediated gene regulation and into the dynamic structure-function relationship contributing to genome regulation derived from an *in vitro* cancer progression model.

RESULTS

Mapping global and local genome organization across breast cancer progression

To understand how genome organization changes during cancer progression, we generated high-resolution (5 kb) genome-wide maps of chromatin contacts using Micro-C in the MCF10A, MCF10AT1, and MCF10CA1a cancer progression series (**Fig. 1A, Supp. Fig. 1A**). We obtained high-quality data with at least 1 billion Micro-C contacts per cell line, spread across 2 biological replicates with 4 technical replicates each (**Supp Table 1**).

We identified features of chromatin organization at several levels, including large-scale reorganizations of compartments, medium-scale changes in topologically associating domains (TADs), and fine-scale changes in chromatin loops (**Fig. 1B**). We assigned A and B compartments using CALDER at a resolution of 10kb, TADs using SpectralTAD combined with FAN-C boundary insulation score calculations at 10kb, and loops using SIP at 5- and 10kb resolution (see Methods for details). Each cell type had similar percentages of the genome assigned to the active A (47.1-50.2%) or inactive B (49.7-52.9%) compartment, with the two cancer cell types MCF10AT1 and MCF10CA1a having a slightly higher proportion of A compartment designations (**Fig. 1C**). Similarly, we detected a similar number of TADs (between 7,459-7,825) and chromatin loops (between 15,713-17,332) in each cell line (**Fig. 1D-E**). Although the three cell lines are karyotypically similar due to their shared genetic background, they contain large-scale chromosomal duplications and translocations which were identified by SKY karyotyping analysis (**Supp. Fig. 1B**), and numerical chromosome aberrations based on SKY and Micro-C sequencing depth analysis were included in the analysis of all chromatin features (see Methods; **Supp. Fig. 1C, Supp. Table 2**). After this correction, chromatin loop counts showed a high degree of reproducibility between technical and biological replicates (**Supp. Fig. 1D**). We used this deep dataset to characterize structural features that are reshaped during cancer progression.

Cancer progression reorganizes compartments, TADs, and chromatin loops

Comparative analysis across the three cell types identified significant changes in all major chromatin features during cancer progression (**Fig. 1, Supp. Fig. 2A-B, Supp. Table 3-5**).

At a large-scale, we detected changes in compartmentalization. We observed a general shift towards the more active A compartment in early cancer progression, with a larger portion of genomic regions becoming more A-like (31.0%) compared to more B-like (26.0%) in the transition from MCF10A to MCF10AT1, while these changes are more balanced in the later transition from MCF10AT1 to MCF10CA1a (30.0% and 30.6%, respectively). Interactions within the most A-like and B-like compartments were predominant in the pre-cancerous MCF10A cells (**Fig 1F**). However, in MCF10AT1 and MCF10CA1a, stronger interactions appear between more intermediate regions, suggesting a greater degree of intermixing that is consistent with increased compartmental heterogeneity which appears to occur early during cancer progression (**Fig. 1F**) (42,66,67).

TAD boundaries represent genomic regions where upstream and downstream sequences are partially insulated from one another, with fewer contacts between them than within (1,5,19). We detected a total of 13,231 TADs across all three cell types, with 17,097 unique boundaries. TADs detected range in size from 190kb to as large as 3.8 Mb, with a mean of 663kb and a median of 460kb (**Supp. Fig. 2C**). Assessing changes in insulation score (IS) at TAD boundaries revealed 3,392 (19.8%) boundaries where the degree of insulation changed significantly over the course of cancer progression. Because individual boundaries may be simultaneously used by multiple TADs, the total number of TADs which changed during progression is 5,084 (38.5%) (**Fig. 1G**). There are nearly three times as many boundary changes between later stages in cancer progression (1,693 differential boundaries between MCF10AT1 and MCF10CA1a) than early stages (567 between MCF10A and MCF10AT1). Interestingly, TAD boundaries that gained or lost insulation during progression showed a significant enrichment for weakened boundaries (71.2%) with far fewer boundaries exhibiting increased insulation strength as cancer progressed (28.8%; p-value 0, permutation test, **Supp. Fig. 2D**). This late-stage weakening of boundaries may reflect a more heterogeneous cell population as cancer progresses (6,68–71).

Chromatin loops are formed by two distal genomic regions that are in more frequent contact than their surrounding or intervening sequences, indicated by higher contact frequency (1,3,10,11). We found 29,205 chromatin loops across all three cell lines, ranging in size from 50 Kb to 2 Mb, with a mean of 402kb and median of 270 Kb in length (**Supp. Fig. 2C**). 77.6% of loop anchors coincided with CTCF peaks, representing 95.0% of loops with at least one anchor bound by CTCF, and CTCF-bound loops were stronger and longer than non-CTCF loops (**Supp. Fig. 2E-F**). Loop boundaries often overlapped with TAD boundaries with 52.4% of TADs consisting of loop domains across all cell lines. However, a majority (73.0%) of chromatin loops did not include TADs (**Supp. Fig. 2G)**. TADs without loop interactions at their boundaries tended to be larger, while loops without TADs can be found at all sizes but are enriched for shorter loops (**Supp Fig. 2H**).

To identify loops that changed significantly during cancer progression, we assessed changes in contact frequency among every loop in each cell type, correcting for karyotypic differences that result in differences in coverage between cell lines (see Methods). We identified 1,469 chromatin loops that change significantly over the course of cancer progression, including both weakened and strengthened contacts (**Fig. 1H**), representing 5.0% of all identified loops. Differential loops were defined by a ≥1.5 fold-change between contact frequencies in any two MCF10 cell lines, and an adjusted p-value of <0.025 when considering variation across biological and technical replicates. Unlike TADs, there was a more balanced number of changes between early (1,004 differential loops between MCF10A and MCF10AT1) and late (1,204 between MCF10AT1 and MCF10CA1a) progression stages, as well as between strengthened (679 loops, 46.2% of all differential loops) and weakened loops (790 loops, 53.8% of all differential loops). Interestingly, only a small portion (19.0%) of differential loops were accompanied by changes in CTCF binding (**Supp Fig. 2I**). Motif analysis of differential loop anchors revealed only weak motifs of various transcription factors enriched at the boundaries of gained and lost loops, although occupancy did not

appear high enough to explain most of the changes we observe (**Supp Fig. 2J**). Weakened loops were often associated with a decrease in H3K27ac, a mark of active enhancers, consistent with the notion that active enhancers can help recruit loop extrusion machinery (**Supp Fig. 2K**; see below) (72,73).

Taken together, these results demonstrate significant global changes in genome organization during cancer progression across multiple scales from chromatin compartments to loops.

Comparison of loops and TADs to other breast cancer cell lines and patient data

To assess how generalizable the static and dynamic structures detected in the progression series are to human tumors, we examined the chromatin loops and TAD boundaries from the MCF10 progression series in non-cancerous mammary epithelial cells (HMEC), five cell lines representing distinct cancer subtypes ranging from less (luminal, HER2+) to more aggressive (triple negative), as well as tissue samples from triple negative breast cancer (TNBC) patients with contralateral normal controls (**Supp. Fig. 3, Supp. Fig. 4**) (49).

We found that most loops and TAD boundaries detected in MCF10 cells had conserved signatures in each of the other cell types, with chromatin loops generally showing high observed-versus-expected contact frequencies and boundaries showing strong dips in insulation score (**Supp Fig. 3A, 4A-B**). When we compared the profiles of differential MCF10 features relative to static MCF10 features within each cell type, we found some cell-specific changes. For example, loops in the "up-early" cluster, which are weaker in MCF10A and stronger in both MCF10AT1 and MCF10CA1a, were significantly stronger than static loops among patient TNBC cells (**Supp. Fig. 3B**; p-value <= 0.05, Wilcoxon rank sum test). This loop cluster was also stronger in HER2+ (HCC1954) and TNBC A (HCC70) cell lines, and the "down-early" loop cluster is significantly weaker than static loops. Among differential TAD boundaries, those that have strong insulation in MCF10A and MCF10AT1 cells, but reduced insulation in MCF10CA1a, showed stronger insulation among all other examined cancer cell types when compared to static loop boundaries, suggesting that the boundaries that weaken in MCF10CA1a are not necessarily consistently weakened in breast cancer cell lines and tissues (**Supp. Fig. 4C**; p-value <= 0.05, Wilcoxon rank sum test). These findings highlight that different model systems indeed have distinct profiles of structural change, just as they have distinct gene expression profiles.

Chromatin loops connect gene promoters to distal regulatory features

We then sought to explore how long-range chromatin interactions relate to gene expression. We identified 17,185 expressed genes across any of the three MCF10 cell types using RNA-Seq and 8,840 differentially expressed genes across all pairwise comparisons in the MCF10 cancer progression (see Methods; **Supp Fig. 5A-B, Supp. Table 6**). A similar number of genes changed in later stages of cancer development (4,968 between MCF10AT1 and MCF10CA1a) compared to early progression (4,773 between MCF10A and MCF10AT1). Reassuringly, as expected from previous studies (41,64), genes associated with epithelial morphogenesis and cell adhesion were upregulated early during progression, whereas regulation of differentiation, tissue development, metabolism, and signal transduction genes was

observed during later stages of progression (**Supp Fig. 5C**). These changes are consistent with the development of an intermediate and diverse pre-cancerous state early on during progression, while late changes are known to facilitate metastasis and support the epithelial-to-mesenchymal-like transition observed phenotypically among the progression series (**Supp Fig. 5D)** (74–77).

To assess clinical significance of the differentially expressed genes in the MCF10 progression system, we overlapped this gene set with data from the Cancer Genome Atlas breast cancer cohort (TCGA-BRCA) (78,79). A total of 1,884 genes that are differentially expressed between any stages of the MCF10 progression series were also differentially expressed between normal and tumor samples from breast cancer patients, but only roughly half change in the same direction (see Methods; **Supp. Fig. 5E**). Interestingly, we found a higher degree of directional agreement between late changing genes (i.e. genes that change in MCF10CA1a compared to MCF10A and MCF10AT1) than early changing genes (i.e. genes that change in MCF10AT1 and MCF10CA1a compared to MCF10A). Of note, several looped genes which are differentially expressed in the MCF10 series showed an effect on overall patient survival (**Supp. Fig. 6A-D)**, and all but one of these genes exhibit significant differences in expression between normal and tumor TCGA-BRCA samples (**Supp. Fig. 6E**; two-sampled Wilcoxon test of normalized counts, see Methods).

To explore the functional role of chromatin loops, we next overlapped chromatin loops with expressed and differential gene promoters and potential regulatory regions. We identified 52,953 potential enhancers as defined by overlapping histone H3K27ac ChIP-Seq and ATAC-Seq accessibility peaks, commonly used to identify enhancer regions (see Methods) (80–82). Approximately 66.8% of chromatin loops featured some combination of active gene promoters and enhancers within 10kb of loop anchors (**Fig. 2A**). We found that chromatin loops that connect two features (either enhancers or promoters; mean length 260-311 kb) are typically shorter than those that contain only one feature (mean length 376-435 kb) or none (**Fig. 2B**; mean length 532 kb; p-value for all comparisons < 2.2e-16, Wilcoxon rank sum test), with promoter-promoter loops being the smallest on average (mean length 260 kb). Interestingly, enhancer-promoter and enhancer-enhancer loops (mean counts 7.9, 7.9, respectively) were stronger than promoter-promoter loops (mean counts 7.2) despite being longer on average, suggesting that epigenetic signatures associated with active enhancers may support stronger contacts (**Fig. 2C**; p-value for both comparisons < 1.3e-5, Wilcoxon rank sum test).

Differentially expressed genes show proximal and distal epigenetic changes at persistent chromatin loops

To understand the regulatory modes of action of genes which were differentially expressed during cancer progression, we determined if they had gained or lost the activity-associated H3K27ac mark at their promoters or at distally looped enhancers. We found that while many genes only featured a corresponding change in H3K27ac at their promoter (57.9% of upregulated and 34.1% of downregulated genes), a large percentage also showed changes in distal enhancer activity (26.5% of upregulated and 15.7% of downregulated genes),

suggesting that enhancer loops may be playing an important functional role in control of these genes (**Fig. 2D, F**). Comparing the direction of fold-change for genes and promoter H3K27ac, distal H3K27ac, or contact frequency with distal enhancers using Fisher's Exact test revealed odds ratios significantly higher than 1 for all comparisons (8.8, 2.1, and 1.2, respectively for changes between MCF10A and MCF10CA1a), but that there was a stronger association with promoters than enhancers or loop strength (**Supp. Fig. 7A-B**). This trend was similar for genes that were differentially regulated both early and late, suggesting that the role of chromatin loops is consistent across all stages of cancer progression. In support of the finding that both distal regulatory changes and changes in contact frequency appear to contribute to changes in gene expression, looped genes that are regulated similarly in MCF10 progression and patient data include both up- and downregulated genes anchored by both static (i.e. *RRM2 and FERMT2)* and dynamic chromatin loops (i.e. *INHBA, PCDH9*; **Supp. Fig. 7C-F)**.

To further assess the contribution of looped regulatory regions, we explored the degree of change in epigenetic marks and contact frequency between proximal and distal regions. Comparing the changes in acetylation at all gene promoters and distal regulatory regions revealed that upregulated genes exhibit a significant increase in H3K27ac at distally looped enhancers, as well as a significant loss of repressive H3K27me3 marks (**Fig. 2E**; p-value < 2.2e-16, one-sample Wilcoxon test). This trend is weaker for downregulated genes, which feature a closer balance of gained and lost H3K27ac at both enhancers and promoters, as well as both gained and lost H3K27me3 at distal regions (**Fig. 2G**). These results suggest that the static chromatin structures observed during the cancer progression process contribute to the control of differentially regulated genes, particularly among upregulated genes.

We observed that genes that are more strongly expressed are more likely to overlap with the anchor of a chromatin loop, ranging from roughly 33.4% overlap for genes with 0 read counts, 47.0% for genes with 10-1,000 counts, and 56.2% for genes with 5,000 and more counts (**Supp. Fig. 8A**). To explore whether differential loops are more prominent among genes that change from a low active state to higher expression bins, we analyzed 108 genes that went from an unexpressed or "off" state (2 or fewer read counts) in one cell line to an expressed "on" state (100 or more read counts) in another (**Supp. Fig. 8B-E**)**.** While these genes were not enriched for differential loops, over 40% overlap with static loops. Similarly, genes that change from a modest "on" state to high expression levels (1000 or more read counts) are not enriched for differential loops, however they do exhibit a higher static loop overlap (61.8%) consistent with higher total gene expression levels (**Supp. Fig. 8F-I)**. For all gene sets examined, looped genes showed strong and similar trends at distal regulatory regions.

Given that only 5% of loops changed significantly during progression (see Fig. 1), it is not surprising that only a small percentage of differentially expressed genes exhibited significant changes in chromatin contacts with distal enhancers (2.1% of upregulated and 2.2% of downregulated genes; **Fig. 2D, F**). This trend was similar between both early- and late-regulated genes (**Supp. Fig. 8J**). On average there is a slight but significant change in contact

frequency between gene promoters and distal enhancers that corresponds to the change in gene expression (**Fig. 2E-F**). However, most differentially expressed genes are in regions where chromatin structure is essentially stable, reinforcing that persistent structural changes are not universally required for gene regulation.

For example, the *SPRY1* gene, which regulates cell growth and differentiation and has been shown to be upregulated in triple-negative breast cancer tumors (83), is upregulated between MCF10AT1 and MCF10CA1a, and is statically looped to distal enhancers that gain H3K27ac (**Fig. 2H**). Similarly, the *SCNN1G* gene, which encodes for a subunit of a sodium channel and is suppressed in head and neck squamous cell cancer (84), is downregulated between MCF10AT1 and MCF10CA1a, and is statically looped to distal enhancers that lose H3K27ac (**Fig. 2I**). In both cases, the contact frequency remains relatively constant despite changes in distal enhancer acetylation. Many additional examples of statically looped, differentially expressed genes were found, including *RRM2* and *FERMT2* (**Supp. Fig. 7E-F)**. Taken together, our results show that changes in gene expression are not necessarily accompanied by structural changes, and they suggest that stable chromatin loops may facilitate functionally relevant gene expression programs by providing a pre-existing structure through which differentially regulated distal enhancers can act on target genes.

Changes in TAD boundary insulation and subcompartments have subtle impacts on gene expression

To explore the impact of changes in insulation at domain boundaries on gene expression, we next examined genes within 50kb of differential boundaries. We found that genes close to weakened boundaries were not enriched for differentially expressed genes, but those near strengthened boundaries were (**Supp. Fig. 9A**; p-values of 0.141 for early and 0 for late strengthened boundaries, and p-values of 1 for weakened boundaries, permutation test). While strengthened boundaries featured both upregulated and downregulated genes, there was a small but significant correlation between the strength of changing boundaries and fold-change of expression of nearby genes when compared to genes at static boundaries, suggesting an average positive impact on gene expression (**Supp. Fig. 9B**; p-value 5.9e-3, Wilcoxon rank sum test). This suggests that changes in TAD boundary insulation have small but noticeable impacts on gene expression.

We also examined how subcompartments change genome-wide and at gene promoters. We see that between any two cell types, a majority of subcompartment changes are small changes, i.e. from A.2.2 to A.2.1 (1 step more A-like) or B.1.1 (1 step more B-like), with larger shifts being rarer (**Supp. Fig. 9C**). The promoters of active genes are enriched for A.1.1 and A.1.2 subcompartments but depleted for all B subcompartments, while inactive gene promoters closely resemble genome-wide distributions (**Supp. Fig. 9D-E**). The promoters of differentially expressed genes have similar subcompartments at lower and higher expression levels, but these changes are more drastic for genes that shift from on to off or on to high, as defined above (**Supp. Fig. 9F-H**). Differentially expressed genes with promoters that shift to more B-like by 2 or more subcompartments or more A-like by 3 or more subcompartments have significant impacts on fold-change, but smaller shifts have

minimal impacts on gene expression (**Supp. Fig. 9I**; p-value <= 0.01, Wilcoxon rank sum test). In summary, small changes in subcompartments are very common but appear to have little impact on gene expression, while larger changes correlate more strongly with changes in gene expression during breast cancer progression but are comparatively rare.

Changes in enhancer acetylation and enhancer-promoter contact are associated with changes in gene expression

To begin to distinguish the effects of enhancer-promoter contact and chromatin looping from enhancer activity effects, we compared gene expression changes at looped and non-looped enhancer-promoter pairs. To do so, we used the activity-by-contact (ABC) model to predict functional enhancer-promoter pairs. ABC combines estimates of enhancer accessibility and activity from ATAC-Seq and H3K27ac ChIP-Seq with enhancer-promoter contact frequency from Micro-C data to generate an estimate of the likelihood of functional enhancer-gene interactions (see Methods) (85). This method allowed us to identify distal regulatory regions that are functionally linked to gene promoters without specifically requiring overlap with chromatin loops. For example, an enhancer and promoter may be in high contact as measured by Micro-C because they overlap with loop anchors, or because they are at close genomic distance. For this analysis, potential enhancers were defined as any overlapping H3K27ac ChIP-Seq and ATAC-Seq peaks 750 bp or farther from the transcription start sites of the target gene.

Applying the ABC model to our data identified 150,056 potential enhancer-promoter pairs across all three cell types. Of these, 53.4% are also promoters of other genes, 23.7% are within the bodies of other genes, and 22.9% are intergenic, and range in distance from 750 to 5M base pairs away from target gene promoters (**Supp Fig. 10A**). To better understand the relationship between contact frequency, enhancer activity, and gene expression, we asked how changes in enhancer activity or contact relate to gene expression at target promoters.

Pairwise comparison of the MCF10 progression lines indicated that changes in both contact frequency and enhancer activity appear to drive changes in enhancer-promoter networks predicted by ABC **(Fig. 3)**. For example, observing potential enhancers with changes in H3K27ac between MCF10CA1a and MCF10A reveals that these enhancers also exhibit a change in contact frequency and are associated with upregulation of target genes (**Fig. 3A**). We also found that changes in contact frequency are associated with increases in H3K27ac and correlate with higher gene expression (**Fig. 3B**). These results show that not only changes in either contact frequency and enhancer activity correlate with increased gene expression, but they also correlate with each other, suggesting a potentially linked functional role during enhancer-promoter communication.

To then relate enhancer-promoter pairs to chromatin loops and to orthogonally assess whether chromatin loops are acting as a functional bridge for active enhancers, we compared looped and non-looped enhancer-promoter pairs. Enhancer-promoter pairs that have changes in distal H3K27ac and are supported by chromatin loops correlated with changes in gene regulation (**Fig. 3C**). This effect was stronger than distance-matched non-

looped enhancer-promoter pairs, but similar to contact-matched non-looped pairs, suggesting that increased contact frequency caused by loop extrusion may contribute to the stronger correlation with gene expression (**Fig. 3D-E**). Contact-matched non-looped pairs were closer on average to the looped pairs of similar contact frequency, while distance-matched non-looped pairs were in less-frequent contact than looped pairs of similar genomic distance (**Fig. 3F-G**).

Comparing the distributions of target gene fold-change for these various sets of enhancer-promoter pairs revealed several trends (**Fig. 3H**). First, pairs with significant changes in contact have a larger mean gene fold-change than pairs with significant changes in activity, suggesting that either can contribute to changes in enhancer-promoter functional pairing but that contact may have a particularly strong impact. Second, looped enhancer-promoter pairs have a comparable or larger mean gene fold-change to pairs with changes in activity or contact, suggesting again that chromatin loops may support functional enhancer-promoter pairs. Lastly, looped pairs have a similar mean gene fold-change as contact-matched pairs, which in turn have a higher mean gene fold-change than distance-matched pairs, suggesting that the increased contact frequency that chromatin loops provide to enhancer-promoter pairs may be a driving force for the functional connection. These trends hold true for all tested pairwise comparisons between cell types (**Supp. Fig. 10B-C**).

Taken together, these findings demonstrate that not all gene regulatory changes are accompanied by chromatin reorganization, but when it occurs reorganization appears to facilitate functional changes.

Changes in chromatin looping are enriched for progression-associated differentially expressed genes

We next explored how changes in chromatin looping may functionally contribute to gene regulation during cancer progression. To do so, we compared changes in chromatin loop contacts to alterations in expression of progression-associated genes. Overall, while only a small subset of gene promoters overlaps with the anchors of differential chromatin loops (507 genes, 3.0% of expressed genes), those that do are significantly enriched for genes that are differentially expressed during cancer progression based on permutation analysis (331 differentially expressed genes, 65.3% of all differentially looped genes; **Supp. Fig. 11A**).

We asked whether there was a relationship between the formation or loss of loops and differential expression of these loop-associated genes. We indeed found that differential loops are more likely to change in the same direction as the gene (i.e. increased contact frequency with distal regions associated with increased gene expression) (**Fig. 4A-B**). The fold-change of differential expressed genes which also showed differential loping were significantly higher than a random sampling of differentially expressed genes (**Fig. 4A**). For example, of 3,261 genes that were differentially upregulated between MCF10A and MCF10CA1a, loops were significantly strengthened at 98 of these genes and significantly weakened at 31. Similarly, of 3,088 downregulated genes, 65 genes overlap weakened loop anchors and 41 genes overlap strengthened loop anchors. In contrast, the number of

expected genes at strengthened or weakened loops for a random sampling of genes this size is 38 and 32, respectively (**Supp. Fig. 11B-C**). We also found a subset of chromatin loops and genes changed in opposite directions (**Fig. 4B**). The genes whose changes in expression correlate with changes in looping are enriched for several cancer-relevant pathways, such as morphogenesis, differentiation, and proliferation (**Fig. 4C**) (86).

In total, we identified 127 unique genes upregulated in areas that experience increased chromatin contacts, either at loop anchors or within existing structures. As an example, the promoter of the *COL12A1* gene, which encodes a collagen protein known to be associated with aggressive and mesenchymal phenotypes, overlaps a loop boundary that is very weak in MCF10A cells where *COL12A1* is not expressed (87–89). As *COL12A1* gene expression is upregulated during progression, contacts increase at a distal region 310 kb away, and H3K27ac and accessibility also increase at these likely distal regulatory regions **(Fig. 4D).**

Similarly, we observe 123 unique genes that are downregulated during oncogenic progression. One example is *WNT5A* which encodes for an important signaling protein and is downregulated in breast cancer and across MCF10A progression (90–92). Similar to *COL12A1*, the promoter of *WNT5A* is involved in many differential distal regulatory contacts, ranging in distance from 240 to 640 kb (**Fig. 4E**). Unlike *COL12A1*, these contacts are strongest in MCF10A cells and severely weaken in MCF10AT1 and MCF10CA1a cells. Accompanying these changes in contact, the distal regulatory regions that appear to support *WNT5A* in MCF10A cells become deacetylated and decrease in accessibility. Another example of an upregulated genes that is looped to a distal enhancer is *RASA1*, which gains contact frequency as gene expression increases, while *ZEB2* and *SDC3* are downregulated genes that lose enhancer contact frequency similar to *WNT5A* (**Supp. Fig. 12A-C**). Other loci showed the opposite behavior; for example, *TNFRSF21* and *HS3ST3A1* are upregulated genes at the anchors of weakened loops, while *NNMT* is a downregulated gene at a weakened loop anchor (**Supp. Fig. 12D-F**).

Genome reorganization occurs at cancer-relevant genes and is accompanied by proximal and distal changes in histone marks

Finally, we aimed to comprehensively explore the genes that are differentially regulated in areas that also have strong changes in chromatin looping, to better understand the potential regulatory mechanisms at play.

A locus-by-locus view of gene and loop fold-change allows us to view the relationship between changes in expression and structure among each pairwise comparison of cells (**Fig. 5A-C**). While we see that a majority of genes have a corresponding change in looping (i.e. up-regulated genes overlapping strengthened loops), we observed that the percentage of corresponding changes increases in the later stages of cancer progression. For example, the percentage of differential loop-gene pairs where the gene overlaps at least one gained loop is 47.5% and 69.9% among genes up-regulated in MCF10A compared to MCF10AT1 and MCF10CA1a, respectively, 66.7% and 79.6% among genes up-regulated in MCF10AT1 compared to MCF10A and MCF10AT1, respectively, and 70.7% and 58.0% among genes up-

regulated in MCF10CA1a compared to MCF10A and MCF10AT1, respectively. This may indicate that the regulatory impacts of changes in chromatin looping occur over longer timescales, or that genes impacted by changes in chromatin structure may be more common in metastatic cells. We did not, however, find any correlation between the magnitude of loop fold-change and gene fold-change.

To better understand how changes in looping are related to gene expression, we compared patterns of gene expression, promoter H3K27 acetylation, and distal enhancer H3K27 acetylation or trimethylation at looped genes that change in either the same or opposite direction as the loops they overlap (**Fig. 5D-E, Supp. Fig. 11D-E**). These two histone H3 modifications are mutually exclusive and have opposite effects on gene expression, marking enhancers and silencers, respectively (93). Both modifications are able to impact distal gene expression via chromatin interactions (94).

Up-regulated genes between MCF10A and MCF10CA1a showed similar epigenetic signatures at both proximal and distal regions, regardless of whether the loop they overlap gets stronger or weaker (**Fig. 5D**). Promoter regions gained H3K27ac, while distal regions gained H3K27ac and lost H3K27me3. There is however a significant difference in the extent of distal changes depending on the loop direction, with strengthened loops exhibiting a significantly higher increase in distal H3K27ac and a decrease in H3K37me3 marks. This behavior further supports the notion that these distally looped regulatory regions are important functional elements. Down-regulated genes showed distinctly different signatures (**Fig. 5E**). Genes that overlap with weakened loop anchors showed decreased promoter H3K27ac and distal H3K27ac and increased distal H3K27me3, consistent with signatures typically associated with reduced gene expression (93). Interestingly, genes that overlap strengthened loop anchors showed different patterns, with a gain in promoter H3K27ac and loss of distal H3K27me3 repressive marks. We conclude that expression of a subset of progression-associated genes is strongly correlated with loop formation. These trends are similar but weaker for genes that change between different cell types (**Supp. Fig. 11D-E**).

Taken together, our genome-wide analysis of structural and regulatory changes during MCF10A cancer progression has highlighted hundreds of restructured regions where cancer-relevant genes are differentially regulated. These findings suggest that, while relatively rare, changes in chromatin looping may be associated with regulatory changes that drive expression of hundreds of oncogenes during cancer progression.

DISCUSSION

Multiple levels of chromatin organization and integration with epigenetic parameters contribute to regulation of gene expression. We have identified dynamic chromatin organizational changes on multiple scales across breast cancer progression using the MCF10A model system. By comparing both a pre-malignant and metastatic cell line to non-cancerous epithelial cells we were able to detect both early- and late-stage changes as well as similarities in genome structure and relate them to gene expression changes, including many cancer-relevant genes.

We found that compartmental shifts occur more often in early stages of cancer development. This behavior is consistent with previous studies that have shown intermixing of chromatin A and B compartments in cancer cells (42,50). The general shift towards more active compartments during cancer progression may reflect a broader epigenetic landscape and greater heterogeneity in gene expression within cancer cell populations (6,67,95–97).

On a finer scale, structural changes in TADs occured more often during later stages of metastasis. We also found an abundance of weakened TAD boundaries in MCF10 breast cancer progression compared with boundaries that gain insulation. This finding is in line with previous studies which have shown TAD weakening in triple-negative breast cancer (49,70), but is contrary to other observations which have suggested that prostate cancer is associated with the formation of many additional TADs (46), or findings of minimal changes in TAD structure in colorectal or breast cancer (42,50). The differences in TAD changes across cancer studies may reflect differences in the types of cancer samples analyzed or the analysis methods used. For example, we detected subtle changes in insulation score at TAD boundaries in the MCF10 model, while other methods tailored towards detecting more drastic restructuring events such as the formation of TAD cliques or TAD fusions found different behaviors (50,71). Recent studies comparing chromatin structures across patient samples have also revealed a high degree of TAD and loop heterogeneity among tumors even of the same cancer type, and it is thus not surprising that patterns observed between different cancer types and in a cell culture model could vary even further (71). The weakened TAD boundaries in the MCF10 progression system are typically strong boundaries in cancer types, suggesting that the same boundaries are not weakened universally in cancer. The weakening of TAD boundaries is intriguing in that it may point to higher heterogeneity in chromatin structure in more aggressive cancer stages, possibly contributing to more extensive gene misregulation (67,70,71). Further studies using single-cell analysis in the MCF10CA1a population will be necessary to confirm if this is the case.

Comparing the structural and functional features of the MCF10 progression system to other breast cancer datasets revealed many common chromatin loops, TADs, and differentially expressed genes, as well as differences between various cancer progression models and patient data. There are hundreds of genes that change over the course of MCF10 progression that are also significantly different in breast cancer tumors from patients compared to healthy controls, including many that overlap with chromatin loops; however, this represents

only roughly a quarter of all differential genes in the MCF10 progression model. Despite this transcriptional diversity, chromatin loops and TAD boundaries detected in MCF10 cell lines showed similar average signatures in mammary epithelial cell lines and triple-negative breast cancer patients. However, cell type-specific differences do appear among differential features based on their timing and direction of change in the MCF10 progression model. For example, loops that are stronger in metastatic MCF10CA1a cells were also significantly stronger than static MCF10 loops among TNBC patients, making them a particularly interesting subset of regions to explore further.

Chromatin loops functionally connect gene promoters to distal regulatory elements. In the MCF10 model, many genes differentially regulated during cancer progression are associated with chromatin loops shared between all three cell lines, but which show changes in distal enhancer H3K27ac and H3K27me3. These trends were stronger with up-regulated genes, suggesting that we may need to explore different epigenetic signatures to better understand how chromatin structure may influence repression. We also observed that more highly expressed genes are more likely to overlap with chromatin loops. Activity-by-contact analysis showed that looped enhancer-promoter pairs also exhibit greater correlation between distal enhancer H3K27ac and gene expression than non-looped enhancers due to their increased contact frequency. These findings suggest that persistent chromatin loops that do not change during cancer progression nevertheless have functional relevance and that they do so by bridging enhancers to target gene promoters.

Activity-by-contact analysis also revealed that subtle changes in chromatin contact can contribute to the rewiring of enhancer-promoter regulatory connections. Enhancer-promoter pairs that exhibit changes in contact correlated with stronger changes in target gene expression than those with only changes in activity, in line with the concept that contact with distal regulatory elements is an important component of gene regulation (98). We also found that changes in chromatin contact are associated with more modest changes in activity, and vice-versa. This correlation between enhancer activity and enhancer-promoter contact further points to a functional link between the two. Together, these results suggest that both contact frequency and activity contribute to enhancer-promoter connectivity.

Strong changes in contact at chromatin loops are relatively rare across cancer progression, but many clear examples do exist, and they are notably enriched for differentially expressed genes. Interestingly, only a small portion of differential loops can be explained by changes in CTCF binding at anchors, suggesting other forces may be influencing their contact frequency. Loss of H3K27ac can be seen at the anchors of lost loops, consistent with the idea that active enhancers can help recruit cohesin to chromatin (72,73). Therefore, some weakened loops might be explained by a loss of H3K27ac which leads to a loss of cohesin at that region.

We also found that genes at differential loops are more likely to be differentially regulated in the same direction as the change in loop strength. This supports the notion that a subset of genes may be regulated in a structure-dependent manner (71,99,100). This interpretation is

in line with previous observations which have shown a subset of genes to be sensitive to cohesin or NIPBL depletion which disrupts chromatin loops (11,17,20,25,32,101–103). Importantly, these findings suggest this subset of structure-sensitive genes may include many with relevance to breast cancer progression.

Epigenetic signatures at gene promoters and distal regions differed based on the direction of gene change. Up-regulated genes consistently showed a gain of active-associated H3K27ac marks at both promoters and distal regions, and a loss of distal repressive H3K27me3 marks. These changes were shared between static, strengthened, and weakened loops, although up-regulated genes at strengthened loops had stronger distal changes. These findings are consistent with the loops functionally supporting interactions with distal enhancers via increase contact. However, down-regulated genes showed more complex patterns. Fewer down-regulated genes could be explained by changes in H3K27ac or H3K27me3 at the promoter or distal regions, with over half of down-regulated genes having no clear epigenetic driver, compared to only 15% of up-regulated genes. This suggests that these histone marks are not sufficient to explain down-regulated genes as well as they can explain up-regulated genes. Down-regulated genes at differential loops also showed opposite patterns based on the direction of loop change; weakened loops showed loss of distal H3K27ac and gain of H3K27me3 consistent with an inactivated enhancer, while strengthened loops showed the opposite. Additional studies will be required to fully understand the repressive mechanisms in this system and how they relate to chromatin structure.

Our study has several limitations. While the MCF10 progression system is a well-established and -characterized model, it does not reflect the complexities of in vivo tumor formation. In addition, the MCF10 cell lines used here does not fully represent all stages of cancer progression as it encompasses normal, premalignant and metastatic states, but lacks a separate malignant but non-metastatic state such as MCF10DCIS. Furthermore, based on the largely correlative nature of our analysis, we are unable to determine the causal relationships of chromatin structure changes at the loci of differentially expressed genes. Follow-up studies exploring any of the hundreds of gene loci identified here as possibly being contact-dependent could help elucidate these relationships. Additionally, as a population-level assay, Micro-C is only able to provide aggregate data across an entire population of cells. To address how cell-to-cell heterogeneity contributes to some of the functional relationships we observe, and whether that heterogeneity is occurring at a cellular or population level, we would need to apply single-cell sequencing or imaging-based approaches. These questions will be the subjects of future studies.

In summary, using the powerful MCF10 breast cancer model, we have generated a rich genomic dataset of structural and functional changes in the genome during breast cancer progression. Our data uncovers new insights into the structure function relationship in gene regulation and into the role of genome organization during malignant breast cancer progression.

METHODS

Cell culture

MCF10A cells were obtained from AATC (CRL-10317). MCF10AT1 (MCF-10AneoT) and MCF10CA1a (MCF10CA1a.cl1) cells were obtained from the Karmanos Cancer Institute. All cells were received frozen, thawed in media and grown to a confluence of 80% after 2-5 passages. Stock solutions were frozen down to be used for all experiments.

MCF10A and MCF10AT1 cells were cultured in standard high calcium medium with growth factors, consisting of DMEM/F12 (Invitrogen, 21041025) supplemented with 1.05 mM CaCl2, 10 mM HEPES, 10 ug/ml insulin (Sigma, I1882), 20 ng/ml EGF (Peprotech, AF-100-15), 0.5 ug/ml Hydrocortisone (Sigma, H0135), 100 ng/ml cholera toxin (Sigma, C8052), and 5% horse serum (Invitrogen, 16050122). MCF10CA1a cells were cultured in standard high calcium medium without growth factors, consisting of DMEM/F12, 1.05 mM CaCl2, 10 mM HEPES, and 5% horse serum.

Karyotypic analysis

Metaphase chromosomes were prepared by incubating cells with 100 ng/ml Colcemid (Roche, Brighton, MA) for two hours, followed by mitotic shake-off. The mitotic cells were then treated with a hypotonic solution (0.075M KCl) for 20 minutes at 37°C. After this treatment, the cells were centrifuged, the supernatant was extracted, and the cells were initially fixed with a methanol/acetic acid solution (3:1). This step was repeated three times. Finally, the cells were fixed onto a slide using a humidity-controlled Thermotron (Thermotron, Holland MI).

SKY probes were purchased from Applied Spectral Imaging (Carlsbad, CA) and hybridized onto slides that were aged at 37°C for 3 days. Detection was then carried out according to the protocol provided by Applied Spectral Imaging, using the following purchased antibodies: Avidin Cy5 (Rockland Immunochemicals, Limerick, PA), Mouse Anti-Digoxin antibody (Sigma-Aldrich), and a goat anti-mouse antibody conjugated to Cy5.5 (Rockland Immunochemicals). Slides were then mounted and counterstained with DAPI (Vector Laboratories, Newark, CA). Hybridization occurred over a period of two days at 37°C. Approximately 20-25 metaphases were imaged and karyotyped using the ASI GenASIS 8.2.2 software on an Olympus BX63 microscope (Evident, Tokyo, Japan) equipped with a Spectral Cube (Applied Spectral Imaging, Carlsbad, CA).

Composite karyotypes for the three cell lines, listing abnormalities found in three or more metaphases, are as follows:
MCF10A: 41-53, XX, idic(1)(q10), +1, der(3)t(3;9)(p21;p22), i(8)(q10), der(9)t(3;5;9)(p14;p23;q34;p21)
MCF10AT1: 42-49, XX, der(3)t(3;9)(p21;p22), t(3;17)(p14;p11.2), t(6;19)(p25;q12q13.3), der(9)t(3;5;9)(p14;p23;q34;p21), t(19;9)(q34;p11),+20
MCF10CA1a:43-63, XX, der(3)t(3;9)(p21;p22), +der(3)t(3;9)(p21;p22), t(3;17)(p14;p11.2), der(9)t(3;5;9)(p14;p23;q34;p21), +10

Micro-C library preparation

Cells were frozen and pellets of 1M cells were used for Micro-C library preparation. The Micro-C library was prepared using the Dovetail® Micro-C Kit according to the manufacturer's protocol. Briefly, the chromatin was fixed with 0.3M disuccinimidyl glutarate (DSG) and 37% formaldehyde in the nucleus. The cross-linked chromatin was then digested in situ with micrococcal nuclease (MNase). Following digestion, cells were lysed with 20% SDS to extract the chromatin fragments and the chromatin fragments were bound to Chromatin Capture Beads. Next, the chromatin ends were repaired and ligated to a biotinylated bridge adapter followed by proximity ligation of adapter-containing ends. After proximity ligation, the crosslinks were reversed, the associated proteins were degraded, and the DNA was purified then converted into a sequencing library using Illumina-compatible adaptors. Biotin-containing fragments were isolated using streptavidin beads prior to PCR amplification. Each library was sequenced on an Illumina Novaseq platform to generate 240-340 million 2 x 150 bp read pairs (average depth of 280M reads). Mico-C libraries were prepared by Dovetail Genomics (Scotts Valley, CA).

RNA-Seq library preparation

Total RNA was isolated from cells using Trizol (Life Technologies) and purified using the Direct-zol RNA kit (Zymo Research, Irvine, CA, USA: R2050). RNA quality and quantity were assessed using the RNA 6000 Nano Kit with the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). RNA quantity was further assessed using a Nanodrop2000 (Thermo Scientific, Lafayette, CO) and Qubit HS RNA assay (Thermo Fisher Scientific). Total RNA was depleted of ribosomal RNA (New England Biolabs, NEB #7400), reverse transcribed using a random hexamer strategy, and strand-specific adapters were added following the NEBNext Ultra II RNA-seq library prep kit (New England Biolabs, E7770). Paired-End sequencing was used to generate high depth coverage ranging from 80 to 120 million paired-end reads.

ChIP-Seq library preparation

ChIPseq for CTCF (Cell Signaling Technology, catalog number 3418) and histone marks H3K27ac (Abcam, ab4729) and H3K27Me3 (Abcam ab6002). Independent biological replicates of each cell line (MCF10A, MCF10AT1, MCF10CA1a) were performed as described, including the optional step of snap freezing of chromatin and excluding the optional third washing step (104). Additionally, the chromatin was precleared with Pierce protein A/G beads for 2 hours at 4°C prior to incubation with antibodies. For CTCF we used 20ul of antibody (Cell Signaling Technologies, 3418S) and 150ug of chromatin for each sample. For H3K27ac 10ug of chromatin was used and 2ug of antibody, while for H3K27me3 15ug of chromatin was used with 4ug of antibody.

ATAC-Seq library preparation

The OMNI-ATACseq protocol was followed as previously described, with an optimized 5 minutes of nuclear extraction rather than 3 minutes (105,106).

Micro-C processing

Micro-C data from each technical replicate (library) was processed from raw reads into contact maps using guidelines outlined by Dovetail Genomics (107). Paired reads were aligned to the hg38 human genome assembly (NCBI GRCh38) using BWA mem (version 0.7.17; settings: -5SP -T0) (108). Pairtools (version 0.3.0) was then used to parse ligations, sort reads, and remove PCR duplicates using the parse (settings: --min-mapq 40 --walks-policy 5unique --max-inter-align-gap 30), sort, and dedup (settings: --mark-dups) commands (109). Alignment files were generated using pairtools split to create .pairs and .sam files, and samtools view (version 1.17; settings: -bS) to create .bam files (110). Final .bam alignment files were sorted and indexed using samtools sort and index commands. The .bam files were used to extract QC metrics using a script from Dovetail Genomics, and calculate complexity using preseq lc_extrap (settings: -bam -pe -extrap 2.1e9 -step 1e7 -seg_len 1000000000). Pairs files were used to generate contact maps using juicer_tools pre (version 1.22.01), and normalized using SCALE (111).

For merged contact maps, we first merged pairs files using pairtools merge, and then ran juicer_tools pre command on the resulting .pairs files. In total, we generated contact maps for each library (technical replicates), each sample (biological replicates), each cell type, and we created one fully merged map including all reads for a precise map of shared features. Micro-C maps for individual technical replicates and cell types are available on GEO (accession GSE254045).

Micro-C copy number variation analysis

Structural variant and copy number variations were detected from contact maps using NeoLoopFinder calculate-cnv at 1Mb resolution (-e uniform; **Supp. Table 2**) (112). These values were averaged over the course of large-scale variations and used as correction factors for differential loop analysis (see **Differential loop, TAD, gene, and peak analysis**, below).

Regions excluded in analysis

Denylists of regions to ignore for feature calling were generated based on regions where SCALE normalization factors were unable to be calculated at 5kb or 10kb, ignoring single bins and merging continuous areas within 100 kb. Activity-by-contact analysis also combined this denylist with the ENCODE denylist for hg38 (113).

Micro-C compartment calling

Compartments were called using CALDER (version 2.0; R version 4.2.1) at 10kb resolution and SCALE normalization (114). We also used FAN-C (version 0.9.21) to calculate eigenvectors at 100kb using SCALE normalization to build saddle plots, and oriented eigenvectors manually based on overlap with active genes (115). A table of subcompartments for each cell type is available on GEO (accession GSE254045).

Micro-C topologically associated domain calling
Topologically associated domains (TADs) were called using SpectralTAD (version 1.18.0) at 10kb resolution with SCALE normalization, a window size of 300, and a level of 3, with a quality filter applied (116). We then created a unified TAD list by merging TADs that overlap at both ends between cell types, with a maximum gap of 10kb (1 pixel). We then combined this analysis with continuous insulation score (IS) calculations from FAN-C insulation command at 10kb resolution with SCALE normalization. We then used the FAN-C boundaries command to detect IS boundaries and only kept TADs that also overlapped with an IS boundary. Tables of TADs and boundaries are included in **Supp. Tables 4 and 5**. Insulation score tracks for each cell type are available on GEO (accession GSE254045).

Micro-C chromatin loop calling
Chromatin loops were called using SIP (version 1.6.2), run at 5kb and 10kb and then merged to 10kb (-g 2 -fdr 0.05). Loops were called in each cell type map using SCALE normalization, in addition to the combined map, and then merged similarly to TADs to create a unified loop list (**Supp. Table 3**) (117). Cell type loops are available on GEO (accession GSE254045).

RNA-Seq processing
All RNA-seq data was analyzed using the nf-core/rnaseq pipeline (118). Adapter and quality trimming was performed with Trim Galore (119). Reads were aligned to the hg38 reference genome using STAR (120) and gene expression was quantified using Salmon (121). Differentially expressed genes were called using DESeq2 (122). An adjusted *p*-value of 0.05 and a log2fold change of 1 were used as thresholds to select significant differential expression. A table of all differentially expressed genes and their DESeq2 metrics are listed in **Supp. Table 6**. A full table of expression data for all genes can be found on GEO (accession GSE320216).

ChIP-Seq processing and peak calling
ChIP-Seq data for CTCF, H3K27ac, and H3K27me3 was processed as detailed previously (41,123). In summary, adapters were cut (cutadapt v1.11) and low quality reads trimmed (Galaxy FASTQ Quality Trimmer 1.0.0; window 10, step 1, minimum quality 20). Reads were mapped to the human genome (hg38 canonical) using STAR version 2.4 (120) with splicing disabled (–alignIntronMax 1). Enriched regions (narrowPeak calls for CTCF and H3K27me3, broadPeak calls for H3K27ac) for each replicate were generated using MACS2 (Feng et al., 2012) and replicates were then evaluated using deepTools (124) to correlate alignments and IDR (125) to evaluate peak call reproducibility. After pooling replicates, MACS2 (126) was used to call peaks at high stringency (P-value <10e-5), these peaks were further filtered according to IDR cutoffs. ChIP-seq data is available on GEO (accession GSE98551 for CTCF, and GSE229295 for H3K27ac and H3K27me3).

ATAC-Seq processing and peak calling
Read trimming and quality filtering was performed using FastQC (127) and TrimGalore (119). The filtered fastq were then downsampled to approximately 50 million reads per

replicate. Reads were aligned to the hg38 reference genome using Bowtie2 (128). Mitochondrial, Multi-mapped, and low quality reads, and duplicates were removed using samtools (110). MACS2 (126) was used to call narrowPeaks, followed by IDR (125) to generate sample level peak sets.

Enhancer and promoter definitions

Gene promoters were defined as regions between 2000 bp upstream and 500 bp downstream of gene transcription start sites.

Potential enhancer regions were identified based on regions that contained both an ATAC-Seq and H3K27ac ChIP-Seq peak. For activity-by-contact analysis, potential enhancers were defined as 150,000 ATAC-Seq peaks with the highest levels of H3K27ac signal, but were subset for regions with H3K27ac peaks after running (see below).

Compartmental saddle plots

Saddle plots were made manually in R. We selected three chromosomes that had no major karyotypic differences between cell lines and had high correlation of eigenvectors between replicates with the same signage (chr2, chr12, chr17). For each chromosome and cell type, we sorted eigenvectors into 20 bins. We then calculated the mean observed/expected values (using SCALE normalization) between regions belonging to different bins, and plotted it as a heatmap.

Differential loop, TAD, gene, and peak analysis

Differential genes were calculated using DESeq2 (version 1.42.0) (122). Each cell type had 3 replicates, and a design of ~cellType was used. No fold-change cutoff was applied. Genes with an adjusted p-value below 0.01 were considered significant.

Differential H3K27ac within ATAC-Seq peaks were calculated using a similar design, but with a p-value cutoff of 0.05.

Differential CTCF peaks were called using the CSAW package following the workflow described in the original publication (41), and an adjusted p-value cutoff of 0.1.

Differential loops were also identified using DESeq2, but with additional adjustments. Raw, un-normalized loop counts were pulled from each technical replicate map, resulting in 8 values per cell type (4 technical replicates for each of 2 biological replicates per cell type). An LRT design was used of ~technicalRep + biologicalRep + cellType, with ~technicalRep + biologicalRep as the comparison. Size factors were provided manually based on NeoLoopFinder output (**Supp. Table 2**). Regions of duplication were identified based on NeoLoopFinder CNV output and spectral karyotypes. For each duplicated region, the CNVs across the region were averaged and used as custom size factors to correct for differences in chromosomal copy number between cell lines. Differential loops were identified based on a fold-change cutoff of 1.5 and an adjusted p-value cutoff of 0.025.

Differential TAD boundaries were detected using an alternative method. Insulation scores were pulled from all TAD boundaries at the technical replicate level (8 values per cell type). A T-test was then applied for each pairwise comparison of cell types, and p-values were adjusted using FDR. Boundaries with an adjusted p-value below 0.01 were considered significant.

Differential loops, TAD boundaries, and genes were clustered based on their patterns of change across the three cell types using k-means clustering of centered and scaled normalized counts.

Integration of Cancer Genome Atlas data

RNA-Seq data from the breast cancer (BRCA) cohort of the Cancer Genome Atlas was used to analyze whether genes differentially expressed in the MCF10 progression model are also differentially expressed in breast cancer tumor tissues (n=1102) from patients comparted to normal control tissues (n = 113). Normalized gene transcript counts from RSEM were downloaded from Xena (129). Protein coding genes that are differentially expressed between tumors and paired peritumors in TCGA-BRCA data were identified using GEPIA3 using DESeq2 with a log2 fold-change cutoff of 1 and an adjusted p-value cutoff of 0.05. A total of 5,289 differential genes were identified from the TCGA-BRCA data: 3,168 with significantly higher expression in tumor samples, and 2,121 with significantly higher expression in normal samples (130). Overall survival curves for select genes were generated using GEPIA (130).

Integration of additional breast cancer structural data

Chromatin loops and TAD boundaries identified in the MCF10 progression model were also examined in other triple-negative breast cancer cell lines and patient samples using data from GEO (accession GSE167150) (49). Insulation score tracks were calculated as for MCF10 cell lines, using FAN-C at 10kb resolution and KR normalization. Chromatin loop counts were extracted using KR normalization at 10kb resolution.

Micro-C feature overlap and aggregate analysis

Micro-C feature overlap and analysis was conducted in R primarily using the *GenomicRanges*, *InteractionSet*, and *mariner* packages (131–133). To overlap chromatin loops with other features such as promoters and enhancers, we used loop anchors at 10kb resolution and allowed features within 10kb (±1 Hi-C pixel) of the loop anchors to be considered overlapping. We used this broad definition for loop overlap to account for the fact that loops are non-punctate and show increased contact frequency within several bins on either side of the called loop pixel, as evidenced by both individual loci and aggregate loop analysis. Aggregate matrices of loop pixels and TAD boundaries were generated using *mariner* and visualized using *plotgardener* (133,134).

Activity-by-contact

The activity-by-contact was used based on Fulco et al. 2019 with slight adjustments (135). To allow for direct comparison of all enhancer-promoter pairs across cell lines, we manually defined potential enhancer regions and used the same input for each cell type. These

potential enhancer regions were defined as they typically are in the pipeline, by finding 150,000 ATAC-Seq peaks with the highest H3K27ac levels. The output was filtered using a suggested ABC score cutoff of 0.025 to identify likely enhancer-promoter pairs. To allow for direct comparison with our other enhancer analysis, we filtered the output based on the enhancer regions also overlapping H3K27ac, which still represented a majority of the valid pairs identified.

Matched enhancer-promoter sets

Covariate-matched subset selection among non-looped enhancer-promoter pairs was performed using the *matchRanges* function from the *nullranges* package (136,137). Enhancer-promoter pair distance or mean Micro-C contact frequency were used as covariates. Matching was done with the stratified matching method without replacement.

Gene ontology and gene set enrichment analysis

Gene ontology (GO) term enrichment was performed in R using the *gprofiler2* package (137). Gene set enrichment analysis (GSEA) was performed with the *GSEA* software, using size factor normalized RNA-seq counts as input (138) and the Hallmark H1 gene set.

ATAC-Seq motif analysis

Motif analysis of ATAC-Seq peaks within strengthened and weakened loop anchors was conducted using the HOMER suite findMotifsGenome.pl script with size given(139). ATAC-Seq peaks within the anchors of static loop anchors were used as background.

H3K27ac peak pileup analysis

H3K27ac ChIP-Seq analysis in the anchors of gained, lost, and static loops was conducted using deeptools (124). Alignment files were normalized using RPGC with the bamCoverage function, and adjusted using scale factors generated from edgeR TMM normalization factors of counts from overlapping H3K27ac and ATAC-Seq peaks (140). Aggregate profile plots were then created using the plotProfile command.

Genomic data visualization

Micro-C contact frequency maps, aggregate analysis plots, gene annotations, and genomic signal tracks (RNA-Seq, ChIP-Seq, ATAC-Seq) were visualized and plotted in R using the *plotgardener* package (134).

ACKNOWLEDGEMENTS

We thank Kathleen Quinn and Joseph Boyd (ADD INSTITUTION) for their work building the ChIP-Seq libraries and processing the ChIP-Seq data, respectively. We thank Darawalee Wangsa and Danny Wangsa in the CCR Genetics Branch and OMICS Technology Facility at the NCI for their expert SKY analysis. We thank Alquassem Abuorquob (ADD INSTITUTION) for his assistance in ChIP-Seq library preparation. We thank Jordan Zhang, Misha Gattengo, and Sierra Wilson at Dovetail Genomics for coordinating Micro-C library preparation and preliminary analysis. STR analyses and Next Generation Sequencing was in part done with the assistance of the Vermont Integrated Genomics Resource (VIGR) at the Vermont Cancer Center, University of Vermont and the Genomics Sequencing Facility (GSF) at Greehey Children's Cancer Research Institute UT Health San Antonio. The results published here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. We thank the members of the Misteli and Stein labs for feedback.

DECLARATIONS

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Availability of data and materials
The Micro-C, ATAC-Seq, and RNA-Seq datasets generated and analyzed during the current study are available on GEO (SuperSeries GSE320319; SubSeries GSE254045, GSE320215, GSE320216). CTCF ChIP-Seq data was previously published under GEO accession GSE98551. H3K27ac and H3K27me3 ChIP-Seq data was previously published under GEO accession GSE229295.

The code used to generate figures from the current study are available on Github at the following repository: https://github.com/ksmetz/MCF10-MicroC.

Competing interests
The authors declare they have no competing interests.

Funding
This work was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research (grant no. ZIABC010309-24 to T.M.), the Northern New England Clinical Translation Network (grant no. GM115516 to G.S.), funding from the National Cancer Institute (grant no. P01CA240685 to G.S., J.S., and S.F.; grant no. 1F32CA220935-01A1 to A.F.), and funding from the Charlotte Perelman and Arthur Jason Perelman Fund for Cancer Research (G.S. and J.S.).

Authors' contributions
KSMR, TM, AF, HG, JS, and GS conceptualized the project, designed the experiments, and interpreted the results. TM, SF, GS, and JS obtained funding. KSMR and AF executed experiments. KSMR and HG performed computational processing and analysis. SF conducted ChIP-Seq processing. KHH coordinated SKY karyotyping. KSMR and TM drafted the manuscript. KSMR, AF, HG, KHH, JS, GS, and TM participated in reviewing and editing the manuscript.

ADDITIONAL FILES

Additional File 1: TableS1-microcMetrics.xls
Supplementary Table 1. Micro-C library quality metrics for individual technical replicates, biological replicates, and cell types.

Additional File 2: TableS2-CNVs.xls
Supplementary Table 2. CNVs from NeoLoopFinder (100kb resolution). Columns indicate bin coordinates (A-C), and CNV values for MCF10A (D), MCF10AT1 (E), and MCF10CA1a (F).

Additional File 3: TableS3-loops.xls
Supplementary Table 3. Chromatin loop summary table for total loop set (n = 29,205). Columns include loop anchor coordinates (A-J), loop call status by cell type (K-N), SIP AP score (O), loop name (P), loop span length (Q), average and maximum un-normalized counts (R-S), log2 fold-change (T-V), adjusted p-value (W), differential status by pairwise comparison (X-Z), un-normalized counts by technical replicate (AA-AX), maximum log2 fold-change (AY), pairwise comparison with greatest fold-change (AZ), differential cluster (BA), Z-score normalized counts (BB-BE), and variance stabilized counts (BF-BH).

Additional File 4: TableS4-TADbounds.xls
Supplementary Table 4. TAD boundary summary table for total boundary set (n = 17,097). Columns include boundary coordinates (A-C), insulation scores by technical replicate (D-AA), adjusted p-values, difference in insulation score, and differential status for each pairwise comparison (AB-AJ), differential status across all comparisons (AK), differential cluster (AL), average insulation score per cell type (AM-AO), and Z-score normalized insulation scores (AP-AS).

Additional File 5: TableS5-TADs.xls
Supplementary Table 5. TAD summary table for total set of TADs (n = 13,231). Columns include TAD boundary coordinates (A-J), and whether the TAD was called in each cell type (K-M).

Additional File 6: TableS6-differentialGenes.xls
Supplementary Table 6. Differentially expressed genes from the MCF10 progression series (n = 8,840). Columns indicate gene coordinates (A-E), gene identifiers (F-I), base mean expression (J), pairwise log2 fold-change (K-M), adjusted p-values (N-P), and differential status (Q-S), differential cluster (T), un-normalized counts by replicate (U-AC), Z-score normalized counts (AD-AF), and variance stabilized counts (AG-AI).

REFERENCES

1. Dekker J, Mirny L. The 3D genome as moderator of chromosomal communication. Cell. 2016 Mar 10;164(6):1110. doi:10.1016/J.CELL.2016.02.007 PubMed PMID: 26967279.
2. Jerkovic´ I, Cavalli G. Understanding 3D genome organization by multidisciplinary methods. Nat Rev Mol Cell Biol. 2021 May 5;22(8):511–28. doi:10.1038/S41580-021-00362-W PubMed PMID: 33953379.
3. Misteli T. The self-organizing genome: Principles of genome architecture and function. Cell. 2020 Oct 1;183(1):28. doi:10.1016/J.CELL.2020.09.014 PubMed PMID: 32976797.
4. Pombo A, Dillon N. Three-dimensional genome architecture: players and mechanisms. Nature Reviews Molecular Cell Biology 2015 16:4. 2015 Mar 11;16(4):245–57. doi:10.1038/nrm3965 PubMed PMID: 25757416.
5. Rowley MJ, Corces VG. Organizational principles of 3D genome architecture. Nature Reviews Genetics 2018 19:12. 2018 Oct 26;19(12):789–800. doi:10.1038/s41576-018-0060-8 PubMed PMID: 30367165.
6. Corces R, Corces VG, Croce L Di, Shilatifard A. The three-dimensional cancer genome. Curr Opin Genet Dev. 2016;36:1–7. doi:10.1016/j.gde.2016.01.002
7. Peng A, Peng W, Wang R, Zhao H, Yu X, Sun Y. Regulation of 3D Organization and Its Role in Cancer Biology. Front Cell Dev Biol. 2022 Jun 8;10:879465. doi:10.3389/FCELL.2022.879465 PubMed PMID: 35757006.
8. Zhang L, Liu F, Wu L, Fu S, Xing L, Zhang L, et al. Disorder of three-dimensional chromosome structure plays a role in carcinogenesis. Clinical and Translational Discovery. 2021 Oct 1;1(1):e17. doi:10.1002/CTD2.17
9. Amodeo ME, Eyler CE, Johnstone SE. Rewiring cancer: 3D genome determinants of cancer hallmarks. Curr Opin Genet Dev. 2025 Apr 1;91. doi:10.1016/J.GDE.2024.102307 PubMed PMID: 39862605.
10. Davidson IF, Peters JM. Genome folding through loop extrusion by SMC complexes. Nature Reviews Molecular Cell Biology 2021 22:7. 2021 Mar 25;22(7):445–64. doi:10.1038/s41580-021-00349-7 PubMed PMID: 33767413.
11. de Wit E, Nora EP. New insights into genome folding by loop extrusion from inducible degron technologies. Nature Reviews Genetics 2022 24:2. 2022 Sep 30;24(2):73–85. doi:10.1038/s41576-022-00530-4 PubMed PMID: 36180596.
12. Sanborn AL, Rao SSP, Huang SC, Durand NC, Huntley MH, Jewett AI, et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc Natl Acad Sci U S A. 2015 Nov 24;112(47):E6456–65. doi:10.1073/PNAS.1518552112/SUPPL_FILE/PNAS.1518552112.ST01.XLSX PubMed PMID: 26499245.
13. Goel VY, Huseyin MK, Hansen AS. Region Capture Micro-C reveals coalescence of enhancers and promoters into nested microcompartments. Nature Genetics 2023 55:6. 2023 May 8;55(6):1048–56. doi:10.1038/s41588-023-01391-1 PubMed PMID: 37157000.
14. Hsieh THS, Cattoglio C, Slobodyanyuk E, Hansen AS, Rando OJ, Tjian R, et al. Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. Mol Cell. 2020;78(3):539-553.e8. doi:10.1016/j.molcel.2020.03.002

15. Oberbeckmann E, Quililan K, Cramer P, Oudelaar AM. In vitro reconstitution of chromatin domains shows a role for nucleosome positioning in 3D genome organization. Nature Genetics 2024 56:3. 2024 Jan 30;56(3):483–92. doi:10.1038/s41588-023-01649-8 PubMed PMID: 38291333.
16. Rowley MJ, Corces VG. Organizational principles of 3D genome architecture. Nat Rev Genet. 2018;19(12):789–800. doi:10.1038/s41576-018-0060-8
17. Chakraborty S, Wenzlitschke N, Anderson MJ, Eraso A, Baudic M, Thompson JJ, et al. Structural perturbation of chromatin domains with multiple developmental regulators can severely impact gene regulation and development. bioRxiv. 2024 Aug 3;2024.08.03.606480. doi:10.1101/2024.08.03.606480
18. Chakraborty S, Wenzlitschke N, Anderson MJ, Eraso A, Baudic M, Thompson JJ, et al. Deletion of a single CTCF motif at the boundary of a chromatin domain with three FGF genes disrupts gene expression and embryonic development. Dev Cell. 2025 Feb;0(0). doi:10.1016/J.DEVCEL.2025.02.002 PubMed PMID: 40015278.
19. da Costa-Nunes JA, Noordermeer D. TADs: Dynamic structures to create stable regulatory functions. Curr Opin Struct Biol. 2023 Aug 1;81:102622. doi:10.1016/J.SBI.2023.102622 PubMed PMID: 37302180.
20. Xiao J, Hafner A, Boettiger AN. How subtle changes in 3D structure can create large changes in transcription. Elife. 2021 Jul 1;10:e64320. doi:10.7554/ELIFE.64320 PubMed PMID: 34240703.
21. Sood V, Misteli T. The stochastic nature of genome organization and function. Curr Opin Genet Dev. 2022 Feb 1;72:45–52. doi:10.1016/J.GDE.2021.10.004 PubMed PMID: 34808408.
22. D'Ippolito AM, McDowell IC, Barrera A, Hong LK, Leichter SM, Bartelt LC, et al. Pre-established Chromatin Interactions Mediate the Genomic Response to Glucocorticoids. Cell Syst. 2018 Aug 22;7(2):146-160.e7. doi:10.1016/J.CELS.2018.06.007 PubMed PMID: 30031775.
23. Kane L, Williamson I, Flyamer IM, Kumar Y, Hill RE, Lettice LA, et al. Cohesin is required for long-range enhancer action at the Shh locus. Nature Structural & Molecular Biology 2022 29:9. 2022 Sep 12;29(9):891–7. doi:10.1038/s41594-022-00821-8 PubMed PMID: 36097291.
24. Schoenfelder S, Fraser P. Long-range enhancer-promoter contacts in gene expression control. Nat Rev Genet. 2019 Aug 1;20(8):437–55. doi:10.1038/S41576-019-0128-0 PubMed PMID: 31086298.
25. Stolper RJ, Tsang FH, Georgiades E, Hansen LLP, Downes DJ, Harrold CL, et al. Loop extrusion by cohesin plays a key role in enhancer-activated gene expression during differentiation. bioRxiv. 2023 Sep 7;2023.09.07.556660. doi:10.1101/2023.09.07.556660
26. Popay TM, Dixon JR. Coming full circle: On the origin and evolution of the looping model for enhancer–promoter communication. J Biol Chem. 2022 Aug 1;298(8):102117. doi:10.1016/J.JBC.2022.102117 PubMed PMID: 35691341.
27. Benabdallah NS, Williamson I, Illingworth RS, Kane L, Boyle S, Sengupta D, et al. Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. Mol Cell. 2019 Nov 7;76(3):473-484.e7. doi:10.1016/J.MOLCEL.2019.07.038 PubMed PMID: 31494034.

28. Despang A, Schöpflin R, Franke M, Ali S, Jerković I, Paliou C, et al. Functional dissection of the Sox9–Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. Nature Genetics 2019 51:8. 2019 Jul 29;51(8):1263–71. doi:10.1038/s41588-019-0466-z PubMed PMID: 31358994.
29. Hansen KL, Adachi AS, Braccioli L, Kadvani S, Boileau RM, Shah R, et al. Synergy between cis-regulatory elements can render cohesin dispensable for distal enhancer function. bioRxiv. 2024. doi:10.1101/2024.10.04.615095
30. Hsieh THS, Cattoglio C, Slobodyanyuk E, Hansen AS, Darzacq X, Tjian R. Enhancer-promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1. Nat Genet. 2022 Dec 1;54(12):1919–32. doi:10.1038/S41588-022-01223-8 PubMed PMID: 36471071.
31. Rao SSP, Huang SC, Glenn St Hilaire B, Engreitz JM, Perez EM, Kieffer-Kwon KR, et al. Cohesin Loss Eliminates All Loop Domains. Cell. 2017 Oct 5;171(2):305-320.e24. doi:10.1016/J.CELL.2017.09.026 PubMed PMID: 28985562.
32. Cuartero S, Weiss FD, Dharmalingam G, Guo Y, Ing-Simmons E, Masella S, et al. Control of inducible gene expression links cohesin to hematopoietic progenitor self-renewal and differentiation. Nat Immunol. 2018 Sep 1;19(9):932–41. doi:10.1038/S41590-018-0184-1 PubMed PMID: 30127433.
33. Popay TM, Pant A, Munting F, Black ME, Haghani N, Dixon JR. Genome-wide in vivo dynamics of cohesin-mediated loop extrusion and its role in transcription activation. bioRxiv. 2024 Oct 2;2024.10.02.616323. doi:10.1101/2024.10.02.616323
34. Wang M, Sunkel BD, Ray WC, Stanton BZ. Chromatin structure in cancer. BMC Mol Cell Biol. 2022 Dec 1;23(1):1–10. doi:10.1186/S12860-022-00433-6/FIGURES/1 PubMed PMID: 35902807.
35. Guo YA, Chang MM, Huang W, Ooi WF, Xing M, Tan P, et al. Mutation hotspots at CTCF binding sites coupled to chromosomal instability in gastrointestinal cancers. Nat Commun. 2018 Dec 1;9(1). doi:10.1038/S41467-018-03828-2 PubMed PMID: 29670109.
36. Katainen R, Dave K, Pitkänen E, Palin K, Kivioja T, Välimäki N, et al. CTCF/cohesin-binding sites are frequently mutated in cancer. Nat Genet. 2015 Jun 26;47(7):818–21. doi:10.1038/NG.3335 PubMed PMID: 26053496.
37. Cosenza MR, Rodriguez-Martin B, Korbel JO. Structural Variation in Cancer: Role, Prevalence, and Mechanisms. Annu Rev Genomics Hum Genet. 2022 Aug 31;23:123–52. doi:10.1146/ANNUREV-GENOM-120121-101149 PubMed PMID: 35655332.
38. Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. Nat Rev Cancer. 2007 Apr;7(4):233–45. doi:10.1038/NRC2091 PubMed PMID: 17361217.
39. Della Chiara G, Jiménez C, Virdi M, Crosetto N, Bienko M. Enhancers dysfunction in the 3D genome of cancer cells. Front Cell Dev Biol. 2023 Nov 6;11:1303862. doi:10.3389/FCELL.2023.1303862/BIBTEX
40. Barutcu AR, Lajoie BR, McCord RP, Tye CE, Hong D, Messier TL, et al. Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells. Genome Biol. 2015 Sep 28;16(1). doi:10.1186/S13059-015-0768-0 PubMed PMID: 26415882.

41. Fritz AJ, Ghule PN, Boyd JR, Tye CE, Page NA, Hong D, et al. Intranuclear and higher-order chromatin organization of the major histone gene cluster in breast cancer. J Cell Physiol. 2017 Feb 1;233(2):1278. doi:10.1002/JCP.25996 PubMed PMID: 28504305.
42. Johnstone SE, Reyes A, Qi Y, Adriaens C, Hegazi E, Pelka K, et al. Large-Scale Topological Changes Restrain Malignant Progression in Colorectal Cancer. Cell. 2020 Sep 17;182(6):1474. doi:10.1016/J.CELL.2020.07.030 PubMed PMID: 32841603.
43. Li J, Fang K, Choppavarapu L, Yang K, Yang Y, Wang J, et al. Hi-C profiling of cancer spheroids identifies 3D-growth-specific chromatin interactions in breast cancer endocrine resistance. Clin Epigenetics. 2021 Dec 1;13(1):1–13. doi:10.1186/S13148-021-01167-6/FIGURES/6 PubMed PMID: 34535185.
44. Mourad R, Hsu PY, Juan L, Shen C, Koneru P, Lin H, et al. Estrogen Induces Global Reorganization of Chromatin Structure in Human Breast Cancer Cells. PLoS One. 2014 Dec 3;9(12):e113354. doi:10.1371/JOURNAL.PONE.0113354 PubMed PMID: 25470140.
45. Rickman DS, Soong TD, Moss B, Mosquera JM, Dlabal J, Terry S, et al. Oncogene-mediated alterations in chromatin conformation. Proc Natl Acad Sci U S A. 2012 Jun 5;109(23):9083–8. doi:10.1073/PNAS.1112570109/-/DCSUPPLEMENTAL PubMed PMID: 22615383.
46. Taberlay PC, Achinger-Kawecka J, Lun ATL, Buske FA, Sabir K, Gould CM, et al. Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations. Genome Res. 2016 Jun 1;26(6):719–31. doi:10.1101/GR.201517.115 PubMed PMID: 27053337.
47. Yang Y, Choppavarapu L, Fang K, Naeini AS, Nosirov B, Li J, et al. The 3D genomic landscape of differential response to EGFR/HER2 inhibition in endocrine-resistant breast cancer cells. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms. 2020 Nov 1;1863(11):194631. doi:10.1016/J.BBAGRM.2020.194631 PubMed PMID: 32956836.
48. Zhou Y, Gerrard DL, Wang J, Li T, Yang Y, Fritz AJ, et al. Temporal dynamic reorganization of 3D chromatin architecture in hormone-induced breast cancer and endocrine resistance. Nature Communications 2019 10:1. 2019 Apr 3;10(1):1–14. doi:10.1038/s41467-019-09320-9 PubMed PMID: 30944316.
49. Kim T, Han S, Chun Y, Yang H, Min H, Jeon SY, et al. Comparative characterization of 3D chromatin organization in triple-negative breast cancers. Experimental & Molecular Medicine 2022 54:5. 2022 May 5;54(5):585–600. doi:10.1038/s12276-022-00768-2 PubMed PMID: 35513575.
50. Rossini R, Oshaghi M, Nekrasov M, Bellanger A, Domaschenz R, Dijkwel Y, et al. Loss of multi-level 3D genome organization during breast cancer progression. bioRxiv. 2024 Nov 27;2023.11.26.568711. doi:10.1101/2023.11.26.568711 PubMed PMID: 38076897.
51. Mortenson KL, Dawes C, Wilson ER, Patchen NE, Johnson HE, Gertz J, et al. 3D genomic analysis reveals novel enhancer-hijacking caused by complex structural alterations that drive oncogene overexpression. Nature Communications 2024 15:1. 2024 Jul 20;15(1):1–15. doi:10.1038/s41467-024-50387-w PubMed PMID: 39033128.
52. Wang X, Xu J, Zhang B, Hou Y, Song F, Lyu H, et al. Genome-wide detection of enhancer-hijacking events from chromatin interaction data in rearranged genomes. Nature Methods 2021 18:6. 2021 Jun 3;18(6):661–8. doi:10.1038/s41592-021-01164-w PubMed PMID: 34092790.

53. Puleo J, Polyak K. The MCF10 Model of Breast Tumor Progression. Cancer Res. 2021 Aug 15;81(16):4183–5. doi:10.1158/0008-5472.CAN-21-1939 PubMed PMID: 34400468.
54. Soule HD, Maloney TM, Wolman SR, Brenz R, Russo J, Pauley RJ, et al. Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. Cancer Res. 1990 Sep 1;50(18):6075–86. PubMed PMID: 1975513.
55. Dawson PJ, Wolman SR, Tait L, Heppner GH, Miller FR. MCF10AT: a model for the evolution of cancer from proliferative breast disease. Am J Pathol. 1996 Jan;148(1):313. PubMed PMID: 8546221.
56. Santner SJ, Dawson PJ, Tait L, Soule HD, Eliason J, Mohamed AN, et al. Malignant MCF10CA1 cell lines derived from premalignant human breast epithelial MCF10AT cells. Breast Cancer Res Treat. 2001;65(2):101–10. doi:10.1023/A:1006461422273 PubMed PMID: 11261825.
57. Bhardwaj A, Singh H, Rajapakshe K, Tachibana K, Ganesan N, Pan Y, et al. Regulation of miRNA-29c and its downstream pathways in preneoplastic progression of triple-negative breast cancer. Oncotarget. 2017;8(12):19645. doi:10.18632/ONCOTARGET.14902 PubMed PMID: 28160548.
58. Buchanan NS, Zhao J, Zhu K, Patwa TH, Miller FR, Lubman DM. Differential expression of acidic proteins with progression in the MCF10 model of human breast disease. Int J Oncol. 2007 Oct 1;31(4):941–9. doi:10.3892/IJO.31.4.941/HTML PubMed PMID: 17786328.
59. Dorgham MG, Elliott BA, Holley CL, Mansfield KD. m6A regulates breast cancer proliferation and migration through stage-dependent changes in Epithelial to Mesenchymal Transition gene expression. Front Oncol. 2023;13:1268977. doi:10.3389/FONC.2023.1268977/FULL
60. Kadota M, Yang HH, Gomez B, Sato M, Clifford RJ, Meerzaman D, et al. Delineating Genetic Alterations for Tumor Progression in the MCF10A Series of Breast Cancer Cell Lines. PLoS One. 2010 Feb 15;5(2):e9201. doi:10.1371/JOURNAL.PONE.0009201 PubMed PMID: 20169162.
61. Lin HH, Qraitem M, Lian Y, Taylor SR, Farkas ME. Analyses of BMAL1 and PER2 Oscillations in a Model of Breast Cancer Progression Reveal Changes With Malignancy. Integr Cancer Ther. 2019 Apr 1;18:1534735419836494. doi:10.1177/1534735419836494 PubMed PMID: 30943793.
62. Maguire SL, Peck B, Wai PT, Campbell J, Barker H, Gulati A, et al. Three-dimensional modelling identifies novel genetic dependencies associated with breast cancer progression in the isogenic MCF10 model. J Pathol. 2016 Nov 1;240(3):315. doi:10.1002/PATH.4778 PubMed PMID: 27512948.
63. Marella N V., Malyavantham KS, Wang J, Matsui SI, Liang P, Berezney R. Cytogenetic and cDNA Microarray Expression Analysis of MCF10A Human Breast Cancer Progression Cell Lines. Cancer Res. 2009 Jul 15;69(14):5946. doi:10.1158/0008-5472.CAN-09-0420 PubMed PMID: 19584277.
64. Rhee DK, Park SH, Jang YK. Molecular signatures associated with transformation and progression to breast cancer in the isogenic MCF10 model. Genomics. 2008 Dec 1;92(6):419–28. doi:10.1016/J.YGENO.2008.08.005 PubMed PMID: 18804527.

65. So JY, Lee HJ, Kramata P, Minden A, Suh N. Differential Expression of Key Signaling Proteins in MCF10 Cell Lines, a Human Breast Cancer Progression Model. Mol Cell Pharmacol. 2012;4(1):31. PubMed PMID: 24558516.
66. Sen A, Xu Z, Tyndale ST, Yasis J, Cho CY, Bump R, et al. Heterogeneity in chromatin structure drives core regulatory pathways in B-cell Acute Lymphoblastic Leukemia. bioRxiv. 2024 Oct 5;2024.10.04.616668. doi:10.1101/2024.10.04.616668
67. Liu M, Jin S, Agabiti SS, Jensen TB, Yang T, Radda JSD, et al. Tracing the evolution of single-cell 3D genomes in Kras-driven cancers. Nat Genet. 2025 Dec 1;57(12):3075–87. doi:10.1038/s41588-025-02297-w PubMed PMID: 40825871.
68. Bose A, Datta S, Mandal R, Ray U, Dhar R. Increased heterogeneity in expression of genes associated with cancer progression and drug resistance. Transl Oncol. 2024 Mar 1;41:101879. doi:10.1016/J.TRANON.2024.101879 PubMed PMID: 38262110.
69. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. Nature Reviews Clinical Oncology 2017 15:2. 2017 Nov 8;15(2):81–94. doi:10.1038/nrclinonc.2017.166 PubMed PMID: 29115304.
70. van den Brand T, Collier MD, Flach KD, Gregoricchio S, Mayayo-Peralta I, Dauyey Z, et al. A 3D genome compendium of breast cancer progression. iScience. 2025 Sep 19;28(9):113268. doi:10.1016/J.ISCI.2025.113268 PubMed PMID: 40894868.
71. Choppavarapu L, Fang K, Liu T, Ohihoin AG, Jin VX. Hi-C profiling in tissues reveals 3D chromatin-regulated breast tumor heterogeneity informing a looping-mediated therapeutic avenue. Cell Rep. 2025 Apr 22;44(4):115450. doi:10.1016/J.CELREP.2025.115450
72. Georgiades E, Harrold CL, Roberts N, Kassouf M, Riva SG, Francis HS, et al. Active regulatory elements recruit cohesin to establish cell-1 specific chromatin domains [Internet]. doi:10.1101/2023.10.13.562171
73. Rinzema NJ, Sofiadis K, Tjalsma SJD, Verstegen MJAM, Oz Y, Valdes-Quezada C, et al. Building regulatory landscapes reveals that an enhancer can recruit cohesin to create contact domains, engage CTCF sites and activate distant genes. Nature Structural & Molecular Biology 2022 29:6. 2022 Jun 16;29(6):563–74. doi:10.1038/s41594-022-00787-7 PubMed PMID: 35710842.
74. Ferlier T, Coulouarn C. Regulation of Gene Expression in Cancer—An Overview. Cells. 2022 Dec 1;11(24):4058. doi:10.3390/CELLS11244058 PubMed PMID: 36552821.
75. Gavish A, Tyler M, Greenwald AC, Hoefflin R, Simkin D, Tschernichovsky R, et al. Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. Nature. 2023 Jun 15;618(7965):598–606. doi:10.1038/S41586-023-06130-4 PubMed PMID: 37258682.
76. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015 Dec 23;1(6):417–25. doi:10.1016/J.CELS.2015.12.004 PubMed PMID: 26771021.
77. Ma XJ, Salunga R, Tuggle JT, Gaudet J, Enright E, McQuary P, et al. Gene expression profiles of human breast cancer progression. Proc Natl Acad Sci U S A. 2003 May 13;100(10):5974–9. doi:10.1073/PNAS.0931261100/SUPPL_FILE/1261TABLE7.XLS PubMed PMID: 12714683.

78. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, et al. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. Cell. 2015 Oct 8;163(2):506–19. doi:10.1016/j.cell.2015.09.033 PubMed PMID: 26451490.
79. Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, McMichael JF, et al. Comprehensive molecular portraits of human breast tumours. Nature 2012 490:7418. 2012 Sep 23;490(7418):61–70. doi:10.1038/nature11412 PubMed PMID: 23000897.
80. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 2010 Dec 14;107(50):21931–6. doi:10.1073/PNAS.1016071107/SUPPL_FILE/ST01.XLSX PubMed PMID: 21106759.
81. Glossary – ENCODE [Internet]. [cited 2025 Mar 8]. Available from: https://www.encodeproject.org/glossary/
82. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature Genetics 2007 39:3. 2007 Feb 4;39(3):311–8. doi:10.1038/ng1966 PubMed PMID: 17277777.
83. He Q, Jing H, Liaw L, Gower L, Vary C, Hua S, et al. Suppression of Spry1 inhibits triple-negative breast cancer malignancy by decreasing EGF/EGFR mediated mesenchymal phenotype. Sci Rep. 2016 Mar 15;6:23216. doi:10.1038/SREP23216 PubMed PMID: 26976794.
84. Yang Y, Zhong X, Lan Y, Liang P, Huang Y, Wang Y, et al. Aberrant inactivation of SCNN1G promotes the motility of head and neck squamous cell carcinoma. Pathol Res Pract. 2022 Dec 1;240:154175. doi:10.1016/J.PRP.2022.154175 PubMed PMID: 36327816.
85. Fulco CP, Nasser J, Jones TR, Munson G, Bergman DT, Subramanian V, et al. Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. Nature Genetics 2019 51:12. 2019 Nov 29;51(12):1664–9. doi:10.1038/s41588-019-0538-0 PubMed PMID: 31784727.
86. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Cell. 2011 Mar 4;144(5):646–74. doi:10.1016/J.CELL.2011.02.013/ASSET/2067D218-2368-451A-B4DA-CF9EF4807B23/MAIN.ASSETS/GR1.JPG PubMed PMID: 21376230.
87. Hu Y, Huang X, Lu X, Lin S. COL12A1 as a prognostic biomarker in HER2-enriched breast cancer and its association with immune infiltration. Eur J Gynaecol Oncol. 2022;43(5):85–90. doi:10.22514/ejgo.2022.045
88. Papanicolaou M, Parker AL, Yam M, Filipe EC, Wu SZ, Chitty JL, et al. Temporal profiling of the breast tumour microenvironment reveals collagen XII as a driver of metastasis. Nature Communications 2022 13:1. 2022 Aug 6;13(1):1–21. doi:10.1038/s41467-022-32255-7 PubMed PMID: 35933466.
89. Song Y, Wang L, Wang K, Lu Y, Zhou P. COL12A1 Acts as a Novel Prognosis Biomarker and Activates Cancer-Associated Fibroblasts in Pancreatic Cancer through Bioinformatics and Experimental Validation. Cancers (Basel). 2023 Mar 1;15(5):1480. doi:10.3390/CANCERS15051480/S1
90. Borcherding N, Kusner D, Kolb R, Xie Q, Li W, Yuan F, et al. Paracrine WNT5A signaling inhibits expansion of tumor-initiating cells. Cancer Res. 2015 May 15;75(10):1972. doi:10.1158/0008-5472.CAN-14-2761 PubMed PMID: 25769722.

91. Roarty K, Baxley SE, Crowley MR, Frost AR, Serra R. Loss of TGF-β or Wnt5a results in an increase in Wnt/β-catenin activity and redirects mammary tumour phenotype. Breast Cancer Res. 2009 Apr 3;11(2):R19. doi:10.1186/BCR2244 PubMed PMID: 19344510.
92. Zeng R, Huang J, Zhong MZ, Li L, Yang G, Liu L, et al. Multiple Roles of WNT5A in Breast Cancer. Med Sci Monit. 2016 Dec 22;22:5058. doi:10.12659/MSM.902022 PubMed PMID: 28005837.
93. Pasini D, Malatesta M, Jung HR, Walfridsson J, Willer A, Olsson L, et al. Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. Nucleic Acids Res. 2010 Apr 12;38(15):4958. doi:10.1093/NAR/GKQ244 PubMed PMID: 20385584.
94. Cai Y, Zhang Y, Loh YP, Tng JQ, Lim MC, Cao Z, et al. H3K27me3-rich genomic regions can function as silencers to repress gene expression via chromatin interactions. Nature Communications 2021 12:1. 2021 Jan 29;12(1):1–22. doi:10.1038/s41467-021-20940-y PubMed PMID: 33514712.
95. Guo M, Peng Y, Gao A, Du C, Herman JG. Epigenetic heterogeneity in cancer. Biomarker Research 2019 7:1. 2019 Oct 31;7(1):1–19. doi:10.1186/S40364-019-0174-Y
96. Carter B, Zhao K. The epigenetic basis of cellular heterogeneity. Nat Rev Genet. 2020 Apr 1;22(4):235. doi:10.1038/S41576-020-00300-0 PubMed PMID: 33244170.
97. Flavahan WA, Gaskell E, Bernstein BE. Epigenetic plasticity and the hallmarks of cancer. Science. 2017 Jul 21;357(6348). doi:10.1126/SCIENCE.AAL2380 PubMed PMID: 28729483.
98. Yang JH, Hansen AS. Enhancer selectivity in space and time: from enhancer–promoter interactions to promoter activation. Nat Rev Mol Cell Biol. 2024 Jul 1;25(7):574. doi:10.1038/S41580-024-00710-6 PubMed PMID: 38413840.
99. Reyes-Gopar H, Pérez-Fuentes KA, Bendall ML, Hernández-Lemus E. Integration of chromosome conformation and gene expression networks reveals regulatory mechanisms in triple negative breast cancer. Front Cell Dev Biol. 2025 Jul 4;13:1597245. doi:10.3389/fcell.2025.1597245
100. Yost KE, Zhao Y, Hung KL, Zhu K, Xu D, Corces MR, et al. Three-dimensional genome landscape of primary human cancers. Nature Genetics 2025 57:5. 2025 May 12;57(5):1189–200. doi:10.1038/s41588-025-02188-0 PubMed PMID: 40355593.
101. Guckelberger P, Doughty BR, Munson G, Rao SSP, Tan Y, Cai XS, et al. Cohesin-mediated 3D contacts tune enhancer-promoter regulation. bioRxiv. 2024 Jul 22;2024.07.12.603288. doi:10.1101/2024.07.12.603288 PubMed PMID: 39026740.
102. Kubo N, Ishii H, Xiong X, Bianco S, Meitinger F, Hu R, et al. Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. Nature Structural & Molecular Biology 2021 28:2. 2021 Jan 4;28(2):152–61. doi:10.1038/s41594-020-00539-5 PubMed PMID: 33398174.
103. Thiecke MJ, Wutz G, Muhar M, Tang W, Bevan S, Malysheva V, et al. Cohesin-Dependent and -Independent Mechanisms Mediate Chromosomal Contacts between Promoters and Enhancers. Cell Rep. 2020 Jul 21;32(3). doi:10.1016/J.CELREP.2020.107929 PubMed PMID: 32698000.

104. O'Geen H, Frietze S, Farnham PJ. Using ChIP-seq Technology to Identify Targets of Zinc Finger Transcription Factors. Methods Mol Biol. 2010;649:437. doi:10.1007/978-1-60761-753-2_27 PubMed PMID: 20680851.
105. Maor-Nof M, Shipony Z, Marinov GK, Greenleaf WJ, Gitler AD. An optimized ATAC-seq protocol for genome-wide mapping of active regulatory elements in primary mouse cortical neurons. STAR Protoc. 2021 Dec 17;2(4):100854. doi:10.1016/J.XPRO.2021.100854 PubMed PMID: 34647036.
106. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods. 2017;14(10):959–62. doi:10.1038/nmeth.4396
107. Welcome to Micro-C documentation — Micro-C 0.1 documentation [Internet]. [cited 2025 Mar 27]. Available from: https://micro-c.readthedocs.io/en/latest/index.html
108. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754–60. doi:10.1093/BIOINFORMATICS/BTP324 PubMed PMID: 19451168.
109. Open2C, Abdennur N, Fudenberg G, Flyamer IM, Galitsyna AA, Goloborodko A, et al. Pairtools: from sequencing data to chromosome contacts. bioRxiv. 2023 Feb 15;2023.02.13.528389. doi:10.1101/2023.02.13.528389 PubMed PMID: 36824968.
110. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021 Jan 29;10(2):1–4. doi:10.1093/GIGASCIENCE/GIAB008 PubMed PMID: 33590861.
111. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst. 2016 Jul 27;3(1):95–8. doi:10.1016/J.CELS.2016.07.002 PubMed PMID: 27467249.
112. Wang X, Xu J, Zhang B, Hou Y, Song F, Lyu H, et al. Genome-wide detection of enhancer-hijacking events from chromatin interaction data in rearranged genomes. Nature Methods 2021 18:6. 2021 Jun 3;18(6):661–8. doi:10.1038/s41592-021-01164-w PubMed PMID: 34092790.
113. Amemiya HM, Kundaje A, Boyle AP. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Scientific Reports 2019 9:1. 2019 Jun 27;9(1):1–5. doi:10.1038/s41598-019-45839-z PubMed PMID: 31249361.
114. Liu Y, Nanni L, Sungalee S, Zufferey M, Tavernari D, Mina M, et al. Systematic inference and comparison of multi-scale chromatin sub-compartments connects spatial organization to cell phenotypes. Nature Communications 2021 12:1. 2021 May 10;12(1):1–11. doi:10.1038/s41467-021-22666-3 PubMed PMID: 33972523.
115. Kruse K, Hug CB, Vaquerizas JM. FAN-C: a feature-rich framework for the analysis and visualisation of chromosome conformation capture data. Genome Biol. 2020 Dec 1;21(1):1–19. doi:10.1186/S13059-020-02215-9/FIGURES/4 PubMed PMID: 33334380.
116. Cresswell KG, Stansfield JC, Dozmorov MG. SpectralTAD: An R package for defining a hierarchy of topologically associated domains using spectral clustering. BMC Bioinformatics. 2020 Jul 20;21(1):1–19. doi:10.1186/S12859-020-03652-W/FIGURES/4 PubMed PMID: 32689928.
117. Jordan Rowley M, Poulet A, Nichols MH, Bixler BJ, Sanborn AL, Brouhard EA, et al. Analysis of Hi-C data using SIP effectively identifies loops in organisms from C. Elegans to

mammals. Genome Res. 2020;30(3):447–58. doi:10.1101/GR.257832.119/-/DC1 PubMed PMID: 32127418.
118. Patel H, Ewels P, Peltzer A, Hammarén R, Botvinnik O, Sturm G, et al. nf-core/rnaseq: nf-core/rnaseq v3.0 - Silver Shark [Internet]. doi:10.5281/ZENODO.4323183
119. Babraham Bioinformatics. Trim Galore! [Internet]. [cited 2017 Feb 15]. Available from: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
120. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics . 2013 Jan 1;29(1):15–21. doi:10.1093/bioinformatics/bts635 PubMed PMID: 23104886.
121. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017 Apr 6;14(4):417–9. doi:10.1038/nmeth.4197
122. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014 Dec 5;15(12):1–21. doi:10.1186/S13059-014-0550-8/FIGURES/9 PubMed PMID: 25516281.
123. Ghule PN, Boyd JR, Kabala F, Fritz AJ, Bouffard NA, Gao C, et al. Spatiotemporal higher-order chromatin landscape of human histone gene clusters at histone locus bodies during the cell cycle in breast cancer progression. Gene. 2023 Jul 1;872. doi:10.1016/j.gene.2023.147441 PubMed PMID: 37094694.
124. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44(W1):W160-5. doi:10.1093/nar/gkw257
125. Li Q, Brown JB, Huang H, Bickel PJ. Measuring reproducibility of high-throughput experiments. https://doi.org/101214/11-AOAS466. 2011 Sep 1;5(3):1752–79. doi:10.1214/11-AOAS466
126. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 2008 Sep 17;9(9):R137. doi:10.1186/gb-2008-9-9-r137
127. Babraham Bioinformatics. FastQC: A Quality Control tool for High Throughput Sequence Data [Internet]. [cited 2017 Feb 15]. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
128. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Mar 4;9(4):357–9. doi:10.1038/nmeth.1923
129. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol. 2020 Jun 1;38(6):675–8. doi:10.1038/s41587-020-0546-8 PubMed PMID: 32444850.
130. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. 2017 Jul 3;45(W1):W98–102. doi:10.1093/nar/gkx247 PubMed PMID: 28407145.
131. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for Computing and Annotating Genomic Ranges. PLoS Comput Biol. 2013;9(8):e1003118. doi:10.1371/JOURNAL.PCBI.1003118 PubMed PMID: 23950696.

132. Lun ATL, Perry M, Ing-Simmons E. Infrastructure for genomic interactions: Bioconductor classes for Hi-C, ChIA-PET and related experiments. F1000Res. 2016;5:950. doi:10.12688/F1000RESEARCH.8759.2 PubMed PMID: 27303634.
133. Davis ES, Parker SM, Kramer NE, Flores JP, Kiran M, Phanstiel DH. Mariner: explore the Hi-Cs. Bioinformatics. 2024 Jun 3;40(6). doi:10.1093/BIOINFORMATICS/BTAE352 PubMed PMID: 38814811.
134. Kramer NE, Davis ES, Wenger CD, Deoudes EM, Parker SM, Love MI, et al. Plotgardener: cultivating precise multi-panel figures in R. Bioinformatics. 2022 Mar 28;38(7):2042–5. doi:10.1093/BIOINFORMATICS/BTAC057 PubMed PMID: 35134826.
135. Fulco CP, Nasser J, Jones TR, Munson G, Bergman DT, Subramanian V, et al. Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. Nature Genetics 2019 51:12. 2019 Nov 29;51(12):1664–9. doi:10.1038/s41588-019-0538-0 PubMed PMID: 31784727.
136. Love M Davis E, Phanstiel D, Lee S MW. nullranges: Generation of null ranges via bootstrapping or covariate matching [Internet]. 2022. Available from: https://nullranges.github.io/nullranges, https://github.com/nullranges/nullranges LB - 4LTs
137. Peterson H, Kolberg L, Raudvere U, Kuzmin I, Vilo J. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. F1000Res. 2020;9:ELIXIR-709. doi:10.12688/F1000RESEARCH.24956.2 PubMed PMID: 33564394.
138. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005 Oct 25;102(43):15545–50. doi:10.1073/PNAS.0506580102/SUPPL_FILE/06580FIG7.JPG PubMed PMID: 16199517.
139. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010 May 28;38(4):576–89. doi:10.1016/j.molcel.2010.05.004 PubMed PMID: 20513432.
140. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010 Jan 1;26(1):139–40. doi:10.1093/bioinformatics/btp616 PubMed PMID: 19910308.

FIGURES

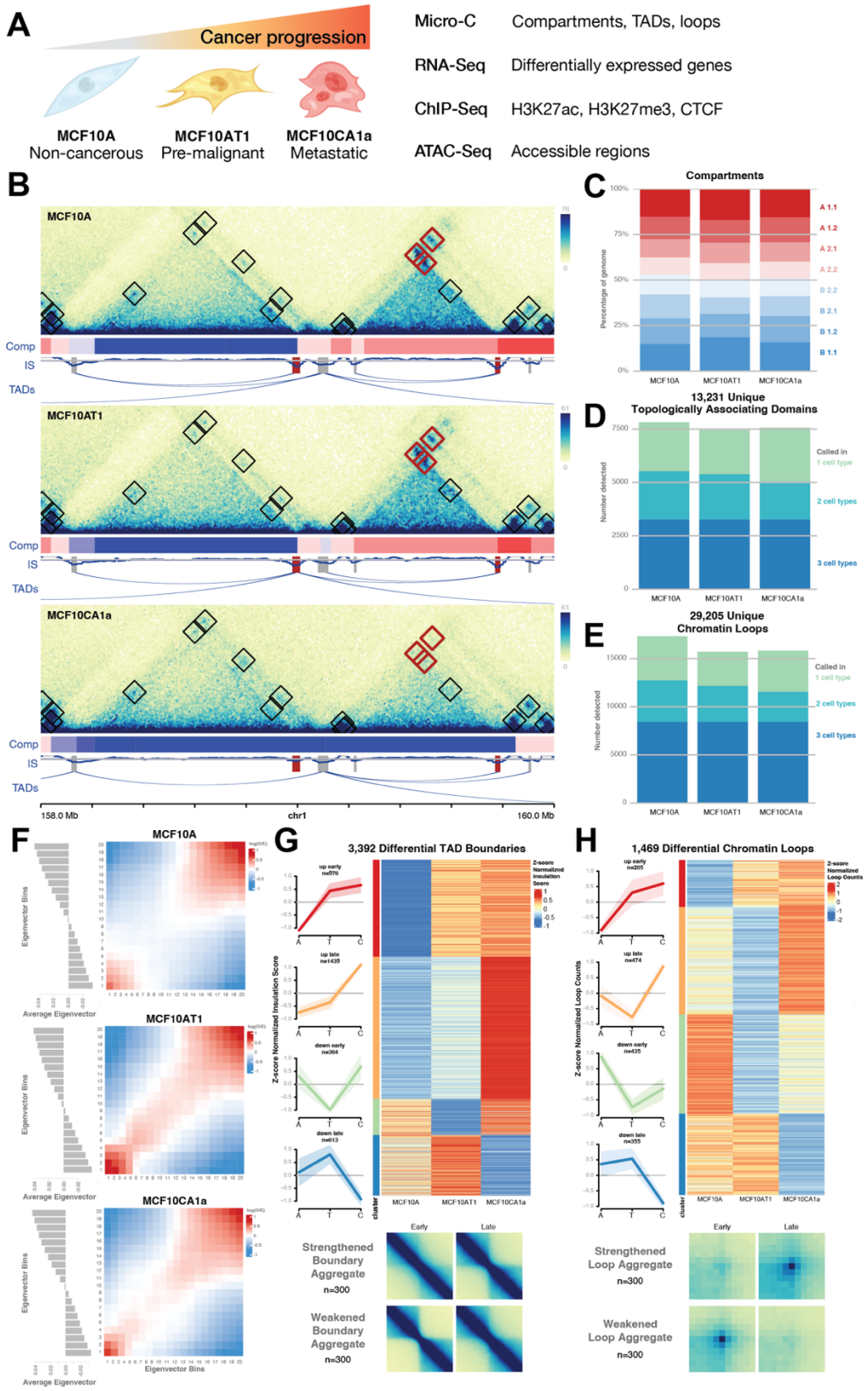

A
Cancer progression
MCF10A
Non-cancerous
MCF10AT1
Pre-malignant
MCF10CA1a
Metastatic
Micro-C
Compartments, TADs, loops
RNA-Seq
Differentially expressed genes
ChIP-Seq
H3K27ac, H3K27me3, CTCF
ATAC-Seq
Accessible regions
B
MCF10A
MCF10AT1
MCF10CA1a
Comp
IS
TADs
158.0 Mb
chr1
160.0 Mb
C
Compartments
Percentage of genome
A 1.1
A 1.2
A 2.1
A 2.2
B 2.2
B 2.1
B 1.2
B 1.1
D
13,231 Unique
Topologically Associating Domains
Number detected
Called in 1 cell type
2 cell types
3 cell types
E
29,205 Unique
Chromatin Loops
F
Eigenvector Bins
Average Eigenvector
G
3,392 Differential TAD Boundaries
Z-score Normalized Insulation Score
cluster
Early
Late
Strengthened Boundary Aggregate
n=300
Weakened Boundary Aggregate
n=300
H
1,469 Differential Chromatin Loops
Z-score Normalized Loop Counts
Strengthened Loop Aggregate
n=300
Weakened Loop Aggregate
n=300

Figure 1. Reorganization of compartments, TADs, and loops during breast cancer progression

(A) A diagram of the experimental design. Three epithelial cell lines represent various stages of breast cancer progression; MCF10A are non-cancerous, MCF10AT1 are pre-malignant, and MCF10CA1a are metastatic. In each cell line we generated 5kb resolution Micro-C to identify features such as compartments, topologically associating domains (TADs), and chromatin loops. We overlapped these features with functional changes in gene expression from RNA-Seq, histone modifications and CTCF binding from ChIP-Seq, and chromatin accessibility from ATAC-Seq. **(B)** Micro-C maps of a 2 Mb region of chromosome 1 in MCF10A (non-cancerous), MCF10AT1 (pre-cancerous), and MCF10CA1a (metastatic) cells at 5 kb resolution. Each map has annotations for loop calls, both static (black boxes) and differential (red boxes). Below each map is a track indicating compartment calls from CALDER (dark red is most A-like, dark blue is most B-like) as well as insulation scores tracks with static (grey) and differential (red) boundaries marked. Ribbons indicate TAD calls for each cell type. **(C)** Lengths of the genome assigned to each compartment in each cell type. **(D)** TAD and **(E)** loop calls from each cell type, colored by the number of maps they were initially detected in. **(F)** Saddle plots of interactions between regions of different compartments in MCF10A, MCF10AT1, and MCF10CA1a. Bottom plots indicate the average eigenvector value for each compartment ventile. Plots shown are for chromosomes 2, 12, and 17 (see Methods). **(G)** Left; Differential TAD boundaries clustered by their timing of change, depicted in line plots and heatmap. Right; aggregate plots of weakened and strengthened TAD boundaries (n=100). **(H)** Left; Differential chromatin loops clustered by their timing of change, depicted in line plots and heatmap. Right; aggregate plots of weakened and strengthened loops (n=100).

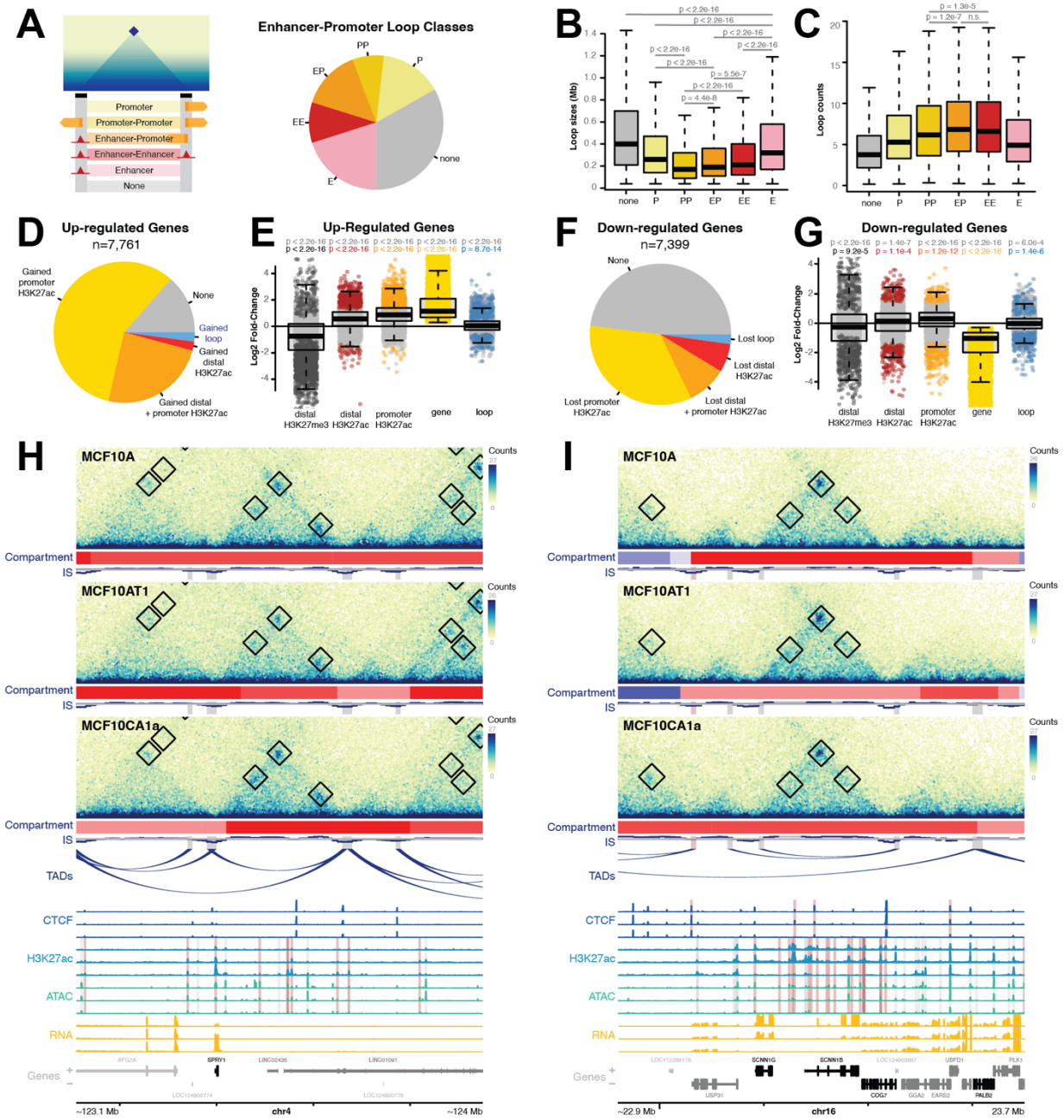


Figure 2. Persistent chromatin loops connect differentially expressed genes to distal enhancers.

(A) Percentages of loops designated as either promoter-promoter, enhancer-promoter, enhancer-enhancer, or single-sided promoter or enhancer loops. **(B)** Distributions of loop sizes by enhancer/promoter designations. P-values represent Wilcoxon tests comparing the means of different loop classes. Boxplots show the median (middle line), 25th and 75th quartiles (box perimeters), and range excluding outliers (dashed line whiskers). Outliers are defined as values that are over 1.5 times the interquartile range beyond the box bounds and are excluded from these plots. P-values represent Wilcoxon tests comparing the means of various loop sets. Non-significant (n.s.) represents p-values above 0.05. **(C)** Distributions of

loop strength by enhancer/promoter designations. P-values represent Wilcoxon tests comparing the means of various loop sets. Non-significant (n.s.) represents p-values above 0.05. **(D)** Percentages of upregulated genes that have gained H3K27ac at promoters, distal enhancers, both, or gained loops. **(E)** Log2 fold-change of distal H3K27me3 (grey), distal H3K27ac (red), promoter H3K27ac (orange), gene expression (yellow), and loop strength (blue), when overlapped. Grey dots indicate features that do not change significantly, while colored points are significantly differential features. Boxplots are defined as in (B). P-values represent Wilcoxon tests comparing the means of each class to 0. Non-significant (n.s.) represents p-values above 0.01. **(F)** Percentages of downregulated genes that have gained H3K27ac at promoters, distal enhancers, both, or gained loops. **(G)** Log2 fold-change of distal H3K27me3 (grey), distal H3K27ac (red), promoter H3K27ac (orange), gene expression (yellow), and loop strength (blue), when overlapped. Boxplot details as defined in (E). **(H)** An example of an upregulated gene (SPRY1) connected to gained enhancers by static loops. Black boxes show loop annotations. Red compartment tracks indicate A compartments, while blue indicates B compartments. In CTCF signal tracks, red highlights indicate differential CTCF peaks. In H3K27ac and ATAC-Seq signal tracks, red highlights indicate differential enhancers as determined by changes in H3K27ac. Genes highlighted in black are differentially expressed. **(I)** An example of downregulated genes (SCNN1G, SCNN1B) connected to lost enhancers by static loops. Plot annotations are as described in (H).

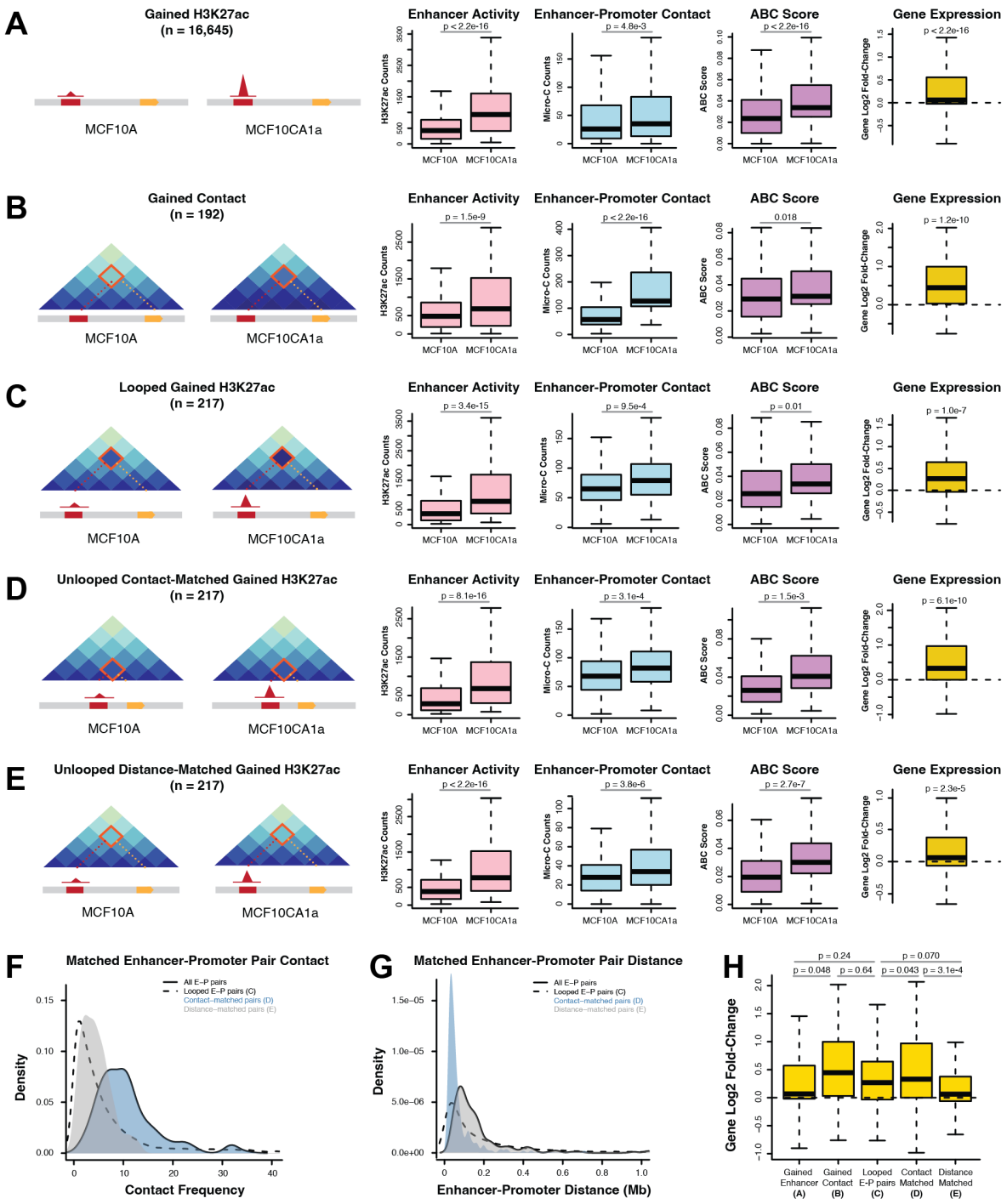


Figure 3. Changes in enhancer acetylation or enhancer-promoter contact are associated with changes in gene expression.

Boxplots of distal enhancer H3K27ac (pink), enhancer-promoter contact (blue), and ABC score (purple), as well as gene log2 fold-change (yellow) for enhancer promoter pairs that feature **(A)** differential H3K27ac among enhancers, **(B)** differential enhancer-promoter contact frequency, and **(C)** differential H3K27ac for enhancer-promoter pairs supported by a chromatin loop. Boxplots in **(D)** and **(E)** represent sets of non-looped enhancer-promoter

pairs with differential H3K27ac that are matched to the looped set in (C) by contact and distance, respectively. Boxplots show the median (middle line), 25th and 75th quartiles (box perimeters), and range excluding outliers (dashed line whiskers). Outliers are defined as values that are over 1.5 times the interquartile range beyond the box bounds and are excluded from these plots. P-values represent T-tests comparing the means of values in MCF10A and MCF10CA1a for enhancer activity, enhancer-promoter contact, and ABC Score, and T-tests comparing the mean of the value to 0 for gene log2 fold-change. **(F)** Contact distribution of all enhancer-promoter pairs (dashed line), compared to the looped enhancer-promoter pairs in (C, solid line), the contact-matched pairs in (D, blue shade), and the distance-matched pairs in (E, grey shade). **(F)** Distance distribution of all enhancer-promoter pairs (dashed line), compared to the looped enhancer-promoter pairs in (C, solid line), the contact-matched pairs in (D, blue shade), and the distance-matched pairs in (E, grey shade). **(G)** Summary boxplot of the gene log2 fold-change for the enhancer-promoter pairs previously shown in figures (A-E). P-values represent T-tests comparing the means of average gene log2 fold-changes values for different sets of enhancer-promoter pairs.

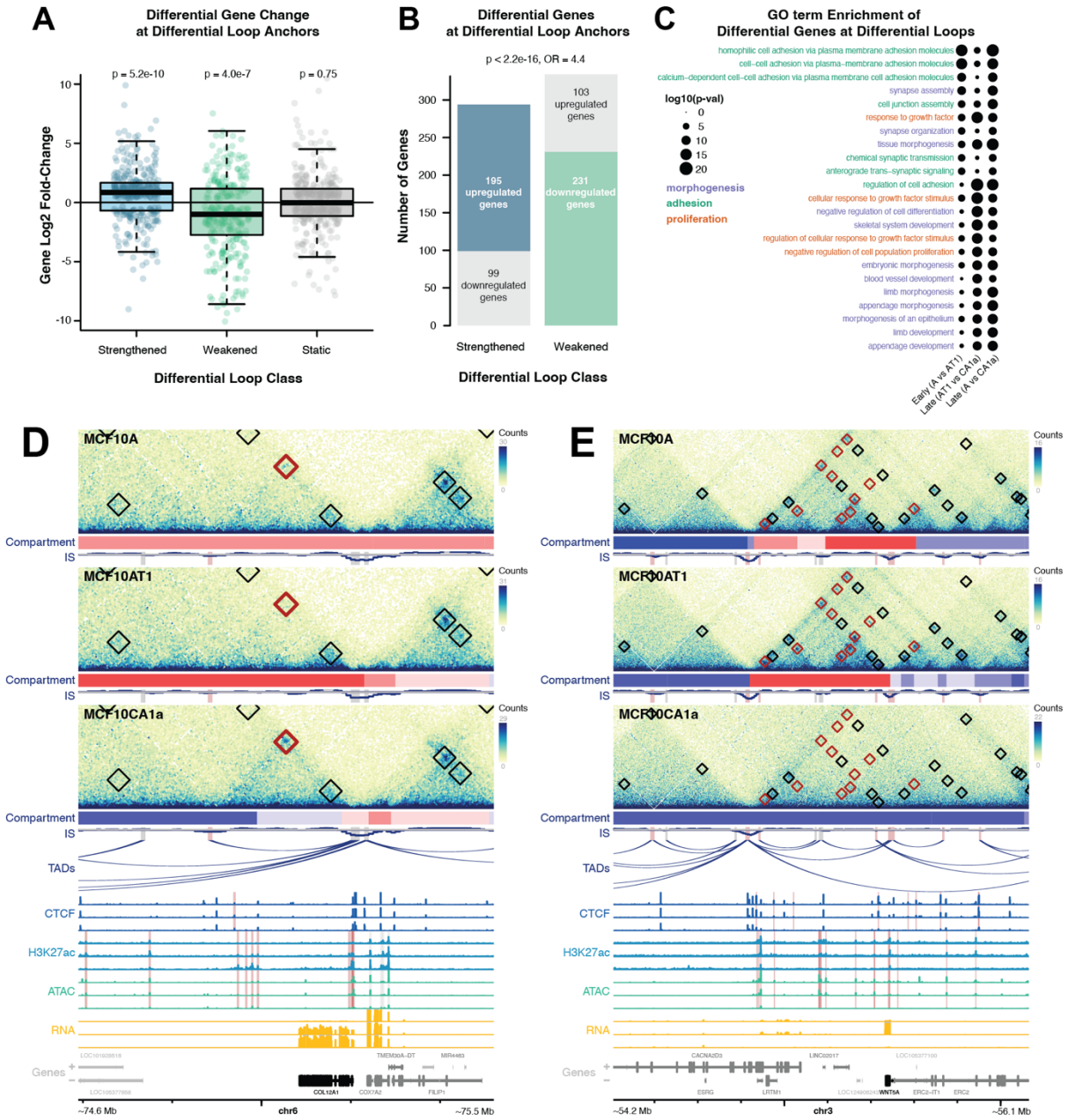


Figure 4. Differential loops are enriched for cancer-relevant differentially expressed genes

(A) Log2 fold-change of differentially expressed genes at the anchors of gained (blue), weakened (green), or static (grey) loops. Boxplots show the median (middle line), 25th and 75th quartiles (box perimeters), and range excluding outliers (dashed line whiskers). Outliers are defined as values that are over 1.5 times the interquartile range beyond the box bounds and are excluded from these plots. P-values represent Wilcoxon tests comparing the mean of each set to 0. **(B)** Bar plot showing the number of differentially expressed genes at strengthened or weakened loop anchors. Bar segments are colored by whether the gene is changing the same (blue for upregulated in strengthened loops, green for downregulated in

weakened loops) or opposite (grey) direction as the loop. P-value represents a Fisher's Exact Test for whether the odds ratio (OR) is greater than 1. **(C)** GO term enrichments for genes upregulated in MCF10A, MCF10AT1, or MCF10CA1a. Size indicates p-value. Terms are color-coded based on gene type; morphogenesis (purple), proliferation (orange), and cell adhesion (teal). **(D)** An example of an upregulated gene (COL12A1) with a promoter that overlaps a strengthened loop with distal enhancers. Black boxes show loop annotations, while red boxes indicate differential loops. Red compartment tracks indicate A compartments, while blue indicates B compartments. In CTCF signal tracks, red highlights indicate differential CTCF peaks. In H3K27ac and ATAC-Seq signal tracks, red highlights indicate differential enhancers as determined by changes in H3K27ac. Genes highlighted in black are differentially expressed. **(E)** An example of a downregulated gene (WNT5A) with a promoter that overlaps with several weakened loops containing distal enhancers that lose H3K27ac. Plots are annotated as in (A).

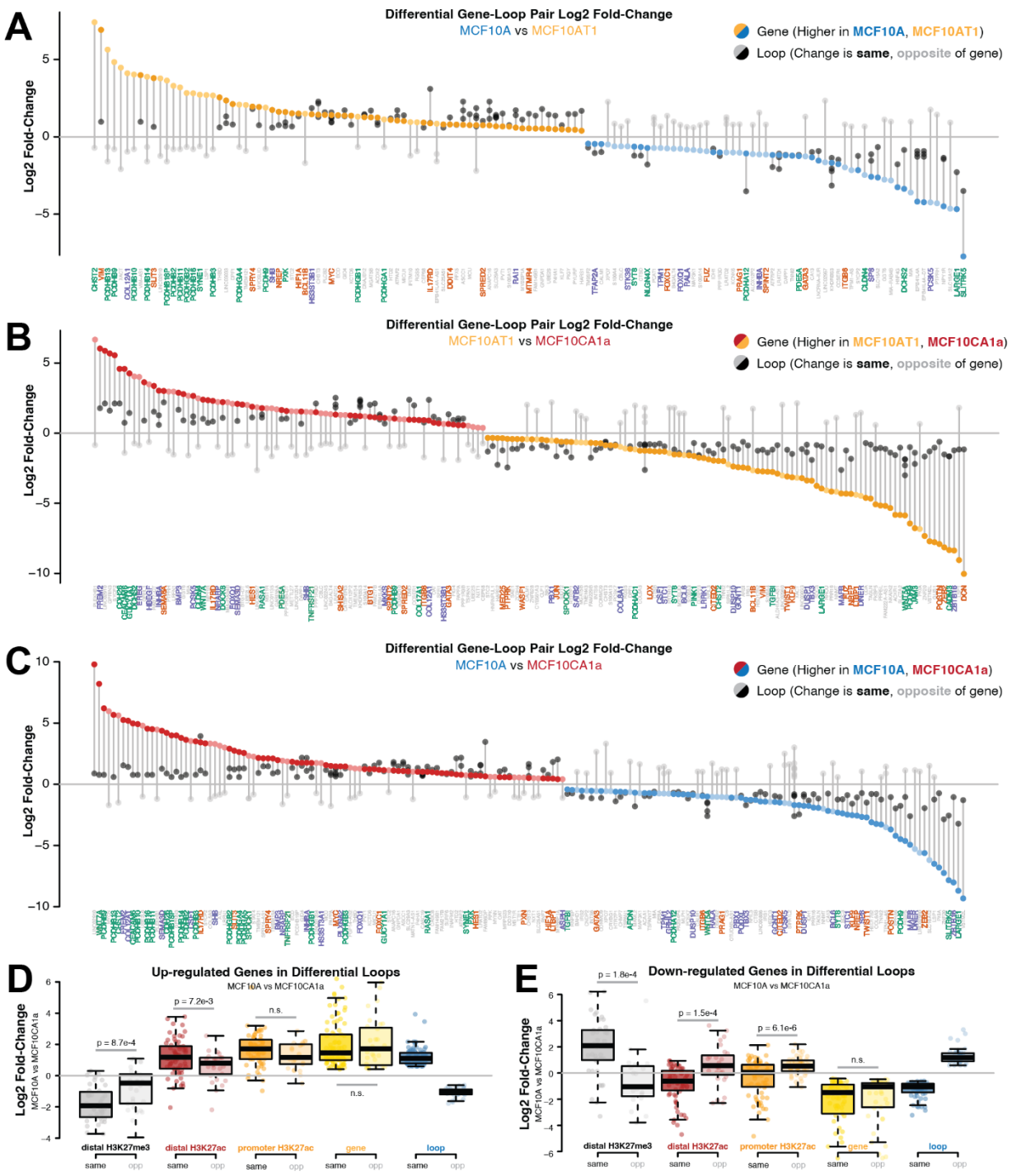


Figure 5. Progression-associated differentially expressed genes exhibit local and distal epigenetic changes at differential loops.

(A-C) Log2 fold-change of genes (colored dots) and the differential loops they overlap with (black/grey dots) for genes and loops that change between **(A)** MCF10A and MCF10AT1, **(B)** MCF10AT1 and MCF10CA1a, and **(C)** MCF10A and MCF10CA1a. Gene labels are below. **(D)** Log2 fold-change between MCF10A and MCF10CA1a of distal H3K27me3 (grey), distal

H3K27ac (red), promoter H3K27ac (orange), gene expression (yellow), and loops (blue) among upregulated genes that overlap with gained loops (darker colors) or lost loops (lighter colors). Boxplots are defined as in (A). P-values represent T-tests comparing the mean values of the features at loops that change in the same and those that change in opposite directions from the differential genes at their anchors. Non-significant (n.s.) p-values are any p- values above 0.01. **(E)** Log2 fold-change of distal H3K27me3 (grey), distal H3K27ac (red), promoter H3K27ac (orange), gene expression (yellow), and loops (blue) among downregulated genes that overlap with gained loops (darker colors) or lost loops (lighter colors). P-values represent T-tests comparing the mean values of the features at loops that change in the same and those that change in opposite directions from the differential genes at their anchors. Non-significant (n.s.) p-values are any p-values above 0.01.

SUPPLEMENTAL FIGURES

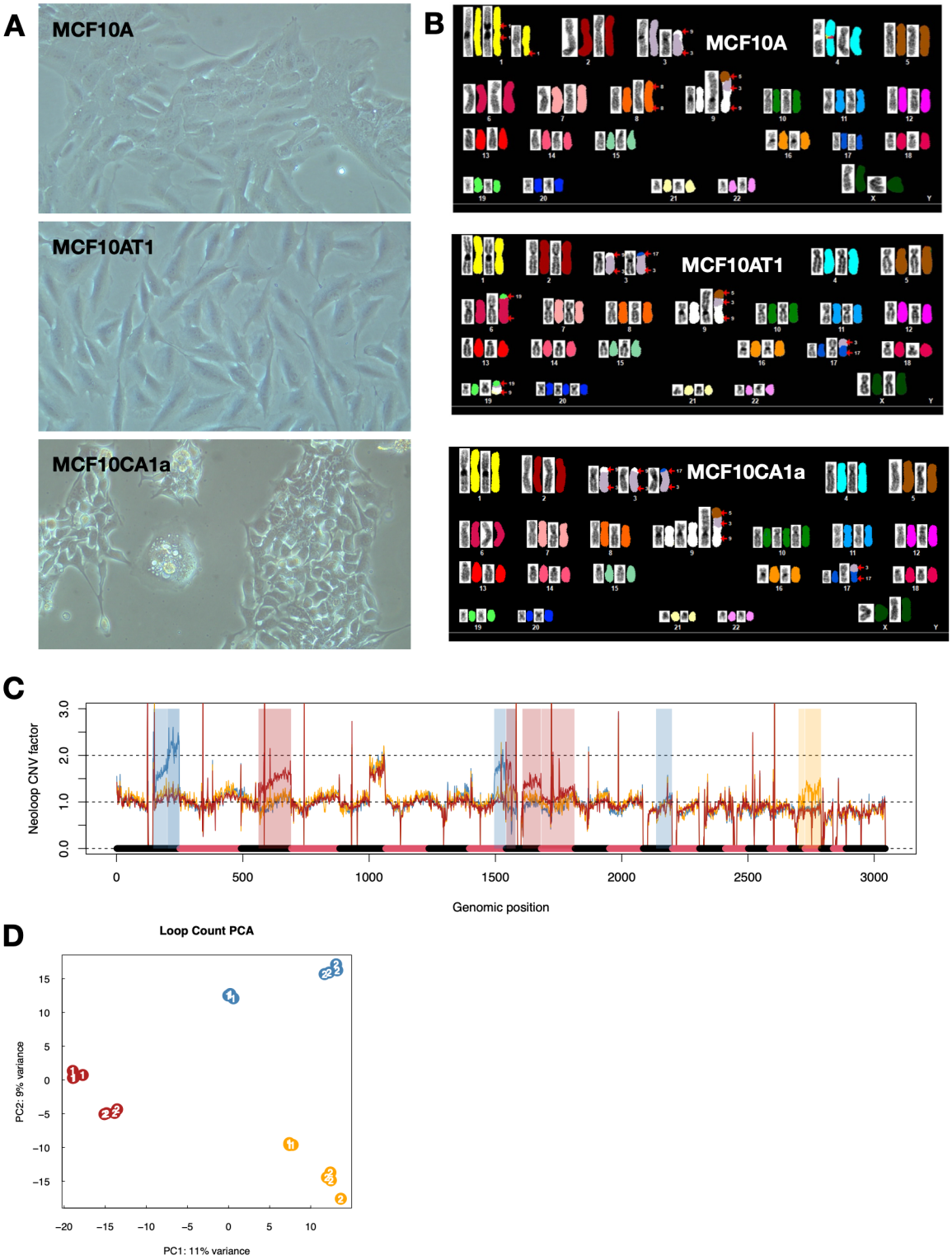
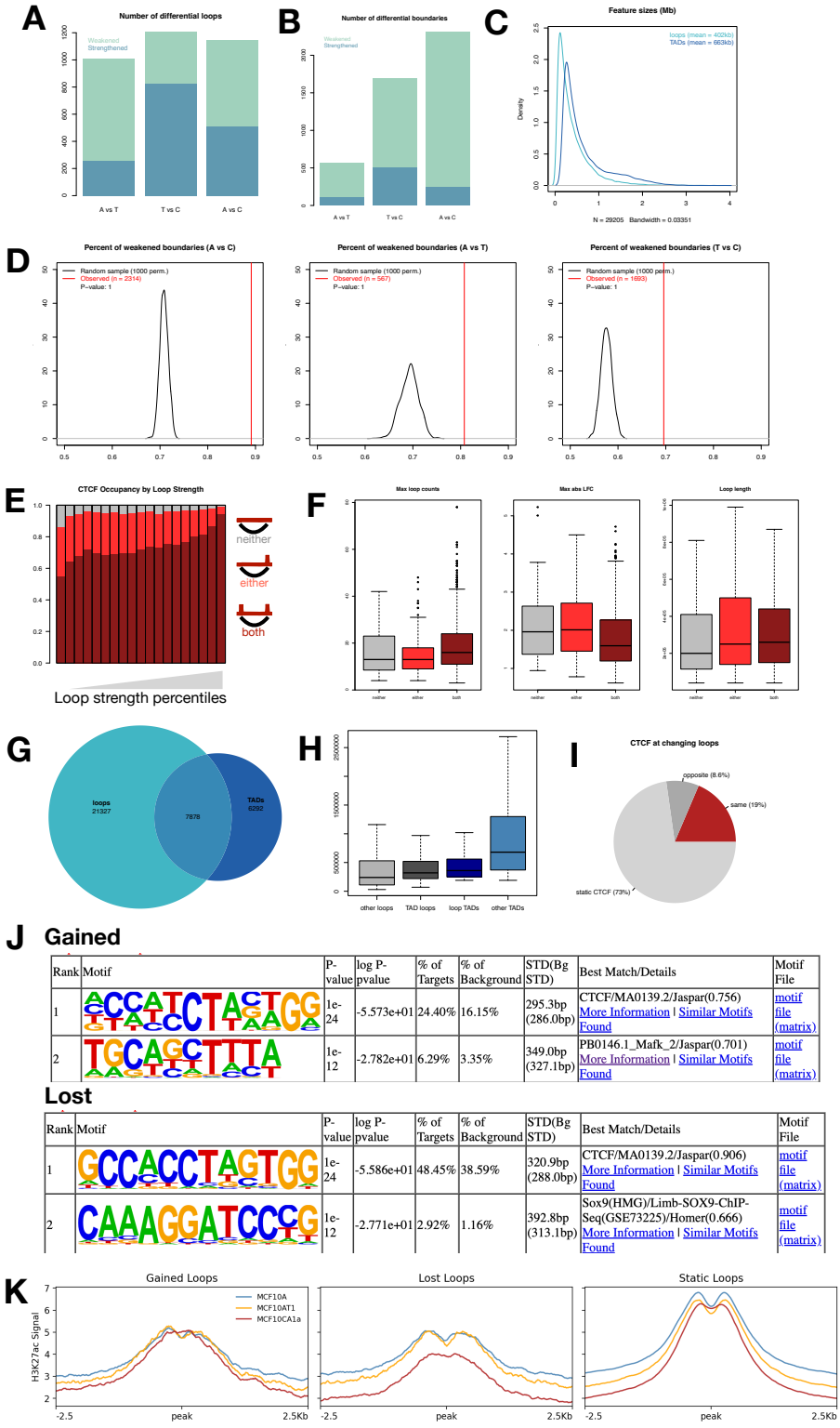


Figure S1. Karyotypic analysis and Micro-C loop reproducibility.

(A) Brightfield microscopy images of MCF10A, MCF10AT1, and MCF10CA1a cell lines. **(B)** Representative SKY karyotype example images from MCF10A, MCF10AT1, and MCF10CA1a cell lines. **(C)** Copy number variation (CNV) factors for loops across the genome as generated from NeoloopFinder. Highlighted regions indicate areas of karyotypic differences that were corrected for in the identification of differential loops; blue is regions with higher CNV in MCF10A, yellow is regions with higher CNV in MCF10AT1, and red is regions with higher CNV in MCF10CA1a. Regions that have shared karyotypic abnormalities among all three cell lines, such as the q arm of chromosome 5 which is duplicated and translocated to chromosome 9 in all cell types, did not require correction. **(D)** Micro-C chromatin loop count principal component analysis (PCA). Blue circles indicate MCF10A samples, yellow indicate MCF10AT1, and red indicates MCF10CA1a.



J Gained

Rank	Motif	P-value	log P-pvalue	% of Targets	% of Background	STD(Bg STD)	Best Match/Details	Motif File
1		1e-24	-5.573e+01	24.40%	16.15%	295.3bp (286.0bp)	CTCF/MA0139.2/Jaspar(0.756) More Information | Similar Motifs Found	motif file (matrix)
2		1e-12	-2.782e+01	6.29%	3.35%	349.0bp (327.1bp)	PB0146.1_Mafk_2/Jaspar(0.701) More Information | Similar Motifs Found	motif file (matrix)

Lost

Rank	Motif	P-value	log P-pvalue	% of Targets	% of Background	STD(Bg STD)	Best Match/Details	Motif File
1		1e-24	-5.586e+01	48.45%	38.59%	320.9bp (288.0bp)	CTCF/MA0139.2/Jaspar(0.906) More Information | Similar Motifs Found	motif file (matrix)
2		1e-12	-2.771e+01	2.92%	1.16%	392.8bp (313.1bp)	Sox9(HMG)/Limb-SOX9-ChIP-Seq(GSE73225)/Homer(0.666) More Information | Similar Motifs Found	motif file (matrix)

Figure S2. Differential loop and TAD feature details.

(A) The number of chromatin loops that exhibit a significant increase (blue) or decrease (green) in contact frequency between each pairwise comparison of cell types. **(B)** The number of topologically associating domain (TAD) boundaries that exhibit a significant increase (blue) or decrease (green) in insulation score between each pairwise comparison of cell types. **(C)** Size distributions of chromatin loops and TADs. **(D)** Permutation test results of weakened boundaries observed in each pairwise comparison as compared to a random sampling. **(E)** Percent of loops with one or both anchors overlapping CTCF ChIP-Seq peaks based on the strength of the loop (20 bins). **(F)** Boxplots of (left to right) maximum loop Micro-C counts, maximum loop count log2 fold-change, and loop length based on loop-CTCF class; neither (grey), one anchor overlap (light red), or both anchors overlap (dark red). **(G)** Venn diagram showing the degree of overlap between chromatin loops and TADs. **(H)** Feature sizes of loops that do not overlap TADs (light grey), loops that do overlap TADs (dark grey), TADs that do overlap loops (dark blue), and TADs that do not overlap loops (light blue). **(I)** Pie chart of the percentage of differential loops that have a correlated change in CTCF binding at either anchor; light grey indicates static CTCF peaks, dark grey indicates CTCF peaks that change in the opposite way as the loop (i.e. a strengthened loop with loss of CTCF binding), and dark red indicates CTCF peaks that change in the same way as the loop (i.e. a strengthened loop with gain of CTCF binding). **(J)** De novo motif results from HOMER for ATAC-Seq peaks within the anchors of gained/strengthened (top) and lost/weakened (bottom) chromatin loops. **(K)** H3K27ac ChIP-Seq aggregate profiles at the anchors of gained/strengthened, lost/weakened, and static loops.

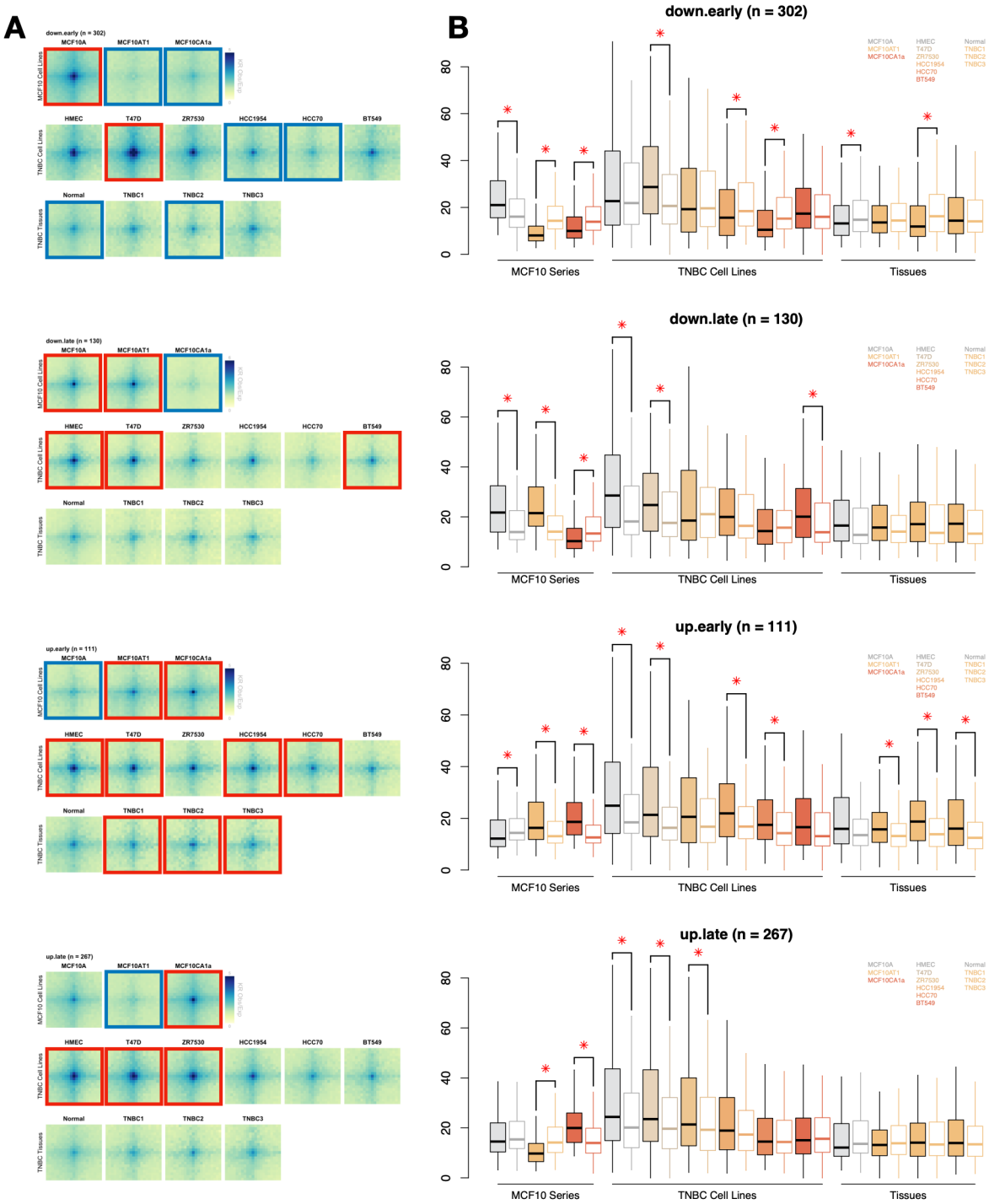


Figure S3. Differential MCF10 loops are conserved in cancer cell lines.
(A) Aggregate loop analysis of chromatin loops from each differential cluster as found in the MCF10 series compared to TNBC cell lines and primary patient tissues (Kim, Han & Chun et al. 2022). Red boxes indicate loops that have significantly higher observed/expected contact frequency among loops in that cluster as compared to a random sampling of an equal number of chromatin loops that are static in the MCF10 series, as determined in (B). **(B)** Boxplots representing quantifications of the aggregate plots shown in (A). Filled boxes

indicate the counts from loops in each differential cluster, while empty boxes indicate a matched sample of static loops that are of similar size. Stars indicate cell lines where there is a significant difference between the counts of the given loops and the static matched set ($p <= 0.05$, Wilcoxon rank sum test).

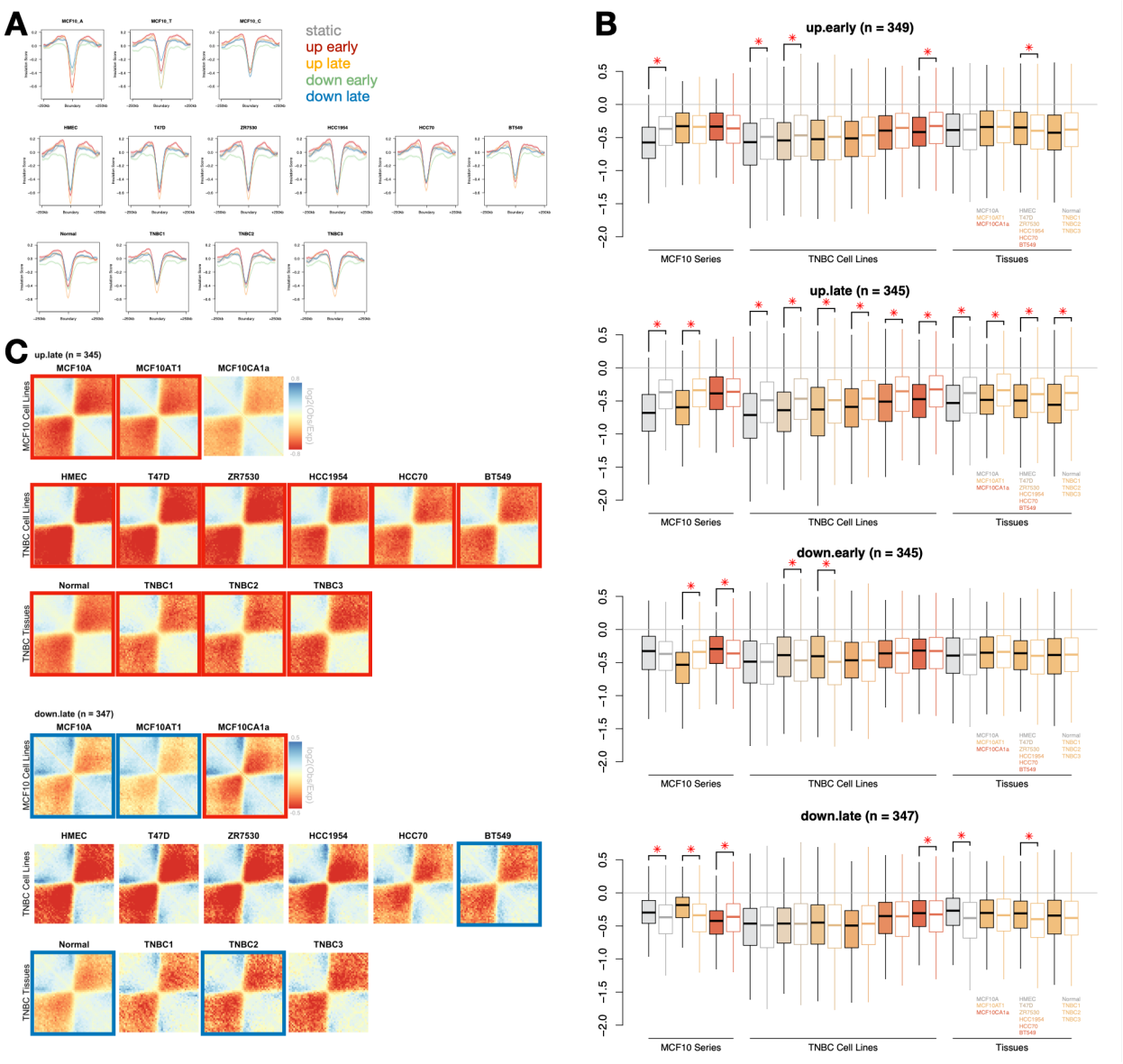


Figure S4. Differential TAD boundaries are conserved in cancer cell lines and tissues. **(A)** Average insulation score profile of TAD boundaries based on their differential cluster in MCF10 progression and in TNBC cell lines and primary patient tissues (Kim, Han & Chun et al. 2022). **(B)** Boxplots showing the distribution of insulation scores at MCF10 boundaries of various differential clusters. Filled boxes indicate the insulation scores from TADs in each differential cluster, while empty boxes indicate the distribution of scores at the same number of static TAD boundaries. Stars indicate cell lines where there is a significant difference between the insulation score of boundaries and the static matched set ($p <= 0.05$, Wilcoxon rank sum test). **(C)** Aggregate plots showing the log2 of observed/expected counts at MCF10 boundaries in MCF10 cell lines as well as cell lines from Kim, Han & Chun et al. 2022.

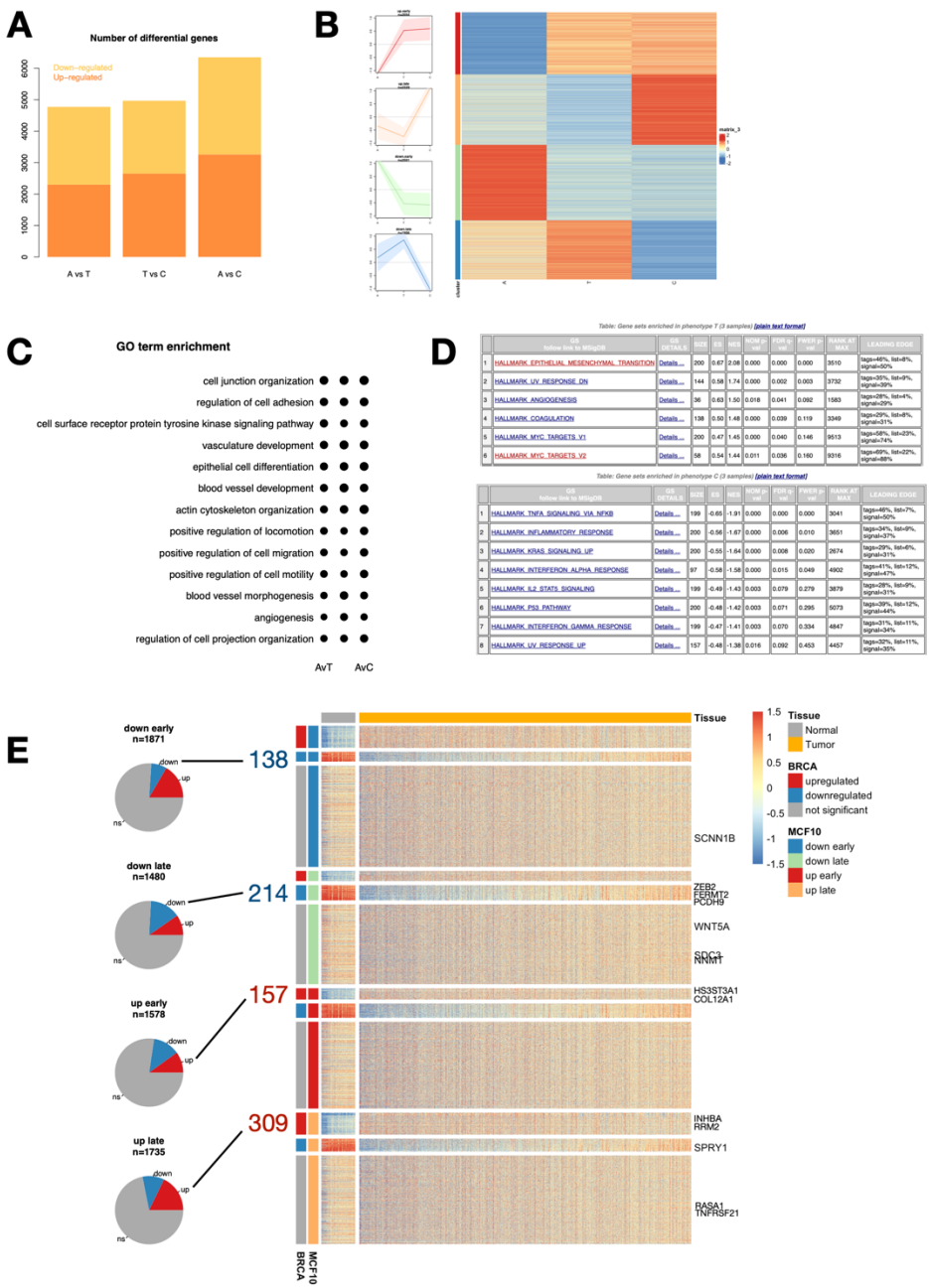


Figure S5. Differential gene expression patterns across breast cancer progression includes cancer-relevant pathways.

(A) The number of genes that exhibit a significant increase (orange) or decrease (gold) in expression between each pairwise comparison of cell types. **(B)** Differential genes clustered by their timing of change, depicted in line plots and heatmap. **(C)** Gene ontology (GO) term enrichment for genes differential expressed in each pairwise comparison of cells. **(D)** Gene set enrichment analysis (GSEA) results showing gene sets enriched among genes differentially expressed early (MCF10AT1, top) or late (MCF10CA1a, bottom) in breast cancer progression. **(E)** A heatmap shows the expression of all differential genes in the MCF10

progression series among normal (grey) and tumor (gold) tissue samples from the Cancer Genome Atlas breast cancer cohort. Genes are clustered first by their pattern of change in MCF10, then their differential status in the patient data. Example genes used in the study are listed on the right. Each cluster also shows a pie chart on the left depicting the percent of genes that are shared in the patient data, with genes that change in the same direction highlighted.

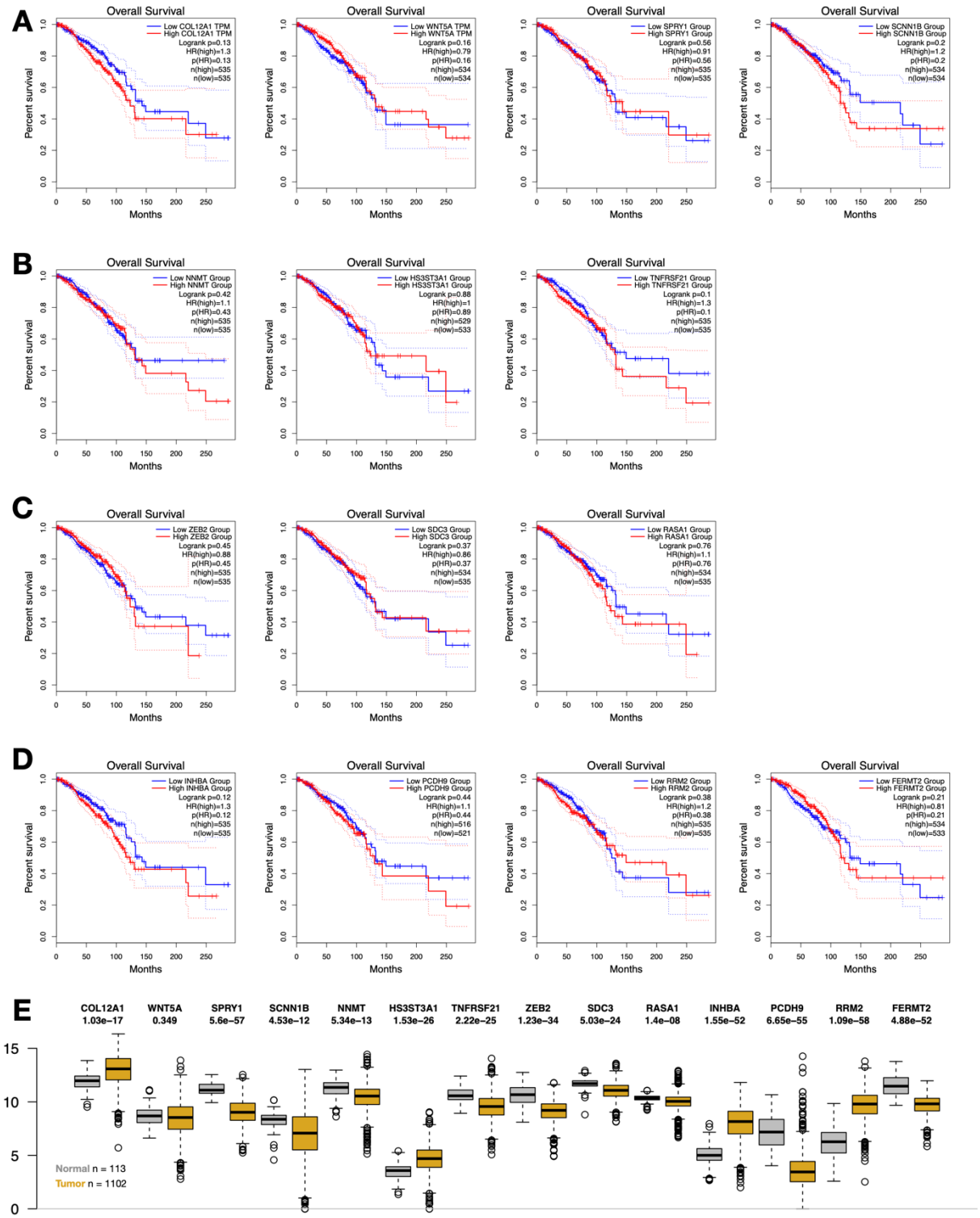


Figure S6. Clinical relevance of select genes misregulated in MCF10.
(A-D) Overall survival curves from the Cancer Genome Atlas for fourteen select genes highlighted in the main figures (A), Supplemental Figure S7 (B-C), and Supplemental Figure S12 (D). **(E)** Gene expression data from the Cancer Genome Atlas for fourteen select genes.

Grey boxplots represent expression values from normal samples and gold represent tumor samples. P-values represent the Wilcoxon rank-sum test comparing the normal and tumor expression levels for each gene.

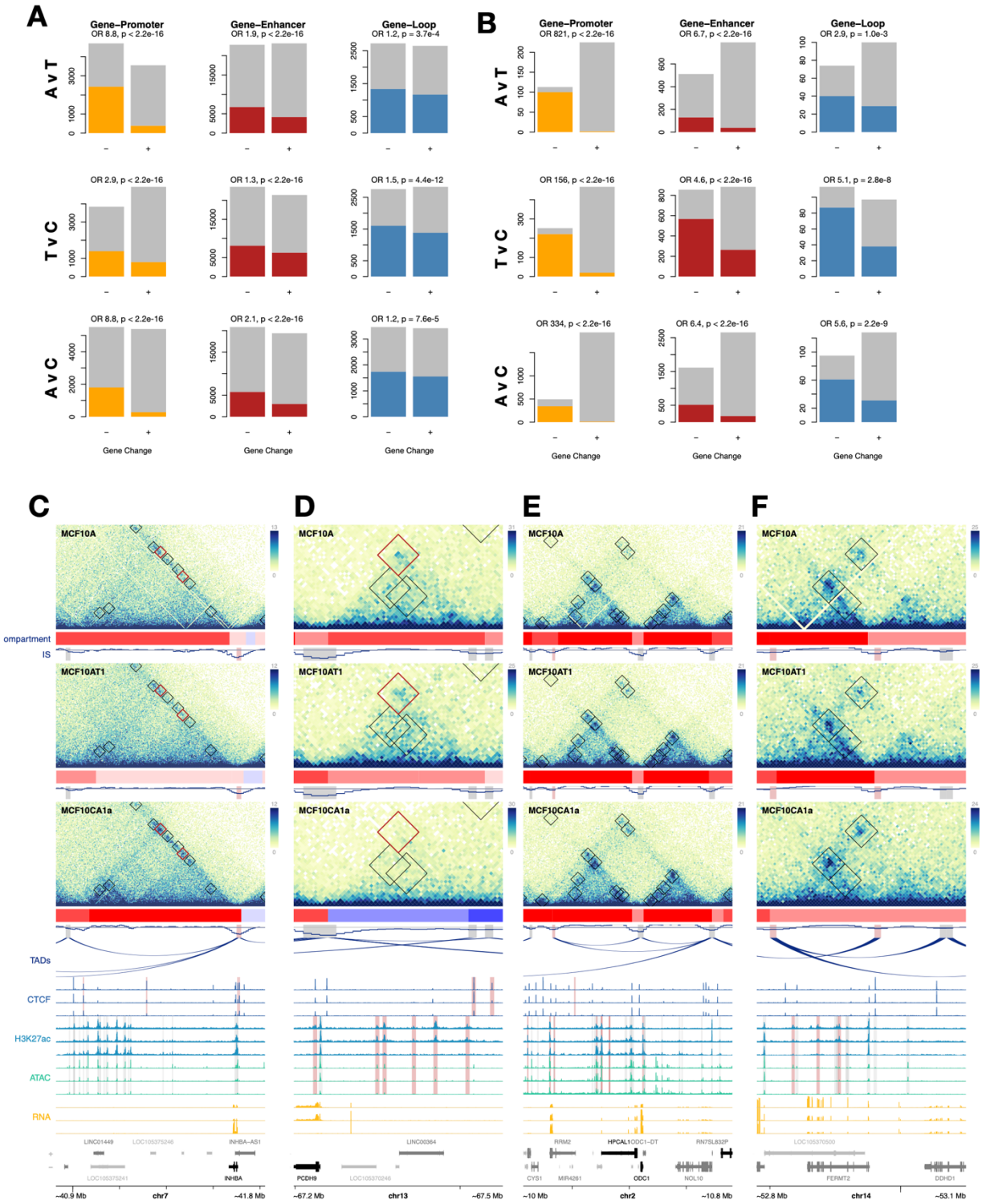


Figure S7. Differentially expressed genes overlap differential and static loops with differential distal regulatory regions.

(A) The number of promoter H3K27ac peaks (left), distal enhancer H3K27ac peaks (middle), or loops (right) that change in a positive (grey) or negative (gold, red, blue) direction based on their overlap with up-regulated or down-regulated genes. **(B)** Same plots as (**E**) but subset

for significantly differential features. **(C-F)** Examples of genes that are differentially regulated in the same direction in MCF10 progression and patient samples and which overlap with chromatin loops in MCF10 cell lines. Black boxes show loop annotations, while red boxes indicate differential loops. Red compartment tracks indicate A compartments, while blue indicates B compartments. In CTCF signal tracks, red highlights indicate differential CTCF peaks. In H3K27ac and ATAC-Seq signal tracks, red highlights indicate differential enhancers as determined by changes in H3K27ac. Genes highlighted in black are differentially expressed.

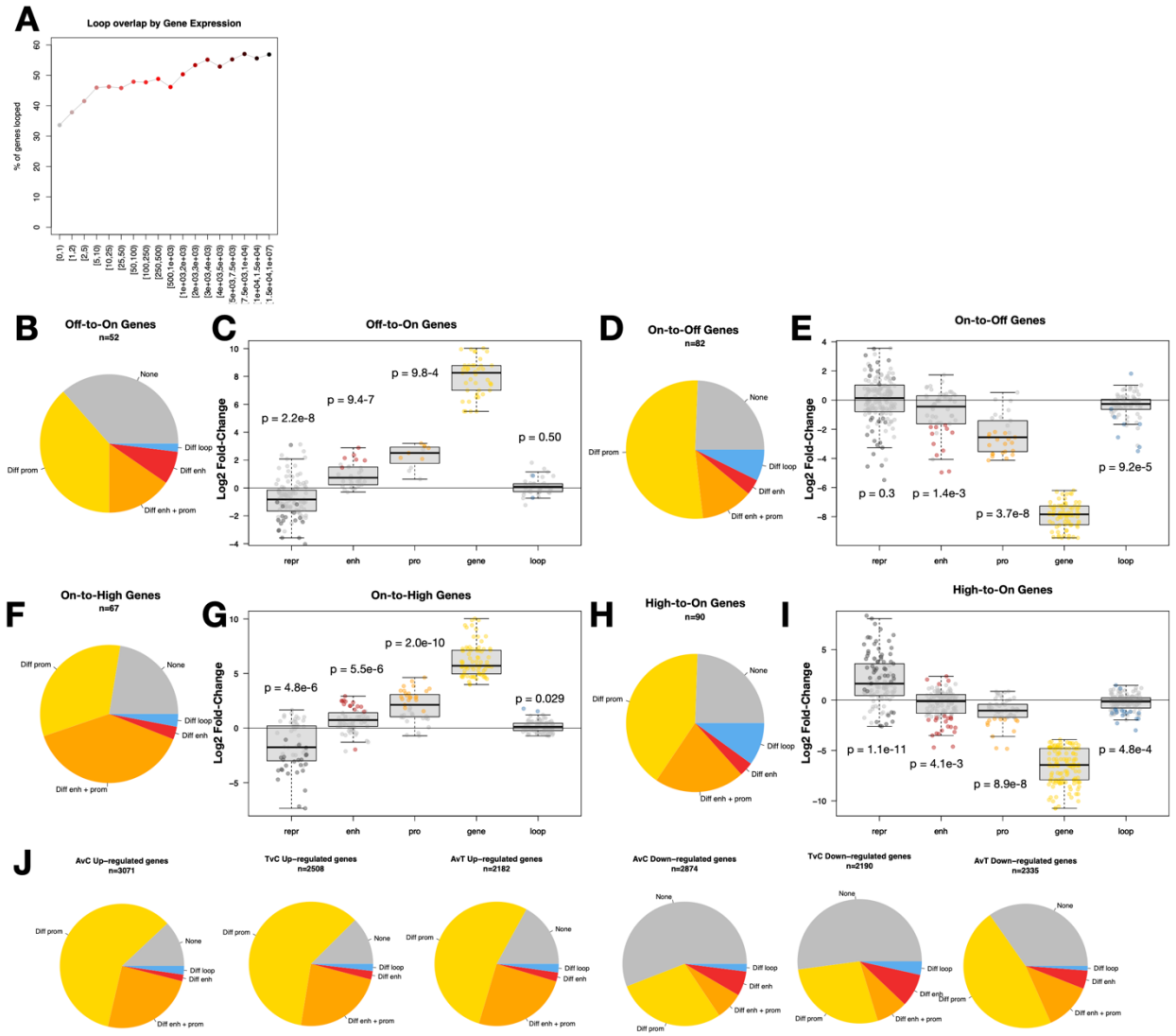


Figure S8. Relationship between differential gene expression, chromatin loops and distal enhancers is consistent based on expression levels and cancer stages.
(A) Percentage of genes that overlap with loop anchors based on gene expression level. **(B)** Percentages of off-to-on genes that have gained H3K27ac at promoters, distal enhancers, both, or gained loops. **(C)** Log2 fold-change of distal H3K27me3 (grey), distal H3K27ac (red), promoter H3K27ac (orange), gene expression (yellow), and loop strength (blue), when overlapped. Grey dots indicate features that do not change significantly, while colored points are significantly differential features. P-values represent Wilcoxon tests comparing the means of each class to 0. Non-significant (n.s.) represents p-values above 0.01. **(D-I)** Same as (B-C), but for on-to-off genes (D-E), on-to-high genes (F-G), and high-to-on genes (H-I). **(J)** Percentages of up-regulated genes (left) and down-regulated genes (right) that have correlated changes in H3K27ac at promoters (gold), distal enhancers (red), both (orange), or loop contact frequency (blue), by pairwise comparison of cell types.

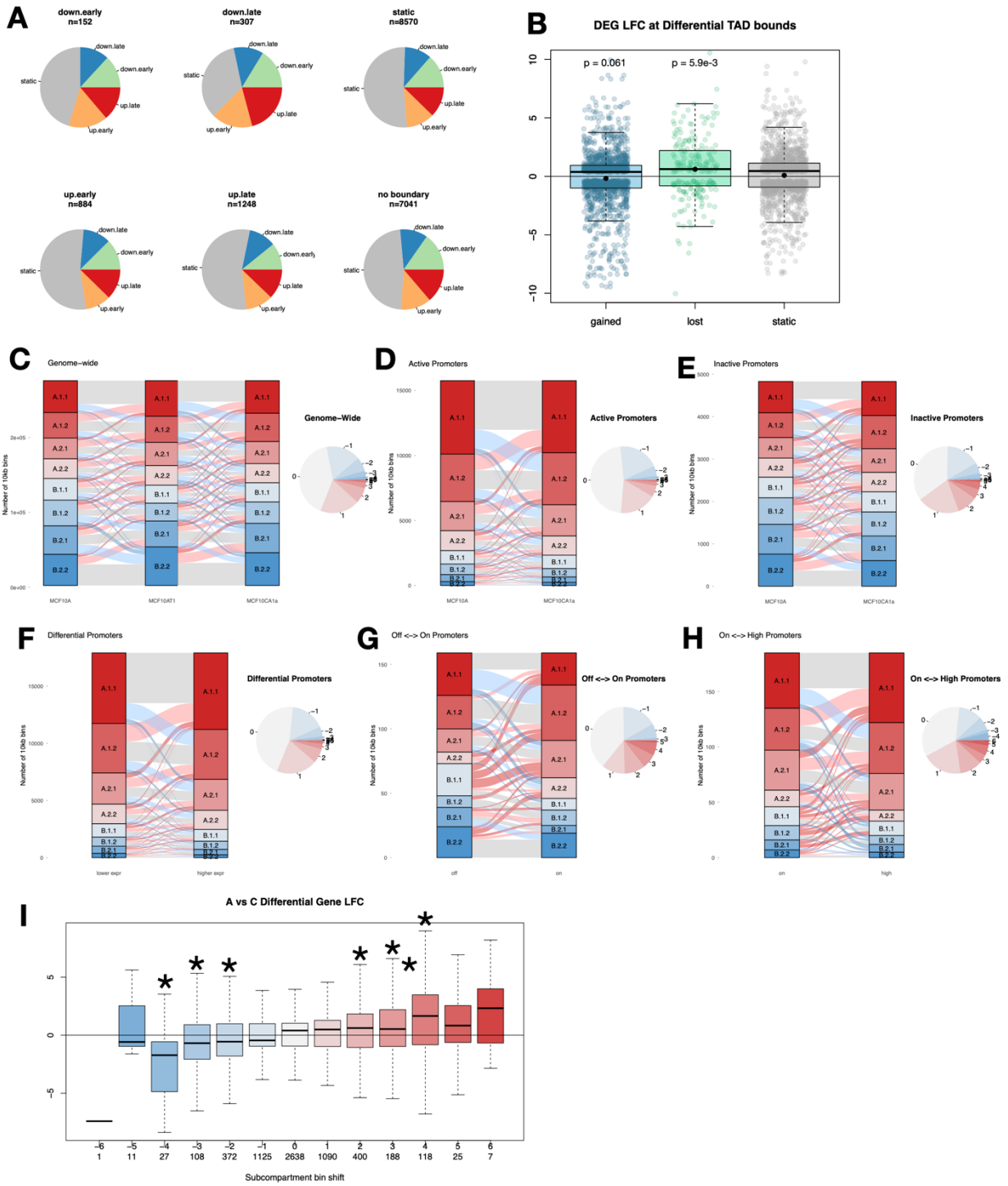


Figure S9. Differential TAD boundaries and subcompartment shifts have a subtle relationship to gene expression.

(A) Pie charts showing the differential status of genes within 50kb of TAD boundaries of various differential clusters. **(B)** Boxplots of the log2 fold-change of genes at the boundaries of TAD boundaries that are either weakened (blue), strengthened (green), or static (grey). P-values represent Wilcoxon rank sum tests comparing the mean of each set to the static set. **(C)** Alluvial plots showing the transition of subcompartments from MCF10A to MCF10AT1,

and from MCF10AT1 to MCF10CA1a. Grey ribbons indicate bins that have the same subcompartments between cell types, while red ribbons shift more A-like and blue ribbons shift more B-like. Darker ribbons shift by more than one subcompartment. A pie chart to the right summarizes the proportion of shifts based on the number of subcompartments they change by. **(D-E)** Alluvial plot showing the difference in subcompartments for the promoters of (D) actively expressed or (E) non-expressed genes that have similar gene expression in MCF10A and MCF10CA1a. Plots and pie charts are colored as in (C). **(F-H)** Alluvial plots showing the difference in subcompartments for the promoters of genes that are differentially expressed between any two cell types (F), change from on to off (G), or change from on to high (H). For each figure, the left bar shows the subcompartments in the cell line with lower expression and the right bar is from the cell line with higher expression. The pie charts summarize the shifts as in (C). **(I)** A boxplot indicating the distribution of gene log2 fold-change between MCF10A and MCF10CA1a for genes with promoters in bins that shift by various numbers of subcompartments. Negative numbers (blue) indicate bins that shift to more B-like subcompartments, while positive numbers (red) indicate shifts to more A-like subcompartments. The numbers below indicate the number of bins that fit into each category. Stars indicate significant differences in mean gene log2 fold-change compared to genes that do not change subcompartments between the two cell types (grey), as indicated by a p-value of 0.01 or less (Wilcoxon rank sum test).

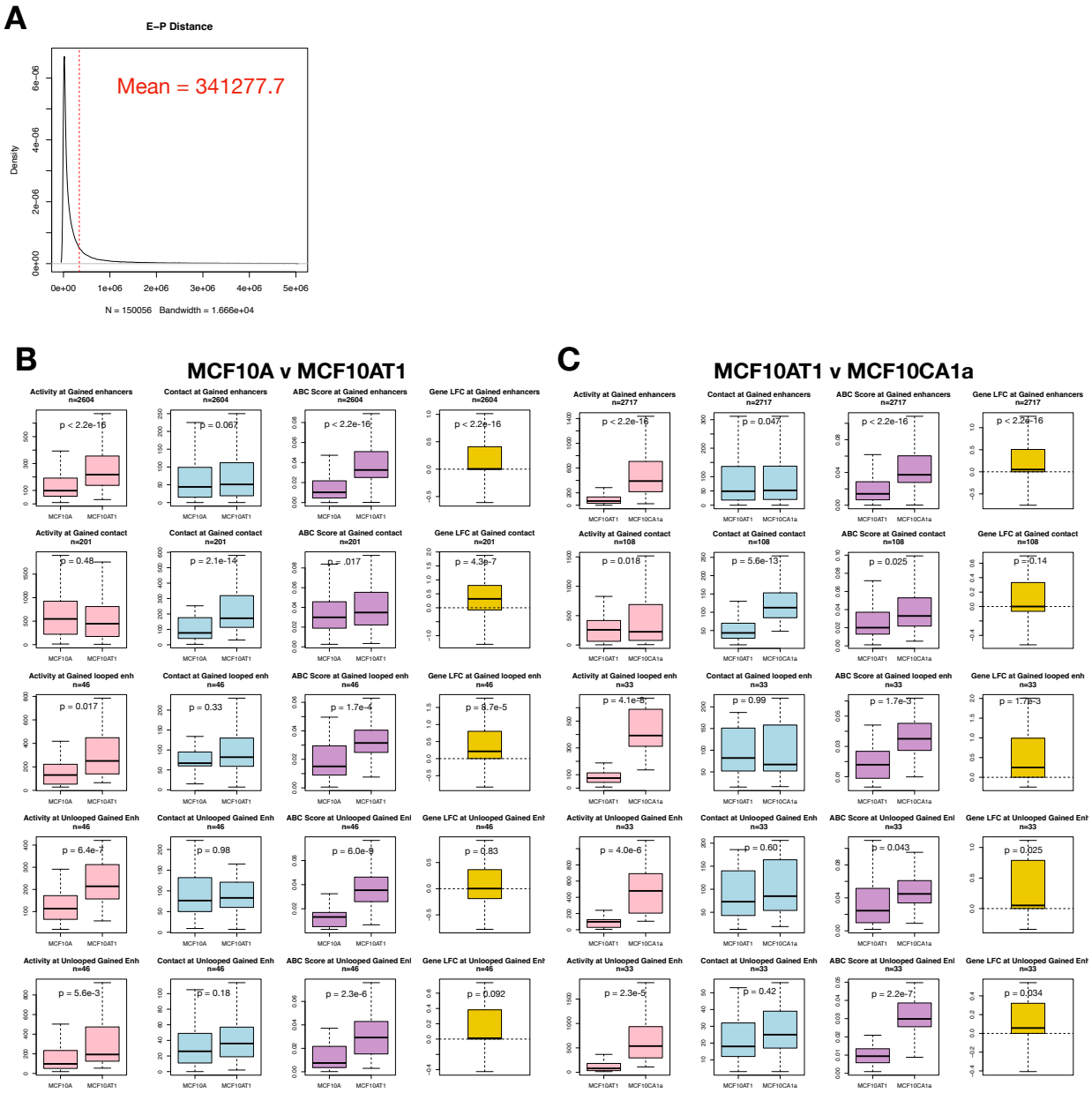


Figure S10. Enhancer-promoter pair details.

(A) Distribution of enhancer-promoter distances as predicted by the activity-by-contact (ABC) model. **(B)** Boxplots of distal enhancer H3K27ac (pink), enhancer-promoter contact (blue), and ABC score (purple), as well as gene log2 fold-change (yellow) for enhancer promoter pairs that feature (top to bottom) differential H3K27ac among enhancers, differential enhancer-promoter contact frequency, differential H3K27ac for looped enhancer-promoter pairs, contact-matched non-looped enhancer-promoter pairs, and distance-matched non-looped enhancer-promoter pairs. Comparison shows changes between MCF10A and MCF10AT1. **(C)** Same as (B) but showing comparisons between MCF10AT1 and MCF10Ca1a.

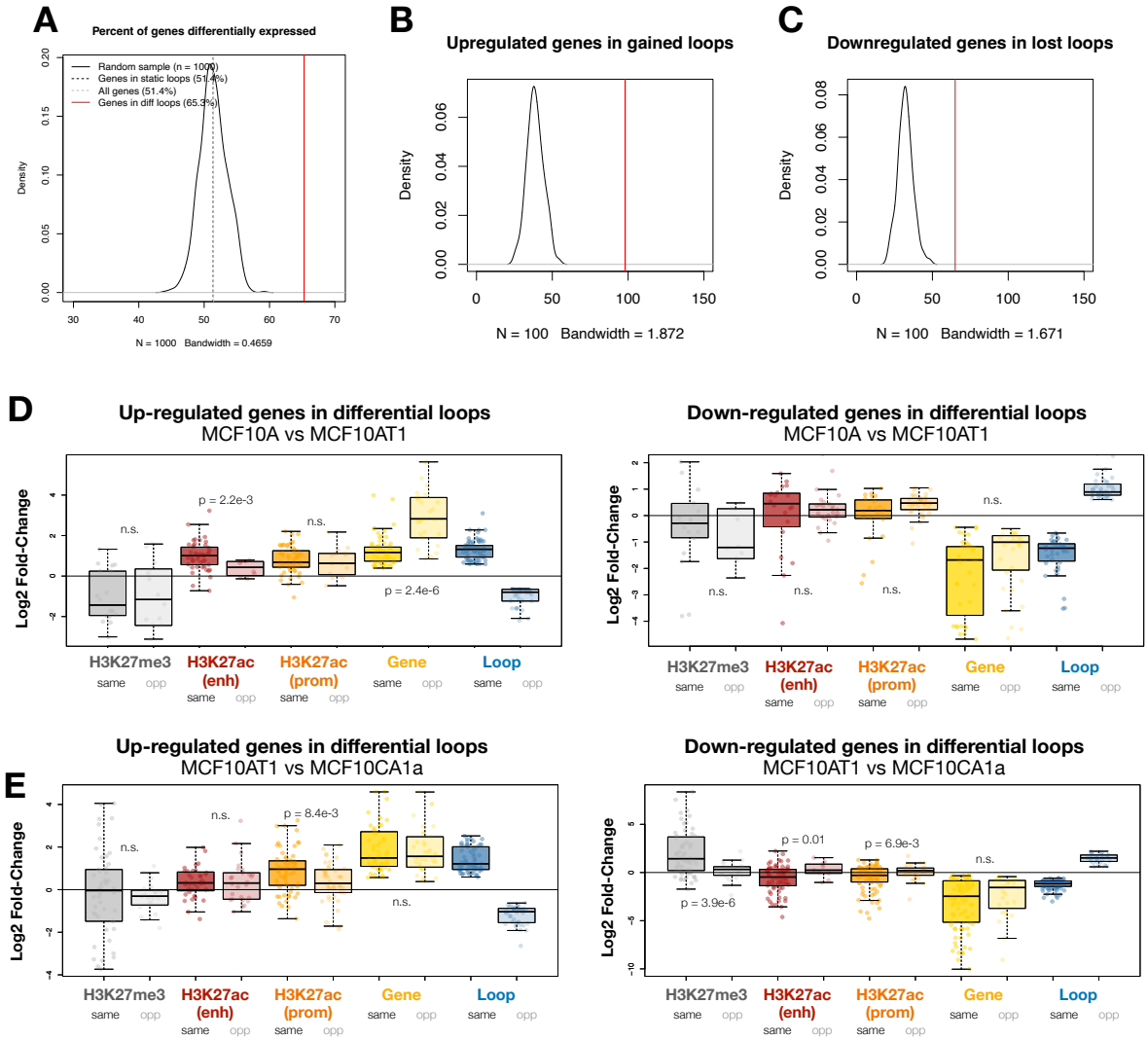


Figure S11. Differential genes within differential loops.
(A) The number of differential expressed genes among all genes (dashed light grey line), genes overlapping static loop anchors (dashed dark grey line), genes overlapping differential loop anchors (red line), and a permutation test of a random sampling of a similar number of genes (black line). **(B)** Permutation test results (n=1,000) for the number of up-regulated genes that overlap with strengthened/gained loops (red) compared to a random sampling (black). **(C)** Permutation test results (n=1,000) for the number of down-regulated genes that overlap with weakened/lost loops (red) compared to a random sampling (black). Boxplots in **(D)** and **(E)** represent log2 fold-change between MCF10A and MCF10AT1 (D) or MCF10AT1 and MCF10CA1a (E) of distal H3K27me3 (grey), distal H3K27ac (red), promoter H3K27ac (orange), gene expression (yellow), and loops (blue) among upregulated or downregulated genes that overlap with gained loops (darker colors) or lost loops (lighter colors. Boxplots are defined as in (A). P-values represent T-tests comparing the mean values of the features at loops that change in the same and those that change in opposite directions from the differential genes at their anchors. Non-significant (n.s.) p-values are any p- values above 0.01.

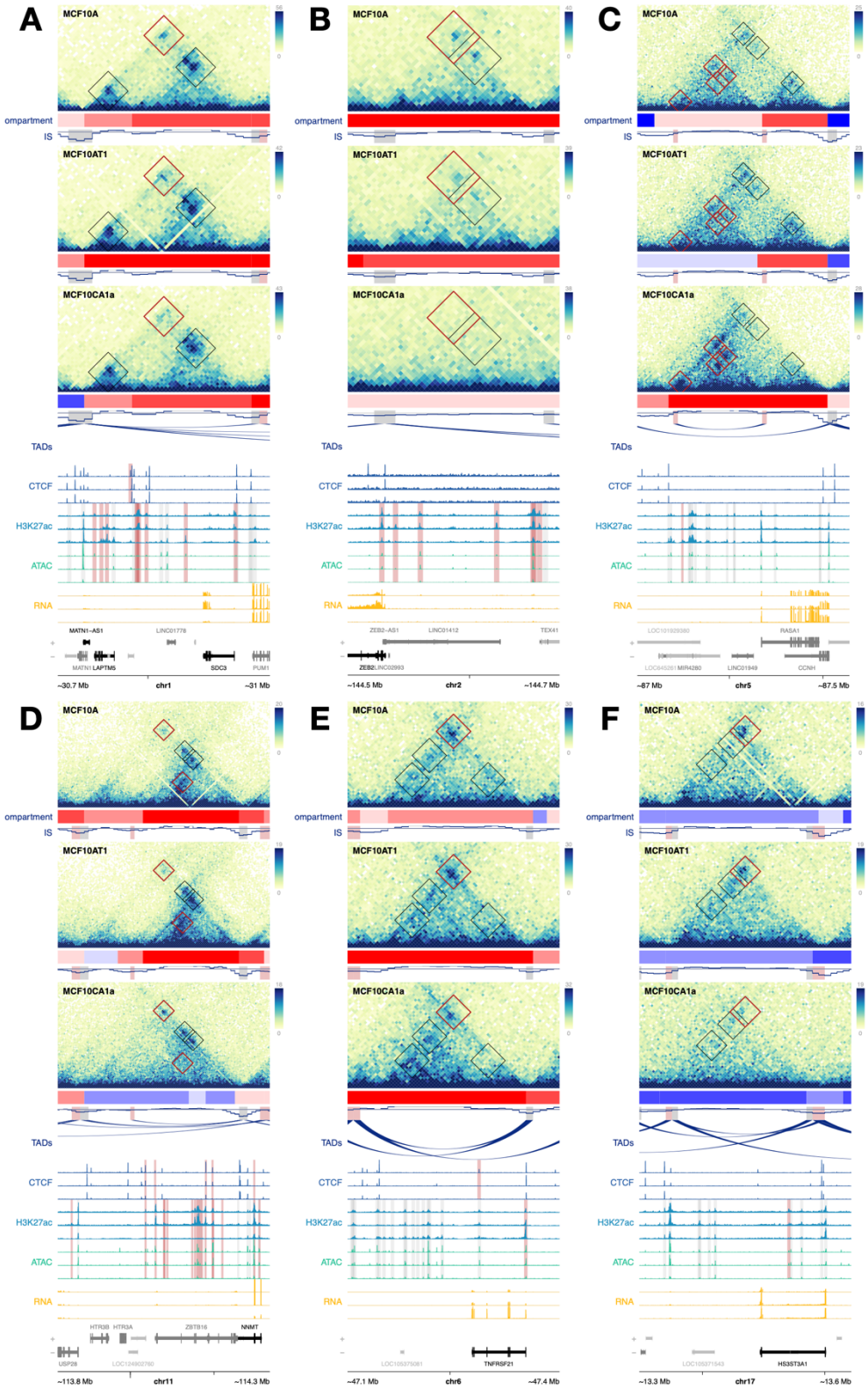


Figure S12. Examples of differentially expressed genes at differential loop anchors.

(A-C) Examples of genes that are differentially expressed in MCF10 progression and overlap with anchors of differential genes that change in the same direction. Black boxes show loop annotations, while red boxes indicate differential loops. Red compartment tracks indicate A compartments, while blue indicates B compartments. In CTCF signal tracks, red highlights indicate differential CTCF peaks. In H3K27ac and ATAC-Seq signal tracks, red highlights indicate differential enhancers as determined by changes in H3K27ac. Genes highlighted in black are differentially expressed. **(D-F)** Examples of genes that are differentially regulated in MCF10 progression and overlap with anchors of differential genes that change in the opposite direction. Plots are annotated as in (A-C).